The Journal of Economic Perspectives

A journal of the American Economic Association

Fall 2015

The Journal of
Economic Perspectives

A journal of the American Economic Association

Editorial offices:
Journal of Economic Perspectives
American Economic Association Publications
2403 Sidney St., #260
Pittsburgh, PA 15203
email: jep@jepjournal.org

The *Journal of Economic Perspectives* gratefully acknowledges the support of Macalester College.
Registered in the US Patent and Trademark Office (®).

Composed by American Economic Association Publications, Pittsburgh, Pennsylvania, USA.
Printed by R. R. Donnelley Company, Jefferson City, Missouri, 65109, USA.

THE JOURNAL OF ECONOMIC PERSPECTIVES (ISSN 0895-3309), Fall 2015, Vol. 29, No. 4. The *JEP* is published quarterly (February, May, August, November) by the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203-2418. Annual dues for regular membership are $20.00, $30.00, or $40.00 depending on income; for an additional $15.00, you can receive this journal in print. E-reader versions are free. For details and further information on the AEA go to https://www.aeaweb.org/. Periodicals postage paid at Nashville, TN, and at additional mailing offices.
POSTMASTER: Send address changes to the *Journal of Economic Perspectives*, 2014 Broadway, Suite 305, Nashville, TN 37203. Printed in the U.S.A.

The Journal of
Economic Perspectives

Contents

Volume 29 • Number 4 • Fall 2015

Statement of Purpose

The *Journal of Economic Perspectives* attempts to fill a gap between the general interest press and most other academic economics journals. The journal aims to publish articles that will serve several goals: to synthesize and integrate lessons learned from active lines of economic research; to provide economic analysis of public policy issues; to encourage cross-fertilization of ideas among the fields of economics; to offer readers an accessible source for state-of-the-art economic thinking; to suggest directions for future research; to provide insights and readings for classroom use; and to address issues relating to the economics profession. Articles appearing in the journal are normally solicited by the editors and associate editors. Proposals for topics and authors should be directed to the journal office, at the address inside the front cover.

Policy on Data Availability

It is the policy of the *Journal of Economic Perspectives* to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Details of the computations sufficient to permit replication must be provided. The Editor should be notified at the time of submission if the data used in a paper are proprietary or if, for some other reason, the above requirements cannot be met.

Policy on Disclosure

Authors of articles appearing in the *Journal of Economic Perspectives* are expected to disclose any potential conflicts of interest that may arise from their consulting activities, financial interests, or other nonacademic activities.

Journal of Economic Perspectives—Volume 29, Number 4—Fall 2015—Pages 3–8

On the Verges of Overconfidence[†]

Ulrike Malmendier and Timothy Taylor

Economists have been concerned about issues of overconfidence at least since Adam Smith (1776, Book I, Chapter X), who wrote in *The Wealth of Nations*: "The over-weening conceit which the greater part of men have of their own abilities, is an ancient evil remarked by the philosophers and moralists of all ages." Titans of modern economics have had similar reactions to the "ancient evil." Daniel Kahneman recently told an interviewer that if he had a magic wand that could eliminate one human bias, he would do away with overconfidence. As Shariatmadari (2015) reports: "Not even he [Kahneman] believes that the various flaws that bedevil decision-making can be successfully corrected. The most damaging of these is overconfidence: the kind of optimism that leads governments to believe that wars are quickly winnable and capital projects will come in on budget despite statistics predicting exactly the opposite." Kahneman argues that overconfidence "is built so deeply into the structure of the mind that you couldn't change it without changing many other things."

Evidence concerning the prevalence of overconfidence is widespread and robust. Some of the results have even become fairly well-known in popular culture, like the findings that most drivers believe they are safer than a typical driver, or that

■ *Ulrike Malmendier is Edward J. and Mollie Arnold Professor of Finance at the Haas School of Business and Professor of Economics at the Department of Economics, University of California, Berkeley, California. Malmendier is also a Co-editor of the* Journal of Economic Perspectives. *Timothy Taylor is Managing Editor,* Journal of Economic Perspectives, *Macalester College, St. Paul, Minnesota. Their email addresses are ulrike@econ.berkeley.edu and taylort@macalester.edu.*

[†] For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.29.4.3

doi=10.1257/jep.29.4.3

the unskilled tend to overestimate their abilities. The finding about driver overconfidence stems from a Svenson (1981) study, a lab experiment using undergraduate students as subjects, which found that 83 percent of American subjects believed that they were in the top 30 percent in terms of driving safety. The finding about overestimation of ability comes from a Kruger and Dunning (1999) study, which reports: "[P]articipants scoring in the bottom quartile on tests of humor, grammar, and logic grossly overestimated their test performance and ability. Although their test scores put them in the 12th percentile, they estimated themselves to be in the 62nd." Overconfidence on both sides of a conflict may be linked to the willingness to fight a war (Wrangham 1999; Johnson 2004) or to the conditions that lead a strike to occur (Neale and Bazerman 1985). Garrison Keillor, long-time host of the popular radio variety show *A Prairie Home Companion*, each week includes a story about the people and events in the mythical town of Lake Wobegon. The monologue always ends this way: "Well, that's the news from Lake Wobegon, where all the women are strong, all the men are good looking, and all the children are above average." A "Lake Wobegon effect" and references to everyone being above average have become something of a *terminus technicus* in discussions of testing (for example, see Cannell 1988; Hartocollis 1999).

In social psychology and organizational behavior, overconfidence and other self-enhancement biases have been studied for decades (for example, Miller and Ross 1975; Langer 1975; Weinstein 1980). Overall, psychologists view above-average effects and comparative optimism as "perhaps the two most robust and widely replicated phenomena from the literature on social comparative judgments" (Chambers and Windschitl 2004, p. 1). Two prominent behavioral economists, Werner DeBondt and Richard Thaler (1995, p. 389) have stated: "Perhaps the most robust finding in the psychology of judgement is that people are overconfident."

Economic researchers have started to explore the implications of this bias. In economics, the key question is how overconfidence affects decision makers: consumers, investors, chief executive officers, and others. Moreover, economists study these individual motivations in the context of a market, where interactions can lead to unexpected results. For example, we know that the interaction of self-interested individuals in a market can, as if led by an invisible hand, have desirable incentive effects. But what about a market consisting of economic agents who are both self-interested and overconfident? For example, if one or both parties are overconfident, they may underestimate the likelihood or the costs of not reaching an agreement, and thus make it harder for a market to function.

This symposium provides several examples of overconfidence in certain economic contexts. Michael Grubb looks at "Overconfident Consumers in the Marketplace." Ulrike Malmendier and Geoffrey Tate consider "Behavioral CEOs: The Role of Managerial Overconfidence." Kent Daniel and David Hirshleifer discuss "Overconfident Investors, Predictable Returns, and Excessive Trading." A number of insights and lessons emerge for our understanding of markets, public policy, and welfare. How do firms take advantage of consumer overconfidence? Might government attempts to rule out such practices—say, by disallowing certain types of

offers by credit card companies that lead overconfident consumers along a welfare-decreasing path—end up providing benefits to some consumers but imposing costs on others? How can researchers use the timing of when chief executive officers choose to exercise stock options as a way of measuring overconfidence? How are empirical measures of CEO overconfidence related to investment and the capital structure of firms? Can overconfidence among at least some investors help to explain prominent anomalies in stock markets like high levels of trading volume and certain seemingly predictable patterns in stock market returns?

In addition, these essays on the economics of overconfidence offer a window into some key recurring issues in the current generation of studies of behavioral economics. One issue is that terminology about heuristics and biases borrowed from the psychology literature can turn out to be rather broad, at least initially. To pin down testable implications, we need to delineate a theory that specifies more precisely how such biased behavior differs from the traditional neoclassical approach. For example, overconfidence can manifest itself in various ways. It can involve *overoptimism*, which involves overestimating positive outcomes, either in terms of magnitude or in terms of frequency, and including outcomes that are not under one's own control ("number of gold medals the United States will win at the Olympics"). When the outcomes are (perceived to be) under one's control, such as one's own abilities or prospects, the term *overconfidence* is often used directly. This meaning is closely related to the "better-than-average" effect, in which individuals tend to overestimate their acumen relative to the average (Larwood and Whittaker 1977; Alicke 1985; Taylor and Brown 1988). Alternatively, overconfidence can take the form of *overprecision*, in which economic actors may hold accurate economic beliefs on average but underestimate the variance of possible outcomes—and in this way underestimate the risks they are actually facing. A number of researchers have demonstrated that people tend to have a strong belief in the effectiveness of their own intuition as opposed to statistical rules, even when there is strong evidence to the contrary (Dawes, Faust, and Meehl 1989; Grove, Zald, Lebow, Snitz, and Nelson 2000). Clearly, differing aspects of overconfidence like overoptimism and overprecision require rather different theoretical approaches—depending on whether the issue involves beliefs about the central tendency of a distribution, or beliefs about the underlying distribution of outcomes around that central tendency—and generate different empirical predictions.

The discussion of behavioral biases inevitably raises the question of how the bias can persist over time. Sure, a consumer or an investor or a top executive might be overconfident at certain times. But under the incentives created by market experience, shouldn't overconfidence be learned away? This question of why behavioral biases can persist has (at least) three interrelated answers.

First, learning may be more difficult in some settings than in others. For example, people tend to be more overconfident about their ability to perform tasks for which feedback is infrequent and inconclusive. The "hard/easy effect" postulates that people tend to be more overconfident about difficult problems and even underconfident about easy ones (Lichtenstein and Fischhoff 1977). As it is for the case of any belief-based bias, overconfidence can persist as long as the individual

can avoid learning about one variable by instead updating beliefs about another stochastic variable, for instance: "The reason why I am not going to the gym isn't that I underestimated my self-control problems; it's just that I am busier than I thought I would be" (as in DellaVigna and Malmendier 2006).

Second, different behavioral biases may reinforce each other. For example, overconfidence might persist in part because of hindsight bias, which is the tendency to believe with hindsight that something that has already happened was highly predictable, even when there would have been little basis beforehand for such a belief (Roese and Vohs 2012). It may also be strenthened by self-attribution bias, which arises when people who experience successful outcomes regard it as a result of their own skill, but blame unsuccessful outcomes on bad luck (Shefrin 2000, p. 101). The theory of cognitive dissonance, going back to Festinger (1957), argues that people experience stress when confronted with information that calls existing beliefs into question, and thus try to avoid facing such belief-changing information. Hence, cognitive dissonance also allows overconfident beliefs to persist. Other biases can be thought of as aspects of overconfidence, or ways in which overconfidence can manifest itself. For example, the illusion of control is a phenomenon experienced by most people who play a game of chance, in which they feel as if they can control the next roll of the dice or the next card they will receive (Langer 1975; Presson and Benassi 1996).

Third, a common theme across behavioral biases is that they may be a useful rule of thumb in some contexts and then spill over to other contexts where they are less useful—hence, the literature in this area sometimes refers to a dual perspective of "heuristics and biases" (as in Gigerenzer 1991). For example, overconfidence may help in bargaining situations or as argued by Schelling (1960 [1980], p. 23) in the context of international conflict and wars. Those who are overconfident about their own perceptions or abilities, or perceive risks as smaller than they actually are, may also be more likely to veer away from usual or accepted group perspectives, which can foster innovation (Bernanrdo and Welch 2001). In fact, even if overconfidence is not individually optimal it might be socially optimal in certain contexts, as Bernardo and Welch (2001) illustrate in the context of entrepreneurship.

An understanding of overconfidence thus matters both for its direct effects and its market implications in many arenas: consumer behavior and firm responses; investor behavior and financial markets; chief executive officer behavior and the market for corporate control; and even labor strikes and declarations of war. In addition, overconfidence is important throughout the study of behavioral economics because it helps to answer the question of why we all tend to hold so tightly to our own views, even when the rational part of our brains has been quite well-informed—whether through real-world experience or academic studies—that our judgment is diverging from textbook rationality.

■ *This introduction draws in part upon earlier drafts of the papers prepared for this symposium.*

References

Alicke, Mark D. 1985. "Global Self-Evaluation as Determined by the Desirability and Controllability of Trait Adjectives." *Journal of Personality and Social Psychology* 49(6): 1621–30.

Bernardo, Antonio E., and Ivo Welch. 2001. "On the Evolution of Overconfidence and Entrepreneurs." *Journal of Economics & Management Strategy* 10(3): 301–330.

Cannell, John Jacob. 1988. "Nationally Normed Elementary Achievement Testing in America's Public Schools: How All 50 States Are Above the National Average." *Educational Measurement: Issues and Practice* 7(2): 5–9.

Chambers, John R., and Paul D. Windschitl. 2004. "Biases in Social Comparative Judgments: The Role of Nonmotivated Factors in Above-Average and Comparative-Optimism Effects." *Psychological Bulletin* 130(5): 813–38.

Dawes, Robyn M., David Faust, and Paul E. Meehl. 1989. "Clinical versus Actuarial Judgment." *Science,* March 13, 243(4899): 1668–74.

DeBondt, Werner F. M., and Richard Thaler. 1995. "Financial Decision-making in Markets and Firms: A Behavioral Perspective." In *Finance: Handbooks in Operations Research and Management Science,* vol. 9, edited by R. Jarrow, V. Maksimovic, and W. T. Ziemba, 385–410. Amsterdam: North Holland.

DellaVigna, Stefano, and Ulrike Malmendier. 2006. "Paying Not to Go to the Gym." *American Economic Review* 96(3): 694–719.

Einhorn, Hillel J. 1980. "Overconfidence in Judgment." In *New Directions for Methodology of Social and Behavioral Science,* vol. 4, 1–16. San Francisco: Jossey-Bass.

Festinger, Leon. 1957. *A Theory of Cognitive Dissonance.* California: Stanford University Press.

Gigerenzer, Gerd. 1991. "How to Make Cognitive Illusions Disappear: Beyond 'Heuristics and Biases.'" *European Review of Social Psychology* 2(1): 83–115.

Grove, William M., David H. Zald, Boyd S. Lebow, Beth E. Snitz, and Chad Nelson. 2000. "Clinical versus Mechanical Prediction: A Meta-Analysis." *Psychological Assessment* 12(1): 19–30.

Hartocollis, Anemona. 1999. "The Nation: Grade Inflation; New Math: No One Is Below Average." *New York Times,* June 20.

Johnson, Dominic D. P. 2004. *Overconfidence and War: The Havoc and Glory of Positive Illusions.* Harvard University Press.

Kruger, Justin, and David Dunning. 1999. "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments." *Journal of Personality and Social Psychology* 77(6): 1121–34.

Langer, Ellen J. 1975. "The Illusion of Control." *Journal of Personality and Social Psychology* 32(2): 311–28.

Larwood, Laurie, and William Whittaker. 1977. "Managerial Myopia: Self-Serving Biases in Organizational Planning." *Journal of Applied Psychology* 62(2): 194–98.

Lichtenstein, Sarah, and Baruch Fischhoff. 1977. "Do Those Who Know More Also Know More about How Much They Know?" *Organizational Behavior and Human Performance* 20(2): 159–83.

Miller, Dale T., and Michael Ross. 1975. "Self-Serving Biases in the Attribution of Causality: Fact or Fiction?" *Psychological Bulletin* 82(2): 213–25.

Neale, Margaret A., and Max H. Bazerman. 1985. "The Effects of Framing and Negotiator Overconfidence on Bargaining Behaviors and Outcomes." *Academy of Management Journal* 28(1): 34–49.

Presson, Paul K., and Victor A. Benassi. 1996. "Locus of Control Orientation and Depressive Symptomatology: A Meta-Analysis." *Journal of Social Behavior and Personality* 11(1): 201–212.

Roese, Neal J., and Kathleen D. Vohs. 2012. "Hindsight Bias." *Perspectives on Psychological Science* 7(5): 411–26.

Schelling, Thomas C. 1960 [1980]. *The Strategy of Conflict.* Harvard University Press.

Shariatmadari, David. 2015. "Daniel Kahneman: 'What Would I Eliminate If I Had a Magic Wand? Overconfidence.'" *The Guardian.* July 18. http://www.theguardian.com/books/2015/jul/18/daniel-kahneman-books-interview.

Shefrin, Hirsh. 2000. *Beyond Greed and Fear: Understanding Behavioral Finance and the Psychology of Investing.* Oxford University Press.

Smith, Adam. 1776 [1904]. *The Wealth of Nations.* Library of Economics and Liberty. http://www.econlib.org/library/Smith/smWN.html.

Svenson, Ola. 1981. "Are We All Less Risky and More Skillful than Our Fellow Drivers?" *Acta Psychologica* 47(2): 143–48.

Taylor, Shelley E., and Jonathon D. Brown. 1988. "Illusion and Well-Being: A Social Psychological Perspective on Mental Health." *Psychological Bulletin* 103(2): 193–210.

Weinstein, Neil D. 1980. "Unrealistic Optimism about Future Life Events." *Journal of Personality and Social Psychology* 39(5): 806–820.

Wrangham, Richard. 1999. "Is Military Incompetence Adaptive?" *Evolution and Human Behavior* 20(1): 3–17.

Journal of Economic Perspectives—Volume 29, Number 4—Fall 2015—Pages 9–36

Overconfident Consumers in the Marketplace[†]

Michael D. Grubb

When consumers sign contracts, expectations about future usage of the product or service matter. For instance, the value provided by car insurance depends on how likely a consumer believes she is to file a claim; the value provided by a gym membership depends on how often a consumer anticipates going to the gym; and the value provided by a cellular phone contract depends on how many gigabytes of data a consumer anticipates using. The standard modeling paradigm makes the expedient assumption that consumers have rational expectations. Imposing rational expectations drastically simplifies models and eliminates the need to directly measure beliefs as they coincide with the distribution of observed outcomes. Yet a large literature shows that consumer beliefs often deviate substantially from rational expectations in systematic ways. This has important consequences for contract design, firm profits, consumer welfare, and public policy.

The term overconfidence is used broadly in the psychology literature, referring to both *overoptimism* and *overprecision*. Overoptimistic individuals overestimate their own abilities or prospects, either in absolute terms or in comparison to others. In contrast, overprecise individuals place overly narrow confidence intervals around forecasts, thereby underestimating uncertainty. These biases can lead consumers to misforecast their future product usage, or to overestimate their abilities to navigate contract terms. In consequence, consumer overconfidence causes consumers

■ *Michael D. Grubb is Assistant Professor of Economics, Boston College, Chestnut Hill, Massachusetts. His email address is michael.grubb@bc.edu.*

[†] For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.29.4.9 doi=10.1257/jep.29.4.9

to systematically misweight different dimensions of product quality and price. Poor choices based on biased estimates of a product's expected costs or benefits are the result.

For instance, overoptimism about self-control is a leading explanation for why individuals overpay for gym memberships that they underutilize (DellaVigna and Malmendier 2006). Overoptimism leads consumers to overestimate gym attendance and hence to overweight the membership benefit of avoiding per-visit gym fees. Similarly, overprecision is a leading explanation for why individuals systematically choose the wrong calling plans, racking up large overage charges for exceeding usage allowances in the process (Grubb 2009; Grubb and Osborne 2015). Overprecision leads consumers to underestimate the variance of future calling, and hence to underweight the cost of calling plans under both low and high usage.

While overconfidence necessarily leads to suboptimal choices, this paper addresses three additional questions about how consumer overconfidence alters market outcomes. First, what will firms do to exploit consumer overconfidence? Firms in many sectors seek to exploit consumer overconfidence by introducing complicated pricing features—and these pricing features are robust to competition. Second, what are the equilibrium welfare consequences of consumer overconfidence for consumers, firms, and society? Not unexpectedly, the welfare consequences for firms and consumers depend on whether overconfident consumers over- or undervalue the offered contracts, as well as on market structure and the nature of consumer heterogeneity. It turns out that competition may fail to protect overconfident consumers from their own mistakes, and that sophisticated consumers often exacerbate the resulting harm to overconfident consumers. Third, what are the implications of consumer overconfidence for public policy? The analysis suggests that policymakers should try to anticipate firms' equilibrium responses to consumer protection measures. Banning contractual terms that exploit overconfidence need not help consumers if firms respond by raising up-front payments an offsetting amount. Yet the analysis also shows that two important market statistics can help policymakers distinguish markets in which regulation may have limited benefit (for example, cellular services, Grubb and Osborne 2015) from those in which regulation may substantially help overconfident consumers (for example, credit cards, Agarwal, Chomsisengphet, Mahoney, and Stroebel 2015).

Overconfidence is just one of several biases that lead consumers to systematically misweight the various dimensions of product price or quality when computing the expected utility of a purchase. For instance, projection bias leads consumers to overweight current preferences, overvaluing unhealthy snacks when hungry (Loewenstein, O'Donoghue, and Rabin 2003) and winter coats when cold (Conlin, O'Donoghue, and Vogelsang 2007). Similarly, individuals with limited attention underweight nonsalient attributes such as eBay shipping fees (Hossain and Morgan 2006; Einav, Kuchler, Levin, and Sundaresan 2015), commodity taxes not included in posted prices (Chetty, Looney, and Kroft 2009), and hidden add-on fees (Gabaix and Laibson 2006). While I focus on overconfidence, and in particular on two important mistakes made by overconfident consumers—misforecasting

consumption and overestimating ability to navigate contract terms—the lessons in the paper apply more broadly to all such biases. In contrast, behavioral biases that limit search, generate inertia, or confuse choices in a manner uncorrelated across consumers have starkly different market implications that I discuss in Grubb (2015b).

Are Consumers Overconfident?

The evidence for overconfidence—both in the form of overoptimism and overprecision—is rooted in more than half a century of psychology research (Lichtenstein, Fischhoff, and Phillips 1982). While several strands of early work on overoptimism must be partially discounted in light of recent criticism (for instance, by Benoît and Dubra 2011), the overall case for overconfidence remains robust, as described as described in Malmendier and Taylor's introduction to this symposium. Here, I focus more narrowly on evidence that consumer overconfidence is important in the marketplace. Such evidence is particularly important for overoptimism, which is known to be context dependent (Weinstein 1980).

Evidence for consumer overconfidence comes from three sources. First, there are experiments in market-relevant settings. For instance, Silk's (2004) experiments suggest that individuals are overoptimistic about the likelihood of redeeming mail-in rebates. Ericson's (2011) related experiment identifies a plausible reason. His experiment asked students to decide between a small payment that would be received automatically and a larger payment that needed to be claimed in six months. Their choices imply that subjects anticipated claiming the payment in six months' time with an average probability of 76 percent, but only 53 percent actually claimed the payment. Ericson (2011) concludes that individuals are overoptimistic about their *prospective memory*, or their ability to remember to take planned actions.

Second, field data on consumer choices provides evidence of consumer overconfidence. Consumer beliefs can be inferred from contract choices (or elicited by survey) and compared to later usage to identify bias. For example, in a study of a New England health club, DellaVigna and Malmendier (2006) show that users who chose a monthly membership could have saved an average of more than 40 percent by foregoing a membership and paying per-visit. This finding is consistent with overoptimism about gym attendance due to overoptimism about self-control. In a study of consumer responses to credit card offers, Ausubel (1999) finds that "consumers are at least three times as responsive to changes in the introductory interest rate as compared to dollar-equivalent changes in the post-introductory interest rate." This finding is consistent with overoptimism about the likelihood of repaying or refinancing debt in time to avoid paying post-introductory rates, perhaps due to overoptimism about self-control (Shui and Ausubel 2005; Heidhues and Kőszegi 2010). In a study of cellular service demand, my coauthor and I use a structural model to estimate that cellular phone consumers are overprecise; specifically, they

underestimate the noise in their forecasts of their own future demand for calls by 62 percent (Grubb and Osborne 2015).

Third, in addition to direct evidence from consumers' contract choice mistakes, field studies can also provide indirect evidence of consumer overconfidence from firm pricing strategies. For example, in Grubb (2009), I show that given observed heterogeneity in calling patterns, cellular carriers' service pricing can be explained as a response to consumer overprecision but not as optimal price discrimination between unbiased consumers. In a similar vein, Gottleib and Smetters (2014) convincingly argue that the contractual terms of whole life insurance policies can only be profit maximizing if customers underweight the risk of allowing their policies to lapse—consistent with overoptimism about liquidity shocks.

For economists, a common question is "Doesn't overconfidence go away with learning?" On the one hand, learning can mitigate overconfidence with appropriate feedback (Bolger and Önkal-Atay 2004). On the other hand, Gabaix and Laibson (2006) show that competition need not give firms an incentive to educate or de-bias consumers, and lab experiments show that outcome feedback of the sort consumers likely receive in practice is often ineffective (Subbotin 1996). Moreover, field studies of consumer choice confirm that learning is no panacea for avoiding past mistakes. Choices may improve slowly (Grubb and Osborne 2015), lessons may be forgotten (Agarwal, Driscoll, Gabaix, and Laibson 2013), and after peaking in financial sophistication in middle age, individuals succumb to contractual traps at an increasing rate as they grow old (Agarwal, Driscoll, Gabaix, and Laibson 2009).

Firms Complicate Contracts to Exploit Overconfidence

In the standard common-prior framework, firms design contracts with two goals in mind: to create surplus from trade and to extract a share as profit. In other words, firms aim to bake a large pie and keep a large share for themselves. Relaxing the common-prior assumption to allow for consumer overconfidence adds a new consideration. Overconfident consumers may overvalue or undervalue offered contracts relative to the true value they deliver. Firms naturally prefer consumers to overvalue contracts as much as possible, because that allows firms both to charge more for contracts and to sell more contracts. Hence, firms selling to overconfident consumers design contracts with an additional goal in mind: either to maximize the amount by which consumers overvalue contracts or to minimize the amount by which consumers undervalue contracts, depending on the situation. In other words, firms aim to bake a large pie, keep a large share for themselves, and make the piece served to customers appear larger than it is.

To understand the implications of consumer overconfidence for firm pricing practices, I focus on two important ways in which overconfidence may cause consumers to misvalue offered contracts. First, overconfident consumers may misforecast their future usage of services. For instance, an overconfident driver

may undervalue car insurance because he underestimates the likelihood of filing a claim (Sandroni and Squintanti 2007, 2013). When consumers misforecast future usage, firms have an incentive to distort marginal prices and/or product quality to exploit the mistake (for example, DellaVigna and Malmendier 2004; Elaiz and Spiegler 2006, 2008; Grubb 2009; Heidhues and Kőszegi 2010). Second, overconfident consumers may be overoptimistic about their own abilities of self-control, prospective memory, or attention. Such consumers overvalue contracts because they overestimate their abilities to navigate contract terms to take advantage of contract benefits or avoid contract costs. For instance, an overconfident consumer may overestimate the likelihood of remembering to mail in a rebate. Firms have an incentive to complicate their contracts with precisely those terms that consumers overestimate their own abilities to navigate (for example, DellaVigna and Malmendier 2004; Gilpatric 2009; Holman and Zaidi 2010; Grubb 2015a).

When Overconfidence Involves Misforecasting Usage

Let us begin by considering implications for pricing when overconfident consumers misforecast their usage. A consumer who underestimates future usage also underestimates the chance of paying marginal fees, and hence underestimates the cost of any increases in marginal prices. When this occurs, firms have an incentive to inflate marginal prices above marginal cost to increase contract overvaluation or mitigate contract undervaluation.

For example, consider a service contract that specifies an up-front fixed fee plus an additional marginal price charged per unit of the service used. Suppose that an overconfident consumer will use exactly one unit of the service with probability 1, but underestimates his usage, believing that he will only use the unit with probability 1/2. Moreover, suppose for a moment that these probabilities are fixed, independent of marginal price. Consider increasing the marginal price by $2, but lowering the contract fixed fee by $1. This change raises the total cost of the contract by $1. From the consumer's perspective, however, the expected cost remains the same because the consumer underestimates the likelihood of paying the $2 increase in marginal price. Thus, the consumer overvalues the contract by an additional dollar, which translates into an additional dollar of profit for the firm.

Ignoring risk aversion or liquidity constraints, the argument in the preceding example for increasing marginal price by $2 (coupled with a $1 reduction in the fixed fee) could be repeated ad infinitum. This leads to the implausible prediction that the optimal marginal price is infinite. Ruling out such implausible predictions is in fact an important reason that the common-prior assumption is standard in economics. However, a common prior is often superfluous for ruling out infinite bets from consumer contracts. This is because contracts typically specify payments conditional on usage or other consumer behavior that is endogenous (rather than exogenous, as implausibly assumed in the example above).

For instance, suppose that the contract in question is a car lease. If the marginal price per mile were too high, lessors would simply avoid driving their cars and the contracts could not be optimal. Thus, the size of the optimal pricing

distortion is endogenously limited. Alternatively, consider the cartoon character Dogbert's offer of a product for $1,000,029 with a mail-in rebate for $1,000,000 (Adams 2003). While in principal this offer could exploit consumers who typically overestimate the likelihood of remembering to mail in rebates, such a high rebate value is self-defeating if it encourages consumers to adopt effective memory aids. These or related arguments can endogenously limit the terms of all of the exploitative contracts that I discuss below. Hence, for practical purposes we need worry no further about million dollar rebates, infinite marginal prices, and other implausibly large contract terms.

When consumers underestimate usage, the marginal prices specified by a contract serve two roles. First, marginal prices affect consumers' usage decisions and, hence, the surplus created by the sale of a contract. To maximize total surplus, firms would like to set marginal prices equal to marginal cost.[1] Second, marginal prices serve as the stakes of a speculative bet about how much the consumer will use. When the consumer uses more than expected, the firm wins the bet, receiving the marginal payments in compensation. To maximize the contract overvaluation created by this speculative bet, firms would like to set marginal price as high as possible. The optimal marginal price is chosen to balance these two considerations—it is set at the point above marginal cost where the additional gains from exploiting consumers' mistaken usage forecasts are offset by the additional costs from distorting consumers' true usage choices.

When overconfident consumers overestimate rather than underestimate future usage, the logic is similar but reversed: firms have an incentive to discount marginal prices below marginal cost. As consumers overestimate the chance of paying marginal fees, they also overestimate the value of any discount to marginal prices. Hence discounting marginal prices in this context either contributes to contract overvaluation or mitigates contract undervaluation.[2]

Will overconfidence distort marginal prices up or down? The answer depends on whether overconfident consumers under- or overestimate future usage. That depends in turn on context and the nature of overconfidence (and would be reversed were consumers underconfident rather than overconfident). For instance, a natural conjecture is that overoptimism about driving ability will lead drivers to underestimate the likelihood of filing a car-insurance claim.[3] If so, insurance companies should distort the price of filing a claim upwards, which means raising deductibles. The prediction that overconfidence increases deductibles is specific to insurance markets. In many cases, predictions about the consequences of overconfidence may be equally context-specific. However, overoptimism about self-control and overprecision each lead to more general predictions.

[1] Alternatively, when consumers have self-control problems, firms would like to set marginal price equal to the *commitment price* that induces efficient consumption (DellaVigna and Malmendier 2004).
[2] The result that marginal prices are distorted upwards when usage is underestimated and downwards when usage is overestimated appears in DellaVigna and Malmendier (2004) and Grubb (2009).
[3] Sundström (2008) and Benoît and Dubra (2011) discuss evidence for overoptimism about driving skill.

Misforecasting Usage Due to Overoptimism about Self-Control

First, consider overoptimism about self-control. DellaVigna and Malmendier (2004) define *investment goods* as those that require costly effort at the point of consumption but yield future benefits. In contrast, *leisure goods* are those that yield an immediate payoff upon consumption but are costly later. For example, a gym workout is an investment good but credit card borrowing is a leisure good. Those who are overoptimistic about their own levels of self-control will overestimate their consumption of investment goods (overestimating their gym attendance) but underestimate their consumption of leisure goods (underestimating credit card borrowing). As a result, DellaVigna and Malmendier (2004) predict that marginal prices of investment goods will be discounted below marginal cost but that marginal prices of leisure goods will be inflated above marginal cost. This prediction is consistent with the fact that many gyms do not charge per-visit fees to members, despite per-visit marginal costs of $3 or more. It is also consistent with evidence that high interest rates on credit card debt do not merely reflect the costs of default, but are substantially above marginal cost: Ausubel (1991) finds that banks are able to resell credit card debt for an average premium of 20 percent.

Misforecasting Usage Due to Overprecision

Next, consider how firms may exploit overprecision. A car lessor who exhibits overprecision may correctly forecast her median mileage but underestimate the variance of her driving needs around the median. Such a consumer will overestimate the likelihood of driving the qth mile if it is below median total mileage, but underestimate the likelihood of driving the qth mile if it is above median total mileage. Therefore, if Q is median mileage, the lease contract should price the first Q miles below marginal cost and all later miles above marginal cost. If consumers can freely dispose of miles (for instance by lending the car to a friend for a weekend trip) then the mileage fee for the first Q miles should not be reduced below zero. While the optimal contract will be fully nonlinear, a three-part tariff is a good approximation (Grubb 2009). A three-part tariff consists of: 1) a fixed fee; 2) an included allowance of units at zero marginal price; and 3) a constant marginal price for additional units. This coincides exactly with observed car leasing contracts, which typically offer an allowance of 36,000 miles with a three-year lease and charge 15 cents per mile for additional mileage. Overconfidence may explain the structure of car lease contracts as well as three-part tariffs in a variety of other settings. Table 1 gives some examples.

The literature on exploitative contracting focuses on how consumer bias affects firm pricing.[4] However, contracts often specify aspects of quality as well as price, and overconfidence should affect these terms as well. I conjecture that the preceding conclusions about marginal price distortions extend naturally to quality distortions. If a consumer overestimates the likelihood of using the qth unit, then she will

[4] Spence (1977) is an exception. See Kőszegi (2014) for a survey.

Table 1
Three-Part Tariffs

Product or service	*Consumers must forecast*	*Example contract*
Car lease	Mileage	Fixed price for 36 months and 36,000 miles plus 15 cents per additional mile (Toyota 2014).
Smart phone service	Data usage	$60/month for unlimited talk and text and 1GB of data plus $15 per additional 500MB of data (Verizon Wireless 2014).
Credit card	Loan duration	Introductory offer with an initial balance transfer: Banks charge a balance transfer fee that is independent of the loan duration. Then there are zero financing charges for the first six months but a high interest rate thereafter (Chase 2014).

Source: Author.

overvalue an increase in its quality, and the firm should overinvest in its quality. Similarly, if a consumer underestimates the likelihood of using the qth unit, then she will underestimate the cost of a quality reduction, and the firm should underinvest in its quality. Just as overprecision leads firms to charge zero marginal price up to a usage allowance followed by high marginal charges thereafter, overprecision could also lead firms to offer high-quality service up to a usage allowance followed by low quality thereafter.

These conjectures about optimal product quality may explain why T-Mobile offers cellular data plans that include an allowance of data at high speed but provide additional data beyond the allowance at slow speed. If overconfident consumers underestimate the variance of their total data usage, then inefficiently high speed should be provided up to an allowance after which inefficiently low speed should be provided. They may also explain coverage limits on car insurance. If overconfident consumers underestimate the variance of their accident losses, then they underestimate the likelihood of large losses and the firm should respond by reducing coverage quality for large losses. Coverage limits accomplish this objective. Table 2 summarizes both examples.

When Overconfidence Involves Overoptimism about Navigating Contract Terms

Now return to the case where overconfident consumers overvalue contracts because they overestimate their abilities to navigate contract terms. Taking full advantage of a contract often requires follow-through—remembering and then completing a costly task in the future, whether it be mailing in a rebate or canceling service once an introductory rate expires. Any contract term offering a future benefit after a costly task is completed presents two challenges. First, such terms create a *memory hurdle*. Overconfidence about prospective memory leads consumers to overestimate the likelihood of remembering to complete tasks and overvalue the contract to firms' benefit (Hollman and Zaidi 2010). Second, such terms serve as a *self-control trap* that exploits

Table 2
Product Quality Distortions

Product or service	*Consumers must forecast*	*Example contract*
Smart phone service	Data usage	$50/month for unlimited talk and text and 1GB of data at high speed. Additional data provided at slow speed (T-Mobile 2014).
Car insurance	Losses	Premium, deductible, and coverage limit beyond which 0 percent of additional loss is covered (Liberty Mutual 2014).

Source: Author.

overconfidence about self-control (DellaVigna and Malmendier 2004). If the cost and benefit of the task are correctly balanced (such that the ratio of benefit to cost is large enough but not too large), then an overconfident consumer will anticipate having the self-control to complete the task in a timely fashion, but in practice will either procrastinate and delay completing the task or fail to complete it altogether (O'Donoghue and Rabin 1999). This is a second reason that overconfident consumers overvalue the contract to the firm's benefit. Table 3 lists a few of the many contract terms that require follow-through. In each case, while there may be other explanations, firms may include the contract terms simply to exploit overconfidence about prospective memory, overconfidence about self-control, or both.

Beyond remembering tasks and exercising self-control, navigating contract terms as intended often also requires consumers to pay attention. For instance, a checking account customer who does not pay attention to her account balance can easily and unintentionally pay her bank $35 to buy a cup of coffee with her debit card, not realizing that her account balance is exhausted and that an overdraft fee applies. Stango and Zinman (2014) find that "60 percent of overdrafters reported overdrafting because they 'thought there was enough money in my account.'" Importantly, if banks designed overdraft fees simply to reflect the marginal cost of extending credit, then they should want to disclose overdraft fees at the point of sale to increase the social efficiency of consumers' choices. Card-processing terminals could be designed to ask a consumer, "Overdraft fee applies. Continue Yes/No?"—but they do not. Instead, the lack of transparency creates an *attention hurdle*, which consumers who are overoptimistic about their attention levels overestimate the likelihood of clearing, leading to contract overvaluation rather than increased efficiency.

Hidden fees are those that consumers may be unaware of when signing a contract because firms disclose them only in the fine print if at all. Whether or not overdraft fees are hidden fees, they are an example of something distinct: *surprise penalty fees*. These are fees that firms impose as a penalty for crossing a consumption threshold, but make a surprise by choosing not to notify customers when they reach the relevant threshold. Consumers can thus only avoid them by paying attention to approaching thresholds themselves. *Surprise loyalty discounts*, which lower rather than raise marginal

Table 3
Barriers to Follow-Through: Memory Hurdles and Self-Control Traps

Contract term	*Explanation*
Mail-in rebates	Overconfident consumers may overestimate the likelihood of remembering to mail in a rebate or of having the self-control to avoid procrastinating until the deadline is missed.
Free trials or teaser rates with switching or cancelation costs	When a free trial or teaser rate expires, the consumer must go to the trouble of switching or canceling to avoid the newly higher fees. Overconfidence about self-control or prospective memory both lead one to overestimate the likelihood of switching or canceling and achieving the benefits rather than procrastinating or forgetting.
Auto-renewal	Auto-renewal makes switching or quitting relatively more costly and can lead to overestimation of switching or quitting.
Bonus cash back (quarterly activation required)	Some credit cards offer additional cash back conditional on customers actively opting in each quarter. Overconfident consumers may overestimate the likelihood of opting in.

Source: Author.
Note: See DellaVigna and Malmendier (2004, 2006), Eliaz and Spiegler (2006), and Heidhues and Kőszegi (2010) for more examples.

price after crossing a consumption threshold, are also attention hurdles. For instance, the last flight required to achieve elite status is effectively discounted by the value of elite rewards. Yet frequent flyers must keep track of their mileage balances to know whether the implicit discount applies to a current trip. Consumers who are overoptimistic about their attention levels will underestimate the likelihood of paying surprise penalty fees but overestimate the likelihood of collecting surprise loyalty discounts. In either case, consumers overvalue contracts with attention hurdles to firms' benefit (Grubb 2015a). More examples are shown in Table 4.

More generally, if overconfident consumers overestimate their ability to clear hurdles and avoid traps in contract terms, then this makes it profitable to add such hurdles and traps to contracts. If consumers overestimate their ability to avoid fees of some kind then firms should add them to contracts. If consumers overestimate their ability to collect discounts, such as mail-in rebates, then firms should add them to contracts. While overconfidence is prevalent, consumers may also be underconfident about their abilities to complete some tasks. Including such tasks in contract terms would only lead consumers to undervalue contract surplus to firms' detriment. Thus, in equilibrium, we should only expect firms to complicate contracts with the sorts of tasks that consumers are overconfident they can complete.

Alternative Explanations for Complex Contracts

It is important to recognize that price-discrimination models with rational expectations can explain almost any pricing pattern, such as rebates (Narasimhan 1984),

Table 4
Attention Hurdles

Product or service	*Consumers must attend to*	*Source of returns to attention.*
Checking account	Account balance	Debit card transaction fee rises from $0 to $35 when balance falls to zero.
Credit card	Account balance	Crossing credit limit triggers over-limit fee*
Smart phone service (prior to 2013**)	Data usage	Marginal price of data rises from $0 to $15 per 500MB after 1GB of usage.
Frequent flyer program	Mileage balance	Perks are awarded upon crossing various mileage thresholds.

Source: Author.
* US over-limit fees are restricted by the 2009 Credit Card Accountability Responsibility and Disclosure (CARD) Act.
** By agreement reached with the Federal Communications Commission (FCC) in 2011, US carriers began alerting customers when they approached or exceeded usage allowances in 2013 (CTIA-The Wireless Association 2011).

menus of three-part tariffs (Grubb 2009), or (allowing for inattention) surprise penalty fees (Grubb 2015a). Moreover, there are often still more explanations for complex pricing strategies, other than either overconfidence or price discrimination. However, careful empirical work has shown that, in a variety of market settings discussed above, such models with unbiased consumers fail to simultaneously explain consumer behavior. As a result, consumer overconfidence can be a better explanation for observed pricing than any rational-expectations-based alternative.

For example, consider that three-part tariffs approximate the optimal contract for exploiting overprecise demand forecasts. Can we conclude that consumers are overprecise whenever firms offer three-part tariffs? The answer is "no," because three-part tariffs can also be optimal for price discrimination among unbiased consumers. However, I find that in the 2002–2005 US cellular phone service market, consumers' calling patterns are inconsistent with the price discrimination explanation.[5] This bolsters direct evidence of overprecision from consumers' calling plan choice mistakes and leads me to conclude that consumer overprecision is the best explanation in this market (Grubb 2009).

Such a conclusion can only be drawn on a market-by-market basis, and it remains a challenge to test the importance of consumer overconfidence across a broader array of markets. Nevertheless, the existing evidence is strong enough that consumer overconfidence should be a leading hypothesis in markets with pricing features like those described in Tables 1–4.

[5] Chao's (2013) alternative explanation of three-part tariffs is not applicable because it is based on a multistop shopping model, while most consumers contract with only one cellular phone service provider.

Complex Pricing is Robust to Competition

Competition does not eliminate complex forms of pricing that are designed to exploit overconfidence. Instead, its primary effect is to lower fixed fees.[6] When competition forces firms to offer more value to consumers, lowering fixed fees is optimal because, unlike lowering marginal fees or adjusting other contract terms, this does not diminish either the true surplus generated by a contract or the amount by which consumers overvalue contracts.

In fact, in a competitive market for a homogeneous good, a firm that simply priced at cost would fail to capture any market share. An example adapted from Grubb (2009) illustrates this point. Suppose that cellular service providers have marginal costs of 10 cents per minute and fixed costs of $40 per customer. Consumers value each minute of calling at 45 cents up to some satiation point, which they do not learn until after signing a contract. Consumer satiation points will either be 100, 400, or 700 minutes, each equally likely. However, overconfident consumers underestimate their uncertainty and believe they will be satiated at 400 minutes with probability 1. Cost-based pricing yields true and perceived utility of $400(\$0.45 - \$0.10) - \$40 = \$100$ to consumers. However, another firm could charge $60 for an allowance of 400 included minutes followed by 45 cent per minute charges for additional calling. In reality, this contract shifts $25 of surplus from consumers to firms, yielding expected profits of $25 and expected utility of $75. Absent overconfidence, no one would choose the contract. However, overconfident consumers perceive their expected utility to be higher, at $400(\$.45) - \$60 = \$120$, overvaluing the contract by $45, and will choose the contract over cost-based pricing.

Figure 1 shows the two pricing schemes graphically. An overconfident consumer expects to be in the shaded region of the figure where the three-part tariff is below cost. The firm recognizes, however, that two-thirds of the time the consumer will actually end up in the non-shaded regions where three-part tariff revenues are above cost. The contract serves as a bet about which region the consumer will end up in, and due to consumer overconfidence, both firm and consumers believe they will win on average.

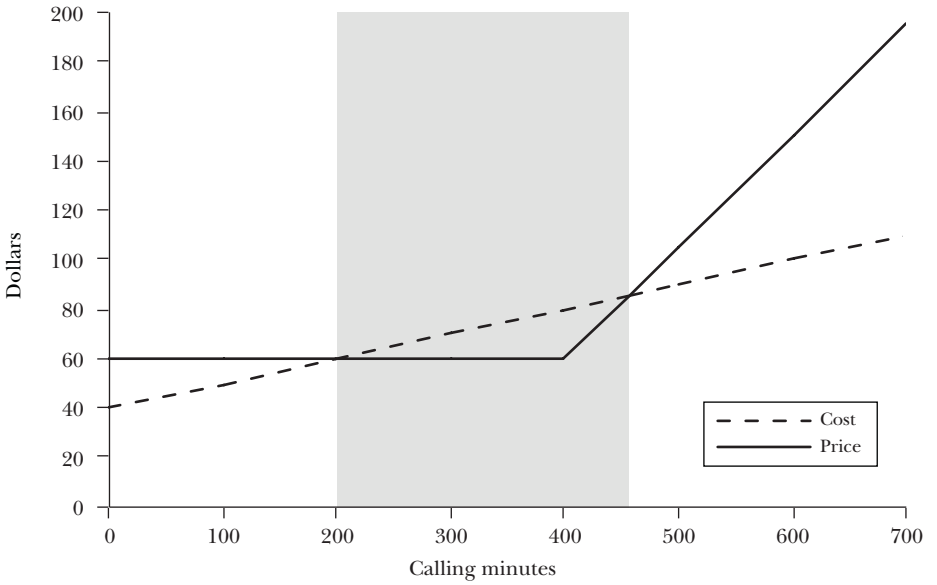
What Are the Equilibrium Welfare Consequences of Consumer Overconfidence?

How do firms, overconfident consumers, and society fare compared to a counterfactual world in which consumers have rational expectations? [7] Answering this question is useful for at least two reasons. First, doing so helps us better understand

[6] I assume one-stop shopping or exclusive contracting, where consumers buy from only one firm. For nonexclusive contracts, competition ensures that marginal prices do not exceed marginal cost (Gottlieb 2008).

[7] I measure welfare of an overconfident consumer as expected utility with respect to the true distribution of outcomes. See Bernheim (2009) and Beshears, Choi, Laibson, and Madrian (2008) for alternative views on the best approach.

Figure 1
Three-Part Tariff Pricing



Source: Author.
Notes: Suppose that cellular service providers have marginal costs of 10 cents per minute and fixed costs of $40 per customer. One firm prices at cost while another firm offers the three-part tariff shown, charging $60 for an allowance of 400 included minutes plus 45 cents per minute for additional calling. In the example described in the text, overprecise consumers choose the three-part tariff because they anticipate their usage always falling inside the gray-shaded interval, where the contract is below cost. The three-part tariff is profitable on average, however, because consumers turn out to consume in the profitable non-shaded regions more often than not.

how markets operate with overconfident consumers. Second, while no magic wand exists to wave away overconfidence, feasible policies can sometimes reduce contract overvaluation, thereby having the same effect on market outcomes as reducing consumer overconfidence.

Answering the welfare question requires first addressing another: Will overconfident consumers over- or undervalue equilibrium contract offers? The answer affects whether overconfidence raises or lowers industry profits, harms or helps consumers, and whether overconfidence expands or contracts the market. If overconfident consumers are overoptimistic about their own levels of self-control, prospective memory, or attention, and hence their ability to take advantage of a contract's potential value, then they overvalue contracts. For example, an individual overconfident about his own self-control overestimates his future gym attendance and hence his value of a gym membership. In contrast, if overconfident consumers misforecast their future usage because they misforecast their future valuations for the service, then they might undervalue contracts. Undervaluing car insurance due to overoptimism about

driving ability is a clear example. Less obvious is the fact that overprecision in demand forecasts can also lead to contract undervaluation (Grubb 2009).

As overconfidence in different contexts can lead to either overvaluation or undervaluation of contracts, both are relevant possibilities. From a firm's perspective, contract overvaluation is equivalent to an upward shift in demand. Typically this increases both prices and sales of contracts, to the benefit of firms but the detriment of infra-marginal consumers who pay the price increase. The reverse is true for contract undervaluation: prices and sales typically fall to the detriment of firms but the benefit of infra-marginal consumers. In either case, consumers on the margin are harmed. Contract overvaluation lures new consumers into paying a price above the value of the contract, while contract undervaluation discourages marginal consumers from accepting a good deal.

The social costs of overconfidence are less dependent on whether consumers overvalue or undervalue contracts. In either case, overconfidence distorts allocations, creating deadweight losses in competitive markets. In particular, overconfidence distorts allocations on both intensive and extensive margins, which Heidhues and Kőszegi (2015) dub the *exploitation* and *participation* distortions, respectively. The exploitation distortion arises because the pricing policies that optimally exploit overconfidence also distort consumer quantity choices on the intensive margin. For instance, overage rates on a cellular service plan may exploit overprecision, but also inefficiently suppress calling or data usage that is valued above its marginal cost. The participation distortion arises because overvaluation causes too many consumers to sign contracts, while undervaluation causes too few consumers to sign contracts.

In competitive market settings in which outcomes would be efficient absent overconfidence, then both distortions on intensive and extensive margins unambiguously lower social welfare. A caveat is that overconfidence could be welfare-improving if the distortions due to overconfidence are countervailing to other distortions already present in the marketplace. In particular, if market participation is already inefficiently low due to market power, then market expansion due to contract overvaluation can increase social welfare, albeit at consumers' expense.

Contract Overvaluation Benefits Firms but Hurts Consumers and Society in Competitive Markets

The informal discussion above claims that overconfidence with contract overvaluation typically benefits firms, harms consumers, and in competitive markets, harms society. Next, I describe a framework that makes this claim precise, and shows how the magnitude of the effects depend on two important market statistics—the elasticity of demand and the pass-through rate, which measures the fraction of an infinitesimal per-unit cost increase that is passed-through to consumers as a price increase.

Figure 2 illustrates the consequences of overconfidence with contract overvaluation in a perfectly competitive market. As is common in this kind of partial-equilibrium analysis, I omit income effects by assuming quasi-linear utility,

and I do not allow for firm entry or exit. Moreover, I proceed via the use of two tricks.[8] Pricing contracts are described by a fixed payment P and a vector of additional terms $\boldsymbol{p}$. For instance, an insurance contract charges a fixed premium (P) but also specifies coverage limits, deductibles, and co-insurance rates ($\boldsymbol{p}$). My first trick is standard: Given rational expectations (RE), I focus on the fixed payment (P_{RE}) as the "price" of the contract. Additional terms I treat like dimensions of product quality; I fix them at their equilibrium values ($\boldsymbol{p}_{RE}$) and suppress them from the graphical analysis. One can think of the contract simply as a product that delivers true expected utility U for price P_{RE}. Given rational expectations, this yields familiar demand and supply curves for contracts in Figure 2.

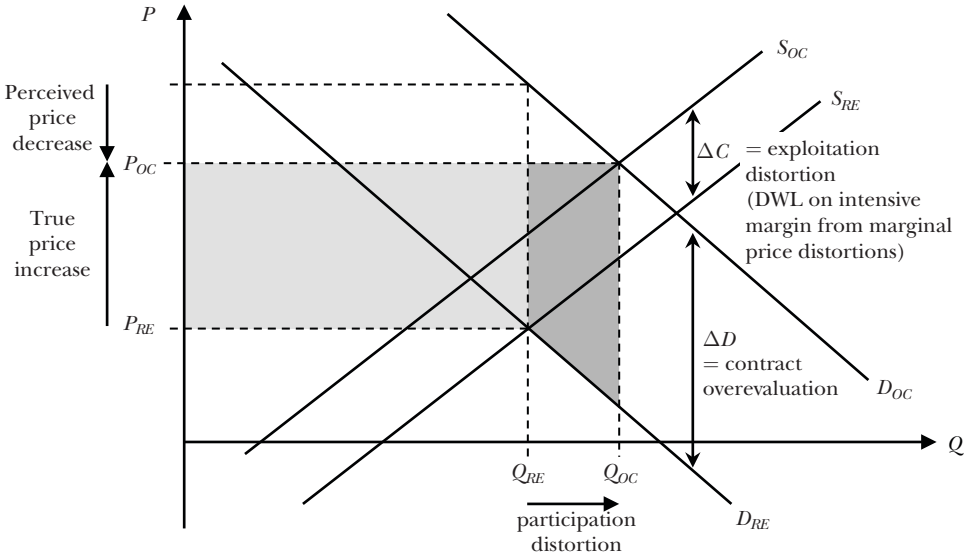
Analyzing demand and supply curves on the same figure for the case of overconfidence (OC) requires a second trick. As discussed earlier, consumer overconfidence will lead firms to complicate a contract's additional terms to include memory hurdles, self-control traps, or attention hurdles, or to create three-part tariffs or quality distortions. To make these more complicated contracts comparable to those offered under rational expectations, I partition the contract's fixed payment into two parts. These are the "price" of the contract (P_{OC}) and a second fixed fee F to be included with the additional terms $\boldsymbol{p}_{OC}$ that are suppressed from the analysis. The second fixed fee is chosen so that, gross of the price P_{OC}, the additional contract terms offered to overconfident consumers (F, $\boldsymbol{p}_{OC}$) yield the same true expected utility U as the additional contract terms ($\boldsymbol{p}_{RE}$) offered under rational expectations. As a result, regardless of whether one considers the case of rational expectations or overconfidence, one can treat offered contracts simply as products that deliver true expected utility U.

A numerical example can help to clarify how this method works. Given rational expectations, a wireless plan might charge \$50 per month for unlimited talk time. In this case the "contract" is "unlimited talk" and its "price" is \$50. Given overconfidence, a wireless plan might charge \$40 per month for 500 included minutes and charge \$0.50 per minute thereafter. This might lead consumers to make fewer calls, thereby losing \$5 in value from foregone calls, and to pay \$10 in additional calling charges. In this case the "contract" is "500 included minutes and \$0.50 thereafter with a \$15 refund" and its "price" is \$55. The \$15 adjustment to the definition of the "contract" offsets the \$5 in foregone calls and the \$10 in additional calling charges. Thus, both contracts offer consumers comparable value when considered gross of their respective prices.

Figure 2 plots contract demand and supply curves under both rational expectations and consumer overconfidence. On the demand side, and given rational expectations, a consumer's value for a contract is the difference between its utility and her outside option. The rational expectations demand curve is downward sloping, rather than a horizontal line, because consumers have heterogeneous outside options. Given overconfidence, the demand curve is shifted upwards by

[8] Previous graphical treatments of welfare given distinct curves describing demand and consumer valuations include Bernheim and Rangel (2009), Madrian (2014), and Spinnewijn (2014).

Figure 2
Consequences of Contract Overvaluation Due to Overconfidence in a Competitive Market



Source: Author.
Notes: The figure illustrates the consequences of overconfidence with contract overvaluation in a perfectly competitive market, plotting contract demand and supply curves under both rational expectations (RE) and consumer overconfidence (OC). "DWL" is deadweight loss. See text for details.

ΔD, the amount that consumers overvalue the offered contract. While the overconfident demand curve determines price and quantity with overconfidence, the rational expectations demand curve is still relevant for welfare calculations with overconfidence because it describes consumers' true valuations.

On the supply side, overconfidence leads to the exploitation distortion—a deadweight loss on the intensive margin equal to ΔC for each customer served. Overconfidence therefore shifts the supply curve upward by this amount, which reflects firms' increased cost of delivering the same true utility U to consumers. (If contract overvaluation results from overoptimism about ability to navigate contract terms, then this shift must be smaller than the shift in demand, as otherwise the contract terms would not be optimal for firms.)

The presence of overconfidence causes demand and supply curves to shift. It causes the equilibrium price to rise from P_{RE} to P_{OC}, and equilibrium quantity to increase from Q_{RE} to Q_{OC}. There are two losses to consumer surplus. First is the area shaded light gray, corresponding to the true price increase borne by existing customers. Second is the area shaded dark gray, corresponding to the amount new customers pay above their true valuations. Deadweight loss to society includes waste on the intensive margin ($\Delta C \cdot Q_{RE}$) captured by the upward shift in the supply

curve, and waste on the extensive margin (the dark gray area between S_{OC} and D_{RE}) due to inefficiently high contract sales.

The magnitudes of consumer harm and deadweight loss depend on how much price and quantity increase, which in turn depends on two important market statistics: the elasticity of demand and the pass-through rate. In perfectly competitive markets, the pass-through rate varies between 0 and 1 according to the elasticities of supply and demand: $\rho = \epsilon_S/(\epsilon_S - \epsilon_D)$. Given a constant market pass-through rate, ρ, the true price increase for a contract offering utility U is a weighted average of the shifts in demand and supply:

$$P_{OC} - P_{RE} = (1 - \rho)\Delta D + \rho\Delta C.$$

Due to contract overvaluation, however, consumers perceive a price decrease of $\rho(\Delta D - \Delta C)$, which increases sales according to the elasticity of demand.[9]

When the pass-through rate is zero, firms raise the price of a contract offering utility U by the amount consumers overvalue the contract, ΔD. Consumers perceive no effective change in the price. Firm profits per customer, however, rise by the price increase less the deadweight loss from distortions on the intensive margin, $(\Delta D - \Delta C)$. When the pass-through rate is positive, firms pass fraction ρ of these potential profits back to consumers. Consumers perceive this to be a price cut of $\rho(\Delta D - \Delta C)$ when in truth it merely reduces the price increase to $(1 - \rho)\Delta D + \rho\Delta C$.

As highlighted in Table 5, there are several implications for welfare. First, a high pass-through rate near 1 protects infra-marginal consumers from their overconfidence by limiting true price increases but, for the same reason, tempts more consumers on the extensive margin to buy at prices above their true valuations by increasing the perceived price drop. In contrast, a zero pass-through rate eliminates the participation distortion (unless demand is perfectly elastic), but maximizes the harm to infra-marginal consumers. Next, while deadweight loss on the intensive margin from the exploitation distortion is always $\Delta C \cdot Q_{RE}$, the participation distortion relies both on consumers perceiving a price drop and on them being responsive to it. Thus, there is no participation distortion either when the pass-through rate is zero and demand is not perfectly elastic, or when demand is perfectly inelastic. In contrast, the participation distortion (and hence the total social cost of overconfidence) is largest when pass-through is high and demand is elastic.

If firms have market power, neither Figure 2 nor the conclusion about total deadweight loss in Table 5 applies. However, the expressions which relate true and perceived price changes to the market pass-through rate are the same. Thus, given a market pass-through rate less than one, contract overvaluation leads to a true price increase but additional sales due to a perceived price drop, which is good for

[9] In similar analyses using the market pass-through rate, Agarwal, Chomsisengphet, Mahoney, and Stroebel (2014) describe the effect of regulating hidden fees on up-front prices, and Farrell (2008) describes the welfare loss when consumers underestimate aftermarket costs.

Table 5
Welfare Consequences of Contract Overvaluation Due to Overconfidence in a Perfectly Competitive Market

Setting	*Outcome*
$\rho = 0$ and $\epsilon_D > -\infty$	No participation distortion or harm to marginal consumers. Minimum deadweight loss ($\Delta C \cdot Q_{RE}$). Maximum price increase (ΔD) to infra-marginal consumers.
$\rho = 1$ and $\epsilon_D = 0$	No participation distortion or harm to marginal consumers. Minimum deadweight loss ($\Delta C \cdot Q_{RE}$). Minimum price increase (ΔC) to infra-marginal consumers.
$\rho = 1$ and $\epsilon_D \ll 0$	Maximum participation distortion and harm to marginal consumers. Maximum deadweight loss. Minimum price increase (ΔC) to infra-marginal consumers.

Source: Author.

firms but bad for consumers, whether firms have market power or not. A difference from the competitive case is that the additional contract sales may be socially valuable if they offset otherwise inefficiently low sales due to market power. Finally, and unlike in a competitive market, the market pass-through rate can exceed one with market power, which implies that contract overvaluation could conceivably benefit infra-marginal consumers by lowering prices.[10]

Contract Undervaluation Harms Firms and Society but May Benefit Some Consumers in Competitive Markets

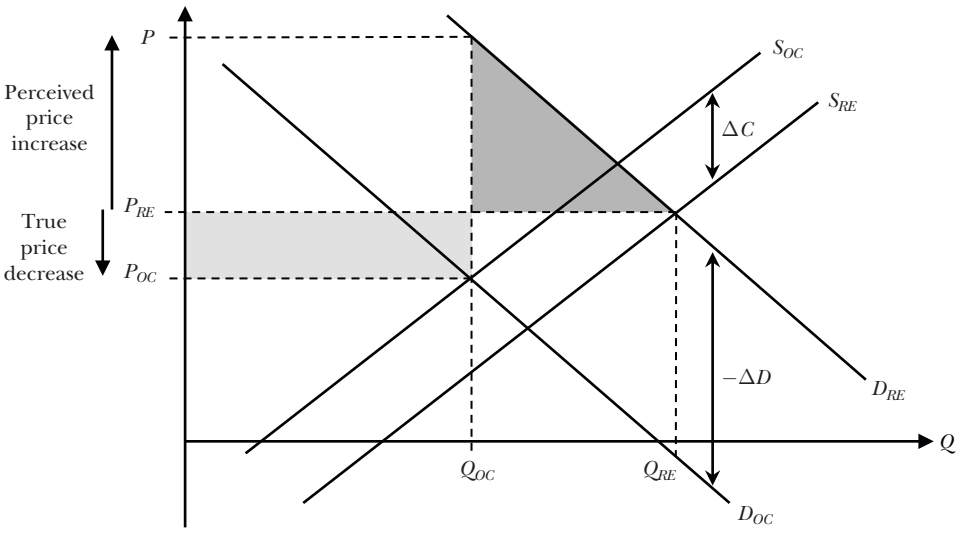
When overconfident consumers misforecast their future valuations for services, they may undervalue contracts. In this case, overconfidence shifts demand downwards relative to rational expectations and may depress the market price.[11] This is bad for firms but good for infra-marginal consumers who enjoy the lower price. On the extensive margin, however, some consumers with true contract valuations above the price will stop buying to their own detriment. Thus, the overall effect of overconfidence on consumer surplus may still be negative. Moreover, absent any source of inefficiency other than overconfidence, overconfidence remains unambiguously bad for social welfare in a perfectly competitive market.

Figure 3 is similar to Figure 2 but depicts the case in which overconfidence leads to sufficient *under*valuation of the contract to depress its market price.

[10] Given a constant pass-through rate, this occurs if $\rho > \Delta D/(\Delta D - \Delta C)$. Whether ρ exceeds this threshold depends on the curvature of the logarithm of demand, which increases pass-through given market power (Weyl and Fabinger 2013).

[11] Expressions for price changes from the preceding section continue to hold, so price falls if $\Delta D < -\Delta C \frac{\rho}{(1-\rho)}$.

Figure 3
Consequences of Contract Undervaluation Due to Overconfidence in a Competitive Market



Source: Author.
Notes: The figure illustrates the consequences of overconfidence with contract undervaluation in a perfectly competitive market, plotting contract demand and supply curves under both rational expectations (RE) and consumer overconfidence (OC). See text for details.

Overconfidence causes the equilibrium price to fall from P_{RE} to P_{OC}, and equilibrium quantity to fall from Q_{RE} to Q_{OC}. Firm profits fall with demand. Deadweight loss to society includes the area between the supply curves (waste on the intensive margin) and the lost surplus due to inefficiently low contract sales (waste on the extensive margin). There are two changes to consumer surplus. First, infra-marginal customers benefit from the true price decrease, benefiting by the area shaded light gray. Second, consumers on the extensive margin lose the area shaded dark gray because their undervaluation causes them to forego purchasing contracts with true expected value above their price. The net effect for consumer surplus may be positive or negative. Nevertheless, overconfidence remains socially costly, as lost profits more than offset any consumer benefit.

Does Competition (Partially) Protect Consumers from Overconfidence?

Suppose that overconfidence leads consumers to overvalue contracts. Then in competitive markets, as discussed above, overconfident consumers are worse off than they would be in a counterfactual world with rational expectations. Thus, competition does not completely protect consumers from overconfidence. However, one may still ask whether competition partially protects consumers from overconfidence. First, does increased competition benefit overconfident consumers? Second,

consider policies that directly reduce contract overvaluation, for instance by limiting the use of contract terms described in Tables 1–4. Does competition reduce the consumer harm from overconfidence and thereby limit the potential benefit from such policies?

A common assumption in the literature is that competition increases the market pass-through rate and that there is full market coverage, meaning that all consumers buy and are infra-marginal.[12] Under this assumption, the answer to both questions is "yes." Infra-marginal consumers always benefit from lower prices and so benefit from competition. Moreover, if competition raises the pass-through rate it also limits the amount by which contract overvaluation raises prices, and hence limits the cost of overconfidence to infra-marginal consumers. Thus competition reduces returns to consumer protection policies that directly reduce contract overvaluation.

Perhaps surprisingly, however, when the full-market coverage assumption is relaxed, the answer to both questions is "not necessarily." The reason is that competition and low prices have very different welfare consequences for marginal consumers than for infra-marginal consumers. In particular, while lower prices always benefit infra-marginal consumers who pay less, lower prices can actually harm marginal consumers who overvalue contracts. As a result, competition can have counterintuitive effects on consumer welfare.

Accounting for contract overvaluation, consumers just indifferent to buying are actually strictly worse off buying than not. These consumers would be protected by high prices that keep them out of the market. Reduced prices due to competition can tempt them to buy when they should not. It is therefore possible that policies that successfully increase competition, lower prices, and expand sales also lower consumer surplus.

Moreover, while competition that raises the pass-through rate reduces the true price increase that results from contract overvaluation, it simultaneously increases any perceived price decrease due to contract overvaluation. Thus while competition mitigates the cost of overconfidence to infra-marginal consumers, it increases the costs to marginal consumers, more of whom are lured into the market to pay more than their true valuations.[13] Therefore, while we might say that competition partially protects infra-marginal consumers from overconfidence, we might also say that market power partially protects marginal consumers from overconfidence (by pricing them out of the market). If demand is elastic, then the latter effect may be important, and competition may increase returns to policies that reduce contract overvaluation.

[12] For example, DellaVigna and Malmendier (2004) assume marginal cost is constant and all consumers have the same outside option. Thus all consumers buy, and pass-through is zero for monopoly but one for perfect competition.

[13] This prediction is reversed if competition decreases rather than increases the pass-through rate, as is certainly possible (Bulow and Pfleiderer 1983; Weyl and Fabinger 2013).

Pooling and Cross-Subsidization

Thus far in my discussion, it may seem that consumer harm is limited when the pass-through rate is high and market demand is very inelastic (Table 5, row 2). This is only true, however, when all consumers are equally overconfident. Consider an example inspired by Gabaix and Laibson (2006). There is a competitive market for checking accounts with a perfectly elastic supply of accounts. A bank's marginal cost of processing an overdraft transaction is zero but the typical fee is $35. All consumers opening checking accounts believe they will pay attention to their balance and avoid overdraft fees. Half of them have rational expectations and do avoid fees. Half, however, are overconfident about their attention and in fact incur $100 in overdraft fees due to inattention.

Given the assumption that supply is perfectly elastic, banks cannot profit from consumer overconfidence. If all consumers were overconfident, annual fees for checking accounts would be $100 below cost and banks would break even with overdraft fee revenue. The market pass-through rate of one ensures that 100 percent of the overdraft fees, which overconfident consumers fail to anticipate paying, are passed back to consumers through lower account fees. The only harm to overconfident consumers would be on the extensive margin. Some who valued checking accounts below cost would be lured to open an account by underestimating its true cost.

How is the welfare of overconfident consumers affected when half of consumers are rational? One might hope that the presence of rational consumers might provide a positive externality that helps protect overconfident consumers. We know, for instance, that consumers with low search costs can benefit others by lowering equilibrium prices for all. Unfortunately, however, the presence of rational consumers does not protect the overconfident from harm. On the contrary, rational consumers can exert what Armstrong (2015) dubs a negative *rip-off externality* on the overconfident, raising the prices they pay.

In this example, when equal numbers of consumers are rational and overconfident, annual checking account fees are priced only $50 below cost. Banks still pass 100 percent of overdraft fee revenue back to consumers in lower fees. Now, however, the account fee reduction is shared equally between overconfident consumers who pay the overdraft fees and rational consumers who do not. As a result, infra-marginal overconfident consumers now pay $50 above cost for their accounts. This overpayment does not accrue to firms, who still make zero profit, but to rational consumers, who receive the $50 as a cross-subsidy. The only overconfident consumers to benefit from the presence of rational consumers are those on the extensive margin who are dissuaded from opening an account when the account fee is $50 below cost rather than $100 below cost.

In short, even when a high pass-through rate prevents firms from exploiting infra-marginal overconfident consumers, they may still be in danger. Rational consumers who choose the same contracts and receive cross-subsidies may exploit them instead. In practice, this redistribution between consumers can be large: in 2013, US overdraft fees totaled $32 billion (Andriotis 2014), but Stango and Zinman (2009, 2014) find that these fees are paid by less than half of account holders. In fact,

just 16 percent of account holders pay over 70 percent of overdraft fees. Because this group tends to be lower income than the general population, the cross-subsidies from the fee payers to the free riders are regressive (Parrish and Frank 2011).

Consumer Protection Policy

Overconfidence may harm consumers in many ways. Indeed, firms continually invent new ways to exploit consumer overconfidence because the returns to such *exploitative* innovation equal or exceed returns to traditional innovation that increases product value (Heidhues, Kőszegi, and Murooka forthcoming). By describing the costs of overconfidence, the previous section shows that policies that mitigate its effects would be valuable, even in competitive markets. Researchers have proposed at least three practical approaches for addressing overconfidence: 1) requiring disclosure to guide contract choice, 2) requiring disclosure to aid contract navigation, and 3) restricting pricing terms.

Bar-Gill (2012) makes the case for the first approach. As overconfident consumers misforecast costs and benefits of contracts, the proposal is to require firms to disclose expected costs and benefits up front. For example, consider a cellular service plan that charges a monthly fee plus additional charges for usage beyond allowances. Disclosing contract terms does not prevent overconfident consumers from misforecasting its cost because the additional charges are a function of future usage. Three possible additional disclosure requirements are imagined. First, the seller could disclose the average monthly bill across all existing customers of the plan (which I find in Grubb, 2009, to be 90 percent higher than the monthly fee in a 2002–2005 panel of students' cellular phone bills). Second, if the seller already has a relationship with a customer, it could personalize the disclosure by reporting what the average monthly bill would be if evaluated at the customer's own past usage levels. Third, as advocated by Thaler and Sunstein (2008), sellers could let customers easily share their entire usage and billing history with third parties who could then give personalized advice about the expected costs of all calling plans in the market. Plans are already underway in the UK to implement this third disclosure requirement in retail energy markets (Department of Energy & Climate Change and Davey 2014) and the retail banking market (Competition & Markets Authority 2014).

The second approach to addressing overconfidence—requiring disclosure to aid contract navigation—has been implemented in both EU and US cellular phone service markets. EU and US consumers must be alerted to roaming rates upon crossing national borders, and US consumers must be alerted as they approach or exceed usage allowances (European Commission 2011; CTIA-The Wireless Association 2011). Such disclosures can make overoptimism about attention to roaming boundaries or usage allowances irrelevant, as the alerts can substitute for consumers' own attentiveness (Grubb 2015a). Extending this policy to retail banking, by requiring point-of sale overdraft warnings, could have substantial benefits (Armstrong and Vickers 2012; Grubb 2015a).

The third approach to addressing overconfidence—regulating prices—is naturally more controversial than a disclosure requirement. In the context of overoptimism about self-control, DellaVigna and Malmendier (2004) argue that price regulation requires too much information (about firm costs, market structure, and consumer preferences) to be practical. Heidhues and Kőszegi (2010) are more optimistic, however, and advocate "prohibiting disproportionately large penalties for deferring small amounts" of repayment in credit markets. They argue that a prohibition would limit exploitation of overoptimism about self-control and would have no downside—as there is unlikely to be any efficiency rationale for the fees. Similar bans of any of the contract terms outlined in Tables 1–4 might also be considered but would have to be carefully weighed against possible drawbacks. For instance, banning three-part tariffs in car leasing (and forcing firms to use two-part tariffs with a constant per-mile charge) would prevent exploitation of overprecise mileage forecasts but could be socially costly if depreciation is nonlinear in mileage.

I hope that regulatory intervention can help ameliorate negative consequences of overconfidence. In evaluating regulatory proposals, however, it is crucial to anticipate equilibrium responses of firms. Correcting individual decision-making errors will often have smaller benefits when equilibrium considerations are taken into account. In some cases, de-biasing consumers can actually cause consumer harm by increasing the prices they pay. For instance, eliminating contract undervaluation can raise prices. Alternatively, de-biasing a fraction of overconfident consumers can exacerbate cross-subsidies paid by those who remain overconfident.[14] Importantly, as identified in the previous section, market pass-through rate and elasticity of demand are useful market statistics for understanding what the equilibrium effects will be. To illustrate this point, consider contrasting cases of intervention in the cellular phone service and credit card markets.

First, by agreement reached with the Federal Communications Commission (FCC) in 2011, US carriers began alerting customers when they approached or exceeded usage allowances in 2013 (CTIA-The Wireless Association 2011). Using a structural model estimated with 2002–2005 student billing data, counterfactual simulations predict that, had the agreement been implemented during the sample period, it would have benefited the average consumer by about $100 annually were prices held fixed. However, when firms' equilibrium pricing response is taken into account, average monthly fees increase to offset lost overage revenue, and the prediction is that the average consumer does not benefit (Grubb and Osborne 2015).

Second, the US 2009 Credit Card Accountability Responsibility and Disclosure (CARD) Act required increased disclosure by credit card lenders and restricted a variety of fees, including over-limit and late-payment fees. The fee restrictions limit the ability of credit card lenders to exploit overoptimism about self-control

[14] See also Mullainathan, Schwartzstein, and Congdon (2012); Handel (2013); Spiegler (forthcoming); and Grubb (2015a) for examples of improved individual decisions worsening equilibrium outcomes.

(Heidhues and Kőszegi 2010) as well as overoptimism about attention to balances and dates. Agarwal et al. (2015) estimate that the CARD act lowered over-limit and late-payment fees of US credit card users by $11.9 billion annually. Moreover, they estimate that there was zero offsetting increase in interest rates or other fees, and zero reduction in available credit. As a result, US consumers really are better off by $11.9 billion per year.

Both cellular usage alerts and restrictions on credit card over-limit fees should reduce any contract overvaluation that arises due to overprecise calling or borrowing forecasts, or due to overoptimism about attention to calling or borrowing levels. Why then should the CARD act benefit consumers so much more than the FCC's agreement with cellular carriers? The answer follows from Table 5. In Grubb and Osborne (2015), we estimate that the market for cellular service is highly inelastic (few would give up their cellphones if industry prices went up 1 percent) and has a high pass-through rate near 1. This corresponds to row 2 of Table 5, the case for which consumer benefits of regulation are most limited. In contrast, as Agarwal, Chomsisengphet, Mahoney, and Stroebel (2014) point out, Ausubel (1991) finds almost no pass-through of changes in credit card lenders' costs of funds to credit card interest rates. A zero pass-through rate corresponds to row 1 of Table 5. In this case, regulation is expected to have large benefits to infra-marginal consumers without affecting marginal consumers—exactly as found in Agarwal, Chomsisengphet, Mahoney, and Stroebel (2015).

Why should the pass-through rate differ so much between the cellular and credit card markets? One possibility is that the pass-through rate is zero in the credit card market because binding price floors prevent lenders lowering up-front fees—they may be unable to lower annual fees below zero or raise cash back above receipts from interchange fees without attracting arbitrageurs (Heidhues, Kőszegi, and Murooka forthcoming). Otherwise, as explained by Weyl and Fabinger (2013), pass-through rates vary with the shape of industry cost curves and demand curves, so should be expected to vary across markets.

Conclusion

Overconfidence causes consumers to misweight different contractual terms and product attributes, and so to misforecast costs and benefits of offered products and services. Firms respond by designing contractual terms to increase consumer overvaluation of their products. Consumer harm and deadweight losses result. Happily, practical policies mandating disclosure or restricting contract terms may reduce contract overvaluation—if not overconfidence itself. Naturally, regulation that restricts pricing has the potential for harmful unintended consequences and should be pursued cautiously. Moreover, in some cases, the potential upside from intervention may be limited. Yet, two market statistics, the pass-through rate and the elasticity of demand, can help identify markets most conducive for intervention. Finally, the remarkable success of policies such as the 2009 Credit Card

Accountability Responsibility and Disclosure (CARD) Act suggest that policymakers with a consumer protection mandate should look for more opportunities to ameliorate consequences of overconfidence—and in fact any other bias, such as projection bias or salient thinking, that leads to systematic misweighting of contractual terms.

■ *I thank Mark Armstrong, Daniel Gottlieb, Rani Spiegler, Juan Dubra, Scott Fulford, and Paul Cichello for their comments on earlier drafts.*

References

Adams, Scott. 2003. *Dilbert*, November 7. Comic strip, http://dilbert.com/strip/2003-11-07.

Agarwal, Sumit, Souphala Chomsisengphet, Neale Mahoney, and Johannes Stroebel. 2014. "A Simple Framework for Estimating Consumer Benefits from Regulating Hidden Fees." *Journal of Legal Studies* 43(S2): S239–52.

Agarwal, Sumit, Souphala Chomsisengphet, Neale Mahoney, and Johannes Stroebel. 2015. "Regulating Consumer Financial Products: Evidence from Credit Cards." *Quarterly Journal of Economics* 130(1): 111–64.

Agarwal, Sumit, John C. Driscoll, Xavier Gabaix, and Laibson, David. 2009. "The Age of Reason: Financial Decisions over the Life Cycle and Implications for Regulation." *Brookings Papers on Economic Activity*, Fall, pp. 51–101.

Agarwal, Sumit, John C. Driscoll, Xavier Gabaix, and David I. Laibson. 2013. "Learning in the Credit Card Market." April 24. Available at SSRN: http://ssrn.com/abstract=1091623.

Andriotis, Annamaria. 2014. "Overdraft Fees at Banks Hit a High, Despite Curbs." *Wall Street Journal*, April 1. http://online.wsj.com/news/articles/SB10001424052702304157204579475573602576630.

Armstrong, Mark. 2015. "Search and Ripoff Externalities." *Review of Industrial Organization* 47(3): 272–302.

Armstrong, Mark, and John Vickers. 2012. "Consumer Protection and Contingent Charges." *Journal of Economic Literature* 50(2): 477–93.

Ausubel, Lawrence M. 1991. "The Failure of Competition in the Credit Card Market." *American Economic Review* 81(1): 50–81.

Ausubel, Lawrence M. 1999. "Adverse Selection in the Credit Card Market." June 17, http://www.ausubel.com/larry/creditcard-papers/adverse.pdf.

Bar-Gill, Oren. 2012. *Seduction by Contract: Law, Economics, and Psychology in Consumer Markets*. Oxford: Oxford University Press.

Benoît, Jean-Pierre, and Juan Dubra. 2011. "Apparent Overconfidence." *Econometrica* 79(5): 1591–1625.

Bernheim, B. Douglas. 2009. "Behavioral Welfare Economics." *Journal of the European Economic Association* 7(2–3): 267–319.

Bernheim, B. Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *Quarterly Journal of Economics* 124(1): 51–104.

Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian. 2008. "How Are Preferences Revealed?" *Journal of Public Economics* 92(8–9): 1787–94.

Bolger, Fergus, and Dilek Önkal-Atay. 2004. "The Effects of Feedback on Judgmental Interval Predictions." *International Journal of Forecasting* 20(1): 29–39.

Bulow, Jeremy I., and Paul Pfleiderer. 1983. "A Note on the Effect of Cost Changes on Prices." *Journal of Political Economy* 91(1): 182–85.

Chao, Yong. 2013. "Strategic Effects of Three-Part Tariffs Under Oligopoly." *International Economic Review* 54(3): 977–1015.

Chase. 2014. "Chase Slate Credit Card." https://creditcards.chase.com/credit-cards/slate2.aspx (accessed June 23, 2014).

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99(4): 1145–77.

Competition & Markets Authority. 2014. *Personal Current Accounts, Market Study Update.* London, UK: Competition & Markets Authority. https://assets.digital.cabinet-office.gov.uk/media/53c834c640f0b610aa000009/140717_-_PCA_Review_Full_Report.pdf.

Conlin, Michael, Ted O'Donoghue, and Timothy J. Vogelsang. 2007. "Projection Bias in Catalog Orders." *American Economic Review* 97(4): 1217–49.

CTIA-The Wireless Association. 2011. "CTIA-The Wireless Association, Federal Communications Commission and Consumers Union Announce Free Alerts to Help Consumers Avoid Unexpected Overage Charges." Press Release October 17. http://www.ctia.org/resource-library/press-releases/archive/ctia-federal-communications-commission-consumers-union-announce-free-alerts (accessed August 31, 2015).

DellaVigna, Stefano, and Ulrike Malmendier. 2004. "Contract Design and Self-Control: Theory and Evidence." *Quarterly Journal of Economics* 119(2): 353–402.

DellaVigna, Stefano, and Ulrike Malmendier. 2006. "Paying Not to Go to the Gym." *American Economic Review* 96(3): 694–719.

Department of Energy & Climate Change (UK), and Edward Davey. 2014. "QR Codes on Energy Bills Put Consumers in Control." Press release March 10. https://www.gov.uk/government/news/qr-codes-on-energy-bills-put-consumers-in-control (accessed December 14, 2014).

Einav, Liran, Theresa Kuchler, Jonathan Levin, and Neel Sundaresan. 2015. "Assessing Sale Strategies in Online Markets Using Matched Listings." *American Economic Journal: Microeconomics* 7(2): 215–47.

Eliaz, Kfir, and Ran Spiegler. 2006. "Contracting with Diversely Naive Agents." *Review of Economic Studies* 73(3): 689–714.

Eliaz, Kfir, and Ran Spiegler. 2008. "Consumer Optimism and Price Discrimination." *Theoretical Economics* 3(4): 459–97.

Ericson, Keith M. Marzilli. 2011. "Forgetting We Forget: Overconfidence and Memory." *Journal of the European Economic Association* 9(1): 43–60.

European Commission. 2011. "More Transparency and Cheaper Roaming Prices for EU Citizens." http://ec.europa.eu/information_society/activities/roaming/regulation/archives/current_rules/index_en.htm (accessed August 29, 2011).

Farrell, Joseph. 2008. "Some Welfare Analytics of Aftermarkets." Unpublished paper.

Gabaix, Xavier, and David Laibson. 2006. "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets." *Quarterly Journal of Economics* 121(2): 505–540.

Gilpatric, Scott M. 2009. "Slippage in Rebate Programs and Present-Biased Preferences." *Marketing Science* 28(2): 229–38.

Gottlieb, Daniel. 2008. "Competition over Time-Inconsistent Consumers." *Journal of Public Economic Theory* 10(4): 673–84.

Gottlieb, Daniel, and Kent Smetters. 2014. "Lapse-Based Insurance." https://bepp.wharton.upenn.edu/files/?whdmsaction=public:main.file&fileID=8485.

Grubb, Michael D. 2009. "Selling to Overconfident Consumers." *American Economic Review* 99(5): 1770–1807.

Grubb, Michael D. 2015a. "Consumer Inattention and Bill-Shock Regulation." *The Review of Economic Studies* 82(1): 219 –57.

Grubb, Michael D. 2015b. "Failing to Choose the Best Price: Theory, Evidence, and Policy." *Review of Industrial Organization* 47(3): 303–340.

Grubb, Michael D., and Matthew Osborne. 2015. "Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock." *American Economic Review* 105(1): 234–71.

Handel, Benjamin R. 2013. "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts." *American Economic Review* 103(7): 2643–82.

Heidhues, Paul, and Botond Kőszegi. 2010. "Exploiting Naïvete about Self-Control in the Credit Market." *American Economic Review* 100(5): 2279–2303.

Heidhues, Paul, and Botond Kőszegi. 2015. "On the Welfare Costs of Naiveté in the US Credit-Card Market." *Review of Industrial Organization* 47(3): 343–56.

Heidhues, Paul, Botond Kőszegi, and Takeshi Murooka. Forthcoming. "Exploitative Innovation." *American Economic Journal: Microeconomics.*

Holman, Jeff, and Farhan Zaidi. 2010. "The Economics of Prospective Memory." Available at SSRN: http://ssrn.com/abstract=1662183.

Hossain, Tanjim, and John Morgan. 2006. "... Plus Shipping and Handling: Revenue (Non)Equivalence in Field Experiments on eBay." *B. E. Journals in Economic Analysis and Policy: Advances in Economic Analysis and Policy* 6(2): Article 3.

Kőszegi, Botond. 2014. "Behavioral Contract

Theory." *Journal of Economic Literature* 52(4): 1075–1118.

Liberty Mutual. 2014. A summary of Michael Grubb's auto insurance coverage. Message from the company sent electronically to Michael Grubb. August 2, 2014.

Lichtenstein, Sarah, Baruch Fischhoff, and Lawrence D. Phillips. 1982. "Calibration of Probabilities: The State of the Art to 1980." In *Judgment Under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Paul Slovic, and Amos Tversky, 306–334. New York: Cambridge University Press.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin. 2003. "Projection Bias in Predicting Future Utility." *Quarterly Journal of Economics* 118(4): 1209–48.

Madrian, Brigitte C. 2014. "Applying Insights from Behavioral Economics to Policy Design." *Annual Review of Economics* 6(1): 663–88.

Mullainathan, Sendhil, Joshua Schwartzstein, and William J. Congdon. 2012. "A Reduced-Form Approach to Behavioral Public Finance." *Annual Review of Economics* 4(1): 511–40.

Narasimhan, Chakravarthi. 1984. "A Price Discrimination Theory of Coupons." *Marketing Science* 3(2): 128–47.

O'Donoghue, Ted, and Matthew Rabin. 1999. "Doing It Now or Later." *American Economic Review* 89(1): 103–124.

Parrish, Leslie, and Josh Frank. 2011. "An Analysis of Bank Overdraft Fees: Pricing, Market Structure and Regulation." *Journal of Economic Issues* 45(2): 353–62.

Sandroni, Alvaro, and Francesco Squintani. 2007. "Overconfidence, Insurance and Paternalism." *American Economic Review* 97(5): 1994–2004.

Sandroni, Alvaro, and Francesco Squintani. 2013. "Overconfidence and Asymmetric Information: The Case of Insurance." *Journal of Economic Behavior & Organization* 93: 149–165.

Shui, Haiyan, and Lawrence M. Ausubel. 2005. "Time Inconsistency in the Credit Card Market." http://www.ausubel.com/creditcard-papers/time-inconsistency-credit-card-market.pdf.

Silk, Timothy Guy. 2004. "Examining Purchase and Non-Redemption of Mail-in Rebates: The Impact of Offer Variables on Consumers' Subjective and Objective Probability of Redeeming." Ph.D. Dissertation, University of Florida. http://etd.fcla.edu/UF/UFE0004380/silk_t.pdf.

Spence, Michael. 1977. "Consumer Misperceptions, Product Failure and Producer Liability." *Review of Economic Studies* 44(3): 561–72.

Spiegler, Ran. Forthcoming. "On the Equilibrium Effects of Nudging." *Journal of Legal Studies.*

Spinnewijn, Johannes. 2014. "Heterogeneity, Demand for Insurance and Adverse Selection." http://personal.lse.ac.uk/spinnewi/perceptions_welfare.pdf.

Stango, Victor, and Jonathan Zinman. 2009. "What Do Consumers Really Pay on Their Checking and Credit Card Accounts? Explicit, Implicit, and Avoidable Costs." *American Economic Review* 99(2): 424–29.

Stango, Victor, and Jonathan Zinman. 2014. "Limited and Varying Consumer Attention: Evidence from Shocks to the Salience of Bank Overdraft Fees." *Review of Financial Studies* 27(4): 990–1030.

Subbotin, Vadim. 1996. "Outcome Feedback Effects on Under- and Overconfident Judgments (General Knowledge Tasks)." *Organizational Behavior and Human Decision Processes* 66(3): 268–276.

Sundström, Anna. 2008. "Self-Assessment of Driving Skill—A Review from a Measurement Perspective." *Transportation Research Part F: Traffic Psychology and Behaviour* 11(1): 1–9.

T-Mobile. 2014. "Simple Choice Plan." http://www.t-mobile.com/cell-phone-plans/individual.html (accessed June 23, 2014).

Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness.* New Haven: Yale University Press.

Toyota. 2014. "Car, Truck, SUV & Hybrid Lease Deals." http://www.toyota.com/deals/lease/ (accessed June 23, 2014.)

Verizon Wireless. 2014. "Single Line Cell Phone Plan from Verizon - Verizon Wireless." http://www.verizonwireless.com/wcms/consumer/shop/shop-data-plans/single-line-data-plans.html (accessed June 23, 2014.)

Weinstein, Neil D. 1980. "Unrealistic Optimism about Future Life Events." *Journal of Personality and Social Psychology* 39(5): 806–820.

Weyl, E. Glen, and Michal Fabinger. 2013. "Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition." *Journal of Political Economy* 121(3): 528–83.

Journal of Economic Perspectives—Volume 29, Number 4—Fall 2015—Pages 37–60

Behavioral CEOs: The Role of Managerial Overconfidence[†]

Ulrike Malmendier and Geoffrey Tate

Chief executive officers of major corporations have often had larger-than-life images in American culture. CEOs regularly grace magazine covers and news headlines. Four times in the last 25 years *Time* magazine has selected CEOs of US corporations as their "Person of the Year": Ted Turner, Andrew Grove, Jeff Bezos, and Mark Zuckerberg. Prominent advertising campaigns have linked the images of companies with the images of their high-profile CEOs—think Lee Iacocca and Chrysler, or Dave Thomas and Wendy's. CEOs also author books to share their stories and advice with the world. Andrew Grove of Intel, for example, has written four such books: *Only the Paranoid Survive: How to Exploit the Crisis Points That Challenge Every Company*; *One-on-One with Andy Grove: How to Manage Your Boss, Yourself, and Your Coworkers*; *High Output Management*; and *Swimming Across: A Memoir*. They even appear on popular television shows (for example, Bill Gates on *Frasier* and Lee Iacocca on *Miami Vice*) and run for major political offices including President of the United States (for example, Ross Perot of Electronic Data Systems and Perot Systems in 1992 and Carly Fiorina of HP in 2016).

All of these examples suggest an image of CEOs that is deeply rooted in self-confidence, and many CEOs actively cultivate their public personas. Thus, it is not surprising that when investment decisions turn out badly, the CEO is often

■ *Ulrike Malmendier is Edward J. and Mollie Arnold Professor of Finance at the Haas School of Business and Professor of Economics, University of California, Berkeley, California. Geoffrey Tate is Associate Professor of Finance, Kenan-Flagler Business School, University of North Carolina, Chapel Hill, North Carolina. Their email addresses are ulrike@berkeley.edu and Geoffrey_Tate@kenan-flagler.unc.edu.*

[†] For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.29.4.37

doi=10.1257/jep.29.4.37

accused of overconfidence. A famous example is Steve Case's merger of AOL with Time Warner in 2000. Initially, the proposed merger was hailed as two firms "impeccably complementing each other's businesses" (reported in Dugan and Cha 2000). However, as stock prices dropped, AOL's operating performance suffered (−30 percent by mid-2002) and the merged company had to write down $54 billion in goodwill related to the deal (Peers and Angwin 2003), the failure was swiftly attributed to Steve Case being "over-ambitious," and to his "hubris" and "magical thinking" (*The Economist* 2002). The deal is now sometimes considered the "worst merger of all time" (McGrath 2015).

Other famous examples revolve around merger contests. Extended bidding wars often lead to value losses for the shareholders of the acquiring firms (Malmendier, Moretti, and Peters 2015). For example, when Sumner M. Redstone of Viacom won an extended competition with QVC to acquire Paramount in 1994, the target was deemed to be so overpriced that Viacom's stock plummeted and the company found itself under a devastating debt burden (Fabrikant 1994; Hietala, Kaplan, and Robinson 2003; Kamar 2009).

Are the accusations of managerial hubris, as voiced by practitioners and often by politicians, justified? While such events carry a whiff (perhaps even a stench) of overconfidence, they offer no systematic evidence that these CEOs in particular, or CEOs as a group, are indeed overconfident, nor do they suggest what consequences might flow from that overconfidence.

In the social and experimental psychology literatures, overconfidence and other self-serving biases have had a prominent position for many decades (for example, Svensson 1981; Miller and Ross 1975; Alicke 1985; Larwood and Whittaker 1977). By comparison, managerial biases have only begun to receive serious attention in the economics and finance literatures. Building on the early work of Roll (1986), Shefrin (2001), Statman and Sepe (1989), and Heaton (2002), the last decade has seen considerable growth in research taking this perspective. By our count, about two dozen articles in top economics and finance journals have been published on the topic since the publication of our paper Malmendier and Tate (2005), along with the appearance of a myriad of other publications and working papers. Reversing earlier resistance against the notion that top managers could make systematically biased decisions, corporate finance research has started to catch up.[1]

[1] Beyond overconfidence, studies have also analyzed a number of other decision biases of top executives. For example, Baker, Pan, and Wurgler (2012) consider the role of reference points and anchoring and show that prior stock price peaks affect mergers and acquisitions through offer prices, deal success, and bidders' announcement effects. Baker and Wurgler (2013) survey literature on how bounded rationality could explain the adoption of capital budgeting criteria, and how loss aversion could explain investment patterns such as "throwing good money after bad." Other research on managerial biases and distortions in decision-making include work on managerial risk aversion (Lewellen 2006), military experience (Malmendier, Tate, and Yan 2011; Benmelech and Frydman 2015), industry expertise (Custodio and Metzger 2013), general ability and execution skills (Kaplan, Klebanov, and Sorensen 2012), age and closeness to retirement (Jenter and Lewellen forthcoming; Yim 2013), MBA degree and working experience (Beber and Fabbri 2012), and personal frugality (Davidson, Dey, and Smith 2013; Cronqvist, Makhija, and Yonker 2012).

In this paper, we provide a theoretical and empirical framework that allows us to synthesize and assess the burgeoning literature on CEO overconfidence. We also provide novel empirical evidence that overconfidence matters for corporate investment decisions in a framework that explicitly addresses the endogeneity of firms' financing constraints.

We begin by describing the empirical approach used in much of the existing literature to measure CEO overconfidence. The most common approach, which we first established in Malmendier and Tate (2005), involves drawing implications from when CEOs choose to exercise their executive stock options. We discuss how to apply this approach to more recent data sources, and also some of the recent efforts to measure CEO overconfidence in other ways.

In considering how CEO overconfidence might affect decision-making, it is important to model an explicit decision-making framework that offers predictions about how rational CEOs will differ from overconfident CEOs. As an example of the pitfalls that can arise with a model-free approach, we point to the common intuition equating hubris with corporate overinvestment—whether in terms of internal investment (measured in capital expenditures) or external investment (merger and acquisition spending). We sketch a theoretical framework to explain why this intuition is not quite right. Instead, CEO overconfidence implies overinvestment only when the firm is flush in internal funds, and it implies a heightened sensitivity of investment to the availability of cash and other forms of capital that the CEO perceives to be relatively cheap.

We then turn to the empirical findings that have been established using the correlations between measures of CEO overconfidence and decisions that CEOs make. We review the evidence about the effects on investment, mergers, choices of internal or external financing, dividend payment levels, and other policies. We also use the workhorse regression models that are most common in this line of research to update some of our own results from a decade ago about the relationship among CEO overconfidence, available cash flow, and investment levels. Because many of the existing empirical studies, including our own, are based on correlations between overconfidence measures and corporate policies, they are subject to the standard concern that correlation does not imply causation. Given that the measurement of biases themselves is challenging, it is important to limit the sources of endogeneity that can complicate the interpretation of empirical results wherever possible. We offer some new empirical results that use the financial shock to interest rates in fall 2007 as an exogenous shock to the pricing and availability of capital. Using this shock for identification, we revisit the prediction that investments of overconfident CEOs will be particularly sensitive to the availability or cost of debt financing.

While our discussion focuses on the decision-making of biased managers in (otherwise) efficient markets and facing rational investors, there is a large parallel literature in behavioral finance that studies the decisions of rational managers operating in inefficient capital markets in which prices are affected by the decisions of irrational investors (for a survey, see Baker and Wurgler 2013). The two literatures

make opposite baseline assumptions on the source of the behavioral friction, and one may wonder how to reconcile the two literatures conceptually. We show that the implications of the two approaches are not inconsistent with each other but complement each other well. A more realistic model would allow for both types of friction and explore how they interact. We discuss ways in which future work can move in this direction.

Measuring CEO Overconfidence

In the behavioral finance literature, the most common approach to measuring CEO overconfidence has been to use decisions that the executive makes on his or her personal portfolio of company stock options. This approach, first introduced in Malmendier and Tate (2005), builds on the following logic: Since the 1980s (and particularly during the 1990s), top US executives have received increasingly large stock and option grants as part of their compensation (Hall and Murphy 2003). As a result, they commonly find themselves in a position where they are under-diversified with respect to company-specific risk. Moreover, the value of a CEO's human capital is tied to the success of the firm, heightening the under-diversification problem. CEOs and other executives have a limited ability to address this issue. For example, stock awards may be restricted such that they can only be sold after the grant "vests"—that is, either after a number of years (time-based vesting) or after certain long-term performance targets are met (performance-based vesting). Stock options are not tradeable and typically also take years to vest before they can be exercised to purchase (and then sell) the underlying stock. Moreover, executives are contractually prohibited from taking short positions in the company's stock.

Under these conditions, a rational, risk-averse executive should seek to exercise stock options, once they are vested, before expiration in order to diversify. The exact timing of optimal option exercise will vary depending on the "moneyness" of the options (that is, on how far the current price of the underlying stock exceeds the strike price at which the executive has an option to purchase the stock), the risk aversion of the executive, and the extent to which the executive is under-diversified (Lambert, Larcker, and Verrechia 1991; Hall and Murphy 2002). Overconfident executives, however, overestimate the future performance of their firms and are therefore more willing to hold options, expecting to profit from future stock price appreciation. Building on this logic, in Malmendier and Tate (2005), we proposed the systematic tendency to hold options longer before exercise as a measure of overconfidence.

More concretely, executive options typically have a ten-year lifespan at grant. While exact vesting schedules vary from one option package to the next, they are almost always fully vested after four years. Thus, CEOs who hold options all the way to expiration have taken a long-term bet on the future performance of their company's stock, despite their under-diversification. Classifying such CEOs as

overconfident is the basic idea of the "Longholder" measure we developed in earlier papers (Malmendier and Tate 2005; 2008).

In this earlier work, we classify a CEO as overconfident if that executive ever held vested options until the year of expiration, provided that the options were at least 40 percent "in the money" at the start of the final year.[2] Imposing the 40 percent threshold ensures that we do not classify CEOs as overconfident who hold "underwater" options to expiration: when an option is "underwater" the current market price of the underlying stock is lower than the strike price, so, of course, a rational executive in any model will decide against exercising such an option. In Malmendier and Tate (2005; 2008), we extensively discuss other factors that might explain why a CEO could choose to exercise an option very late during its duration—in particular, the possibility of (persistent) positive private information about future firm performance. As we show, this and other factors either do not drive the variation in the measure or do not predict the same correlations with corporate policies as overconfidence.

In our earlier work, we construct the "Longholder" measure described above using data from Hall and Liebman (1998), which covers CEO stock and option holdings in Forbes 500 companies from 1980 to 1994. In this paper, we update the Longholder measure to more recent time periods. First, we use Thomson Reuters' Insider Filings database for the 1996–2012 time period and, second, we use the detailed data on executive stock and option holdings from Compustat's Execucomp database in the format available after 2006. Our discussion provides a blueprint for updating these measures to more recent data.

Before presenting the details, we note that there are prior efforts to update the option-based overconfidence measures by other researchers. Several researchers, including Campbell, Gallmeyer, Johnson, Rutherford, and Stanley (2011), Hirshleifer, Low, and Teoh (2012), and Banerjee, Dai, Humphery-Jenner, and Nanda (2015a, b), have used the annual data on CEO vested option holdings in Execucomp's pre-2006 reporting format, beginning in 1992, to construct a measure of CEO overconfidence. Though the exact implementations vary, the general approach is to classify CEOs as overconfident when they have large enough holdings of vested options that are sufficiently in-the-money. One problem is that the pre-2006 Execucomp data does not contain details about individual option packages. For example, we do not know individual grant dates, expiration dates, or strike prices. As a result, researchers cannot assess the timing of exercise relative to expiration (or grant) dates. Instead, they typically calculate some form of approximate "average moneyness." But variation in moneyness is a direct function of stock prices. In these approaches, it is therefore crucial to include careful controls for stock returns to avoid entangling measurements of overconfidence with stock market returns.

[2] For example, options would be 40 percent "in the money" if the strike price is $100, but the stock is currently trading at $140. The 40 percent threshold comes from a calibration of the Hall and Murphy (2002) model of rational option exercise using a constant relative risk aversion coefficient of three and assuming the CEO holds 67 percent of personal wealth in company stock.

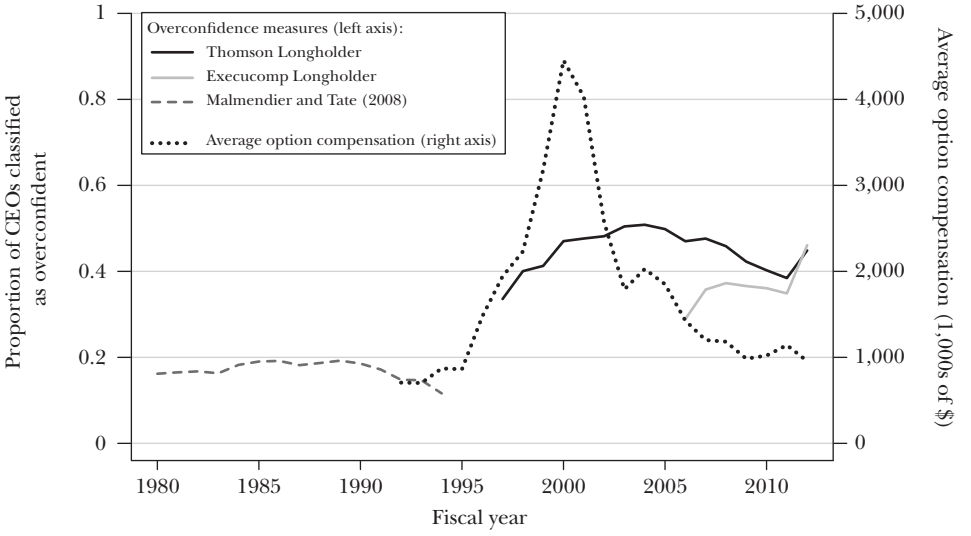
Thomson Reuters data and post-2006 Execucomp data allow researchers to avoid these potentially confounding factors. The first data source, Thomson Reuters data, provides transaction-level data on when options are exercised. Corporate insiders are required by the Securities and Exchange Commission to file what is called "Form 4" when they undertake a transaction that changes their ownership position in the company. From this data source, we retrieve all filings by CEOs that report derivative transactions (in Table 2 of Form 4) from January 1996 through 2012. We isolate the records that identify call option exercises, strike prices at which the option allows executives to purchase the stock, and exercise dates. We then merge the resulting records with monthly stock price data from the Center for Research on Security Prices (CRSP). We identify option exercises that meet two criteria: 1) the exercise occurred within one calendar year of option expiration and 2) the option was at least 40 percent in the money 12 months prior to expiration. We translate the data into a firm-year panel using Compustat's Execucomp database, which provides us with the identity of the CEO in all Standard & Poor's 1500 firms year-by-year, together with information on the CEO's aggregate holdings of company stock and options. We code the binary variable Longholder as 1 if someone working as a CEO exercised an option meeting the criteria above. Remaining CEOs are the control sample and have Longholder set equal to 0.

The second data source, Compustat's Execucomp database, includes grant-level data on CEO option holdings. In this case, we set the variable Longholder equal to one if someone working as a CEO continues to hold options during the final year before expiration, provided, again, that the options are at least 40 percent in-the-money entering the final year. The Execucomp version of the measure cannot be constructed pre-2006 because the required data became available only after a change in reporting requirements (specifically, on the DEF14A form in FAS123R). Thus, there is a relatively short time series for which we can use this measure.

Figure 1 shows the proportion of CEOs who are overconfident using these two data sources as well as the Hall and Liebman data from Malmendier and Tate (2005; 2008). Assuming that CEOs do not allow "in-the-money" options (that is, options where the strike price at which the executive can purchase the stock is lower than the market price) to expire without exercising them and that there are no errors in reporting, the two Longholder variables should be equivalent. We find that the percentage of CEOs identified as overconfident lies around 40 percent using either source. However, we also find that the correlation between the two variables is 0.56, and the source of the measurement error is difficult to identify. It is comforting to see, though, that the two measures appear to converge in the most recent years. For the empirical work later in the paper, we will carry out the exercise with each of these datasets.

The figure also reveals that the parallel measure of Longholder for the earlier period from 1980 to the mid-1990s, with the same 40 percent threshold, identifies a lower percentage of overconfident CEOs, around 20 percent in that sample. While there is no particular economic reason to expect that the proportions of CEOs classified as overconfident under the measures would remain constant over a 30-year period, some of the differences might be attributable to differences in the distributions of option-based compensation between the earlier and the later

Figure 1
Option-based Overconfidence Measure



Notes: In earlier work (Malmendier and Tate 2005; 2008), we classified a CEO as overconfident if that executive ever held vested options until the year of expiration, provided that the options were at least 40 percent "in the money" at the start of the final year. We show the "Longholder" measure from Malmendier and Tate (2008) and update it to more recent time periods, first using Thomson Reuters' Insider Filings database for the 1996–2012 time period (Thomson Longholder), and second using the detailed data on executive stock and option holdings from Compustat's Execucomp database in the format available after 2006 (Execucomp Longholder). Average option compensation is computed using data from Compustat's Execucomp database.

samples. The use of executive options was relatively limited in the 1980s and early 1990s compared to the later sample years. Both the absolute level of option compensation and its relative level (percent of compensation paid via options) increased dramatically from about 1995 to 2000 (Murphy 2013). It is conceivable that the frequency of biases differs in the broader (and younger) set of CEOs. The change in compensation may also explain the jump between the Malmendier and Tate and Thomson Longholder measures in Figure 1 on a purely mechanistic level: the increase provides for significantly more opportunities to identify a CEO as overconfident in the late 1990s and early 2000s than in the 1980s and early 1990s. The subsequent decline of option compensation[3] is also consistent with the right half of the Thomson-based graph, though less with the Execucomp measure.

Alternatively, the experience of a prolonged up-market might have instilled overconfidence in CEOs. In fact, building on the argument from Murphy (1999), the

[3] In more recent years, option compensation has been declining, while the use of restricted stock grants has started to rise. Shareholder demands, tax-law changes, and the experience of worthless options in the wake of the financial crisis are typically cited as reasons (Chasan 2013).

explosion in stock option grants might be attributable to the nearly two decades of bull market conditions after 1987. Executives who had not personally experienced a major market downturn were eager to receive option compensation, especially after having witnessed the substantial fortunes it created for many of their colleagues. As Murphy points out, there is a clear correlation between market conditions and option programs. During the decline in stock prices in the early 1970s, companies systematically replaced their "underwater" option programs with accounting-based performance plans, and they then reversed the trend during the following market upturn. If these factors are indeed a significant factor in the demand for and supply of option compensation, they would automatically be correlated with excessively optimistic beliefs, and their reversal.

Regardless of the underlying mechanism explaining the relative increase in measured overconfidence, the imperfect correlation of the parallel measures underscores the usefulness of seeking other ways to measure overconfidence. In fact, if the recent decline of option-based compensation continues, it may become impractical to use option exercise to measure beliefs in the future.

Several promising recent approaches to measuring CEO overconfidence are not tied to option exercise decisions. As a first example, Otto (2014) and Hribar and Yang (forthcoming) suggest managerial forecasts of earnings as a possible lens through which researchers can observe overconfident beliefs. Specifically, Otto (2014) measures CEO overconfidence by using the fraction of a firm's voluntary earnings forecasts that later turned out to exceed the realized earnings. He shows that this measure is positively correlated with option-based measures of overconfidence. Hribar and Yang (forthcoming) also show that CEO overconfidence captured by option-exercise behavior affects management forecasts. Unlike the failure to exercise an option over its ten-year duration, however, there are no (implicit) repeated decisions built into a management forecast. Thus, in operationalizing this measure, it will be important to require repeatedly optimistic forecasts to separate CEO bias from mistakes or simple unfavorable resolutions of uncertainty. Nevertheless, these papers provide a clear foundation for exploring earnings forecasts as a measure of overconfidence.

Second, a number of papers use portrayal in the business press as a way to measure CEO beliefs. A common approach, first used in our paper Malmendier and Tate (2008), is to count past articles in prominent business publications like the *Wall Street Journal* or *BusinessWeek* that refer to CEOs using words that suggest overconfidence ("confident"/"confidence," "optimistic"/"optimism") relative to the number of articles that refer to CEOs with words that are unlikely to suggest overconfidence ("cautious," "conservative," "practical," "reliable," and "steady"). This alternative measure tends to positively correlate with option-based measures like "Longholder," and also with the forecast-based measures (Hribar and Yang forthcoming). Given recent advances in text analysis in the empirical literature, more sophisticated implementations of this approach hold great promise for future research.

Third, some papers use survey instruments to elicit a direct measure of overconfidence. For example, Ben-David, Graham, and Harvey (2013) measure executive

biases using 10 years of a quarterly survey conducted by Duke University between 2001 and 2011, which contains projections by US chief financial officers. They use the 80 percent confidence interval of the CFOs' predictions of one- and ten-year market-wide stock returns to illustrate the miscalibration of senior executives' beliefs. They find that realized stock-market returns fall within executives' 80 percent confidence interval only 36 percent of the time, implying that financial executives are, on a whole, too confident in their projections. Graham, Harvey, and Puri (2013) directly measure the psychological traits and attitudes of senior executives. The authors designed and administered an anonymous psychometric personality test of risk-aversion, optimism, time preferences, loss aversion, and other personality traits. Survey participants included CEOs and CFOs who subscribed to *Chief Executive* magazine, *CFO* magazine, and executives who attended the World Economic Forum in Davos. (Most surveys were administered online, though some were administered through fax.) This approach has great potential to add important refinement and color to our understanding of how biases matter in firms because the questions we can answer are not limited by the availability of accounting data or even to the sample of public firms. Given that this type of methodology has a longer tradition in neighboring fields in the social sciences, there is also potential in cross-disciplinary research to exploit best practices in survey design and implementation.

Modeling the Implications of Overconfident CEOs

Following the approaches outlined above, a large body of research has used these overconfidence measures as the right-hand-side variables of interest to explain various corporate outcomes. While such an analysis seems straightforward, the theoretical predictions for how CEOs' overestimation of their own abilities (and, consequently, of the future performance of the firm under their leadership) affect corporate outcomes are oftentimes more subtle than they may at first appear. For example, the seemingly straightforward argument that an overconfident CEO will tend to overinvest is incomplete in a way that could be misleading. Here, we will sketch a simple decision-making model to illustrate that hubris does *not* necessarily imply "too much investment" or "too many mergers" (which could be mis-inferred from Roll 1986). Instead, it induces a tradeoff between misperceived gains from investment and the costs of external financing. As empirical work progresses in this area, it may prove useful to model more explicitly the decision-making problem of the CEO in other contexts to avoid similar potential pitfalls.

We consider a unifying framework that seeks to capture how managerial overconfidence affects corporate investment, including both internal and external projects as well as the firm's capital structure.[4] Because our goal is to demonstrate the distortionary power of overconfidence, we abstract from informational

[4] An algebraic exposition of this model in the spirit of Malmendier and Tate (2005) appears in the online Appendix available with this paper at http://e-jep.org.

asymmetries and agency problems and assume that the firm's manager maximizes current shareholder value in an efficient capital market.

We define overconfidence as the overestimation of the value a manager believes he or she can create. This bias manifests itself in two forms. First, an overconfident manager believes that the company's current assets are undervalued by the market. Second, an overconfident manager overestimates the value of future potential investments he or she might pick.

The CEO makes two decisions: 1) the level of investment, including both internal investment (capital expenditure) and external investment (mergers); and 2) whether financing of the investment will rely on internal cash flow (or, equivalently, on riskless debt) versus external equity capital.

In this setting, an overconfident CEO believes that the market price for riskless debt financing is accurate, but disagrees with potential new shareholders as to what an equity stake in the firm is worth, and hence on the appropriate price of newly issued shares. As a result, the model generates two main testable predictions about the difference in behavior between a rational and an overconfident CEO.

Prediction 1: The investment of overconfident CEOs is more sensitive to the availability of internal cash flow than the investment of CEOs who are not overconfident.

Rational CEOs are indifferent between all available sources of capital for an investment project. They view the price of equity (and, in a more general model, of risky debt) as set appropriately in the market. Thus it doesn't matter whether an investment is paid for with internal cash, or if the firm raises funds by issuing stock (or, in a more general model, risky corporate bonds) at a fair market price. Rational CEOs' investment decisions exclusively depend on whether the project generates value for existing shareholders.

Overconfident CEOs, instead, shy away from risky external equity capital, because they believe that the market is setting the price of equity in the firm too low and they are trying to protect existing shareholders from (perceived) dilution. As a result, overconfident CEOs first exhaust internal sources of financing before turning to risky external financing, and investment choices deviate from the first best as they depend to a greater extent on whether internal cash flow is available.

A key insight is that overconfident CEOs do not necessarily overinvest even though, by definition, they overestimate the returns to their internal investment or merger projects. It is true that this overestimation unambiguously induces the CEOs to prefer a higher level of investment than optimal. However, they will invest according to this preference only if they can access internal funds or riskless debt, the pricing of which they do not perceive to be distorted. Their calculation changes when they have to finance the investment by issuing equity. The need to access financing that they misperceive to be more costly than it is curbs their desire to overinvest. Depending on whether a CEO is more overconfident about what the current stock price of the firm should be or what the investment return will be, the ultimate decision could be to invest either above or below the first-best level. In other words, overconfident CEOs only overinvest if their company is flush with internal funds.

Rational CEOs do not make this mistake. They understand the equivalence of external and internal financing, and thus choose the value-maximizing level of investment regardless of the type of financing available to their firm.

While it is difficult to test for over- and underinvestment directly since it is hard to measure the right (optimal) level of investment, it is feasible to test the investment–cash flow prediction spelled out above.

We can also derive a second testable prediction, which further links the cash-flow dependence to the firm's financial constraints:

Prediction 2: The investment–cash flow sensitivity of overconfident CEOs is more pronounced in equity-dependent firms.

As spelled out above, the overconfident CEO overinvests when internal sources of finance are abundant. Hence, in a cash-rich situation, variations in investment will be less correlated with the exact amount of (abundant) internal funds. It is only when the limitation of internal funding is binding, relative to the desired level of investment, that the extra-degree of investment–cash flow sensitivity kicks in.

Although the simple version of our model does not include a separate parameter for the riskiness of debt, the argument can easily be extended. Moreover, it can be used directly to illuminate the importance of a firm's debt capacity. Think of available debt financing as essentially riskless debt—maybe in the form of a credit line. The higher is the amount of riskless debt the firm can utilize, the more likely are such nondiluting forms of financing to be sufficient for any desired level of investment. Thus, we expect to observe that investment–cash flow sensitivity is more pronounced in equity-dependent firms with overconfident CEOs than in firms with untapped debt capacity.

In short, the main takeaway of this model is that the positive relation between CEO overconfidence and investment is not unambiguous once we also explicitly model the CEO's interactions with (potentially) unbiased agents in external capital markets. In this richer environment, the CEO faces a tradeoff between the perceived benefits of investment and the perceived costs of raising unduly expensive financing. Instead, the clear prediction of the model is that investment is sensitive to the availability of cash or other forms of capital that the CEO perceives to be relatively cheap.

The Evidence on Overconfidence and Corporate Investment

A decade of empirical work using various measures of CEO overconfidence is now yielding an intriguing array of results. We begin this section by reviewing some of the most prominent papers, including studies that test the predictions of the basic model in the previous section, along with other implications of overconfidence for corporate policies. We then reanalyze one concrete example of empirical work in this area, the prediction that the investment of overconfident CEOs is more sensitive to the costs of external finance. We first revisit our investment–cash flow sensitivity model from Malmendier and Tate (2005) using more modern data, from 1996 to 2012. This analysis will use the kind of regression model that has been a workhorse in

this literature. We then exploit a natural-experiment design in the spirit of Almeida, Campello, Laranjeira, and Weisbenner (2012) to isolate plausibly exogenous exposure to a shock to external financing costs, in order to demonstrate that the model can be verified using more robust and current econometric techniques.

Some Main Results in the Previous Literature

The model of decision-making by overconfident CEOs in the previous section predicts that corporate investment in firms run by overconfident executives should be more sensitive to the availability of internal resources than the investment of other firms, in particular at times when the firm is facing financial constraints. In Malmendier and Tate (2005), we test this prediction by applying the traditional investment–cash flow sensitivity model of Fazzari, Hubbard, and Peterson (1988) and Kaplan and Zingales (1997). The key result is that the relation between investment and the interaction of various overconfidence measures with cash flow is estimated to be significantly positive, implying a significantly higher sensitivity of investment to internal resources among overconfident CEOs than among rational peers, which is consistent with Prediction 1 in the previous section. The cash flow sensitivity also appears most prominently among equity-dependent firms—measured using a variety of common proxies from the literature, including firm size, firm age, and the Kaplan-Zingales index—which is consistent with Prediction 2.

In Malmendier and Tate (2008), we use a similar dataset in the context of mergers, and show that overconfident CEOs also tend to have a higher tendency to undertake mergers and, particularly, diversifying deals, especially if they have access to internal financing. This is again consistent with Prediction 2. The market reaction to merger announcements of overconfident CEOs is also significantly more negative than that of non-overconfident CEOs, pointing to the suboptimality of overconfident CEOs' chosen level of (external) investment. Outside the merger context, the optimality of investment is hard to assess.

Several papers in the literature consider capital-structure decisions. Deshmukh, Goel, and Howe (2013) examine the interaction of overconfidence and dividend policy and show that firms with overconfident CEOs generally pay out smaller dividends, which is consistent with overconfident CEOs viewing external financing as costly and wishing to build financial slack for future investment needs. In Malmendier, Tate, and Yan (2011), we show that overconfident managers use less external finance and, conditional on accessing external capital, issue less equity (more debt) than their peers, consistent with overconfident managers' aversion to equity financing.

Another strand of this work relates CEO overconfidence to earnings management and the incidence of earnings misstatements. Because overconfident CEOs overestimate future earnings, they borrow more aggressively against future earnings to avoid missing current earnings forecasts and generally practice less-conservative accounting practices—for example, in delaying recognition of losses (Bouwman 2014; Ahmed and Duellman 2013). Schrand and Zechman (2012) find that overconfident executives are more likely to exhibit an optimistic bias that leads to

nonintentional earnings misstatement, which can raise their incentive to misstate earnings intentionally in subsequent periods.

Though the studies described thus far all point to potential value-destruction as a result of CEO overconfidence, another strand of this work finds evidence of a "bright side" to overconfidence. Galasso and Simcoe (2011) and Hirshleifer, Low, and Teoh (2012) link managerial overconfidence with innovation, finding that overconfidence is generally beneficial for a firm's innovation performance, and especially so in competitive and innovative industries. These benefits can help us begin to understand why we consistently observe overconfidence among CEOs over time.

Building on this point, another strand of literature considers explicitly the implications of overconfidence for firms' hiring and compensation practices. Though one possibility is that firms cannot recognize overconfidence in CEO candidates, Banerjee, Dai, Humphery-Jenner, and Nanda (2015a) find that firms choose overconfident CEOs at times when the predictable consequences of overconfidence on policies—like high levels of investment given abundant internal funds—are likely to benefit the firm. They provide evidence that overconfident senior executives are indeed more likely to be selected as CEOs and that they tend to enhance firm performance during the period when there is a change in strategy for more mature firms. Building on the theoretical insight of Goel and Thakor (2008), who argue that moderate biases can increase value by mitigating the effects of risk aversion, Campbell, Gallmeyer, Johnson, Rutherford, and Stanley (2011) show that CEOs with relatively low or relatively high optimism face a higher probability of forced turnover than moderately optimistic CEOs, which is consistent with the view that there is an interior optimum level of managerial optimism that maximizes firm value. Otto (2014) finds that firms are able to profit from the overconfidence of CEOs when providing equity-linked or bonus compensation: they provide CEOs who overestimate the future value of the firm's equity with smaller grants and fewer bonus payments—which illustrates one way in which sophisticated principals can take advantage of biased agents.

The diversity of applications in which these measures of CEO overconfidence are associated with the predicted outcome assuages concerns about how overconfidence is being captured in any one particular empirical setting. Several recent papers also take the promising direction of looking beyond broad data aggregates—like net debt issuance or corporate investment—and instead study micro-level variation in contract features. For example, Adam, Burg, Scheinert, and Streitz (2014) link overconfidence with the corporate choice to issue performance-sensitive debt. Sunder, Sunder, and Tan (2010) study the inclusion of covenants in the debt contract, and provide evidence that creditors appear to recognize behavioral biases of the managers with whom they interact in the market.

Overconfidence and Investment–Cash Flow Sensitivity since 1996

Below, as an illustration, we run through a test for the effect of CEO overconfidence on corporate outcomes in the context of corporate investment decisions. First, we use a regression approach of the kind commonly used in this literature and update the investment–cash flow sensitivity findings from Malmendier and Tate

(2005) using data from 1996 to 2012. Second, we provide one example of how to combine the best available overconfidence measures with better identification of corporate outcomes, following the supply-shock approach of Almeida et al. (2012).

We obtain data on firm fundamentals from the Compustat database and data on stock prices from the CRSP database. The dependent variable in the analysis is corporate investment, which we measure using capital expenditures scaled by the lagged value of net property, plants, and equipment. The main independent regressors are: cash flow (measured as the sum of earnings and depreciation scaled by the lagged value of net property, plants, and equipment); Tobin's q (measured as the ratio of the market value to the book value of assets); firm size (measured as the natural logarithm of total assets); CEO stock ownership (measured as the percentage of company stock held by the CEO); CEO vested option holdings (measured as the number of CEO vested options scaled by the number of outstanding common shares); and an indicator variable that equals one if the board of directors has between four and twelve members. The last-mentioned proxy builds on evidence that smaller boards of directors are associated with higher market valuations (for example, Yermack 1996) and provides a crude but readily available proxy for the within-firm variation in corporate governance over time that we cannot capture by including firm fixed effects. We include the latter five regressors both in levels and interacted with cash flow. We use lags of all of the independent regressors to ensure that their values predate the firm's investment choice. We also include year fixed effects and their interaction with cash flow to capture variation in macroeconomic conditions during the 17-year sample period that might correlate both with investment and the presence (or measurement) of CEO overconfidence. Similarly, we include firm fixed effects to capture unobserved time-invariant differences between firms. Because the data contain severe outliers, we drop observations in the upper and lower 1 percent of the distributions of cash flow, Tobin's q, and investment. We cluster standard errors at the firm level.

The full results of this regression appear in the online Appendix available with this paper at http://e-jep.org. Table 1 presents estimated coefficients for the main variables. Our main explanatory variables of interest are the Longholder variable measuring CEO overconfidence and its interaction with cash flow. Remember that the model presented earlier predicts a positive coefficient on the interaction, since overconfident CEOs will tend to overvalue investment projects but also view external finance as unduly costly. Here we use the version of Longholder constructed from Thomson Reuters data to be able to exploit the 17-year sample period on which it is available.[5]

In column 1 of Table 1, we provide estimates of the regression model. Consistent with prior research, we find firms with higher values of Tobin's q and firms with higher cash flows invest significantly more than other firms. We also confirm

[5] We do not find a significant effect of overconfidence in the investment–cash flow sensitivity model using the Execucomp-based measure during the Great Recession years. However, it is likely that all firms were investment–cash flow sensitive during this period, reducing the ability to differentiate between firms in the cross-section.

Table 1
CEO Overconfidence and the Sensitivity of Corporate Investment to Cash Flow

			Firms with:			
	Full sample		*Speculative-grade rating and unrated debt*		*Investment-grade debt rating*	
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Q	0.046***	0.036***	0.046***	0.034***	0.038***	0.035***
	(0.004)	(0.005)	(0.005)	(0.007)	(0.005)	(0.006)
Size	−0.046***	−0.042***	−0.059***	−0.050***	−0.025***	−0.015
	(0.007)	(0.008)	(0.009)	(0.012)	(0.007)	(0.010)
Stock Ownership	0.054	−0.018	0.033	−0.056	0.150	0.177
	(0.049)	(0.061)	(0.053)	(0.065)	(0.095)	(0.163)
Vested Options	0.273	0.226	0.217	0.273	0.528	−0.083
	(0.284)	(0.348)	(0.332)	(0.446)	(0.516)	(0.534)
Efficient Board Size	−0.001	−0.007	−0.010	−0.031	−0.001	−0.005
	(0.006)	(0.007)	(0.014)	(0.022)	(0.006)	(0.006)
Longholder	−0.006	−0.006	−0.011	−0.029*	−0.002	0.005
	(0.007)	(0.008)	(0.011)	(0.015)	(0.007)	(0.009)
Cash Flow (CF)	0.057**		0.036		0.056*	
	(0.022)		(0.035)		(0.031)	
Longholder × CF	0.012**	0.020*	0.019**	0.046***	0.000	−0.004
	(0.006)	(0.011)	(0.008)	(0.015)	(0.008)	(0.019)
Controls × CF fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year × CF fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm × CF fixed effects	No	Yes	No	Yes	No	Yes
R^2	0.195	0.447	0.189	0.473	0.233	0.440
N	14,615	14,615	9,088	9,088	5,527	5,527
Number of firms	2,341	2,341	1,808	1,808	858	858

Notes: The sample consists of firms in Compustat's Execucomp database between 1997 and 2012. The table reports linear regression coefficients. The dependent variable is investment, measured as capital expenditures scaled by the lag of net property plants and equipment. *Q* is Tobin's *q*, the market value of assets scaled by the book value of assets. Stock Ownership (Vested Options) is the CEO's holdings of company stock (vested options) scaled by common shares outstanding. Efficient Board Size is an indicator variable equal to 1 if the firm's board has between 4 and 12 members. Cash Flow is the sum of earnings and depreciation scaled by the lag of net property plants and equipment. Longholder is an indicator variable equal to 1 if the CEO during his tenure held an option to the last year before expiration, provided it was at least 40 percent in-the-money entering its final year. We define the Longholder variable using exercise data from the Thomson Reuters insider filings database. All standard errors are adjusted for firm-level clustering.
***, **, and * indicate statistical significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

the prediction of the theoretical model (Prediction 1) that investment–cash flow sensitivity is heightened among overconfident CEOs. Here, the magnitude of the estimates is modest: a one standard deviation increase in cash flow increases

investment by 8 percent of a standard deviation more among overconfident CEOs than among rational peers.

In column 2, we augment the regression specification by allowing the sensitivity of investment to cash flow to vary by firm—that is, we include the interactions of firm fixed effects with cash flow. This specification allows us to separate variation in investment–cash flow sensitivity that is due to managerial overconfidence from variation that may be driven by time-invariant firm effects of firms with overconfident managers. In the augmented specification, we identify the interaction of Longholder and cash flow using only cases in which we can compare the sensitivity of CEOs with different Longholder values who have operated the same firm. In Malmendier and Tate (2005), we were not able to estimate this specification due to the shorter time series and much smaller cross-section of firms in the data. We continue to estimate a heightened sensitivity of investment to cash flow among overconfident CEOs in this more conservative specification; indeed, the magnitude of the effect is nearly double what we find in column 1.

In Malmendier and Tate (2005), we argue that the overconfidence model also produces a nuanced prediction that CEOs with access to cheap debt financing should not exhibit heightened sensitivity of investment to the availability of cash in the firm because they can also finance desired (over)investment by tapping unused debt capacity. It is instead "equity-dependent" or financially constrained firms in which we should observe the effects of overconfidence on investment financing, as also spelled out in Prediction 2 of our model. In the remainder of Table 1, we test this second prediction by splitting the sample of firms based on whether the firm has an investment-grade long-term debt rating. We then rerun the specifications from columns 1 and 2 on the resulting subsamples of the data. In columns 3 and 4 of Table 1, we present the results for the subsample of firms without access to investment grade debt. We find, as predicted, a positive effect of the overconfidence measure on the sensitivity of investment to internal resources that is even stronger than the estimates in the first two columns, both economically and statistically. The estimate in column 4, which includes the interaction of firm fixed effects with cash flow, is roughly 4 times the size of the baseline estimate in column 1. In the last two columns, we provide the complementary estimates for firms that have investment-grade debt ratings. Here we do not find any difference between the investment to cash flow sensitivities of firms with and without CEOs we classify as overconfident. Apparently, access to cheap debt capital already allows CEOs the ability to invest at desired levels or, put differently, perceived financing constraints do not bind.

Taken as a whole, this regression evidence corroborates and in some ways strengthens the evidence from Malmendier and Tate (2005). For example, here we are able to separate definitively the effect of the Longholder measure from time-invariant differences across firms in cash flow sensitivities. It is intriguing that the estimated effect of cash flow on investment is substantially smaller over the last two decades than estimates using earlier data: for comparison, see Kaplan and Zingales (1997) for evidence from the 1970s and 1980s. Whether this reflects an

increase in the efficiency of financial markets, changes in CEO choices, or a failure of the econometric model is an open question.

It is natural to have some concerns about the overall validity of this type of model formulation and, in particular, about the endogenous availability of internal financing sources. Moreover, we note that our choice to measure cash flow over the prior fiscal year—to ensure that our measure of cash coming into the firm predates the firm's investment choice—appears to matter for the results. The connection between overconfidence, cash flow, and firm investment is no longer statistically significant if we measure investment and cash flow simultaneously, as many prior papers in the literature have done. This variation of the key results from what may seem to be innocuous choices in specification underscores the importance of taking measurement and endogeneity seriously. In practice, this means attempting to identify new and robust settings in which to measure the effects of CEO biases on corporate choices.

Overconfidence and an Exogenous Shock

We examine one such setting, using a recent shock to corporate debt markets to provide exogenous variation in the availability of debt financing. Almeida et al. (2012) describe a supply shock that occurred in corporate credit markets beginning in August 2007, at the outset of the financial crisis. Over the three prior years, average spreads on both investment-grade and high-yield bonds were generally stable and, if anything, mildly decreasing. However, in August 2007, the shock to credit markets that started with a decline in housing prices in 2006 and a wave of subprime mortgage defaults began to show up in the prices of long-term corporate bonds. By early 2008, spreads on investment-grade corporate bonds had risen from roughly 1 percent to nearly 3 percent; similarly, spreads on high-yield corporate bonds rose from roughly 3 percent to the range of 7–8 percent. To be clear, these changes occurred before the Great Recession had a chance to take hold, and well before the Lehman bankruptcy and the other economic catastrophes in September 2008. To identify the effect of a shock to financing constraints on corporate investment, Almeida et al. (2012) exploit differences across firms in the portion of long-term debt that matured just after the credit-market shock hit. Specifically, they identify as "treated" those firms that had at least 20 percent of their long-term debt maturing in 2008 and thus faced the choice of refinancing or raising alternative sources of finance during the credit market disruption in late summer and fall of 2007. For their key test, Almeida et al. compare treated firms to a matched sample of firms without large amounts of debt maturing in the crisis period, finding that firms affected by the shock to financing constraints significantly cut investment, in absolute terms and relative to control firms.

For our purposes, a simpler approach will suffice. We exploit variation within the set of firms affected by the credit market shock, computing the difference-in-differences of investment between firms run by overconfident and rational CEOs. To measure overconfidence, we employ again the Longholder measure—this time using both the Thomson Reuters and the Execucomp data discussed earlier. We also allow for variation in the threshold percentage of maturing long-term debt required

for treatment, measuring the difference in treatment effects for thresholds of 15, 20, 25, and 30 percent.

We estimate differences-in-differences between firms run by overconfident and rational CEOs using a panel regression. Each firm accounts for two observations in the regression, one prior to the shock and one after. Following the specification in Almeida et al. (2012), the dependent variable is the average of a firm's quarterly capital expenditures relative to a lagged value of capital, for the average of the first three quarters of 2007 and 2008 in the prior and after periods, respectively. We measure the difference in investment by including an indicator variable called "After," which takes the value one for the "After" observation. The difference-in-differences is captured by the interaction of After with the Longholder overconfidence measure. We cluster the standard errors by firm.

In panel A of Table 2, we present the results using the version of Longholder computed from Thomson Reuters insider filings data. Looking at the interaction of the After and the Longholder variables, we find a larger drop in investment expenditures following the financing shock among overconfident CEOs that increases in magnitude monotonically as we raise the threshold for sample inclusion—that is, as we compare firms that had greater exposure to the credit market shock. The difference-in-differences are statistically significant for thresholds of 25 and 30 percent of long-term debt maturing just after the shock, with p-values of 0.059 and 0.056, respectively. The effects are also economically large. The economic magnitude of the difference-in-differences increases monotonically with the threshold, from 7.5 percent of a standard deviation of investment in column 1 to 55 percent of a standard deviation in column 4.

In panel B, we report the results using the version of Longholder computed from Execucomp data. We find the same qualitative pattern. Here, the larger drop in investment following the shock among overconfident CEOs is statistically significant for all the thresholds for sample inclusion. The results are also stronger economically: the difference-in-differences ranges from 34 percent of a standard deviation of investment in column 5 to 58 percent of a standard deviation in column 8.

Overall, the results confirm the theoretical prediction that the investment of overconfident CEOs is more sensitive to external financing costs than is the investment of rational CEOs. This empirical methodology has the advantage of allowing us to sidestep the endogeneity of cash flow in standard investment–cash flow regressions. Nevertheless, the approach has some limitations. First, we are using only a small subsample of the data on which we observe clean variation around the shock, raising the question of whether the estimates are more broadly valid. Second, exogenous shocks are likely to be relatively rare, meaning that similar opportunities for identification in other samples are not guaranteed. Third, while our approach is an improvement in the sense that it uses an exogenous financing shock and thus clarifies the identification issues, it does not address potential error in the measurement of overconfidence. Despite these limitations, we view the analysis as representative of the type of approach that is necessary to demonstrate the importance of behavioral factors within the broader corporate finance and economics literatures.

Table 2

CEO Overconfidence and the Response of Corporate Investment to the Credit Crisis of 2007

% Long-term debt maturing in 2008:	*Panel A. Thomson Reuters data*				*Panel B: Execucomp data*			
	15% *(1)*	*20%* *(2)*	*25%* *(3)*	*30%* *(4)*	*15%* *(5)*	*20%* *(6)*	*25%* *(7)*	*30%* *(8)*
After	−0.001	0.000	0.004	0.008	0.002	0.000	0.002	0.005
	(0.003)	(0.004)	(0.006)	(0.005)	(0.003)	(0.004)	(0.006)	(0.005)
Longholder	0.009	0.007	0.010	0.011	0.011	0.008	0.017	0.026
	(0.009)	(0.012)	(0.018)	(0.025)	(0.011)	(0.015)	(0.021)	(0.028)
Longholder × After	−0.003	−0.011	−0.021*	−0.029*	−0.013*	−0.018**	−0.025*	−0.030*
	(0.006)	(0.007)	(0.011)	(0.014)	(0.007)	(0.009)	(0.013)	(0.017)
Constant	0.052***	0.055***	0.061***	0.060***	0.052***	0.054***	0.059***	0.055***
	(0.006)	(0.008)	(0.010)	(0.011)	(0.005)	(0.006)	(0.009)	(0.010)
R^2	0.010	0.009	0.015	0.020	0.011	0.015	0.021	0.033
N	185	131	92	64	177	127	88	61
Number of firms	95	67	47	33	92	66	46	32

Notes: The sample contains two observations, one from 2007 and one from 2008, for each firm in Compustat's Execucomp database that had x percent of its long-term debt expiring in 2008, where x is indicated in the column heading. See Almeida et al. (2012) for additional sample filters. The table reports linear regression coefficients. The dependent variable is the average of quarterly investment scaled by the lag of net property, plants, and equipment for the first three quarters of the year. After is an indicator variable equal to one in 2008. Longholder is an indicator variable equal to 1 if the CEO during his tenure held an option to the last year before expiration, provided it was at least 40 percent in-the-money entering its final year. In panel A, we define Longholder using exercise data from the Thomson Reuters insider filings database. In panel B, we define Longholder using the data on outstanding stock option grants from the Execucomp database. Standard errors are adjusted for firm-level clustering.

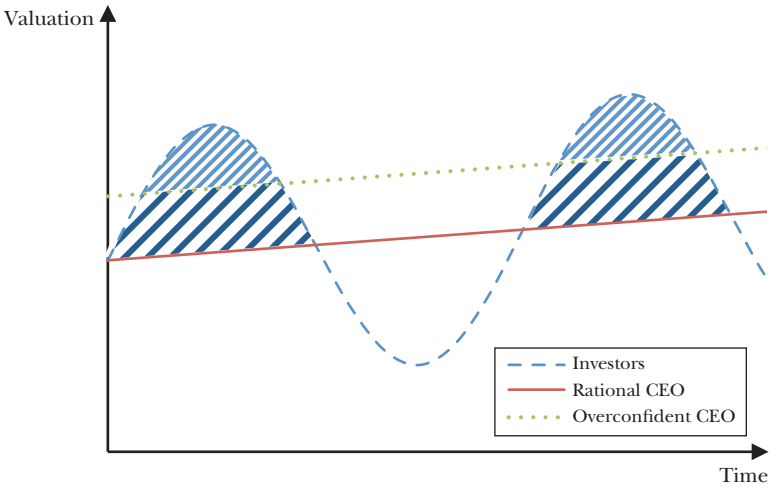
***, **, and * indicate statistical significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Biased Managers or Biased Investors?

Our discussion has focused entirely on the biases of top executives. Early research in behavioral finance, however, tended to focus almost exclusively on the biases of individual investors (for example, Odean 1998; Odean 1999; Barber and Odean 2000; and also the Daniel and Hirshleifer contribution in this symposium). Later studies considered how these individual biases could be exploited by rational market participants—including rational managers of firms. Examples include studies of how financial analysts exploit the naiveté of small investors with excessively positive recommendations; of how financial advisors cater to the return-chasing behavior of investors by offering stocks with strong recent performance; and of how credit companies exploit consumers' self-control problems with teaser rates that automatically increase after six months (Malmendier and Shanthikumar 2007; Li 2015; Ausubel 1999; see also Grubb's contribution to this symposium).

Indeed, this line of argument suggested that a number of corporate finance puzzles can be understood as an outcome of rational managers exploiting investor biases. For example, if investors sometimes overestimate and sometimes underestimate the value of a firm, a rational manager who maximizes value to existing shareholders should concentrate equity issues to occur during periods of positive

Figure 2
Illustration of Differences in Firm Valuation



Source: Authors.

sentiment. Baker and Wurgler (2000; 2002) provide evidence that firms tend to issue relatively more equity than debt just before periods of low market returns and shun equity, in favor of debt, before periods of high returns. Their results suggest that managers exploit the inefficiency of the stock market and that market timing has large and persistent effects on capital structure.

But if top corporate executives seek to exploit irrational investors, are such findings in conflict with the hypothesis that some managers are overconfident and persistently overestimate the value of their firm? The answer is no. With Figure 2, we aim to illustrate intuitively that the two biases can go hand in hand. The figure shows a stylized, hypothetical example of the discrepancies between a firm's true value over time, as it would be assessed by a rational CEO (solid line); an overconfident CEO's assessment (dotted line), which persistently lies above the rational assessment; and investors' assessment (dashed line), which lies sometimes above the rational assessment (positive investor sentiment) and sometimes below the rational assessment (negative investor sentiment).

The figure illustrates that both types of biases—persistent managerial overconfidence and time-varying investor sentiment—can be combined to capture a broader range of facts. Overconfidence induces the CEO to overestimate the value of the firm relative to the market much of the time, and relative to its true value all the time. Hence, the CEO's investment and financing choices will be biased in the ways we discussed above. At the same time, during periods of very positive investor sentiment, investors' assessment of the firm might be even more optimistic than that of the CEO. Such peak periods of investor sentiment produce distorted access to equity financing and allow for periods of market overvaluation (narrowly shaded

areas), as predicted by Baker and Wurgler (2000, 2002), albeit shorter ones than in a world with (only) rational managers (narrowly shaded plus widely shaded areas).

In short, there is no inherent inconsistency between these two strands of behavioral corporate finance. For further discussion of these biases and how they can interact, see also the survey articles by Barberis and Thaler (2003), Baker, Ruback, and Wurgler (2007), and Baker and Wurgler (2013).

Discussion and Conclusion

A large and growing body of evidence suggests that a substantial share of top corporate executives exhibit symptoms of overconfidence in their decisions. The main measure of CEO overconfidence used here has been the willingness of CEOs to keep their personal wealth undiversified by holding stock options until very close to their expiration. Other measures of CEO overconfidence include earnings forecasts, survey responses, and even psychometric tests. The presence of CEO overconfidence—that is, of a belief by the top executive that the price of the firm's stock should be higher than it is—seems to matter for a variety of firm decisions and the choice of financing for those decisions. Notably, it matters for the extent to which investment choices, both those involving internal investment and external mergers, track the available cash and easy-to-obtain debt available to firms. But CEO overconfidence also appears to be correlated with other choices like paying less in dividends and relying less on external equity-based finance. Other research suggests that firms are able to identify overconfident CEOs, and firms that plan to undertake a change in strategy or to vigorously pursue innovation may prefer a degree of overconfidence. Moreover, firms can offer overconfident CEOs lesser amounts of company stock as part of their compensation packages. This body of research keeps expanding, with additional measures of CEO overconfidence, theoretical models linking overconfidence to various practices and outcomes, and empirical tests that pay more attention to exogeneity and identification of cause and effect.

The immediate focus of this literature is often to identify specific instances in which CEO overconfidence matters. But in a broader sense, this literature is proposing and exploring a different approach to the key agency problem of the firm, involving the separation of ownership and control. The classic statement of this problem is that managers of firms have an incentive to pursue their own private benefits, while shareholders and other claimholders of the firm would like the manager to act as an agent for their interests. The evidence here suggests that even when managers intend to maximize claimholder value, they can fail to do so because they hold overconfident beliefs. To the extent that CEO overconfidence (and other potential biases) matter, then some approaches to corporate governance will be more robust than others. Providing equity-linked compensation may mitigate traditional misalignment of CEO incentives but is unlikely to affect the choices of biased managers, who already believe they are maximizing value. On the other hand, measures to improve active monitoring of the firm—to the extent that the

monitors themselves are rational—may address both concerns. Recent work has begun to look seriously at how and when governance can mitigate managerial biases (Banerjee, Humphery-Jenner, and Nanda 2015b; Kolasinski and Li 2013), which seems to offer a promising direction for additional research. Ultimately, behavioral biases like CEO overconfidence matter not just for the choices and outcomes of the agents who are subject to them, but also for the (potentially rational) agents with whom they interact, transact, and contract in the market place.

References

Adam, Tim R., Valentin Burg, Tobias Scheinert, and Daniel Streitz. 2014. "Managerial Optimism and Debt Contract Design: The Case of Syndicated Loans." Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?.

Ahmed, Anwer S., and Scott Duellman. 2013. "Managerial Overconfidence and Accounting Conservatism." *Journal of Accounting Research* 51(1): 1–30.

Alicke, Mark D. 1985. "Global Self-Evaluation as Determined by the Desirability and Controllability of Trait Adjectives." *Journal of Personality and Social Psychology* 49(6): 1621–30.

Almeida, Heitor, Murillo Campello, Bruno Laranjeira, and Scott Weisbenner. 2012. "Corporate Debt Maturity and the Real Effects of the 2007 Credit Crisis." *Critical Finance Review* 1(1): 3–58.

Ausubel, Lawrence M. 1999. "Adverse Selection in the Credit Card Market." Working paper, June 17. http://www.ausubel.com/creditcard-papers/adverse.pdf.

Barber, Brad, and Terry Odean. 2000. "Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." *Journal of Finance* 55(2): 773–806.

Barberis, Nicholas, and Richard Thaler. 2003. "A Survey of Behavioral Finance." Chap. 18 in *Handbook of the Economics of Finance* Vol. 1B, edited by G.M. Constantinides, M. Harris, and R. Stulz. North Holland: Elsevier Science B.V.

Baker, Malcom, Xin Pan, and Jeffrey Wurgler. 2012. "The Effect of Reference Point Prices on Mergers and Acquisitions." *Journal of Financial Economics* 106(1): 49–71.

Baker, Malcom, Richard S. Ruback, and Jeffrey Wurgler. 2007. "Behavioral Corporate Finance. Chap. 4 in *Handbook of Corporate Finance: Empirical Corporate Finance Vol. 1,* edited by B. Espen Eckbo. New York: Elsevier/North-Holland.

Baker, Malcom and Jeffrey Wurgler. 2000. "The Equity Share in New Issues and Aggregate Stock Returns." *Journal of Finance* 55(5): 2219–57.

Baker, Malcom and Jeffrey Wurgler. 2002. "Market Timing and Capital Structure." *Journal of Finance* 57(1): 1–32.

Baker, Malcolm and Jeffrey Wurgler. 2013. "Behavioral Corporate Finance: An Updated Survey." Chap. 5 in *Handbook of the Economics of Finance Vol. 2A*, edited by George M. Constantinides, Milton Harris, Rene M. Stulz. North Holland: Elsevier.

Banerjee, Suman, Lili Dai, Mark Humphery-Jenner, and Vikram Nanda. 2015a. "Top Dogs: Overconfident Executives and New CEO Selection." Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2371435.

Banerjee, Suman, Mark Humphery-Jenner, and Vikram Nanda. 2015b. "Restraining Overconfident CEOs through Improved Governance: Evidence from the Sarbanes-Oxley Act." *Review of Financial Studies* 28(10): 2812–58.

Beber, Alessandro, and Daniela Fabbri. 2012. "Who Times the Foreign Exchange Market? Corporate Speculation and CEO Characteristics." *Journal of Corporate Finance* 18(5): 1065–87.

Benmelech, Efraim, and Carola Frydman. 2015. "Military CEOs." *Journal of Financial Economics* 117(1): 43–59.

Ben-David, Itzhak, John R. Graham, and Campbell R. Harvey. 2013. "Managerial Miscalibration." *Quarterly Journal of Economics* 128(4): 1547–84.

Bernardo, Antonio E., and Ivo Welch. 2001. "On the Evolution of Overconfidence and Entrepreneurs." *Journal of Economics & Management Strategy* 10(3): 301–30.

Bouwman, Christa. 2014. "Managerial Optimism and Earnings Smoothing." *Journal of Banking and Finance* 41(C): 283–303.

Campbell, T. Colin, Michael Gallmeyer, Shane A. Johnson, Jessica Rutherford, and Brooke W. Stanley. 2011. "CEO Optimism and Forced Turnover." *Journal of Financial Economics* 101(3): 695–712.

Chasan, Emily. 2013. "Last Gasp for Stock Options?" *Wall Street Journal: CFO Journal*, August 26.

Cronqvist, Henrik, Anil K. Makhija, Scott E. Yonker. 2012. "Behavioral Consistency in Corporate Finance: CEO Personal and Corporate Leverage." *Journal of Financial Economics* 103(1): 20–40.

Custódio, Cláudia, and Daniel Metzger. 2013. "How Do CEOs Matter? The Effect of Industry Expertise on Acquisition Returns." *Review of Financial Studies* 26(8): 2008–47.

Davidson, Robert, Aiyesha Dey, Abbie J. Smith. 2013. "Executives' 'Off-the-Job' Behavior, Corporate Culture, and Financial Reporting Risk." Chicago Booth Research Paper No. 12-24. Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2096226.

Deshmukh, Sanjay, Anand M. Goel, and Keith M. Howe. 2013. "CEO Overconfidence and Dividend Policy." *Journal of Financial Intermediation* 22(3): 440–63.

Dugan, Ianthe Jeanne, and Ariana Eunjung Cha. 2000. "AOL to Acquire Time Warner in Record $183 Billion Merger." *Washington Post*, January 11.

Economist, The. 2002. "After Hubris, Nemesis." July 19.

Fabrikant, Geraldine. 1994. "Executives Say That Viacom Has Won Paramount Battle." *New York Times*, February 15.

Fazzari, Steven, M., R. Glenn Hubbard, and Bruce C. Peterson. 1988. "Financing Constraints and Corporate Investment." *Brookings Papers on Economic Activity*, no. 1, pp. 141–95.

Galasso, Alberto, and Timothy S. Simcoe. 2011. "CEO Overconfidence and Innovation." *Management Science* 57(8): 1469–84.

Goel, Anand M., and Anjan V. Thakor. 2008. "Overconfidence, CEO Selection, and Corporate Governance." *Journal of Finance* 63(6): 2737–84.

Graham, John R., Campbell R. Harvey, and Manju Puri. 2013. "Managerial Attitudes and Corporate Actions." *Journal of Financial Economics* 109(1): 103–21.

Hall, Brian J., and Kevin J. Murphy. 2002. "Stock Options for Undiversified Executives." *Journal of Accounting and Economics* 33(2): 3–42.

Hall, Brian J., and Kevin J. Murphy. 2003. "The Trouble with Stock Options." *Journal of Economic Perspectives* 17(3): 49–70.

Hall, Brian J., and Jeffrey B. Liebman. 1998. "Are CEOs Really Paid Like Bureaucrats?" *Quarterly Journal of Economics* 113(3): 653–91.

Heaton, J. B. 2002. "Managerial Optimism and Corporate Finance." *Financial Management* 31(2): 33–45.

Hietala, Pekka, Steven N. Kaplan, and David T. Robinson. 2003. "What is the Price of Hubris? Using Takeover Battles to Infer Overpayments and Synergies." *Financial Management* 32(3): 5–31.

Hirshleifer, David, Angie Low, and Siew Hong Teoh. 2012. "Are Overconfident CEOs Better Innovators?" *Journal of Finance* 67(4): 1457–98.

Hribar, Paul, and Holly Yang. Forthcoming. "CEO Overconfidence and Management Forecasting." *Contemporary Accounting Research*.

Jenter, Dirk, and Katharina Lewellen. Forthcoming. "CEO Preferences and Acquisitions." *Journal of Finance*.

Kamar, Ehud. 2009. "The Story of Paramount Communications v. QVC Network: Everything is Personal." Chap. 10 in *Corporate Law Stories*, edited by J. Mark Ramseyer. New York: Foundation Press.

Kaplan, Steven, Mark Klebanov, and Morten Sorensen. 2012. "Which CEO Characteristics and Abilities Matter?" *Journal of Finance* 67(3): 973–1007.

Kaplan, Steven N., and Luigi Zingales. 1997. "Do Investment–Cash Flow Sensitivities Provide Useful Measures of Financing Constraints?" *Quarterly Journal of Economics* 112(1): 169–215.

Kolasinski, Adam C., and Xu Li. 2013. "Can Strong Boards and Trading Their Own Firm's Stock Help CEOs Make Better Decisions? Evidence from Corporate Acquisitions by Overconfident CEOs." *Journal of Financial and Quantitative Analysis* 48(4): 1173–1206.

Lambert, Richard, A., David F. Larcker, and Robert E. Verrecchia. 1991. "Portfolio Considerations in Valuing Executive Compensation." *Journal of Accounting Research* 29(1): 129–49.

Larwood, Laurie, and William Whittaker. 1977. "Managerial Myopia: Self-Serving Biases in Organizational Planning." *Journal of Applied Psychology* 62(2): 194–98.

Lewellen, Katharina. 2006. "Financing Decisions When Managers Are Risk Averse." *Journal of Financial Economics* 82(3): 551–90.

Li, Sheng. 2015. "The Influence of Financial Advisers on Return Chasing." Job Market Paper. https://dl.dropboxusercontent.com/u/19386549/shengli_jmp.pdf.

Malmendier, Ulrike, Enrico Moretti, and Florian Peters. 2015. "Winning by Losing: Evidence on the Long-Run Effects of Mergers." Unpublished paper.

Malmendier, Ulrike, and Devin Shanthikumar. 2007. "Are Small Investors Naive about Incentives?" *Journal of Financial Economics* 85(2): 457–89.

Malmendier, Ulrike, and Geoffrey Tate. 2005. "CEO Overconfidence and Corporate Investment." *Journal of Finance* 60(6): 2661–2700.

Malmendier, Ulrike, and Geoffrey Tate. 2008. "Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction." *Journal of Financial Economics* 89(1): 20–43.

Malmendier, Ulrike, Geoffrey Tate, and Jon Yan. 2011. "Overconfidence and Early-Life Experiences: The Effect of Managerial Traits on Corporate Financial Policies." *Journal of Finance* 66(5): 1687–1733.

McGrath, Rita Gunther. 2015. "15 Years Later, Lessons from the Failed AOL-Time Warner Merger." *Fortune*, January 10.

Merrow, Edward W., Kenneth E. Phillips, and Christopher W. Myers. 1981. *Understanding Cost Growth and Performance Shortfalls in Pioneer Process Plants*. Santa Monica, CA: Rand.

Miller, Dale T., and Michael Ross. 1975. "Self-Serving Biases in the Attribution of Causality: Fact or Fiction?" *Psychological Bulletin* 82(2): 213–25.

Murphy, Kevin J. 1999. "Executive Compensation." Chap. 38 in *Handbook of Labor Economics Vol. 3B*, edited by O. Ashenfelter and D. Card, 2485–2563. Elsevier.

Murphy, Kevin J. 2013. "Executive Compensation: Where We Are, and How We Got There." Chap. 4 in *Handbook of the Economics of Finance Vol. 2A,* edited by George M. Constantinides, Milton Harris, and René M. Stulz. North Holland: Elsevier.

Odean, Terry. 1998. "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance* 53(5): 1775–98.

Odean, Terry. 1999. "Do Investors Trade Too Much?" *American Economic Review* 89(5): 1279–98.

Otto, Clemens A. 2014. "CEO Optimism and Incentive Compensation." *Journal of Financial Economics* 114(2): 366–404.

Peers, Martin, and Julia Angwin. 2003. "AOL Posts a $98.7 Billion Loss on New Goodwill Write-Down." *Wall Street Journal*, January 30.

Roll, Richard. 1986. "The Hubris Hypothesis of Corporate Takeovers." *Journal of Business* 59(2): 197–216.

Schrand, Catherine M., and Sarah L. C. Zechman. 2012. "Executive Overconfidence and the Slippery Slope to Financial Misreporting." *Journal of Accounting and Economics* 53(1–2): 311–29.

Shefrin, Hersh. 2001. "Behavioral Corporate Finance." *Journal of Applied Corporate Finance* 14(3): 113–26.

Statman, Meir, and James F. Sepe. 1989. "Project Termination Announcements and the Market Value of the Firm." *Financial Management* 18(4): 74–81.

Sunder, Jayanthi, Shyam V. Sunder, and Liang Tan. 2010. "The Role of Managerial Overconfidence in the Design of Debt Covenants." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1595007.

Svenson, Ola. 1981. "Are We All Less Risky and More Skillful Than Our Fellow Drivers?" *Acta Psychologica* 47(2): 143–48.

Yermack, David. 1996. "Higher Market Valuation of Companies with a Small Board of Directors." *Journal of Financial Economics* 40(2): 185–211.

Yim, Soojin. 2013. "The Acquisitiveness of Youth: CEO Age and Acquisition Behavior." *Journal of Financial Economics* 108(1): 250–73.

Journal of Economic Perspectives—Volume 29, Number 4—Fall 2015—Pages 61–88

Overconfident Investors, Predictable Returns, and Excessive Trading[†]

Kent Daniel and David Hirshleifer

The last several decades have witnessed a shift away from a fully rational paradigm of financial markets towards one in which investor behavior is influenced by psychological biases. Two principal factors have contributed to this evolution: a body of evidence showing how psychological bias affects the behavior of economic actors; and an accumulation of evidence that is hard to reconcile with fully rational models of security market trading volumes and returns. In particular, asset markets exhibit trading volumes that are high, with individuals and asset managers trading aggressively, even when such trading results in high risk and low net returns. Moreover, asset prices display patterns of predictability that are difficult to reconcile with rational expectations–based theories of price formation.

In this paper, we discuss the role of overconfidence as an explanation for these patterns. Overconfidence means having mistaken valuations and believing in them too strongly. It might seem that actors in liquid financial markets should not be very susceptible to overconfidence, because return outcomes are measurable, providing extensive feedback. However, overconfidence has been documented among

■ *Kent Daniel is Professor of Finance, Graduate School of Business, Columbia University, New York City, New York, and Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. David Hirshleifer is Professor of Finance, Merage School of Business, University of California at Irvine, Irvine, California. Their email addresses are kd2371@columbia.edu and david.h@uci.edu.*

[†] For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.29.4.61 doi=10.1257/jep.29.4.61

experts and professionals, including those in the finance profession. For example, overconfidence is observed among corporate financial officers (Ben-David, Graham, and Harvey 2013) and among professional traders and investment bankers (Glaser, Langer, and Weber 2013). People tend to be overoptimistic about their life prospects (Weinstein 1980), and this optimism directly affects their financial decisions (Puri and Robinson 2007).

We do not mean to suggest that overconfidence is the only phenomenon worth considering in behavioral finance, nor that it should serve as an all-purpose explanation for all financial anomalies. But overconfidence seems likely to be a key factor in financial decision-making. Overconfidence is a widespread psychological phenomenon (as discussed by Malmendier and Taylor in their overview for this symposium), and is associated with a cluster of related effects. For example, it includes *overplacement*—overestimation of one's rank in a population on some positive dimension—and *overprecision*—overestimation of the accuracy of one's beliefs. An example is overestimation of one's ability to predict future stock market returns. A cognitive process that helps support overconfident beliefs is *self-attribution bias*, in which people credit their own talents and abilities for past successes, while blaming their failures on bad luck.

To evaluate the importance of overconfidence for financial markets, we proceed as follows. We start by reviewing two sets of empirical findings seemingly at odds with rational agent asset-pricing theories: the arguments that trading volumes are excessive and the evidence that security returns are predictable. We then sketch a sequence of models of investor trading and security prices that include various aspects of overconfidence, with increasing complexity, and discuss the empirical implications of each of these models. We hope that this presentation will clarify which aspects of the model are important in delivering specific empirical implications. Finally, we offer some conclusions about how overconfidence contributes to our understanding of financial markets.

Evidence on Trading Patterns and Return Predictability

The notion of market efficiency, as explained in Fama (1970), is based on the idea that when investors in frictionless asset markets compete with one another, securities will be priced to fully reflect all publicly available information. More generally, rationality on the part of investors has some strong implications.

With surprisingly mild theoretical assumptions, one can show that rational individuals should not agree to disagree. Intuitively, if we start with the same prior beliefs, yet now we disagree, this suggests that at least one party has information that the other party should be taking more fully into account (Aumann 1976). In a similar spirit, rational investors should not place bets with each other; the fact that another investor is willing to take the opposite side of my trade should suggest to me that this investor knows something I do not know (Grossman 1976;

Milgrom and Stokey 1982; Tirole 1982). For this reason, leading rational frictionless models of asset pricing—at least in their most simple versions—imply that after a single round of trading, everyone should hold the market portfolio. Investors should not bet against each other, each expecting to beat the counterparties. However, we clearly observe high volumes of trade in financial markets.

Moreover, in an efficient market, a trading strategy based on existing information cannot be used to earn abnormal profits. If such trading strategies do exist, there is a *return anomaly*: such opportunities suggest either that rational agents are not fully exploiting available profit opportunities, or that risk aversion or market frictions constrain their ability to do so. However, it is now a well-accepted empirical finding—even by those who adhere to a rational-actors explanation—that asset markets do display strong patterns of return predictability. This finding poses a challenge to the hypothesis that investors are rational, because it suggests that investors are making mistakes: they are throwing away money buying overpriced securities that subsequently do poorly, and are missing out on buying underpriced securities that subsequently do well. An alternative explanation for return predictability is that it results from some kind of risk premia—risky assets predictably return more than less-risky ones. This explanation then raises the question of the actual source of risk and whether plausible levels of risk aversion are high enough to explain the size of the predictability, a question we address below.

In this section, we will explore the evidence on high trading volumes and predictable returns in greater depth and discuss how overconfidence-based explanations provide some insight into these patterns.

Disagreement, Speculative Trade, and Trading Volume

A financial trade requires that two parties agree to disagree in the sense that at a given price one party believes it is a good idea to sell the asset while the other party believes it is a good idea to buy it. Of course, there are possible reasons for informed agents to trade other than disagreement, such as liquidity motives (for example, sending a child to college) or to rebalance to achieve a more diversified portfolio (for example, after a shock to one's labor income or human capital). Speculative trade can arise in rational models if investors in securities markets are periodically required to sell or buy securities as a result of liquidity shocks. Several models starting with Grossman and Stiglitz (1976) have shown that if there are random shocks to security supply (designed to capture the idea that there are investors whose need to cash out of their positions is unpredictable to others), this can add enough noise to make room for some speculative trading.

But such motives for trade are relatively limited, and do not seem to explain the magnitudes of trade, or the willingness of investors to incur the large transaction costs that they sometimes need to pay to make such trades. The total volume of trade in financial markets is vast. Over the period 1980–2014, the annualized average turnover for the 500 largest US stocks has averaged 223 percent, or just over $100 billion per day. Over the year 2014, the total dollar trade in these top 500 stocks

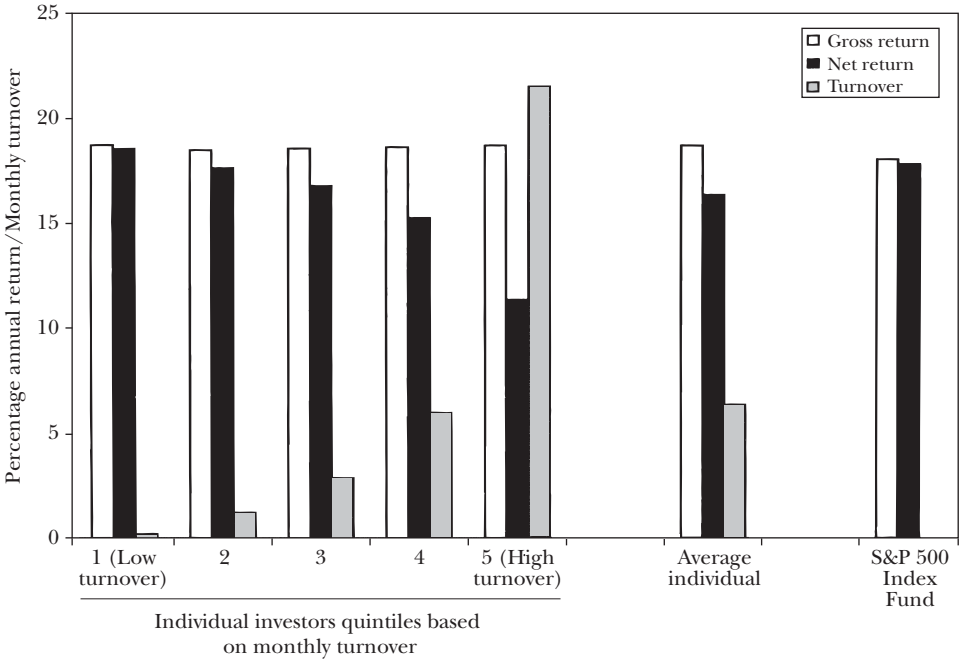
was $29.5 trillion (Collin-Dufresne and Daniel 2014)—nearly double the US GDP. Trade in foreign exchange is even larger. Froot and Thaler (1990) report that, as of 1989, average trading in the foreign exchange market was about $430 billion per day as compared to daily US GDP during that year of $22 billion and daily trades in goods and services of $11 billion.

Theories constructed to explain how rational traders should react to liquidity shocks don't seem sufficient to explain the magnitudes of financial trades that we observe, or the patterns in trading volume. Rather, several findings point to overconfidence as a likely explanation. Everyday experience suggests that there is considerable disagreement across individuals in the economy, with each individual believing that he or she is correct. In overconfidence-based models, investors who are overconfident form judgments about the value of a security that put too much weight on their own views and insufficient weight on the views of other investors (as reflected in the security's price). As a result, overconfident investors expect high profits from trading on their opinions.

The excessive trading of individual investors can be called *the active investing puzzle*. Individual investors trade individual stocks actively, and on average lose money by doing so. The more actively investors trade, the more they typically lose (Odean 1999). In particular, Barber and Odean (2000) find that in a sample of trades of 78,000 clients of a large discount brokerage firm from 1991 to 1996, some households trade much more than others. The turnover and gross and net returns to the clients in different turnover quintiles are summarized in Figure 1, reproduced from their paper. The gray bars give the average monthly turnover of the accounts in each quintile. Strikingly, the average monthly turnover in the fifth quintile is over 20 percent per month. The white bars give gross returns (that is, without accounting for the costs associated with trading) and show that, across quintiles, there is little variation in average gross returns. However, the black bars show that the net returns are quite different. The high-turnover investors pay large fees, given their high volume of trade, which drives down their net returns. The net returns of all quintiles except the lowest are lower than the net return from investing in a Standard & Poor's 500 index fund.

Tests that aggregate across individual investors also find that the stocks that individual investors buy tend to subsequently underperform. Investor losses can be astonishingly large; in the aggregate, the annual losses of Taiwanese individual investors amount to 2.2 percent of Taiwan's gross domestic product and 2.8 percent of total personal income (Barber, Odean, and Zhu 2009). In experimental markets as well, some investors overestimate the precision of their signals, are more subject to the winner's curse, and have inferior trading performance (Biais, Hilton, Mazurier, and Pouget 2005). Greater ease of trading gives investors free rein to harm themselves by more aggressive trading, as occurred with the rise of online trading (Barber and Odean 2002; Choi, Laibson, and Metrick 2002). A similar point applies to individuals who invest in active mutual funds instead of index funds for better net-of-fees performance. Indeed, the existence of actively managed mutual

Figure 1
Monthly Turnover and Annual Performance of Individual Investors



Source: Barber and Odean (2000, Figure 1).
Notes: The white bar (black bar) represents the gross (net) annualized geometric mean return for February 1991 through January 1997 for individual investor quintiles based on monthly turnover (grey bar). The net return on the S&P 500 Fund is that earned by the Vanguard Index 500.

funds that charge high fees without providing correspondingly high gross performance provides evidence that a number of individual investors are overconfident about their ability to select the high-performing active fund managers (French 2008; Malkiel 2013).

A range of evidence from a wide variety of sources suggests that overconfidence provides a natural explanation for the active investing puzzle because it causes investors to trade more aggressively even in the face of transactions costs or adverse expected payoffs (Odean 1998). In one of the rare studies of investor trading that measures overconfidence directly, Grinblatt and Keloharju (2009) associate the trading behavior of Finnish investors with the results of a psychometric test given to all Finnish males at age 19 or 20. The study finds that overconfident investors (as well as investors who are prone to sensation-seeking) trade more often. In a different study consistent with overconfidence as an explanation for the active investing puzzle, Kelley and Tetlock (2013) construct a structural model of market trading that includes informed rational investors as well as uninformed investors who trade either for hedging reasons or to

make an (overconfident) bet on perceived information. They estimate this model using a dataset on trades, prices, and information releases for US traded firms and conclude that, without overconfidence-based trading, volumes would be smaller by a factor of 100. Finally, motivated by psychological evidence that men are more overconfident than women in decision domains traditionally perceived as masculine, such as financial matters, Barber and Odean (2001) compare the trading behavior and performance of men and women. The gross (benchmark-adjusted, before fees) returns of the men in the sample are lower, though this difference is not statistically significant. However, consistent with higher confidence on the part of men, the average turnover for accounts opened by men is about 1.5 times higher than accounts opened by women, and as a result men pay 0.94 percent per year in higher transaction costs. As a result, the net-of-fees returns of men are far lower.

Other aspects of investor trading behavior are also consistent with overconfidence and the psychological processes that accompany it. Individual investors tend to trade more after they experience high stock returns. For example, early adopters of online trading tended to make the switch after unusually good personal performance, and subsequently traded more actively (Barber and Odean 2002; Choi, Laibson, and Metrick 2002). This connection may help to explain why stock market trading volume increases after high returns, as has been documented in a large number of countries (Griffin, Nardari, and Stulz 2007). For example, annualized turnover in US common stocks was at levels of over 100 percent late in the bull market of the 1920s, fell through the 1930s and 1940s, and then rose dramatically from the 1990s up through the financial crisis of 2007–2008 (Collin-Dufresne and Daniel 2014). Statman, Thorley, and Vorkink (2006) find that US market turnover is positively correlated with lagged monthly market returns and that turnover of individual securities is positively associated with lagged market turnover (after controlling for past values of turnover and returns in each security).

How can these patterns of overconfidence and high turnover persist over time, despite the high risks and costs they impose upon investors? Overconfidence in general is supported by bias in self-attribution, as modeled in Daniel, Hirshleifer, and Subrahmanyam (1998) and Gervais and Odean (2001); that is, investors who have experienced high returns attribute this to their high skill and become more overconfident, while investors who experience low returns attribute it to bad luck rather than experiencing an offsetting fall in their overconfidence level.

Overconfidence is likely to be especially important when security markets are less liquid and when short-selling is difficult or costly. When short-selling is constrained, pessimists about a stock find it harder to trade on their views than optimists. If some of the optimists do not adequately take into account that pessimists are sidelined by short-sale constraints, the optimists will overvalue the stock, resulting in equilibrium overpricing. Thus, when overconfidence is combined with short sales constraints, we expect the security to become overpriced (Miller 1977).

Motivated by this hypothesis, Diether, Malloy, and Scherbina (2002) document that firms about which the analysts disagree more—measured by the dispersion in

the analysts' forecasts of the firm's future earnings—on average earn lower returns. This finding is usually interpreted as evidence that investor disagreement matters; overconfidence provides a natural explanation for *why* disagreement exists and matters. Because volatility creates greater scope for disagreement, this approach also suggests overpricing of more volatile stocks. Consistent with this insight, Ang, Hodrick, Xing, and Zhang (2006, 2009) and Baker, Bradley, and Wurgler (2011) show that high idiosyncratic-volatility stocks earn lower subsequent returns than low-volatility stocks. This hypothesis is also consistent with the finding that stocks and other assets with high systematic risk (that is, high market beta) typically earn too low a return premium relative to the risk-return tradeoff implied by equilibrium models such as the Capital Asset Pricing Model (Black, Jensen, and Scholes 1972; Frazzini and Pedersen 2014).

During the high-tech boom at the turn of the millennium, there were episodes of strong disagreement in which, remarkably, the market value of a parent firm was sometimes substantially less than the value of its holdings in one of its publicly traded divisions (Lamont and Thaler 2003). Such patterns reflected the fact that an optimistic set of investors were excited about the prospects of a glamorous division, and the relatively pessimistic investors who were setting the price of the parent firm found it too costly or troublesome to short-sell the glamorous division to bring its price in line with that of the parent. Also consistent with overvaluation induced by investor disagreement, stocks with tighter short-sale constraints have stronger return predictability (Nagel 2005). Such asymmetry between the long and the short side of return anomalies is especially strong during optimistic periods, when overvaluation is most severe (Stambaugh, Yu, and Yuan 2012).

Overconfident disagreement, combined with short-sale constraints, can also cause dynamic patterns of increasing overpricing. Building on Harrison and Kreps (1978), Scheinkman and Xiong (2003) present a model in which overconfidence generates disagreement among agents regarding asset fundamentals. Owing to short-sale constraints, investors buy stocks that they know to be overvalued in the hope of selling at even higher prices to more optimistic buyers. This magnifies the pricing effects of disagreement. Such bubbles should be more severe in markets with lower available supply of shares, or "float" (Hong, Scheinkman, and Xiong 2006), as seems to have occurred during a bubble in Chinese warrants (Xiong and Yu 2011).

Although overconfidence causes problems in markets, it may bring some benefits, as well. Overconfidence can induce investors to investigate more, and/or to trade more aggressively based on their signals. This sometimes results in greater incorporation of information into price (Hirshleifer, Subrahmanyam, and Titman 1994; Kyle and Wang 1997; Odean 1998; Hirshleifer and Luo 2001). Furthermore, overconfidence encourages investors to participate in asset classes, such as the stock market or international investing, that they might otherwise neglect (owing to concerns such as fear of the unfamiliar). Empirically, a greater feeling of

competence about investing is associated with more active trading and with greater willingness to invest in foreign stock markets (Graham, Harvey, and Huang 2009).

Return Predictability

Here, we lay out certain documented patterns in return predictability that are at odds with the efficient markets hypothesis and potentially attributable to overconfidence. We first concentrate primarily on the nature and direction of the patterns, as opposed to their magnitudes. Of course, it is possible that the abnormal returns generated by "anomaly portfolios" based on patterns of predictable returns are not anomalous at all. A strategy may earn high returns relative to some benchmark by virtue of exposure to some systematic risk factor that the benchmark does not capture. (A factor in the asset pricing literature refers to a statistical source of common variation in security returns—usually the return on a portfolio. For example, the returns of individual stocks can be explained in part by realizations of the stock market as a whole, as is verified by regressing stock returns on the market portfolio.) In the next subsection, we will argue that the large premia earned by a combination of these anomaly-based strategies is too large to be explained plausibly in this way. We consider evidence on return predictability of three types: 1) predictability based on the market price of the firm, scaled by measures of fundamental value; 2) predictability based on a recent history of past returns (momentum and reversal); and 3) predictability based on underreaction to, or neglect of, public information about fundamentals.

One of the earliest anomalies uncovered in academic research was the size anomaly (Banz 1981; Keim 1983)—the phenomenon that "small" firms, defined in terms of low market-capitalization, earn higher returns than large firms. Even stronger predictability is obtained when scaling the firm's market capitalization by a measure of the firm's fundamental value. Fama and French (1992) find that the book-to-market ratio—that is, the book-value of equity, scaled by the firm's market capitalization—predicts returns. In particular, so-called "value firms" with high book-to-price ratio substantially outperform "growth firms" with low book-to-price ratios. Many other fundamental-to-price measures, including earnings-to-price, sales-to-price, and cash-flow-to-price ratios, also positively forecast future returns (Lakonishok, Shleifer, and Vishny 1994).

A pattern of long-term price reversal (DeBondt and Thaler 1985) can also be understood as related to the fundamental-to-price ratio. Intuitively, a stock that is mispriced now probably did not share the same mispricing years ago. Daniel and Titman (2006) add an additional dimension to this point; if past long-term returns are decomposed into a component associated with public information and an orthogonal component, a long-term reversal of prices is only observed for the orthogonal component. The component of the past return associated with public information does not reverse.

Post-earnings announcement drift or *earnings momentum* is the phenomenon in which firms that announce high earnings relative to forecasts, or whose price

jumps up on an announcement date, tend to earn high returns over the subsequent 3–6 months (Bernard and Thomas 1989; 1990). *Price momentum* is the tendency for returns over the past 3–12 months to continue in the same direction in the future. The overconfidence explanation for momentum involves a pattern of continuing overreaction and slow correction.

More specifically, price momentum in the US stock market has several key features. First, it is predominantly associated with lagged price changes that can be attributed to public information releases. In contrast, price changes that cannot be associated with news tend to exhibit reversal rather than continuation (Chan 2003; Tetlock 2011). Second, in the long-run, momentum tends to reverse (Griffin, Ji, and Martin 2003; Jegadeesh and Titman 2011). Third, momentum effects are weak for value stocks, but strong for growth stocks (Daniel and Titman 1999). Fourth, momentum strategies generate especially strong returns in calm periods when the past return on the market is high (Cooper, Gutierrez, and Hameed 2004; Daniel and Moskowitz 2014), but exhibit strong negative skewness and earn lower returns in turbulent (high volatility) bear markets (Daniel and Moskowitz 2014; Daniel, Jagannathan, and Kim 2012).

Asness, Moskowitz, and Pedersen (2013), among many others, document strong value and momentum anomalies in non-US data and in other asset classes including currencies, commodity futures, and government bonds. Moskowitz (2015) shows that the same momentum and value/reversal patterns observed in other asset classes are also present in sports betting venues. Sports betting markets are a useful test-bed for overconfidence-based theories because the outcomes of these contests are unlikely to be interdependent with other economic outcomes that may affect the marginal utility of individuals. Moskowitz argues that the presence of value and momentum effects in sports betting markets is consistent with delayed overreaction theories of asset pricing. He finds that higher ambiguity predicts stronger momentum and value returns, consistent with what is observed in financial markets and consistent with the overconfidence-based model of Daniel, Hirshleifer, and Subrahmanyam (2001).

Many items reported in financial statements can be useful in forecasting future earnings, but investors do not appear to make full use of such information. One prominent example is "operating accruals," which are the accounting adjustments made to a firm's cash flows to obtain earnings, a standard measure of profitability. Such adjustments may include sales transactions whose payments have not yet arrived or expense transactions for which actual payments have not yet been made. Sloan (1996) shows that market prices don't fully reflect the extent to which earnings arise from cash flows or accruals.

A common pattern in event studies is continuation of the event-date return, so that events that are on average good news predict high subsequent returns, and the opposite for bad news events (see the summary in Hirshleifer 2001). For example, the issuance of new securities tends to convey bad news about future cash flows, while the repurchase of existing securities tends to convey good news. Consistent

with return continuation, repurchases tend to be followed over a long period by high returns (Ikenberry, Lakonishok, and Vermaelen 1995), and equity and debt issues in many countries by negative abnormal returns (Loughran and Ritter 1995; Spiess and Affleck-Graves 1995; Henderson, Jegadeesh, and Weisbach 2006). Daniel and Titman (2006) and Pontiff and Woodgate (2008) develop more comprehensive measures of share issuance over a given time period, and show that a lagged measure of issuance strongly forecasts returns. At the aggregate level as well, the share of equity issues in total new equity and debt issues has been a negative predictor of US stock market returns (Baker and Wurgler 2000).

Return Predictability—Magnitudes

The patterns of returns documented in the preceding section might reflect certain kinds of rational risk premia, rather than mistakes or biases on the part of investors. Here, we summarize evidence on the risk and rewards of strategies based upon these effects to see if this explanation is plausible. A portfolio that simultaneously exploits several of the patterns of return predictability documented in the preceding section generates an exceptionally high reward-to-risk ratio. Using insights from Hansen and Jagannathan (1997), accommodating these premia within any frictionless rational expectations model would require extreme (and we will argue, unrealistic) variation in investor marginal utility across states of the world.

We start with a set of seven "zero-investment" portfolios designed to capture the return predictability patterns described in the preceding section.

First, the "Small Minus Big" or SMB portfolio, proposed by Fama and French (1993), captures the difference in average returns between small and large market-capitalization firms. This portfolio, at the beginning of each month, takes a long position in $1 worth of small-market-capitalization stocks, financed by taking a short position in $1 worth of large market-capitalization stocks. Historically, investors should have been able to capture the returns of this zero-investment or $1-long/$1-short portfolio with minimal transaction costs (despite the need to sell short). This portfolio has been used in numerous academic studies, and yearly, monthly, and daily returns from 1926 on are available on Kenneth French's website.

Second, the "High Minus Low" or HML portfolio is formed to exploit the persistently higher returns of stocks with high book-to-market ratios relative to those with low book-to-market ratios. The portfolio involves buying value stocks—stocks with ratios of book-value of equity to market-value of equity in the top 30 percent of all stocks on the New York Stock Exchange—and shorting growth stocks, with book-to-market ratios in the bottom 30 percent.

Third, the "Up Minus Down" or UMD portfolio is a price momentum portfolio (Carhart 1997; Fama and French 1993). It is formed by buying stocks that rose in price in the previous time period (often 12 months) and taking a short position in stocks that declined in price in the previous time period. Thus, it is based on momentum in stock prices.

Fourth, the "ISsUance" or ISU portfolio buys a value-weighted portfolio of firms that over the preceding three years repurchased stock, and shorts a portfolio of stocks that issued new equity, based on the Daniel and Titman (2006) measure.

Fifth, the "ACcRual" or ACR portfolio goes long on a portfolio of firms that had the lowest ratio of accruals to earnings over the past year, and goes short on the firms that had the highest accruals based on the measure of Sloan (1996).

Sixth, the "Betting-Against-Beta" or BAB portfolio is constructed following the description in Frazzini and Pedersen (2014). The long side of the portfolio is a leveraged portfolio of low-beta stocks. The portfolio takes a short position in high-beta stocks.

Finally, the "Idiosyncratic-VOLatility" or IVOL portfolio takes a long position in the set of firms that had the lowest idiosyncratic volatility of daily returns over the preceding one month, and shorts the highest idiosyncratic volatility stocks, measured following the procedure specified in Ang et al. (2006).

In working with these portfolios, remember that the "Sharpe ratio" of a portfolio is the ratio of its reward to its risk. More specifically, we define it here to be the ratio of the annualized excess return on the portfolio to the annualized return standard deviation of the portfolio. To summarize how an investor might exploit these anomalies, it is useful to examine the Sharpe ratios achieved by combining the anomaly portfolios into super-portfolios.

Table 1 presents Sharpe ratios for portfolios consisting of the US market portfolio—specifically the Center for Research in Security Prices value-weighted index return—along with various mixtures of the seven candidate anomaly portfolios.[1] Each row of Table 1 represents a different combination of the set of anomaly portfolios designed to achieve a high Sharpe ratio. The first eight columns show the weights on each of the anomaly portfolios, and the number in the ninth (and last) column gives the annualized Sharpe ratio of the overall portfolio that combines them. The component portfolios are normalized so that each has the same volatility over the 1963:07–2014:05 sample period. Thus, the weights given in the table are proportional to the volatility of that component.

The first row of the table shows that during this sample period, a portfolio that was 100 percent invested in the market index (Mkt-Rf) experienced an annualized Sharpe ratio of 0.39. Specifically, the annualized return, net of the one-month Treasury-bill rate, was 6 percent, and the annualized volatility was 15.5 percent. The second row shows how much an investor could have improved on the market Sharpe ratio by also investing in the size-based SMB and value-based HML portfolios. The optimal combination of these three portfolios results in a Sharpe ratio of 0.76, a vast improvement relative to the market portfolio on its own. The next few lines of the table show that the ability to invest in the momentum factor

[1] Mkt-Rf is the notation used by Fama and French (1993) for the excess return of the CRSP value-weighted index, relative to the one-month US treasury-bill return in the same month. "Mkt-Rf" meaning Market-minus-Riskfree is a long-short portfolio that is long on the market and short on the risk-free asset.

Table 1
Anomaly-based Strategy Sharpe Ratios

Portfolio weights (%)								*Sharpe ratio*
Mkt-Rf	*SMB*	*HML*	*UMD*	*ISU*	*ACR*	*BAB*	*IVOL*	
100.0	—	—	—	—	—	—	—	0.39
34.9	18.7	46.4	—	—	—	—	—	0.76
25.8	10.5	33.0	30.7	—	—	—	—	1.07
8.0	4.5	33.9	17.7	26.8	9.1	—	—	1.37
7.7	12.4	13.8	4.5	18.0	10.2	9.5	24.0	1.78

Notes: This table presents the realized optimal strategy Sharpe ratios from 1963:07 to 2014:05 for a set of long-short portfolios based on a set of anomalies taken from the finance literature: Market-minus-Riskfree (Mkt-Rf) is the market index portfolio from Fama and French (1993); "Small Minus Big" (SMB) and "High Minus Low" (HML) are two other Fama and French (1993) portfolios; "Up Minus Down" (UMD) is the Carhart (1997) price momentum portfolio; "ISsUance" (ISU) and "ACcRual" (ACR) are long-short portfolios based on the Daniel and Titman (2006) cumulative issuance and Sloan (1996) accruals measures, respectively; "Betting-Against-Beta" (BAB) is a Frazzini and Pedersen (2014) portfolio; and finally "Idiosyncratic-VOLatility" (IVOL) is a Ang et al. (2006) portfolio. See text for details.

brings the Sharpe ratio up to 1.07, and the ability to invest in the issuance and accrual portfolios brings it up further to 1.37. Finally, if the investor had been free to invest in any of these eight portfolios, and knew beforehand the distribution of returns over this period (not the returns themselves but only the distribution), that investor could have earned a Sharpe ratio of 1.78, more than four times higher than that of the market.

The numbers presented in this table are the Sharpe ratios for the optimal portfolios, calculated as if investors knew up front the realized distribution of returns. But our main conclusions still apply if investors do not have full foreknowledge of distributions. For example, an equal-weighted combination of the eight portfolios (weights which do not require assumptions about future performance of the portfolios) earns an annualized Sharpe ratio of 1.54. Similarly, Asness, Moskowitz, and Pedersen (2013) document that a 50/50 combination of only the value and momentum portfolios, but diversified across different regions and asset classes, produces an annualized Sharpe ratio of 1.59.

Any asset-pricing model—whether rational or behavioral—needs to explain why investors are apparently passing up these very high-return, low-volatility investments. In a rational expectations setting, asset premia arise only when the asset's returns are risky, meaning that returns are high when the investor is relatively rich (and marginal utility of wealth is low) and are low when the investor is poor (and marginal utility is high). To explain such a large Sharpe ratio, marginal utility must be quite variable. The Hansen and Jagannathan (1991) bound shows that, to explain the existence of a portfolio with a Sharpe ratio of 1.8 requires that the annualized standard deviation

in marginal utility growth be almost as large or larger—that is, greater than approximately 170 percent. Both casual observation and macroeconomic data suggest that marginal utility growth does not vary nearly this much. For example, the annualized volatility of aggregate US consumption growth is 100 times smaller (at about 1.8 percent). We note that the macroeconomics profession continues to wrestle with the equity premium puzzle—the finding that the Sharpe ratio of the equity market portfolio—which is about 0.4 (annualized)—is "too high" given the low covariance of market returns with consumption growth (Hansen and Singleton 1983; Mehra and Prescott 1985; Weil 1989). Thus, the Sharpe ratio of 1.78 in Table 1 is far harder to reconcile with a rational investor model.

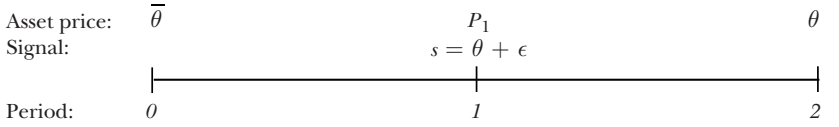
Perhaps an answer to these puzzles can be found in trading frictions that make it costly for rational investors to trade to exploit perceived profit opportunities. However, the magnitude of such frictions, as captured by bid-ask spreads, is too small to explain why investors would forego the combination of return and risk described here. For moderate-sized trades in large firms, such as those used to construct the zero-investment portfolios described here, such spreads are relatively small. Alternatively, maybe these results arise from data mining, and if one looked at different time periods, or a limited set of these portfolios, or weighted the portfolios differently, then the pricing anomalies would disappear. One can tinker with different time periods, or different portfolios, or different weights. But the opportunities presented by these anomaly portfolios appear robust.

What other theories can explain the patterns in Table 1? Could it be that the decision processes or beliefs of investors are biased in ways that induce the seven pricing anomalies listed earlier? Overconfidence-based models suggest that the answer is "yes." In these models, investors continue to optimize, but do so based on incorrect beliefs about the state probabilities. Under this explanation, investors think that the state probabilities are such that the expected returns of the anomaly portfolios are not abnormally high, despite the evidence in Table 1. This explanation need not presume that all investors are overconfident. There could still be rational investors who correctly perceive the high available Sharpe ratios, but if such investors are relatively small in number, and capital-constrained, their trading to exploit the profit opportunities will not fully eliminate them. How might overconfidence generate the anomalies that underlie Table 1 so that overconfident investors do not believe that these portfolios outperform? In the next section, we lay out overconfidence models that can potentially explain these patterns.

Overconfidence-Based Models of Asset Price Formation

In the standard frictionless rational expectations framework, investors process information perfectly. Thus, asset prices are always equal to rationally discounted expected cashflows, where discount rates are equal to rational expectations of returns. Investors earn returns that are, on average, exactly what they expect.

Figure 2
Model 1: Basic One-Signal Model—Timeline



Source: Authors.
Notes: This figure illustrates the timeline for the basic three-date, one-signal model. At time 0, the investor knows only the prior distribution from which the time 2 final payoff is drawn. At time 1, the investor observes a noisy signal s. At $t = 2$, the payoff is revealed.

As discussed in the previous section, so-called zero investment portfolios constructed to reveal anomalies have produced high Sharpe ratios—high average excess returns with low volatility—which have low correlation with macroeconomic shocks that might plausibly represent risk. Thus, some researchers have turned to behavioral models in an attempt to explain these patterns. The behavioral models rely on either nonstandard preferences or biased beliefs. In models with nonstandard preferences, investors correctly expect that high excess returns are achievable with these anomaly-based portfolios. In these models investors choose not to invest more in these portfolios because they find certain kinds of risk extraordinarily painful to bear. In contrast, biased belief models posit that investors make mistakes in the way that they form expectations about asset payoffs. Overconfidence-based models fall into the second category.

We now provide a sequence of models that illustrate some insights of the overconfidence-based approach. The first model is a bare-bones setting which captures the fact that an overconfident investor overreacts to a signal that is perceived as private, resulting in overreaction and correction, consistent with evidence of long-run return reversals. We then present models that show how refinements to this basic model, grounded in the psychological evidence on overconfidence, can plausibly generate other anomalies described above.

Model 1: One Signal

Consider a static overconfidence model involving a three-date, one-signal example. Figure 2 provides a timeline. For the moment, assume that the representative investor in the model is risk neutral. There are three dates, $t \in \{0, 1, 2\}$, and two securities: a risk-free asset with a risk-free rate of zero, and a security which will pay an uncertain liquidating dividend θ at time 2. The prior distribution for θ is known. At $t = 1$, the investor receives a private signal of the form $s = \theta + \epsilon$. In a representative agent model, signals cannot of course be truly private, but the model can be viewed as one where there is also a very small mass of investors who do not receive the signal. Also, labeling this signal as "private" captures this idea that this agent believes that she has used her skill to process information and generate new and

unique insights about the payoff θ. The psychology literature suggests that agents will be more overconfident about these "private" signals than they will be about "public" information such as earnings announcements that are more easily translated into estimates of future firm value. The date 1 price in this setting is a weighted average of the prior expectation and the signal, with relative weights proportional to the investor's perceived precisions of the prior and of the signal.

We are interested in whether the asset return is forecastable. If the investor is rational (not overconfident), then the price P_1 is equal to the rational expectation of the payoff $\mathbb{E}[\theta]$, and in this case the price change from date 1 to date 2 is unforecastable—that is, it is not correlated with any price change from time 0 to time 1, nor with the signal received in time 1. The information the signal provides for fundamental value is correctly impounded into price at date 1, and the market is efficient.

However, if an investor overestimates the precision of the signal, the investor will overreact to that signal. Thus, a positive signal will cause the date 1 price to be too high, resulting in too high a price change between dates 0 and 1. On average the price then falls back between dates 1 and 2, which is a pattern of return reversal. In contrast, if the investor underestimates the precision of the signal, the date 1 price will underreact, and the subsequent price change will on average be positive again for a second period, which would be a case of return momentum.

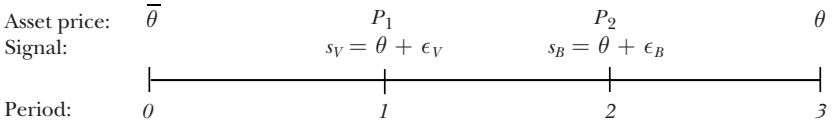
What might cause the investor's estimate of the signal precision to differ from the true precision? One possible answer is investor overconfidence. In the simplest version of the Daniel, Hirshleifer, and Subrahmanyam (1998) model, the representative investor observes only a private signal and is overconfident about that signal, resulting in price reversal. Alternatively, in Eyster, Rabin, and Vayanos (2013), investors make a different error; they fail to infer fully the private signals received by other investors from the price. Effectively, the representative investor underestimates the information implicit in price—namely, the precision of the aggregate private signal. In consequence, the investor underreacts to this signal, which implies that in equilibrium price underreacts to new information. Their model implies price momentum, but because this is a result of pure underreaction to information, there is no reversal.

Neither pure underreaction nor pure overreaction, as reflected in Model 1, fully captures the return-predictability evidence discussed earlier, in which there is momentum at shorter horizons and reversal at longer horizons. In addition, Model 1 does not allow for *public* information signals prior to the terminal date, and therefore does not allow consideration of whether returns can be predicted based on public information such as the news of a new equity issue by the firm. To capture these patterns, we need to move to a richer model.

Model 2: Public and Private Signals

In Model 2, we introduce separate public and private signals. This model is the "static overconfidence" model in Daniel, Hirshleifer, and Subrahmanyam (1998). The timeline for Model 2 is given in Figure 3. There are now four dates and two signals: s_V is a priVate signal, and s_B is puBlic. The investor is overconfident, and

Figure 3
Model 2: Separate Public and Private Signals—Timeline



Source: Authors.
Notes: This figure presents the timeline for the four-date, two-signal model. Now the investor observes distinct pri*V*ate and pu*B*lic signals s_V and s_B at $t = 1$ and $t = 2$, respectively. At $t = 3$, the asset payoff θ is revealed.

therefore overestimates the precision of the private signal at time 1. However, the investor correctly estimates the precision of the public signal and the prior.

This approach delivers several additional features. First, as in Model 1, the market overreacts to the private signal, therefore the price change from time 1 to time 3 is in the opposite direction of the price change from time 0 to time 1. In addition, the market underreacts to the public signal: that is, price changes are autocorrelated so that $\text{cov}(R_{2,3}, R_{1,2}) > 0$. Given this positive return autocorrelation, it is tempting to jump to the conclusion that following the public release of good news at date 2, the share price will continue rising between dates 2 and 3, but this turns out to be incorrect. Intuitively, consider the rationally updated expectation of the fundamental θ conditional on the public signal. We want to see if, on average, the date 2 price differs from this expectation. If so, the public signal can be used to predict the subsequent return. For example, assuming that the precisions of the prior and public signal are equal, and the prior is 0 and the public signal is 100, then the rational updated expectation of the payoff will be 50. On average, the unbiased private signal will be 50 (the expected fundamental plus mean zero noise). So even though the private signal is overweighted relative to the public signal in market price, there is no mispricing on average, conditional on the public signal. On average the private signal has zero effect on the expectation, which is a weighted average of 50. So on average there is no conditional mispricing.[2]

To explain the evidence that share prices underreact to corporate announcements documented earlier, a further refinement is needed. Suppose that a good- or bad-news public signal is an event chosen by the firm or some other party in opposition to the private signal. For example, perhaps the firm announces a new equity issue—a bad news event—when the firm is overvalued (that is, overconfident investors received a positive private signal). There is evidence that firms that issue equity are indeed overvalued (Loughran and Ritter 1995; Dong, Hirshleifer, and Teoh 2012). In a similar way, evidence suggests that firms engage

[2] Thus, knowing the public signal does not allow one to forecast the future return from time 2 to time 3. For the interested reader, a formal proof of this assertion is given in Appendix B.

in repurchase—a good-news event—in response to undervaluation (Ikenberry, Lakonishok, and Vermaelen 1995). We call such public signals "selective." To the extent that public signals are selectively undertaken in opposition to preexisting mispricing, such signals will show return continuation, wherein the long-run return after the event is on average of the same sign as the initial market reaction to the event. This implication is consistent with the strong performance of the ISU (issuance) portfolio described earlier.

However, Model 2 still does not deliver the key empirical predictions that there will be both medium-term price momentum (Jegadeesh and Titman 1993) and long-term reversal (DeBondt and Thaler 1985). To deliver these implications we need to consider the psychology of how overconfidence changes over time as people receive feedback from their environments.

Model 3: Dynamic Overconfidence

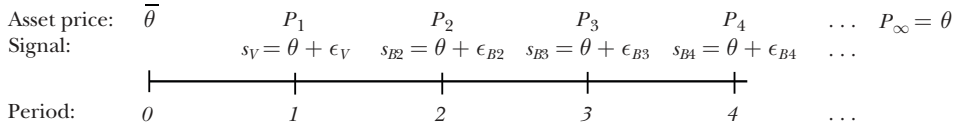
Confidence changes over time as people receive feedback about their judgments and decisions. When people learn that their recent forecasts were accurate, they tend to revise their confidence upward, and when they learn that they were wrong they tend to revise it downward. However, this process is not symmetric, owing to self-attribution bias, which is the tendency of people to treat successes as mainly a reflection of their own skills and failures as mainly a matter of bad luck—the "heads I win, tails it's chance" fallacy (Langer and Roth 1975). Self-attribution bias explains how overconfidence can persist over time.

Incorporating the dynamics of overconfidence into our price formation model allows us to derive more realistic predictions for patterns of return continuation and reversal. To do so, we need to give investors opportunities to update their estimate of their private signal precision. Thus we adopt the structure illustrated in Model 3 (Figure 4), with the change that there are now an unlimited number of public signals arriving at times 2, 3, 4, . . .

Consistent with findings from the psychology literature, we specify that the investor's estimate of private signal precision shifts through time as a function of whether the investor's private signal proves to be consistent with subsequently arriving public signals. This specification for confidence updating is admittedly ad hoc, but is roughly consistent with the psychology literature. In particular, investors update their estimates of their signal accuracy based on their historical forecast success, but in a biased way.

Think of the "cumulative public signal" as the average of all previous public signals. The investor's perceived precision evolves over time based on public signal arrival. The updating rule is that when the arrival of the next public signal pushes the cumulative public signal (and market price) in the direction of the investor's private signal, then the investor becomes more confident in her private signal. So the investor's estimated signal precision increases by a factor of $1 + \overline{k}$. In contrast, if the new public signal pushes the price away from the investor's valuation, the investor loses confidence, and the investor's estimate of her private signal precision

Figure 4
Model 3: Dynamic Overconfidence Model—Timeline



Source: Authors.
Notes: This figure illustrates the dynamic overconfidence model timeline. At time 1, the informed investors receive a private signal s_V. At each subsequent time (2, 3, . . .), the investor receives additional public signals, with uncorrelated noise terms.

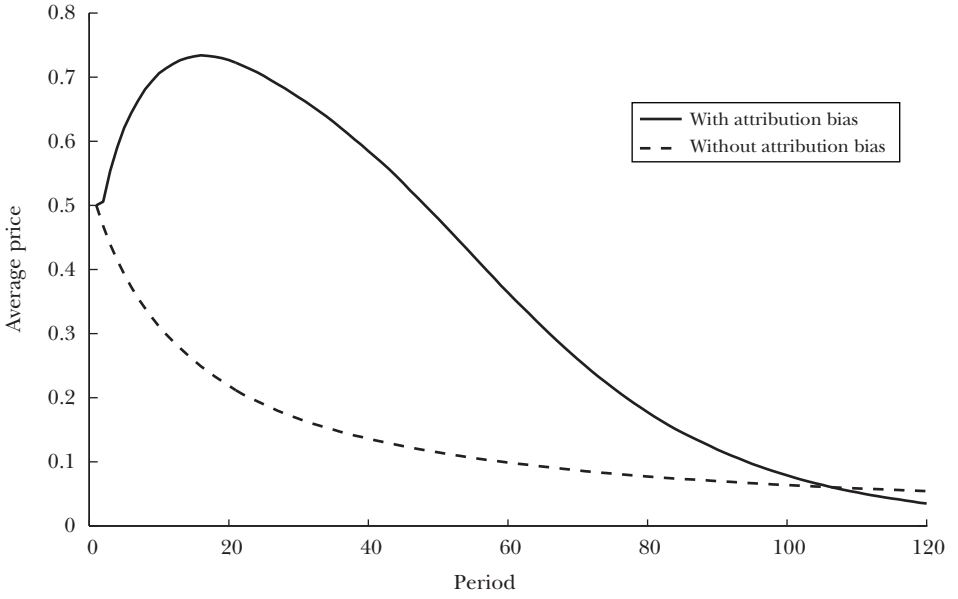
falls by a factor of $(1 - \underline{k})$. Biased self-attribution is captured by the assumption that $\overline{k} > \underline{k}$: the investor's estimated precision increases more with a good outcome than it decreases with a bad outcome.

Figure 5 illustrates the impulse response to a private signal of $s_V = 1$ at time 1, when the mean of the prior distribution is $\overline{\theta} = 0$, the true security value is $\theta = 0$, and the prior and private signal precisions are equal. The dashed line illustrates the price path with static overconfidence (as presented earlier in Model 2). Here, because of the equal precisions, the price at time 1 is 0.5—the average of $\overline{\theta}$ and s_V. However, starting at $t = 2$, with the arrival of the first public signal, the price on average starts to decline, as the average public signal is equal to $\theta = 0$, and converges to the true security value of $\theta = 0$.

The solid line in Figure 5 illustrates the average price path with dynamic overconfidence (as in Model 3). As in the static overconfidence setting, the price initially moves to $P_1 = 0.5$. However now, on observing the sequence of (noisy) public signals, the investor's estimate of private signal precision increases, resulting in continuing overreaction to this original signal—in this example up to about 15 periods. Eventually, as more public signals arrive, the cumulative public signal becomes more precise and the mispricing necessarily converges to zero. The result is a hump-shaped impulse response function. If instead we began with a private signal that was negative, there would be a trough-shaped impulse response function—the reflection across the x-axis of the solid line in Figure 5.

This shape implies momentum at short lags and reversal at long lags. To build some intuition on this point, consider the hump-shape (the long side). The upward slope in the overreaction phase indicates that positive returns tend to be followed by positive returns. The downward slope in the correction phase indicates that negative returns tend to be followed by negative returns. Similar reasoning applies on the short side. In contrast, with a long lag, a positive return on the left side of the hump tends to be followed by a negative return in the correction phase. In sum, a model with self-attribution and dynamic shifts in confidence implies positive short-lag autocorrelations and negative long-lag autocorrelations and is therefore

Figure 5
Response to a Unit Private Signal—Static and Dynamic Overconfidence Models



Source: Daniel, Hirshleifer, and Subrahmanyam (1998).
Notes: This figure illustrates the impulse response to a private signal $s_V = 1$ at time 1 when the true security value is $\theta = 0$. In this simulation, the prior and private-signal precisions are equal. The dashed line illustrates the impulse response in the static overconfidence setting. The solid line is the impulse response in the dynamic overconfidence setting.

consistent with evidence of momentum and long-run reversal discussed earlier. It is also consistent with the strong performance of the UMD (Up Minus Down) momentum-based portfolio described earlier.

Models with Both Rational and Overconfident Investors

In the models so far, prices are set by overconfident investors. How would these conclusions change were we to introduce a mass of rational investors into these models? These investors would act as arbitrageurs, pushing prices toward fundamental values.

In Daniel, Hirshleifer, and Subrahmanyam (2001), we explore such a setting as an extension of the three-date static-overconfidence model explored earlier. In this approach, the market has a continuum of risk-averse investors who start identical to each other. There are N securities, and the joint distribution of their fundamental payoffs is common knowledge. At time 1, investors receive different private signals. Some receive signals about what we call "factor realizations"—common influences that affect the returns of all securities—while others receive signals about what we call "residual payoff components"—the pieces of security payoffs that are not

explained by common factors. Investors are overconfident about the signal they receive: they believe that the precision of that signal is higher than it is actually is. However, the investors who do not receive a signal instead infer the signal as it manifests itself through prices, assess precision correctly, and act as arbitrageurs. Owing to risk aversion, these arbitrageurs eliminate only some of the mispricing.

This setting yields a number of implications for the relationships between risk and return. First, just as in the Model 2 setting, size and fundamental/price ratios are predictors of future security returns. Size is a negative predictor, because a firm that is large in market value will on average be large in part because it is overvalued. This ability of size to predict returns can help to explain the performance of the SMB (Small Minus Big) portfolio described earlier. For a similar reason, fundamental/price ratios (such as earnings-to-price or book-value-to-price) are positive return predictors. Indeed, scaling of price by a fundamental measure can improve return predictability, because a firm can have high price for fundamental reasons, not just because of mispricing. These effects can explain the performance of the HML (High Minus Low) book-to-market-based portfolio described earlier.

A second key implication is that the amount of mispricing will be constrained by the return factor structure, meaning the set of random variables ("factors") that affect the returns of different stocks, and the sensitivities of returns to the different factors. The factor structure affects how risky it is to arbitrage mispricing. When all investors are overconfident, relatively extreme mispricing is feasible. However, when there are arbitrageurs with rational perceptions, high Sharpe ratios become an attractive opportunity to exploit. Such exploitation acts as a constraint on possible mispricing.[3] In particular, in the limit as the number of securities in the market becomes arbitrarily large, it is possible to form portfolios that hedge away factor risk and exploit any mispricing of residual payoff components. Such portfolios are virtually risk-free. This implies that, owing to arbitrage activity, there will be almost no security-specific mispricing (with the possible exception of a small number of securities).

In contrast, to arbitrage the mispricing of a factor (such as the excess return on the market portfolio, or the return on the HML portfolio, both discussed at Table 1), an investor must bear substantial factor risk—the risk that the factor portfolio return could turn out high or low. This implies that in equilibrium, the factor portfolio can remain substantially mispriced. This contrast between almost perfect arbitrage of idiosyncratic mispricing, but not of factor mispricing, comes in part from the assumption that markets are perfectly liquid. For illiquid stocks, arbitrage is more costly, so all stocks can have some idiosyncratic mispricing.

In this setting, regressing across stocks on β (the classic risk measure of the Capital Asset Pricing Model) as well as the fundamental-to-price ratio generally helps disentangle risk premium versus mispricing effects. If overconfidence

[3] More precisely, the flow of wealth from irrational to rational investors becomes arbitrarily large, which clearly is not sustainable.

about signals is extreme and the fundamental is measured perfectly, even though β is priced, it has no incremental power to predict future returns. Intuitively, the fundamental-to-price ratio captures both standard risk effects and mispricing effects—both drive market price down relative to expected future cash flows. In the limiting case in which the firm-specific signal the overconfident investors receive is pure noise, and the fundamental proxy is perfect (the best rational forecast of future cash flows), β does not provide any additional useful information to predict returns. The fundamental/price ratios will eliminate β in a multiple regression when forecasting the cross-section of future returns. This implication is consistent with empirical studies mentioned earlier in which book-to-market eliminates β in predicting returns.

Finally, this model displays excessive disagreement because overconfident investors insist on relying too heavily on the signals they possess and then will trade against rational arbitrageurs who do not possess those signals and do not overweight the signals' precision. An excessively large volume of trade will result. In this way, overconfidence helps to explain the remarkably high volumes of trade in liquid securities.

Summing Up: Linking the Models to the Trading Strategies

We have already discussed how the models in this section can explain the strong performance of the first four trading strategies summarized in Table 1. We close this section by discussing whether overconfidence can help explain the performance of the remaining three trading strategies: ACR (long on low-accrual firms, short on high accrual firms), BAB (long on low-beta stocks, short on high-beta stocks), and IVOL (long on stocks with low idiosyncratic volatility, short on stocks with high idiosyncratic volatility).

We begin with the strong performance of the BAB and IVOL portfolios, which reflects, respectively, the underperformance of stocks with high systematic and idiosyncratic risk. As noted earlier, in a model such as that of Miller (1977) in which there is both investor disagreement about firm value and short-sales constraints, irrational optimists dominate price setting. This implies that when investors disagree more about a firm's future prospects, that firm will be more overpriced and thus will earn lower returns on average. Overconfidence provides a natural explanation for the irrational tendency for investors to be too insistent in disagreeing, and for optimists to fail to fully adjust for the fact that there are pessimists who have been sidelined by short-sale constraints. High-risk firms have greater scope for overconfidence and disagreement, so we expect this source of overpricing to be greatest for high-risk firms. In these ways, overconfidence provides a natural explanation for the idiosyncratic volatility and betting-against-beta effects.

ACR (the accrual anomaly) is usually understood as arising from limited investor attention. The earnings for a firm are the sum of its cash flow and accrual components. The cash flow component of earnings is a much more favorable

indicator than the accrual component of high future profits (Sloan 1996). Investors who do not delve into earnings to evaluate these components separately will tend to overvalue firms with high accruals and undervalue firms with low accruals.

In our view, overconfidence is an important part of understanding return anomalies that are usually attributed solely to limited investor attention. Limited attention has a much bigger effect on price if the investors are overconfident and so fail to recognize that the information they are neglecting is important. A similar point is made by Kahneman (2011), who discusses the tendency of people to be overconfident about fast heuristic judgements (which he calls "System 1").

Cursedness: A Related Approach to Asset Pricing

We make no attempt at a systematic review of behavioral approaches to investment here, but one alternative, cursedness, is notable for its potential overlap with the overconfidence approach.[4] Indeed, Eyster, Rabin, and Vayanos (2013) point out that cursedness can potentially explain several financial economic phenomena that are often understood in terms of overconfidence.

In cursedness, a game-theoretic equilibrium concept developed in Eyster and Rabin (2005), individuals underweight the information implicit in the actions of others. An example is provided by the winner's curse—the phenomenon that those who win a sealed-bid auction often have submitted too high a bid, in which the very fact of winning is an indication that others do not value the object as highly. A sophisticated bidder will make a subtle inference: if I win, others have information that is more adverse than mine. Someone who understands the winner's curse will then tend to bid more conservatively to adjust for the danger of overbidding, or at times choose not bid at all and thus receive a safe outcome of zero.

An overconfident individual who overweights his own signal will, accordingly, also underweight the information implicit in the actions of others, so the overconfidence and cursedness approaches yield overlapping implications. However, the cursedness approach does have some distinct implications. The behavior of an overconfident individual is too aggressive even when others have no signals; in contrast, cursedness only arises when others have signals that the cursed individual might fail to take into account. These distinctions matter for a key argument of Eyster, Rabin, and Vayanos (2013) in favor of cursedness over overconfidence as an explanation for overly aggressive trading. According to their argument, an

[4] Other behavioral approaches include representativeness and conservatism (Edwards 1968; Kahneman and Tversky 1972; Barberis, Shleifer, and Vishny 1998); realization utility (Barberis and Xiong 2012); mental accounting and prospect theory (Kahneman and Tversky 1979; Thaler 1985; Barberis and Huang 2001; Grinblatt and Han 2005); limited attention (Kahneman 1973; Hirshleifer and Teoh 2003; Peng and Xiong 2006); and anchoring (Tversky and Kahneman 1974; George and Hwang 2004).

overconfident investor should still trade little, because the investor should recognize that one personal signal is minor relative to the aggregated signals of millions of other investors (some of whom might be highly expert). In contrast, a cursed investor ignores those other signals, and hence trades too readily.

However, an investor could be overconfident about the uniqueness of a personal signal, not just its quality. Consider a setting where a security's payoff will be $\theta = \theta_1 + \theta_2$, that the investor believes he has a unique signal about θ_1, but that millions of others are observing signals about θ_2. Even with only a moderate level of overconfidence about signal precision, such an investor may trade quite aggressively, despite being fully aware that there are many other informed players in the market.

Furthermore, we believe that cursedness does not go far in explaining the phenomenon of aggressive trading. Many financial economists now believe that the great bulk of individual investors—those who are not insiders, financial professionals, or remarkable amateurs—have little or no useful private information that would allow them to trade profitably in individual stocks. But a poorly informed investor who is only cursed, not overconfident, understands perfectly well that the expected profitability of making a trade is quite small, and moreover, is costly, owing to brokerage fees, time costs, and risk. These frictions or a modest degree risk aversion should easily deter aggressive trading by investors who are cursed but understand that they are ill-informed.

Finally, the empirical evidence summarized earlier in this paper documents short-term return momentum and long-term return reversal in numerous markets. The model of cursedness in Eyster, Rabin, and Vayanos (2013) explains momentum as a pure underreaction phenomenon. As such, it explains momentum but not long-run reversal. The overconfidence approach, in contrast, explains momentum and reversal jointly as parts of a phenomenon of continuing overreaction and sluggish correction (Daniel, Hirshleifer, and Subrahmanyam 1998).

In summary, we believe that cursedness offers a rich approach for understanding economic phenomena. We do not, however, see cursedness, at least taken in isolation, as offering an explanation for the key patterns presented here—excessive trading, short-term momentum, long-term reversal, and the other anomalies summarized in Table 1—that have motivated the use of overconfidence in models of securities markets.

Conclusion

This essay has two main themes: 1) There are anomalies in financial markets—unprofitable active trading, and patterns of return predictability—that are puzzling from the perspective of traditional purely rational models; and 2) models of overconfidence, and of the dynamic psychological processes that underlie overconfidence, can plausibly explain why these patterns exist and persist.

For those readers who are uncomfortable with an explanation for anomalies based on imperfect rationality, we would point out that the empirical patterns of unprofitable active trading and of return predictability are more-or-less agreed upon both by the leading fans of the efficient markets hypothesis and those with a more behavioral bent. For example, the data underlying the three- and five-factor models of Fama and French (1993; 2015) suggest that portfolios can be built that provide high returns with relatively low volatility. The main disagreement is not over the empirical facts described in this paper, but about what components should be added to an asset pricing model to describe them.

We believe that overconfidence offers a useful component, both because of how it explains the agreed-upon facts emphasized here and also because overconfidence promises to help integrate other elements of behavioral finance theory. For example, some authors have emphasized the importance of investor disagreement in understanding financial markets (Hong and Stein 2007). Overconfidence provides a natural explanation for why investors who process the same public information end up disagreeing so much. Limited investor attention has also recently been offered as an explanation for various empirical patterns in trading and prices. Overconfidence explains why investors who neglect important information would nevertheless trade aggressively, so that such neglect can influence price. In these ways and others, overconfidence offers a microfoundation for other important building blocks of behavioral finance models.

■ *We thank Gordon Hanson, Ulrike Malmendier, Enrico Moretti, Timothy Taylor, and Paul Tetlock for helpful discussions and comments and thanks to Chen Wang for excellent research assistance.*

References

Alti, Aydogan, and Paul C. Tetlock. 2013. "Biased Beliefs, Asset Prices, and Investment: A Structural Approach." *Journal of Finance* 69(1): 325–61.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xioayan Zhang. 2006. "The Cross-Section of Volatility and Expected Returns." *Journal of Finance* 61(1): 259–99.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang. 2009. "High Idiosyncratic Volatility and Low Returns: International and Further U.S. Evidence." *Journal of Financial Economics* 91(1): 1–23.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen. 2013. "Value and Momentum Everywhere." *Journal of Finance* 68(3): 929–85.

Aumann, Robert J. 1976. "Agreeing to Disagree." *Annals of Statistics* 4(6): 1236–39.

Baker, Malcolm P., Brendan Bradley and Jeffrey Wurgler. 2011. "Benchmarks as Limits to Arbitrage: Understanding the Low-Volatility Anomaly." *Financial Analysts Journal* 67(1): 40–54.

Baker, Malcolm, and Jeffrey Wurgler. 2000. "The Equity Share in New Issues and Aggregate Stock Returns." *Journal of Finance* 55(5): 2219–57.

Banz, Rolf W. 1981. "The Relationship between

Return and Market Value of Common Stocks." *Journal of Financial Economics* 9(1): 3–18.

Barber, Brad M., and Terrance Odean. 2000. "Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." *Journal of Finance* 55(2): 773–806.

Barber, Brad M., and Terrance Odean. 2001. "Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment." *Quarterly Journal of Economics* 116(1): 261–92.

Barber, Brad M., and Terrance Odean. 2002. "Online Investors: Do the Slow Die First?" *Review of Financial Studies* 15(2): 455–88.

Barber, Brad M., Terrance Odean, and Ning Zhu. 2009. "Do Retail Trades Move Markets?" *Review of Financial Studies* 22(1): 151–86.

Barberis, Nicholas, and Ming Huang. 2001. "Mental Accounting, Loss Aversion, and Individual Stock Returns." *Journal of Finance* 56(4): 1247–92.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. "A Model of Investor Sentiment." *Journal of Financial Economics* 49(3): 307–43.

Barberis, Nicholas, and Wei Xiong. 2012. "Realization Utility." *Journal of Financial Economics* 104(2): 251–71.

Ben-David, Itzhak, John R. Graham, and Campbell R. Harvey. 2013. "Managerial Miscalibration." *Quarterly Journal of Economics* 128(4): 1547–84.

Bernard, Victor L., and Jacob K. Thomas. 1989. "Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium?" *Journal of Accounting Research* 27: 1–36.

Bernard, Victor L., and Jacob K. Thomas. 1990. "Evidence that Stock Prices Do Not Fully Reflect the Implications of Current Earnings for Future Earnings." *Journal of Accounting and Economics* 13: 305–40.

Biais, Bruno, Denis Hilton, Karine Mazurier, and Sebastien Pouget. 2005. "Judgmental Overconfidence, Self-monitoring and Trading Performance in an Experimental Financial Market." *Review of Economic Studies* 72(2): 287–312.

Black, Fischer, Michael Jensen, and Myron Scholes. 1972. "The Capital Asset Pricing Model: Some Empirical Tests." In *Studies in the Theory of Capital Markets*, edited by Michael C. Jensen, 79–121. New York: Praeger.

Carhart, Mark M. 1997. "On Persistence in Mutual Fund Performance." *Journal of Finance* 52(1): 57–82.

Chan, Wesley S. 2003. "Stock Price Reaction to News and No-News: Drift and Reversal after Headlines." *Journal of Financial Economics* 70(2): 223–60.

Choi, James J., David Laibson, and Andrew Metrick. 2002. "How Does the Internet Affect Trading? Evidence from Investor Behavior in 401(k) Plans." *Journal of Financial Economics* 64(3): 397–421.

Collin-Dufresne, Pierre, and Kent Daniel. 2014. "Liquidity and Return Reversals." Columbia GSB working paper.

Cooper, Michael J., Roberto C. Gutierrez Jr., and Allaudeen Hameed. 2004. "Market States and Momentum." *Journal of Finance* 59(3): 1345–65.

Daniel, Kent D., David Hirshleifer, and Avanidhar Subrahmanyam. 1998. "Investor Psychology and Security Market Under- and Overreactions." *Journal of Finance* 53(6): 1839–86.

Daniel, Kent D., David Hirshleifer, and Avanidhar Subrahmanyam. 2001. "Overconfidence, Arbitrage, and Equilibrium Asset Pricing." *Journal of Finance* 56(3): 921–65.

Daniel, Kent D., Ravi Jagannathan, and Soohun Kim. 2012. "Tail Risk in Momentum Strategy Returns." Kellogg School working paper. https://faculty.kellogg.northwestern.edu/models/faculty/m_download_document.php?id=103.

Daniel, Kent D., and Tobias J. Moskowitz. 2014. "Momentum Crashes." Columbia Business School working paper.

Daniel, Kent D., and Sheridan Titman. 1999. "Market Efficiency in an Irrational World." *Financial Analysts' Journal* 55(6): 28–40.

Daniel, Kent D., and Sheridan Titman. 2006. "Market Reactions to Tangible and Intangible Information." *Journal of Finance* 61(4): 1605–43.

DeBondt, Werner F. M., and Richard H. Thaler. 1985. "Does the Stock Market Overreact?" *Journal of Finance* 40(3): 793–808.

Diether, Karl B., Christopher J. Malloy, and Anna Scherbina. 2002. "Differences of Opinion and the Cross Section of Stock Returns." *Journal of Finance* 57(5): 2113–41.

Dong, Ming, David A. Hirshleifer, and Siew Hong Teoh. 2012. "Overvalued Equity and Financing Decisions." *Review of Financial Studies* 25(12): 3645–83.

Edwards, W. 1968. "Conservatism in Human Information Processing." In *Formal Representation of Human Judgment*, edited by Benjamin Kleinmuntz, 17–52. New York: John Wiley & Sons.

Eyster, Erik, and Matthew Rabin. 2005. "Cursed Equilibrium." *Econometrica* 73(5): 1623–72.

Eyster, Erik, Matthew Rabin, and Dimitri Vayanos. 2013. "Financial Markets Where Traders Neglect the Informational Content of Prices." http://personal.lse.ac.uk/vayanos/WPapers/FMTNICP.pdf.

Fama, Eugene F. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance* 25(2): 383–417.

Fama, Eugene F., and Kenneth R. French. 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance* 47(2): 427–65.

Fama, Eugene F., and Kenneth R. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* 33: 3–56.

Fama, Eugene F., and Kenneth R. French. 2015. "A Five-Factor Asset Pricing Model." http://personal.lse.ac.uk/vayanos/WPapers/FMTNICP.pdf.

Frazzini, Andrea, and Lasse Heje Pedersen. 2014. "Betting Against Beta." *Journal of Financial Economics* 111(1): 1–25.

French, Kenneth R. 2008. "Presidential Address: The Cost of Active Investing." *Journal of Finance* 63(4): 1537–73.

Froot, Kenneth A., and Richard H. Thaler. 1990. "Anomalies: Foreign Exchange." *Journal of Economic Perspectives* 4(3): 179–92.

George, Thomas, and Chuan-Yang Hwang. 2004. "The 52-Week High and Momentum Investing." *Journal of Finance* 59(5): 2145–76.

Gervais, Simon, and Terrance Odean. 2001. "Learning to Be Overconfident." *Review of Financial Studies* 14(1): 1–27.

Glaser, Markus, Thomas Langer, and Martin Weber. 2013. "True Overconfidence in Interval Estimates: Evidence Based on a New Measure of Miscalibration." *Journal of Behavioral Decision Making* 26(5): 405–17.

Graham, John R., Campbell R. Harvey, and Hai Huang. 2009. "Investor Competence, Trading Frequency, and Home Bias." *Management Science* 55(7): 1094–106.

Griffin, John M., Xiuqing Ji, and J. Spencer Martin. 2003. "Momentum Investing and Business Cycle Risk: Evidence from Pole to Pole." *Journal of Finance* 58(6): 2515–47.

Griffin, John M., Federico Nardari, and René M. Stulz. 2007. "Do Investors Trade More When Stocks Have Performed Well? Evidence from 46 Countries." *Review of Financial Studies* 20(3): 905–51.

Grinblatt, Mark, and Bing Han. 2005. "Prospect Theory, Mental Accounting, and Momentum." *Journal of Financial Economics* 78(2): 311–39.

Grinblatt, Mark, and Matti Keloharju. 2009. "Sensation Seeking, Overconfidence, and Trading Activity." *Journal of Finance* 64(2): 549–78.

Grossman, Sanford J. 1976. "On the Efficiency of Competitive Stock Markets Where Trades Have Diverse Information." *Journal of Finance* 31(2): 573–85.

Grossman, Sanford J., and Joseph E. Stiglitz. 1976. "Information and Competitive Price Systems." *American Economic Review* 66(2): 246–53.

Hansen, Lars Peter, and Ravi Jagannathan. 1991. "Implications of Security Market Data for Models of Dynamic Economies." *Journal of Political Economy* 99(2): 225–62.

Hansen, Lars Peter, and Ravi Jagannathan. 1997. "Assessing Specification Errors in Stochastic Discount Factor Models." *Journal of Finance* 52(2): 557–90.

Hansen, Lars Peter, and Kenneth J. Singleton. 1983. "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns." *Journal of Political Economy* 91(2): 249–65.

Harrison, J. Michael, and David M. Kreps. 1978. "Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations." *Quarterly Journal of Economics* 92(2): 323–36.

Henderson, Brian J., Narasimhan Jegadeesh, and Michael S. Weisbach. 2006. "World Markets for Raising New Capital." *Journal of Financial Economics* 82(1): 63–101.

Hirshleifer, David. 2001. "Investor Psychology and Asset Pricing." *Journal of Finance* 56(4): 1533–97.

Hirshleifer, David A., and Guo Ying Luo. 2001. "On the Survival of Overconfident Traders in a Competitive Securities Market." *Journal of Financial Markets* 4(1): 73–84.

Hirshleifer, David A., Avanidhar Subrahmanyam, and Sheridan Titman. 1994. "Security Analysis and Trading Patterns when Some Investors Receive Information before Others." *Journal of Finance* 49(5): 1665–98.

Hirshleifer, David A., and Siew Hong Teoh. 2003. "Limited Attention, Information Disclosure, and Financial Reporting." *Journal of Accounting & Economics* 36(1–3): 337–86.

Hong, Harrison, José A. Scheinkman, and Wei Xiong. 2006. "Asset Float and Speculative Bubbles." *Journal of Finance* 61(3): 1073–117.

Hong, Harrison, and Jeremy C. Stein. 2007. "Disagreement and the Stock Market." *Journal of Economic Perspectives* 21(2): 109–28.

Ikenberry, David, Josef Lakonishok, and Theo Vermaelen. 1995. "Market Underreaction to Open Market Share Repurchases." *Journal of Financial Economics* 39(2–3): 181–208.

Jegadeesh, Narasimhan, and Sheridan Titman. 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance* 48(1): 65–91.

Jegadeesh, Narasimhan, and Sheridan Titman. 2011. "Momentum." *Annual Review of Financial Economics* 3(1): 493–509.

Kahneman, Daniel. 1973. *Attention and Effort.* Englewood Cliffs, New Jersey: Prentice Hall.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow.* New York, NY: Farrar, Straus and Giroux.

Kahneman, Daniel, and Amos Tversky. 1972. "Subjective Probability: A Judgment of Representativeness." *Cognitive Psychology* 3(3): 430–54.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47: 263–91.

Keim, Donald B. 1983. "Size-Related Anomalies and Stock Return Seasonality: Further Empirical Evidence." *Journal of Financial Economics* 12(1): 13–32.

Kelley, Eric K. and Paul C. Tetlock. 2013. "Why Do Investors Trade?" https://www0.gsb.columbia.edu/faculty/ptetlock/papers/Kelley_Tetlock_May13_Why_Do_Investors_Trade.pdf.

Kyle, Albert, and F. Albert Wang. 1997. "Speculation Duopoly with Agreement to Disagree: Can Overconfidence Survive the Market Test?" *Journal of Finance* 52(5): 2073–90.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny. 1994. "Contrarian Investment, Extrapolation and Risk." *Journal of Finance* 49(5): 1541–78.

Lamont, Owen A., and Richard H. Thaler. 2003. "Can the Market Add and Subtract? Mispricing in Tech Stock Carve-Outs." *Journal of Political Economy* 111(2): 227–68.

Langer, Ellen J., and Jane Roth. 1975. "Heads I Win, Tails It's Chance: The Illusion of Control as a Function of the Sequence of Outcomes in a Purely Chance Task." *Journal of Personality and Social Psychology* 32(6): 951–55.

Loughran, Tim, and Jay R. Ritter. 1995. "The New Issues Puzzle." *Journal of Finance* 50(1): 23–52.

Malkiel, Burton G. 2013. "Asset Management Fees and the Growth of Finance." *Journal of Economic Perspectives* 27(2): 97–108.

Mehra, Rajnish, and Edward C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics* 15(2): 145–61.

Milgrom, Paul, and Nancy Stokey. 1982. "Information, Trade and Common Knowledge." *Journal of Economic Theory* 26(1): 17–27.

Miller, Edward M. 1977. "Risk, Uncertainty, and Divergence of Opinion." *Journal of Finance* 32(4): 1151–68.

Moskowitz, Tobias J. 2015. "Asset Pricing and Sports Betting." Chicago Booth Research Paper 15-26.

Nagel, Stefan. 2005. "Short Sales, Institutional Investors and the Cross-Section of Stock Returns." *Journal of Financial Economics* 78(2): 277–309.

Odean, Terrance. 1998. "Volume, Volatility, Price and Profit When All Traders are Above Average." *Journal of Finance* 53(6): 1887–934.

Odean, Terrance. 1999. "Do Investors Trade Too Much?" *American Economic Review* 89(5): 1279–98.

Peng, Lin, and Wei Xiong. 2006. "Investor Attention, Overconfidence, and Category Learning." *Journal of Financial Economics* 80(3): 563–602.

Pontiff, Jeffrey, and Artemiza Woodgate. 2008. "Share Issuance and Cross-Sectional Returns." *Journal of Finance* 63(2): 921–45.

Puri, Manju, and David T. Robinson. 2007. "Optimism and Economic Choice." *Journal of Financial Economics* 86(1): 71–99.

Scheinkman, José A., and Wei Xiong. 2003. "Overconfidence and Speculative Bubbles." *Journal of Political Economy* 111(6): 1183–1219.

Sloan, Richard G. 1996. "Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?" *Accounting Review* 71(3): 289–315.

Spiess, D. Katherine, and John Affleck-Graves. 1995. "Underperformance in Long-Run Stock Returns Following Seasoned Equity Offerings." *Journal of Financial Economics* 38(3): 243–68.

Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan. 2012. "The Short of It: Investor Sentiment and Anomalies." *Journal of Financial Economics* 104(2): 288–302.

Statman, Meir, Stephen Thorley, and Keith Vorkink. 2006. "Investor Overconfidence and Trading Volume." *Review of Financial Studies* 19(4): 1531–65.

Tetlock, Paul C. 2011. "All the News That's Fit to Reprint: Do Investors React to Stale Information?" *Review of Financial Studies* 24(5): 1481–1512.

Thaler, Richard H. 1985. "Mental Accounting and Consumer Choice." *Marketing Science* 4(3): 199–214.

Tirole, Jean. 1982. "On the Possibility of Speculation under Rational Expectations." *Econometrica* 50(5): 1163–81.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185(4157): 1124–31.

Weil, Philippe. 1989. "The Equity Premium Puzzle and the Risk-Free Rate Puzzle." *Journal of Monetary Economics* 24(3): 401–21.

Weinstein, Neil D. 1980. "Unrealistic Optimism about Future Life Events." *Journal of Personality and Social Psychology* 39(5): 806–20.

Xiong, Wei, and Jialin Yu. 2011. "The Chinese Warrants Bubble." *American Economic Review* 101(6): 2723–53.

Journal of Economic Perspectives—Volume 29, Number 4—Fall 2015—Pages 89–112

The Ongoing Evolution of US Retail: A Format Tug-of-War[†]

Ali Hortaçsu and Chad Syverson

The past 15–20 years have seen substantial and visible changes in the way US retail business is conducted, with many formerly dominant companies and formats in the sector—for example, Sears, Radio Shack, JCPenney, Circuit City, and a number of shopping malls—struggling to adjust and sometimes suffering fatal blows. Some parts of retail, like traditional department stores as well as book and music stores, have seen large declines in sales and employment. Explanations about what is happening in the retail sector have been dominated by two powerful and not fully consistent narratives: a prediction that retail sales will migrate online and physical retail will be virtually extinguished, and a prediction that future shoppers will almost all be heading to giant physical stores like warehouse clubs and supercenters.

Online e-commerce in retail has been a cultural phenomenon and target of substantial attention in the business and technology media since the late 1990s; many of the most famous "dotcom" busts of the late 1990s were e-retailers. E-commerce has doubtlessly affected important elements of technology, demand, and market structure in the retail sector. Extensive research in the economics literature has explored the rise of e-commerce and its effects on various retail markets (for a survey, see Lieber and Syverson 2012). While physical retail hasn't been killed off by online

■ *Ali Hortaçsu is the Ralph and Mary Otis Isham Professor of Economics, University of Chicago, Chicago, Illinois. Chad Syverson is the J. Baum Harris Professor of Economics, University of Chicago Booth School of Business, Chicago, Illinois. Both authors are Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are hortacsu@uchicago.edu and chad.syverson@chicagobooth.edu.*

[†] For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.29.4.89

doi=10.1257/jep.29.4.89

retail yet, the possibility is often raised. The "death of retail," a term that according to Google Trends emerged in 2009, has been declared in multiple forums.

Although online retail will surely continue to be a force shaping the sector going forward and may yet emerge as the dominant mode of commerce in the retail sector in the United States, its time for supremacy has not yet arrived. Retail sales through the physical format of warehouse clubs and supercenters offer large product lines of goods such as apparel, furniture, and appliances as well as a full line of groceries. Examples include the well-known warehouse clubs Costco and Sam's Club as well as the grocery-plus-department-store formats found in Walmart Supercenters. This segment of the retail sector is just plain large. Its four largest firms accounted for almost 8 percent of total retail sales in 2012. This is almost 50 percent more than *all* e-commerce retail sales in that year. We discuss evidence below indicating that this segment has had a greater effect on the shape of retail over the past 15–20 years than has e-commerce. The current scale and influence of this single sector of physical retail relative to all of e-commerce suggests that while physical retail is likely to continue evolving in the coming years, it is unlikely to meet its demise soon. At the very least, it suggests the potential for an extensive future role for "bricks-and-clicks" hybrids that combine e-commerce and physical platforms.

In this essay, we review changes that have taken place in US retail along these and other dimensions. We begin with an overview of the retail sector as a whole, which over the long term has been shrinking as a share of total US economic activity and in terms of relative employment share. The retail sector has experienced stronger-than-average productivity growth, but this has not been accompanied by commensurate wage growth. We then turn to specific discussions of the aforementioned two main forces shaping the retail landscape in recent decades: e-commerce and warehouse clubs/supercenters. We then look more broadly at changes across the structure of the retail sector, including scale, concentration, dynamism, and degree of urbanization. We conclude with a discussion of the likely future course of the retail sector.

Overview of the US Retail Sector

When we refer to "retail," we are abiding by the sector definitions used by statistical agencies in the United States and many other countries. The North American Industry Classification System (NAICS) used by the statistical agencies of the United States, as well as Canada and Mexico, defines retail trade as entities "engaged in retailing merchandise, generally without transformation, and rendering services incidental to the sale of merchandise." Similarly, the International Standard Industrial Classification (ISIC) defines the sector as "re-sale (sale without transformation) of new and used goods to the general public, for personal or household consumption or utilization." These definitions have two important commonalities. First, retail sells "merchandise" or "goods"—that is, physical objects as well as digital items. Second, it sells them without transformation. These two conditions rule out economic activities that many might think of as retail or at least occur in retail-like settings like strip malls

or sidewalk storefront shops. For example, the definition rules out restaurants and bars (the objects provided are transformed) as well as personal services like barbers, nail salons, repair shops, and the like.[1] The establishments that *are* included in the sector are stores that sell untransformed goods ranging from automobiles to zippers as well as nonstore retailers, who by definition are "organized to serve the general public, but their retailing methods differ." The nonstore retailer definition lists examples of these different methods: "they reach customers and market merchandise with methods, such as the broadcasting of 'infomercials,' the broadcasting and publishing of direct-response advertising, the publishing of paper and electronic catalogs, door-to-door solicitation, in-home demonstration, selling from portable stalls (street vendors, except food), and distribution through vending machines. Establishments engaged in the direct sale (nonstore) of products, such as home heating oil dealers and home delivery newspaper routes are included here."

The Long Arc

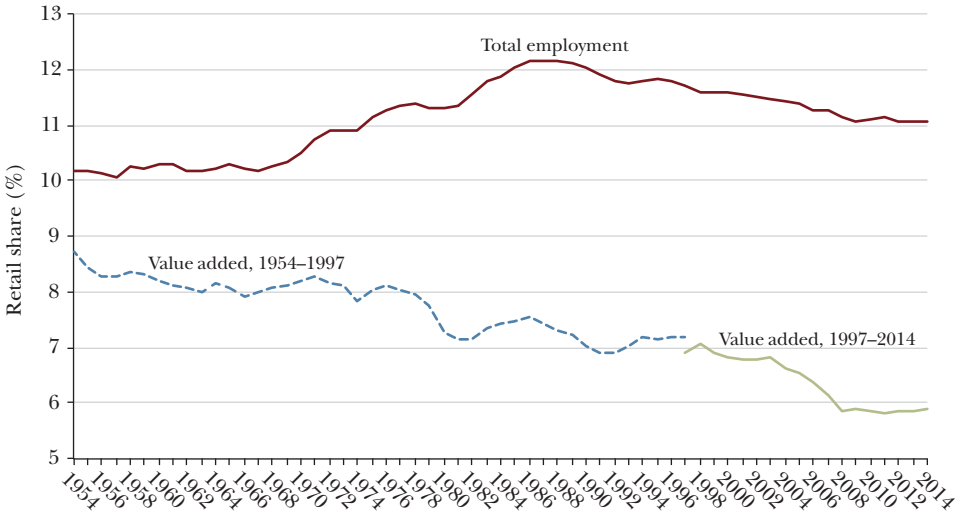
The retail sector's share of total (nonfarm) employment was slightly above 10 percent in 1954. It stayed near this level until around 1970, at which point it started to rise steadily toward a peak of 12.2 percent in 1987. Since then it has fallen back to its current level near 11 percent. Retail's share of value added has not followed this up-and-down pattern. Instead, it experienced a secular decline throughout the period (though at varying rates), dropping from its 1954 start at 8.7 percent to its current level just under 6 percent. Figure 1 shows the evolution from 1954–2014 of retail's share of US economic activity in terms of both employment and value added.

One thing to keep in mind regarding interpretation of these long-run patterns is that, as discussed above, retail by definition sells goods as opposed to services. Goods consumption as a share of the economy has seen a long-run decline. But relative to total goods consumption, retail has *not* been shrinking over the long run. In 1954, retail value added equaled 21.7 percent of the value added of the private goods-producing sectors (agriculture, forestry, fishing, and hunting; mining; construction; and manufacturing) minus net exports of goods. In 2014, this share was about 24 percent, though it has fallen from a peak of 29 percent in the late 1990s. It is also worth noting that despite any drops in overall shares of economic activity, retail's employment and value-added *levels* trended upward throughout this period.

Regardless of longer-run trends, the retail sector has seen little change in its share of economic activity since the onset of the Great Recession. Its employment and value-added shares have held steady at 11.1 percent and 5.9 percent, respectively, since 2008. An impending "death of retail" certainly does not reveal itself in the aggregates.

[1] Restaurants and bars *were* included in the retail sector under the older Standard Industrial Classifications (SIC) taxonomy used in the United States before 1997. Unless otherwise noted, statistics reported below from that period have been adjusted to remove these establishments.

Figure 1
Retail's Share of Total Employment and Value Added, 1954–2014



Source: Authors' calculations from US Bureau of Labor Statistics Current Employment Survey data and the US Bureau of Economic Analysis value-added-by-industry data. The Bureau of Economic Analysis data have a series break in 1997; values for both series are shown in 1997.

The fact that the retail sector's value-added share has been consistently smaller than its employment share indicates that value added per employee—a measure of labor productivity—is lower in retail than in the economy overall. The magnitude of this difference is substantial. In 2014, value added per employee in the nonfarm economy was $124,000, while in retail it was roughly half this level at $66,000. While some of this difference reflects lower average hours per employee in retail, the hours difference is not large enough to explain the labor productivity gap. Bureau of Labor Statistics (BLS) data indicate retail workers averaged about 31.4 hours per week in 2014, about 10 percent below the 34.5-hour average for all nonfarm workers. Value added per worker-hour in retail is therefore still about 40 percent lower than its level in the economy overall.

This difference in labor productivity is reflected in part in an average earnings gap between retail and the overall economy. Total labor compensation per employee in 2013 (the latest year for which data are available at this writing) in the retail sector was just above $35,000, as compared to the analogous non-farm-economy-wide value of $65,000. Again, even adjusting for the fact that the typical employee in retail works fewer hours, retail compensation per hour is still 40 percent lower in retail, commensurate with the value-added gap.

Retail's labor productivity gap was even larger three decades ago, however. The sector has been catching up. While there is some disagreement across data sets in the patterns of productivity growth since 2003—value added per worker reported

by the Bureau of Economic Analysis levels off around 2003 while the Bureau of Labor Statistics output-per-hour labor productivity metric continues to rise through 2014—both series agree that labor productivity growth in the retail sector has outpaced that in the broader economy since 1987. Specifically, real value added per worker in retail in the BEA data grew about 80 percent between 1987 and 2014 (2.2 percent per year). Value added per worker rose only 50 percent over the same period (1.5 percent per year) for the entire nonagricultural economy. Retail output per hour according to the BLS productivity data grew 110 percent (2.9 percent per year) from 1987–2013, in contrast to a 70 percent gain (2.1 percent per year) for all nonfarm private businesses.

Some research has delved into the possible microfoundations of this sectorwide productivity growth. Foster, Haltiwanger, and Krizan (2006) document that within-store productivity growth accounts for a relatively minor portion of sectorwide productivity growth in US retail. Instead, the reallocation of activity across stores drives most of the gains in overall retail productivity, which in turn occurs both through the entry of new, more-efficient firms replacing a set of less-efficient exiting ones, as well as through successful firms adding new stores (rather than expanding their existing ones).[2] Doms, Jarmin, and Klimek (2004) find the productivity levels and growth rates of retail establishments are correlated with their rates of investment in information technologies. This potential for productivity growth driven by information technology evokes Basker's (2012) examination of earlier retail-sector productivity gains harnessed in the sector through the introduction of barcodes. The Institute for Competitiveness and Prosperity (2010) finds, using the World Management Survey data (for example, see Bloom, Lemos, Sadun, Scur, and Van Reenen 2014), that larger retailers employ better management practices than do smaller ones in both the United States and Canada. The increase in scale in the sector discussed below has also coincided with greater product variety in many settings. This too could be a source of productivity growth, and could also be especially relevant for e-commerce as noted by Brynjolfsson, Hu, and Smith (2010).

While it is unclear whether these relationships between technology, management, variety, and productivity are causal, the patterns do suggest possible channels through which productivity shapes the success and survival of retailers. It is also interesting to note that while many of these proposed productivity drivers involve digital and other information technologies, they are as likely to be operating on the "back end" of retail (selection of offerings, distribution, inventory management, and so on) as on the customer-facing "front end" (websites, online advertising, and the like). Thus, the productivity gains of information technologies need not be harnessed exclusively or even primarily though e-commerce retailing. The continued importance of physical operations is a theme to which we return below.

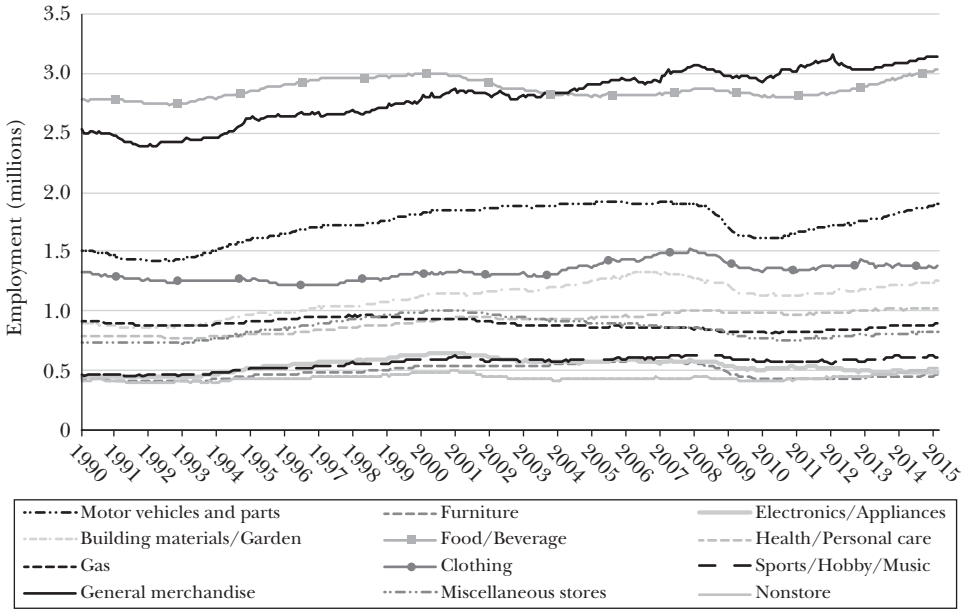
[2] This dominance of reallocation of activity across heterogeneous-productivity stores rather than productivity growth within stores also appears to exist in other countries' retail sectors, as Bronnenberg and Ellickson explain in this issue.

Whatever the sources of retail sector productivity changes, labor earnings growth in the sector has not kept up with them. Total real labor compensation per employee in retail, as reported by the US Bureau of Economic Analysis (BEA) industry accounts, rose an average of 1.2 percent per year from 1987–2003, but fell from 2003–09 and has changed little since, resulting in a decline in real compensation of −0.4 percent per year from 2003–2013. In this sense, the qualitative pattern in retail is not unlike the divergence between economy-wide productivity and some measures of labor compensation during the past several decades (for example, as discussed in Mishel 2012). The productivity–compensation gap is more extreme in retail. In the entire (nonagricultural) economy, total real labor compensation per employee rose 1.4 percent per year from 1987 to 2003 and 0.7 percent per year from 2003 to 2013. Thus, retail compensation growth has been even slower than compensation growth in the economy overall, which itself has been lagging behind productivity. Data from the BLS industry-occupation wage data, available only since 1997, are consistent with this interpretation. Average real annual compensation in the industry across all occupations rose from $28,200 in 1997 to $32,300 in 2003, which is 2.3 percent per year on average, and began to fall afterward, averaging −0.5 percent growth per year from 2003 to 2014.[3]

All in all, while retail labor productivity growth averaged around 2.5 percent per year between 1987 and 2014, average labor earnings growth was only about 0.6 percent during those years, with earnings actually falling after 2003. Labor compensation's share of value added in the retail sector dipped from 57.6 to 54.5 percent between 1997 and 2013. The combination of productivity gains and drops in labor compensation has reduced the sector's unit costs. These cost drops have been captured by two parties: consumers pay a lower retail margin on goods they buy, and payments to the sector's capital holders have risen. Regarding consumers, according to the annual input-output tables from the US Bureau of Economic Analysis, retail margins as a share of sales of all commodities fell from 5.0 percent to 4.7 percent over 1997–2013, and margins on personal consumption expenditures in particular dropped from 11.9 to 10.6 percent during the period. (For reference, total commodity sales were $29.7 trillion in 2013 and personal consumption expenditures were $11.5 trillion.) As to capital owners, gross operating surplus—the part of industry value added not paid as labor compensation or taxes on production—rose from 23.2 percent of value added in 1997 to 24.9 percent in 2013.

[3] These compensation measures are on a per employee basis, so they will not reflect differences in hours per employee across time or sectors. As noted above, however, data from the US Bureau of Labor Statistics indicate that average weekly hours per worker in retail are only about 10 percent less than for workers in the overall economy. To explain the differential growth in compensation per worker since 1987, hours in retail would have needed to have dropped an average of 0.6 percent per year relative to those in the overall economy from 1987–2013.

Figure 2
Retail Employment by Its NAICS 3-digit Component Industries, 1990–2015



Source: US Bureau of Labor Statistics Current Employment Survey data.

Retail Subindustry Changes

These sectorwide trends hide variations in the fortunes of specific industries within the retail sector. Figure 2 shows the evolution of employment since 1990 in each of the 12 three-digit NAICS retail industries, the largest industrial subcategorization within the sector.

Total retail employment has grown 17 percent since 1990. Every component three-digit industry in the sector but one also saw employment growth. The exception was gasoline stations, which saw employment drop by about 2 percent. The industries with the fastest growth rates were building materials and garden stores (39 percent employment growth over the period); sports, hobby, and music stores (32 percent); and health and personal care stores (30 percent). Of the sector's total employment growth of 2.3 million since 1990, from 13.3 to 15.6 million, the three industries contributing the largest portion of these gains were general merchandise stores (gained 630,000 employees); motor vehicles and parts sellers (+400,000); and building materials and garden stores (+360,000). Nonstore retailers, the industry in which the vast majority of online retail occurs, saw 27 percent employment growth over the period. However, the industry's relatively small size meant that this robust growth rate still only accounted for

5 percent of overall retail employment growth. We return to the role of retail e-commerce below.

The three-digit industry experiencing the largest drop in its employment *share* within retail was food and beverage stores, dropping from 21.0 percent to 19.5 percent of retail sector employment between 1990 and 2014. Gas stations' share fell to 9.0 percent from 10.0 percent. On the other hand, general merchandise stores experienced the largest share gain, from 19.0 to 20.3 percent, followed by motor vehicles and parts (despite a drop during the Great Recession), which saw its share rise from 11.3 to 12.1 percent.

While many of these employment patterns are mirrored in these industries' share of total retail sales, one interesting distinction is that the two industries that initially accounted for the largest share of sales—motor vehicles and parts dealers and food and beverage stores—have both seen substantial drops in sales shares over the past 25 years. After peaking at 27 percent of retail sales, the motor vehicles and parts share dropped as low as 19 percent during the Great Recession, before partially recovering to 22 percent. The decline of share in food stores was steadier, with their share falling from 20 percent in 1992 to 14 percent in 2013.

An obvious question to ask is what factors drove these changes, both the sector-wide aggregates and differences across component industries. The postulated effect of e-commerce on bricks-and-mortar retailers could well have differential effects across industries. So could the growth of large-format retail outlets like warehouse clubs. We explore retail e-commerce in more detail in the next section, and then turn to large-format retail in the following section.

E-commerce in Retail

The growth of e-commerce has received extensive attention in the business media and academic literature. Between 2000 and 2014, the fraction of all retail sales accounted for by e-commerce has risen steadily from 0.9 to 6.4 percent, according to figures from the US Census Bureau.[4] The increasing share reflects an 11-fold increase in nominal annual e-commerce sales from 2000 to 2014, in contrast to a 55 percent increase in nominal retail sales during that time. However, even with its recent rapid growth, the miniscule base from which this expansion grew means online commerce is still a small part of retail activity.

The vast majority of retail e-commerce sales—about 85 percent in 2013—occur in the Electronic Shopping and Mail-Order Houses industry (NAICS 45411), a

[4] The US Census Bureau defines e-commerce as "transactions sold on-line whether over open networks such as the Internet or proprietary networks running systems such as Electronic Data Interchange (EDI)," where EDI is itself defined as "the structured transmission of data between organizations by electronic means . . . to transfer electronic documents or business data from one computer system to another computer system. . . . without human intervention." EDI is more applicable to B2B e-commerce than the B2C transactions that define the retail sector (Lieber and Syverson 2012).

subindustry of nonstore retailers, commonly abbreviated ESMOH.[5] This subindustry accounted for only 7.8 percent of the retail sector's shipments and 2.2 percent of its employment. Online sales were 19 percent of ESMOH activity in 2000 but 63 percent by 2013, showing the rise in e-commerce has been substantial. (The division of 2013 employment shares between "Electronic Shopping" and "Mail-Order Houses" subindustries were around 55 and 45 percent, respectively.) Keep in mind that there are several types of retail activities that don't happen in a physical store and also are not e-commerce, even in the retail subindustry (that is, ESMOH), where e-commerce is most dominant.

The three product categories that account for the most online retail sales by Electronic Shopping and Mail-Order Houses are clothing, accessories, and footwear (18 percent of ESMOH e-commerce sales); an "other merchandise" catch-all that includes collectibles, souvenirs, auto parts and accessories, hardware, lawn and garden equipment and supplies, and jewelry (15 percent); and furniture (10 percent).

The remaining 15 percent of online retail sales not taking place within Electronic Shopping and Mail-Order Houses are made by establishments whose primary activities are physical in nature. Among these, the largest share of e-commerce sales comes from motor vehicles and parts dealers. They account for 11 percent of total retail e-commerce sales (and about 70 percent of non-ESMOH e-commerce). The large volume in the motor vehicles and parts stores industry means that these e-commerce sales still only account for 2.9 percent of that industry's sales.[6] Clothing and accessories stores are the only other industry that accounts for more than 1 percent of retail e-commerce sales. Here too this online activity is a small share—1.4 percent—of the industry's total sales.

We combined the product-specific data on e-commerce sales within the Electronic Shopping and Mail-Order Houses category along with e-commerce sales by establishments outside ESMOH to compute online sales as a fraction of total sales for a number of specific products. The shares for 2013 are reported in Table 1 in decreasing order of e-commerce intensity along with the 2013 total sales of the product (e-commerce and otherwise). The total e-commerce sales of these product classes accounted for 65 percent of all retail e-commerce sales in 2013. (Some products sold in high volume online, such as airline tickets, are not considered retail sales as they do not fit the goods-based definition of "merchandise.")

[5] The Electronic Shopping and Mail-Order Houses category is described in the NAICS classification manual in this way: "An industry group comprising establishments primarily engaged in retailing all types of merchandise using non-store means, such as catalogs, toll free telephone numbers, or electronic media, such as interactive television or computer. Included in this industry are establishments primarily engaged in retailing from catalog showrooms of mail-order houses." According to at least one financial information website YCharts, Amazon is classified as operating primarily in this industry (see https://ycharts.com/companies/AMZN).

[6] Note that franchise law restrictions make it extremely difficult for new auto dealers to actually make sales using online-only platforms. See Scott Morton, Zettelmeyer, and Silva-Risso (2001) for more discussion.

Table 1
Product-Specific E-commerce as a Share of Product Total Sales

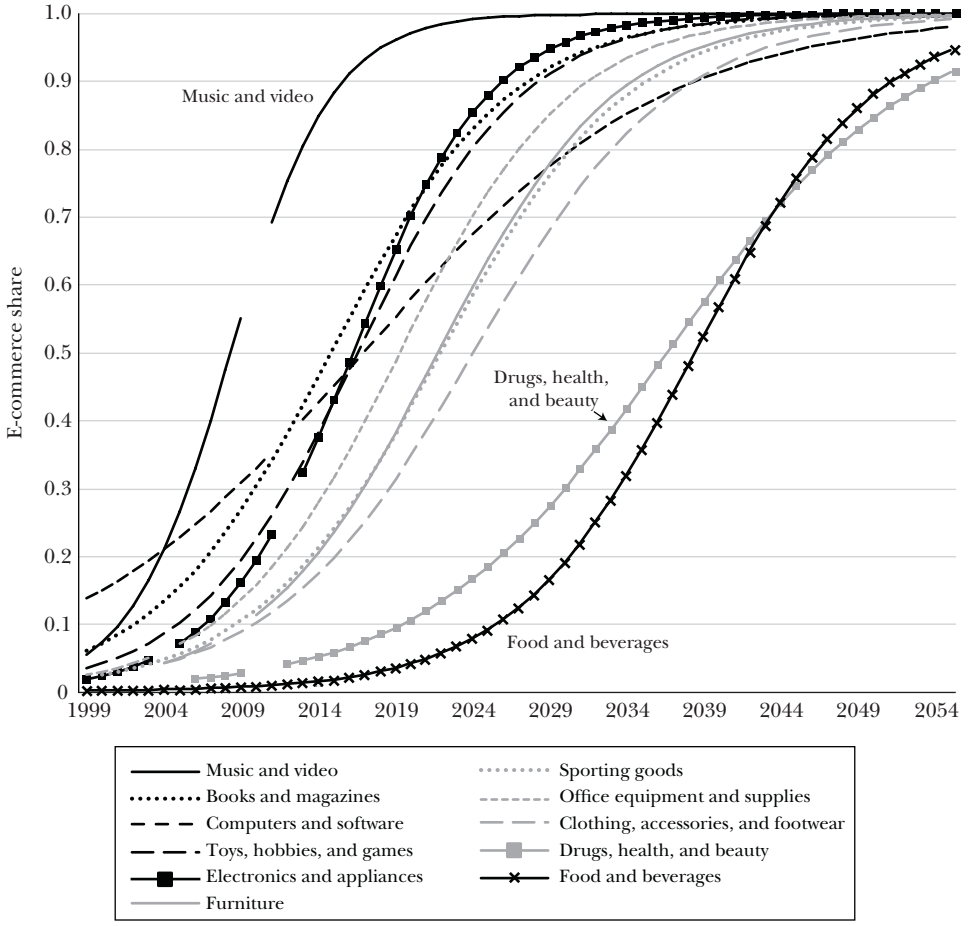
Product category	*E-commerce share of retail sales, 2013*	*Total retail sales (e-commerce and not), 2013 ($ billions)*	*Projected year that product's e-commerce share will be (**bold italics** means reached in the data already):*			
			25 percent	*50 percent*	*75 percent*	*90 percent*
Music and videos	79.6%	$11.8 B	***2005***	***2009***	***2012***	2016
Books and magazines	44.2%	$23.9 B	***2009***	2015	2021	2028
Computer hardware and software	32.9%	$62.3 B	***2006***	2017	2028	2038
Toys, hobbies, and games*	28.8%	$25.5 B	***2011***	2017	2023	2028
Electronics and appliances	23.1%	$102.6 B	***2013***	2017	2021	2026
Furniture	17.5%	$118.0 B	2016	2022	2028	2035
Office equipment and supplies*	17.3%	$24.6 B	2014	2020	2026	2032
Sporting goods	16.9%	$54.1 B	2016	2022	2029	2035
Clothing, accessories, and footwear	14.9%	$291.1 B	2017	2024	2031	2038
Drugs, health, and beauty	4.7%	$374.5 B	2028	2037	2045	2054
Food and beverages	0.9%	$650.9 B	2032	2039	2045	2051

Source: Authors' calculations from US Census Retail E-stats. We computed shares by dividing the sum of the product category's e-commerce sales within and outside Electronic Shopping and Mail-Order Houses (ESMOH) by the sum of total ESMOH sales of the product and total sales of the product's corresponding retail industry. Most product categories in the ESMOH breakout correspond directly to a NAICS store-based retail industry; when not, we apportioned non-ESMOH e-commerce sales proportional to that product's share of total sales within the industry. Computer hardware and software numbers, reported separately in the ESMOH data, were combined for the sake of comparability with the figures from computer and software stores (NAICS 44312). An asterisk denotes that the product's 2013 data is extrapolated from changes during 2010–2012 due to missing 2013 data. The projections are predictions from a logistic diffusion model fit to products' observed e-commerce shares through 2013. Figures in italics were reached within the sample. The model assumes the saturation (asymptotic) share of each product is 100 percent; earlier attempts to estimate the saturation share as an additional parameter gave unrealistically low long-run shares.

In a result that will surprise no one, the most e-commerce-intensive product category in the data is the music and videos category, with 79.6 percent of all sales in 2013 conducted via e-commerce. Books and magazines were the second most dependent on online retail platforms, with 44.2 percent of their sales online. Following that were computer hardware and software at 32.9 percent and toys, hobbies, and games at 28.8 percent.

Music and videos therefore appear to have almost saturated their e-commerce potential. But online retail still has considerable ability to expand in other categories in the future. Indeed, the e-commerce shares of some of the largest product classes—such as clothing, accessories, and footwear; drugs, health, and beauty; and food and beverages—are the lowest. To attempt to quantify the likely expansion of

Figure 3
Diffusion Curves for Shares of E-Commerce Sales in Eleven Product Categories



Source: Authors' calculations using data from US Census Retail E-stats.
Notes: To attempt to quantify the likely expansion of e-commerce shares for various product categories listed in Table 1, we fit S-curves (specifically, logistic diffusion curves) to the e-commerce shares for each product category using available annual data from 1999–2013. Projections are not plotted for years before 2014 with missing data. We assume a saturation parameter of 1; that is, we assume that given enough time, potentially all of these products may be sold completely on line. This is a highly speculative exercise.

e-commerce in the various product categories, we fit S-curves (specifically, logistic diffusion curves) to the products' e-commerce shares using available annual data from 1999–2013 (see Figure 3). Of course, this exercise is highly speculative. We do not have many data points on which to fit the curves. For some products the problem is worse because e-commerce sales are not reported in some years due to confidentiality or data quality reporting restrictions. Further, we assume a saturation

parameter of 1; that is, we assume that given enough time, potentially all of these products may be sold completely online.[7]

With our reservations duly stated, the diffusion curve estimates suggest that many of the product categories could see considerable share growth over the next decade, as shown in the last four columns of Table 1. Of the product classes, all but two (drugs, health, and beauty, as well as food and beverages) are projected to hit 50 percent e-commerce shares by 2024 (the music and videos category has already surpassed this level, of course). The same products are all projected to reach 75 percent e-commerce shares by 2031. Some of these product classes have sales that are quite substantial, with a few categories having total sales of over $100 billion in 2013. These results suggest it is not outlandish to believe that annual online sales might increase by hundreds of billions of dollars over the next decade.

On the other hand, the potential growth by 2024 of e-commerce relative to *all* retail sales is still modest. Total e-commerce retail sales in 2013 were $260 billion, or 5.8 percent of total retail. Even supposing all categories projected to reach 50 percent online sales by 2024 actually hit 100 percent, that is only the 2013 equivalent of an additional $570 billion of online sales. The share of e-commerce in total retail sales would still be less than 20 percent in that (rather extreme) case. Recall that, to date, the largest tracked categories in terms of total retail sales (drugs, health, and beauty; and food and beverages) have the smallest online shares. Until online sales diffuse more deeply into these categories, the bulk of retail will remain physically based. (Our diffusion curves predict 50 percent shares by the late 2030s for the drugs and food product classes, though we are extrapolating extensively in these cases.)

We are wary about putting too much weight on these results due to the caveats mentioned above, but a conservative interpretation would be that there will be considerable across-product variation in the timing and depth of the growth of e-commerce. More broadly, predictions of an impending demise of physical retail have been greatly exaggerated. Even rather optimistic projections about e-commerce growth still leave a considerable amount of activity to physical establishments in coming years.

The Rise of the Warehouse Club/Supercenter Retail Format

Not only should we expect physical formats to remain a substantial factor in the retail sector over the foreseeable future, over the most recent past decades a

[7] We also estimated a specification that fit products' saturation levels (that is, asymptotic shares) as separate parameters to allow for the possibility that sales of certain products never completely move to online platforms. However, this yielded unrealistically low estimates of asymptotic share. Perhaps this was because we were often extrapolating diffusion curves before an obvious inflection point. To our eye, that estimation routine appeared to take any excuse for an inflection point in the data as bona fide, with the routine typically projecting a product's asymptotic share as less than 5 percentage points above its observed share in 2013.

particular physical format has arguably had an even greater impact on retail than has e-commerce. That format is the warehouse club/supercenter. Some basic summary statistics offer prima facie evidence of the outsized role of this emerging format in the sector.

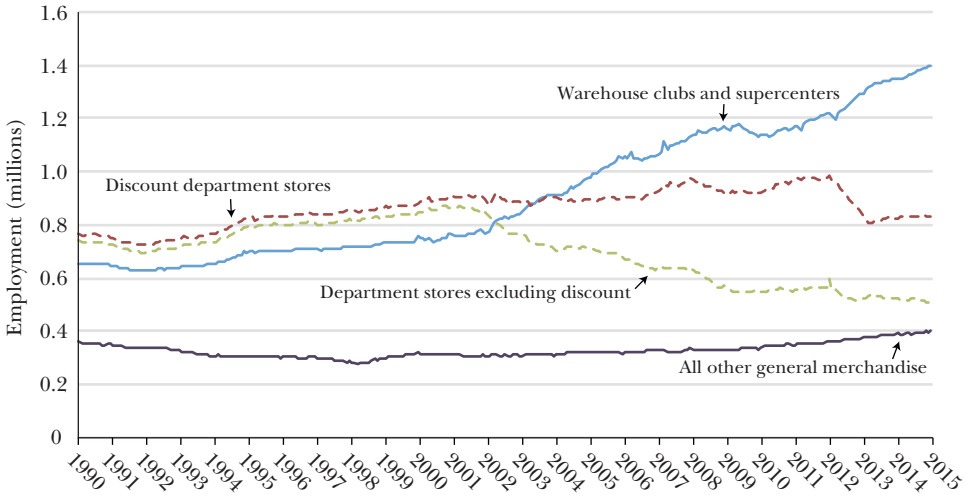
The retail sector is divided by the NAICS taxonomy into twelve three-digit industries, those shown in Figure 2. In turn, these three-digit retail industries are subdivided into 27 four-digit and 58 five-digit subindustries. Electronic Shopping and Mail-Order Houses had the second-largest growth rate in nominal sales between 1992 and 2013 among the five-digit subindustries, experiencing a tenfold rise from $35 billion to $348 billion. However, the fastest growth rate among the five-digit subindustries was observed in Warehouse Clubs and Supercenters (NAICS 45291). The NAICS manual describes the industry comprising "establishments known as warehouse clubs, superstores or supercenters primarily engaged in retailing a general line of groceries in combination with general lines of new merchandise, such as apparel, furniture, and appliances." While the federal statistical agencies cannot report the industry classification of any specific identifiable establishment or firm, clearly the well-known discount warehouse clubs like Costco and Sam's Club fit this definition. It also appears that the newer, larger-format Walmarts ("Walmart Supercenters"—those that carry a full line of groceries), Targets ("SuperTargets"), and Kroger Marketplace and Meijer stores fit in here as well. However, this category requires that the store offer a general line of groceries, which means many "big-box" format stores do *not* fall within this industry. Sales in the warehouse clubs and supercenters subindustry grew 10.5 times between 1992 and 2013 from $40 billion to $420 billion. In both growth rates and actual dollars, then, the expansion of this subindustry outstripped growth in ESMOH.

A direct comparison of some of the major players in each segment bolsters the notion that warehouse club/supercenter growth has exceeded the astounding growth in e-commerce. Amazon, perhaps the largest company operating in Electronic Shopping and Mail-Order Houses in terms of revenues, reported in annual financial filings an increase in US sales of about $38 billion between 2000 and 2013.[8] The largest warehouse club chain, Costco, saw its US sales rise by $50 billion over the same period. The Sam's Club warehouse club division of Walmart added $32 billion in growth during this time.

More specific elements of the timing of the warehouse club/supercenter boom also point to its role in driving the decline of alternative specific retail formats. Figure 4 plots employment for the four component industries of the general merchandise stores industry (NAICS 452), of which warehouse clubs and supercenters is one component, along with discount department stores, non-discount department stores, and a residual "other general merchandise" category. Employment in each of the first three industries grew at roughly the same pace throughout the 1990s. At the turn of the millennium, however, the series diverge. Warehouse club/supercenter

[8] Amazon only offers geographic breakouts of revenues into North America and International categories. We assumed 90 percent of North American sales are in the US market.

Figure 4
General Merchandise Retailing Employment by Industry Group, 1990–2015



Source: US Bureau of Labor Statistics, Current Employment Survey data.

employment starts climbing swiftly, roughly doubling between 2000 and 2014, with only a mild hiccup in 2009. Employment at traditional (nondiscount) department stores, on the other hand, began to shrink. In total, warehouse clubs/supercenters have added 660,000 jobs between 2000 and the start of 2015 even as traditional department stores have shed 350,000 jobs. The patterns in sales revenues are even starker. Between 1992 and 2013, as warehouse clubs/supercenters saw a 10.5-fold increase in nominal sales, traditional department stores revenues *fell* by 18 percent, down 37 percent in nominal revenues from their 1999 peak. (For context, the Consumer Price Index rose 66 percent during this period.)

The retail sector has therefore seen a major shift in the way that stores selling multiple varieties of merchandise operate, with a shift from the traditional service-oriented department store toward a lower-cost model that in some dimensions borrows the logistics techniques of the wholesale sector. While some high-end department stores have been able to stave off decline by focusing on higher-income shoppers desiring extensive service, those in the middle have struggled.

The coincident timing of the expansion of warehouse clubs/supercenters and the contraction of traditional department stores points to the possibility that the former grew at least partially at the expense of the latter. Some of the most substantial changes within the retail sector may be largely incidental to the growth in online commerce rather than a result of it.

The geographic and across-product patterns in the data also point to the expansion of warehouse clubs/supercenters as a key driver of the contraction of the old-line department stores. Using County Business Patterns (CBP) data from 2003 and 2013, we regress the ten-year change in a county's number of establishments classified as

department stores on the county's change in warehouse club/supercenter stores. We also included the county's change in its logged total establishments across all industries to control for overall economic growth in the county, as well as the change in the log number of retail establishments to control for any county-specific changes across the entire retail sector. The results indicate that counties that saw larger increases in the number of warehouse club/supercenter stores (relative to growth in the size of the county's overall economy and its retail sector specifically) saw larger declines in their number of department stores. The coefficient indicates that every extra warehouse club/supercenter store is associated with a drop in the number of department stores of 0.686 (standard error = 0.086). Mean changes in counties' store counts were 0.67 for warehouse clubs/supercenters and −0.40 for department stores, so the mean increase in warehouse clubs/supercenters quantitatively predicts the mean change in department stores reasonably closely.[9]

To gauge the broader local effects of warehouse clubs/supercenters across retail segments, we repeated this exercise while replacing the county's change in number of department stores with its change in store counts for nine of the product categories explored above in the e-commerce diffusion analysis (music and video stores and computer and software stores were dropped as separate industries in the 2013 County Business Patterns data, so we do not include them in our analysis). In four of the nine product categories (books and magazines; toys, hobbies, and games; furniture; and office equipment and supplies), warehouse club/supercenter expansion in a county had a negative and significant (5 percent level) association with the growth in that product's stores in the same county. Four categories exhibited an insignificant relationship: electronics and appliances, sporting goods, clothing and accessories, or food and beverages. There was a positive and significant relationship between warehouse club/supercenter growth and drugs, health, and beauty stores.

Shifts in Retail Market Structure

The patterns in the retail sector involving e-commerce and warehouse club/supercenter stores have been accompanied by a number of systematic changes in retail market structure that we document in this section. These include increases

[9] For the data and full results of the regressions described here, see the online Appendix available with this paper at https://www.aeaweb.org/jep/. Our sample contained 3,136 counties. There are 3,196 counties in the County Business Patterns data; 60 were dropped because they did not have at least one retail establishment in either 2003 or 2013. If we run the specification using changes in logged department store and warehouse club/supercenter establishments, which limits the sample to 835 counties with nonzero establishment counts of both types in both years, the estimated elasticity is −0.193 (standard error = 0.031). Both results are robust to also including the change in the number of ESMOH establishments in the county, though the magnitude of the warehouse club/supercenter count coefficient falls by about one-third in the levels specification. While one might at first glance suspect it is unlikely for the ESMOH sector to have local effects given the nature of their business, Hortaçsu, Martinez-Jerez, and Douglas (2009) show that a disproportionate amount of online-platform-based transactions take place between a buyer and seller living in the same narrowly defined geographic region.

in the average scale of retail operations, increasing concentration within the industry, a reduction in business dynamism, and a modest shift in retail activity toward more populated areas.

Increase in Scale

The average scale of operations has been increasing in retail. Based on the comprehensive Statistics of US Businesses data, between 1998 and 2012 average retail firm size (measured by employment) grew by 18 percent, from 19.3 to 22.8 employees per firm. This rise considerably exceeded the more modest 4 percent change, from 19.4 to 20.2 employees per firm, in the overall US economy. Just over half of this increase in size came from a larger scale of operations at the individual retail establishment (that is, the individual store), which grew from 12.8 to 13.9 employees per establishment. The remainder came from an increase in the average number of establishments per retail firm from 1.51 to 1.63. In contrast, virtually all of the increase in average firm size throughout the broader economy was the result of an increase in establishments per firm rather than employees per establishment. This increase in retail operational scale is part of a global trend (as discussed by Bronnenberg and Ellickson in this journal issue).

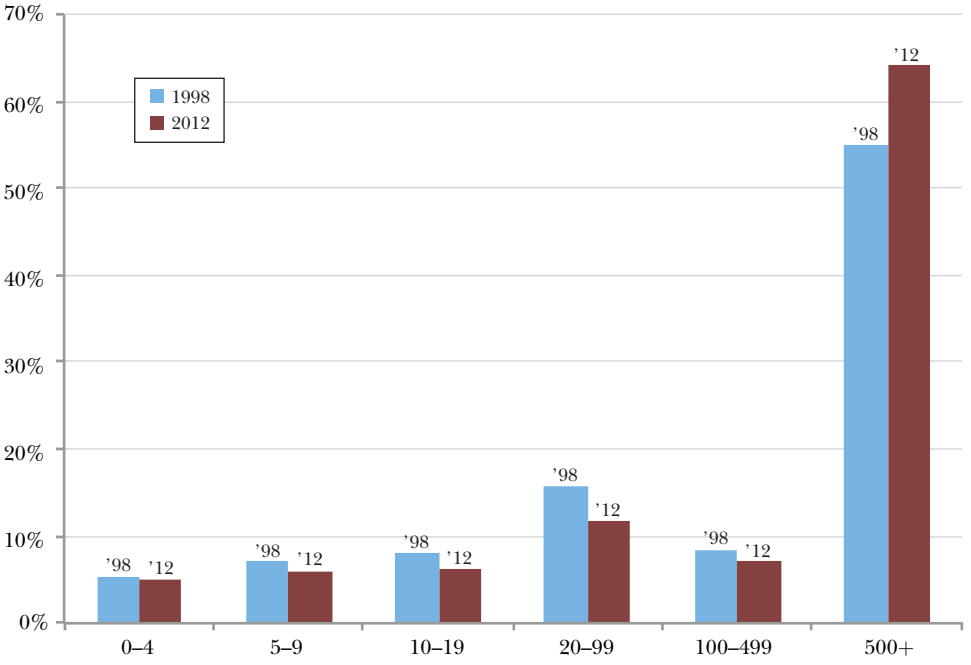
This shift in mean size of retail firms was completely due to growth in the upper tail of the firm size distribution. Figure 5 shows the fraction of retail employment accounted for by firms of various size categories. All size categories of fewer than 500 employees, which include 99.7 percent of all retail firms, saw drops in their share of retail employment. Of the 8.9 percentage point gain in 500+ employee firms' share, 3.9 percentage points came from a drop in the share of firms with between 20 and 99 employees. Again, these patterns echo qualitatively similar but quantitatively smaller shifts in the overall economy, where 500+ worker firms saw a share gain of 2.6 percentage points.

The scale of shopping centers—collections of retail establishments owned by different firms—has also increased over the long run, though it has been stagnant for the past decade. Based on figures from the International Council on Shopping Centers, the average shopping center size in the US grew from about 77,000 square feet in 1970 to 92,000 square feet in 2014.[10] However, it has been at that 92,000 square foot level since 2004, so much of the scale up occurred before the growth in establishment and firm sizes discussed above.

One potential factor that could be pushing toward greater scale in the retail sector is the increasing importance of network economies among chain stores. For example, economies of scale in procurement, logistics, or brand, would all encourage a larger scale of operations, at least at the firm level. There has been

[10] The Council defines a shopping center as "a group of retail and other commercial establishments that is planned, developed, owned and managed as a single property, typically with on-site parking provided." The Council provides data on shopping center counts by size category. We computed the overall average center size by assuming the average size center within each category was at the simple mean between the category's endpoint square footages. Centers in the largest category (over 1,000,000 square feet) were assumed to have an average size of 1,250,000 square feet.

Figure 5
Share of Retail Employment by Firm Size, 1998 and 2012



Source: Authors' calculations from the Statistics of US Businesses, US Census data.

extensive research on these network mechanisms in retail (for example, Holmes 2001; Ellickson 2007; Jia 2008; Holmes 2011; Ellickson, Houghton, and Timmins 2013; Nishida 2014). In addition to the potential effect of network economies on productivity growth in the retail sector, Bertrand and Kramarz (2002) show that the absolute scale of the sector itself might be affected. They demonstrate that entry regulations in France that discouraged large retail formats stunted the growth of the retail sector overall.

Again, the warehouse clubs and supercenters retail format plays an important role in explaining these sectorwide patterns. In 1998, employment in the warehouse club/supercenter industry was just under 450,000, already a nontrivial 3.2 percent of overall retail employment. By 2012, employment in the warehouse club/supercenter industry was nearly 1.4 million, almost 10 percent of the sector's 14.8 million total. Average firm employment in the industry rose 13-fold from 1998 to 2012, though most of this was through expansions in the format's number of stores per firm (from 13 to 161) rather than employees per store (251 to 270). Scale growth in retail would have been notably less in absence of the expansion of warehouse club/superstore companies. Excluding warehouse clubs/superstores, average employment per retail firm grew only 10 percent as opposed to the 18 percent gain once warehouse clubs/superstores are included.

Table 2
Average Annual Payroll per Employee, by Firm Size, 2012

Firm size category (number of employees)	*Retail sector*	*Overall economy*
Total	**$25,500**	**$46,700**
0–4	$24,200	$40,300
5–9	$23,600	$34,400
10–19	$26,300	$36,500
20–99	$32,800	$40,400
100–499	$35,500	$44,900
500+	$23,200	$52,600

Source: Authors' calculations from the Statistics of US Businesses, US Census data.

The story is less clear when it comes to considering how growth of the Electronic Shopping and Mail-Order Houses sector has affected the size of retail operations. We found in previous work with coauthors that the advent and diffusion of e-commerce skewed the size distribution to the right in the two retail industries we examined: bookstores and auto dealers (Goldmanis, Hortaçsu, Syverson, and Emre 2010). The mechanism leading to this change was that e-commerce technologies reduced search costs and led to a concomitant increase in consumers' ability to substitute among sellers. This favored lower-cost, higher-quality firms within the industry, pushing a greater share of activity toward them. While we only empirically tested this mechanism for two industries in the retail sector, conceptually it could act more broadly across other retail markets. This finding would imply that e-commerce also had a part in increasing the typical scale of operations in the sector. On the other hand, the more direct, compositional effect of ESMOH on scale in retail pushes in the direction opposite the sectorwide trend. Average scale in Electronic Shopping and Mail-Order Houses has *fallen* as the industry has grown. While average employment per firm in the industry was 25.7 in 1998, it had dropped by more than half, to 12.6, by 2012. Almost all of this change was the result of a decline in average employment per establishment from 23.5 to 12.1 rather than a reduction in establishments per firm. Companies with 500 or more employees accounted for 58.7 percent of the industry's employment in 1998, but only 48.9 percent in 2012. Thus the upscaling of the typical retail business has happened not because of scale changes within ESMOH, but in spite of them. These numbers indicate that the representative ESMOH firm isn't Amazon; it is instead more likely to be a small vendor selling its wares using Amazon's platform.

Interestingly, and evoking the earlier results that labor compensation has lagged productivity growth in the retail sector, there is no clear firm-size wage premium in retail, at least as measured via payroll per employee from the Statistics of US Businesses data (which is itself compiled from tax data). As shown in Table 2, payroll per retail employee was $25,500 in 2012. For comparison, average

Table 3
Changes in Concentration in the Retail Sector, 1997–2007

Share of sector sales accounted for by:	*1997*	*2007*
4 largest firms	7.9 %	12.3%
8 largest firms	11.7%	17.5%
20 largest firms	18.5%	25.4%
50 largest firms	25.7%	33.3%

Source: US Economic Census.

compensation for workers at retail firms with more than 500 employees—the segment where the sector's growth is concentrated—was only $23,200. There is a positive wage gradient among smaller retail firms; average compensation at retail firms with between 100 to 499 employees was $35,500 while it was $24,200 for firms with fewer than five employees, for example. But the level falls considerably at the largest firms. This pattern is not driven by fewer hours per employee in large retail companies. Bureau of Labor Statistics hours data don't offer breakouts by firm size, but the 2013 Current Population Survey indicates that average weekly hours (among both full and part-time workers) were 38.0 hours in retail firms with fewer than 500 employees and 36.0 hours in larger firms. This 6 percent gap is not enough to close the over 30 percent difference in compensation per employee. This reversal of the wage gradient with firm size also stands in contrast to the overall private economy, where average payroll in 2012 ranged from a low of $34,400 for firms with 5 to 9 employees to $52,600 for those with more than 500 employees. This pattern is consistent with a large set of research documenting a firm size wage premium (for example, Troske 1999).[11]

Concentration

As retail firms have become larger, the sector has also become more concentrated. While comprehensive data from the 2012 Economic Census (the most recent) are not yet available, Table 3 indicates a clear trend toward concentration based on the changes observed between the 1997 and 2007 Economic Censuses. The largest four firms in the retail sector accounted for 7.9 percent of total retail sales in 1997. By 2007, the four largest firms accounted for 12.3 percent. The market shares accounted for by the largest 8, 20, and 50 firms also increased substantially over the period.

[11] The reversal of the firm-size wage gradient that we find here is qualitatively similar to, though much larger than, what Cardiff-Hicks, Lafontaine, and Shaw (2015) found in Current Population Survey data. They also found a nonmonotonic pattern in retail wages with firm size, even controlling for standard worker observables. While they defined the sector more expansively than we do, we found similar results in their subsample that overlaps with our sector definition. We thank them for making their data available to us.

Focusing again on warehouse clubs/supercenters specifically, while concentration did mildly increase during the period, what is most notable is how concentrated the industry already was by 1997. The four-firm concentration ratio at that time was 89.6 percent, and the eight-firm ratio was 99.4 percent. While the subindustry expanded, it also became more concentrated by 2007, with the four- and eight-firm concentration ratios having risen to 93.9 and 99.9 percent, respectively. In 2007, the four largest warehouse club/supercenter companies accounted for 7.8 percent of *all* retail sales (up from 3.0 percent in 1997).

Again, the changes observed in Electronic Shopping and Mail-Order Houses move opposite the sectorwide trends. The industry is considerably less concentrated than warehouse clubs/supercenters and seems to have become even less concentrated, if only slightly, over time. Its four-, eight-, 20-, and 50-firm concentration ratios in 1997 were 24.4, 32.0, 47.3, and 63.3 percent. These values shifted to 21.1, 32.4, 46.2, and 59.1 percent in 2007. These changes are likely related to the drop in the average scale of ESMOH businesses discussed above.

Declining Dynamism

US retail has also seen a downward trend in business dynamism, at least as measured by firm entry and exit rates or the amount of job reallocation across firms. In this journal, Decker, Haltiwanger, Jarmin, and Miranda (2014) show that, echoing patterns observed in the broader economy, the share of the retail sector's employment accounted for by young firms has been shrinking since 1982. Firms that were under six years old (and thus relatively recent entrants into the sector) accounted for 27 percent of retail employment in 1982. That had fallen to 20 percent by 1992, 16 percent by 2002, and 14 percent by 2012. This rate of decline is larger than the entry slowdown observed in the overall economy during the same period.

One potential explanation for this reduction in the number and size of young retail firms, consistent with the results above, is suggested by Jarmin, Klimek, and Miranda (2005): that activity in the sector has shifted away from small companies, especially those with only one store—the proverbial "mom and pop" operations. Much of the entry activity in earlier years was likely due to these types of operations. However, it should be noted that the slowdown in dynamism is happening economy-wide, which indicates that additional factors might be at play within retail.

Urbanization

One other shift in market structure, quantitatively less notable than the changes in scale or concentration, is a move in the retail sector's activity toward more populated areas. In 2003, 1.3 percent of retail establishments were located in the smallest quintile of counties (as measured by total employment across all sectors). The fractions in the second, third, and fourth quintiles were 3.1, 6.2, and 12.4 percent, respectively. The remaining 77.0 percent were in the largest quintile of counties. By 2013, this fraction had grown to 78.0 percent, while dropping in each of the other quintiles. The fractions in the first through fourth quintiles were 1.2, 2.8, 5.8, and 12.2 percent.

Over the same period, the warehouse club/supercenter stores and Electronic Shopping and Mail-Order Houses industries, both already much more likely to locate in more populated areas in 2003 (with 74.8 and 86.1 percent of their respective establishments located in the largest quintile of counties), saw slight changes in these fractions by 2013, but in opposite directions. The share of warehouse club/supercenter establishments in the largest quintile fell to 72.6 percent as the total number of establishments in this industry rose 69 percent. For ESMOH, the share in the largest quintile rose to 88.3 as the total number of establishments increased by almost 94 percent.

What's Next for US Retail?

The future trajectory of the retail sector can be broken down into specific questions about overall growth, growth of subindustries like e-commerce and warehouse club/supercenter, productivity, payments to factors, and costs to final goods consumers.

For overall growth, the key question is whether the long-run trend of the retail sector shrinking relative to the rest of the economy will continue, or whether instead the shorter-run stable share seen since 2008 will hold.

For some product categories, the online retail component of retail shows no sign of slowing yet, but may reach saturation within the next decade. The growth of the warehouse club and supercenter format has equaled that of e-commerce since 2000, and evidence on the timing, location, and market structure changes in retail suggest that the format in recent years has played an even stronger role in shaping the sector than has online retail. That said, sales growth of this subindustry since 2007 has fallen somewhat relative to the fairly constant growth of e-commerce.

If the retail sector continues to see labor productivity gains in excess of the economy-wide average, its employment share will fall even if its value-added share remains constant. Continued productivity growth for the retail sector as a whole is certainly plausible: after all, average sales per employee is considerably higher in Electronic Shopping and Mail-Order Houses (at $1.17 million in 2013) than in retail overall ($296,000).

One can imagine the future of the retail sector as being pulled in one direction by the growth of e-commerce, which involves smaller employment firms, less market concentration, more geographical dispersion, and higher productivity. At the same time, the sector is being pulled in another direction by the warehouse clubs and supercenters, with higher employment firms, very high market concentration, location near population centers, and lower productivity relative to online channels. While warehouse clubs/supercenters have had more influence on the sector to this point, e-commerce has had its own effects and may be growing in relative importance. Perhaps this concurrent expansion and strength of e-commerce and a physical format portends a retail future not dominated by either, but rather with a substantial role for a "bricks-and-clicks" hybrid. The formats may end up

being as much complements as substitutes, with online technologies specializing in product search and discovery, and physical locations facilitating consumers' testing, purchase, and returns of products (A.T. Kearney 2014).

Whichever retail format eventually predominates will not just shape a considerable share of economic activity but will also sculpt the look and feel of our public spaces. Physical retail is a necessarily social and public process. The archetypical look of an historic era depends in no small part on the look of the retail space, from the town squares and downtown streets of the early and mid 20th century, to the malls of the 1980s, to the more recent big-box store islands floating in parking lot seas. The market will determine much about the space in which we will shop, including whether we will shop with others or mostly shop alone at home.

■ *We thank Gordon Hanson, Enrico Moretti, and Timothy Taylor for comments. We also thank Mattie Toma for excellent research assistance, and we are grateful to Jeff Severts who provided both excellent research assistance and helpful conversations.*

References

A. T. Kearney. 2014. "On Solid Ground: Brick-and-Mortar Is the Foundation of Omnichannel Retailing." https://www.atkearney.com/documents/10192/4683364/On+Solid+Ground.pdf/f96d82ce-e40c-450d-97bb-884b017f4cd7.

Basker, Emek. 2012. "Raising the Barcode Scanner: Technology and Productivity in the Retail Sector." *American Economic Journal: Applied Economics* 4(3): 1–27.

Bertrand, Marianne, and Francis Kramarz. 2002. "Does Entry Regulation Hinder Job Creation? Evidence from the French Retail Industry." *Quarterly Journal of Economics* 117(4): 1369–1413.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, Daniela Scur, and John Van Reenen. 2014. "EEA-FBBVA Lecture 2013: The New Empirical Economics of Management." *Journal of the European Economic Association* 12(4): 835–76.

Brynjolfsson, Erik, Yu (Jeffrey) Hu, and Michael D. Smith. 2010. "Research Commentary—Long Tails vs. Superstars: The Effect of Information Technology on Product Variety and Sales Concentration Patterns." *Information Systems Research* 21(4): 736–47.

Cardiff-Hicks, Brianna, Francine Lafontaine, and Kathryn Shaw. 2015. "Do Large Modern Retailers Pay Premium Wages?" *ILR Review* 68(3): 633–65.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda. 2014. "The Role of Entrepreneurship in US Job Creation and Economic Dynamism." *Journal of Economic Perspectives* 28(3): 3–24.

Doms, Mark, Ron S. Jarmin, and Shawn Klimek. 2004. "Information Technology Investment and Firm Performance in U.S. Retail Trade." *Economics of Innovation and New Technology* 13(7): 595–613.

Ellickson, Paul B. 2007. "Does Sutton Apply to Supermarkets?" *RAND Journal of Economics* 38(1): 43–59.

Ellickson, Paul B., Stephanie Houghton, and Christopher Timmins. 2013. "Estimating Network Economies in Retail Chains: A Revealed Preference Approach." *RAND Journal of Economics* 44(2): 169–93.

Foster, Lucia, John Haltiwanger, and C. J. Krizan. 2006. "Market Selection, Reallocation, and Restructuring in the US Retail Trade Sector in the 1990s." *Review of Economics and Statistics* 88(4): 748–58.

Goldmanis, Maris, Ali Hortaçsu, Chad Syverson, and Onsel Emre. 2010. "E-commerce and the

Market Structure of Retail Industries." *Economic Journal* 120(545): 651–82.

Holmes, Thomas J. 2001. "Bar Codes Lead to Frequent Deliveries and Superstores." *RAND Journal of Economics* 32(4): 708–25.

Holmes, Thomas J. 2011. "The Diffusion of Wal-Mart and the Economies of Density." *Econometrica* 79(1): 253–302.

Hortaçsu, Ali, F. Asís Martínez-Jerez, and Jason Douglas. 2009. "The Geography of Trade in Online Transactions: Evidence from eBay and MercadoLibre." *American Economic Journal: Microeconomics* 1(1): 53–74.

Institute for Competitiveness and Prosperity. 2010. "Management Matters in Retail." Working Paper 13.

Jarmin, Ronald S., Shawn Klimek, and Javier Miranda. 2005. "The Role of Retail Chains: National, Regional and Industry Results." CES Working Paper 05-30, Center for Economic Studies, US Census Bureau.

Jia, Panle. 2008. "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retail Industry." *Econometrica* 76(6): 1263–1316.

Lieber, Ethan, and Chad Syverson. 2012. "Online versus Offline Competition." In *Oxford Handbook of the Digital Economy*, edited by Martin Peitz and Joel Waldfogel, 189–223. Oxford University Press.

Mishel, Lawrence. 2012. "The Wedges Between Productivity and Median Compensation Growth." Economic Policy Institute Issue Brief 330.

Nishida, Mitsukuni. 2014. "Estimating a Model of Strategic Network Choice: The Convenience-Store Industry in Okinawa." *Marketing Science* 34(1): 20–38.

Scott Morton, Fiona, Florian Zettelmeyer, and Jorge Silva-Risso. 2001. "Internet Car Retailing." *Journal of Industrial Economics* 49(4): 501–519.

Troske, Kenneth R. 1999. "Evidence on the Employer Size-Wage Premium from Worker-Establishment Matched Data." *Review of Economics and Statistics* 81(1): 15–26.

Journal of Economic Perspectives—Volume 29, Number 4—Fall 2015—Pages 113–134

Adolescence and the Path to Maturity in Global Retail[†]

Bart J. Bronnenberg and Paul B. Ellickson

Distributing goods from producers to consumers constitutes a large fraction of overall economic activity. Using the United Nations National Accounts database, the distributive trades—retailing, wholesaling, and transportation—account for a constant 20–21 percent of global GDP going back at least as far as 1970, amounting to 1.3 times the GDP share of manufacturing in 2013. The share does not vary much across continents. The size of the distributive trades alone suggests that productivity growth in retailing could have a substantial impact on consumer welfare.

We argue that, over the past several decades, the adoption and diffusion of "modern retailing technology" represents a substantial advance in productivity, providing greater product variety, enhanced convenience, and lower prices. While the impact of this retail revolution is seen most clearly in the world's most developed countries, we conjecture that its impact may be even more profound for consumers in the developing world, where modern retailing is just starting to spread.

Distributive trades move goods from producer to consumer. Efficient production of transactions minimizes the overall cost of the distribution channel by allocating the costs of storage, handling, and transport to the most efficient provider, which can

■ *Bart Bronnenberg is Professor of Marketing and CentER Research Fellow, School of Economics and Management, Tilburg University, Tilburg, Netherlands and Research Fellow, Centre for Economic Policy Research, London, United Kingdom. Paul Ellickson is Professor of Marketing and Economics, Simon Business School, University of Rochester, Rochester, New York. Their email addresses are bart.bronnenberg@tilburguniversity.edu and paul.ellickson@simon.rochester.edu.*

[†] For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.29.4.113

doi=10.1257/jep.29.4.113

include the consumer. The benefits of retail innovation are notoriously difficult to quantify. Nonetheless, it seems clear that modern retail technologies provide a wider variety of products and services at lower cost, both in terms of price and the opportunity cost of time, yielding substantial increases in consumer welfare. In their analysis of Walmart's impact on the US grocery industry, Hausman and Leibtag (2007) conclude that the short-term benefit to consumers (mainly due to lower prices) of Walmart's entry is on the order of 25 percent of food expenditures. In this journal, Basker (2007) claimed that, by at least one measure, Walmart alone may have accounted for almost half of the 35.5 percent increase in the productivity of the US general merchandise sector between 1982 and 2002, echoing a McKinsey Global Institute (2001) report attributing the bulk of the acceleration in overall US productivity growth in the mid-1990s to innovations introduced by Walmart and subsequently adopted by its rivals. In their analysis of retail globalization in Mexico, Atkin, Faber, and Gonzalez-Navarro (2015) conclude that entry by foreign supermarkets led to welfare gains on the order of 6.2 percent of initial household income. In this article, we seek to explain the source of such gains and the likelihood that they will be replicated in other developing markets.

We first describe modern retailing, highlighting the role of modern formats, scale (often transcending national boundaries), and increased coordination with upstream and downstream partners in production and distribution. In developed markets, the transition to modern retailing is nearly complete. In contrast, many low-income and emerging markets continue to rely on traditional retail formats, that is, a collection of independent stores and open air markets supplied by small-scale wholesalers, although modern retail has begun to spread to these markets as well. E-commerce is a notable exception: the penetration of e-commerce in China and several developing nations in Asia has already surpassed that of high-income countries for some types of consumer goods. To understand the forces governing the adoption of modern technology and the unique role of e-commerce, we propose a framework that emphasizes the importance of scale and coordination in facilitating the transition from traditional to modern retailing. We conclude with some conjectures regarding the likely impact of increased retail modernization for the developing world.

Developments in Global Retail Since 2000

In this section, we use country-level retail volume data obtained from Euromonitor[1] to study the evolving market share of different store-based formats in food retailing between 2000 and 2014. We then look at the extent to which Internet-based retailing is replacing brick-and-mortar sales.

[1] The retail volume data used in this study were collected by Euromonitor and are available from their Passport database. The data used in this study were collected directly from retailers, store checks, surveys of the retail trade, desk research, and public data sources. The data were subject to validation and a consistency check of separately collected micro and macro data. The Passport database claims to be fully cross-country comparable. For more detailed notes, see http://www.euromonitor.com/research-methodology. We have data for 50 large economies spread over all continents.

Format Innovation in Store-Based Retailing

In the late 19th and early 20th century, food and nonfood retailing took place in individually owned stores with limited selection, high margins, and low turnover of the items on their shelves. The rise of chain stores, which occurred in the 1920s in the United States with a sharp increase in the number of stores run by companies like A&P, Walgreen's, and JC Penney, brought a combination of lower prices and higher volume, in part by sharply increasing store size and in that way taking over the economic function of traditional wholesalers (Tedlow 1990). Studies covering grocery prices from the 1920s and 1930s reveal that chain stores grocery prices were 4.5 to 15 percent lower than mainly single-store rivals. Modern mass-market retailing continues this evolution by integrating not only into wholesaling but also into production and distribution and by fostering increased differentiation through the specialization of formats. Big box retailers such as supermarkets, hypermarkets, club stores, and supercenters/mass-merchandisers combine scale and logistical advantages in distribution with the store-level convenience of one-stop shopping. Using data spanning 1998–2003, Hausman and Leibtag (2007) find that Walmart's prices were 15–25 percent lower than traditional supermarkets, suggesting that much of the benefit of modern retail technology is passed through to consumers. Although supermarkets, department stores, and general merchandisers initially targeted the whole population, recent growth in more specialized formats suggests an increasing role for segmentation and differentiation aimed at customers who place differing emphases on variety, price, or convenience, especially in the most-developed countries.

Table 1 examines some effects of this move toward modernization and differentiation in retail store types, presenting 2014 revenue shares and annual 2000–2014 share growth rates for each store format across a selection of high-income countries. We focus on grocery products because they are relatively well defined compared to other retail categories and are among the first to experience modernization. Of course, there is significant variation across high-income countries, and we will point out some specific examples.[2]

The top part of Table 1 shows revenue shares of traditional retail formats. Retailers in the "Independent/Other" category have 10 or fewer outlets. These traditional "mom and pop" markets are primarily family-owned, and include a diverse collection of kiosks, open-air markets, and souvenir stores selling food and drink items. The table shows that in 2014, traditional retailers in high-income nations have a joint revenue share of just 18 percent on average. Some nations—like Japan, the United Kingdom, and the United States—are approaching a complete transition out of traditional retail formats, whereas others still have a larger presence, like Italy and other nations in southern Europe. In these countries, revenue shares of traditional retail formats declined at a rate of 2.5–2.7 percent per year on average from 2000–2014. This downward trend is consistent with the pattern documented

[2] Shares and growth rates across all rich and developing nations in our database are provided in an Online Appendix available with this paper at http://e-jep.org.

Table 1

Share and Annual Growth by Store Type in the Grocery Trade in High-Income Nations

(share in 2014 in percent; annual percentage growth from 2000–2014 underneath in parentheses)

Store type	*All high-income nations*	*Germany*	*Italy*	*Japan*	*United Kingdom*	*United States*
Traditional formats						
Independent/other	9.4	6.2	11.3	9.6	5.4	6.6
	(−2.7)	(−3.9)	(−0.5)	(−2.7)	(−5.0)	(−2.4)
Food/drink/tobacco specialists	8.9	10.3	14.5	6.8	5.9	6.0
	(−2.5)	(−1.5)	(−2.7)	(−6.4)	(−4.7)	(−1.3)
Modern formats						
Supermarkets	33.5	28.1	34.6	37.4	20.9	29.6
	(−0.1)	(−0.5)	(1.0)	(0.8)	(−2.1)	(−1.4)
Hypermarkets/mass merchandisers	24.9	17.0	17.5	17.8	41.8	32.5
	(0.9)	(−0.9)	(0.5)	(−0.1)	(1.6)	(1.5)
Discounters	7.2	33.8	10.4		5.7	1.3
	(2.8)	(3.0)	(2.7)		(5.2)	(1.3)
Warehouse clubs	3.9			0.6		10.5
	(3.9)			(21.6)		(3.7)
Convenience stores	7.3		11.6	27.8	17.1	2.0
	(2.1)		(−0.5)	(3.6)	(4.0)	(0.0)
Forecourt retailers attached to gas stations	5.0	4.3			3.1	11.4
	(−0.3)	(−0.5)			(−0.7)	(−0.0)

Source: Data from Euromonitor 2015.

Notes: 2014 shares report on fraction of national grocery revenue. Whereas some revenue of mass merchandisers and warehouse clubs is nongrocery, these formats constitute a major source of grocery retailing and are included. In parenthesis, average annual percentage growth is computed over the years 2000–2014: if $s_{f,y}$ is the revenue share of a store format f in year y then define $g_{f,t} = s_{f,t}/s_{f,t-1}$. Next, define g_f as its geometric mean, computed over the years that sales are recorded. The reported percentage change is $100 \times (g_f - 1)$. No share or growth rates are reported if the 2014 regional share of a particular store type is less than one half of a percent of the market.

by Foster, Haltiwanger, and Krizan (2006), who examined the major restructuring of the US retail trade sector during the 1990s. They identify the key role of reallocation dynamics—replacing small, low-productivity firms with large, high-productivity entrants—in explaining the sharp increase in US retail productivity over this time period. In particular, they find that almost all of the robust growth in labor productivity over the 1990s (about 14 percent over this 10-year period) is due to new entrants displacing inefficient incumbents. These new entrants operate modern store formats almost exclusively. We discuss these modern store formats in three groups, roughly organized by whether they appeal to consumers mostly seeking variety, low prices, or convenience.

The first group consists of formats that offer substantial variety: supermarkets and hypermarkets/mass merchandisers (appearing in two rows in Table 1).[3] This format is the dominant method of selling grocery products to consumers in high-income regions, collectively accounting for an estimated 58 percent of the grocery trade in these markets. Given its mass appeal, this specialized format is the one most likely to be introduced first as retail markets develop. Between 2000 and 2014, hypermarkets grew at a steady pace, whereas the revenue share of supermarkets essentially remained constant. Hypermarkets (like Walmart) grew strongly while the pure supermarket format contracted somewhat in the United Kingdom and the United States. This pattern is consistent with the increased role of format specialization in the most advanced retail economies.

The second group consists of formats that offer low prices with limited variety, known as discounters and warehouse/club stores. This group includes firms like Costco in the United States and Lidl or Aldi in Europe. Discounters and club stores capture a modest 11 percent of the grocery trade in high-income nations, but their revenue share has been rising at an annual rate of 2.8 and 3.9 percent on average, the highest among any retail format. In the United States, the rise of club stores occurred alongside the rapid expansion of supercenters (mostly Walmart) and may represent an attempt to differentiate from this dominant firm. Costco, in particular, targets primarily high-income, suburban consumers who have ample space to store large "club packs" and less need for high-touch service. Aldi or Lidl are discounters that attract low-income consumers by focusing on unbranded goods and offering little in the way of service. The discounter format is very popular in Germany relative to the United States, suggesting that discounters and warehouse clubs (and to some extent hypermarkets/mass merchandizers) are alternative business models that cater to a similar need. In Germany, land use restrictions and a population more clustered in central cities favors the smaller footprint of the discount store, whereas the suburbanized United States is better served by more remote "big box" formats like hypermarkets and clubs.

The final group consists of retail formats that offer convenience in location or opening hours, including convenience stores and "forecourt" retailers attached to gas stations. These formats are not as new as hypermarkets, discounters, and club stores, but are considered "modern" in the sense that they are specialized and exploit the scale economies offered by a chain of stores to drive down costs. The trend toward larger, more remote supermarkets and hypermarkets (and the subsequent exit of mid-sized outlets) evidently yields pockets of underserved

[3] The distinction between supermarket and hypermarkets/mass merchandisers is not particularly sharp. Euromonitor International defines a hypermarket as a retail outlet with a selling space of at least 2,500 square meters (27,000 square feet) and a primary focus on groceries. In the United States, most supermarkets would easily fall into this category. Mass merchandisers like Walmart and Target combine grocery with general merchandise and, to accommodate these additional product categories, tend to be significantly larger than supermarkets. In Europe and other parts of the world, a hypermarket is closer to what would simply be considered a large supermarket in the United States. For this reason, it is easiest to view this collection as a single category.

local demand, especially amongst consumers with limited transportation options, creating new markets for these older formats. Convenience and forecourt formats account for 12 percent of the grocery trade on average in high-income nations and are especially common in countries with dense urban populations (like Japan and the United Kingdom). Overall, the convenience store format grew at an annual rate of 2 percent. Forecourt retailers in the United States account for a large share of retail sales relative to convenience stores. American cities are designed to accommodate car traffic, and chains like 7-Eleven, Circle-K, and am/pm often bundle food service with gas (landing them in the forecourt category).

Taken together, these trends reveal a grocery industry in the process of shifting from independent stores to a collection of formats specialized at serving different consumer needs but organized under the banner of a chain to exploit economies of scale. The migration of retailing, from traditional single-store proprietorships to modern chains of specialized formats, took place mainly across firms rather than within firms, consistent with the capital reallocation hypothesis proposed by Foster, Haltiwanger, and Krizan (2006). Clearly, some of these efficiency gains are due to the replacement of skilled full-time workers with less-skilled part-time workers, which could impact the overall wage structure. However, Basker (2007) notes that the existing empirical evidence reveals that Walmart's net impact on jobs in the US retail and wholesale sectors is small, and likely positive, though small negative effects have been found for retail wages. In their study of Mexico, Atkin, Faber, and Gonzalez-Navarro (2015) find that entry by foreign supermarkets has little impact on average municipality-level income or unemployment, while the impact on consumer welfare is positive and significant.

How do markets in the lower-income nations compare with these high-income nations in terms of format evolution? Table 2 shows 2014 shares and 2000–2014 annual share growth of store-based retail formats in UN-designated developing countries, first as an aggregate and next for a selection of nations chosen for their size and geographic dispersion: Argentina, India, Nigeria, South Korea, and Turkey. We discuss the so-called BRICS nations—Brazil, Russia, India, China, and South Africa—in more depth later. Relative to high-income countries, lower-income countries continue to support a large fraction of traditional retailers. In 2014, 57 percent of grocery revenue took place through the two traditional formats. However, this pattern has been rapidly changing. Starting at a combined share of 80 percent of grocery revenue in 2000, the shares of these traditional retail formats have been declining at 1.2 and 2.8 percent per year.

There are also striking patterns in the evolving composition of modern retail formats. The joint share of high-variety formats—supermarkets and hypermarkets/mass merchandisers—was 37 percent in 2014 and growing at approximately 6 percent a year. The growth rates are particularly strong in nations in Africa and the Middle East, where the supermarket revolution occurred most recently, while quite a bit lower in Latin America countries, which experienced the "first wave" of retail modernization in the 1990s (Reardon and Timmer 2012). The prevalence of the second modern retail format group, low-price/limited assortment formats, currently lags far behind

Table 2

Share and Annual Growth by Store Type in the Grocery Trade in Emerging Economies

(share in 2014 in percent; annual percentage growth from 2000–2014 underneath in parentheses)

Store type	*All developing nations*	*Argentina*	*India*	*Nigeria*	*South Korea*	*Turkey*
Traditional formats						
Independent/other	44.4	47.4	81.0	86.7	15.6	47.4
	(−2.8)	(0.7)	(0.1)	(−0.3)	(−2.4)	(−3.4)
Food/drink/tobacco specialists	12.2	12.7	17.3	6.7	13.9	15.6
	(−1.2)	(−0.7)	(−1.1)	(−0.9)	(−4.4)	(1.1)
Modern formats						
Supermarkets	24.9	22.6	0.8	2.9	14.7	20.1
	(5.4)	(−0.6)	(8.6)	(7.3)	(−3.8)	(6.5)
Hypermarkets/mass merchandisers	11.8	12.3	0.8	1.4	35.8	2.7
	(6.6)	(−1.1)	(29.3)	(86.0)	(3.4)	(13.2)
Discounters	2.6	2.3				11.6
	(4.0)	(8.6)				(19.9)
Warehouse clubs	0.5				5.8	
	(7.6)				(15.2)	
Convenience stores	2.7	0.8			14.2	1.7
	(9.5)	(32.2)			(10.9)	(33.6)
Forecourt retailers attached to gas stations	0.9	1.9		1.9		1.0
	(4.5)	(−3.3)		(4.6)		(5.8)

Source: Data from Euromonitor 2015.
Notes: See note to Table 1.

the high-income countries. Recall that this format group was the last to develop in most high-income countries as well. Discounters and warehouse/club stores only have a 3 percent share in emerging markets. However, their growth rates are especially strong in Latin America (here exemplified by Argentina, but growth is even stronger in Brazil, Chile, and Columbia), consistent with a transition to a second phase of modernization of retail and increased specialization. The third group, convenience stores and forecourt formats has 4 percent of the market in developing nations. Forecourt retailers are virtually absent outside of Latin America, presumably because low car ownership per capita in many developing countries limits the market for the food/gas pairing. The convenience store segment is much smaller than that in high-income regions but is growing more rapidly.

Hausman and Leibtag (2007) provide evidence that retail modernization leads to lower food prices. To investigate how or whether this translates into expenditures, we collected Euromonitor data on per capita expenditures on food and nonalcoholic beverages and expressed these as a percentage of total per capita

consumption expenditure. These measures are available for 50 countries across 15 years (2000–2014). We then relate these expenditure shares to the fraction of food retailing that takes place in the six modern formats in Table 1 and 2 as a measure of retail modernization. Pooling across 750 observations using a two-way (country and time) fixed effects regression, we find that a full transition to modern retail formats is associated with a drop of 8.2 percent of the food expenditure share (t-statistic = −6.2) from 26.3 to 18.1 percent of consumption expenditure.

The First 20 Years of Online Retailing

Along with the proliferation of modern store formats, the other transformative innovation in modern retailing has been the introduction of online sales platforms. By selling items online, firms can offer a wider selection of products at reduced cost, primarily by aggregating demand across a larger set of consumers and removing key links of the supply chain. The elimination of physical stores is the most obvious example of the latter, but the practice of "drop-shipping" directly from a manufacturer's warehouse also removes the need for a distribution center to hold that inventory that supplies retail stores, and reduces the cost and risk associated with offering products for which demand is locally thin (Lieber and Syverson 2012).

Online purchases have benefits and costs that vary by product category. For example, online purchase of physical goods introduces a delay between purchase and delivery, but also gives consumers a greater opportunity to comparison-shop by lowering search costs and travel time and provides a seamless method of gathering information on the experience of previous customers (through online reviews). On the other hand, online retail offers less ability to inspect goods before purchase (and adds the risk of not having a product delivered at all), which renders the reputation of the firm all the more important. Whether a purchase is made online or in-store clearly depends on the frequency of purchase, the homogeneity of the product, and the number of products typically purchased in a given occasion, amongst other factors. Books fall at one end of this spectrum, and thus, in modern retailing systems, are primarily bought online, while groceries fall at the other end, and are typically bought in-store. Here, we provide an overview of the relative penetration of online retailing across several types of goods in both high-income and developing markets. For a fuller treatment of the tradeoffs in online retailing, Lieber and Syverson (2012) provide a useful starting point.

Table 3 contains the share of total retail revenue from online transactions in high-income countries, along with its annual 2004–2014 growth. To focus the discussion, we present the two retail categories with the greatest online presence: apparel and footwear, and electronics and appliances. For comparison, we also report on a composite category consisting of all retailed goods, along with a separate breakout for groceries.

First, after 20 years of e-commerce growth and the ubiquitous presence of firms like Amazon and eBay, online sales in high-income nations still represent only a small fraction of overall retail sales, accounting for 6 percent of all retail sales on average. However, this average is somewhat misleading, because there are big

Table 3

Share of Online Purchases and Annual Percentage Growth Rates by Region and Industry for Selected High-Income Nations

(share in 2014 in percent; annual percentage growth from 2000–2014 underneath in parentheses)

Subcategory	*All high-income nations*	*Germany*	*Italy*	*Japan*	*United Kingdom*	*United States*
All	5.7	5.6	1.5	6.7	9.9	7.1
	(13.7)	(15.4)	(16.4)	(11.2)	(17.7)	(12.5)
Apparel and footwear	14.8	21.3	2.8	11.2	23.4	18.4
	(15.7)	(16.3)	(34.1)	(15.4)	(16.5)	(12.2)
Electronics and appliances	20.2	21.5	9.7	6.6	28.6	30.3
	(13.0)	(20.0)	(14.4)	(10.1)	(28.3)	(9.7)
Grocery items	1.0			1.8	3.3	0.6
	(11.6)			(9.2)	(12.3)	(13.5)

Source: Data from Euromonitor 2015.

Notes: The subcategory "All" is the sum of six subcategories: apparel and footwear; electronics and appliances; grocery; health and beauty; home and garden; and personal goods. Personal goods are a miscellany consisting of media products, personal accessories, eye-wear, games and toys, sports goods, and pet care. The shares reported are the percentage of regional sales purchased online. In parenthesis, average annual percentage growth is computed over the years 2004–2014: If $s_{c,y}$ is the online share of a subcategory f in year y then define $g_{c,t} = s_{c,t}/s_{c,t-1}$. Next, define g_c as its geometric mean, computed over the years that online sales are recorded. The reported percentage growth is $100 \times (g_c - 1)$. No share or growth rates are reported if the 2014 regional share of a particular transaction type is less than one half of a percent of the market.

differences across countries. For instance, Italy has an online retail market share of less than 2 percent, compared to 10 percent in the United Kingdom.

Across consumer goods industries, e-commerce is heavily concentrated in categories such as apparel and footwear, and electronics and appliances, where online sales account for a substantial 15 and 20 percent of retail volume in high-income countries, reaching 18 and 30 percent in the United States.[4] The share of online sales across all high-income nations has witnessed double-digit growth in the period from 2004 to 2014.

In contrast, online sales remain an insignificant fraction of grocery retailing, with an estimated 1 percent of sales transacted online. This pattern is perhaps not surprising, given that groceries are purchased more frequently than any other retail category, in part because consumption often closely follows purchase. Moreover, physical search (for example, for fresh produce) remains a key aspect of

[4] Lieber and Syverson (2012) report broadly consistent fractions for online purchases in the United States circa 2007 (that is, similar but lower than our data, which are 2014), based on data from a Forrester survey. They also provide detailed breakouts by product category, finding the highest fractions of online purchase in books, apparel, footwear, and consumer electronics.

Table 4

Share of Online Purchases and Annual Percentage Growth Rates by Region and Industry for Selected Emerging Nations

(share in 2014 in percent; annual percentage growth from 2000–2014 in parentheses underneath)

Subcategory	*All developing nations*	*Argentina*	*India*	*Nigeria*	*South Korea*	*Turkey*
All	4.2 (25.5)	1.5 (25.9)	0.8 (50.1)	0.6 (57.5)	11.6 (9.3)	1.7 (27.6)
Apparel and footwear	13.5 (32.9)	0.6 (30.6)	3.2 (51.1)	1.5 (64.0)	23.5 (14.9)	2.9 (53.4)
Electronics and appliances	15.0 (22.1)	14.9 (21.0)	2.5 (30.5)	1.0 (67.3)	9.8 (0.8)	6.9 (16.1)
Grocery items	0.8 (26.9)				6.3 (15.8)	

Source: Data from Euromonitor 2015.
Note: See note for Table 3.

grocery shopping, since many perishable products are not of uniform quality and the ingredients for a particular meal might depend on what is currently on hand. Furthermore, consumers are often shopping for an entire basket of goods rather than a single product or handful of items, and this form of shopping is likely to be more challenging online. From the retailer's perspective, impulse purchases likely play a lesser role online, eliminating a key source of revenue.

How do developing markets compare to high-income ones in terms of online penetration? Table 4 examines the diffusion of online retailing in the developing world. The overall average share of online retailing across all categories is, as in high-income nations, quite low at 4 percent.

Interestingly, we now see even larger differences across nations. Across developing nations, migration to online sales in the low-income regions has primarily taken place in Asia, represented in Table 4 by South Korea (we discuss China, for which this pattern is also true, momentarily). Indeed, whereas African and South American nations generally remain far behind high-income nations in terms of share of online sales, several Asian countries have closed in on 2014 levels of US online retailing. The difference between Asia and South America is not explained by Internet penetration, which is lower in the former (35 percent) than the latter (52 percent), as reported at http://www.internetworldstats.com (accessed on 7/1/2015).

Summary

Comparing a typical supermarket in the 1970s to the retail outlets of 2014, we first see a number of market changes. First, there have been vast improvements in both the supply and availability of fresh products and the diversity and quality of

products on offer: for example, consider the quality and variety of bread or coffee, or the much greater year-round availability of fresh produce in a modern store. Moreover, investments in information technology have led to increases in labor productivity that are at least partly passed on to consumers through lower prices (Basker 2015). Second, the dominant retailers have diversified into a more varied set of store formats, each specialized to better deliver variety (hypermarkets), quantity (discounters), or time savings (convenience stores, online stores) to a more diverse and time-constrained consumer base. While such specialization may lead to a decrease in the level of service experienced by some consumers, it seems likely that the overall shopping experience has improved markedly over this period for almost all consumers. Indeed, the limited empirical evidence currently available (for example, Hausman and Liebtag 2007), suggests that the associated increase in consumer surplus is sizable. Third, the retail sector continues to make it easier for consumers to find and buy goods both online and off.

Although constructing a clear productivity measure to quantify these utility gains is challenging, the retail sector of 2014 surely provides far more utility—via superior match value and greater variety—than its 1970s counterpart. Moreover, the challenges inherent in constructing productivity measures with these kinds of qualitative changes suggest that estimated productivity gains may be vastly understated.

Next we summarize the necessary and often dramatic transitions that must take place if a nation's retail sector is to move from traditional to modern, with all the associated consumer benefits in terms of lower price, increased variety, and greater convenience.

The Structure of Modern Retailing

Modern retailing technology includes a constellation of activities encompassing format modernization and specialization, scale economies moderated by spatial competition, supply chain integration, information technology investment, and forward integration into direct delivery. In this section, we describe these activities and the investments required to undertake them. This discussion sets out the basis of what it will take for countries, especially those in emerging markets, to develop modern retail practices.

Format Modernization and Differentiation

Modern retail-store formats have steadily replaced traditional "mom and pop" outlets in both high- and low-income markets. Two distinct aspects of format evolution are especially important for understanding modern retailing. First, large-footprint stores can offer a much wider variety of products than traditional outlets, and can do so while exploiting scale economies at the level of both outlet and chain (see below), thus offering a price level that sole proprietorships do not have the cost structure to match. This description applies not just to supermarkets carrying groceries, but also to mass merchandisers like Walmart offering general

merchandise and "category killers" like Home Depot, Best Buy, and Staples, which offer a deep selection within a narrow set of products. Their ability to do so reflects innovations in information technology and logistics that have increased labor productivity and reduced inventory costs.

Second, retail outlets have shown increasing specialization in targeting their customers, particularly in the grocery category. In high-income countries, retailers target segments of the population that differ in either their willingness or ability to pay for variety and services. For example, high-quality service-oriented firms like Whole Foods attract wealthy, time-constrained urban and suburban professionals with organic products and prepared meals. At the other end of the spectrum, limited assortment chains like Save A Lot and Aldi target the urban poor with small footprint stores offering unbranded products at very low prices. Diverse formats also allow firms to adjust to local market conditions. For example, European hypermarkets combine food and nonfood products in the same store, which allows firms to accommodate government restrictions on the number of large footprint stores. Similarly, German limited assortment stores are partly a response to tight zoning restrictions. More broadly, as populations become more diverse by preference or income, the set of formats in this "retail ecosystem" expands accordingly.

The same patterns are in play in emerging markets. While much of the growth in modern retailing in these countries is amongst the high-variety formats (as shown earlier in Table 2), the remaining modern formats develop as well. For example, Walmart operates multiple formats in Mexico. In addition to its classic superstore format, it carries a mini-grocer format called "Bodega Aurrera Express," that is aimed at consumers with incomes too low to shop in Walmart's flagship stores. Moreover, real estate development in many Mexican cities is too dense to accommodate a big box outlet. In France, Carrefour operates hypermarkets, supermarkets (Carrefour Marché), convenience stores (Carrefour Express), and multichannel retail (Carrefour Drive) where consumers can pick up baskets of products bought online if they choose not to have them home-delivered.

Economies of Scale and Supply Chain Integration

Economies of scale contribute to modern retail technology at two levels: the outlet and the chain. The importance of outlet-level scale is evident in the physical footprint of the stores themselves, which have been increasing in size across many retail sectors for many decades. This increase in store size can be partly explained by what Oi (1992) refers to as the "economies of massed reserves." Oi viewed the primary goal of retailing as minimizing the frictions resulting from transport and inventory costs. Thus, he noted that economies of scale arise naturally from the connection between the arrival rate of consumers and the flow capacity of the outlets built to serve them. In particular, as consumers' storage and transportation capabilities improve—say, as a result of increased automobile ownership and relocation to large suburban homes—average transactions sizes also increase. The result is efficiency gains for the larger footprint stores, which have faster inventory turns and therefore lower storage costs.

Economies of store size lead to large firms because logistics systems and information technology allow stores to provide an ever-expanding array of products. A combination of information technology and the use of barcode-based restocking algorithms favor higher-frequency delivery schedules, which in turn contributes to larger optimal store sizes (Holmes 2001). Distribution costs are often spread across many stores, but typically rise with distance. In the US retail sector, the increasing importance of information technology and logistical innovations coincided with a sharp increase in the number of products (Messinger and Narasimhan 1995). In turn, this change shifted the focus of competition from price to variety, provided an additional justification for increased selling space, and created a link between outlet scale and the overall size of the chain (Ellickson 2007).

One key linkage between scale at the store and chain level is the regional distribution center, which is tasked with replenishing the stores in its catchment area with the bulk of the products carried by the chain (Ellickson 2007). Almost all modern supermarket, club, mass merchandise, and big box retail chains are vertically integrated into distribution, operating their own distribution centers. This suggests it is more difficult to coordinate delivery and replenishment schedules through a third-party logistic system. Moreover, many of the data interchange protocols employed by these firms are proprietary and take place over closed networks.

The size of the catchment area determines the geographic scope of the market and, consequently, the efficient scale of the chain. The catchment area, in turn, is determined by how far products can be cost-effectively transported. For grocery products, the maximum distance is on the order of a few hundred miles. For general merchandise and durables, the distance is much farther. As a consequence, the US retail sector is served by a large number of regional supermarket firms (only a handful of which are national), while European supermarkets typically cross national boundaries. In contrast, national or international chains dominate the mass merchandise market on both sides of the Atlantic, largely because their products are less costly to ship.

In addition to distribution-related economies, large chains can also exploit quantity discounts from manufacturers and economies of scale in advertising. Unlike economies related to distribution, these other efficiencies can scale up indefinitely with the size of the chain (or market), ultimately providing an economic rationale for global retailing (even for groceries). In addition, the largest firms may also choose to integrate further upstream into production, by offering store-branded "private-label" products that compete directly with the national brands. While employing this strategy requires a consumer base that is loyal to the retailer, it gives that retailer an additional source of bargaining power with manufacturers, and constitutes another manifestation of scale. Private-label programs have been particularly effective in Europe, where consumers are extremely loyal to a single chain and product advertising has been limited historically (Dobson 2006).

Economies of scale are apparently also significant at the global level. In 2012, the 100 largest retailers in the world by revenue were active in 11.8 countries on average (based on data from Deloitte 2014). For instance, in 1970, Carrefour

confined its business area to its native France and neighboring Belgium. It then spread to other European nations and currently operates 10,000 stores in 34 countries on four continents. Walmart didn't expand outside the United States until 1991 but now operates in 27 countries on five continents.

Still, these firms are exceptional—for most multinational retailers, international expansion is more geographically focused. Of the top 100 retailers, 50 are in five or fewer markets (including their country of origin). For food retailers such as Portugal's Jerónimo Martins entering Poland, or Carrefour entering Brazil, trade areas are not necessarily adjacent or even proximate. This lack of a clear geographic pattern seems consistent with the existence of multiple equilibria at the global level (for example, such an outcome can arise when there is some limit to the benefits of scale, implying that many potential market structures can be supported).

At the same time, increasing scale makes modern retail chain stores also face significant coordination costs. For example, grocery supermarkets must procure a variety of (often fresh) goods at a scale that is commensurate with their expanding reach, which requires coordination with several independent suppliers and manufacturers. For retail firms, coordination involves investment in information technology systems, storage facilities, and packaging technologies. Some of these investments are relationship-specific with other firms, creating the potential for hold-up. Coordination can also involve large financial commitments in inventories, which are sometimes financed with short-term debt. For a retailer, the need for coordination implies supply chain risk, such as disruptions in quantity and quality, which are costly in this low-margin business. As we will argue later, these are critical concerns in developing markets.

In addition to backward coordination, there is also a need for coordination and integration with forward links in the supply chain. Oi (1992) notes that, because the consumer's time is an input to the retail production function, it is natural for innovative retailers to explore methods of affecting the consumer's cost of transaction or home production if they can profitably do so (see also Bronnenberg 2015). This process results in forward integration of at least two forms. First, modern retailers complement their business with delivery services that are either owned (for example, in the case of the grocery chain Peapod) or contracted (for example with UPS or Fedex). Second, retailers such as Whole Foods offer ready-to-eat products, cutting food preparation time for time-constrained consumers, while bouquets of flowers can now be purchased at local supermarkets, eliminating a separate trip to a florist. Meal kit services such as startups Hello Fresh and Blue Apron, each shipping more than a million fresh meal kits per month as of 2014, are another manifestation of a continuing trend of retailers forward-integrating their services into the consumer's kitchen.

The Expansion of Online Retailing

Because the vast majority of retail sales still occur in physical stores, we have primarily emphasized the aspects of retail modernization that involve reallocating economic activity from small, low-productivity, single-store outlets to large,

high-productivity chains. However, many retail categories are experiencing a shift to online sales, which reduces the need for physical outlets but continues to require sophisticated upstream integration. In a sense, online retailing is a natural extension of the move toward forward integration mentioned a moment ago. E-commerce firms, which replace the physical store with an online marketplace, eliminate the consumer's need to travel to the store, shifting the role (but not the cost) of distribution entirely onto the firm. Distributing directly to the consumer relies on the same information and logistic technology that facilitates store-based retailing, suggesting that, at least on the supply side, these channels might be complements, rather than substitutes.

Online sales are growing quickly and already have a strong position in several retail categories, as illustrated in Tables 3 and 4. Goldmanis, Hortaçsu, Chad, and Emre (2010) argue that the primary competitive impact of online retailing falls on the small, high-cost retailers (that is, the "mom and pop" stores) rather than large, low-cost chains. In fact, low-cost "big box" retailers can in some cases become more profitable as e-commerce expands. At least in some retail categories, store-based retailing appears to hold its own against online. In others, it may play a complementary role. Pozzi (2013) analyzes the introduction of an online shopping service by a large bricks-and-mortar US supermarket chain and finds that the new online channel led to a 13 percent increase in revenue, with little cannibalization of existing sales. He attributes the increase in sales from online expansion to a mix of market expansion (due to a reduction in travel costs) and taking business from competing chains. Similarly, Einav, Levin, Popov, and Sundaresan (2014) find that an expansion of mobile commerce (using mobile devices) tends to increase the total sales for firms. While for some product categories, online sales seem destined to dominate other sales channels, many other categories are likely to continue to be sold through a mix of both online and offline channels.

Explaining Retail Modernization

A Coordinated Exploitation of Scale

What determines the timing of adoption and speed of diffusion of modern retail technology? To structure the argument, we frame the phenomenon as the coordinated exploitation of scale economies. Murphy, Shleifer, and Vishny (1989) suggest a similar mechanism to account for industrialization of manufacturing. In particular, we define retail industrialization as the application of scale economies and supply chain integration to the production of retail sales. We focus on modernization by country, which parallels the emphasis of Murphy, Shleifer, and Vishny on the importance of relatively distinct domestic markets in constraining the push towards industrialization. Reardon and Timmer (2012) note that "at present only 10 percent of global processed food output is traded across national borders." The perishable or fragile nature of many products, including food products, places even tighter constraints on the geography over which scale economies can be spread.

In the Murphy, Shleifer, and Vishny (1989) framework, the failure to industrialize is caused by the inability to coordinate on a Pareto superior equilibrium, which arises out of an externality problem stemming from the need to modernize several sectors at once. In retail, a coordination problem of that type occurs along the vertical chain from production to wholesaling to retailing to consumption. To take full advantage of scale, all elements of the chain must act in concert. With this framing in mind, we restate the retail modernization question in two parts: 1) What slows or speeds the leveraging of scale economies in retailing? 2) Why do certain countries remain stuck in a traditional equilibrium with outdated technology?

Three key players determine the speed of retail industrialization: consumers, governments, and firms. We start with consumers. A major impediment to retail modernization in many countries around the world is low consumer income: that is, consumers must achieve a minimal level of disposable income to facilitate a substantial shift from home to market production. Reardon and Timmer (2012) note that the "share of packaged food in food expenditures is 7 percent in low-income countries, 30 percent in lower-middle-income countries, and 45 percent in upper-middle-income countries." To understand the role of consumer demand in driving the adoption of modern retail technology, it's important to recognize that (unlike with manufacturing), the consumer's time is a key input to the retail production function (Oi 1992). As the opportunity cost of time becomes more valuable, technologies that minimize it (for example, by allowing fewer trips with larger basket sizes) become more dominant. Urbanization tends to reduce the time input of shopping. Higher female labor force participation tends to increase the opportunity cost of time. The big box retail format also requires a co-investment on the part of consumers in both transportation and storage technology—that is, cars and refrigerators. Lagakos (forthcoming) demonstrates the importance of car ownership in driving the adoption of modern grocery formats in six developing countries. He argues that the continued reliance on traditional formats is an optimal supply response when ownership of complementary durables is sufficiently limited.

The government provides a set of inputs to the retail production function via shared infrastructure. Physical retailing is inherently local, and the scale economies that drive it are constrained by the logistics of physical transportation. The local nature of retailing makes this a country-specific investment: distributing goods in Argentina requires good roads in Argentina, and road building is usually a task left to the government. Among the earliest emerging-market countries to modernize their retail industries were the former Soviet-controlled countries of Eastern Europe, exploiting the transportation infrastructure built during the cold war.

Governments also play a pivotal role in setting regulations on the use of land, the ease of obtaining building permits, the regulation of corruption, the availability of autos (through policies allowing the imports of used cars), and the minimum wage structure. In discussing a report from the McKinsey Global Institute on productivity, Baily and Solow (2001) noted in this journal that "the teams working on the case studies also concluded that a proliferation of regulations—land use, business hours, permits—created a much more frequent problem for productivity

in domestic service industries and construction than in manufacturing." They give the example of Korea, whose reliance on outdated retail formats throughout the 1990s was attributed to a mix of national investment policy, land use restrictions, and entry regulations.

In many emerging markets, another way in which government affects the retail sector lies in its ability to set policies regarding foreign direct investment. In an empirical study aimed at explaining the share of modern retail formats in 42 countries, Traill (2006) identifies GDP per capita and openness to foreign investment as the two most important explanatory factors. In their study of Walmart in Mexico, Iacovone, Javoric, Keller, and Tybout (2015) note that the impact of foreign direct investment in retail goes beyond simply the retailers themselves, as it "can induce structural changes in the domestic manufacturing industries that supply retailers with consumer goods." The increased level of competition can yield large efficiency gains. Studying the case of Romania, Javorcik and Li (2013) find that a 10 percent increase in the number of retail outlets run by foreign chains is associated with a 2.4 to 2.6 percent increase in the total factor productivity of the supplying industries. In their analysis of Chinese exports, Head, Jing, and Swenson (2014) find that increased exposure to multinational retailers led to rising exports from the regions those retailers entered, suggesting that the increased familiarity with the quality control and product requirements necessary to export goods is transferable across firms.

Firms are clearly the foremost strategic players driving the adoption of modern retailing technology. A modern chain of vertically integrated, large-format stores relies on an upstream distribution system of local producers, third-party logistics firms, and either third-party or integrated wholesalers who must all modernize together. Transactions that were often historically informal must be formalized through contracts with local suppliers and intermediaries. In a case study of Chile, Berdegué (2001) found that small farming cooperatives had to incur significant costs to deliver products of homogeneous quality, to coordinate harvest cycles, and to grade, sort, and package in a manner that met the downstream chain's requirements. Also, adopting formal accounting processes makes previously informal transactions subject to taxes.

Among the toughest coordination problems is the joint adoption of commonly used technology. Iacovone, Javoric, Keller, and Tybout (2015) note that to work with Walmart, "suppliers often need to make complementary investments in office technologies and computerized tracking systems." They must also invest in modern warehousing facilities, cold storage capability, and standardized packaging equipment. All of these investments require sufficient scale to cover their costs, which means that upstream and downstream firms need to modernize together. The problem of broad-based coordination (and the existence of many levels of externalities) suggests the potential importance of a centrally positioned decisionmaker to coordinate these decisions.

Particularly in the smaller countries, which do not have the population size to warrant local provision of large-scale retail distribution, it may often be most

efficient for an outside firm to play the central role in coordinating the vertical chain, often by relying on existing relationships with regional suppliers and global logistics firms. In this situation, rules that limit foreign direct investment may be the main constraint on developing a modern retail sector in such countries. For instance, Reardon, Henson, and Berdegué (2007) note that "a fear and complaint of a retailer entering a developing country is often the lack of a developed logistics sector." Rather than waiting for a logistics sector to develop within the country, international retailers frequently rely on third party logistics firms via a method known as "follow-sourcing," in which large international firms convince partner firms to co-locate in new markets. In addition, large retailers often need to upgrade the capacity of farmers and suppliers to meet their requirements. In the mid 2000s, Walmart worked to help establish micro-finance for farmers in Central America and India, and Carrefour was instrumental in building third-party supply chains in Indonesia and Thailand.

Emerging economies with larger domestic markets, like India or China, are clearly big enough to support the organic growth of domestic firms. However, they still face the issues of scale and the ability of third-party support networks to keep up. Firms that initially rely on local sources eventually turn to regional and later global sourcing. Firms that start out by relying on third-party wholesalers eventually integrate into more and more aspects of distribution and possibly production.

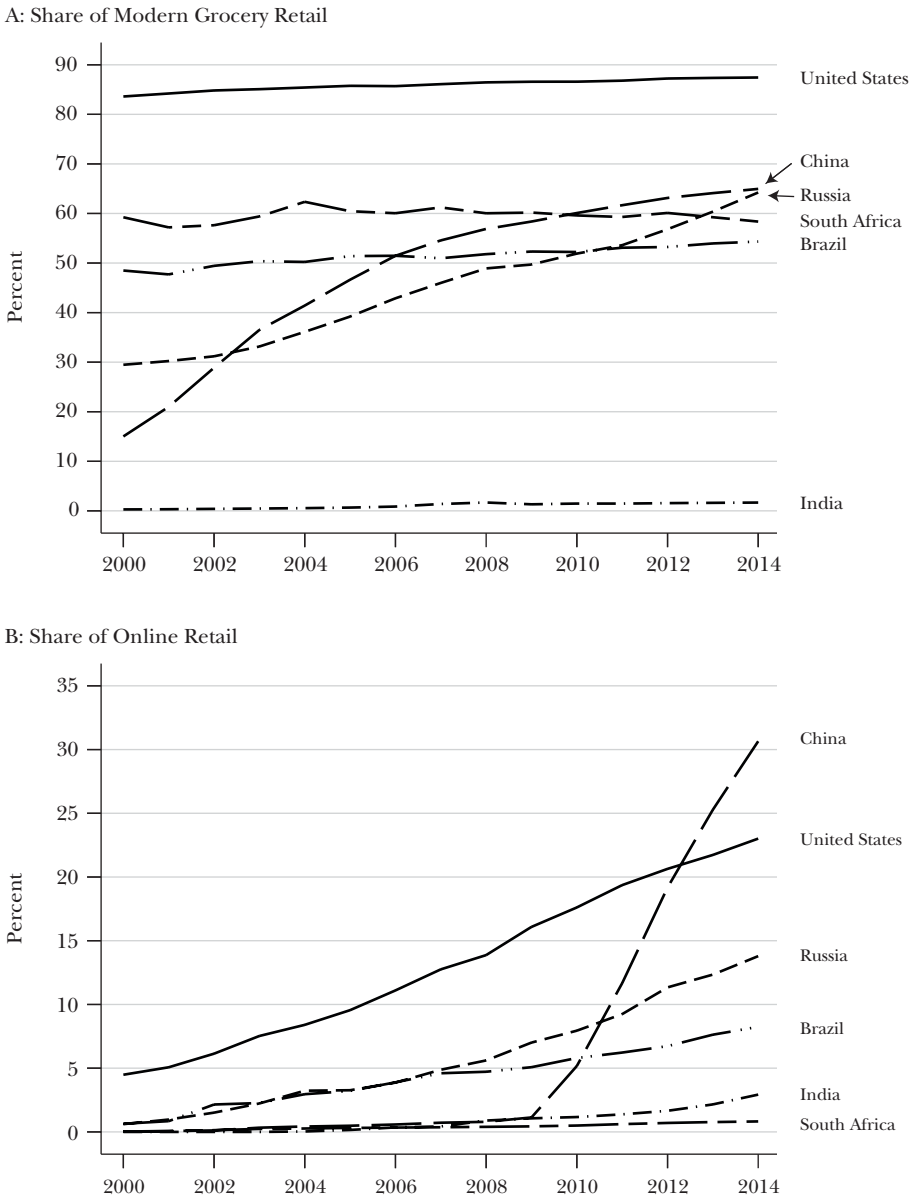
Apart from food retailing, minimum efficient scale in other areas of retail is less constrained by space. Because nonperishable goods can be shipped much greater distances, the efficient chain size of firms specializing in these products can be much larger. The coordination problem across a distribution chain may be substantially mitigated in product categories amenable to e-commerce distribution, which requires only an Internet-connected consumer base and a sufficiently developed third-party logistics system. In particular, in developing countries that do not have the infrastructure or local supply chains to support efficient store-based retailing for goods that are purchased infrequently (consumer electronics, clothing), e-commerce may become the dominant mode of retail commerce.

Modernization of Retail in the BRICS Nations

Depending on the success of all required channel partners modernizing in concert, the process of moving from an uncoordinated equilibrium (of traditional retailers) to a modern, integrated system may be quick or might not begin at all. To illustrate, we provide some anecdotal evidence of adoption of modern retailing systems in Brazil, Russia, India, China, and South Africa (BRICS), using the United States as a point of reference.

The top panel of Figure 1 shows the 2000–2014 joint share of all "modern formats" included in Tables 1 and 2. The United States has all but completely converted to modern grocery retailing. Brazil and South Africa had mostly modern systems, or were almost to that point, even before the period began and experienced

Figure 1
Modern Retailing Systems in BRICS versus United States, 2000–2014



Source: Data is from Euromonitor 2015.
Notes: Figure 1A shows the grocery revenue share of all modern retailing formats (convenience stores, discounters, forecourt retailers, hypermarkets, mass-merchandisers, supermarkets, and warehouse clubs) as a fraction of the total retail revenue, which also includes traditional retail formats (food/drink/tobacco specialists, independent stores, and other stores). Figure 1B shows the share of online retailing as a fraction of total retail revenue for selected consumer durables consisting of "Apparel/Footware" plus "Electronics/Appliances."

little additional change over the period. India, on the other hand, shows very low penetration of any form of modern retailing and still operates in a traditional equilibrium. In contrast, Russia and China have jumped from a complete traditional equilibrium like in India to the widespread adoption of new retail formats covering 65 percent of grocery retailing since the year 2000.

The lower panel of Figure 1 portrays the development of e-commerce as measured by online sales of two categories of durable goods often sold online: apparel and footwear plus electronics and appliances. For reference, the share of online sales in the United States in these two categories grew from 5 to 23 percent of total category value between 2000 and 2014. Over this time span, online sales in India and South Africa lag far behind, never reaching more than 3 percent of total category value. In contrast, Russia and Brazil show strong growth in e-commerce sales. At historic growth rates, Russia will reach the 2014 US level in just five or six years. However, the big story here is the spectacular development of e-commerce in China. In just five years, Chinese online retailing leapfrogged that of the United States, going from essentially nonexistent online in 2009 to 31 percent of total retail value in 2014 (representing more than $320 billion in revenue by Euromonitor's data) in 2014. Indeed, we speculate that with this type of growth in online retail, big box retailing in these categories in China may be skipped in favor of a dominant e-commerce retail channel.

What explains the rapid modernization of retail in China and its near absence in India? In keeping with the framework laid out above, we propose that a coordinated set of events allowed Chinese retail entrepreneurs to achieve scale and vertical coordination at a surprisingly rapid rate. At the *consumer* level, GDP per capita in China grew to US$13,130 by 2014 (according to the World Development Indicators database). Increases in income, combined with a high savings rate, meant that between 2000 and 2014 many Chinese consumers passed the wealth threshold to purchase relatively expensive durable goods. Urbanization in China doubled from 26 percent of the population in 1990 to 54 percent in 2014, according to data from the UN World Urbanization Prospects. For comparison, US Census Bureau data shows that the same transition took 55 years in the United States from 1870 until 1925. As a result, China's new middle class was living in metropolitan areas with modern roads and widespread Internet access, presenting the perfect conditions for e-commerce to blossom. At the *government* level, Chinese economic policy has recently promoted household consumption and provision of domestic services, relative to export-based growth. This pro-consumption policy encourages consumers to purchase big-ticket items such as cars, electronics, and appliances. In turn, car and appliance ownership drives demand for storable food products and larger durables. The Chinese government also initiated large-scale infrastructure projects that facilitate more efficient mass distribution by sellers. China's expressway system, which connects cities and supports intercity supply chains, tripled in length between 2004 and 2013, according to the National Bureau of Statistics of China. At the *firm* level, large sellers, both Chinese and foreign, have entered food and consumer durable markets sourcing their products via a

relatively reliable supply chain. These large firms are investing in shared resources. Launched in 2013, Alibaba and eight other companies invested the equivalent of billions of US dollars in the China Smart Logistic Network (CSN), consisting of fulfillment centers, distribution centers, and various forms of transportation, with the goal of delivering parcels to the entire country in 24 hours.

In contrast, consumers in India have lower incomes on average—about half the GDP per capita of their Chinese counterparts (in the World Development Indicators database). Moreover, India's policies concerning foreign direct investment have been among the world's most restrictive and were only recently liberalized in 2012. Local regulations across India continue to severely impede investment. Road construction and infrastructure development in India—for example, as measured by expressway miles—greatly lags China's. Firms in India find their growth bogged down by these restrictions. As one example, a recent article in the *Wall Street Journal* describes how Indian e-commerce giant Flipkart struggles to deliver products in the traffic-jammed streets of most Indian cities and has been forced to develop a different delivery technology that relies heavily on scooters and motorbikes—thus sharply limiting the set of products that can be home delivered (McLain 2015).

Conclusion

Modern retailing technology depends on the close coordination of many interdependent agents in the supply chain. Moving from a traditional retail system consisting of small-scale wholesalers and retailers to a modern one with large-scale integrated firms requires a coordinated set of complementary investments by consumers, government, and firms.

We argue that retail modernization offers important potential for welfare growth. In the absence of modern retailing, consumers face implicit tariffs in the form of inefficient transportation, distribution, wholesaling, and retail. Consumers pay these tariffs in terms of high prices, less variety, and less leisure (inefficient home production). Evidence suggests that the magnitude of these effects is large. Indeed, if one uses differences in market shares of modern retail as a proxy for the productivity gap, the static productivity difference between high-income and emerging nations is responsible for a large fraction of the differences in GDP. This proxy almost certainly understates the full productivity gains, as the dynamic gains arising from producers operating in more competitive and flexible markets are also important.

■ *We thank Emek Basker, Marnik Dekimpe, Bryan Ellickson, Katrijn Gielens, Els Gijsbrechts, Gordon Hanson, Enrico Moretti, and Timothy Taylor for helpful comments and suggestions. Bart Bronnenberg gratefully acknowledges financial support from the Netherlands Organization for Scientific Research (NWO).*

References

Atkin, David, Benjamin Faber, and Marco Gonzalez-Navarro. 2015. "Retail Globalization and Household Welfare: Evidence from Mexico." NBER Working Paper 21176.

Baily, Martin Neil, and Robert M. Solow. 2001. "International Productivity Comparisons Built from the Firm Level." *Journal of Economic Perspectives* 15(3): 151–72.

Basker, Emek. 2007. "The Causes and Consequences of Wal-Mart's Growth." *Journal of Economic Perspectives* 21(3): 177–98.

Basker, Emek. 2015. "Change at the Checkout: Tracing the Impact of a Process Innovation." *Journal of Industrial Economics* 63(2): 339–370.

Berdegué, Julio A. 2001. *Cooperating to Compete: Associative Peasant Business Firms in Chile.* Wageningen, NLD: Thesis Wageningen University.

Bronnenberg, Bart J. 2015. "The Provision of Convenience and Variety by the Market." *RAND Journal of Economics* 46(3): 480–98.

Deloitte. 2014. "Global Powers of Retailing 2014." Research Report. http://www2.deloitte.com/content/dam/Deloitte/global/Documents/Consumer-Business/dttl-CB-GPR14STORES.pdf.

Dobson, Paul W. 2006. "Exploiting Buyer Power: Lessons from the British Grocery Trade." *Antitrust Law Journal* 72(2): 529–62.

Einav, Liran, Jonathan Levin, Igor Popov, and Neel Sundaresan. 2014. "Growth, Adoption, and Use of Mobile E-Commerce." *American Economics Review* 104(5): 489–94.

Ellickson, Paul B. 2007. "Does Sutton Apply to Supermarkets?" *Rand Journal of Economics* 38(1): 43–59.

Foster, Lucia, John Haltiwanger, and C. J. Krizan. 2006. "Market Selection, Reallocation, and Restructuring in the U.S. Retail Trade Sector in the 1990s." *Review of Economics and Statistics* 88(4): 748–58.

Goldmanis, Maris, Ali Hortaçsu, Chad Syverson, and Önsel Emre. 2010. "E-Commerce and the Market Structure of Retail Industries." *Economic Journal* 120(545): 651–82.

Hausman, Jerry, and Ephraim Leibtag. 2007. "Consumer Benefits from Increased Competition in Shopping Outlets: Measuring the Effect of Wal-Mart." *Journal of Applied Econometrics* 22(7): 1157–77.

Head, Keith, Ran Jing, and Deborah L. Swenson. 2014. "From Beijing to Bentonville: Do Multinational Retailers Link Markets?" *Journal of Development Economics* 110(C): 79–92.

Holmes, Thomas J. 2001. "Bar Codes Lead to Frequent Deliveries and Superstores." *RAND Journal of Economics* 32(4): 708–725.

Iacovone, Leonardo, Beata Smarzynska Javoric, Wolfgang Keller, and James R. Tybout. 2015. "Supplier Responses to Wal-Mart's Invasion of Mexico." NBER Working Paper 17204.

Javorcik, Beata S., and Yue Li. 2013. "Do the Biggest Aisles Serve a Brighter Future? Global Retail Chains and Their Implications for Romania." *Journal of International Economics* 90(2): 348–63.

Lagakos, David. Forthcoming. "Explaining Cross-Country Productivity Differences in Retail Trade." *Journal of Political Economy.*

Lieber, Ethan, and Chad Syverson. 2012. "Online versus Offline Competition." In *Oxford Handbook of the Digital Economy*, edited by Martin Peitz and Joel Waldfogel. Oxford University Press.

McKinsey Global Institute. 2001. *U.S. Productivity Growth 1995–2000: Understanding the Contribution of Information Technology Relative to Other Factors.* Washington, DC: McKinsey Global Institute.

McLain, Sean. 2015. "Why E-Commerce Can Be a Big Pain for India's Deliverymen." *Wall Street Journal,* March 29.

Messinger, Paul R., and Chakravarthi Narasimhan. 1995. "Has Power Shifted in the Grocery Channel?" *Marketing Science* 14(2): 189–223.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny. 1989. "Industrialization and the Big Push." *Journal of Political Economy* 97(5): 1003–26.

Oi, Walter Y. 1992. "Productivity in the Distributive Trades: The Shopper and the Economies of Massed Reserves." In *Output Measurement in the Service Sectors* edited by Zvi Griliches, 161–191. University of Chicago Press/NBER.

Pozzi, Andrea. 2013. "The Effect of Internet Distribution on Brick-and-Mortar Sales." *RAND Journal of Economics* 44(3): 569–83.

Reardon, Thomas, Spencer Henson, and Julio Berdegué. 2007. "'Proactive Fast-Tracking' Diffusion of Supermarkets in Developing Countries: Implications for Market Institutions and Trade." *Journal of Economic Geography* 7(4): 7399–7431.

Reardon, Thomas, and C. Peter Timmer. 2012. "The Economics of the Food System Revolution." *Annual Review of Resource Economics* 4: 4225–64.

Tedlow, Richard S. 1990. *New and Improved: The Story of Mass Marketing in America.* New York, NY: Basic Books.

Traill, W. Bruce. 2006. "The Rapid Rise of Supermarkets?" *Development Policy Review* 24(2): 163–74.

Journal of Economic Perspectives—Volume 29, Number 4—Fall 2015—Pages 135–154

Online Higher Education: Beyond the Hype Cycle[†]

Michael S. McPherson and Lawrence S. Bacow

When two Silicon Valley start-ups, Coursera and Udacity, embarked in 2012 on a bold effort to supply college-level courses for free over the Internet to learners worldwide—regardless of whether they were enrolled in a traditional college or university—the notion of the Massively Open Online Course (MOOC) captured the nation's attention. The Harvard- and MIT-based non-profit edX quickly joined the field. Reporters and analysts debated not so much whether MOOCs would transform the landscape of higher education, but only how quickly they would do so—and how many existing universities, if any, would survive the onslaught.

But while the leading MOOC firms remain in business, they have had their struggles, as start-ups often do. Rather than posing an imminent threat to traditional universities, MOOCs now seem like an example of what the consulting firm Gartner (n.d.) calls the "hype cycle," which follows a five-step process of initial trigger, inflated expectations, and trough of disillusionment, before reaching the more productive, final stages of "slope of enlightenment" and "plateau of productivity," when the technology's broad applicability leads to widespread mainstream adoption. In the case of MOOCs, the inflated expectations include predictions that

■ *Michael S. McPherson is President of the Spencer Foundation, Chicago, Illinois. From 1996 to 2003, he was President of Macalester College, St. Paul, Minnesota. Lawrence S. Bacow is Leader-in-Residence, Center for Public Leadership, Harvard Kennedy School, Cambridge, Massachusetts. From 2001 to 2011, Bacow was President of Tufts University, Medford, Massachusetts. He is also a Senior Adviser to Ithaka S&R, New York, New York. The authors' email addresses are mmcpherson@spencer.org and lawrence_bacow@hks.harvard.edu.*

[†] For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.29.4.135

doi=10.1257/jep.29.4.135

MOOCs will make quality higher education globally available at near-zero marginal cost and with little need for many currently employed faculty members.

One challenge facing MOOCs and other forms of instruction intended for use outside institutional frameworks is the need for users to exert considerable self-discipline to stay with a program that is undertaken individually. MOOCs in particular have very low completion rates. Banerjee and Duflo (2014) documented this point in an imaginative way. The authors offered a MOOC on global poverty, and they set a registration deadline that they didn't in fact enforce. The lack of enforcement produced a kind of accidental regression discontinuity analysis that allowed a comparison of results between those who made the deadline and those who missed it. They found that those who made the deadline, even by a day or two, did better on grades and completion than those who missed it. Interpreting the failure to register on time as a measure of self-discipline, the authors conclude that the "noncognitive" capacity for self-control was significant in accounting for success in online learning.

Although MOOCs are an interesting experiment with a role to play in the future of higher education, they are in fact a surprisingly small part of the online higher education scene, given the attention they have received. We believe that online education, at least online education that begins to take full advantage of the interactivity offered by the web, is still in its infancy (Bowen 2013). Thus, we begin by sketching out the several faces of online learning—asynchronous, partially asynchronous, the flipped classroom, and others—as well as how the use of online education differs across the spectrum of higher education. We then turn to some of the main issues posed by the growth of online education, which are how it will affect cost and convenience, how it will affect student learning, and how it will affect the role of faculty and administrators. We argue that the process by which online education spreads through higher education is likely to be slower than many commenters expect. Furthermore, while we hope that online education will bring substantial benefits, there is also the possibility that it could lead to less-attractive outcomes. This could happen if legislators use the existence of online education as an excuse for sharp cuts in higher education budgets that lead to lower-quality education for many students, at the same time that richer, more-selective schools are using online education as one more weapon in the arms race dynamic that is driving costs higher.

The Current Status of Online Learning

In evaluating statistics about online education, it's important to recognize that "college" has become a capacious term in the United States, encompassing essentially any form of education or training high school graduates receive, from studying late Victorian poetry to learning to drive a truck. Moreover, what is described as "online education" involves many different uses of technology to facilitate learning.

During the last decade, estimates of the prevalence of online education have often drawn from the annual studies of the Babson Survey Research Group, which

Table 1
Student Participation in Online Courses by Type and Selectivity of Institution, 2013

	All courses online	*Some courses online*	*No courses online*
Type of institution			
Total	11%	15%	75%
Public, 4-years	6%	17%	77%
Private, not-for-profit, 4-years	10%	8%	82%
Public, 2-years	10%	18%	72%
Private for-profit	34%	6%	60%
Carnegie classification			
Community college	10%	18%	72%
4-year, public, non-research institutions	8%	17%	75%
Public research universities	4%	17%	80%
Private not-for-profit research universities	2%	5%	93%
Liberal arts colleges	0%	2%	98%
For-profit institutions	34%	6%	60%

Source: US Department of Education, NCES, Integrated Postsecondary Education Data System (IPEDS)
Notes: The column "All courses online" shows the percentage of undergraduate students exclusively enrolled in distance education courses as defined in IPEDS. The column "Some courses online" shows the percentages of undergraduate students with some but not all of their enrollment in distance education courses. And the column "No courses online" shows the share of undergraduate students who are not enrolled in any distance education course. "Community college" is defined as all public institutions in Carnegie's Associates category, and "4-year, public, non-research institutions" are defined as 4-year or above public institutions not classified as "Doctorate" nor as "Associate" by Carnegie. "Public research universities" are all public institutions in Carnegie's Doctorate classification, and "Private not-for-profit research universities" are all those in Carnegie's Doctorate classification that are private not-for-profit. "Liberal arts colleges" correspond to all those private not-for-profit institutions in Carnegie's "Baccalaureate Colleges—Arts & Sciences" classification. The sample is restricted to Title IV eligible institutions; branches are not included.

regularly surveys chief academic officers at a large number of schools: for example, the 2014 data includes responses from 2,800 higher education institutions. By these estimates, about one-third of all students enrolled in college in a given year take at least one course in which 80 percent or more of the material is provided online (Allen and Seaman 2015). The US Department of Education recently began to conduct its own survey of online education as part of its Integrated Post-Secondary Education Data System (IPEDS), with full coverage of the roughly 4,900 US institutions of higher education. As shown in Table 1, IPEDS data indicates that as of 2013, about 26 percent of all students took at least one course that was entirely online, and about 11 percent received all of their education online. Although the IPEDS data is drawn from a larger sample, the Babson data remains useful because it asks a broader set of questions about attitudes and trends regarding aspects of online education.

The federal IPEDS data provides information on online course-taking in different parts of the higher education system. At liberal arts colleges and private not-for-profit research universities, the use of online education is minimal. There is somewhat

greater use at public research universities. At less-prestigious and less-selective institutions, including the "non-research" public universities and community colleges, there is greater use of online instruction. And the greatest use, by a considerable margin, is in for-profit institutions. Deming, Golden, Katz, and Yuchtman (2015) show that use of online instruction is particularly prevalent in for-profit colleges that operate as chains, in contrast to stand-alone or mom-and-pop institutions, suggesting that economies of scale are important to the economics of online learning. In general, use of online learning appears to be inversely proportional to prestige and selectivity, a point to which we will return below.

Online courses employ a variety of formats. Most existing online courses for undergraduates are offered "asynchronously," meaning that students individually determine when they interact with the online material. Even with fully online asynchronous courses, the amount of interaction between students and the course's computer technology varies substantially. At one end of the spectrum, there is no interactivity: Austin Community College, for example, offers a number of courses which entirely consist of recorded lectures broadcast online and through cable television (Austin Community College District 2015; Lack 2013). At the other end of the spectrum, in certain MOOCs in advanced computer programming, students submit computer applications they have designed and receive immediate machine-generated debugging feedback that pushes the application to extremes as a way of pointing out potential weaknesses. In such cases, the feedback is intense and immediate.

There is growing interest in "blended" or "hybrid" courses that combine face-to-face instruction with digitized online instruction. In some respects, almost every course taught today is a hybrid and incorporates at least some online component. For example, instructors routinely distribute readings and assignments electronically. Many encourage their students to view videos and other supplementary materials online. Students often submit papers and problem sets electronically and receive feedback on their assignments in the same format. Email has displaced traditional office hours at many institutions. While students may not be formally enrolled in "online courses," the influence of digital content in the academy is ubiquitous. One of the most popular methods of exploiting digital learning is the idea of "flipping the classroom." In a flipped classroom, lectures are recorded and viewed by students asynchronously. The class time that is freed up from the lecture is then used for more intense, interactive exercises and discussion. In theory, the flipped classroom allows more class time to be devoted to more-active learning. We describe the cost implications of flipped classrooms later in this essay.

Although it may seem natural to assume that online course delivery (especially in asynchronous mode) is best-suited to subjects where there are objectively right and wrong answers, in fact the range of online courses and even fully online degrees available is quite wide. Penn State University, for example, offers fully online introductory courses in subjects that include anthropology, comparative literature, economics, history, philosophy, psychology, and sociology (Penn State World Campus n.d.; Lack 2013).

A significant growth area for online learning is in professional education at the graduate level. It may seem counterintuitive that online instruction could thrive in areas like teacher preparation or nursing, where live interaction with students or patients appears to be central. Yet some of the largest programs in the country rely heavily on online instruction and are frequently offered with no residency requirements.[1] Two factors seem vital to the effectiveness of such programs. First, the programs typically require either arrangements with local schools and hospitals to cooperate in providing clinical experience of documented quality for participants in the program, or else brief visits to campus are required for that purpose. Second, synchronous online discussion sessions frequently play an important role in instruction in these programs.

We lack rich descriptive data about the extent of the use of online instruction of these varying kinds in different subject matters and different segments of the higher education system. This lack of data is partly explained by the rapid speed of change, but it also reflects the fact that the study of instructional methods in higher education has long been neglected among both scholars and administrators.

We noted earlier that more-selective and prestigious colleges and universities make less use of fully online courses than other institutions do. What explains this pattern of adoption? A natural explanation is that more-selective institutions compete on the basis of personal service, prestige, and brand while less-selective places are offering something closer to a commodity product (Brewer, Gates, and Goldman 2001; Zemsky 2009). Consumers cannot judge the quality of an education by inspection—it is a classic example of an "experience" good (Nelson 1970)—so symbolic indicators like the degree of selectivity, the quality of the facilities, and the difficulty of admission are emphasized. For institutions that are more selective and wish to present themselves that way, pedagogical innovation may be risky, especially if it appears to trade off the opportunity for intimate personal interaction with faculty against a technology that may appear to "cheapen" the product. Ironically, the highest-end institutions may be reluctant to adopt online learning unless they can demonstrate that it is actually *more* expensive than existing methods—a point to which we will return. These considerations weigh less heavily at less-selective institutions, whose main selling point is their ability to deliver a credential that has a demonstrable value in the marketplace. It is at the less-selective institutions that innovations in online education that can lower cost, expand availability, and/or increase convenience of access have a good chance of succeeding, at least in the near term. (If it is the case that more-traditional training is more effective than online training—the evidence to date is inconclusive—in the long run, the online strategy could come into question: employers would eventually notice a productivity

[1] See, for example, "Master of Arts in Teaching," USC Rossier Online http://rossieronline.usc.edu/academics/master-of-arts-in-teaching-program/ (accessed August 25, 2015) and "College of Nursing," Medical University of South Carolina, http://academicdepartments.musc.edu/nursing/academics/masters/msn_faq.htm (accessed August 25, 2015).

difference between traditionally trained and online-trained employees. However, this would be a slow-working corrective mechanism at best.)

Cost and Convenience

What lies behind the rapid growth in online education over the last decade, from about 10 percent of students taking at least one online course in 2002 to 33 percent in 2012 (as measured in the Allen and Seaman 2014 data)? The two most obvious explanations involve cost and convenience. The Internet, at least in its asynchronous use, affords delivery of instructional material that scales.

It is worth noting that the Internet is not the first, nor the second, instructional technology that offered the promise of vast scale at low cost. Berland (1992), citing a popular commentator named Waldeman Kaempffert writing in 1924, reported that "there were visions of radio producing 'a super radio orchestra' and 'a super radio university' wherein 'every home has the potentiality of becoming an extension of Carnegie Hall or Harvard University.'" Craig (2000) reports that "the enthusiasm for radio education during the early days of broadcasting was palpable. Many universities set up broadcast stations as part of their extension programs and in order to provide their engineering and journalism students with experience in radio. By 1925 there were 128 educational stations across the country, mostly run by tertiary institutions" (p. 2831). The enthusiasm didn't last—by 1931 the number of educational stations was down to 49, most low-powered (p. 2839). This was in part the result of cumbersome regulation, perhaps induced by commercial interests; but the student self-control problem, similar to that observed by Banerjee and Duflo (2014), likely played a role as well. As NBC representative Janice Waller observed, "Even those listeners who clamored for educational programs, Waller found, secretly preferred to listen to comedians such as Jack Benny. These "intellectually dishonest" people "want to appear very highbrow before for their friends . . . but down inside, and within the confines of their own homes, they are, frankly, bored if forced to listen to the majority of educational programs" (as quoted in Craig 2000, pp. 2865–66).

The excitement in the late 1950s about educational television outshone even the earlier enthusiasm for radio. An article by Schwarzwalder (1959, pp. 181–182) has an eerily familiar ring: "Educational Television can extend teaching to thousands, hundreds of thousands and, potentially, even millions. . . . As Professor Siepman wrote some weeks ago in *The New York Times,* 'with impressive regularity the results come in. Those taught by television seem to do at least as well as those taught in the conventional way.' . . . The implications of these facts to a beleaguered democracy desperately in need of more education for more of its people are immense. We shall ignore these implications at our national peril." Schwartzman goes on to claim that any subject, including physics, manual skills, and the arts can be taught by television, and even cites experiments that show "that the discussion technique can be adapted to television."

Clearly neither radio nor television has fulfilled its early promise (hype?) as a tool for college instruction. Yet even in the Internet age, similar approaches survive.

Thus, a minimal and inexpensive form of online learning is the rebroadcast of taped live lectures. At the University of Florida, for example, the "Principles of Microeconomics" course uses this approach: otherwise, the course enrollment of 1,500 students would substantially exceed the size of the largest lecture hall on campus (Lack 2013; Gabriel 2010). Such one-way transmission of information uses the Internet as a means of delivering video, but with the advantage that students can stop, rewind, replay, and fast forward (but not ask questions or receive feedback online). Some MOOCs take (more or less) this form, providing a series of mini-lectures, sometimes interrupted by brief quizzes to provide an incentive for students to keep looking at the screen.

Asynchronous online courses are attractive to institutions because of their low marginal cost and their potential to expand markets substantially by offering credit-bearing courses to students in distant locations. Of course, "distance learning" itself is anything but new. Queen Victoria in 1858 authorized the University of London to offer degrees through its International Programmes to students throughout the world who could not reside at universities. For example, Nelson Mandela studied law under the university's International Programmes while in prison.[2] Countless students have benefited from these and other correspondence programs, with degrees offered through proctored examinations.[3] With the arrival of online education, such courses have become core products at a number of for-profit colleges and universities including, for example, the University of Phoenix and Capella University. There has also been considerable entry to this market from some public and private nonprofit universities. Arizona State University has recently used an advertising campaign to help expand its role in this market, with more than 70 degree programs offered entirely online. Southern New Hampshire University has risen from a relatively obscure small, private nonprofit institution to become a major player nationally in fully online degree programs.

Asynchronous online courses are an appealing substitute for correspondence courses delivered by physical mail. Moreover, the ability to market courses online may change the grounds of competition for place-based public universities substantially, with potentially far-reaching effects on quality and price. Such courses may also offer opportunity to improve time-to-degree in large institutions. A common problem in large public universities is that students are unable to enroll in gateway subjects needed to complete their degree due to limited enrollment. Online courses may be a means to improve speed or rates of completion at large institutions, although we are not aware of any hard evidence on this point.

[2] See University of London, International Programmes, "Timeline." http://www.londoninternational.ac.uk/our-global-reputation/our-history/timeline/ (accessed August 25, 2015).

[3] The Open University, established in the United Kingdom in 1966, today has a student enrollment of over 200,000. Over the years, it has used a variety of distance learning technologies to reach its students including radio, TV, and online learning.

However, the key difference in potential effectiveness between Internet technology and educational TV or radio is not just a more-convenient distribution channel, but the former's capacity for two-way interaction between student and instructor (or virtual instructor), and interactions among students. (Internet technology, as noted earlier, also has the advantage of allowing students to start, stop, and rewind video content, something that traditional educational radio and TV did not.) Instructional systems that provide automated feedback to students based on their progress can enable self-paced designs that permit faster progress for more adept or industrious students.[4] This opens the possibility of so-called "adaptive learning" systems, in which not only the pace but also the content and pedagogy of lessons might adjust automatically in response to evidence about a student's comprehension revealed by her interaction with the software. In principle, sufficiently advanced versions of such technologies could reproduce at least some of the kinds of sensitive give-and-take that skilled teachers and responsive students produce together.

Computer-adaptive techniques are now in wide use for examinations, including the Graduate Record Examinations supplied by the Educational Testing Service and the Common Core state-level examinations at the K-12 level developed by the Smarter-Balanced Consortium. But designing computer-adaptive instruction appears to be a much harder problem than designing examinations. Examinations simply involve sampling from within a given educational domain, a task that is easy to improve on with relatively simple rules about what multiple choice or short answer question to ask next. Instruction involves working with students to help them gain competence over a particular domain of knowledge or skills, which involves the very difficult challenge of diagnosing the reasons for their mistakes. Intensive work on the instructional challenge has been undertaken at Carnegie Mellon University and other places, but sophisticated computer-adaptive instruction is not currently in widespread use in online courses.

Richly interactive online instruction is obviously much more expensive than Internet-delivered television. The development costs for Carnegie Mellon's sophisticated but far from fully computer-adaptive courses in statistics and other fields have been estimated at about $1 million each (Parry 2009). Although future technical developments will reduce the costs of providing a course of a fixed level of quality over time, those future technical developments will also encourage the provision of additional features. Universities can invest in improving the production values of such television programs at the margin in ways that range from multiple camera angles to the incorporation of sophisticated graphics

[4] Self-paced instruction of course preceded the Internet. As one example, in the late 1950s, a company called Science Research Associates introduced a color-coded system of reading cards that was widely used as a supplement in elementary schools. This popular innovation helped create the fortune that SRA's founder Lyle Spencer used to endow the Spencer Foundation, where one of the authors of this article now works. Recently the original SRA reader system has been digitized and is still sold (SRA Reading Laboratory 2015; *Wikipedia* entry on "Science Research Associates," last modified February 3, 2015, https://en.wikipedia.org/wiki/Science_Research_Associates).

and live location video. Many interactive courses could also conceivably benefit from regular updating based on recent events or scholarship, although we note that currently, it is quite cumbersome to update most online courses. While an instructor in a traditionally taught course can easily drop new material into the syllabus or even an individual class, modifying an online course usually requires reshooting video, editing existing content, modifying software, and so on. These changes cannot be done quickly or in real time. They often require coordination of multiple parties including the instructor, instructional designer, camera operators, editors, web designers, and software engineers. Our point is that while online courses offer the potential for constant modification and updates, realizing this potential may in fact be expensive, leading to less-frequent updates than for traditionally taught subjects.

Highly sophisticated and therefore expensive online courses are likely to be financially feasible only when offered at a scale far beyond that of the individual university. Those who foresee the widespread adoption of adaptive learning technology often underestimate the cost of producing it. Stanford President John Hennessey, in a recent lecture to the American Council of Education, estimated that the cost of producing a first-rate highly interactive digital course to be in the millions of dollars (Jaschik 2015). Few individual institutions have the resources to make such investments. Furthermore, while demand may be substantial enough to support such investments for basic introductory courses in fields that easily lend themselves to such instruction, it is unlikely that anyone will invest in the creation of such courses for upper-level courses unless they can be adopted at scale.

It's also important to remember that any effects of online education on costs will occur as part of the overall system of US higher education. Selective institutions of higher education have long known how to make education cheaper. It involves larger classes, less student–faculty contact, less-intensive hands-on learning, fewer curricular options, and less in the way of student services and amenities. Some of the features that make elite colleges expensive may well be essential to the learning experience, but others pretty transparently are not and are still well-loved. In our time as presidents of Macalester and Tufts (and previous time as senior administrators at Williams and MIT), not once did we have a student or their parents ask us to do any of the above and lower the price. If anything, people always wanted more.

To put it bluntly, selective institutions of higher education actually compete to be among the least cost-effective providers of educational services. As anyone knows who has taken a prospective student on a college tour, elite colleges routinely advertise small classes, frequent student-faculty contact, and lots of opportunities for hands-on learning. Similarly, colleges compete by offering a vast array of curricular options: multiple majors and minors, again with faculty support. Curricular entropy adds to cost by committing institutions to offer a diverse array of courses over time, even when demand for certain courses or majors dwindles. This approach clearly appeals to students choosing these schools, to the parents who are typically footing the bill for a substantial share of the cost, and to faculty choosing where to work. Many of the most selective institutions are also among the best endowed. Wealthy

colleges can in fact gain competitive advantage by raising expectations for how resource-intensive a "good" college should be.

Indeed, there is a real chance that at least in selective higher education, technology will actually be used to raise rather than lower cost. There are obvious ways to use online materials to complement rather than to substitute for in-person instruction. Flipping the classroom, as we will explain further, is one. Instructors can also import highly produced video material—either purchased or homemade—to complement their classes, and there could easily emerge a market in modular lessons aimed at allowing students to extend material farther or to get a second take on a difficult set of concepts. If individual faculty members are authorized to make these choices, and universities agree to subsidize expensive choices, costs seem likely to rise. With no reduction in faculty input, but an enhancement in other resources, the worry of parents that technology is undermining their children's experience is quelled even as faculty are assured they are not losing either their autonomy or their jobs.

While the absolute number of highest-prestige institutions of education is of course small, they cast a long shadow as market leaders and as institutions of origin for many faculty at less-prestigious places. If online technology stimulates higher costs at the most selective institutions, it will set expectations for how higher education will function elsewhere.

While flipping the classroom and using time previously allocated to lecture for more-intensive student-faculty interaction may improve instruction (a topic discussed in the next section), it offers little promise of actually reducing cost. Lectures are relatively cheap. Smaller courses that focus on more discussion-oriented teaching require lots of space and lots of instructors and much more intense student–instructor interaction. Moreover, in the typical flipped classroom, faculty are substituted for graduate instructors, further increasing costs. Finally, the early returns from experiments at Harvard suggest that students at that institution are not wild about the flipped classroom. They find that they spend considerably more time viewing the recorded lecture (and responding to questions about their comprehension) than simply sitting passively through a live lecture. This may be due to the fact that students are being asked to take more responsibility for their learning in a format that is still unfamiliar, or it may be attributable to poor course design.[5]

Many observers envision that the savings in large lecture courses will come from replacing the live lecturer with an online version of that person. But most of the cost of large lecture courses comes from staffing discussion sections with graduate students and obtaining classrooms for the sections and, in science and engineering, substantial lab space as well. Real savings would come from finding a way to replace the discussion sections with some version of interactive sessions run by technology. Sophisticated courses like those developed by Carnegie Mellon

[5] Faculty are also still learning how to teach in this new format. Students report redundancy between material covered online, and that covered in discussion section (Derek Bok Center for Teaching and Learning 2014).

University in the Open Learning Initiative have the ambition to eliminate or reduce the need for discussion sections. Because section teaching relies so heavily on the presence of graduate students, undergraduate education at research-oriented institutions effectively subsidizes graduate education. Indeed, we believe this subsidy may cause many institutions to inflate the size of their graduate programs, resulting in the overproduction of PhDs in some fields. Shrink the number of teaching assistants, and many departments will struggle to support as many graduate students. Of course, we can easily imagine the complaints from our colleagues if online technology becomes a mechanism to reduce graduate student enrollment.

Changes in Student Outcomes and Pedagogy

What do we know about the educational effectiveness of online instruction in all its varieties compared to traditional methods? Not much. Thorough surveys of this literature find very few examples of well-designed experiments and quasi-experiments, and these generally do not find statistically significant differences in student outcomes between online or hybrid courses and traditional courses.[6] Similarly null findings emerge from a small number of regression-based studies with reasonably good statistical controls on comparative student quality and other variables (Means, Toyama, Murphy, Bakia, and Jones 2009; Bell and Federman 2013).

The fact that existing studies have not yielded decisive "victories" for traditional forms of education is surely a source of disquiet, if not alarm, for many faculty defenders of traditional practice. Indeed, based on the fact that a number of these studies have not so far generated clear winners and losers, it is tempting to conclude that online, hybrid, and traditional courses yield essentially the same outcomes in most circumstances. This conclusion would be overreaching. More plausibly, the evidence suggests that pedagogy is multidimensional in such a way that it cannot be reduced to "online" and "offline." As we have seen, instruction with a significant online component takes place in a wide variety of settings and subject matters, and the variation in the ways that "traditional" instruction is delivered is also great. There is also a wide range of learning outcomes that may be of interest. Tests that assess recall of facts or computational skill may yield quite different results from ones that assess conceptual grasp or problem-solving ability.

[6] In one of the strongest studies, Bowen, Chingos, Lack, and Nygren (2014) contrast traditional and hybrid versions of a statistics class in a randomized controlled trial and find no significant differences in learning outcomes, with small standard errors and reason to believe there would be cost savings from widespread use of the hybrid method. Several experimental and quasi-experimental studies have found small negative effects of hybrid versus traditional instruction in various contexts (for example, see Joyce, Crockett, Jaeger, Altindag, and O'Connell 2014, Kwak, Menezes, and Sherwood 2015, and Olitsky and Cosgrove 2014, as reported in Wu 2015.) These findings say little about fully online instruction nor about likely outcomes in courses with different subject matters. Figlio, Rush, and Yin (2010) undertook a randomized controlled trial comparing traditional and fully online microeconomics classes and found modest evidence that students in the "live" course learned more.

Ultimately, there are limits to what can be learned from piling up longer lists of A to B comparisons between online and traditional versions of the same course particularly when the definitions of "online" and "traditional" vary from one example to the next. (This is not to say that we cannot learn from such comparisons, only that we have to be careful about generalizations.) What we really lack is an adequate understanding of what makes for effective instruction in particular settings, with students who have particular characteristics, with effectiveness judged by the achievement of well-defined and valued outcomes. The Pittsburgh Science of Learning Center, a joint effort of Carnegie Mellon University and the University of Pittsburgh, is an important locale for such work.

Obviously, experimenting with various approaches to online learning will and should continue in parallel with more theoretical research; equally obviously, universities will continue to act even in the absence of strong evidence. After all, universities have been quite willing to offer traditional instruction in various formats for centuries with remarkably little attention to comparative educational effectiveness!

Institutions of higher education have had centuries to perfect traditional chalk-and-talk pedagogy. It seems unlikely that great gains in the speed or effectiveness of education remain to be had with this approach. Whatever one believes about the effectiveness of online education today, it is likely to get better—and probably dramatically better—over time.

Shifts in the Role of College and University Faculty

Traditionally, college and university teachers are accorded nearly complete autonomy in how they conduct their classes: indeed, we have sometimes heard assertions that academic freedom should protect faculty from any demand for information about how they teach. However, the growth of online education seems likely to shift this historical pattern in ways that could give greater power to administrators.

One perhaps ironic byproduct of the interest in online learning is that administrators and scholars are beginning to ask more questions about actual instructional practices, including "traditional" approaches, and how we can assess their quality and effectiveness. Also, online education allows administrators to "peer over the shoulder" of instructors to monitor not just what is being taught, but a range of other metrics: for example, how long it takes an instructor to respond to student questions, the turn-around time for assignments, and instant breakdowns of data on how different groups of students perform in a given class.

The use of technology to import classes from other institutions—highly desirable in theory as a way to achieve economies of scale in producing high-quality online instruction—raises thorny issues of its own. While college and university faculty are not reluctant to employ a text authored by someone else, they are reluctant to be perceived as little more than facilitators of someone else's course (Bacow, Bowen, Guthrie, Lack, and Long 2012). Thus, faculty may resist the wholesale importing of lectures by well-known scholars from other institutions. For example, in 2013,

the philosophy department at San Jose State University rebelled against efforts to import Harvard philosopher Michael Sandel's well-regarded course on "Justice" (Hartnett 2013). By contrast, in 2014 Yale agreed to import Harvard's most popular course: "Introduction to Computer Science" (Bernhard 2014). Yale students will watch the Harvard lectures online and will be taught in sections by Yale faculty and graduate students, whose efforts will be coordinated with colleagues from Harvard. However, this second example is likely to be atypical. We suspect that computer scientists are outliers in their openness to digital learning and are likely to remain so for a good while.

Unlike written materials (including textbooks) that can be mashed up, reordered, and supplemented, most online courses are not currently produced in formats that lend themselves to customization. By contrast, those who are attracted to teaching MOOCs are often seeking a larger audience for their particular approach to the material. As a result, they often are not interested in reducing their integrated approach to the material into a series of short modules that can be reordered or customized by others. To do so would lose the integrity of the whole that they seek to convey. Until digital content is designed in a more flexible way, faculty are likely to be slow to adopt it, because they cannot incorporate it easily into their personal conception of how a particular body of material should be taught.

Also, at least some faculty fear that technology may weaken their relationships with their students (Bacow, Bowen, Guthrie, Lack, and Long 2012). Many value these relationships, and fear that they would isolate themselves from students by embedding their course in a digital environment—even if we reach the point where highly sophisticated and responsive robots provide really excellent instruction. Many students also enjoy face-to-face interaction with their professors, at least at places where such interaction is common and expected.

Finally, we note that widespread adoption of online learning will require the resolution of a number of potentially divisive intellectual property issues that will often pit faculty against administration. Standard practice on most campuses is that faculty own the copyright for their course materials and lectures. For example, even though a textbook may have been produced using substantial institutional resources, the copyright is owned by its author. Institutions are unlikely to make the substantial investments needed to create high-quality interactive online courses if they are unable to recover their costs. Similarly, faculty are unlikely to invest the time necessary to produce this content—which many report to be substantially in excess of what is required to teach a conventional course—if they are not going to be compensated for their efforts. Clearly, a revenue sharing model is in order, but we know of no conventions yet established governing the relative rights of faculty and sponsoring institution to the revenue generated from online learning.

Could Coursera, Udacity, edX, and others fill this gap? At least so far, their offerings are not especially close to an adaptive learning technology with the high production values that would impress a typical college student in the second decade of the 21st century. The original business model for MOOCs (whether offered by for-profit or nonprofit ventures) was predicated on the possibility of generating

profitable revenue streams from courses offered free online outside the framework of existing colleges and universities, a strategy that has not panned out so far. Alternative options would be for MOOCs to market their wares to existing universities or for universities themselves to develop cross-institutional arrangements that would permit sharing development costs. Griffiths, Chingos, Mulhern, and Spies (2014) report findings from a set of studies imbedding elements of MOOCs into the offerings of a university system. Again, these alternatives require that colleges and universities solve a set of quite difficult economic and governance problems.

Textbook publishers are also entering the market for online content. Many modern textbooks already contain some digital content, whether it is a CD-ROM with supplementary materials or access to a website that curates the same. Furthermore, textbook publishers have access to production capacity and distribution channels that many academic institutions lack. Time will tell whether textbook publishers succeed in providing online content on their own without the brand identity provided through partnership with a sponsoring academic institution.[7]

Similarly, because digital learning technologies make it possible for faculty to "teach" elsewhere without being physically present, we can foresee challenges to conventional conflict-of-interest and conflict-of-commitment policies. For example, can a faculty member employed at one institution produce an online version of his or her course and sell it for adoption at a competitor school? At present, institutions are happy to have their faculty's textbooks adopted elsewhere. How will they feel if textbooks are ultimately replaced by digital course packs with embedded exercises and lectures from their own faculty?

We also note the potential for conflict over control of future versions of online courses. While an author may control the production of future editions of a textbook, should a sponsoring institution be able to modify an online course without the permission of the faculty member who appears on screen? What if the faculty member is no longer employed by the institution? No longer alive? Similarly, if the faculty member decamps to another institution, can that faculty member take the course along to the new institution? Can a faculty member who is moving between institutions reproduce a course from the first institution, at least in some form, and then have that course compete with the former institution? Again, these issues and others will need to be resolved before institutions and their faculty are willing to make the necessary investments to produce the highest-quality interactive and updated online content that could truly disrupt conventional forms of instruction. These issues raise thorny legal and governance challenges where progress will be hard won and, we suspect, not quick (Bowen and Tobin 2015).

[7] For an example of a sophisticated attempt to enter this market, see Pearson's efforts to market MyLabs, a series of interactive online modules for teaching a full range of subjects. See "My Lab™ & Mastering™," http://www.pearsonmylabandmastering.com/northamerica/errors/index.html, accessed August 25, 2015.

Summing Up: Dystopian and Optimistic Possibilities

Online education offers the tantalizing possibility of comparable learning outcomes at potentially lower cost. Some commentators like Clayton Christiansen predict that online education will be a dramatically disruptive technology (for example, Christiansen and Horn 2011). For example, Christiansen has written that 15 years from now, a new class of institutions, or extra-institutional provision of cheap high-quality education, will drive half the universities in the country out of business.

We are skeptical of such extreme predictions. We have suggested that a range of issues must be addressed before online education is likely to threaten the traditional structure of higher education. It will take some time before the technical problems of achieving high-quality interactive instruction at scale are solved, partly because it is not at all clear who will make the big investments necessary to do so. A large range of institutional and practical obstacles will stand in the way of rapid substitution of cost-saving technology-mediated education for traditional modes of instruction at selective colleges and universities

Like all professors who seek to study higher education, we are handicapped by being deeply experienced in the kinds of places where we have studied and worked, and often deeply ignorant of the vast heterogeneity of the other parts of the so-called "system" of higher education. Nonetheless, we offer some thoughts about potential futures for online classes in different kinds of higher education institutions.

Broad-access unselective institutions are already among the largest users of online instruction. These institutions are responsible for the education of many students—at least half of all those enrolled in postsecondary education—and they disproportionately educate lower-income students and students of color. Enabling technological advances to support improvement in the educational success of these institutions at manageable cost is an important goal, arguably the most important goal for using technology to improve American higher education. (Of course, the implications of these technologies for global learning would be potentially gigantic.) There is especially high potential for online education to cater to the large number of nontraditional students, which includes adult learners and those who have a very high opportunity cost of attending college whether at the undergraduate or graduate level. For this group of students, asynchronous online learning can be a godsend. Opportunities surely exist for technology to penetrate this market further, and quality is likely to improve as faculty and others figure out how to take better advantage of new educational technology. As the technology improves and as more institutions adopt it, more of these students are likely to receive all or at least some of their education online.

Yet this great opportunity is accompanied by considerable risk. It is all too easy to envision legislators who see a chance to cut state-level or national-level spending that supports higher education by imposing cheap and ineffective online instruction on institutions whose students lack the voice and political influence to demand quality. It's equally easy to imagine for-profit institutions proffering online courses in a way that takes advantage of populations with little experience with college in a marketplace where reliable information is scarce.

What about the large number of institutions with a medium level of selectivity? These institutions already deliver the first two years of undergraduate education relatively cheaply in very large lecture classes, often with support of brigades of teaching assistants. As we have previously noted, technology has the capacity to reduce costs here not so much by replacing live with online lectures as by reducing the demand for section teaching, with consequent reduction in the demand for graduate student teaching assistants and graduate enrollment (which would be a disruptive but perhaps not a bad thing). It remains to be seen whether online education in the form of interactive pedagogy (as opposed to the use of video clips and machine-based testing) will penetrate the upper-division courses in these schools, because of the large and differentiated number of such courses and their relatively small enrollments. Also, not every subject is equally amenable to online instruction, so entire disciplines may be relatively untouched for the foreseeable future by the development of more-advanced educational technology.

For the most selective institutions of higher education—less than 5 percent of enrollments but a vastly larger percentage of both popular and academic attention—some of these investments in technology-enhanced learning experience may provide benefits that outweigh their costs. Yet for those, like us, who think the elite colleges already spend more than can be justified on educational grounds, the idea of technology as a cost enhancer is unwelcome. One route toward heading off this potential trend is to recognize the collective action problem that underlies it. In the topsy-turvy world of prestige competition, raising spending relative to the existing norm can be a winning competitive strategy. But when all elite institutions raise spending together, the arms race may well result in welfare loss and political damage compared to a world where colleges jointly exercise restraint. Any selective college or university that makes a bold move to economize by using online instruction to substitute for faculty in some measure, risks harming its reputation. However, if a group of institutions move in a coordinated and carefully designed way to explore such possibilities, both cooperating with each other to share development costs and emphasizing that they will guard against reductions in quality, such a group might be seen as innovators rather than cheapskates. Coordinated activity like this requires some level of trust among the institutions involved and raises a risk of violating antitrust policy. After all, back in 1992, MIT and a number of other colleges were sued by the US Department of Justice for alleged price fixing because of efforts to coordinate the determination of need in granting of financial aid awards. (Other institutions settled with the Justice Department and signed consent decrees, while MIT litigated the issue and ultimately prevailed on appeal, as reported in Bangs 1994.) But as William Bowen (1992) argued persuasively in his legal brief and testimony in that case, coordination among nonprofit institutions can at times be justified on public policy grounds.[8]

[8] In his testimony, Bowen (1992) noted that before he closed a Slavic languages program at Princeton, he called colleagues at Columbia and Yale to make sure they were enthusiastic about their programs, thereby ensuring that such an educational program would not disappear from the land. Bowen argued that this type of cooperation should be allowed among nonprofits.

Finally, for those who believe that brilliantly produced online courses taught by a handful of the very best faculty in the world will eliminate the demand for live versions of the same courses, we note the continuing vibrant and growing market for live concerts, theatrical productions, and sporting events. Cheap digital downloads of music have not eliminated the demand for live concerts, nor has the availability of live sports on TV (often with better viewing angles, instant replay, and simplified access to bathroom facilities) eliminated the demand for tickets to live sporting events. Even more puzzling to technology fans, a rising proportion of live-theater shows on Broadway display a kind of technological regression by adopting the plots of expensively produced movies seen by millions to recycle them with extremely labor-intensive live reproductions. If anything, the increased availability of digital programming may have even stimulated demand for the live experience. For example, Michael Sandel's popular Harvard course, "Justice," is now available in multiple formats, including as a PBS series and in an online version. These formats have not reduced the demand for Sandel's live lectures. To the contrary. Sandel has lectured to tens of thousands of fans in soccer stadiums in Asia largely because he is known there through his earlier digital presence (Cheng 2014).

While online education is here to stay, so are traditional bricks and mortar colleges and universities. Excepting only the Roman Catholic Church, universities are the longest-lived institutions in Western society. Having survived such disruptive innovations as the printing press, radio, and television, we suspect that universities will survive this most recent disruption as well. They will adapt and change in response to this new technology as they have adapted and changed in the past to other pressures.

However, it is possible to envision both dystopian and optimistic possibilities. If technology is used in broad-access institutions to drive cost down without regard to quality, and at the same time is used in elite higher education to further increase the cost and restrict the availability of the "best" education, we will wind up with a society both more unequal and less-productive than it could be. If the new digital technology is used in broad-access institutions to extend education to a wider population and in top-level institutions to reduce the cost and expand the availability of exceptionally good education to more of those who can benefit from it, we can view the future with more optimism. The eventual outcome will be determined not by the irresistible force of technology alone, but also by the exercise of judgment by citizens and by educational, business, and political leaders.

■ *The authors thank Esperanza Johnson of the Spencer Foundation for able research assistance.*

References

Allen, I. Elaine, and Jeff Seaman. 2014. "Grade Change: Tracking Online Education in the United States." For 2013. Babson Survey Research Group; Pearson; Sloan-C.

Allen, I. Elaine, and Jeff Seaman. 2015. "Grade Level: Tracking Online Education in the United States." 2014. Survey of Online Learning. Babson Survey Research Group.

Austin Community College District. N.d. "Telecourse Viewing Options." Webpage accessed August 25, 2015. http://dl.austincc.edu/ITV/.

Bacow, Lawrence S., William G. Bowen, Kevin M. Guthrie, Kelly A. Lack, and Matthew P. Long. 2012. "Barriers to Adoption of Online Learning Systems in US Higher Education." Ithaka S+R, May 1.

Banerjee, Abhijit V., and Esther Duflo. 2014. "(Dis)organization and Success in an Economics MOOC." *American Economic Review* 104(5): 514–18.

Bangs, Elizabeth T. 1994. "MIT Settlement Won't Save Overlap." *Harvard Crimson*, June 9. http://www.thecrimson.com/article/1994/6/9/mit-settlement-wont-save-overlap-pdid/.

Bell, Bradford S., and Jessica E. Federman. 2013. "E-learning in Postsecondary Education." *Future of Children* 23(1): 165–85.

Berland, Elaine Prostak. 1992. "'Up in the Air' Re-considering the Cultural Origins of Broadcasting and the Myth of Entertainment During the 1920s." *American Journalism* 9(2): 54–65.

Bernhard, Meg P. 2014. "Yale Faculty Approves CS50 Venture; Harvard Mum." *Harvard Crimson*, November 9. http://www.thecrimson.com/article/2014/11/9/yale-faculty-approves-cs50.

Bowen, William G. 1992. "Affidavit of William G. Bowen." US District Court for the Eastern District of Pennsylvania. Civil Action No. 91-CV-3274, April 29.

Bowen, William G. 2013. *Higher Education in the Digital Age*. Princeton University Press.

Bowen, William G., Matthew M. Chingos, Kelly A. Lack, and Thomas I. Nygren. 2014. "Interactive Learning Online at Public Universities: Evidence from a Six-Campus Randomized Trial." *Journal of Policy Analysis and Management* 33(1): 94–111.

Bowen, William G., and Eugene Tobin. 2015. *Locus of Authority: The Evolution of Faculty Roles in the Governance of Higher Education*. Princeton University Press.

Brewer, Dominic J., Susan M. Gates, and Charles A. Goldman. 2001. *In Pursuit of Prestige: Strategy and Competition in US Higher Education*. Transaction Publishers.

Cheng, Jonathan. 2014. "Harvard Professor Michael Sandel Adds New Title: Seoulite." *Korea Real Time,* December 5. http://blogs.wsj.com/korearealtime/2014/12/05/harvard-professor-michael-sandel-adds-new-title-seoulite/.

Christensen, Clayton, and Michael Horn. 2011. "Colleges in Crisis: Disruptive Change Comes to American Higher Education." *Harvard Magazine,* July–August, 7(21): 1–4.

Craig, Douglas B. 2000. *Fireside Politics: Radio and Political Culture in the United States, 1920–1940*. Kindle edition. John Hopkins University Press.

Deming, David J., Claudia Goldin, Lawrence F. Katz, and Noam Yuchtman. 2015. "Can Online Learning Bend the Higher Education Cost Curve?" *American Economic Review* 105(5): 496–501.

Department of Education, National Center for Education Statistics. 2012. *IPEDS*. Fall enrollment, Distance education status and level of student [ef2012a_data_stata.csv]. Retrieved from http://nces.ed.gov/ipeds/datacenter/DataFiles.aspx.

Department of Education, National Center for Education Statistics. 2013. *IPEDS*. Fall enrollment, Distance education status and level of student [ef2013a_data_stata.csv]. Retrieved from http://nces.ed.gov/ipeds/datacenter/DataFiles.aspx.

Department of Education, National Center for Education Statistics. 2012. *IPEDS*. Institutional characteristics, Directory [hd2012_data_stata.csv]. Retrieved from http://nces.ed.gov/ipeds/datacenter/DataFiles.aspx.

Department of Education, National Center for Education Statistics. 2013. *IPEDS*. Institutional characteristics, Directory [hd2013_data_stata.csv]. Retrieved from http://nces.ed.gov/ipeds/datacenter/DataFiles.aspx.

Derek Bok Center for Teaching and Learning. 2014. "Blended Learning in Harvard College: A Pilot Study of Four Courses." Harvard University.

Figlio, David N., Mark Rush, and Lu Yin. 2010. "Is It Live or Is It Internet? Experimental Estimates of the Effects of Online Instruction on Student Learning." NBER Working Paper 16089.

Gabriel, Trip. 2010. "Learning in Dorm, Because Class Is on the Web." *New York Times*, November 4.

Gartner. N.d. "Gartner Hype Cycle." http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp (accessed August 25, 2015).

Griffiths, Rebecca, Matthew Chingos, Christine Mulhern, and Richard Spies. 2014. *Interactive Online Learning on Campus: Testing MOOCs and Other Platforms in Hybrid Formats in the University System of Maryland*. Ithaka S+R, July 10.

Hartnett, Kevin. 2013. "San Jose State to Michael Sandel: Keep your MOOC off our Campus." *Boston Globe*, May 3.

Jaschik, Scott. 2015. "Not a Tsunami, But . . ." *Inside Higher Ed*, March 16, 2015. https://www.insidehighered.com/news/2015/03/16/stanford-president-offers-predictions-more-digital-future-higher-education/.

Joyce, Theodore J., Sean Crockett, David A. Jaeger, Onur Altindag, and Stephen D. O'Connell. 2014. "Does Classroom Time Matter? A Randomized Field Experiment of Hybrid and Traditional Lecture Formats in Economics." NBER Working Paper 20006.

Kwak, Do Won, Flavio M. Menezes, and Carl Sherwood. 2015. "Assessing the Impact of Blended Learning on Student Performance." *Economic Record* 91(292): 91–106.

Lack, Kelly A. 2013. "Current Status of Research on Online Learning in Postsecondary Education." Ithaka S+R, March 21.

Means, Barbara, Yukie Toyama, Robert Murphy, Marianne Bakia, and Karla Jones. 2009. "Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies." US Department of Education.

Medical University of South Carolina (MUSC). N.d. "College of Nursing." Webpage. http://academicdepartments.musc.edu/nursing/academics/masters/msn_faq.htm (accessed August 25, 2015).

Nelson, Phillip. 1970. "Information and Consumer Behavior." *Journal of Political Economy* 78(2): 311–29.

Olitsky, Neal H., and Sarah B. Cosgrove. 2014. "The Effect of Blended Courses on Student Learning: Evidence from Introductory Economics Courses." *International Review of Economics Education* 15: 17–31.

Parry, Mark. 2009. "Obama's Great Course Giveaway." *Chronicle of Higher Education*, August 3.

Pearson. N.d. "My Lab™ & Mastering™." A series of interactive online modules for teaching a full range of subjects. Webpage. http://www.pearsonmylabandmastering.com/northamerica/errors/index.html (accessed August 25, 2015).

Penn State World Campus. N.d. "Online Degrees and Certificates." Webpage. http://www.worldcampus.psu.edu/degrees-and-certificates (accessed August 25, 2015).

Schwarzwalder, John C. 1959. "The Promise of Teaching by Educational Television." *College English* 20(4): 180–84.

SRA Reading Laboratory. N.d. Website and reading program. http://www.srareadinglabs.com/ (accessed August 25, 2015).

University of London, International Programmes. N.d. "Timeline." Webpage. http://www.londoninternational.ac.uk/our-global-reputation/our-history/timeline/ (accessed August 25, 2015).

USC Rossier Online. N.d. "Master of Arts in Teaching." Webpage. http://rossieronline.usc.edu/academics/master-of-arts-in-teaching-program/ (accessed August 25, 2015).

Wu, D. Derek. 2015. "Online Learning in Postsecondary Education: A Review of the Empirical Literature (2013–2014)." Ithaka S+R, March 11.

Zemsky, Robert. 2009. *Making Reform Work: The Case for Transforming American Higher Education*. Rutgers University Press.

Journal of Economic Perspectives—Volume 29, Number 4—Fall 2015—Pages 155–176

How Economics Faculty Can Survive (and Perhaps Thrive) in a Brave New Online World[†]

Peter Navarro

The academy in which we toil is moving rapidly towards a greater role for online delivery of higher education, and both fans and skeptics offer strong reasons to believe this technological shock will have substantial disruptive effects on faculty. How can we as economic educators continue to provide sufficient value-added to justify our role in a world where much of what we now do is effectively being automated and commoditized?

In their annual survey of online education trends, Allen and Seaman (2015) cite evidence that over 70 percent of all degree-granting institutions "have some distance offerings" while that number jumps to 95 percent for institutions with 5,000 or more students. Online learning in higher education also seems to be spreading much faster than faculty acceptance. On the one hand, the proportion of American "academic leaders" citing online learning as "critical to their institution's long term strategy" has risen from 49 percent in 2002 to an all-time high of 71 percent. On the other hand, "only 28% of chief academic officers say that their faculty members accept the 'value and legitimacy of online education,' a rate substantially the same as it was in 2003."

At its most extreme, online higher education involves an elegant digital platform capable of seamlessly delivering multimedia content while engaging and evaluating students across the globe at (close to) zero marginal costs. A competition to develop this kind of platform is at the heart of the so-called "MOOC" or "massive open online course" phenomenon. Private sector MOOC leaders include Coursera,

■ *Peter Navarro is Professor of Economics and Public Policy, Paul Merage School of Business, University of California at Irvine, Irvine, California. His email address is navarrop@uci.edu.*

[†] For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.29.4.155

doi=10.1257/jep.29.4.155

edX, Udacity, and Udemy, but there are many other options including even an "iTunesU." On the university side, "free course" leaders range from Carnegie Mellon, Duke, Harvard, and MIT to Stanford, Yale, and University of California campuses including Berkeley, Los Angeles, and Irvine. The availability of high-end online course content provided by an array of educational entrepreneurs is also rising quickly. Access to innovative software applications designed to provide an engaging and interactive learning environment and experience is on the rise, too.

In this brave new online world, many successful and resilient faculty will add value (and differentiate their product) not by producing costly and elaborate multimedia lectures in which they become a superstar professor-celebrity (a model discussed, for example, in Delbanco 2013; Glader 2012), but rather through the careful, clever, and innovative choices that they (and their departments) make regarding both the adoption of the online content of other providers and the forms of online interactions they integrate into their course designs. Possible forms of faculty-to-student and student-to-student interactions run the digital gamut from discussion boards and electronic testing to peer assessments, games and simulations, virtual office hours, and other forms of interactive experiences that may yet be only a gleam in the eye of some entrepreneurial developer.[1]

The degree to which online education may eventually substitute for traditional classroom delivery will likely depend on both the course level and the nature and size of the institution itself. For example, the traditional large lecture hall principles courses may go the way of all flesh. However, intermediate college and graduate-level economics courses may survive and thrive in a "flipped class" format in which online content is used to free up the traditional classroom for higher-level discussions and interactions.

Similarly, large public universities in which a substantial share of students can find it difficult to take the courses that they need to graduate on time may move more quickly to build their online catalogues, in part as a way of addressing their capacity constraints. However, small private institutions may see the "personal classroom touch" as a key differentiator in the market and largely maintain the traditional classroom and/or "blend" online content into partially online "hybrid" courses more judiciously.

This article explores some of the basic descriptive questions economic educators and their administrators are likely to face as the online education tide rises: for example, how much does it cost to develop online content and how much time does it take? This article also tackles some important prescriptive questions as well. For example, what are the key "ingredients" for a pedagogically sound online course? Should you produce your own full service MOOC or adopt the content of others?

[1] For an overview of discussion boards as a collaborative learning tool, see, for example, Curtis and Lawson (2001). On the effectiveness of games and simulations in Kolb's experiential learning cycle relative to traditional teaching methods, see, for example, Herz (1998). The benefits of and technologies to implement "peer assessment" are reviewed in Luxton-Reilly (2009).

Throughout this exploration, I will draw on both the extant literature as well as my own experience at the University of California, Irvine, where the online evolution is advancing rapidly. At UCI, the economics department has already put its entire principles sequence online. Other large enrollment courses now offered in an online format include Academic Writing; both Basic and Classical Physics; Criminology, Law, and Society; Introduction to Business and Management; Preparation for General Chemistry; Pre-Calculus; Philosophy 1; Principles of Operating Systems (in the computer science department); and Principles of Materials Science and Engineering. Within my own unit, the Merage School of Business, we have uploaded and offloaded all the core courses in our undergraduate business minor to the cloud, and online course options are being rapidly developed to service both core and elective courses across our MBA programs.

What should faculty be doing to cost-effectively meet the challenges of the online education shock while best serving not just their own interests but those of their institutions and consumer-students as well? This article will help answer this prescriptive question. While acknowledging there may be sharp, short-run divergences between stakeholders,[2] we focus here on "joint utility maximization." As online education is rapidly becoming the new norm, we assume that most economic educators will have greater professional success over the long run by deftly and skillfully adapting to this technological shock rather than resisting it.

The Global Reach of MOOCs

As a starting point, it may be useful to note the astonishing scale and scope of MOOCs: For example, my "Power of Economics" two-course principles sequence and personal finance course for undergraduates delivered by the Coursera platform has, in less than three years, seen more than 400,000 students enroll from over 190 countries.[3] Like virtually all MOOCs, only a small fraction of those enrolling—about 5 percent—actually complete my MOOC courses. While these low completion numbers are rightly viewed as the Achilles' heel of MOOCs, it is worth remembering that a modest fraction of a very large number can still be a large number. Of the top nine courses listed by lifetime enrollment on Coursera, all exceed 100,000 in signups, while a single Social Psychology course from Wesleyan University is approaching the quarter-million mark (Leek 2013).

[2] The short-run politics are admittedly difficult: the more senior faculty members there are, the more power they are likely to have but the less likely they will be able to adapt because of a skill set developed in a traditional classroom era. For those faculty caught in this "Mr. Chips" world, utility maximization may simply equate to full resistance to the adoption and diffusion of online education in any form. However, as Nobel Laureate Paul Samuelson often quipped in a Keynesian context, "funeral by funeral, science makes progress" (for example, Economics U$A, 2012).

[3] The principles of economics courses for macroeconomics and microeconomics are available at https://www.coursera.org/learn/principles-of-macroeconomics and https://www.coursera.org/learn/principles-of-microeconomics respectively. The personal finance course for undergraduates is at https://www.coursera.org/learn/managingmoney.

This is not to say that the MOOC specifically, or online education in general, necessarily spells the end of the traditional lecture hall. Rather, in thinking about your own choice set, it is important to understand not just how online education works but how a MOOC—or, more generally, online content—might fit into your own strategy of delivering more customized, differentiated, and interactive education to your students.

While you can certainly develop your own online content to compete in the new marketplace, a properly designed MOOC can provide high-quality content that you otherwise might spend vast amounts of time (and money!) developing on your own. (And even if you did spend all that time and money, you could easily end up with an inferior product.) But how will the availability of such online content affect your classes?

The Death of the Large Lecture Hall Is Not Greatly Exaggerated

The large lecture hall classes often used to teach basic principles courses across many disciplines is the format most threatened by this new technology. In the existing "live theater" large lecture hall paradigm, students must assemble in synchronous fashion at assigned geographic coordinates, where they sit passively to receive content from a live body at a lectern. In the new online "movie" paradigm, students can receive content asynchronously almost anywhere on the planet, while the live body no longer has to perform the same content repeatedly. Rather, some variation on this performance simply has to be captured and "bottled" in some type of video, audio, or multimedia format.

Of course, the live-theater paradigm has its advantages: for example, the excitement of a live performance and a focal point in time and space for students to meet in-person. But over time, technological advances will allow the bottled performance to become ever more high-quality, well-presented, flexible, and interactive. While it takes neither a rocket scientist nor an economist to predict which side is likely going to win,[4] it is worth considering where this battle may best be fought. One can choose a position at one of three main locations along the online education continuum: the flipped, hybrid, and fully online classrooms.

At one end of the online education continuum, there is the "flipped classroom." In this model, the physical classroom remains a staple of the academy. However, students are tasked with mastering some or all of the basics *outside* the classroom while classroom time is reserved for an interactive discussion of higher-order concepts, ideas, and applications based on the principles.

[4] In Navarro (forthcoming), I offer a more in-depth treatment of the relevance of the large lecture hall in today's e-learning world. One interesting piece of survey data indicates that 37 percent of the respondents from my online undergraduate class in personal finance said that "at least half" of the large lecture hall courses at UCI "could be delivered in an online format," while 33 percent answered "most" and 11 percent said "all."

Online content can be particularly valuable for delivering such basics;[5] and the benefits of this flipped model have been extensively studied. They include: students take more responsibility and are more active learners, and they cultivate a deeper understanding of the content and how to apply it (for example, see Zappe, Leicht, Messner, Litzinger, and Lee 2009; Bergman and Sams 2012; Berrett 2012).

I have used this flipped model in conjunction with my own online and digital content (in the early years, using CD-ROMs) to teach basic economics going back almost 20 years, and I've seen these benefits first hand. While MOOC courses are well-suited for delivery in a totally online format with virtually no personal touch, my executive MBA students who have gathered together in a 60-seat amphitheater expect a more complex and personalized experience for their tuition dollars—and that's one key area where the flipped approach excels.

At the other end of the online content continuum, there is the fully online course. While non-online assets such as textbooks and printed articles may continue to play a role in this world, the fully online course is an immersive digital experience for students—and an uncomfortable environment for at least some "Mr. Chips" faculty weaned on live performance education and reliant on the cues, customs, and feedback of the traditional classroom.

In between the flipped and fully online classrooms, there is the partially online *hybrid* model, also called the *blended* model. It substitutes online content for some substantial fraction of the time spent in traditional classroom meetings, while the remainder of the time occurs in the classroom. Together with the flipped model, this hybrid/blended model may offer the most promising 21st century combination of high-quality, personalized, and noncommoditized education in a form that both usefully leverages online content and helps to justify one's salary and the continued existence of the broader physical academy. The hybrid and flipped models are not mutually exclusive. For example, half of your classes in a course might be fully online while the other half might be delivered in a flipped, rather than traditional, classroom format.

The research on this point is encouraging, particularly for the integration of MOOCs into various blended curricula. For example, in the Griffiths, Chingos, Mulhern, and Spies (2014) study of the University of Maryland system that incorporated MOOC content into the delivery of hybrid courses, side-by-side comparisons found that: "Students in the hybrid sections . . . did as well or slightly better than students in the traditional sections in terms of pass rates and learning assessments, a finding that held across disciplines and subgroups of students."

[5] It may be useful to distinguish here between the "online education" model of today and the "distance education" model that has been around since the days of correspondence courses. In the survey mentioned at the start of this paper, Allen and Seaman (2015) equate "distance learning" with "online learning," but this need not be the case. While distance education can certainly be delivered in an online format, it can also be delivered in purely analog fashion. In contrast, online education is delivered primarily in a digital framework, but may still feature analog elements like the printed textbook.

We are experimenting heavily with this hybrid/blended model within my own business school to leverage the skill sets of our teaching corps while offering students more scheduling flexibility. The goal is to continue to personalize the MBA experience and convey the complexities of the MBA curriculum in at least a partially live theater format and, thereby—in all frankness and in the very real world of profit and loss—at least partially justify the high cost of an MBA degree.

We are far from the first to pursue this hybrid/blended model. For example, Kolowich (2015) reports that 2U, a company that helps universities develop online and blended degree programs, has already assisted clients ranging from Georgetown and George Washington University to Simmons and Yale. Other major players in this "enabler market" include Bisk Education and Pearson Embanet. A typical arrangement usually involves substantial revenue sharing (often 50 percent or more) with the partnering university. For example, Bisk receives "80 percent of the gross revenue from the three-course, online certificate program in executive education the company runs with the University of Florida. The contract, which was signed in 2012, projects that after five years the program will have made about $1.6-million for the university and $6.3-million for Bisk" (Kolowich 2014). According to Howard (2014): "There are now in excess of 2,000 online degree programs in the U.S. About half of the schools rely on a third-party facilitator . . . [F]acilitators currently bring in an estimated $1 billion a year in tuition revenue. That market is expected to double in four years . . ."

The "Net Generation" Bids Adieu to Mr. Chips

In my view, both the flipped and hybrid class formats leveraging MOOC and other online content are likely to be relatively more useful than fully online courses as one moves up the ladder of higher education, both from lower to upper division courses and from the undergraduate to graduate level. One reason supporting such speculation may be found in surveys of American and British students that have found a high degree of boredom with traditional class lectures. On the American front, as reported in Berk (2009), a national survey conducted by UCLA's Higher Education Research Institute of nearly 250,000 college freshman at more than 500 colleges and universities (Pryor et al. 2008) found that over 40 percent of the students report "they are frequently bored in class. As Berk (2009) reports on the British experience, of the 211 students surveyed "59% found lectures boring in at least half of their classes and 30% find most or all of their lectures boring" while "the use of PowerPoint® slides was the most important factor contributing to boredom." Most interestingly from a hybrid and flipped class perspective, "the least boring methods were found in seminars, practical sessions, and group discussions, where students could interact and actively participate."

One underlying cause of this pervasive boredom with the traditional lecture hall appears be a generation gap between older vintage "sage on the stage" economic educators (King 1993) and a new Internet generation now streaming into our

classrooms with very different expectations (and perhaps shorter attention spans). In a survey of more than 7,000 US college students, Junco and Mastrodicasa (2007) profile a "net gen" world where computers, cell phones, web searches, instant messaging, blogging, music downloading, and multitasking are second nature, and net gens themselves find some of their deepest connections in the virtual rather than physical world.

In serving this net gen market, Berk (2009) argues that the traditional, talking-head broadcaster and transmitter of information must morph, as a matter of both sound pedagogy and survival, into group facilitator and orchestrator of collaborative knowledge creation (Brown and Adler 2008), thus taking the role of a "guide on the side,'" with a concomitant shift from "'teacher-centered' to 'learner-centered' teaching methods." (DeAngelo et al. 2009)

In fact, the online education world provides abundant opportunities to hit all points of this net gen compass while avoiding the Achilles heel of the traditional classroom—understimulation. At the anecdotal level, I have heard more than one of my sage-on-the-stage, PowerPoint-dependent colleagues complain (sometimes bitterly) that while they are teaching as well as they ever have, their teaching evaluation scores continue to fall. There appears to be little understanding of the increasing mismatch between their own skill sets and the net gen consumer-students they increasingly serve.

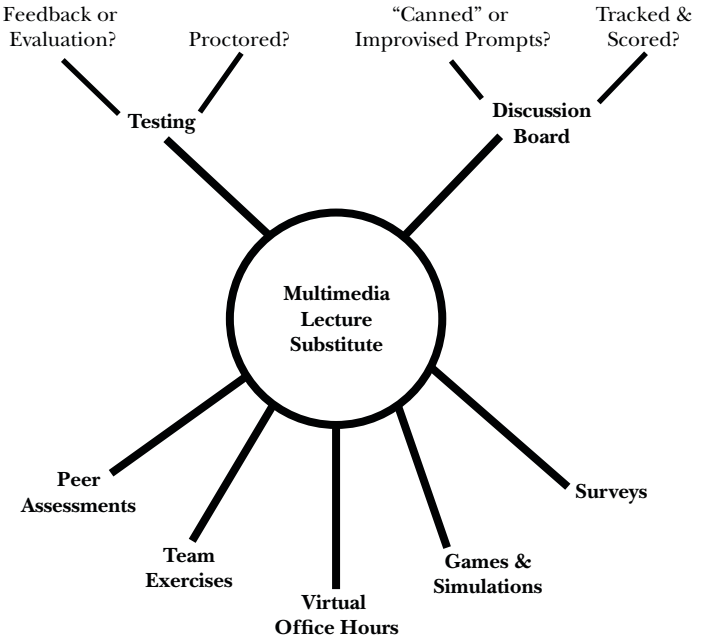
From a labor market perspective, there is also this observation, which follows from the discussion in this journal by Autor (2015): online education technologies will both substitute labor *and* complement labor. For example, while MOOCs may spell doom for some type of teaching like traditional lectures that cover the basics of a discipline, a shift to more hybrid courses might increase the demand for other types of teaching, like personalized in-class discussions of examples and applications. While the overall effect on labor demand is unclear, there certainly will be distributional consequences, with winners and losers among educators depending on their skills, willingness to adapt, and ability to innovate.

What's in Your MOOC?

Let's suppose now you want to create your own online course. How should you go about it—and what will it cost in both human and financial capital? There certainly is no shortage of books, articles, and manuals offering tips on proper design and delivery: for example, see Clark and Mayer (2011) for a comprehensive "how to" book. Figure 1 provides an illustrative overview of the various elements you may choose to include in your particular production.

At the center of this online course blueprint is some form of multimedia substitute for the traditional lecture. In addition, the top half of the figure lists both electronic testing and a discussion board as essential elements, while the bottom half of the figure adds nonessential "bells and whistles" that may be used to improve pedagogy, to help differentiate one's teaching product, and

Figure 1
Illustrative Elements of a Pedagogically Sound MOOC



Source: Author.

thereby to add further value to the online experience.[6] Each of these elements in the figure is examined now in more detail starting with the critical multimedia lecture substitute.

A Continuum of Production Value Complexity

The multimedia lecture substitute can involve a continuum of complexity, defined by the range of the production and associated production values.

At the simplest end, a single camera records your lectures, which are memorialized as videos on the web. The next step past the talking head video is to superimpose a set of your lecture slides over all (or some portion) of your recorded lectures. To maintain a personal touch, you can have your talking head video appear in a little box super-imposed on the slides. Alternatively, you may choose to by-pass the recorded video lecture altogether and go right to a "voiceover" of your lecture notes, which serves as the narrative element for the sequence of the slides you might ordinarily use in a lecture. Here, the slides from, say, your

[6] For a more skeptical view than the one offered here, some of the main research studies calling into question the efficacy of online education include Carpenter, Brown, and Hickman (2004), Zavarella (2008), and Xu and Jaggars (2011).

PowerPoint or Keynote lecture presentation can be easily converted into either a video format or a set of images that can then be added to the timeline in the video editing process.

For additional levels of complexity, one can add various forms of so-called "b-roll" to the video editing process. Such b-roll ranges from simple photos and images and video clips to more complex graphics and full-blown animations. One can also add visual effects. For example, "green screen" technology allows a video editor to replace a green background with still images or video using a process called "Chroma keying." While this technology used to be the exclusive preserve of weather reporting on TV, today the cost of high-quality Chroma keying has gone from the tens of thousands of dollars to virtually zero. Other editing effects include transitions like the spinning cube or cross dissolve between clips, various sound effects, and perhaps even a musical score.

Based on my own experience and feedback from students,[7] I believe both administrators and educators should strive to include as much complexity in their online content as budget and time allow. I have regularly conducted exit surveys for my online courses at UCI; and one result has been consistent across both undergraduate and MBA populations: my high production value multimedia lecture content delivered through the Coursera MOOC platform is regularly rated as the most essential instructional technology of the class.

A Sole or Joint Venture?

If you decide to develop some form of multimedia lecture substitute for use in a flipped, hybrid, or fully online format (rather than adopting an already-existing MOOC or other online content), you face a practical choice: Should you produce the content as a do-it-yourself "stand-alone venture" in which you are the sole developer, or as a "joint venture" in which you partner with, or hire, a team of developers?

The do-it-yourself model is the most daunting. To achieve high production values will require a significant capital investment in the necessary equipment and software. In my case, I have more than $15,000 invested in a studio that features state-of-the-art video cameras and audio recorders; top-shelf software for animation, graphics, and editing; a professional grade lighting setup, and a green screen background that accommodates "full body shots" rather than the more typical waist up look. ("Full body shots" allows an editor to place the subject in any environment in creative ways that avoid the trap of the "talking head" presentation.)

Beyond this financial outlay, you will also have to invest in your own human capital if you truly want to "do it yourself." The requisite skills range from learning how to perform well on camera and executing compelling voiceovers to mastering

[7] For example, in my inaugural core curriculum class delivered online in Summer 2014, 86 percent rated the Coursera lectures and quizzes as "absolutely essential" (53 percent) or "essential" (33 percent) while an additional 10 percent found the content "useful" or "very useful." In my standard "value proposition" question, a supermajority of my MBA students also have rated my online macro class as more or much more valuable than the average value of the other core classes they had taken; and much of the credit may be assigned to the multimedia experience students are exposed to.

basic camerawork, sound and video editing, and perhaps producing your own advanced graphics and animation. Most of you will conclude that this was not what you got your PhD or Master's degree to do and either abandon the idea entirely or quickly opt for the "joint development" option.

With joint development, you must find some facilitating entity to assist you. This can be a private sector facilitator like the aforementioned 2U or Bisk. However, at least based on some of the revenue-sharing "deals" in the public record to date (as documented earlier), the cost of using a private sector provider is likely to be higher relative to another possibility—an existing unit on your own campus to assist faculty with the development of online and hybrid courses.

To provide an idea of what joint venturing with such an in-house production team entails, it may be useful to summarize a few highlights from a document distributed by the UCI Distance Learning Center (2014) on my own campus. According to this analysis, "faculty members can expect to spend between 60 and 120 hours to prepare existing content for online delivery" and "an additional 80 hours" to develop a new fully online course. The steps involved include: selecting the appropriate instructional technologies; adjusting existing or creating new content suitable for online delivery; preparing supplementary materials, notes, references, articles, and other resources for posting online; creating discussion forum prompts; and developing the various learning assessment tools.

To assist in these tasks, the Distance Learning Center assigns each faculty member an instructional designer, a videographer, a video editor, an animator and/or graphics artist, and a web designer. The price tag for a single one-quarter, four-credit course ranges from $4,000 to $12,000 (and does not include faculty labor costs), with the final price depending on the level of service sought and the production values of the multimedia content itself.

Production values can range from the simple "turn the camera on and talk" approach to highly sophisticated lessons featuring an abundance of graphics and animations along with sophisticated editorial cuts and sound effects. As you add complexity, you quickly (and exponentially) add costs. For example, video editing and animation services are billed out at $50 an hour by UCI's Distance Learning Center, and both graphics and animations are highly labor-intensive. Similarly, the costs of "b-roll" to cover your own talking head entail not just labor search costs to find suitable images and video clips, but often royalty payments for use of the content as well.

Ten (More) Commandments

How should your online content look and feel? Guo, Kim, and Rubin (2014) offer some insights. In a large study involving 6.9 million views of 862 lectures by almost 128,000 students enrolled in the edX MOOC platform, they found evidence that: "Videos produced with a more personal feel could be more engaging." They also found that students were more engaged by shorter videos and by a mixture of talking heads and graphics rather than either alone. This lesson also held true for high-quality recorded classroom lectures edited into shorter segments.

Based on such research and my own experience in the trenches—and always keeping in mind that today's net gen "consumer-student" is accustomed to fancy and punchy video—I offer "Ten Commandments" of developing engaging and pedagogically sound online content.

1: Do NOT put yourself on camera for any extended period of time. Today's students abhor "talking heads" like nature abhors a vacuum. Even if you are a better-than-average classroom lecturer, what works in the classroom doesn't necessarily work online.

2: DO stress high production values. Online platforms for distributing content are agnostic about quality. They can deliver a low-resolution PowerPoint slide show choreographed to scratchy audio—"garbage in, garbage out"—just as easily as they can deliver high-definition, surround sound multimedia presentations with 3D animations. If you are to compete successfully with other content providers—and thereby successfully compete for the attention of your students—you must strive for the highest-end production values your budget allows. This means a well-lit set, full high-resolution video, rich sound, crisp graphics, and high-end animation. It is pointless to hope that good content will overcome poor production values—which is perhaps a sound reason to consider adopting the content of others rather than producing your own.

3: DO break each of the presentations up into short modules. A good guideline is that such modules should be 3 to 7 minutes, and never exceed that limit. Such modules appeal to the time constraints facing today's students and also to the customs and styles of their interconnected social media world. From an instructor's point of view, organizing your material in this way also helps you better orchestrate the flow of the content, even as it forces you to think about the right breaks in the material and find the appropriate segues.

4: DO include interactive elements in the presentation. For example, I include a low-tech "question wizard" in my own presentations. Specifically, at key interactive nodes, I instruct students to pause in viewing the material and take time to provide a possible response to a question posed or task assigned—for example, illustrating how a rise in a factor price will cause a supply curve to shift. Higher-tech alternatives continue to emerge that allow you to embed quizzes in your videos or to add interactive questions.[8]

5: DO make your multimedia presentation content "timeless." Focus on economic principles that are consistent over time and minimize the use of material with dates on it that is not of an historical nature. Handle current events within the context of other modalities like the discussion board. It will be costly and time-consuming for you and any staff to update and reload multimedia presentations that will otherwise

[8] Entities like Camtasia, Google, and TedEd appear to be fruitfully pushing the envelope on this. Other interesting webtools include HapYak, Educanon, Blubbr, and Vialogues: for example, HapYak helps you add links and text to your videos while Blubbr helps you create quizzes around YouTube videos. For more information, visit http://www.educatorstechnology.com/2014/02/8-good-web-tools-to-create-video.html.

quickly become dated. That said, there are some workaround possibilities here. For example, you might reference the latest data on the unemployment rates in Europe and the United States in your voiceover and then periodically update only the actual graphic in your video presentation. This can easily be done if you are producing your own "mini-MOOC" and post your videos online on YouTube or Vimeo for your classes. Another alternative is to include a sprinkling of more contemporary modules that you redo each year. In a MOOC format, you may wish to manage this periodically updated content yourself on an external site like YouTube or Vimeo, and then just list links to these modules on the course page; this avoids burdening MOOC administrators with the uploading of the revised videos.

6: DO proof everything very carefully. It's one thing to screw up a calculation on a PowerPoint slide in front of a few hundred students. It's quite another to screw up a multimedia presentation with 80,000 eyeballs on it. Here, the pedagogical literature strongly suggests that technical or other problems experienced by students will translate into lower levels of satisfaction (for example, Carswell, Thomas, Petre, Price, and Richards 2000; Carswell and Ventaktesh 2002). As an example, I mentioned earlier that I codeveloped an online course on personal finance for undergraduates in 2014 that debuted to over 40,000 Coursera enrollees and close to 500 UCI undergrads. In a key module, the voiceover did not match the graphic example, which resulted in a flurry of negative email and activity on the discussion board. Fixing the problem required re-recording the audio, re-editing the video, outputting a new video file, and replacing the flawed video with the fixed video on the Coursera platform—hours of work to fix a 20-second error. As they say in carpentry: "Measure twice, cut once."

7: DO pay careful attention to copyright issues. A common tendency these days is to rip off images and video from the web with little regard for intellectual property. This practice is risky, particularly when done by faculty who are affiliated with universities that can be sued. Therefore, rely as heavily as you can on images that are free to the commons—Wikipedia is often a useful source—and cite everything you use. You can also use the advanced search engine on Google to find royalty-free images; but if you truly want higher-end imagery, build it into your budget and use a service like Getty Images to help you search for and acquire such imagery.

8: Do NOT wing it. I always write scripts for everything I record either on camera or as voiceover—umms, ahs, and awkward pauses or rambling threads just don't cut it with today's discerning students. For video camera work, I use a teleprompter software called "Presentation Prompter" loaded onto my laptop computer, and I place the laptop strategically beneath the lens so it appears like I'm looking right into the camera. (The teleprompters of old were clunky contraptions costing $10,000 or more, while teleprompter software on a laptop can be found for $100 or less (and even for free with limited options). The trick is to write your copy as if you were speaking it—after all, it *is* a lecture—and then read that copy as if you *are* speaking it. If a newscaster can do it smoothly, so can you with a little practice—so warm up your voice and do one or more practice runs before hitting the "record" button.

9: DO have accurate scripts or transcriptions of your presentations and use them for captioning your course. Captioning is important both for the hearing-impaired and for students for whom English is not their first language. It is also possible to translate the captions, so that students can see subtitles in their preferred language. In a "tip sheet" for foreign students, the University of Canberra (2015) acknowledges a common set of problems: "The lecturer speaks too fast," "Different lecturers have different pronunciation," "Some lecturers use strange idioms," and "I don't know enough vocabulary." The advice includes that students seek "recorded lectures."

10: DO keep your user interface as simple as possible. I have not strictly followed this rule. The instructional environment has become technologically rich, and so I have used multiple platforms to build my online courses. While the requirement that students engage in multiple logins on different platforms has led to some student complaints in my exit surveys, I have thought it to be a worthwhile tradeoff. That said, seek for yourself, and perhaps you will find better ways to integrate competing and complementary technologies and content into a single seamless platform for your course.

The Essential Elements of Testing and Discussion Boards

Let's turn now from the multimedia lecture substitute and consider two of the most essential elements of a hybrid or fully online course: testing and the use of discussion boards (also known as online forums). With the testing element, it is useful in addition to distinguish in an online world between the use of testing for students' own feedback and the need by instructors to evaluate student performance.

By way of example, the approach Coursera takes with my MOOC courses, which I believe is a good one, is to allow students to take the quizzes attached to the lessons multiple times. In this way, students are allowed to test their retention and mastery of the material. Evaluation and grading can then be achieved through proctored exams like a mid-term and a final.

This distinction between unproctored "feedback quizzes" and proctored evaluation exams is important. In an asynchronous online environment, cheating is often much easier. For example, online quizzes taken asynchronously can be "cut and pasted" into a document or captured as images with screen shot technology, and this information can then be shared with other test takers—Student A takes a quiz on Monday and emails it to Student B who takes it later in the week. It may even be that Student A is simply a stalking horse for other students, and Student A has no intention of completing the class.

The emergent industry of digitally proctored exams can address this academic integrity issue to some extent. Competitors include Proctor Free, Software Secure, and, the one that I have relied on, ProctorU. With ProctorU, students must have a computer with sound and a webcam. Each student registers in advance for the exam in question and ProctorU verifies the student's identity. On "game day," the student logs in via ProctorU. At that point, a live human proctor watches both the student (via the student's own webcam) and the student's computer display during the entire exam. This makes it much harder for a student to use cheat sheets, get

assistance from friends, or otherwise use forbidden test aids. I have used this tool for students to log in for my final exams from all over the world. The primary downside is that online proctoring entails an additional student cost—around $15 for a one-hour exam to about $30 for a three-hour exam. Another negative, according to my student exit surveys, is that a small but vocal minority find the process of being watched over by a nanny-cam to be Orwellian. (And speaking of Orwell, students of mine who ventured to China during the exam period couldn't get through China's firewall to reach ProctorU.)

Another approach to reducing cheating on midterm and final exams is to specify a very tight window of time during which the exam can be taken. But of course, that decision defeats at least some of the advantage of the flexibility of the online course. Like fighting ants and roaches, dealing with cheaters is a never-ending battle; and the need for effective one-on-one or in-class monitoring of cheating will likely continue as a serious limitation on taking degree-granting, for-credit MOOCs to large scale.

Discussion Boards and Online Forums

When class sizes are large, classroom discussion becomes logistically difficult and faculty office hours can't hope to allow meetings with more than a small fraction of the enrolled students. Perhaps for this reason, the electronic discussion board has become a best practice and staple of online courses (and many traditional classroom courses as well). Suler (2004) and TeacherStream LLC (2009), for example, offer an overview of the various methods to facilitate discussion board interactions.

A typical online board will feature a "prompt" from the professor often tied to some news article, assigned reading, or portion of a lecture. Student "discussants" can respond both to the prompt itself and to other responses by students in a system of discussion threads with possible feedback from the instructor. More sophisticated boards can introduce a "points system" whereby posts are graded by one's peers and/or the instructor. In these ways, discussion boards provide peer-to-peer learning opportunities along with personalized instructor-to-student interactions in a collaborative learning environment.[9]

Your choice of possible discussion prompts spans the spectrum from prepackaged "canned prompts" synchronized to the course material to "improvised prompts" tied to the current news of the day that feed back into the course material. For example, a typical canned generic prompt in a macroeconomics class might be "what factors move exchange rates?" Alternatively, for an improvised prompt, you might post a current events question (with a related article link) like "how is today's announcement by the European Central Bank to cut interest rates likely to affect the value of the euro?"

The virtue of canned prompts is that they can be automated by the MOOC platform into the delivery of the course and monitored by teaching assistants given

[9] Moore (1989) formalized the key types of interaction useful in a learning context. For a literature survey of the virtues of various forms of interaction, see Su, Bonk, Magjuka, Liu, and Lee (2005). The *Journal of Interactive Learning* publishes a steady stream of research in this area.

a "cheat sheet" of likely responses, with the instructor only participating when a student question has stumped the staff. The advantage of the improvised prompt is that it weaves immediate real world applications into the fabric of economic theory. Either way, it is essential to monitor any discussion board; anything from irrelevance and irreverence to profanity and pontificating can infect the discussion. In addition, if you want students to pay attention to the board, you need to provide appropriate incentives. Here, a points system may be helpful and can be implemented with higher-end software, for example, at the time of this writing I am experimenting with Yellowdig.

More Value-Added For Your Online Course: Peer Assessment and Team Exercises

Let's turn briefly now to the nonessential elements that may be included in an online course as a way of adding further value and differentiation to your teaching product. For example, both the peer assessment and team exercise can be used to enhance student-to-student interactions and help overcome some of the alienation and social isolation that can occur in an online learning format (for example, Rovai and Wighting 2005).

With peer assessment, increasingly sophisticated software tools such as Blackboard, Moodle, and WebPA allow you to subdivide a class into microcells and to have these cells internally evaluate and perhaps grade assignments. Topping (1998, 2009, 2010) offers an interesting look at the development of peer assessment theory and practice. In their technology review, Honeychurch, Barr, Brown, and Hamer (2012) find that peer assessment not only helps overcome the hurdle of having an instructor grade large numbers of assignments; they argue the "real value . . . resides not in the feedback itself (the product) but in the process of constructing the feedback."

As an example, suppose you want to assign a paper or YouTube-style video presentation[10] on quantitative easing (a macro application) or a game-theoretic analysis of competition in a specific industry of the student's choosing (a micro application). You divide your class into microcells of five students per cell. Within each cell, each student reads and comments upon the papers or presentations of the others—and perhaps assigns a grade. In the process, as Luxton-Reilly (2009) notes, "a high degree of individualized feedback for students can be maintained by engaging them in tasks that promote learning by interacting with each other." Watters (2012) discusses the constraints on using peer assessment in a MOOC world of very large numbers. Alternatively, in a team exercise format, teams can be tasked with the assignment rather than individual students.

In my own case, I regularly subdivide my classes into small teams and encourage them to develop their own support groups for the class. I also offer students the opportunity to meet in virtual space with both me and with their teams using conferencing softwares like Google Hangout and VirBela. For example, VirBela

[10] In an online world, the traditional "paper" is sometimes giving way to more multimedia productions. Such productions still require students to write—in this case a script—but also help develop other soft presentation skills increasingly valued in today's labor market.

offers a complete virtual environment that allows students and instructors to meet as "avatars" in a virtual lecture hall or team room and communicate verbally and/or through messaging. In this way, softwares such as VirBela offer a high-tech alternative to the traditional office hours and help mitigate one of the major downsides of online learning—the loss of the traditional face-to-face sit-down with an instructor or teaching assistant.[11]

As for simulations and games, entire journals are dedicated to their applications and uses in an experiential learning environment: for example, *Simulation & Gaming* and the *International Journal of Gaming and Computer-Mediated Simulations*. Such tools allow students to come at the material from a different direction, learn the material with a more hands-on perspective, and perhaps introduce a competitive element into the pedagogical equation.[12] I use a stock market simulation via a software called Stocktrak to inform the complex relationships that may exist between stock prices and macroeconomic events and allocate points on the basis of contest winners and losers.

The broader point here is that technologies continue to evolve that help you better "touch" students in ways that could hitherto only be done in the traditional classroom or an office hours setting.

The University as "Company Town"

As an additional consideration regarding whether you want to develop your online content personally, there is the critical issue of the assignment of property rights. The assignment of these rights can have significant impact on your own welfare—as well as on the pace of online content development at your institution.

In one scenario, you develop your own content on your own time and with your own resources, you own the intellectual property rights, and the only person who can use your content without your permission (and possible compensation) is yourself. In the opposite scenario, you develop online content, but the university owns the property rights by contract because you are in its employ. There is a debate on the University of California, Irvine campus, and I'm sure on other campuses as well, as to whether development of course materials has now become part of one's job description, such that no additional incentives should be required. One can imagine a "no good deed goes unpunished" scenario in which other faculty and lecturers free ride on your effort, and you wind up with fewer classes to teach—or even possibly without a job.

In this latter scenario, the overarching problem is that the faculty labor market may be prone to monopsony power from universities (for discussion, see Ransom

[11] In the early development of online education, this loss was at least partially offset by the obligatory string of email correspondences between students and professors as well as the occasional phone call. However, instructors can quickly get buried in an avalanche of email from students all asking the same or similar questions. In this scenario, email is simply a time suck rather than an effective teaching tool, and discussion boards go a long way towards solving this problem.

[12] For the use of games and simulations in economics education, see, for example, Greenlaw (1999), Gremmen and Potters (1997), Lean, Moizer, and Towler (2006), and Schmidt (2003).

1993; Hallock 1995). In the presence of such monopsony power, universities are more likely to impose conditions on new hires that reflect the emergence of online education options. Within my own business school, for example, all new ladder faculty sign contracts committing them to possible online teaching—and by implication developing online content on behalf of the campus.

If you are going to expend the time, resources, and energy to produce substantial online content or even a full-fledged MOOC, know the rules in advance. This is not just about being selfish; the unchecked exercise of monopsony power by the academy will almost certainly lead to an undersupply of quality online content. Campus administrators who see the benefit in having their faculty involved in generating high-quality online content should therefore tread lightly, rather than trample, the intellectual property rights of their academic cadres.

On-Demand versus Phased Delivery

The advent of online education affects not just the space in which we teach (physical or virtual) and classroom times (synchronous or asynchronous). It also opens up the traditional quarter/semester structure of content delivery to disruption. Should online courses be offered in a phased delivery/cohort format that mimics the traditional classroom model and calendar? Or alternatively, should the academy more heavily emphasize "on-demand" education—with all the implications for "college life" as we know it?

As an example of the most regimented form of the phased delivery/cohort model, I originally offered my economics principles sequence on the Coursera MOOC platform in a ten-week quarter format. Each week, a new topic would be introduced (for example, production theory, the Keynesian model); and the student cohort had to complete the multimedia presentation and pass the requisite quiz in a timely manner. In addition, the discussion board was refreshed with new prompts tied to the weekly topic. Students could go back and review previous material, but they could not jump ahead.

The MOOC community quickly let me—and Coursera!—know that this regimented delivery is anchored in an old school traditional classroom mentality. Effectively, the cohort model chains a powerful digital tool to the artificial bounds of time in the same way that for decades, network television remained bound to the formula of releasing one episode a week of a prime-time series. After cascades of student complaints, Coursera decided to experiment with the on-demand format, with me as one of their first guinea pigs. With this approach, eager beavers can now "binge" their way through my Coursera courses as fast as Netflix users have gone through a season of *House of Cards*. At the other end of this on-demand spectrum, slow pokes can turn a normal ten-week race into a six-month marathon—and thereby better avoid contributing to the high drop-out rate symptomatic of MOOCs.

While many students appreciate the flexibility of an on-demand model, this structural shift also creates significant pedagogical problems. One of the biggest downsides of an on-demand approach is the loss of synchronicity for student-to-student and professor-to-student interactions. In the regimented cohort

format, Week Three of the macro principles course might feature fiscal policy, while Week Four could feature monetary policy. During those weeks, all students as a cohort and community could be presented with discussion prompts geared to the topic at hand. Once the on-demand approach kicks in, the class discussion boards and virtual office hours lose their sharp focus, and the courses devolve, at least to a certain extent, into mastery of basic timeless content. Another downside is that "binge learning" may lead to lower longer-term retention of the material—a prime topic for future research as this topic is now currently an empirical black hole.

On the plus side, an undervalued virtue of the on-demand model may well be its ability to move more highly motivated students (perhaps with greater aptitude for the topic) more quickly through the curriculum. This on-demand virtue may play particularly well at many public universities where certain key courses have had enrollment limitations that can make it difficult for students to fulfill both major and graduation requirements in a timely manner. For this reason alone, the on-demand format may (largely) win out, warts and all, at least for a number of core courses and particularly for public universities struggling to graduate their students on time.

Some Final Career Advice

How can an economics educator continue to provide significant value (and thereby stay gainfully employed) in a world where online education is automating and commoditizing many of the tasks that have traditionally been part of the job? How you proceed from this point in time will depend in large part on your current stage of career and position. While I am certainly not tooled up to be anyone's career counselor, here is how I see the chessboard.

If you are primarily a research economist with some teaching responsibilities in the mid to later stage of your career at a higher-ranked research institution, don't bother with developing online content on your own. The most you may want to do is explore flipping or blending your traditional classroom by leveraging already existing online content either from a MOOC or elsewhere. By following this seemingly modest path, your students will get a better education. At the same time, you may find teaching to be a fresh pleasure because you are no longer spending so much time explicating basic material and instead will have more time to share actual theoretical and empirical complexities with your students (as well as perhaps more of your own research).

If you are a research economist much earlier in your career, you should, at least as a hedging strategy, learn more about technology-mediated education, and perhaps build some competencies in this area. If you fail to do that, your hope for academic survival will rely on prodigious research output in top tier journals—an option that only works for a chosen few.

If you are primarily an economic educator with less demand that you publish, it is essential that you make the transition to the virtual world and take to both heart

and head all of the online ways to add value to the classroom and differentiate your product. This will require a degree of retooling similar to what America's blue-collar manufacturing workers have experienced in adapting to a world of automation and robots, but it will likely be a very worthwhile investment as the live performance "sage on the stage" steadily gives way to the "cyber-guide on the side."

If you are a department chair or a dean and want to meet the competitive challenge posed by online education, you must set up appropriate incentives for online course development. This likely means the carrot of proper funding and reduced teaching loads in the development stage (and perhaps any gentle sticks at your disposal). Particularly at research institutions, it means providing your faculty with the necessary assurances that contributions in helping an institution move to an online and/or flipped class world will carry significant weight when it comes time for merit, tenure, and promotion decisions.

In addition, department chairs and deans hoping to innovate rapidly should consider moving beyond the "volunteer model" of offering incentives and hoping for a response and instead move to a "recruitment model." We have learned this lesson painfully at my business school: if you wait for volunteers, the response will likely be tepid, the faculty best suited for the task may not respond, and you may wind up with more online clunkers than hits. Instead of waiting for volunteers, identify the faculty you think will do the best job helping you move into the online space and vigorously woo and recruit them.

Finally, for chancellors and other central administrators, my perhaps counterintuitive advice is not to get caught up in the hype over online education. Assess your institution's skills and build your program to amplify its strengths. For example, if you are a small liberal arts college or an elite MBA school, posting everything online may diminish the competitive and comparative advantages you might now have in the marketplace. In such cases, follow the blended and flipped routes and leverage your strengths. Perhaps most importantly, as your faculty develop online content, be zealous about quality control. Don't let the current frenzy to produce online content result in low-quality offerings that diminish rather than build your reputational capital over time.

References

Allen, Elaine I., and Jeff Seaman. 2015. *Grade Level: Tracking Online Education in the United States.* Babson Survey Research Group. http://www.onlinelearningsurvey.com/reports/gradelevel.pdf.

Autor, David H. 2015. "Why Are There Still So Many Jobs? The History and Future of Workplace Automation." *Journal of Economic Perspectives* 29(3): 3–30.

Bergmann, Jonathan, and Aaron Sams. 2012. *Flip Your Classroom: Reach Every Student in Every Class Every Day.* Washington, DC: International Society for Technology in Education.

Berk, Ronald. 2009. "Teaching Strategies for the Net Generation." *Teaching and Learning Journal* 3(2): 1–23.

Berrett, Dan. 2012. "How 'Flipping' the Classroom Can Improve the Traditional Lecture." *Chronicle of Higher Education*, February 19.

Bouhnik, Dan, and Tali Marcus. 2005. "Interaction in Distance-Learning Courses." *Journal of the American Society for Information Science and Technology* 57(3): 299–305.

Brown, John Seely, and Richard P. Adler. 2008. "Minds on Fire: Open Education, the Long Tail, and Learning 2.0." *EDUCAUSE Review* 43(1): 16–32.

Carpenter, Trudy G., William L. Brown, and Randall C. Hickman. 2004. "Influences of Online Delivery on Developmental Writing Outcomes." *Journal of Developmental Education* 28(1): 14–16, 18, 35.

Carswell, Alan D., and Viswanath Venkatesh. 2002. "Learner Outcomes in an Asynchronous Distance Education Environment." *International Journal of Human-Computer Studies* 56(5): 475–94.

Carswell, Linda, Pete Thomas, Marian Petre, Blaine Price, and Mike Richards. 2000. "Distance Education via the Internet: The Student Experience." *British Journal of Education Technology* 31(1): 29–46.

Clark, Ruth Colvin, and Richard E. Mayer. 2011. *E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*. 3rd Edition. San Francisco, Pfeiffer.

Curtis, David D, and Michael J. Lawson. 2001. "Exploring Collaborative Online Learning." *Journal of Asynchronous Learning Networks* 5(1): 21–34.

DeAngelo, Linda, Sylvia Hurtado, John H. Pryor, Kimberly R. Kelly, José Luis Santos, and William S. Korn. 2009. *The American College Teacher: National Norms for the 2007–2008 HERI Faculty Survey*. Los Angeles: Higher Education Research Institute, UCLA. http://heri.ucla.edu/PDFs/pubs/FAC/Norms/Monographs/TheAmericanCollegeTeacher2007To2008.pdf.

Delbanco, Andrew. 2013. "MOOCs of Hazard." *New Republic*, March 31.

Economics U$A. 2012. "The Great Depression and the Keynesian Revolution: What Did We Learn." Program #17 airscript. 21st Century Edition. Educational Film Center & Annenberg Foundation. https://www.learner.org/series/econusa/pdfs/unit17_keynes_videotranscript.pdf.

Glader, Paul. 2012. "Will MOOCs Promote Superstar Teaching over Superstar Research at Princeton and Other Ivy Universities? *Wired Academic,* May 18. http://www.wiredacademic.com/2012/05/will-moocs-promote-superstar-teaching-over-superstar-research-at-princeton-and-other-ivy-universities/.

Greenlaw, Steven A. 1999. "Using Groupware to Enhance Teaching and Learning in Undergraduate Economics." *Journal of Economic Education* 30(1): 33–42.

Gremmen, Hans, and Jan Potters. 1997. "Assessing the Efficacy of Gaming in Economic Education." *Journal of Economic Education* 28(4): 291–303.

Griffiths, Rebecca, Matthew Chingos, Christine Mulhern, and Richard Spies. 2014. *Interactive Online Learning on Campus: Testing MOOCs and Other Platforms in Hybrid Formats in the University System of Maryland*. July 10. Report by Ithaka S+R. http://sr.ithaka.org/sites/default/files/reports/S-R_Interactive_Online_Learning_Campus_20140716.pdf.

Guo, Philip J., Juho Kim, and Rob Rubin. 2014. "How Video Production Affects Student Engagement: An Empirical Study of MOOC Videos." Published at ACM Conference on Learning at Scale (L@S). http://groups.csail.mit.edu/uid/other-pubs/las2014-pguo-engagement.pdf.

Hallock, Kevin. 1995. "Seniority and Monopsony in the Academic Labor Market: Comment." *American Economic Review* 85(3): 654–57.

Herz, Bernhard, and Wolfgang Merz. 1998. "Experiential Learning and the Effectiveness of Economic Simulation Games." *Simulation & Gaming* 29(2): 238–50.

Honeychurch, Sarah, Niall Barr, Craig Brown, and John Hamer. 2012. "Peer Assessment Assisted by Technology." Computer Assisted Assessment. http://www.academia.edu/1765620/Peer_Assessment_Assisted_by_Technology.

Howard, Caroline. 2014. "No College Left Behind: Randy Best's Money-Making Mission to Save Higher Education." *Forbes,* March 3.

Junco, Reynol, and Jeanna Mastrodicasa. 2007. *Connecting to the Net.generation: What Higher Education Professionals Need to Know about* Today's *Students*. Washington, DC: National Association of Student Personnel Administrators (NASPA). http://blog.reyjunco.com/pdf/NetGenerationProof.pdf.

King, Allison. 1993. "From Sage on the Stage to Guide on the Side." *College Teaching* 41(1): 30–35.

Kolowich, Steve. 2014. "What You Need to Know about Companies That Run Online Programs for Colleges." *Chronicle of Higher Education,* September 30.

Kolowich, Steve. 2015. "Yale Announces 'Blended' Online Master's Degree." *Chronicle of Higher Education,* March 10.

Lean, Jonathan, Jonathan Moizer, and Mike Towler. 2006. "Simulations and Games." *Active Learning in Higher Education* 7(3): 227–42.

Leek, Jeff. 2013. "Data Analysis in the Top Nine Courses in Lifetime Enrollment at Coursera." Blog post. Simplystatistics.org, September 16. http://simplystatistics.org/2013/09/16/data-analysis-in-the-top-9-courses-in-lifetime-enrollment-at-coursera/.

Luxton-Reilly, Andrew. 2009. "A Systematic Review of Tools that Support Peer Assessment." *Computer Science Education* 19(4): 209–32.

Mitchell, Josh. 2014. "A Tough Lesson for College Textbook Publishers." *Wall Street Journal,* August 27.

Moore, Michael G. 1989. "Three Types of Interaction." *American Journal of Distance Education* 3(2): 1–7.

Navarro, Peter. Forthcoming. "The Death of the Large Lecture Hall, the Rise of Peer-to-Peer Course Delivery?" *American Journal of Distance Learning.*

Pearce, Kyle. 2013. "The 50 Most Popular Courses on Coursera." DIYGenius website. http://www.diygenius.com/the-50-most-popular-courses-on-coursera/.

Pryor, John H., Sylvia Hurtado, Linda DeAngelo, Jessica Sharkness, Laura C. Romero, William S. Korn, and Serge Tran. 2008. *The American Freshman: National Norms for Fall 2008.* Los Angeles: Higher Education Research Institute, UCLA. http://heri.ucla.edu/PDFs/pubs/TFS/Norms/Monographs/TheAmericanFreshman2008.pdf.

Ransom, Michael R. 1993. "Seniority and Monopsony in the Academic Labor Market." *American Economic Review* 83(1): 221–33.

Rovai, Alfred P, and Mervyn J. Wighting. 2005. "Feelings of Alienation and Community among Higher Education Students in the Virtual Classroom." *The Internet and Higher Education* 8(2): 97–110.

Schmidt, Stephen J. 2003. "Active and Cooperative Learning Using Web-based Simulations." *Journal of Economics Education* 34(2): 151–67.

Su, Bude, Curtis J. Bonk, Richard J. Magjuka, Xiaojing Liu, and Seung-hee Lee. 2005. "The Importance of Interaction in Web-Based Education: A Program-level Case Study of Online MBA Courses." *Journal of Interactive Online Learning* 4(1).

Suler, John. 2004. "In Class and Online: Using Discussion Boards in Teaching." *CyberPsychology & Behavior* 7(4): 395–401.

TeacherStream, LLC. 2009. "Mastering Online Discussion Board Facilitation: Resource Guide." Edutopia. https://www.edutopia.org/pdfs/stw/edutopia-onlinelearning-mastering-online-discussion-board-facilitation.pdf.

Topping, Keith. 1998. "Peer Assessment between Students in Colleges and Universities." *Review of Educational Research* 68(3): 249–76.

Topping, Keith J. 2009. "Peer Assessment." *Theory into Practice* 48(1): 20–27.

Topping, Keith J. 2010. "Peers as a Source of Formative and Summative Assessment." Chap. 4 in *Handbook of Formative Assessment,* edited by Heidi Andrade and Gregory J. Cizek. New York: Taylor and Francis.

UCI Distance Learning Center. 2014. "ILTI Proposal Preparation: Support Materials for UCI Faculty." Memorandum circulated electronically to UCI faculty.

University of Canberra. 2015. "Participating Effectively in Lectures and Tutorials." Available at http://learnonline.canberra.edu.au/mod/book/view.php?id=180718&chapterid=451.

Watters, Audrey. 2012. "The Problems with Peer Grading in Coursera." *Inside Higher Ed,* August 27. https://www.insidehighered.com/blogs/hack-higher-education/problems-peer-grading-coursera.

Xu, Di, and Shanna Smith Jaggars. 2011. "Online and Hybrid Course Enrollment and Performance in Washington State Community and Technical Colleges." CCRC Working Paper No. 31.

Zappe, Sarah, Robert Leicht, John Messner, Thomas Litzinger, and Hyeon Woo Lee. 2009. "Flipping the Classroom to Explore Active Learning in a Large Undergraduate Course." In Proceedings, American Society for Engineering Education Annual Conference & Exhibition. AC 2009-92.

Zavarella, Carol. 2008. "Computer-Based Instruction and Remedial Mathematics: A Study of Student Retention at a Florida Community College." Doctoral dissertation, University of South Florida.

Journal of Economic Perspectives—Volume 29, Number 4—Fall 2015—Pages 177–198

Rewriting Monetary Policy 101: What's the Fed's Preferred Post-Crisis Approach to Raising Interest Rates?[†]

Jane E. Ihrig, Ellen E. Meade, and Gretchen C. Weinbach

For several decades prior to the global financial crisis that started in 2007, the Federal Reserve through its Federal Open Market Committee (FOMC) primarily implemented monetary policy in a certain way: It set a target for the federal funds rate, which is an overnight interbank borrowing rate—that is, an interest rate paid when banks borrow from other banks in the very short-term. The Fed pursued its desired federal funds interest rate target through "open market operations" that involved modest purchases and sales of Treasury securities. However, in the aftermath of the financial crisis and with a superabundant level of reserve balances in the banking system having been created as a result of the Federal Reserve's large-scale asset purchase programs, implementing monetary policy through this traditional approach will no longer work. Instead, the Fed intends to affect the federal funds interest rate by using policy tools like the interest rate paid on excess reserves and a facility to extend overnight reverse repurchase agreements.

Being able to explain and to understand this fundamental change in the Fed's main tools for the implementation of monetary policy has implications for a number of groups. It obviously matters for the Fed itself; in particular, the Federal Reserve has been influenced in recent years by academic research showing that communication and transparency have substantial effects on the credibility and strength of monetary policy. Many investors and market-watchers seek to look below the

■ *Jane E. Ihrig, Ellen E. Meade, and Gretchen C. Weinbach are Deputy Associate Director, Senior Adviser, and Associate Director, respectively, in the Division of Monetary Affairs, Board of Governors of the Federal Reserve System, Washington, DC. Their email addresses are jane.e.ihrig@frb.gov, ellen.meade@frb.gov, and gweinbach@frb.gov.*

[†] For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.29.4.177

doi=10.1257/jep.29.4.177

surface of Fed decisions—like the announced target for the federal funds interest rate—and to understand how such decisions are actually implemented. The shift in policy tools also affects the task of some of society's explainers, including journalists and teachers of economics, because most of the past textbook descriptions of how monetary policy works will not be accurate for years to come.

Of course, the Federal Reserve is not the only central bank that will face the challenge of tightening monetary policy while holding a much larger balance sheet than it held in the past. The Fed has seen its assets rise from about $900 billion in 2006 to about $4.5 trillion today, or from 6 percent of nominal gross domestic product (GDP) to about 26 percent of nominal GDP. Other central banks have had similar or larger increases. For example, assets of the Bank of Japan have increased from about 20 percent of nominal GDP to more than 60 percent of nominal GDP over this period, and assets of the Swiss National Bank have increased from 20 percent of nominal GDP to more than 80 percent of nominal GDP. The net increase in assets of the European Central Bank has so far been more modest, with assets increasing from less than 10 percent of nominal GDP for the euro zone to more than 20 percent of nominal GDP—but its quantitative easing program is still underway.

Though other central banks also will be confronted with similar issues, this paper focuses on the Federal Reserve's past, present, and future approach to implementing monetary policy. In particular, we provide a primer on how the Federal Reserve will implement monetary policy when the Federal Open Market Committee decides it is time to raise interest rates. We begin with the standard textbook model of reserve balances to illustrate the approach used by the Federal Reserve before the financial crisis to keep the federal funds rate near its desired target. We explain why that pre-crisis approach will not work in the current environment. We then discuss the policy tools available to implement monetary policy, and explain the approach that the Committee intends to take when it decides to begin raising short-term interest rates. For additional detail on the issues discussed in this paper, a useful starting point is our discussion paper Ihrig, Meade, and Weinbach (2015).

How Did the Fed Implement Monetary Policy Prior to the Financial Crisis?

The textbook explanation of open market operations is based on two key features: 1) requirements that banks hold reserve balances in amounts determined by the Federal Reserve; and 2) banks trying to keep these balances to a minimum, in part because before the financial crisis the balances earned no return.

The original Federal Reserve Act (as amended by the Monetary Control Act of 1980) and the International Banking Act of 1978 impose reserve requirements on most deposit-taking institutions in the United States, requiring that commercial banks, savings banks, thrift institutions, and credit unions—as well as most US branches and agencies of foreign banks—(hereafter "banks," for simplicity) are assessed reserve requirements against certain deposit liabilities. For example, as of January 22, 2015,

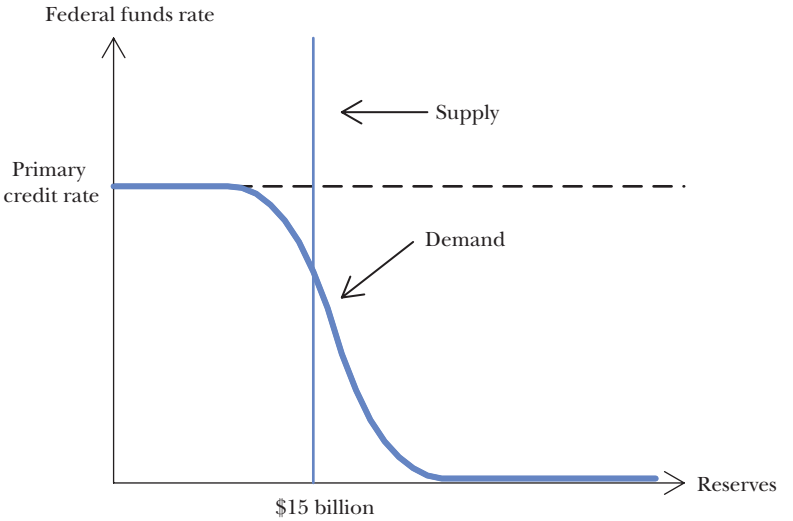
institutions needed to hold reserves equal to 3 percent of any net transaction accounts between $14.5 million and $103.6 million, and 10 percent of any net transaction accounts above $103.6 million (http://www.federalreserve.gov/monetarypolicy/reservereq.htm#table1). Banks are required to satisfy their reserve requirements in the form of vault cash, which they hold primarily to meet the liquidity needs of their customers and, if the quantity of vault cash held is insufficient, also in the form of a balance maintained at the Federal Reserve. Prior to the financial crisis, many banks in the United States satisfied their reserve requirement with vault cash, though about 900 banks did not and so also needed to maintain reserve balances at the Fed. The balances that banks maintain at the Federal Reserve that are necessary for meeting reserve requirements are "required reserve" balances; any reserve balances held in excess of what is necessary to meet reserve requirements are termed "excess reserve" balances.[1] Before the financial crisis and recession that started in 2007, total reserve balances in the US banking system hovered around $15 billion, with excess balances making up less than $2 billion of this total. As discussed in greater detail below, reserve balances have grown tremendously since the financial crisis.

The combination of Federal Reserve–created demand for reserve balances and the desire of banks to limit such balances drove an active interbank market, known as the federal funds market, in which banks borrowed from and lent funds to each other on a daily basis at an interest rate known as the federal funds rate. With reserve balances generally scarce, the Federal Reserve could affect the market-determined level of the federal funds rate and keep it close to target level by a combination of announcing a target level for the federal funds rate and making small changes in the supply of aggregate reserves as needed.

Figure 1 presents the standard demand and supply framework for reserve balances shown in many textbooks. The demand by banks for reserves is downward sloping because of the opportunity cost of holding reserve balances (which in the past paid no interest). Conversely, as the price of overnight borrowing falls, banks are generally inclined to hold more reserves in order to satisfy their reserve requirements and also possibly to leave themselves with modest excess balances to protect against unexpected outflows that can cause reserve balance deficiencies—for which banks are charged a penalty. The upper left-hand side of the demand curve becomes horizontal at the "primary credit rate," which is the interest rate that the Fed charges banks to borrow overnight (as part of the Fed's discount window). Borrowing at the primary credit rate provides banks with a source of back-up funding at an interest rate that is well above

[1] In practice, banks meet their required reserve balances (also referred to as "reserve balance requirements") with some leeway. A penalty-free band is used to create a range on both sides of the required reserve balance within which a bank needs to maintain its average balance over a given period. For more information on reserve requirements, see the Federal Reserve Board's "Reserve Maintenance Manual" at http://www.federalreserve.gov/monetarypolicy/2015-reserve-maintenance-manual-about-this-manual.htm or its web page on "Reserve Requirements" at http://www.federalreserve.gov/monetarypolicy/reservereq.htm. Data on reserve balances are published weekly on the H.3 Statistical Release at http://www.federalreserve.gov/releases/h3/current.

Figure 1
Banks' Demand for and the Fed's Supply of Reserve Balances before the Financial Crisis



Source: Authors.

the Fed's target federal funds rate.[2] Although, in theory, banks should be unwilling to pay more than the primary credit rate for overnight funding, they sometimes do. Borrowing from the Fed involves higher transactions costs as well as possible reputational effects (termed "stigma") in which banks fear that borrowing from the Fed sends a signal that they are not regarded by other financial institutions as a good credit risk. For these reasons, some banks may choose to borrow from other institutions in the federal funds market at interest rates that exceed the primary credit rate.

The Fed's supply curve for reserve balances is vertical because the Fed is a monopolistic supplier of reserves; the supply curve shifts to the right or left when the Fed adds or subtracts reserves from the banking system using open market operations. The intersection of the demand and the supply curves occurs at the market federal funds rate.

Prior to the financial crisis, the supply and demand curves for bank reserves intersected on the downward-sloping portion of the demand curve. As a result, if the market federal funds rate was above the target federal funds rate, then the Fed would execute purchases of securities that would add reserve balances to the banking system and shift the supply curve to the right. Conversely, if the market federal funds rate was below the target federal funds rate, then the Fed

[2] Data on banks' aggregate borrowings from the Fed are published weekly on the H.3 Statistical Release at http://www.federalreserve.gov/releases/h3/current. For more information on the Fed's discount window programs, see *Purposes & Functions* (Federal Reserve System 2005).

would execute sales of securities that would drain reserve balances from the banking system and shift the supply curve to the left. (Of course, when banks trade existing reserve balances among themselves in the federal funds market, that trading leaves the aggregate amount of reserve balances unchanged; see the online Appendix available with this paper at http://e-jep.org for a discussion of this point.) Each business day, the Federal Reserve examined demand and supply conditions and, informed by staff models, determined whether an adjustment to reserve supply was needed, including which kind was suitable and the approximate size that would be appropriate. Judson and Klee (2010) discuss how forecasts were used to determine open market operations.

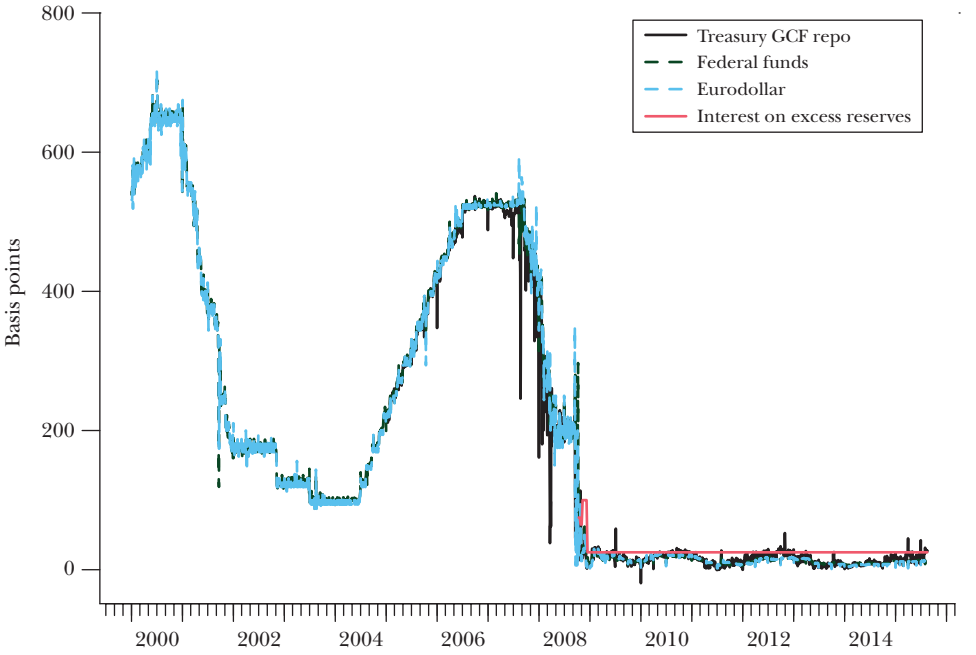
Prior to the financial crisis, the kind of open market operation that the Fed would use to produce the desired movement in reserve supply depended on its assessment of conditions in the market for reserves. For example, suppose the goal was to reduce the federal funds interest rate. In this situation, the Federal Reserve—more specifically, the Open Market Trading Desk at the Federal Reserve Bank of New York—would purchase a security from the private sector, a transaction that cleared through banks and resulted in reserve balances being added to the banking system. This purchase could be permanent or it could be temporary (the latter transaction is termed a repurchase agreement). In Figure 1, this transaction would shift the supply curve to the right for as long as the Fed owned the security, and thereby put downward pressure on the market federal funds rate. (The online Appendix available with this paper at http://e-jep.org describes the mechanism by which increases in the Fed's securities holdings result in a commensurate increase in the amount of reserve balances held by the banking system.)

The Fed's monetary policy targeted the federal funds interest rate, and then other short-term market interest rates tended to move with that rate. For example, Figure 2 shows three different overnight market interest rates. The federal funds interest rate that is targeted by the Fed reflects, as we have already discussed, a market in which banks are the borrowers, and a mixture of banks, securities dealers, and government-sponsored enterprises (financial services corporations created by Congress, such as Fannie Mae, Freddie Mac, and the Federal Home Loan Banks) are the potential lenders.

The Eurodollar market, although it started in London, is now a large global market. "Eurodollars" is a general term for large (often in the millions of dollars) US dollar-denominated deposits in banks outside the United States, usually held for a period of less than six months. Such deposits avoid regulations applicable to US-based deposits. The Eurodollar market is a place for money market funds and various financial and nonfinancial lenders to store funds for relatively short periods of time.

The repurchase agreement market, or repo market, involves a two-part transaction in which one party first sells a security to another and simultaneously agrees to repurchase that security in the near future. The original buyer of the security is in effect lending money on a short-term basis, and earns a rate of return for doing so, while the original seller of the security obtains additional cash in the short-term. The difference between the sale price and the repurchase price of the security, together with the length of time between the sale and purchase steps

Figure 2
Overnight Market Interest Rates



Source: Authors using data from Depository Trust & Clearing Corporation, Federal Reserve Bank of New York, Bloomberg, and Federal Reserve Board of Governors.
Note: GCF = General Collateral Finance.

of the transaction, implies the rate of interest earned by the party that purchased the security and loaned the funds. The repurchase market typically involves banks and securities dealers taking the role of cash borrowers—that is, they are typically sellers of securities in the first stage of a repurchase agreement. Money market funds, hedge funds, government-sponsored enterprises, and securities dealers are the lenders in this market, essentially holding the securities while lending cash for a short time until the repurchase agreement expires or is renewed. Before the financial crisis, many of the Fed's daily open market purchases of securities were structured as repo transactions.

The market for repos is complex. There are two basic types of repo transactions: "bilateral" and "tri-party," referring to the number of participants involved in the transaction. Within the tri-party market there is a segment called the GCF (General Collateral Finance) repo market, mostly used by securities dealers and serviced by the Fixed Income Clearing Corporation. (For more information on the structure of repo markets, see Copeland, Duffie, Martin, and McLaughlin 2012.) The term "general collateral" means that the party lending the money—that is, the party buying the security that will later be repurchased—is willing to accept a range of bonds issued by the US Treasury and by government-sponsored enterprises as collateral for the loan.

We will return to a discussion of repurchase agreements and how the Federal Reserve plans to make use of their cousin, reverse repurchase agreements, later in this paper. Here, we only wish to emphasize that overnight market interest rates tend to track each other. This pattern reflects, in part, the fact that many of the same financial institutions are active participants in the markets for various money market instruments. For example, banks are active borrowers in all three of the money markets depicted in Figure 2, and while the lenders vary a bit across the markets, there is also notable overlap. All in all, arbitrage generally works well to keep short-term interest rates highly correlated.

In broad terms, persistent changes in the level of short-term interest rates are transmitted to other, longer-term interest rates as well, including those commonly faced by businesses and households—although this connection from changes in short-term to long-term interest rates is not a simple one-to-one process. Ultimately, the Federal Reserve conducts monetary policy in order to achieve its statutory mandate of maximum employment, stable prices, and moderate long-term interest rates as prescribed by the Congress and laid out in the Federal Reserve Act. (The Federal Reserve's statutory mandate is often referred to as a "dual mandate" of maximum employment and price stability, because of the belief that moderate long-term interest rates will result if inflation is expected to be low and stable.) Thus, as economic conditions change over time, the Federal Open Market Committee adjusts monetary policy accordingly, typically by raising or lowering its target for the federal funds rate, so as to foster economic conditions it judges to be consistent with achieving its statutory goals.

How Did the Financial Crisis Affect the Fed's Operational Framework?

The first event commonly associated with the global financial crisis took place on August 9, 2007, when the French bank BNP Paribas suspended withdrawals from three of its investment funds due to problems in the US subprime mortgage market. At the onset of the financial crisis, the Federal Open Market Committee began reducing its target for the federal funds interest rate, and implementing policy using the conventional open market operations discussed in the previous section. The target federal funds interest rate moved down from 5¼ percent in August 2007, to its effective lower bound of 0 to 25 basis points in December 2008, where it remained in early fall of 2015.[3]

[3] The target or "intended" federal funds rate is published on the Federal Reserve Board's website at http://www.federalreserve.gov/monetarypolicy/openmarket.htm. The Federal Reserve also responded to the financial crisis with a number of credit and liquidity programs designed to support the liquidity of financial institutions and foster improved conditions in financial markets. Although these programs led to significant increases in the Federal Reserve's balance sheet, the programs have expired or were concluded, and they are not boosting the Fed's balance sheet today. Details of these liquidity programs are available on the Federal Reserve Board's website at http://www.federalreserve.gov/monetarypolicy/bst_crisisresponse.htm.

Table 1
A Simplified Federal Reserve Balance Sheet: Before and After the Financial Crisis
(billions of dollars)

Before: August 8, 2007				*After: December 24, 2014*			
Assets		*Liabilities*		*Assets*		*Liabilities*	
Securities	791	Reserve balances	14	Securities	4,247	Reserve balances	2,610
Other assets	78	Currency	777	Other assets	262	Currency	1,294
		Other	45			Other	548
		Capital	33			Capital	57
Total	**869**	**Total**	**869**	**Total**	**4,509**	**Total**	**4,509**

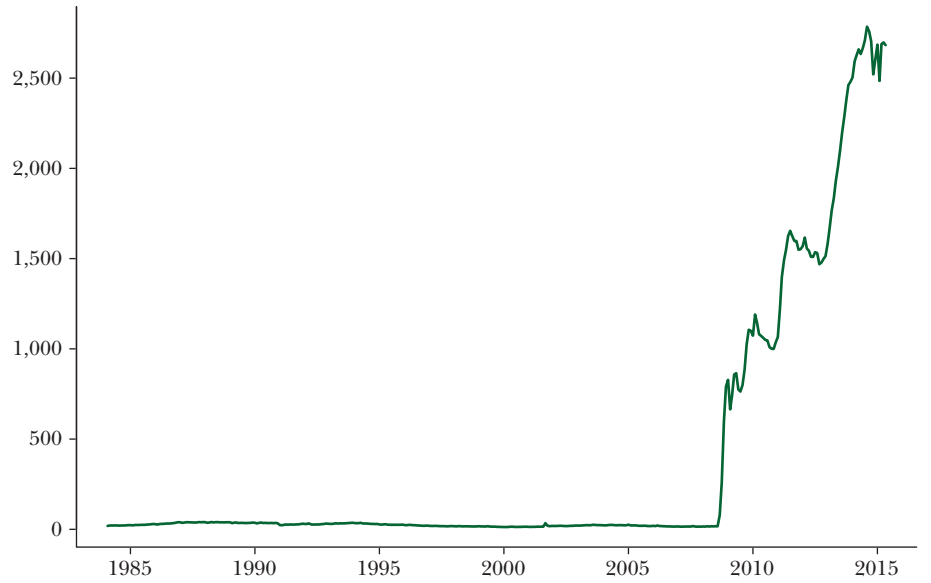
Source: Authors using data from Federal Reserve Board of Governors H.4.1 Statistical Release, titled "Factors Affecting Reserve Balances."

As short-term interest rates reached near-zero, the Federal Open Market Committee carried out a series of large-scale asset purchase programs between November 2008 and October 2014 in which the Fed purchased in the secondary market about $1,690 billion in Treasury securities, $2,070 billion in agency mortgage-backed securities, and $170 billion in debt issued or guaranteed by government-sponsored enterprises.[4] These operations were unprecedented and their effects uncertain. The programs were intended to put downward pressure on longer-term interest rates in the economy—the purchases reduced the available supply of securities in the market, leading to an increase in the prices of these securities and a reduction in their yields. Academic studies provide varying estimates of the magnitude of downward pressure that these operations have put on longer-term interest rates (Fischer 2015, table 1, titled "Empirical Studies of LSAPs"). The purchase programs taken together are estimated to have reduced longer-term interest rates by roughly 100 basis points, as reported in Ihrig, Klee, Li, Schulte, and Wei (2012). The large-scale asset purchase programs also helped to support mortgage markets (Krishnamurthy and Vissing-Jorgensen 2011).

For the purposes of this paper, the key issue isn't how these large-scale asset purchase programs affected interest rates or mortgage markets, but rather that their legacy is a dramatic alteration of the Federal Reserve's balance sheet. Table 1 shows a simplified version of the Federal Reserve's balance sheet before

[4] In addition, from September 2011 through December 2012, the Fed conducted a maturity extension program where it sold or redeemed $667 billion in shorter-dated Treasury securities and purchased the same amount of longer-dated Treasury securities, as reported on Federal Reserve Board's website at http://www.federalreserve.gov/monetarypolicy/bst_openmarketops.htm. Mortgage-backed securities are a type of asset-backed security that is secured by a package of mortgage loans and for which interest and principal payments associated with the mortgages are passed through to the holders of the securities; agency mortgage-backed securities are those issued by government-sponsored enterprises.

Figure 3
Total Reserve Balances held by Banks
(billions of dollars)



Source: Authors using data from Federal Reserve Board of Governors, H.3 Statistical Release, titled "Aggregate Reserves of Depository Institutions and the Monetary Base."

and after the financial crisis. The left panel shows that on August 8, 2007, the Federal Reserve's assets were comprised principally of Treasury securities holdings of $791 billion; its liabilities were mainly currency ($777 billion), with banks holding $14 billion in reserve balances at the Federal Reserve. As the Fed made its purchases of securities, the Fed generally also reinvested payments of principal and interest to keep its portfolio of securities from shrinking. As a result, by late December 2014, the Fed's securities holdings rose to nearly 5½ times their pre-crisis level, as shown in the right panel of Table 1. In addition, reserve balances became the Fed's largest liability, amounting to $2.6 trillion, and, as shown in Figure 3, these balances have remained in that neighborhood since then, with excess reserves making up all but about $90 billion of this total.

Another important factor affecting the federal funds market (and thus the implementation of monetary policy going forward) is that since October 2008, the Federal Reserve has paid interest on banks' reserve balances. The Financial Services Regulatory Relief Act of 2006 authorized interest payments on reserve balances beginning in 2011, and the Emergency Economic Stabilization Act of 2008 advanced the effective date of this authority to October 2008. The Federal Reserve has designated two rates of interest on reserve balances, one rate for required reserve balances and a separate rate for excess reserve balances; the

interest rates that the Fed pays on reserves that are required and those that are excess are currently the same, although they could be set at different levels. In the discussion in this paper, for simplicity and given the predominance of excess reserve balances, we focus on the interest rate on excess reserve balances. All else equal, an increase in the interest rate on excess reserves would be expected to put upward pressure on the federal funds rate because banks would have an incentive to borrow in the federal funds market at rates below the interest rate on excess reserves and place those balances at the Fed.

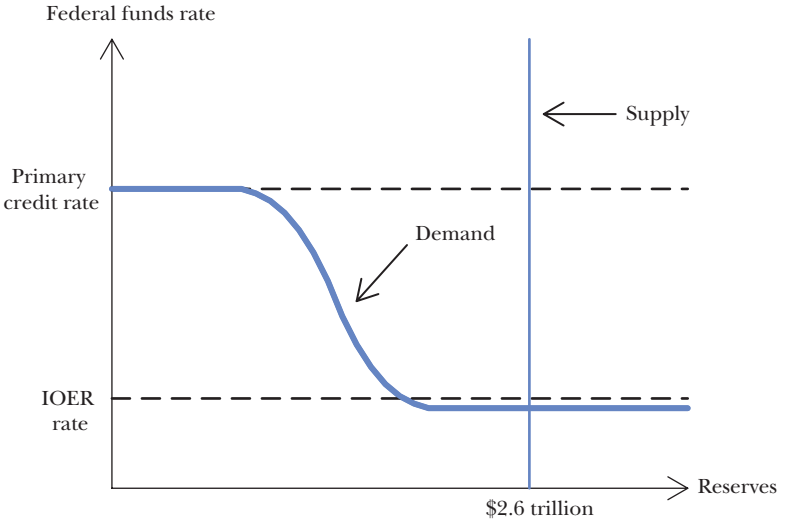
Since the Fed began paying interest on reserves, the market federal funds interest rate has generally been *below* the interest rate on excess reserves. One might think that interest on excess reserves (IOER) (see Figure 2) should provide a floor for the federal funds interest rate because banks would not lend at rates below what they could receive at the Fed. However, this situation has arisen because, in addition to banks not needing to borrow actively from each other because of the high quantity of reserves already in the banking system, the *nonbank* lenders in the federal funds market have an incentive to lend reserves at any rate above zero because they are not eligible to earn the interest rate on excess reserves on the balances they keep at the Fed. As explained above, the nonbanks that are active in the market for federal funds are government-sponsored enterprises. Banks borrow from these nonbanks to earn the spread between the market interest rate at which they borrow funds and the interest rate they earn from the Fed by holding those funds as excess reserves.

Figure 4 shows the market for reserve balances in the last few years after the expansion of bank reserves and illustrates two key differences from Figure 1. First, the supply curve for reserves is far to the right on the x-axis, representing the superabundant level of reserves in the banking system. The supply and demand curves now intersect on the flat portion of the demand curve. With the supply curve for reserves in its current position, the traditional steps to put upward pressure on market interest rates—announcing a higher target level for the federal funds rate and being prepared to conduct the appropriate open market operation by selling a small amount of securities into the market and draining an equally small amount of reserves—will no longer suffice. Second, with the Fed paying interest on reserves, the lower portion of banks' demand curve flattens out near the interest rate on excess reserves, reflecting the arbitrage activity just described. In this situation, when the time arrives to raise the target range for the federal funds interest rate, how will the desired increase be accomplished?

What Tools Could the Fed Use to Raise Interest Rates?

The Federal Reserve has a number of policy tools—some traditional, some new—that it can use to help raise the federal funds interest rate in a situation of superabundant reserves. In this section, we discuss the available policy tools, along with the ways in which those policy tools are expected to influence the federal funds rate.

Figure 4
Banks' Demand for and the Fed's Supply of Reserve Balances Today



Source: Authors.
Note: IOER = interest on excess reserves.

Channels of Influence of the Policy Implementation Tools

It is useful to summarize the three main channels through which the Fed's policy tools are generally expected to affect the market-determined federal funds interest rate and broader interest rates in the economy: encouraging arbitrage, increasing the scope of influence, and increasing reserve scarcity. We will refer to these concepts below in describing how each of the available policy tools is thought to work.

A policy tool can *encourage arbitrage* in money markets when it offers an interest rate that acts as a reservation rate—that is, the lowest rate of return that a financial institution would be willing to accept for investing its funds when assessing available investment opportunities.[5] Generally speaking, financial institutions with access to a given policy tool have an incentive to borrow funds in money markets at rates that are below the interest rate that the Federal Reserve offers on the policy tool and invest the funds in the policy tool, putting upward pressure on money market rates.

If a policy tool establishes a reservation rate for a broader set of financial institutions than banks, we say it has an *increased scope of influence* in money markets. Access to this tool will narrow the set of institutions that might lend money below the rate earned on the policy tool and put upward pressure on the lowest interest rates in money markets.

[5] Note that the Fed is a risk-free counterparty. Because there is no risk that the Federal Reserve will be unable to return money, banks do not require additional compensation for default risk in the rates they receive from the Fed.

Use of a policy tool can *increase reserve scarcity* by draining reserve balances and moving the level of aggregate reserves closer to its traditional position. If the aggregate level of reserve balances were reduced sufficiently, banks would need to resume borrowing federal funds to meet their demand for reserve balances, leading them to put upward pressure on the market federal funds rate.

Available Policy Tools

Rate of Interest on Excess Reserve Balances. As noted earlier, most transactions in the federal funds market today reflect arbitrage activity between banks that earn interest on reserves and nonbanks that do not (Goodfriend 2015 offers more detail). An increase in the interest rate on excess reserves should pull up the federal funds rate in these arbitrage transactions. Similarly, other money market rates should increase as banks arbitrage between holding excess reserve balances and alternative money market instruments.

Of course, banks need to be willing and able to actively perform this arbitrage for these effects to be realized. As shown in Figure 2, the federal funds rate has been highly correlated with other money market rates, which suggests that such arbitrage does happen. However, our understanding of the potential strength of these arbitrage effects may be incomplete, in part because interest on excess reserve balances has only been in effect over a period of time in which short-term interest rates have been kept near zero.

Overnight Reverse Repurchase Operations. In this type of open market operation, the Open Market Trading Desk would *sell* a security to the private sector, a transaction that would initially result in a decline in the quantity of reserve balances in the banking system, shifting the supply curve to the left. As with a repo transaction, this transaction would include a second step in which the transaction is unwound—the Desk would repurchase the security at a specified price at an agreed-upon time in the future and return the funds it had been holding, leaving reserve balances back where they started.

In the past, the Fed has conducted relatively small-dollar amounts of overnight reverse repurchase agreements with "primary dealers," which are institutions that buy and sell Treasury securities directly from and to the Fed with the intention of acting as the "middleman" between the Fed and market participants in the private sector. A full list of primary dealers is available at http://www.newyorkfed.org/markets/pridealers_current.html. Some well-known examples include firms like Cantor Fitzgerald & Co., Citigroup Global Markets, Credit Suisse Securities (USA), Daiwa Capital Markets America, Deutsche Bank Securities, and Goldman, Sachs & Co.[6]

[6] The Fed also regularly conducts overnight reverse repurchase agreements for international organizations, and the amount of outstanding reverse repurchase agreements reported on the Fed's balance sheet recently includes about $150 billion of these transactions (for more detail, see the H.4.1 statistical release at http://www.federalreserve.gov/releases/h41).

At present, the Fed is testing a somewhat different reverse repo transaction known as overnight reverse repurchase operations. These operations have three key differences relative to the small-scale open market operations used in the Fed's monetary policy operations before the financial crisis.

First, the Fed is currently offering overnight reverse repurchase agreements on a daily basis at a pre-announced "offering rate," which is the maximum interest rate the Fed is willing to pay in the operation. Counterparties will compare the Fed's overnight reverse repurchase offering rate to other money market rates and determine whether to bid in the Fed's overnight reverse repurchase operation. The Fed initially took the approach of offering a rate of return (typically 5 basis points) and accepting bids from all eligible counterparties willing to accept that rate. Currently, the Fed specifies an offering rate of return and the size of each operation is limited, both in terms of the amount each individual institution can bid (currently up to $30 billion) and the aggregate amount of the operation (currently set at $300 billion). The Federal Reserve has been reporting the results of its daily overnight test operations, including the bid amounts submitted and accepted, as well as the high, low, and awarded bid rates. On September 30, 2014, for example, demand for the Fed's overnight reverse repurchase operation was more than $400 billion. The results of the most recent operation may be found on the Federal Reserve Bank of New York's website at http://www.newyorkfed.org/markets/omo/dmm/temp.cfm. In testing overnight reverse repurchase operations, the Fed has varied the offering rate, and this has generally demonstrated that demand for the Fed's overnight reverse repurchase operations is indeed sensitive to the pattern of interest rates.

Second, the set of counterparties that are eligible to participate in the Fed's overnight reverse repurchase operations is much broader than it was in the past. The Fed conducted traditional open market operations with primary dealers; today, the institutions that are eligible to participate in the Fed's overnight reverse repurchase operations include about two dozen banks as well as a large number of money market funds under the management of 29 different firms, 22 primary dealers, and 13 government-sponsored enterprises (including 10 separate Federal Home Loan Banks). A full list of the counterparties that are eligible to participate in the Fed's reverse repurchase operations appears at http://www.newyorkfed.org/markets/expanded_counterparties.html. The eligible nonbank institutions, which are unable to earn interest on reserves, may be encouraged to engage in arbitrage activity relative to the rate the Fed offers on its overnight reverse repurchase operations because they have little incentive to lend funds in money markets at interest rates below the one they can receive directly by engaging in a reverse repurchase agreement with the Fed. For this reason, the Fed's reverse repurchase operations can provide a floor under the level of money market interest rates.

Third, unlike in the past, the Fed's overnight reverse repurchase agreements could be used in relatively large scale to increase the scarcity of reserves. When the Fed announces an overnight reverse repurchase test operation, it also announces

an aggregate offering amount—the total amount of dollars the Fed is willing to accept at the operation. The Fed's testing of overnight reverse repurchase operations has demonstrated that these operations can set a soft floor under the level of the federal funds rate, and other short-term market interest rates, as long as market participants are confident that the aggregate capacity of the operations is large enough to meet demand. If the Fed wanted to increase the scarcity of reserves in the banking system, it could set the offering amount on its operations relatively high, and possibly also adjust the offering rate, to encourage demand for these operations. (The online Appendix available with this paper at http://e-jep.org describes the mechanism by which increases in overnight reverse repurchase operations result in a commensurate decline in the amount of reserve balances held by the banking system.) However, the Fed has discussed concerns associated with having a persistently large overnight reverse repurchase program, a topic to which we will return. Thus, the role that overnight reverse repurchases may play in increasing reserve scarcity over time is likely to be limited.

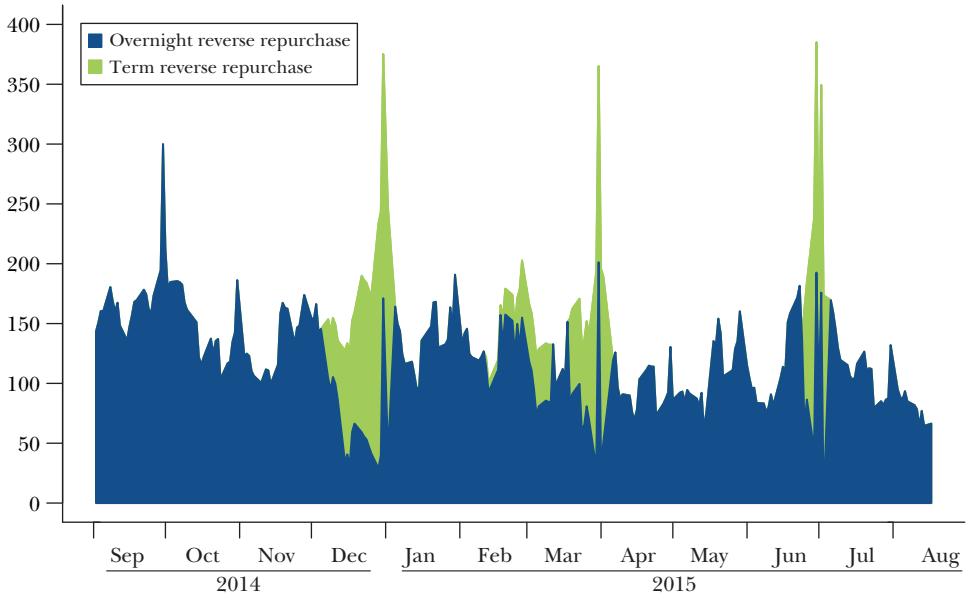
Term Reverse Repurchase Operations. In addition to overnight reverse repurchase operations, the Fed can conduct the same type of open market operation but have the second repurchasing stage occur more than one business day later; such transactions are known as "term reverse repurchase agreements." As was the case with the overnight version of these operations, in the past the Fed has conducted relatively small-dollar amounts of term reverse repurchase agreements with primary dealers.

The Fed has been testing term reverse repurchase operations, although less regularly than the overnight operations, as shown in Figure 5. During the testing period, the Fed has offered its term operations at varying amounts, rates, and maturities, with most operations timed to cover quarter-end dates. The reason for this timing is that investment options for major cash lenders tend to dwindle at quarter-ends because some large banking institutions reduce the size of their balance sheets—that is, they tend to borrow less and accommodate less investment activity of other institutions—at that time in light of regulatory reporting requirements. About $200 billion of term reverse repurchase agreements were outstanding on the Fed's books at year-end 2014 and March-end 2015. Testing has also showed that term operations serve in part as a substitute for overnight operations.

Term Deposit Facility. The Federal Reserve may also choose to offer interest-bearing deposits to banks through its Term Deposit Facility. When a bank elects to place funds in this facility, the funds are moved out of reserve balances for the life of the term deposit. (The online Appendix available with this paper at http://e-jep.org describes the mechanism by which increases in term deposits result in a commensurate decline in the amount of reserve balances held by the banking system.) Thus, the Term Deposit Facility acts to increase reserve scarcity and it also encourages arbitrage as banks compare the yield the Fed offers on a term deposit with other investments opportunities.

The Fed has been testing the functionality of two types of term deposit operations since June 2010. In the first type, the Fed offers a given dollar amount of term

Figure 5
Total Overnight and Term Reverse Repurchase Test Operations
(billions of dollars outstanding daily)



Source: Authors using data from Federal Reserve Bank of New York, "Temporary Open Market Operations" https://apps.newyorkfed.org/markets/autorates/tomo-results-display?SHOWMORE=TRUE.

deposits; banks then bid for the size of the deposit they want and specify the interest rate on the deposit. The Fed accepts bids beginning with the lowest bid rate and proceeding to higher bid rates until the total offered amount is exhausted. In this type of term deposit operation, all banks receive an interest rate identical to the rate paid to the last bank whose bid was accepted—that is, all banks receive the highest bid rate accepted. In the second type of term deposit operation, the Fed offers an interest rate and allows banks to deposit the amount of funds they desire, up to a predetermined maximum.

In its test operations, the Fed has varied some features, including the length of the term, the offering rate, and whether banks are permitted to withdraw their deposits prior to the end of the term, subject to a penalty. The option to withdraw deposits early has proven to be particularly attractive to banks in making their cash management decisions. During testing in February 2015, term deposits outstanding grew to about $400 billion on the Fed's balance sheet. Results of all term deposit operations can be found on the Federal Reserve Board's website at http://www.federalreserve.gov/monetarypolicy/tdf_2014.htm.

Sell Federal Reserve Securities Holdings. One might expect that the Federal Reserve would choose to sell some of its securities holdings. That is, large-scale

asset purchases created the current situation of superabundant reserves, so why not engage in large-scale asset sales to reverse the process? Such sales could be seen as a way to return to a situation in which reserves are sufficiently scarce so that the Fed's traditional approach to implementing changes in the target federal funds rate may again be used, or as a way to unwind the current accommodative stance of monetary policy directly. For either reason, sales might be seen as the most obvious course of action in current circumstances.

However, these motivations for possibly selling securities are unattractive to policymakers. First, with excess reserves having accumulated over the course of several years and standing well in excess of $2 trillion, a policy to substantially reduce the supply of reserves in the banking system would have to be enacted in a very large size and even then would take a significant amount of time to complete. For example, if the chosen pace of the Fed's large-scale asset *purchases* is any guide, it would take a number of years to sell a sufficient quantity of securities so as to cause a meaningful inward shift in the reserves supply curve. Such a tactic is not sufficiently nimble for implementing monetary policy. Also, using a heftier pace of sales might bring about unwanted effects in financial markets—for instance, the upward pressure on market interest rates that would accompany sizable sales of the Fed's holdings of securities could be hard to gauge and control.

Second, and more broadly, the monetary policies of central banks around the globe have generally been implemented in short-term financial markets. The Fed has many years of experience affecting conditions in the federal funds market, which is an overnight money market, and prefers to continue to implement its policy through this market as it raises short-term interest rates to more-normal levels while its balance sheet is large.

For these reasons, the Federal Open Market Committee has indicated that it plans to reduce its securities holdings in a gradual and predictable way primarily by altering its policy of reinvesting maturing and prepaying securities at some point after the start of the policy normalization process. In particular, the Committee has indicated that securities sales will not be part of the initial package of steps that it intends to take to begin to raise interest rates, and that it does not anticipate selling agency mortgage-backed securities as part of the policy normalization process, although limited sales might be warranted in the longer run to reduce or eliminate residual holdings. The Committee has emphasized that the timing and pace of any securities sales would be communicated to the public in advance.

Alter Reinvestments of Federal Reserve Securities Holdings. As just noted, the Federal Open Market Committee has said that it plans to alter its current policy of reinvesting maturing and prepaying securities—that is, either cease or commence phasing out its securities reinvestments—at some point after it begins increasing the target range for the federal funds rate. The timing with which the Committee will alter its reinvestments policy will depend on how economic and financial conditions and the economic outlook evolve. The pace at which this

tactic would reduce the Fed's overall holdings of securities is driven in part by the maturity dates of the Fed's holdings of Treasury and agency securities, which are known (and reported by the New York Fed at http://www.newyorkfed.org/markets/soma/sysopen_accholdings.html), and also by the pace at which agency mortgage-backed securities might be prepaid, which can only be estimated. Some examples of when prepayment of agency mortgage-backed securities would occur include when households pay off some or all of a mortgage balance early because they refinance their original mortgage with a lower available mortgage rate, pay off their mortgage when they sell a house to move, or pay down a portion of their mortgage to reduce the level of their debt.

If the Federal Open Market Committee decided to end its reinvestments, how would this affect the Fed's security holdings? As of late December 2014, the Federal Reserve held $4.2 trillion of securities, of which about $2.5 trillion were Treasury securities, $1.7 trillion were agency mortgage-backed securities, and about $39 billion were agency debt. If the Federal Open Market Committee were to end its policy of reinvesting its securities holdings in the near future, nearly $700 billion of securities would mature or roll off of the Fed's portfolio in 2016 and 2017 taken together, comprising about $410 billion of Treasury securities and an estimated $290 billion of agency mortgage-backed securities (using actual holdings of securities as of December 24, 2014, and projected prepayments of agency mortgage-backed securities based on the model in Carpenter, Ihrig, Klee, Quinn, and Boote 2015, along with Blue Chip interest rate projections). Although this step would gradually shift the supply curve of reserves to the left, increasing reserve scarcity, it would take a number of years before such a shift in the supply of reserves would put upward pressure on the federal funds interest rate.

Reserve Requirements. Reserve requirements are calculated as fractions of certain deposit account levels and have traditionally helped to create demand for reserve balances. These reserve ratios are re-evaluated annually, although they have not been adjusted since 1992. In theory at least, the Fed could increase existing reserve ratios in order to require that additional reserve balances be held by banks, thereby contributing to the scarcity of excess reserve balances. The Fed may also, after consulting Congress, impose so-called emergency reserve requirements if it finds that extraordinary circumstances require such action, a step that the Fed has never before taken. Authorization for emergency reserve requirements is given in Section 204.5 of the Federal Reserve Board's Regulation D. In the past, the Federal Reserve has adjusted reserve requirements infrequently and not used them as an active tool in monetary policy implementation. The Fed has not indicated that it is considering adjusting reserve requirements at this time, either. In current circumstances, it would likely take a very substantial rise in reserve requirements to have a material effect on the federal funds interest rate.

Table 2 summarizes how each of the policy tools discussed above could be used to put upward pressure on short-term interest rates. As we discuss in the next section, the Federal Open Market Committee has announced that it plans to rely primarily on the first two channels of influence on interest rates—that is, the Fed

Table 2
The Channels through which Fed Policy Tools Could Put Upward Pressure on Interest Rates

Policy tools	*Encourage arbitrage*	*Increased scope of influence*	*Increase reserve scarcity*
Increase interest on excess reserves rate	✓		
Offer overnight reverse repurchase agreements	✓	✓	✓
Offer term reverse repurchase agreements	✓	✓	✓
Offer term deposits	✓		✓
Sell Fed's securities holdings			✓
Alter reinvestments of Fed's securities holdings			✓
Increase reserve requirements			✓

Source: Authors.

plans to use its policy tools to encourage arbitrage and to increase its scope of influence in short-term money markets.

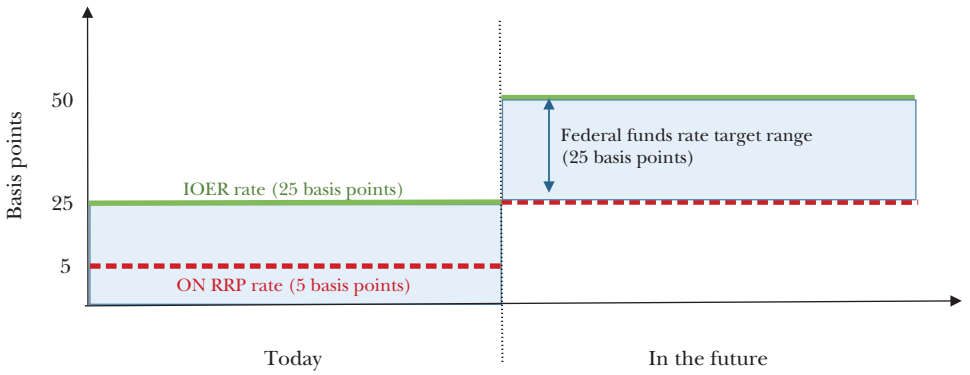
What Is the Fed's Preferred Approach for Raising the Federal Funds Interest Rate?

With the superabundant level of reserve balances in the banking system and with the Federal Open Market Committee having decided not to reverse course and sell off the assets it has purchased, the Fed cannot rely on the reserve scarcity channel to influence short-term interest rates. So what policy tools will the Fed prefer to use when it decides to raise the target range for the federal funds interest rate? The Federal Open Market Committee has formulated and issued plans regarding the approach it intends to take when it decides that the time has come to begin raising short-term interest rates, which are laid out in the "Policy Normalization Principles and Plans" that were issued following the September 2014 Federal Open Market Committee meeting (http://www.federalreserve.gov/newsevents/press/monetary/20140917c.htm) and augmented in the minutes of the Committee's March 2015 meeting (http://www.federalreserve.gov/monetarypolicy/files/fomcminutes20150318.pdf). In addition, Potter (2015) provides an explanation for how what the Fed calls policy "normalization" will be implemented.

When the Federal Open Market Committee decides that economic conditions and the outlook warrant, it will tighten the stance of monetary policy by raising the target range for the federal funds rate, a step that some have nicknamed interest rate "liftoff." Framing monetary policy in terms of the federal funds rate has the advantage that it focuses Fed monetary policy communications on the same policy

Figure 6

The Target Range for the Federal Funds Interest Rate and the Fed's Administered Rates



Source: Authors.
Note: IOER = interest on excess reserves; ON RRP = overnight reverse repurchase agreements.

interest rate as in the past. The Committee plans to set a target range for the federal funds rate that is 25 basis points wide, as it has done since December 2008.

The primary policy tool that the Federal Open Market Committee plans to use to move the federal funds rate into its new target range is the interest rate on excess reserves; at liftoff, this interest rate will be set to the top of the new federal funds target range. As described above, increases in the interest rate on excess reserves will help to pull the federal funds rate and other short-term market interest rates into the target range via arbitrage.

The plan is for an overnight reverse repurchase facility to be used as a supplementary tool to help to push money market interest rates up from below both by encouraging arbitrage and by having an increased scope of influence in money markets: remember, the Fed can undertake its overnight reverse repurchase operations with a different set of money market participants than are eligible to earn the interest rate on excess reserves. At liftoff, the Fed will set an offering interest rate for overnight reverse repurchase agreements at the bottom of the new target range for the federal funds interest rate.

Figure 6 illustrates how this combination of tools is expected to work. The region on the left represents the position of the target range for the federal funds rate and the Fed's two overnight administered rates as they are set prior to liftoff, with a target range of 0 to 25 basis points (the shaded region), the interest rate on excess reserves rate at 25 basis points (the solid line), and the rate on the Fed's overnight reverse repurchase agreements at 5 basis points (the dotted line). For the sake of this illustration, say that the new target range is 25 to 50 basis points so that the interest rate on excess reserves and the overnight reverse repurchase rate would be increased to 50 basis points and 25 basis points, respectively. The increases

in these two administered rates will act to raise the market federal funds rate along with other short-term market interest rates.[7]

The main concern raised by the Federal Open Market Committee in using an overnight reverse repurchase agreement facility is that a large and persistent program could permanently alter patterns of borrowing and lending in repo markets and money markets as a whole—a concern the Committee has referred to as increasing the Federal Reserve's role or size of its "footprint" in money markets. Keep in mind that the Fed's operations in financial markets before the crisis were generally quite small and were aimed at affecting conditions in the federal funds market, a relatively small market. A large overnight reverse repurchase agreement facility could potentially expand the Federal Reserve's role in financial intermediation and reshape the financial industry over time in ways that are difficult to anticipate in advance. In addition, in times of stress in financial markets, demand for a safe and liquid central bank asset might increase sharply, and the Fed's counterparties could shift cash away from financial and nonfinancial corporations in the private sector and place it at the Fed instead, potentially causing or exacerbating disruptions in the availability of funds in money markets.

To mitigate these concerns, the Federal Open Market Committee plans to use its overnight reverse repurchase agreement facility only to the extent necessary to support short-term interest rates, and it will phase the facility out when it is no longer needed (for further analysis of these issues, see Frost et al. 2015). Nonetheless, balancing the need for keeping control over short-term interest rates against the risks associated with a large overnight reverse repurchase agreement facility, the Committee has determined that when policy normalization commences, the aggregate amount offered through its overnight reverse repurchase agreement facility will be temporarily large in order to help move the federal funds rate into its new target range. The Committee has also said that it expects that it will be appropriate to reduce the capacity of the facility fairly soon after it begins raising interest rates.[8]

[7] The Fed will most likely also raise the primary credit rate when it begins raising short-term interest rates. Recall that the primary credit rate is the interest rate at which banks can borrow reserves overnight from the Fed. Since early 2010, the primary credit rate has been set at 75 basis points, 50 basis points above the top of the current range for the target federal funds rate. Given that reserves are now superabundant and will remain so for some time, banks generally will not need to borrow from the Fed and so are unlikely to be influenced by the level of the primary credit rate. In addition, the reputational costs (or "stigma") associated with borrowing from the Fed are likely much higher than was the case prior to the financial crisis in part because the Fed is now required to release information about such borrowing to the public, albeit with a lag.

[8] See the minutes of the Federal Open Market Committee's March 2015 meeting for a discussion of the options the Committee considered for setting the aggregate size of the overnight reverse repurchase agreement facility in the early stages of the policy normalization process (http://www.federalreserve.gov/monetarypolicy/files/fomcminutes20150318.pdf). Details of the Committee's discussion of the footprint and financial stability issues associated with an overnight reverse repurchase agreement facility are available in the minutes of the Committee's April 2014, June 2014, and July 2014 meetings (found here, respectively: http://www.federalreserve.gov/monetarypolicy/files/fomcminutes20140430.pdf; http://www.federalreserve.gov/monetarypolicy/files/fomcminutes20140618.pdf; and http://www.federalreserve.gov/monetarypolicy/files/fomcminutes20140730.pdf).

The 25-basis-point spread between these two administered overnight rates at liftoff is also related to concerns about a persistently large overnight reverse repurchase agreement program. A spread of this size is expected to be narrow enough to allow sufficient control over short-term market interest rates, but wide enough to keep the overnight reverse repurchase agreement facility from becoming so attractive to financial market investors that its potential size invokes concerns about the size of the Fed's footprint in money markets or poses risks to financial stability.

The Federal Open Market Committee has said that it may use other supplementary tools if necessary, which would possibly include term reverse repurchase operations and term deposits. In addition, as noted above, the Fed is planning at some point to allow its securities holdings to decrease gradually by ceasing or reducing reinvestments of principal and interest from its securities holdings. Of course, economic and financial developments will continue to evolve, and the Fed will adjust the details of its approach to policy implementation accordingly.

Conclusion

After the Federal Reserve had used its conventional tools of open market operations to reduce the federal funds rate to near-zero in December 2008, the Fed turned to purchasing securities in the open market as a policy tool for putting downward pressure on longer-term interest rates. These purchases resulted in a superabundant level of reserve balances in the banking system. One legacy of that decision is that the traditional approach to raising the federal funds rate, which was used for many years prior to the financial crisis, will no longer work.

We have described the toolkit available to Federal Reserve policymakers and reviewed the preferred approach of the Federal Open Market Committee to raising the federal funds rate. Testing of the policy tools suggests that the proposed approach should work well. Of course, after policymakers decide that the time has come to begin raising short-term interest rates, they will be vigilant in using their available tools to adjust their approach, as needed, to ensure appropriate control over the federal funds rate and other short-term interest rates.

■ *We thank Miguel Acosta, Jim Clouse, Gordon Hanson, Melanie Josselyn, Joe Kachovec, Lorie Logan, Steve Meyer, Ben Miller, Enrico Moretti, Bill Nelson, and Timothy Taylor for comments. Melanie Josselyn and Joe Kachovec provided excellent research assistance. The views expressed here are those of the authors and do not necessarily reflect the views of other members of the research staff, the Board of Governors of the Federal Reserve System, or the Federal Reserve System.*

References

Carpenter, Seth, Jane Ihrig, Elizabeth Klee, Daniel Quinn, and Alexander Boote. 2015. "The Federal Reserve's Balance Sheet and Earnings: A Primer and Projections." *International Journal of Central Banking* 11(2): 237–83.

Copeland, Adam, Darrell Duffie, Antoine Martin, and Susan McLaughlin. 2012. "Key Mechanics of the U.S. Tri-Party Repo Market." *Economic Policy Review* Federal Reserve Bank of New York 18(3): 17–28.

Federal Reserve System, Board of Governors. 2005. *Purposes & Functions,* 9th edition. Washington DC: Board of Governors of the Federal Reserve System. http://www.federalreserve.gov/pf/pf.htm.

Fischer, Stanley. 2015. "Conducting Monetary Policy with a Large Balance Sheet." Remarks at the 2015 U.S. Monetary Policy Forum, sponsored by the University of Chicago Booth School of Business, New York, New York, February 27. http://www.federalreserve.gov/newsevents/speech/fischer20150227a.htm.

Frost, Josh, Lorie Logan, Antoine Martin, Patrick McCabe, Fabio Natalucci, and Julie Remache. 2015. "Overnight RRP Operations as a Monetary Policy Tool: Some Design Considerations." Finance and Economics Discussion Series 2015-010, Board of Governors of the Federal Reserve System. http://dx.doi.org/10.17016/FEDS.2015.010.

Goodfriend, Marvin. 2015. "The Fed Should Fix the Interest on Reserves Floor." Prepared for the Shadow Open Market Committee Meeting, New York, New York, March 20. http://shadowfed.org/wp-content/uploads/2015/03/GoodfriendSOMC-March2015.pdf.

Ihrig, Jane, Elizabeth Klee, Canlin Li, Brett Schulte, and Min Wei. 2012. "Expectations about the Federal Reserve's Balance Sheet and the Term Structure of Interest Rates." Finance and Economics Discussion Series no. 2012-57, Board of Governors of the Federal Reserve System, July.

Ihrig, Jane E., Ellen E. Meade, and Gretchen C. Weinbach. 2015. "Monetary Policy 101: A Primer on the Fed's Changing Approach to Policy Implementation." Finance and Economics Discussion Series no. 2015-47, Board of Governors of the Federal Reserve System, June.

Judson, Ruth A., and Elizabeth Klee. 2010. "Whither the Liquidity Effect: The Impact of Federal Reserve Open Market Operations in Recent Years." *Journal of Macroeconomics* 32(3): 713–31.

Krishnamurthy, Arvind, and Annette Vissing-Jorgensen. 2011. "The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy." *Brookings Papers on Economic Activity*, Fall, 43(2): 215–87.

Potter, Simon. 2015. "Money Markets and Monetary Policy Normalization." Speech prepared for the Money Marketeers, New York University, April 15, http://www.ny.frb.org/newsevents/speeches/2015/pot150415.html.

Journal of Economic Perspectives—Volume 29, Number 4—Fall 2015—Pages 199–226

Household Surveys in Crisis[†]

Bruce D. Meyer, Wallace K. C. Mok, and James X. Sullivan

Large and nationally representative surveys are arguably among the most important innovations in social science research of the last century. As the leadership of the Committee on National Statistics of the National Academy of Sciences wrote: "It is not an exaggeration to say that large-scale probability surveys were the 20th-century answer to the need for wider, deeper, quicker, better, cheaper, more relevant, and less burdensome official statistics" (Brown, Citro, House, Marton, and Mackie 2014). Household surveys are the source of official rates of unemployment, poverty, health insurance coverage, inflation, and other statistics that guide policy. They are also a primary source of data for economic research and are used to allocate government funds.

However, the quality of data from household surveys is in decline. Households have become increasingly less likely to answer surveys at all, which is the problem of *unit nonresponse*. Those that respond are less likely to answer certain questions, which is the problem of *item nonresponse*. When households do provide answers, they are less likely to be accurate, which is the problem of *measurement error*. We will

■ *Bruce D. Meyer is the McCormick Foundation Professor, Irving B. Harris Graduate School of Public Policy Studies, University of Chicago, Chicago, Illinois. Wallace K. C. Mok is Assistant Professor of Economics, Chinese University of Hong Kong, Hong Kong. James X. Sullivan is the Rev. Thomas J. McDonagh, C.S.C., Associate Professor of Economics, University of Notre Dame, Notre Dame, Indiana. Their email addresses are bdmeyer@uchicago.edu, wallacemok@cuhk.edu.hk, and sullivan.197@nd.edu.*

[†] For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.29.4.199 doi=10.1257/jep.29.4.199

document a noticeable rise in all three threats to survey quality in many of the most important datasets for social science research and government policy.[1]

Of course, if nonresponse arises randomly across the population, survey data would still lead to unbiased estimates of distributions. Thus, we also investigate what is known about the extent to which these problems create bias. However, it can be difficult to verify that nonresponse is independent of survey measures of interest. After all, we typically have very limited information on the characteristics of those who do not respond. The fundamental problem in assessing survey bias due to these problems is the lack of a benchmark measure of the true outcome.

One productive approach to measuring the degree of bias in household surveys, along with addressing potential bias, is to compare survey results with administrative data. In this paper, we focus on the accuracy of survey reporting of government transfers because reliable benchmarks for these programs exist from both aggregate and micro-level administrative data. For example, aggregate administrative data for Temporary Assistance for Needy Families (TANF) are available in annual reports provided by the Administration for Children and Families within the US Department of Health and Human Services, and total Supplemental Nutrition Assistance Program (SNAP) payments are available from the Food, and Nutrition Service of the US Department of Agriculture. Micro-level administrative data for these same program would typically come from a state welfare agency and take the form of records of monthly benefit payment amounts to individuals along with some information on each individual's characteristics. Another advantage of focusing on transfers is that the questions about these programs are often clear and comparable in surveys and administrative sources. We examine the quality of household survey data through comparisons with administrative data from nine large programs that receive considerable attention from both the research and policy community. For example, we compare the total dollar value of food stamp benefits reported, by all respondents in a survey to the total dollar value of food stamp benefits awarded as recorded in US Department of Agriculture, Food and Nutrition Service administrative data.

Our results show a sharp rise in the downward bias in household survey estimates of receipt rates and dollars received for most programs. In recent years, more than half of welfare dollars and nearly half of food stamp dollars have been missed in several major surveys. In particular, this measurement error typically takes the form of underreporting resulting from true program recipients being recorded as nonrecipients. (Throughout this paper we use underreporting as a synonym for understatement or underrecording, since it is likely due to errors by both interviewers and interviewees.) We argue that although all three threats to survey quality are important, in the case of transfer program reporting and amounts, measurement error rather than unit nonresponse or item nonresponse appears to contribute the most bias.

[1] In certain cases, additional measurement issues like coverage error and sampling error will be important. See Groves (2004) and Alwin (2007) for an exhaustive list of types of survey errors.

The underreporting of transfer income in surveys has profound implications for our understanding of the low-income population and the effect of government programs for the poor. We point to evidence from linked administrative and survey data that indicates that this underreporting leads to an understatement of incomes at the bottom, the rate of program receipt, and the poverty-reducing effects of government programs—and thus to an overstatement of poverty and inequality.

The evidence on declining survey quality we present here is likely not unique to transfer income. While evidence comparing other survey variables to administrative benchmarks is scarce, there is evidence suggesting that survey biases in self-employment and pension income, education, pension contributions, and some categories of expenditures have also risen.

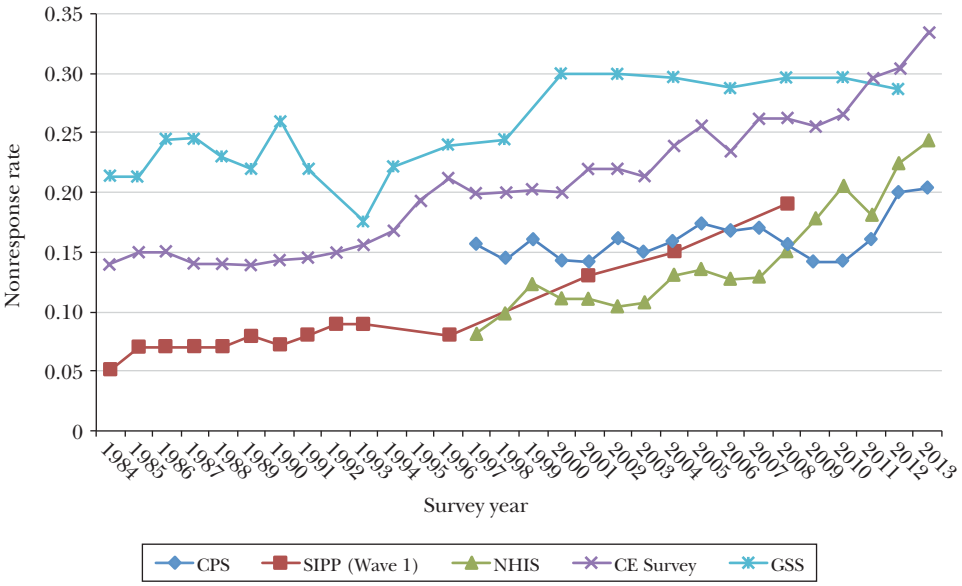
Our results call for more research into why survey quality has declined. Our preferred explanation is that households are overburdened by surveys, leading to a decline in many measures of survey cooperation and quality. The number and breadth of government surveys rose sharply between 1984 and 2004 (Presser and McCulloch 2011), and the number of private surveys has been rising as well. We discuss the limited evidence concerning some alternative explanations including increasing concerns about privacy, a decline in public spirit, less leisure time, or the stigmatizing effect of giving certain answers to questions. We conclude by noting the need for research on ways to improve the quality of household surveys. In particular, more frequent linking of survey data with administrative microdata provides one potentially fruitful avenue for improving the quality of survey data.

Rising Unit Nonresponse Rates

Unit nonresponse, which occurs when a household in a sampling frame is not interviewed at all, has been rising in most surveys. Unit nonresponse rates rose by 3–12 percentage points over the 1990s for six US Census Bureau surveys (Atrostic, Bates, Burt, and Silberstein 2001). In non-Census surveys, the rise in unit nonresponse is also evident, and in some cases even sharper (Steeh, Kirgis, Cannon, and DeWitt 2001; Curtin, Presser, and Singer 2005; Battaglia, Khare, Frankel, Murray, Buckley, and Peritz 2007; Brick and Williams 2013). The National Research Council (2013) provides a thorough summary for US surveys, but the pattern is apparent in surveys in other countries as well (de Leeuw and de Heer 2002).

Indeed, the problem of rising unit nonresponse in major surveys has been a heavily discussed topic in the survey research community. Unit nonresponse was the subject of two National Research Council reports and a special issue of a major social science journal (National Research Council 2011, 2013; Massey and Tourangeau 2013). The Office of Management and Budget (2006) has set a target response rate for federal censuses and surveys, and recommends analysis of nonresponse bias when the unit response rate is less than 80 percent. At least one influential journal, the *Journal of the American Medical Association*, restricts publication of research using low-response-rate surveys (Davern 2013).

Figure 1
Unit Nonresponse Rates of Major Household Surveys



Sources: For the Current Population Survey, see Appendix G of US Census Bureau (various years (a)). For the Survey of Income and Program Participation, see Source and Accuracy Statement of US Census Bureau (various years (b)). For the National Health Interview Survey, see Table 1 of US Department of Health and Human Services (2014). For the Consumer Expenditure Survey, see US Department of Labor (various years). For the General Social Survey, see Table A.6 of "Appendix A: Sampling Design and Weighting," in Smith, Marsden, Hout, and Kim (2013).
Note: We report the unit nonresponse rate for five prominent household surveys during the 1984–2013 period: the Current Population Survey Annual Demographic File/Annual Social and Economic Supplement (CPS), the Survey of Income and Program Participation (SIPP), the Consumer Expenditure (CE) Survey, the National Health Interview Survey (NHIS), and the General Social Survey (GSS).

In Figure 1, we report the unit nonresponse rate for five prominent household surveys during the 1984–2013 period: the Current Population Survey Annual Demographic File/Annual Social and Economic Supplement (CPS), which is the source of the official US poverty rate and income distribution statistics; the Survey of Income and Program Participation (SIPP), which is the best source for the information needed to determine eligibility for and receipt of transfers; the Consumer Expenditure (CE) Survey, which is the main source of data on consumption and provides the weights that are put on price changes when calculating inflation as measured by the Consumer Price Index; the National Health Interview Survey (NHIS), which is the primary source for information on the health status of the US population; and the General Social Survey (GSS), which may be the most used dataset across the social sciences for social and attitudinal information. For other analyses in this paper, we also examine nonresponse rates in the American Community Survey (ACS), which replaced the Census long form, providing detailed

small-area information annually, and the Panel Study of Income Dynamics (PSID), which is the longest-running longitudinal survey (and thus allows tracking specific households over time).

The surveys in Figure 1 show a pronounced increase in unit nonresponse over time, in recent years reaching 16 to 33 percent. Between 1997 and 2013, the unit nonresponse rate in the Current Population Survey rose from 16 to 20 percent while the rate in the National Health Interview Survey rose from 8 to 24 percent.[2] The National Research Council (2013) reports a general decline in response rates for a long list of surveys. The decline in response rates seems to be even more pronounced for public opinion surveys (Pew 2012). Interestingly, response rates appear to be much higher for surveys in developing countries. Mishra, Hong, and Khan (2008) report that recent demographic and health surveys from 14 different African countries all had unit response rates above 92 percent.

One of the few notable exceptions to high nonresponse rates for domestic surveys is the American Community Survey. The ACS's low survey nonresponse rate (about 3 percent in recent years) is due in large part to the fact that the survey is mandatory. A US Census Bureau (2003) study showed that a change to a voluntary standard for the ACS led to a rise in nonresponse rates to the mail version of the survey of more than 20 percentage points. The ACS also contacts potential respondents through multiple modes including mail, telephone, and personal visits. Only about 66 percent of households in most years respond to the initial two modes of contact. A random subsample of nonrespondents at that point is selected for a personal home visit and their responses are given extra weight. Those not selected are given a weight of zero, so they do not contribute to the overall nonresponse rate. The ACS can affect community funding (Reamer 2010), which may also in part account for why the nonresponse rate is so low.

Of the problems with surveys, rising unit nonresponse has gotten the most attention. This emphasis is not surprising given that it is widespread, often easy to measure, and increases survey costs. However, the rate of unit nonresponse is not particularly informative about the accuracy of statistics from a survey. Unit nonresponse only leads to bias if it is nonrandom, with the exact requirement depending on the statistic in question and the weighting method. However, exploring whether unit nonresponse is random can be difficult because researchers typically have only limited information on the characteristics of nonresponders. Even if nonresponders look like responders based on a limited set of characteristics—say, age and geography—this does not mean that these groups are similar along other dimensions, such as willingness to participate in government programs. Evidence on the extent to which unit nonresponse leads to bias differs by survey and question. While there are examples of substantial bias, in other cases the resulting bias is small or can be mitigated by appropriate weighting—certain demographic variables in the survey are weighted to correspond

[2] Regression estimates of a linear time trend over the available years yield a positive coefficient on year for each of the surveys that is strongly significantly different from zero in four of the five cases, and weakly significant in the remaining case. For details, see Online Appendix Table 1.

to the total population (National Research Council 2013, pp. 42–43). Even in public opinion surveys with response rates under 10 percent, researchers have argued that properly weighted responses are largely representative (Pew 2012). In their survey of bias estimates, Groves and Peytcheva (2008) found that bias magnitudes differed more across statistics (such as mean age or gender) within a survey than they did across surveys. Indeed, with the possibilities for weighting adjustments taken into account, unit nonresponse is probably not the main threat to the quality of household survey data.

Several methods have been proposed for improving unit nonresponse—for example, advance notification of the survey through the mail, increasing the number of times the potential respondent is contacted, improving the training of interviewers, or offering financial incentives for participation—but the evidence suggests only small effects from such efforts (National Research Council 2011). Even when such efforts increase response rates, they do not necessarily lead to a reduction in bias. Indeed, if they mainly encourage the groups that are already overrepresented in the survey or who are unmotivated to cooperate, they can even make the bias worse (Groves 2006; Groves and Peytcheva 2008; Tourangeau, Groves, and Redline 2010; Peytchev 2013; Kreuter, Müller, and Trappmann 2014). Also, a tradeoff can arise between different measures of survey accuracy: inducing participation from those who are initially reluctant to complete a survey may lead to greater problems with item nonresponse (missing answers to questions) or measurement error (inaccurate answers).

Rising Item Nonresponse

Even if a household agrees to participate in a survey, responses to key questions may not be obtained due to refusal or inability to answer, or failure of the interviewer to record the response. This item nonresponse is distinct from a respondent misreporting that he did not receive a certain type of transfer income, which would be considered measurement error. Most surveys (and all of those that we examine) typically impute a response in these cases of missing data. Many methods are used to impute, though the Census "hot-deck" procedure, where a missing value is imputed from a randomly selected similar record, is probably the most common (for more information, see Andridge and Little 2010). Surveys impute responses for all sorts of questions, including those related to demographic characteristics such as age and education, employment, and income. Nonresponse rates are typically low for most questions (Bollinger and Hirsch 2006), but they can be quite high for questions related to labor and nonlabor income. For transfer programs, surveys may impute "recipiency" (that is, whether or not a person received a given type of benefit at all) as well as the dollars received or the months of benefits received.

As evidence of the extent of item nonresponse and how it has changed over time, we present imputation rates for survey questions on receipt of transfer income. We calculate the share of dollars recorded in two major household surveys that is imputed for six large transfer programs that all receive considerable attention

from both the research and policy community: Aid to Families with Dependent Children/Temporary Assistance for Needy Families (AFDC/TANF); the Food Stamp Program/Supplemental Nutrition Assistance Program (FSP/SNAP); Supplemental Security Income (SSI); Social Security (OASDI) including both retirement and disability benefits; Unemployment Insurance (UI); and Workers' Compensation (WC). These large national programs provide benefits to tens of millions of individuals—together they distributed almost $1 trillion in 2011. We present the imputation shares for the Current Population Survey in Figure 2A and for the Survey of Income and Program Participation in Figure 2B. These two surveys focus on income and program receipt, and they are a good indicator of the state of the art in survey collection over time. Although not reported here, we have also calculated similar imputation rates for the American Community Survey, the Consumer Expenditure Survey, and the Panel Study of Income Dynamics (Meyer, Mok, and Sullivan 2015).

The imputation rates are quite high, averaging about 25 percent. In 2013, the imputation shares in the Current Population Survey ranged from 24 percent of dollars recorded from Temporary Assistance for Needy Families and the Supplemental Nutrition Assistance Program to 36 percent of Social Security dollars. Overall, the Survey of Income and Program Participation has noticeably higher imputation rates than the Current Population Survey.[3]

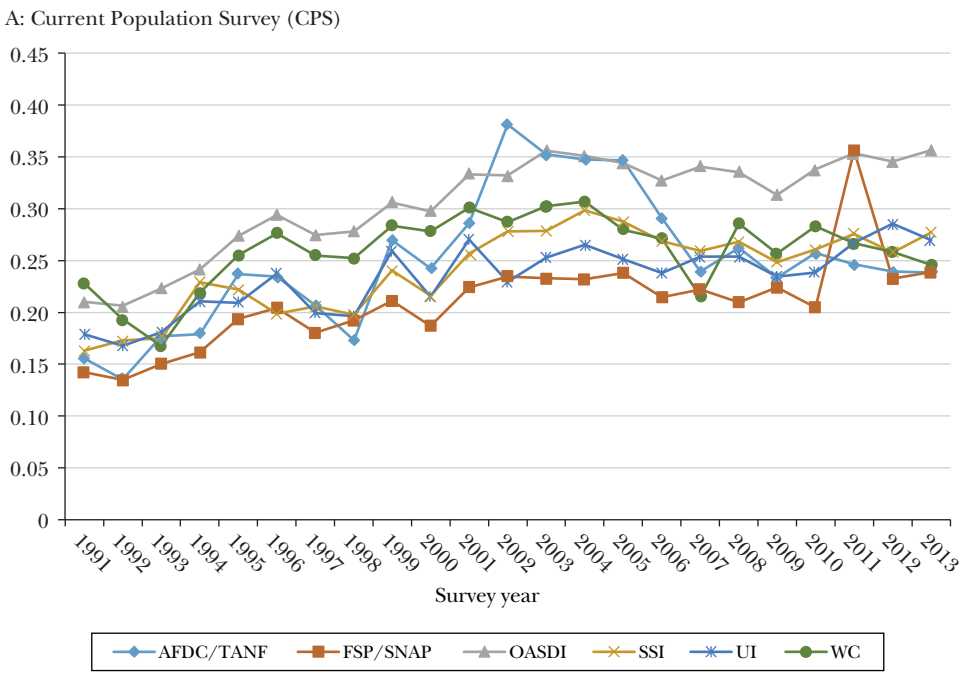
Figures 2A and B also show an increase in imputation rates over the past two and a half decades. This rise is evident in all programs in both the Current Population Survey and Survey of Income and Program Participation. The estimates suggest, for example, that for AFDC/TANF, in the CPS the fraction of dollars imputed is rising by 0.4 percentage points each year.[4] The imputation rates for months of receipt (not reported) are similar to those for dollars reported here. In recent years, at least 10 percent of months are imputed in the CPS for all four programs for which we have months. For the SIPP, month imputation shares are sometimes below 10 percent, but are more typically between 10 and 20 percent. The shares have generally risen over time.

Transfer income may be imputed either when there is missing information on whether the household receives income from a given program, or when there is missing information on the number of dollars received. We have also calculated the

[3] Imputation procedures in the Survey of Income and Program Participation take advantage of information collected in previous waves. For example, beginning with the 1996 panel, missing data were imputed by using the respondent's data in the previous wave (if available). Starting with wave 2 of the 2004 panel, the SIPP began to use "dependent interviewing" in which the interviewers use information from the prior wave to tackle item nonresponse during the actual interview. For the results in Figure 2B and Table 1 we do not include values imputed from prior wave information in our calculation of total dollars imputed. See Meyer, Mok, and Sullivan (2015), Chapter 4 of US Census Bureau (2001), and Pennell (1993) for more information.

[4] We summarize the trends by regressing the imputation share on a constant and a time trend separately for each program and survey. As shown in Online Appendix Table 2, for all six programs and both surveys the coefficient on the time trend is positive. In the case of the Current Population Survey, the upward trend is statistically significant at the 1 percent level for four of the six programs, while the trend is significant in the Survey of Income and Program Participation for five of six programs.

Figure 2
Item Nonresponse Rates in Two Surveys for Various Transfer Programs Calculated as Share of Dollars Reported in the Survey that Is Imputed

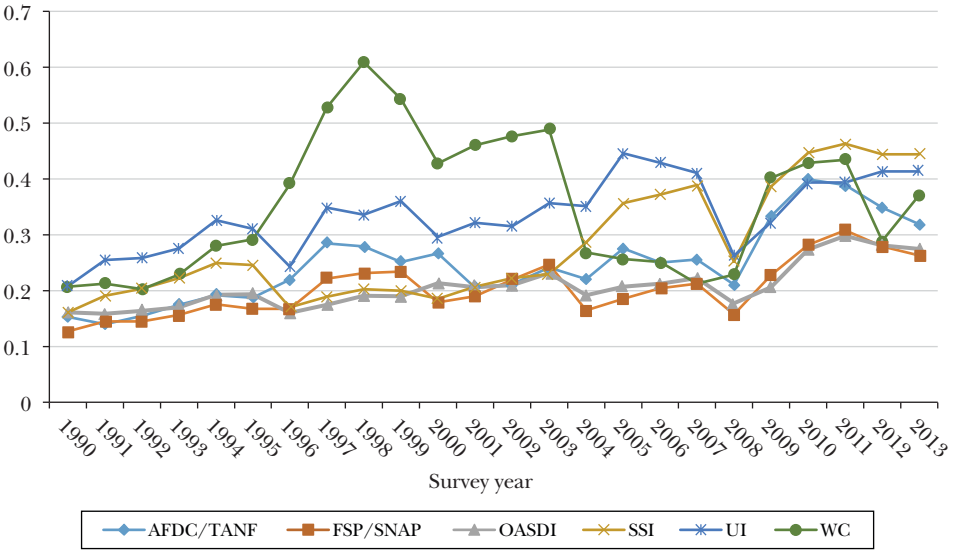


share of total dollars reported attributable to those whose recipiency was imputed. In the Current Population Survey and the Survey of Income and Program Participation this share, is typically on the order of 10 percent but is frequently higher for programs covered by these surveys. There is substantial variation across programs and over time. For most of the years since 2000, recipiency imputation exceeds 20 percent for AFDC/TANF. The rise in recipiency imputation over time is less pronounced than that for overall imputation, which includes not only recipiency imputation but also imputation of dollar amounts when receipt is reported but the dollar amount is not.

While imputation might improve quality assessments based on comparisons to aggregate amounts paid out, there are important limitations associated with imputed values. Studies using linked survey and administrative data show that the rates of false positive and false negative reporting of receipt are almost always much higher among the imputed observations than the nonimputed ones (Meyer, Goerge, and Mittag 2014; Celhay, Meyer, and Mittag 2015). Also, even in cases where imputations may reduce bias for statistics such as the mean, they may create greater bias in other statistics such as measures of inequality or covariances. For example, Bollinger and Hirsh (2006) show that including imputed values of earnings leads to biased coefficients in regressions of earnings on other attributes.

Figure 2 (continued)

B: Survey of Income and Program Participation (SIPP)



Notes: We calculate the share of dollars recorded in the Current Population Survey (CPS) (Figure 2A) and the Survey of Income and Program Participation (SIPP) (Figure 2B) that is imputed, for six large transfer programs: Aid to Families with Dependent Children/Temporary Assistance for Needy Families (AFDC/TANF), the Food Stamp Program/Supplemental Nutrition Assistance Program (FSP/SNAP), Supplemental Security Income (SSI), and Social Security (OASDI) including both retirement and disability benefits, Unemployment Insurance (UI), and Workers' Compensation (WC). For the SIPP, Figure 2B, the share of dollars reported that is imputed excludes imputation using previous wave information.

Measurement Error and Estimates of Bias

Inaccurate responses given that there is a response are called measurement error, and can contribute to bias (the difference between an estimate and the true value) in common statistics calculated from survey data. One way to test for measurement error is to link survey data on the payments that individuals or households say they have received with administrative microdata on the amounts actually provided to each household. Comparisons to administrative microdata on program receipt have been fairly limited in the literature. This approach has often been restricted to a single state, year, program, and dataset (Taeuber, Resnick, Love, Stavely, Wilde, and Larson 2004). Examples of studies that examine more than one program (but still a single dataset) include Moore, Marquis, and Bogen (1996), Sears and Rupp (2003), and Huynh, Rupp, and Sears (2002). A review of earlier studies can be found in Bound, Brown, and Mathiowetz (2001).

An alternative approach to comparisons to administrative microdata is to compare aggregate survey and administrative data. Comparisons to administrative aggregates have been used widely, but results are only available for a few years, for a few transfer

programs, and for some of the key datasets. Important papers include Duncan and Hill (1989), Coder and Scoon-Rogers (1996), Roemer (2000), and Wheaton (2007). These papers tend to find substantial underreporting that varies across programs.

To provide a more comprehensive look at the magnitude of measurement error in many surveys, across many years, and for several programs, we compare aggregate survey and aggregate administrative data. The aggregate administrative data are available through a variety of reports from government agencies. A complete list of sources for these administrative aggregates is available in Meyer, Mok, and Sullivan (2015). The administrative aggregate data that we use have been audited, so we expect bias in these data to be small. There might be some bias that results from different coverage or variable definitions between the administrative aggregates and the survey aggregates, but we make considerable effort to ensure these align closely.

Since administrative data sources are heterogeneous, they should not always be taken as accurate. For example, Abowd and Stinson (2013) model administrative and survey measures of earnings, treating both sources as error-ridden. Their approach is driven in part by conceptual differences between the survey and administrative measures of earnings that mean that there is not a single true earnings measure. This problem is not present when one examines transfers, because the survey and administrative concepts are generally the same.

Through comparisons to aggregate administrative data, we show that survey measures of whether an individual receives income and of how much income is received from major transfer programs are both sharply biased downward, and this bias has risen over time. Although these measures of bias include all three threats to survey quality—unit nonresponse, item nonresponse, and measurement error—in the following section we argue that the bias is largely due to measurement error.

Here, we will focus on two statistics, the mean receipt of certain transfer programs measured in dollars, and the mean receipt measured in months of receipt. Means of transfer receipt are important statistics. Reports of these means affect distributional calculations of inequality and poverty; calculations of the effects of programs on the income distribution; and estimates of what share of those eligible for a certain program receive support from the program. Our analyses focus on how underreporting has changed over time. For a more extensive discussion of how these findings differ across programs and datasets, see Meyer, Mok, and Sullivan (2015).

Below we report estimates of the proportional bias in dollar reporting, which we call Dollar Bias, and in month reporting, which we call Month Bias. These biases can be defined as the net reporting rate minus 1, or more specifically,

$$\textit{Dollar Bias} = \frac{\text{dollars reported in survey, population weighted}}{\text{dollars reported in administrative data}} - 1$$

and

$$\textit{Month Bias} = \frac{\text{months reported in survey, population weighted}}{\text{months reported in administrative data}} - 1.$$

These expressions give us the proportional bias in the mean and therefore can be thought of as the proportional bias in the total dollars or months, or in per person dollars or months. Also note that the reporting rates in the above definitions are net rates reflecting underreporting by true recipients counterbalanced to some extent by overreporting by recipients and nonrecipients.

We calculate the bias in the mean receipt of transfer dollars for the same programs for which we reported imputation rates above, only now we are able to divide Old-Age, Survivors, and Disability Insurance (OASDI) into its retirement (OASI) and disability (SSDI) components.[5] We also calculate months-of-receipt reporting biases for seven programs. Months of receipt are not available in all cases, including Unemployment Insurance and Workers' Compensation, but they are available for some programs for which we do not observe dollars, including the National School Lunch Program (NSLP) and the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC). We do this for as many individual years as are available for five of the most important datasets for analyzing income, its distribution, and receipt of transfers: the Current Population Survey, the Survey of Income and Program Participation, the American Community Survey, the Consumer Expenditure Survey, and the Panel Study of Income Dynamics. If these datasets are understating transfers received in a substantial way—and we will show that they are—it has important implications for our understanding of the economic circumstances of the population and the effects of government programs. We should emphasize that all of the bias estimates we report include imputed values in the survey totals, so the bias calculated here understates the measurement problems. To put it another way, by providing values for households that do not report receipt of transfer income, imputations may lead to smaller estimates of bias in our approach even though these imputations introduce considerable measurement error due to the inaccuracy of imputed values.

The top panel of Table 1 presents the average Dollar Bias over the 2000–2012 period for seven programs from five household surveys. With the single exception of Supplemental Security Income in the Survey of Income and Program Participation, the bias is negative, indicating underreporting of dollars of transfer income. The upward bias in reporting of SSI appears to be due to confusion among recipients between SSI, a Social Security Administration program aimed at low-income people who are blind, disabled, or elderly, and Old-Age and Survivors Insurance (OASI), which is what most people mean by Social Security (Huyhn, Rupp, and Sears 2002; Gathright and Crabb 2014). In most cases, the bias reported in Table 1 is large. For our main cash welfare programs, Temporary Aid to Needy Families (combined with General Assistance in two cases), four of five surveys have a bias of 50 percent or more, meaning that less than half of the dollars given out are

[5] In several of the datasets, Social Security Disability Insurance benefits are in some cases combined with Social Security Retirement and Survivors benefits. To separate these programs, we use data from the Social Security Bulletin (US Social Security Administration, various years) to calculate for each year, age, school status, and gender cell, the proportions of total Social Security dollars that are paid to OASI and SSDI recipients. See Meyer, Mok, and Sullivan (2015) for more details on methodology.

Table 1

Proportional Bias in Survey Estimates of Mean Program Dollars and Months Received, by Program and Survey, 2000–2012

	AFDC/ TANF	*FSP/ SNAP*	*Social Security*		*SSI*	*UI*	*WC*	*NSLP*	*WIC*
			OASI	*SSDI*					
Dollars									
ACS	−0.519	−0.458	−0.165	−0.299	−0.046				
CE	−0.767	−0.587	−0.149	−0.214	−0.283	−0.583	−0.618		
CPS	−0.500	−0.417	−0.086	−0.187	−0.162	−0.325	−0.541		
PSID	−0.619	−0.308	−0.086	−0.176	−0.322	−0.360	−0.646		
SIPP	−0.357	−0.170	−0.070	−0.146	0.164	−0.388	−0.651		
Months									
ACS			−0.154	−0.261	−0.372				
CPS	−0.453	−0.422	−0.147	−0.154	−0.397			−0.503	−0.341
PSID	−0.574	−0.297	−0.114	−0.121	−0.502			−0.470	−0.180
SIPP	−0.232	−0.165	−0.008	0.041	0.023			0.141	−0.197

Notes: Each cell reports the average dollars/months proportional bias for the specified program and survey in the 2000–2012 period. The transfer programs are: Aid to Families with Dependent Children/Temporary Assistance for Needy Families (AFDC/TANF), which is combined with General Assistance in the case of the Consumer Expenditure Survey and the American Community Survey; the Food Stamp Program/Supplemental Nutrition Assistance Program (FSP/SNAP); Social Security, including Old-Age and Survivors Insurance (OASI) and Social Security Disability Insurance (SSDI); Supplemental Security Income (SSI); Unemployment Insurance (UI); Workers' Compensation (WC); National School Lunch Program (NSLP); and Women, Infants, and Children (WIC). The surveys are: American Community Survey (ACS), Consumer Expenditure (CE) Survey, Current Population Survey (CPS), Panel Study of Income Dynamics (PSID), and Survey of Income and Program Participation (SIPP).

captured in surveys. Even in the Survey of Income and Program Participation (SIPP), a survey especially designed to capture transfer program income, more than one-third of TANF dollars are missed. For the Food Stamp Program/Supplemental Nutrition Assistance Program (FSP/SNAP), the bias is at least 30 percent for four of the five surveys. The bias in dollar reporting of Unemployment Insurance (UI) and Workers' Compensation (WC) is also pronounced: it is at least 32 percent for UI and 54 percent for WC in all surveys. The Social Security Administration programs (the retirement program OASI, the disability insurance program SSDI, and support for the low-income elderly, blind, and disabled through SSI) have much less bias, which may, in part, be due to the fact that receipt of these programs tends to be more regular or permanent.

The average Month Bias for this same period is reported in the bottom panel of Table 1. These biases are very similar to the corresponding dollar reporting biases in the top panel. In the case of the Food Stamp Program/Supplemental Nutrition Assistance Program (FSP/SNAP), the similarity is striking, with the bias in the two types of reporting never differing by more than 1.1 percentage points for the three datasets. This pattern suggests that individuals report about the right dollar amount on average, conditional on reporting. Or, put another way, most of the bias is due to not reporting at all, rather than reporting too little conditional on reporting. In

Meyer, Mok, and Sullivan (2015), we report several sets of conditions under which equality of the dollar and month bias implies that dollar amounts are reported correctly on average. The Dollar Bias estimates are only slightly larger in absolute value than the Month Bias estimates, suggesting that there is a small amount of underreporting of dollars conditional on receipt, nevertheless. In the case of the CPS and the PSID, the evidence suggests that total dollars and months are understated by similar amounts, again suggesting that conditional on reporting receipt, the monthly benefits are reported about right on average.

For Old-Age and Survivors Insurance (OASI) and Social Security Disability Insurance (SSDI), we see similar biases for monthly receipt and dollar receipt, with the bias for dollar receipt being slightly larger (in absolute value), again suggesting that most of the downward bias results from failure to report receipt rather than underreporting the dollar amount of benefits conditional on reporting receipt. For Supplemental Security Income, the bias for dollar receipt is actually smaller in absolute value than the bias for monthly receipt for each survey other than the Survey of Income and Program Participation, where the bias is larger but positive, all of which suggests some overreporting of dollars conditional on reporting receipt.[6]

The average biases in monthly participation reporting for the National School Lunch Program (NSLP) and for the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) are also reported in the bottom panel of Table 1. Reporting of NSLP months is quite low for both the Panel Study of Income Dynamics (PSID) and the Current Population Survey (CPS), which both have an average bias of about 50 percent. In the Survey of Income and Program Participation (SIPP), on the other hand, the bias is positive, indicating that more months of participation are reported than we see in the administrative data. This result is likely due in part to our assumptions that all eligible family members (ages 5–18) receive lunches and that they do so for all four of the reference months of a given wave of the SIPP. WIC is also underreported significantly. The average bias for monthly WIC receipt in the CPS, PSID, and SIPP ranges from 19 to 34 percent.

This large bias in mean receipt of transfer programs has been increasing over time. Table 2 reports estimates from regressions of annual estimates of the proportional bias in dollar reporting on a constant and a time trend for various years from 1967 to 2012 for the five surveys and seven programs. Most household reports of transfer programs in the Consumer Expenditure (CE) Survey, Current Population Survey (CPS), and Panel Study of Income Dynamics (PSID) show a significant increase in the downward bias—that is, a decline in dollar reporting over time. The downward bias in mean dollars reported of AFDC/TANF in the CPS, for example, increases by about one percentage point each year. The time trends in bias in the Survey of Income

[6] For Old-Age and Survivors Insurance, Social Security Disability Insurance, and Supplemental Security Income, the surveys other than the Survey of Income and Program Participation do not report monthly participation, only annual participation. Since our administrative numbers are for monthly participation, we use the relationship between average monthly and annual participation calculated in the SIPP to adjust the estimates from the other sources. This adjustment step likely induces some error that accounts for the weaker similarity between the bias for monthly and dollar receipt.

Table 2
Trend in Proportional Bias in Mean Dollars Reported in Survey (Including Those Imputed), by Program and Survey

			Social Security				
	AFDC/TANF	*FSP/SNAP*	*OASI*	*SSDI*	*SSI*	*UI*	*WC*
ACS	−0.96		0.08	−0.68	3.50		
	(0.87)		(0.07)	(0.11)***	(1.11)**		
	12		12	12	12		
CE	−1.87	−1.1	0.07	−0.51	0.05	−0.74	−2.33
	(0.43)***	(0.43)**	(0.23)	(0.23)**	(0.27)	(0.19)***	(0.38)***
	33	33	33	33	33	33	33
CPS	−0.71	−0.59	0.20	−0.61	0.41	−0.39	−0.71
	(0.20)***	(0.09)***	(0.02)***	(0.08)***	(0.12)***	(0.19)*	(0.16)***
	37	34	45	45	38	26	25
PSID	−1.04	−0.93	0.40	−0.62	−0.04	−0.47	−0.46
	(0.12)***	(0.27)***	(0.10)***	(0.23)**	(0.26)	(0.16)***	(0.12)***
	36	38	36	36	34	30	30
SIPP	−0.46	−0.06	0.05	−0.33	1.52	−0.45	−0.50
	(0.34)	(0.15)	(0.18)	(0.49)	(0.37)***	(0.22)*	(0.10)***
	29	30	30	30	30	30	29

Notes: For each cell, we report the year coefficient from a regression of the proportional bias in percentages on a constant and year, with its standard error underneath, followed by the sample size, where each observation is a year. The number of years varies across survey and program with as many 45 years for Old-Age and Survivors Insurance (OASI) in the Current Population Survey (CPS) (1967–2012, 1969 missing) and as few as 12 for the ACS (2000–2011). The regressions correct for first-order autocorrelation using the Prais–Winsten procedure. For titles of all the surveys and transfer programs, see the note to Table 1.
***, **, and *, indicate that the coefficient is statistically significantly different from zero at the 1, 5, and 10 percent levels, respectively.

and Program Participation (SIPP) and the American Community Survey (ACS) are less pronounced. The exceptions to the general rise in bias are the programs Old-Age Survivors Insurance (OASI) and Supplemental Security Income (SSI), which have rising reporting rates in most cases. However, in the case of SSI in the SIPP, rising reporting leads to greater bias because the bias is always positive in recent years.[7]

The implication that measurement error in survey responses to government programs has grown over time is consistent with findings from Gathright and Crabb (2014), who calculate measurement error directly by linking Survey of Income and Program Participation data to Social Security Administration data for the

[7] Estimates consistent with those reported in Tables 1 and 2 are available in previous studies for some surveys for a subset of years and programs including: Coder and Scoon-Rogers (1996) for five of our programs for 1984 and 1990 for the Current Population Survey and the Survey of Income and Program Participation; Roemer (2000) for the same five programs for 1990–1996 for the CPS and the SIPP; Wheaton (2007) for four programs between 1993 and 2005 in the CPS and a shorter period in the SIPP; and Duncan and Hill (1989) for the CPS and Panel Study of Income Dynamics for earlier years.

Supplemental Security Income and the Old-Age Survivors and Disability Insurance programs. An added benefit of such linking is that one can identify false positives and false negatives. Their analysis shows that false positive and false negative rates for reported receipt and the mean absolute deviation of the reported benefit amount from the administrative amount increased between the 1996 and 2008 panels of the SIPP for both SSI and OASDI. During this period, the mean absolute error in the benefit amount increased by 70 percent for OASDI and by 60 percent for SSI.

The underreporting of transfer income in surveys has profound implications for our understanding of the low-income population and the effect of government programs for the poor. Accounting for underreporting of receipt, and substantial reporting and imputation error in amounts conditional on correctly reporting receipt, sharply changes what one learns from the survey data. Meyer and Mittag (2015) link data on four transfer programs (SNAP, TANF, General Assistance, and Housing Subsidies) to the New York data from the Current Population Survey over a four-year period from 2008–2011. In the survey data, 43 percent of SNAP recipients and 63 percent of public assistance recipients are not recorded as receiving benefits. Accounting for the survey errors more than doubles the estimate of the income of those who are reported to have income below half the poverty line. It leads the reported poverty rate to fall by 2.5 percentage points for the entire population and over 11 percentage points for single mothers. It nearly doubles the poverty-reducing effect of the four programs overall, and increases it by a factor of over 1.5 for single mothers. The share of single mothers with no earnings or program receipt is cut in half.

Is the declining quality of survey data unique to transfer income? One might argue that potential reasons for declining quality that might be unique to transfer income, such as rising stigma or less recognition of the general program names, make it a special case. But as we argue below, these reasons do not appear to explain the sharp rise in bias that we find.

The evidence on whether measurement error has grown over time for other outcomes is limited, but this evidence suggests the problem of declining survey quality goes well beyond transfer income.[8] Reporting of self-employment income has worsened, but there is no clear trend for wage and salary income and dividends in the Current Population Survey (CPS) and the Survey of Income and Program Participation (Coder and Scoon-Rogers 1996; Roemer 2000). Comparing earnings aggregates from the Survey of Consumer Finances with those computed using IRS Statistics of Income data, Johnson and Moore (2008) find that respondents overreport earnings, and this overreporting has worsened over time. They also find a sharp increase in pension income underreporting. Barrow and Davis (2012) compare reported postsecondary enrollment in the October CPS to Integrated Postsecondary Education Survey administrative data, showing that CPS reporting of type of college attended has gotten worse, though error in reporting of overall enrollment has remained stable. Other studies have shown that measurement

[8] Many studies document substantial bias in levels for other outcomes; for a summary, see Bound, Brown, and Mathiowetz (2001).

error in pension contributions has grown over time (Dushi and Iams 2010). Also, Bee, Meyer, and Sullivan (2015) show that while reporting rates for some of the biggest components of consumption have remained stable over time, there have been noticeable declines for some categories such as food away from home, shoes and clothing, and alcoholic beverages. Future research on changes in bias in other outcomes would be a valuable extension to this literature.

Decomposing the Overall Bias

The bias estimates we present in Tables 1 and 2 are based on aggregate data, and for that reason they reflect not just measurement error but also coverage error (which arises when the sampling frame does not properly represent the underlying population) and error due to unit and item nonresponse. For several reasons, we argue that the most important source of the overall bias is measurement error.

Coverage error could explain some of the significant underreporting we find if the sampling frame for the surveys we examine (typically based on the noninstitutionalized Census population) does not capture the entire population that receives benefits. This argument about underweighting is essentially an argument about individuals being missed in the Census count.[9] Although we do not have undercount data for those who receive transfer income, estimates of the overall Census undercount are small, particularly relative to most of our bias estimates for transfer reports (Hogan 1993; Robinson, Amed, Das Gupta, and Woodrow 1993). Moreover, undercount estimates have declined over time, and the estimates for the 2010 Census actually suggest an overcount (US Census 2012).

There are also reasons to believe that bias resulting from unit and item nonresponse might be small. While unit nonresponse is surely nonrandom with respect to receipt of transfer income, appropriate weighting may offset much of this bias. Similarly, item nonresponse also appears to be nonrandom, but will not lead to bias in mean reports if imputations are on average accurate.

Empirical studies relying on linked administrative and survey microdata support these arguments. Bee, Gathright, and Meyer (2015), for example, show that for income in the Current Population Survey, unit nonresponse leads to remarkably little bias in the distribution of income. The estimates of bias from studies linking survey and administrative microdata that are most comparable to ours using aggregate data come from Marquis and Moore (1990), which, we should point out, relies on survey data from 30 years ago. Their bias estimates for months of receipt are reported in the first column of Table 3. The second column reports our estimated

[9] We discuss issues related to the institutionalized population that receives transfers in the following section. As a check, for each survey and year, we have confirmed that our weighted population totals are close to Census population estimates. The sample weights in the Panel Study of Income Dynamics are not appropriate for weighting to the complete population in some years. We adjust them in a manner suggested by the PSID staff, and the appendix to Meyer, Mok, and Sullivan (2015) provides details.

Table 3
Proportional Bias Estimates from Microdata and Aggregate Data Compared

Transfer program	*Microdata bias estimate due to unit nonresponse and measurement error*	*Aggregate data bias estimate due to all sources of error*
Aid to Families with Dependent Children (AFDC)	−0.39	−0.21
Food Stamp Program (FSP)	−0.13	−0.15
Old-Age, Survivors, and Disability Insurance (OASDI)	0.01	−0.06
Supplemental Security Income (SSI)	−0.12	−0.14

Notes: The estimates using microdata are from Marquis and Moore (1990) and use data from the Survey of Income and Program Participation (SIPP) over June 1983 to May 1984 for months of receipt in Florida, New York (OASDI and SSI only), Pennsylvania, and Wisconsin. The aggregate data we use for the estimates in column 2 are averages of 1983 and 1984 average monthly participation for the entire United States. We also assume Old-Age, Survivors, and Disability Insurance (OASDI) participation is the sum of Old-Age and Survivors Insurance (OASI) and Social Security Disability Insurance (SSDI) participation.

bias based on comparisons of aggregate survey and administrative data for the same year and survey as Marquis and Moore but not the same months or states. The bias we calculate in the second column is a function of sample weighting, coverage error, unit and item nonresponse, and measurement error, while the bias in the first column is only a function of item nonresponse and measurement error. Thus, if the biases in each of these columns are similar, then this suggests that the combination of sample weighting, coverage error, and unit nonresponse is not that important relative to the other sources of bias. The results in Table 3 suggest that the weights—as well as unit nonresponse and coverage error—are not a substantial source of bias because the bias estimates from the linked microdata are fairly close to our estimates using comparisons to aggregates. Our estimates are particularly close (or higher) for the Food Stamp Program and for Supplemental Security Income, which are programs that target the poor—a group that perhaps is most plausibly thought to be underweighted or underrepresented.

Through linked survey and administrative microdata, one can decompose our bias estimates into three different sources of error: unit nonresponse (combined with coverage error and weighting), item nonresponse, and measurement error. In Table 4 we report this decomposition of our estimates of dollar bias for the Food Stamp Program (FSP) and Public Assistance (combining Temporary Assistance for Needy Families and General Assistance) in three of our surveys in recent years using New York state data for 2007–2012. We find that the bias due to the combination of coverage error, unit nonresponse, and weighting is substantial, the bias due to item nonresponse is small, and the bias due to measurement error is always larger than the other sources of bias combined. The combined coverage, unit nonresponse, and weighting bias varies from −0.049 to −0.096 for the FSP and −0.100 to −0.154 for Public Assistance across the three surveys. The item nonresponse bias varies from −0.020 to −0.067 for the FSP and −0.022 to −0.057 for Public Assistance. The

Table 4
Decomposition of Proportional Bias in Dollars Received into Its Sources Using Microdata

Survey	*Program*	*Bias due to combination of coverage, unit nonresponse, and weighting*	*Bias due to item nonresponse*	*Bias due to measurement error*	*Total bias due to all sources of error*
ACS	Food Stamps	−0.096	na*	na*	na*
	Public Assistance	−0.154	−0.022	−0.529	−0.705
CPS	Food Stamps	−0.049	−0.067	−0.267	−0.382
	Public Assistance	−0.106	−0.057	−0.563	−0.726
SIPP	Food Stamps	−0.056	−0.020	−0.121	−0.197
	Public Assistance	−0.100	−0.043	−0.584	−0.727

Source: Authors' calculations based on New York State data for 2007–2012 from Celhay, Meyer, and Mittag (2015).
Notes: We calculate the bias due to the combination of errors in coverage, weighting, and unit nonresponse as the ratio of weighted administrative program dollars received by all linked households in the Current Population Survey to total administrative dollars paid out minus one. We calculate the bias due to item nonresponse as weighted dollars imputed to those not responding to the benefit question minus the dollars actually received by these households as a share of total dollars paid out. Finally, we calculate the bias due to measurement error as the dollars recorded by non-imputed respondents minus true dollars received as a share of total dollars paid out. The surveys are the American Community Survey (ACS), the Current Population Survey (CPS), and the Survey of Income and Program Participation (SIPP).
*Food stamp dollars received are not reported in these years of the ACS.

bias due to measurement error is substantial for the FSP, ranging from −0.121 to −0.267, and for Public Assistance it is even larger, ranging from −0.529 to −0.584.

Direct evidence of substantial measurement error is not restricted to these two programs. Through linked survey and administrative microdata, Gathright and Crabb (2014) document substantial measurement error in receipt and amounts of Supplemental Security Income and the Old-Age Survivors and Disability Insurance in the Survey of Income and Program Participation, and this measurement error is rising over time.

Methodological Issues when Comparing Aggregate Data

Comparing weighted microdata from surveys to administrative aggregates is an attractive approach for evaluating survey bias because it can be done easily for many years and across many surveys. However, this approach also has some important limitations including possible differences between the survey and administrative populations, and incomplete information on benefit receipt in some surveys. An additional concern that we will not discuss here is that by looking at net measures of bias, we are missing the extent to which a rise in false negative reports could be

counterbalanced by a rise in false positive reports. Most of these problems are not present when linking microdata at the household level.

Survey and Administrative Data Populations Do Not Always Align

Our household survey totals do not include those living outside the 50 states and the District of Columbia, the institutionalized, or decedents. We make a number of adjustments in order to make the administrative and survey data totals comparable (for a full description, see Meyer, Mok, and Sullivan 2015). For example, we exclude from the administrative totals payments to those in US territories and those outside the United States. Where such information is not available, we subtract estimates of the share of such payments obtained from years when this information is available. For most programs, these adjustments are typically small, ranging from 0.02 percent (Supplemental Security Income) to about 3 percent (Social Security Disability Insurance). The notable exception is the Food Stamp Program, where dollars paid to US territories constituted about 10 percent of the total prior to 1982.

As another example, to adjust for the fact that the institutionalized can receive some benefits in the Social Security-related programs, we rely on data from the Decennial Censuses (which include the institutionalized) and the 2006 American Community Survey to determine the share of dollars that are likely missed in household surveys that do not cover the institutionalized. That the surveys do not include decedents is a potential concern because recipients of transfers in one calendar year may subsequently die before being interviewed in a household survey the next year. We do not adjust for decedents, but assuming that the weights for extrapolating the household survey results to the population are well-chosen, we expect the lack of a specific adjustment for decedents to have little effect on our estimates in most cases.[10]

Often the reference period for the administrative data (typically a fiscal year) does not exactly align with that for the survey data. We convert fiscal year administrative data to a calendar basis by weighting the fiscal years. Another noncomparability is that administrative data for transfer income are based on awardees, while the survey data typically provide information on the person to whom the benefit is paid. Awardees and payees may be different people. For example, adults may receive Social Security and Supplemental Security benefits on behalf of their children. Most household surveys provide little information about exactly who is the true awardee of the

[10] Previous studies have adjusted for decedents by applying age-, gender-, and race-specific death rates to the data (Roemer 2000). However, if survey weights have previously been calculated to match survey-weighted population totals with universe population estimates by age, gender, and race, then such an adjustment is unwarranted. A case could be made for adjusting the data if these characteristics are nonstationary (but such an adjustment is likely to be small) or if the adjustments were based on additional individual characteristics that are not used to determine weights but are related to death, such as receipt of Social Security Disability Insurance or Supplemental Security Income or other programs, but we do not have this information. Consequently, our estimates of bias for SSDI and SSI are likely to be overstated somewhat, since recipients likely have a higher mortality rate than the average person of their age, gender, and race, and consequently are more likely to miss the interview the following year.

benefit, although the Survey of Income and Program Participation does provide some partial information about who is the true awardee of Social Security benefits.

Some Surveys Provide Incomplete Information on the Receipt of Benefits

In certain years of the Panel Study of Income Dynamics, for example, we only have information about benefit receipt for the household head and the spouse. We address this issue by using the share of total benefits received by non-head, non-spouse family members in other years and scaling up the aggregates accordingly. This adjustment assumes that these shares change slowly over time. Non-head, non-spouse dollars received are typically under 10 percent of family dollars but exceed 20 percent for Supplemental Security Income in a few years.

Sometimes surveys do not distinguish between different types of benefits received. In some cases, we cannot distinguish between different types of Social Security income. In this situation, we apply the Old-Age Survivors and Disability Insurance dollar proportions from published totals to determine participation in these programs. Applying these proportions essentially assumes that an individual can only receive benefits from one of these programs, but not both. In practice, however, individuals can receive benefits from both programs in a year—most commonly those whose disability benefit switches automatically to an old-age benefit when they reach retirement age. This issue leads to a slight bias downward in our Social Security retirement and disability participation estimates.

Reasons for Nonresponse and Errors

Why are nonresponse and measurement error so prevalent? Why have these threats to survey quality grown over time? Regarding the high rate of unit nonresponse, disinterest or lack of time appear to be important factors. Based on data recorded by interviewers for two household surveys—the 1978 National Medical Care Expenditure Survey and the 2008 National Health Interview Survey—the most common reasons given for unit nonresponse include that potential respondents are not interested, do not want to be bothered, or are too busy, while privacy concerns also seem to be important (Brick and Williams 2013, p. 39; National Research Council 2013). Reasons for unit nonresponse are often divided into three categories: noncontact, refusals, and other reasons (such as language problems or poor health). Failure to contact has also been offered as a possibility by some who have noted the rise of gated communities and the decline of land-line phones, which could make door-to-door or phone surveys more difficult. However, the rise in nonresponse in household surveys has been primarily driven by refusals by those who are contacted (Brick and Williams 2013), and thus we will not emphasize these potential "technological" reasons for noncontact.

One might suspect that the reasons for item nonresponse and measurement error are closely related to those for unit nonresponse, though the literature on survey quality has tended to focus on unit nonresponse separately. One reason the three

sources of error may be related would arise if some potential respondents are just less cooperative so that their participation is worse in many dimensions. Some research has examined this hypothesis. For example, Bollinger and David (2001) show that those who respond to all waves of a Survey of Income and Program Participation panel report participation in the Food Stamp Program more accurately than those who miss one or more waves. Similarly, Kreuter, Muller, and Trappmann (2014) show in a German survey that hard-to-recruit respondents provided less-accurate reports of welfare benefit receipt than those easy to recruit. The reasons for item nonresponse likely differ depending on the nature of the questions. In the case of earnings, Groves and Couper (1998) suggest that the most important reason for nonresponse is concerns about confidentiality but that insufficient knowledge is also important.

The reasons for the underreporting of transfer benefits in household surveys have been catalogued by several authors; Marquis and Moore (1990) provide nice examples for the Survey of Income and Program Participation, while Bound, Brown, and Mathiowetz (2001) and Groves (2004) provide more general discussions. Interviewees may forget receipt or confuse the names of programs. They may misremember the timing of receipt or who are the true recipients of a program within a family. Errors may be due to a desire to shorten the time spent on the interview, the stigma of program participation, the sensitivity of income information, or changes in the characteristics of those who receive transfers. Survey and interviewer characteristics such as the interview mode (in person or by phone) and respondent type (self or proxy) may also matter for the degree of underreporting. Notice that all of these explanations may lead to item nonresponse, measurement error conditional on responding, or both.

Information on the extent of underreporting and how it varies across programs, surveys, and time should help in differentiating among the explanations for underreporting. For example, one standard explanation of underreporting is the stigma of reporting receipt of "welfare" programs, and the inclination to give "socially desirable" answers (Sudman and Bradburn 1974). This explanation is consistent with the low reporting rates of four of the programs most associated with "welfare" or idleness: Temporary Assistance to Needy Families, the Food Stamp Program, Unemployment Insurance, and the Special Supplemental Nutrition Program for Women, Infants, and Children. However, other patterns of reporting by program do not fit with a stigma explanation for underreporting. Workers' Compensation has the greatest bias but is presumably not a program that greatly stigmatizes its recipients, given that the program is for those injured while working.

Another common explanation for underreporting is that interviewees forget receipt, misremember the timing of receipt, or confuse the names of programs. Such issues should arguably be less common for programs that are received regularly, such as Old-Age and Survivors Insurance, Social Security Disability Insurance, and Supplemental Security Income. And, as shown in Table 1, these three programs typically have smaller bias than the other transfer programs we examine. However, the estimates in Table 1 show that the proportional bias for these programs is still large, particularly for SSDI and SSI, although this could be due to greater stigma

for these two programs. Also, all three of these Social Security programs have item nonresponse rates that are no better than for some programs with less-regular receipt (see Figures 2A and B).

Why has survey quality deteriorated over time? Several studies have considered this question, mostly focusing on unit nonresponse. The traditional reasons proposed include increasing urbanization, a decline in public spirit, increasing time pressure, rising crime (this pattern reversed long ago), increasing concerns about privacy and confidentiality, and declining cooperation due to "oversurveyed" households (Groves and Couper 1998; Presser and McCullogh 2011; Brick and Williams 2013). The continuing increase in survey nonresponse as urbanization has slowed and crime has fallen make these less likely explanations for present trends. Tests of the remaining hypotheses are weak, based largely on national time-series analyses with a handful of observations. Several of the hypotheses require measuring societal conditions that can be difficult to capture: the degree of public spirit, concern about confidentiality, and time pressure. The time pressure argument seems inconsistent with the trend toward greater leisure (Aguiar and Hurst 2007) and would suggest that those with higher incomes and less leisure should be less likely to respond to surveys—a pattern that is at best weakly present. We are unaware of strong evidence to support or refute a steady decline in public spirit or a rise in confidentiality concerns as a cause for declines in survey quality. Some of these hypotheses seem amenable to a geographically disaggregated time-series approach, but little work seems to have been done along those lines. Groves and Couper (1998) show that nonresponse rates differ across demographic groups; cooperation is lower among single person households and households without young children, for example. But more research is needed on whether changes in demographic characteristics such as these can account for declining survey quality.

Changes in survey procedures over time can also provide evidence on the reasons for changes in underreporting of receipt of government transfers. The reduction or elimination of in-person interviewing seems to have had little effect on reporting rates. For example, reporting rates do not change much after the 1996 reduction of in-person interviewing in the Survey of Income and Program Participation. This result is consistent with the observation by Groves (2004) that there is no robust evidence of a difference in errors between in-person and phone interviewing. Reporting for transfer programs does not appear to be sensitive to whether or not the interviewer explicitly mentions the name of a program (Meyer, Mok, and Sullivan 2015). There is some evidence that adding "bracketed" responses—for example, starting in 2001, when a specific amount is not provided, the Consumer Expenditure Survey asks interviewees whether the amount falls within certain ranges—leads to increased reporting rates for some programs, but this evidence is not consistent across programs (Meyer, Mok, and Sullivan 2015).

Our own reading of the evidence supports the hypothesis that "oversurveyed" respondents are less cooperative resulting in greater nonresponse and measurement error. Presser and McCullogh (2011) document a sharp rise in the number of government surveys administered in the United States over the 1984–2004

period. They report that a series of random-digit-dial telephone surveys found that the share of Americans surveyed in the past year more than quadrupled between 1978 and 2003 (Council for Marketing and Opinion Research 2003). They also note that real expenditures on commercial survey research increased by more than 4 percent annually for the 16 years ending in 2004. We suspect that talking with interviewers, once a rare chance to tell someone about your life, is now crowded out by an annoying press of telemarketers and commercial surveyors. The decline in unit and item response rates may not fully reflect the secular decline in the willingness of households to cooperate because survey administrators have tried to offset a declining household willingness to be surveyed by altering their methods. For example, Groves and Couper (1998) note cases where the number of attempted contacts with respondents has increased in order to stem the rise in nonresponse.

Taken together, the existing evidence does not provide a complete explanation for why survey quality has deteriorated over time. That some households are oversurveyed seems to contribute to the problem, but other explanations are likely important as well. There is a clear need for further research to fill the gaps in this literature.

The Future of Microdata

As the quality of conventional household survey data has declined, the availability of alternative data for research and policy analysis has increased. For empirical research, while the role of survey data has declined, that of administrative data has risen; Chetty (2012) reports that the share of nondevelopment-microdata-based articles in the "top four" general interest economics journals that relied on survey data fell from about 60 to 20 percent between 1980 and 2010, while the share of articles relying on administrative data rose from about 20 to 60 percent. A number of standard sources of administrative data have already been mentioned in this article; these include data from tax records and from transfer programs. In addition, the use of alternative forms of administrative data has been increasing. For example, the work surveyed in Einav and Levin (2014) offers examples like the use of administrative data on student test scores to measure teacher value added, or data on earnings to assess the effect of the spread of broadband Internet access into different areas.

Administrative data offer a bundle of advantages and disadvantages. The datasets often have large sample sizes and low measurement error, permitting the estimation of small effects and the testing of subtle hypotheses. The data often allow longitudinal measurement, which is not possible in cross-sectional household data and can be difficult in longitudinal surveys with substantial attrition. When changes occur in policy or practice, especially when those changes affect only certain populations or geographic areas, administrative data often enable the use of experimental or quasi-experimental research methods. On the other hand, administrative data sets are typically not designed for academic research, and they can be quite heterogeneous in origin, topic, and quality. Researchers can find it difficult to access these data, whether for

original research or for replication. Administrative data often cover only a limited set of characteristics of individuals, and these variables are often of low quality, especially for types of information that are collected but are not directly needed by the program for administration or other purposes. Also, administrative data sources often have incomplete coverage and are nonrepresentative, making the data unsuitable for drawing generalizable conclusions or examining population trends.

The limitations of administrative data can potentially be addressed by linking to household survey data. Many recent reports by government agencies, advocacy groups, and politicians have pointed to the advantages of administrative data linked to survey data (for example, Burman et al. 2005; Brown et al. 2014; Office of Management and Budget 2014; US House of Representatives 2014). These reports have noted the usefulness of such data for a wide variety of policy analyses. President Obama's 2016 budget calls for $10 million for the Census Bureau "to accelerate the process of acquiring additional key datasets . . . ; expand and improve its infrastructure for processing and linking data; and improve its infrastructure for making data available to outside researchers" (White House 2015). A recent bipartisan bill would establish a commission to recommend the structure of a clearinghouse for administrative and survey data (US Senate 2015).

Linking administrative microdata to survey microdata may improve the quality of survey data by providing more accurate information for some variables or by shortening the interview length and reducing the burden on survey respondents who no longer need to be asked questions they might be reluctant to answer. Such data linking can also be useful for improving the existing stock of data. For example, Nicholas and Wiseman (2009, 2010) and Meyer and Mittag (2015) show how one can use linked data to correct for underreporting of transfer income when calculating poverty rates.

A number of examples already exist of linked survey and administrative data. The Health and Retirement Survey is linked to administrative data on Social Security earnings and claims, as well as to Medicaid data. The National Center for Health Statistics is currently linking several of its population-based surveys to administrative data. Many randomized experiments of welfare and training programs linked household survey instruments to Unemployment Insurance earnings records or other administrative datasets (Grogger and Karoly 2005). Ad hoc examples within government have also produced useful research such as the work by Scherpf, Newman, and Prell (2014) using administrative data on the Supplemental Nutrition Assistance Program.

Surveys have explored alternative methods to improve survey quality, including multi- or mixed-mode methods to collect information from respondents (Citro 2014). Use of the Internet has become an increasingly more common mode. In addition to standard mail, telephone, and face-to-face interview modes, the American Community Survey now allows respondents to respond online. These new methods may, in some cases, reduce costs and have the potential to improve data quality, but whether these approaches reduce bias remains to be seen.

Much of what we know about the conditions of the American public and the information that is used for public policy formation comes from national survey

data. The ongoing deterioration of household survey data documented in this paper seems unlikely to end, especially as surveying for commercial purposes and the feeling of being oversurveyed continues to grow. Without changes in data collection and availability, the information infrastructure to formulate and evaluate public policies and to test social science theories will degrade. Efforts to improve national survey data and to reduce nonresponse bias and measurement error are worthwhile. But perhaps the most productive step toward improving the quality of data available for social science research—rather than just seeking to slow the pace of erosion in the quality of that data—is to increase the availability of administrative datasets and to find additional ways to link them to household survey data and substitute administrative variables for survey questions in a timely fashion.

■ *We are grateful for the assistance of many New York Office of Temporary and Disability Assistance (OTDA) and Census Bureau employees including George Falco, Dave Dlugolecki, Graton Gathright, Joey Morales, Amy O'Hara, and Frank Limehouse. The microdata analysis was conducted at the Chicago Census Research Data Center by researchers with Special Sworn Status and the results were reviewed to prevent the disclosure of confidential information. We would like to thank Pablo Celhay for excellent research assistance. We also thank Dan Black, Constance Citro, Michael Davern, Nikolas Mittag and participants at seminars at the American Enterprise Institute and the Federal Deposit Insurance Corporation for their helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the New York OTDA or the US Census Bureau.*

References

Abowd, John, and Martha Stinson. 2013. "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data." *Review of Economic Statistics* 95(5): 1451–67.

Aguiar, Mark, and Erik Hurst. 2007. "Measuring Trends in Leisure: The Allocation of Time over Five Decades." *Quarterly Journal of Economics* 122(3): 969–1006.

Alwin, Duane F. 2007. *Margins of Error: A Study of Reliability in Survey Measurement.* John Wiley & Sons: Hoboken, NJ.

Andridge, Rebecca R., and Roderick J. A. Little. 2010. "A Review of Hot Deck Imputation for Survey Non-Response." *International Statistical Review* 78(1): 40–64.

Atrostic, B. K., Nancy Bates, Geraldine Burt, and Adrian Silberstein. 2001. "Nonresponse in US Government Household Surveys: Consistent Measures, Recent Trends, and New Insights." *Journal of Official Statistics* 17(2): 209–226.

Barrow, Lisa, and Jonathan Davis. 2012. "The Upside of Down: Postsecondary Enrollment in the Great Recession." *Economic Perspectives,* Federal Reserve Bank of Chicago, 4Q: 117–29.

Battaglia, Michael P., Mina Khare, Martin R. Frankel, Mary Cay Murray, Paul Buckley, and Saralyn Peritz. 2007. "Response Rates: How Have They Changed and Where Are They Headed?" In *Advances in Telephone Survey Methodology* edited by James M. Lepkowski, Clyde Tucker, J. Michael Brick, Edith D. de Leeuw, Lilli Japec,

Paul J. Lavrakas, Michael W. Link, and Roberta L. Sangster, 529–60. New York, NY: Wiley.

Bee, C. Adam, Graton Gathright, and Bruce D. Meyer. 2015. "Bias from Unit Non-Response in the Measurement of Income in Household Surveys." Unpublished paper, University of Chicago.

Bee, C. Adam, Bruce Meyer, and James Sullivan. 2015. "The Validity of Consumption Data: Are the Consumer Expenditure Interview and Diary Surveys Informative?" In *Improving the Measurement of Consumer Expenditures*, edited by Christopher Carroll, Thomas Crossley, and John Sabelhaus, 204–240. University of Chicago Press.

Bollinger, Christopher, and Martin H. David. 2001. "Estimation with Response Error and Nonresponse: Food-Stamp Participation in the SIPP." *Journal of Business and Economic Statistics* 19(2): 129–41.

Bollinger, Christopher R., and Barry Hirsch. 2006. "Match Bias from Earnings Imputation in the Current Population Survey: The Case of Imperfect Matching." *Journal of Labor Economics* 24(3): 483–519.

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement Error in Survey Data." In *Handbook of Econometrics,* vol. 5, edited by James J. Heckman and Edward Leamer. Elsevier: Amsterdam.

Brick, Michael J., and Douglas Williams. 2013. "Explaining Rising Nonresponse Rates in Cross-Sectional Surveys." *Annals of the American Academy of Political and Social Science* 645(1): 36–59.

Brown, Lawrence, Constance Citro, Carol House, Krisztina Marton, and Christopher Mackie. 2014. "The Past, Present, and Future of Federal Surveys: Observations from the Committee on National Statistics." Conference proceedings from JSM (Joint Statistical Meetings) 2014 - Social Statistics Section.

Burman, Len, Dan Feenberg, Austan Goolsbee, Charles Hulten, Bruce Meyer, John Karl Scholz, and Joel Slemrod. 2005. "Report on the State of Publicly Available Data and Statistics For the Study of Public Economics." Report for the Committee on Economic Statistics of the American Economic Association.

Celhay, Pablo, Bruce D. Meyer, and Nicholas Mittag. 2015. "Measurement Error in Program Participation." Unpublished paper.

Chetty, Raj. 2012. "Time Trends in the Use of Administrative Data for Empirical Research." NBER Summer Institute presentation. Available at the author's website.

Citro, Constance F. 2014. "From Multiple Modes for Surveys to Multiple Data Sources for Estimates." *Survey Methodology* 40(2): 137–61.

Council for Marketing and Opinion Research. 2003. "Respondent Cooperation and Industry Image Study." Previously available at http://www.cmor.org/rc/studies.cfm.

Coder, John, and Lydia Scoon-Rogers. 1996. "Evaluating the Quality of Income Data Collected in the Annual Supplement to the March Current Population Survey and the Survey of Income and Program Participation." Housing and Household Economic Statistics Division, Bureau of the Census. https://www.census.gov/sipp/workpapr/wp215.pdf.

Curtin, Richard, Stanley Presser, and Elinor Singer. 2005. "Changes in Telephone Survey Nonresponse over the Past Quarter Century." *Public Opinion Quarterly* 69(1): 87–98.

Davern, Michael. 2013. "Nonresponse Rates Are a Problematic Indicator of Nonresponse Bias in Survey Research." *Health Services Research* 48(3): 905–912.

de Leeuw, Edith, and Wim de Heer. 2002. "Trends in Households Survey Nonresponse: A Longitudinal and International Comparison." In *Survey Nonresponse*, edited by Robert M. Groves, Don A. Dillman, John Eltinge, and Roderick J.A. Little. Wiley.

Duncan, Greg J., and Daniel H. Hill. 1989. "Assessing the Quality of Household Panel Data: The Case of the Panel Study of Income Dynamics." *Journal of Business and Economic Statistics* 7(4): 441–52.

Dushi, Irena, and Howard Iams. 2010. "The Impact of Response Error on Participation Rates and Contributions to Defined Contribution Pension Plans." *Social Security Bulletin* 70(1): 45–60.

Einav, Liran, and Jonathan Levin. 2014. "Economics in the Age of Big Data." *Science* 346 (6210).

Gathright, Graton, and Tyler A. Crabb. 2014. "Reporting of SSA Program Participation in SIPP." Working Paper, US Census Bureau. http://www.eventscribe.com/2014/ASA-JSM/assets/pdf/312130_92347.pdf.

Grogger, Jeffrey, and Lynn A. Karoly. 2005. *Welfare Reform: Effects of a Decade of Change.* Cambridge, MA: Harvard UP.

Groves, Robert M. 2004. *Survey Errors and Survey Costs.* Hoboken, NJ: Wiley.

Groves, Robert M. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70(5): 646–75.

Groves, Robert M., and Mick P. Couper. 1998. *Nonresponse in Household Interview Surveys.* New York, NY: Wiley.

Groves, Robert M., and Emilia Peytcheva. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias." *Public Opinion Quarterly* 72(2): 167–89.

Hogan, Howard. 1993. "The 1990 Post-Enumeration Survey: Operations and

Results." *Journal of the American Statistical Association* 88(423): 1047–60.

Huynh, Minh, Kalman Rupp, and James Sears. 2002. "The Assessment of Survey of Income and Program Participation (SIPP) Benefit Data using Longitudinal Administrative Records." Unpublished paper, Social Security Administration, http://www.oecd.org/std/36232630.pdf.

Johnson, Barry W., and Kevin Moore. 2008. "Differences in Income Estimates Derived from Survey and Tax Data." 2008 SOI Paper Series.

Kreuter, Frauke, Gerrit Müller, and Mark Trappmann. 2014. "A Note on Mechanisms Leading to Lower Data Quality of Late or Reluctant Respondents." *Sociological Methods & Research* 43(3): 452–64.

Marquis, Kent H., and Jeffrey C. Moore. 1990. "Measurement Errors in SIPP Program Reports." In *Proceedings of the 1990 Annual Research Conference*, 721–45. Washington, DC: US Bureau of the Census.

Massey, Douglas S., and Roger Tourangeau, eds. 2013. "The Nonresponse Challenge to Surveys and Statistics," (special title of) volume 645, issue 1 of *Annals of the American Academy of Political and Social Science*, pages 6–236.

Meyer, Bruce D., Robert Goerge, and Nicholas Mittag. 2014. "Errors in Survey Reporting and Imputation and Their Effects on Estimates of Food Stamp Program Participation." Unpublished paper.

Meyer, Bruce D., and Nicholas Mittag. e: Implications for Poverty, Program Effectiveness and Holes in the Safety Net." Unpublished paper, University of Chicago.

Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan. 2015. "The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences." NBER Working Paper 15181. Updated version June 2015.

Mishra, V. B. Barrere, R. Hong, and S. Khan. 2008. "Evaluation of Bias in HIV Seroprevalence Estimates from National Household Surveys." *Sexually Transmitted Infection* 84(Supplement 1): i65-i70.

Moore, Jeffrey C., Kent H. Marquis, and Karen Bogen. 1996. "The SIPP Cognitive Research Evaluation Experiment: Basic Results and Documentation." Working Paper 212, Survey of Income and Program Participation, US Census Bureau.

National Research Council. 2011. *The Future of Federal Household Surveys: Summary of a Workshop.* K. Marton and J.C. Karberg, rapporteurs. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. 2013. *Nonresponse in Social Science Surveys: A Research Agenda*. Edited by Roger Tourangeau and Thomas J. Plewes. Panel on a Research Agenda for the Future of Social Science Data Collection, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, National Research Council. Washington, DC: The National Academies Press.

Nicholas, Joyce, and Michael Wiseman. 2009. "Elderly Poverty and Supplemental Security Income." *Social Security Bulletin* 69(1): 45–73.

Nicholas, Joyce, and Michael Wiseman. 2010. "Elderly Poverty and Supplemental Security Income, 2002–2005." *Social Security Bulletin* 70(2): 1–30.

Office of Management and Budget. 2006. "Standards and Guidelines for Statistical Surveys." September. Executive Office of the President, Washington, DC.

Office of Management and Budget. 2014. "Guidance for Providing and Using Administrative Data for Statistical Purposes." Memorandum for the Heads of Executive Departments and Agencies, M-14-06, February. Executive Office of the President, Washington, DC.

Pennell, Steven G. 1993. "Cross-Sectional Imputation and Longitudinal Editing Procedures in the Survey of Income and Program Participation." Prepared for the US Bureau of the Census. https://www.census.gov/sipp/workpapr/wp9314.pdf.

Pew Research Center. 2012. "Assessing the Representativeness of Public Opinion Surveys," May 15. http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/.

Peytchev, Andrew. 2013. "Consequences of Survey Nonresponse." *Annals of the American Academy of Political and Social Science* 645(1): 88–111.

Presser, Stanley, and Susan McCullogh. 2011. "The Growth of Survey Research in the United States: Government-Sponsored Surveys, 1984–2004." *Social Science Research* 40(4): 1019–2024.

Reamer, Andrew D. 2010. "Surveying for Dollars: The Role of the American Community Survey in the Geographic Distribution of Federal Funds." Brookings Report, July 2010.

Robinson, J. Gregory, Bashir Ahmed, Prithwis Das Gupta, and Karen A. Woodrow. 1993. "Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis." *Journal of the American Statistical Association* 88(423): 1061–71.

Roemer, Marc I. 2000. "Assessing the Quality of the March Current Population Survey and the Survey of Income and Program Participation Income Estimates, 1990–1996." Working Papers, Census Bureau.

Scherpf, Erik, Constance Newman, and Mark Prell. 2014. "Targeting of Supplemental Nutrition

Assistance Program Benefits: Evidence from the ACS and NY SNAP Administrative Records." Agricultural and Applied Economics Association 2014 Annual Meeting, July 27–29, 2014, paper 174295.

Sears, James, and Kalman Rupp. 2003. "Exploring Social Security Payment History Matched with the Survey of Income and Program Participation." Unpublished paper, Social Security Administration. http://www.oecd.org/std/36232603.pdf.

Smith, Tom W., Peter V. Marsden, Michael Hout, Jibum Kim. 2013. *General Social Surveys, 1972–2012 [Cumulative File].* Codebook. ICPSR 34802, Inter-University Consortium for Political and Social Research.

Steeh, Charlotte, Nicole Kirgis, Brian Cannon, and Jeff DeWitt. 2001. "Are They Really as Bad as They Seem? Nonresponse Rates at the End of the Twentieth Century." *Journal of Official Statistics* 17(2): 227–47.

Sudman, Seymour, and Norman M. Bradburn. 1974. *Response Effects in Surveys: A Review and Synthesis.* Chicago: NORC/Aldine Publishing Company.

Taeuber, Cynthia, Dean M. Resnick, Susan P. Love, Jane Stavely, Parke Wilde, and Richard Larson. 2004. "Differences in Estimates of Food Stamp Program Participation Between Surveys and Administrative Records." Working paper, US Census Bureau.

Tourangeau, Roger, Robert M. Groves, and Cleo D. Redline. 2010. "Sensitive Topics and Reluctant Respondents: Demonstrating a Link between Nonresponse Bias and Measurement Error." *Public Opinion Quarterly* 74(3): 413–32.

US Census Bureau. 2001. "Survey of Income Program Participation Users' Guide." Third Edition. United States Department of Commerce, Bureau of the Census, Washington, DC.

US Census Bureau. 2003. "Meeting 21st Century Demographic Data Needs—Implementing the American Community Survey: Testing the Use of Voluntary Methods." United States Department of Commerce, Bureau of the Census, Washington, DC.

US Census Bureau. 2012. "Census Bureau Releases Estimates of Undercount and Overcount in the 2010 Census." United States Department of Commerce, Bureau of the Census, Washington, DC. http://www.census.gov/newsroom/releases/archives/2010_census/cb12-95.html.

US Census Bureau. Various years (a). "Current Population Survey: Annual Social and Economic (ASEC) Survey Codebook." United States Department of Commerce, Bureau of the Census, Washington, DC.

US Census Bureau. Various years (b). "Survey of Income and Program Participation Codebook." United States Department of Commerce, Bureau of the Census, Washington, DC.

US Department of Health and Human Services. 2014. "National Health Interview Survey – Survey Description." Centers for Disease Control and Prevention, Division of Health Interview Statistics, National Center for Health Statistics, Hyattsville, Maryland.

US Department of Labor (various years). "Consumer Expenditure Interview Survey Public Use Microdata Documentation." U.S. Department of Labor, Bureau of Labor Statistics, Division of Consumer Expenditure Surveys.

US House of Representatives. 2014. "Expanding Opportunity in America." House Budget Committee Majority Staff, July 24.

US Senate. 2015. "Murray, Ryan Introduce Bill to Expand Data Use in Evaluating Federal Programs, Tax Expenditures." News Release. http://www.murray.senate.gov/public/index.cfm/2015/4/evidence-based-policy-murray-ryan-introduce-bill-to-expand-data-use-in-evaluating-federal-programs-tax-expenditures.

US Social Security Administration. Various years. "Annual Statistical Supplement to the Social Security Bulletin." US Social Security Administration, Office of Research, Evaluation and Statistics.

Wheaton, Laura. 2007. "Underreporting of Means-Tested Transfer Programs in the CPS and SIPP." 2007 Proceedings of the American Statistical Association, Social Statistics Section.

White House. 2015. "The President's Budget Fiscal Year 2016." https://www.whitehouse.gov/sites/default/files/omb/budget/fy2016/assets/fact_sheets/building-and-using-evidence-to-improve-results.pdf.

Journal of Economic Perspectives—Volume 29, Number 4—Fall 2015—Pages 227–244

Seven Centuries of European Economic Growth and Decline[†]

Roger Fouquet and Stephen Broadberry

John Maynard Keynes (1931) wrote in his essay "Economic Possibilities for Our Grandchildren":

> From the earliest times of which we have record—back, say, to two thousand years before Christ—down to the beginning of the eighteenth century, there was no very great change in the standard of life of the average man living in the civilised centres of the earth. Ups and downs certainly. Visitations of plague, famine, and war. Golden intervals. But no progressive, violent change. Some periods perhaps 50 per cent better than others—at the utmost 100 per cent better—in the four thousand years which ended (say) in A.D. 1700.

Over the decades, his nuanced account of the past has become a great deal more stylized—the lack of "very great" or "violent" change became an absence of any change. Today, received wisdom holds that the western European countries did not experience major phases of economic growth (or decline) prior to the Industrial Revolution. As one example among many that could be cited, Hansen and Prescott (2002, pp. 1214–15) write that "sustained growth has existed for at most the past two centuries, while the millennia prior have been characterized by stagnation with

■ *Roger Fouquet is Principal Research Fellow, Grantham Research Institute on Climate Change and the Environment at the London School of Economics and Political Science, London, United Kingdom. Stephen Broadberry is Professor of Economic History, University of Oxford, United Kingdom. Their email addresses are r.fouquet@lse.ac.uk and stephen.broadberry@nuffield.ox.ac.uk.*

[†] For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.29.4.227

doi=10.1257/jep.29.4.227

no significant permanent growth in living standards." Yet qualitative accounts of European histories seem to indicate that the Renaissance in Italy and the Golden Age in Holland reflected phases of economic development—associated with the expansion of trade and urbanization, as well as developments in art and science—prior to the Industrial Revolution (Goldthwaite 2009; Acemoglu and Robinson 2012; de Vries and van der Woude 1997).

Building in part on Angus Maddison's (1982, 1995, 2003) bold empirical research program to create very long-run data series for many countries, a generation of economic historians has been exploring archives and combining datasets to create more and better evidence. Over the last four years, a number of very long-run time series have been completed for major economies of Europe connecting the late medieval era with the present using annual data. The new time series have been looked at individually by the researchers who produced the data (Broadberry, Campbell, Klein, Overton, and van Leeuwen 2011; Malanima 2011; van Zanden and van Leeuwen 2012; Schön and Krantz 2012; Álvarez-Nogal and Prados de la Escosura 2013; Reis, Martins, and Costa 2013). The main aim of this paper is to present this data together and offer an alternative interpretation of very long-run European economic development.

The first section of this paper rejects the received wisdom that economies in pre–Industrial Revolution Europe were stagnant. The new data shows trends in GDP per capita in the key European economies before the Industrial Revolution, identifying episodes of economic growth in specific countries, often lasting for decades. Ultimately, these periods of growth were not sustained, but they noticeably raised GDP per capita. It also shows that many of these economies experienced periods of substantial economic decline. Thus, rather than being stagnant, pre-nineteenth century European economies experienced a great deal of change.

In the second section, the paper tentatively finds that the likelihood of being in a phase of growth increased and the risk of being in a phase of decline decreased in the nineteenth and twentieth centuries. The discussion in the third section turns to evidence on patterns of divergence and convergence in Europe, showing that divergence occurred when a new economic leader moved ahead, followed by a period of convergence and catch-up by others. The fourth section presents the main data sources and methods used to construct the GDP per capita estimates from the late medieval and early modern eras until the nineteenth century in six European economies: England/Great Britain, Holland, Italy, Spain, Sweden, and Portugal.

Following the publication of successive editions of Maddison's (1982, 1995, 2003) data on long-run economic performance, understanding of long-run economic growth and development advanced, as many economists exploited the new information. Research by Baumol (1986), DeLong (1988), and Pritchett (1997) relating to convergence and divergence is based on Maddison's post-1870 OECD countries dataset, which helped to stimulate endogenous growth theory. Maddison's work was also a key inspiration for new theories of economic growth: for instance, Maddison's later dataset covering the entire world during the second millennium stimulated work by Hansen and Prescott (2002) and Galor (2005).

The production of these new very long-run datasets may spur an equivalent or greater wave of understanding of economic growth and development. For the first time, it is possible to investigate annual changes in economic growth and development over several centuries, potentially identifying multiple major phases of economic growth and decline in one or a series of countries. This more detailed perspective of very long-run European economic growth indicates that the Industrial Revolution was not an isolated event originating in Great Britain. Instead, it was a phase in a much longer process of economic transformation across Europe.

Given current concerns about limited growth in industrialized economies and thoughts about ways to stimulate a new industrial revolution, research based on this data may offer insights about these underlying processes and be of great interest to economists in general. Since economic transformations on the scale of the Industrial Revolution are rare events, an understanding of the processes underlying such transformations will need to look back many centuries. Furthermore, comparisons of past experiences with more recent ones may suggest commonalities in the relationships with and determinants of economic development. This paper focuses on presenting basic facts and patterns to the best of our current abilities, and avoids the temptation to seek more detailed explanations. However, recent work, work in progress, and future papers seem certain to employ this data as the basis for new empirical and theoretical research.

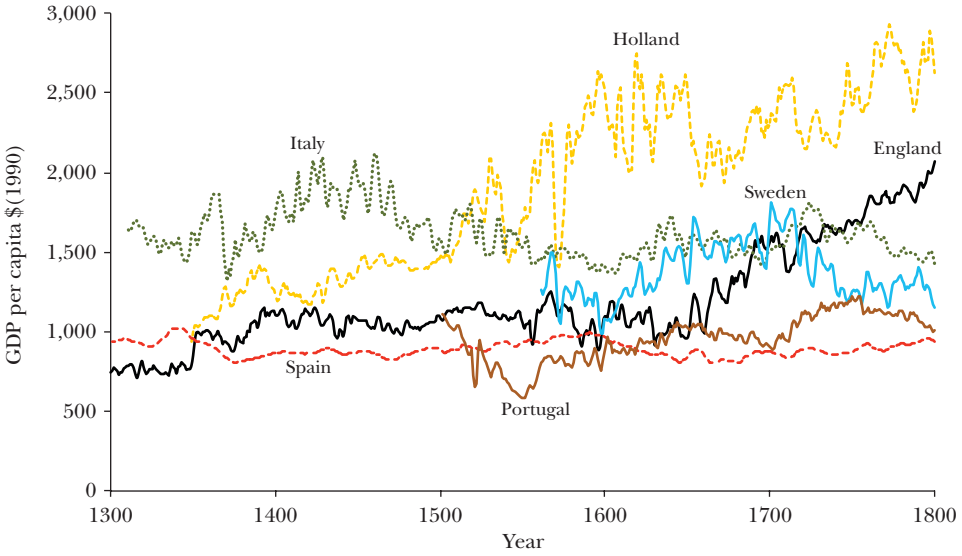
Growth Episodes and Growth Reversals in Europe before 1800

Economic growth before the nineteenth century was not stagnant, but had extended periods of growth and decline. Figure 1 presents GDP per capita for six European countries before the nineteenth century: England (from 1300 until 1700) and Great Britain afterwards, Holland (starting in 1348), Italy (specifically, Central and Northern Italian States from 1310), Spain (since 1300), Sweden (beginning in 1560), and Portugal (from 1500).

Figure 1 depicts four major "growth episodes" in specific European economies. Of the economies shown, Italy was the first to have experienced a per capita growth episode as population declined sharply after the Black Death, leaving survivors with more land and capital per person. Moreover, it was a period in which Italian cities prospered by expanding their pivotal role in trade links between Europe and Asia (Hodgett 2006). Between 1350 and 1420, the level of per capita income rose by 40 percent, which represents a modest but nonnegligible growth rate of 0.8 percent per year over 70 years.

Holland followed with a spectacular sixteenth century. Per capita GDP rose by 70 percent from 1505 to 1595 as Dutch trade expanded rapidly and the economic structure shifted away from agricultural production towards higher-value commodities, which translates into a growth rate of 1.3 percent per annum during this period. A decade later, Sweden started developing through its control of the Baltic trade, and its per capita GDP grew 41 percent in the first half of the seventeenth century.

Figure 1
GDP per Capita in Selected European Economies, 1300–1800
(three-year average; Spain eleven-year average)



Sources: England/Great Britain (Broadberry et al. 2011); Italy (Malanima 2011); Holland (van Zanden and van Leeuwen 2012); Sweden (Schön and Krantz 2012); Spain (Álvarez-Nogal and Prados de la Escosura 2013); Portugal (Reis, Martins, and Costa 2013; Palma and Reis 2014).
Note: Figure 1 presents GDP per capita for six European countries before the nineteenth century: England (from 1300 until 1700) and Great Britain afterwards, Holland (starting in 1348), Italy (specifically, Central and Northern Italian States from 1310), Spain (since 1300), Sweden (beginning in 1560), and Portugal (from 1500).

In the second half of the seventeenth century, England became the next vibrant economy—its per capita income growing by more than 50 percent during this time. This growth episode followed the end of a Civil War that marked an important step on the road to constitutional monarchy, culminating in the Glorious Revolution of 1688 (North and Weingast 1989; Acemoglu and Robinson 2012). However, population stagnated during the second half of the seventeenth century, so it was only after 1700 that Great Britain achieved modern economic growth with the coexistence of population growth and per capita GDP growth.

To ensure that these phases are not artifacts of selecting peaks and troughs in a volatile series, the total growth in GDP per capita is measured as the average value in the decade following the "growth episode" divided by the average value in the decade preceding this phase—although for reasons explained in greater detail later in this paper, one should be careful about over-interpreting these estimates. Some might argue that, starting from a low base, a rise of 40 or even 70 percent may not be a great absolute increase, and indeed, it is not very impressive when spread over 50 years or more. Nevertheless, these are unquestionably extended periods,

covering a number of decades, in which certain economies grew substantially in both relative and absolute terms, providing major improvements in the average standards of living (although ideally, distributional effects of these increases would also be taken into account). Clearly, these pre–nineteenth century European economies were not stagnant.

Interestingly, at practically every point during the entire sixteenth and seventeenth centuries, at least one economy in Europe was experiencing a growth episode. It would be worth investigating in greater detail the scale of spillovers to trade partners and the degree of emulation in these early periods (Reinert 2011). Certainly, England was highly dependent on Swedish iron imports in the seventeenth century (King 2005) and sought to emulate Holland's economic policies (Thirsk 1978). At the same time, until the eighteenth century, no two economies of the six shown here experienced simultaneous major phases of economic growth. While this certainly does not support the dubious mercantilist belief that foreign trade was a zero-sum game—and in particular, the belief that only exporters gained from international trade—it might help to explain why the belief held traction for so long (Smith 1776).

Figure 1 also identifies the periods that can be categorized as economic "growth reversals." Italy suffered most from periods of major economic decline, from its early period of glory. It experienced three periods of substantial decline of around 20 percent of per capita GDP as population growth returned, its markets remained fragmented between small states, and the focus of European trade shifted from the Mediterranean to the Atlantic. Following the collapse of per capita incomes in Italy in the mid-fifteenth century, it took more than 400 years to regain these levels of GDP per capita. Portugal suffered a dramatic collapse of roughly 40 percent of per capita GDP in the first half of the sixteenth century, associated with poor weather conditions (Reis, Martins, and Costa 2013)—though it recovered partially in the subsequent two decades. The Spanish economy also declined from the end of the sixteenth century—which was associated with the resource curse resulting from silver mining in the colonies (Drelichman 2005; Álvarez-Nogal and Prados de la Escosura 2013). Sweden suffered a collapse in the early eighteenth century, as it lost its great power status, with per capita GDP dropping almost 30 percent in three decades. Finally, after a period of growth in the first half of the eighteenth century, Portugal lost 16 percent of per capita GDP in three years and then spiraled downwards following the Great Earthquake of Lisbon in 1755.

There is little understanding of major economic collapses, especially since they are such rare events. In *Anna Karenina*, Leo Tolstoy proposes that "[a]ll happy families are alike; each unhappy family is unhappy in its own way." Certainly, there has been plenty of effort put toward finding in what way successful economies are alike, but little toward understanding the ways in which unsuccessful economies decline. More analyses of economic declines are needed to complement studies of economic failure such as Easterly and Levine (1997), Rodrik (1999), and Acemoglu and Robinson (2012). These new long-run datasets hold the promise of shedding light on economically depressed phases in history.

The Long Road to Sustained Growth

Each of the six European economies discussed here experienced phases of economic growth and decline at different times. Despite the different national patterns, did a general change in growth rates occur over time? In particular, one might provisionally expect that there were more phases of growth and fewer phases of economic decline in later centuries.

In looking at Figure 1, few obvious differences between centuries emerge. The seventeenth and eighteenth century perhaps show greater growth, but they also have more and better data. Many different criteria have been proposed for identifying phases of growth more formally, such as the frequency of consecutive years of growth (Hausman, Pritchett, and Rodrik 2005; Easterly 2006). However, identifying phases of growth is more difficult when analyzing mostly agrarian economies or periods before reliable statistical records existed, because of the high volatility in the GDP per capita series. The volatility can result either from weather-sensitive agricultural production or the estimation methods, which inevitably display a great degree of uncertainty.

Not surprisingly, therefore, a more formal analysis does not identify any difference over these earlier centuries: for example, in each century before 1800, there was only a 1–2 percent chance of a country being in a period of more than 1.5 percent annual growth in four consecutive years, as shown in Table 1, column 2. For comparison, this likelihood increased to a 5 percent chance in the nineteenth century and a 40 percent likelihood during the twentieth century. So, for the six countries observed, there was a substantial difference in the likelihood of sustained periods of growth between the pre-1800 period and the nineteenth and twentieth centuries. As received wisdom suggests, sustained economic growth seems to be a more recent phenomenon.

Looking at the economic downturns, Figure 1 does not seem to display an obvious change in frequencies before 1800. However, Table 1 provides some evidence. Here, we use a criterion of three consecutive years of less than −1.5 percent growth to identify a downturn. Adding across the six countries, there were 47 downturns before the nineteenth century and only eight after 1800. Between the fifteenth and eighteenth century, there was an average of two economic downturns per country per century (10 + 14 + 9 + 12 episodes divided by six countries divided by four centuries), while the nineteenth and twentieth centuries experienced less than one economic downturn per country per century. In the fifteenth and sixteenth centuries, economic downturns occurred about 8 percent of the time; in the seventeenth and eighteenth centuries, they were experienced 4–5 percent of years; and, in the nineteenth and twentieth centuries, downturns occurred 2–3 percent of the time. Thus, there appears to have been a modest reduction in the likelihood of experiencing downturns over the centuries from the fifteenth century.

Of course, caution should be taken in interpreting these values. While, the nine four-year stretches of pre-1800 economic growth in any of the six economies (see column 1 of Table 1) do coincide with broadly agreed-upon periods of economic improvements; the method used here to identify four-year-phases of growth could

Table 1
Periods of Economic Growth and Decline across Six Economies, 1300–2000
(England/Great Britain, Italy, Holland, Sweden, Spain, and Portugal)

	# of phases of 4-year consecutive 1.5% annual growth rate	*% of years in 4-year consecutive 1.5% annual growth rate*	*# of phases of 3-year consecutive −1.5% annual growth rate*	*% of years in 3-year consecutive −1.5% annual growth rate*
1300s	1	1.1%	2	1.6%
1400s	1	1.0%	10	8.0%
1500s	3	2.3%	14	8.7%
1600s	2	1.3%	9	4.3%
1700s	2	1.3%	12	5.8%
1800s	8	5.3%	4	2.0%
1900s	38	40.0%	4	3.2%

Source: Authors.
Notes: Column 1 represents the number of times countries in the group had four consecutive years of at least 1.5 percent growth in GDP per capita. That is, if this were to happen once in one country and twice in another country within a certain century, it would equal three times in that century. Column 2 shows the same as a percentage of the total number of years (for which data exists) during the specified century. Column 3 represents the number of times a country had three consecutive years of −1.5 percent (or lower) growth rates. Column 4 shows the same as column 3 as a percentage of total number of years.

easily include periods in which the economy did not improve or exclude periods in which the economy did improve.

However, this evidence tentatively indicates that, before the nineteenth century, there was little apparent difference in growth phases and a possible reduction in the frequency of downturns. By comparing the data with the last two centuries, there do appear to be substantial increases in growth phases and reductions in the occurrence of downturns in the nineteenth and especially the twentieth centuries. Explaining the source of these differences could prove to be important for understanding how economies managed to generate sustained economic growth.

Very Long-Run Cycles of Convergence and Divergence

The convergence and divergence of GDP per capita in the very long run is a central question in the literature on economic development. For example, the classic Solow (1956) growth model predicts convergence of less-developed economies with leading economies (for discussion, see Mankiw, Romer, and Weil 1992; see also in this journal the Lucas, 2000, model of convergence driven by the "take-off" date). While some economies have caught up since the end of the nineteenth century, many have remained less developed and fallen behind relative to the leading economies

(Easterlin 1981; Abramovitz 1986; Pritchett 1997). From a very long-run perspective, there has been a great deal of debate about the Great Divergence, when European economies overtook Asian economies like China over the period from the sixteenth to the nineteenth century (Pomeranz 2000; Broadberry and Gupta 2006; Allen, Bassino, Ma, Moll-Murata, and van Zanden 2011; Broadberry 2013). What does the very long-run data presented here have to say about the process of convergence?

With evidence for only a small sample of economies around the world, drawing conclusions about very long-run divergence and convergence at a global scale is inappropriate—in fact, DeLong (1988) showed that often the countries for which historical data exist are the successful economies with high GDP per capita and, therefore, drawing global conclusions based on a historical sub-sample can be very misleading. Thus, it is important to emphasize that the focus in the following discussion is on regional European convergence or divergence. Amongst these six European economies, data availability does not reflect relative success, as there was considerable catching-up and falling behind of particular nations over this 500-year period. Thus, at least tentative conclusions about convergence and divergence for European economies may be drawn from the data available.

A "Little Divergence" between Mediterranean and Northwest European economies has been proposed using some of these new datasets. Broadberry (2013) focuses on explaining the Great Divergence between Europe and Asia and the Little Divergence between northwestern Europe and the rest of Europe from the sixteenth century, arguing that economic structure and institutions determined how particular economies reacted to and were affected by the pivotal shocks (or "critical junctures" in Acemoglu and Robinson 2012) associated with the Black Death in the mid-fourteenth century and the new trade routes between Europe, Asia, and the Americas that opened-up at the end of the fifteenth century. Thus, to some, the shocks were curses; to others, they were blessings in the long run.

However, some additional observations related to convergence and divergence across these countries are also possible. By comparing the position of the leading economy's GDP per capita relative to that of the "following economies," it is possible to discern from Figure 1 the degree of convergence. Using a very limited set of countries (England, Holland, Italy, and Spain), we see that for much of the fourteenth and fifteenth century, the average economy in Europe had a GDP per capita of between 50 and 60 percent of the leading economy, Italy. For a slightly broader set of countries, we see that by 1500, the average economy was 75 percent of the leading economy.[1] By 1600, when Holland had emerged as the new leader, the average economy had fallen to 42 percent of the leader. By 1700, with Holland stagnating and England catching up, this average was 61 percent of the leader. By 1800, with the Industrial Revolution and Great Britain's supremacy, the gap increased and

[1] For this calculation and subsequent calculations in this paragraph, we also include German states (Pfister 2011), France (Álvarez-Nogal and Prados de la Escosura 2013; see also Squicciarini and Voigtländer forthcoming) and what would today be known as Belgium (Buyst 2011), for which new data for each century are available.

the average was 50 percent of the leader. By 1900, for these average European countries, per capita GDP had fallen to 41 percent of the leader (Bolt and van Zanden 2014). In 2000, as other studies have shown (Baumol 1986; Pritchett 1997), the gap had once again dropped amongst these European countries, and the relative average was 84 percent of the leader, the Netherlands.

Thus, amongst the group of economies in Figure 1 and a slightly broader set of countries (and extending the evidence to the present), there appear to be cycles of divergence (in the fourteenth, sixteenth, and eighteenth centuries) and convergence (in the fifteenth, seventeenth, and twentieth centuries) over the very long run. These very long-run cycles of economic divergence and convergence are linked with the emergence of new leaders, the waning of their momentum, and the catching-up by followers. The fourteenth century was probably the beginning of a phase of divergence associated with the rise of Central and Northern Italy. By 1600, Holland was the new leading economy and the (average) relative level of followers had declined, suggesting divergence (Broadberry 2013). In 1800, England was catching up with Holland and taking the lead shortly afterwards; thus, divergence occurred in the eighteenth century and was accentuated during the nineteenth century. Phases of convergence reflected the stagnation (or even decline) of the leader and the process of other economies learning from the leader and perhaps gaining from spillovers.

Finally, building on Quah's (1996, 1997) concept of income mobility, it is worth commenting on the ability of economies to move upwards relative to other followers, though not necessarily to become the leader. Figure 1 shows that there was some mobility and opportunity for economies in Europe to improve their positions and status, which may have been important in determining the outcome of geopolitical tensions amongst European rivals. However, it also shows that there was a tendency for the leader to remain in its position as the wealthiest economy for a century or more, and for the poorest to be stuck for very long periods. Thus, within Europe there was a degree of stratification and the formation of clubs, with some mobility between them (as argued in Durlauf and Johnson 1995).

The Data

How can researchers construct these GDP per capita estimates from the late medieval and early modern eras until the nineteenth century? As discussed earlier, Figure 1 presents six original datasets constructed within the last four years: England/Great Britain, Holland, Northern and Central Italy, Spain, Sweden, and Portugal. Each time series starts and ends in different years and uses a different combination of methods to estimate output. Table 2 summarizes the sources and methods of data construction for agricultural and nonagricultural sectors.

Three main methods have been used to construct historical estimates of GDP and GDP per capita: methods based on direct measures of income, methods based on direct measures of output, and indirect methods. The first method involves estimating national income from data on individual incomes or, more commonly,

Table 2
A Broad Classification of Methods for Estimating GDP per Capita in Selected European Countries

	Period	*Agriculture*	*Industry*	*Service*
England/Great Britain	1270–1870	Output	Output	Output/Proxies
Holland	1348–1510	Demand	Proxies	Proxies
	1510–1807	Output	Output	Output
Italy (Central & Northern Regions)	1310–1861	Demand	Proxies	Proxies
Spain	1254–1850	Demand	Proxies	Proxies
Sweden	1560–1800	Demand	Output	Proxies
Portugal	1500–1850	Demand	Proxies	Proxies

Sources: England/Great Britain (Broadberry et al. 2011); Italy (Malanima 2011); Holland (van Zanden and van Leeuwen 2012); Sweden (Schön and Krantz 2012); Spain (Álvarez-Nogal and Prados de la Escosura 2013); Portugal (Reis, Martins, and Costa 2013; Palma and Reis 2014).

wages—particularly by Clark (2007, 2010). However, when using wages, this approach needs to take account of changes in hours per day worked and days per year worked—or else, it offers only limited evidence on variation in living standards through time. Without taking account of all factors, GDP estimates based on income and output approaches are likely to follow divergent trends over time. Angeles (2008) identifies three key factors determining the divergence: income distribution, per capita labor supply, and relative price changes. He shows that there was considerable divergence in English GDP per capita and real wages between 1700 and 1820 due to increased inequality (Saito 2015), the "Industrious Revolution" (a term used to convey that the Industrial Revolution was a longer-run combination of events than is often considered, as in de Vries 1994), and the increasing relative price of food (Broadberry, Campbell, Klein, Overton, and van Leeuwen 2015).

Given these concerns about using income or wage-based approaches, either the output method or indirect methods (focusing on demand or using proxies), were used to create the six time series presented in Figure 1 and Table 1. Here, some examples will illustrate how GDP estimates were generated using these methods.

The generally preferred approach to estimating national income is the output approach, provided sufficient information is available on the main sources of production (de Vries and van der Woude 1997, p. 721; Maddison 2007, pp. 316–19; Broadberry et al. 2011, p. 2). The rich accounts of English and British economic history since late medieval times offered an opportunity to estimate pre-1870 annual GDP using an output approach that separates the agricultural, industrial, and service sectors.

England/Great Britain

For medieval agriculture, three data sources are available: first, the Medieval Accounts Database of Campbell (2000, 2007) is based mainly on a large sample of

manorial accounts, which were drawn up using a common template by the reeve (a servant of the lord), who managed the demesne (the land owned by the lord) under close supervision of the lord's bailiff or steward and recorded detailed information on land use, crops, animals, and livestock products. Second, the Early Modern Probate Inventories Database covering the mid-sixteenth to the mid-eighteenth centuries, assembled by Overton (1991; 2000) and Overton, Whittle, Dean, and Haan (2004), pulls together similar information extracted from inventories drawn up by the Church Commissioners for the estates of farmers. Third, the Modern Farm Accounts Database of Turner, Beckett, and Afton (2001), covering a period from 1720 until 1913, is based on a large sample of accounts produced by farmers and kept in local record offices. Agricultural outputs were calculated by multiplying the acreage for each crop by the yield per acre. Broadberry et al. (2015) estimate the total acreage. The trends in yields were split into three main time periods, based on the data sources available. For pastoral output, a similar procedure was undertaken, multiplying the number of animals by the share producing and their yields. Prices for individual crops and animal products are used to convert the output into current prices and create weights for the agricultural real output index.

Production estimates or indicators existed for the key English industries up to 1700, based on careful reconstruction from archival records by generations of scholars. Crucial sources included Carus-Wilson and Coleman (1963) for wool and woolen cloth, drawing on detailed records of exports of wool and woolen cloth; King (2005) for iron, based on a reconstruction of all blast furnaces, their capacity, and knowledge of when they were in blast; and Hatcher (1973) for tin, based on receipts of coinage dues. Outputs related to leather and food processing were estimated by Broadberry et al. (2015) on the basis of key inputs obtained from the reconstruction of the agricultural sector. For the construction sector, detailed information on cathedral building is combined with an index of housebuilding based on population and urbanization, while the growth of book production is based on titles listed by the British Library. These series are combined to generate an index of industrial production from 1264 to 1700. Crafts and Harley (1992) offer an index from 1700 until 1870, to which Broadberry et al. (2015) add some new series.

For the service sector, an approach developed by Deane and Cole (1962), was followed with some adjustments. The service sector is broken down into commerce, housing, domestic services, and government. The commerce indicator is based on combining estimates of domestic trade (the volume of agricultural and industrial output adjusted for the growing share that was marketed) and international trade (derived from the detailed records of trade that were kept for taxation purposes), freight transport (based on merchant shipping tonnage, distances traveled on the main trade routes, and volumes shipped), and financial services (using the velocity of money, derived by comparing estimates of the stock of money with existing estimates of nominal, as opposed to real, national income). Housing and domestic services were assumed to grow at the same rate as population. Government activity is measured based on its revenue, which exists in detailed annual exchequer accounts back to the early twelfth century (O'Brien and Hunt 1999).

The three real output series for the agricultural, industrial, and service sectors were combined using a set of sectoral weights that capture the changing structure of the economy. The starting point is an input-output table for 1841 from Horrell, Humphries, and Weale (1994). The nominal value-added shares for 1841 are projected back using the sectoral real output series reflated to convert them into nominal series. The principal sources for the price series used include Clark (2004, 2005, 2006), Beveridge (1939), and Thorold Rogers (1866–1902). Value-added shares for each sector are derived in this way at roughly 50-year intervals and used to create a chained index of GDP, following Feinstein (1972). To estimate GDP per capita, this aggregate GDP series is divided by population, taken from Wrigley and Schofield (1989) and Wrigley, Davies, Oeppen, and Schofield (1997) for the period since 1541, and derived from information on the number of tenants in a regionally representative sample of manors using the method of Hallam (1988) for the pre-1541 period.

Italy

The indirect approach to estimating GDP per capita depends on modeling or using proxies to generate indicators of economic output. Particularly for agricultural production, where demand is deemed to be relatively stable, a model of agricultural demand is used. For example, the lack of evidence on agricultural production in Italy prior to the mid-nineteenth century led Malanima (2011) to use the demand approach. Estimates of agricultural production start with the assumption that they are equal to consumption. While there might be some imports and exports, Malanima (2011) argues that the net value of these imports and exports are negligible for Central and Northern Italy. Thus, estimates of agricultural consumption will provide a close indicator of production.

The exercise for Italy involved estimating per capita agricultural consumption based on a model of demand (including income and price elasticities) and data on consumer income levels and real prices of agricultural production and industrial products (as substitutes). A number of other historical studies, pioneered by Crafts (1980) and more recently developed by Allen (2000), have used estimates of income elasticities of agricultural products ranging from 0.3 to 0.9. Guided by these previous studies, and Italian estimates from 1861 to 1910 (Federico 2003), Malanima (2011) selected an income elasticity of 0.4. The previous historical studies reviewed had used a cross-price elasticity of 0.1—in other words, agricultural and industrial products are seen as weak substitutes for one another. The sum of the income, own-price, and cross-price elasticities are assumed equal to 0 (relying on the "adding-up" property in linear models described in Deaton and Muellbauer 1980, p. 16), which helps to guide the value of the own-price elasticity (−0.5). Thus, based on these elasticities, and on data for wages (acting as a proxy for income) and for the real prices of agricultural and nonagricultural products, Malanima (2011) estimated the per capita agricultural consumption and, hence, an indicator of production. Price and wage data were collected systematically from institutions such as schools, hospitals, and government departments for many European countries reaching back to the medieval period.

For industry and service sectors, indirect production estimates often depend on long-run trends in urbanization rates, and Bairoch's (1988) dataset of European towns greater than 5,000 inhabitants going back 1,000 years is crucial for this approach. Urbanization rates offer an indicator of the share of nonagricultural activities, as town and city dwellers are not likely to be involved in crop cultivation or pastoral activities. Naturally, an indirect approach is generally only used when the direct approaches using output or income are not possible due to a lack of data. However, given the lack of detailed income and output data, indirect methods are starting to be used more in this area of research.

For Central and Northern Italy, the share of nonagricultural output between 1861 and 1936 was regressed on urbanization rates. The coefficient of the relationship was key to estimating output before 1861. However, without taking account of non-urban industry over the centuries, there would have been a risk of overestimating late medieval output. Thus, combining the coefficient and the urbanization rates with an index of the share of non-urban workers (based on Allen 2000), the share of nonagricultural output was estimated back to 1310. With an estimate of per capita agricultural output and of the share of nonagricultural output, it was possible to construct a GDP per capita series from 1310 until 1861. For consistency, this series was linked to a series for Central and Northern Italy (Daniele and Malanima 2007). While Bolt and van Zanden (2014, p. 635) raise some questions about these estimate, this series provides a valuable (and the only) indicator of long-run growth rates related to Italy.

Putting these various estimates together with an estimate of per capita agricultural output and of the share of nonagricultural output (sometimes separated into industrial and service sectors), it is possible to construct a GDP series. This value is then divided by the geographical boundary's population to produce per capita GDP.

General Considerations

As one might expect, there is considerable uncertainty about the margins of error surrounding the estimates of per capita GDP given here. Readers seeking a guide can begin by considering the methodologies used. For instance, van Zanden and van Leeuwen (2012) propose that the data for Holland before 1510 is far less reliable than the data for Holland between 1510 and 1807 because the former was based on modeling agricultural demand and proxies for industry and services, whereas the latter was based on output measures. Thus, data based on output measures, such as those for England/Great Britain, Holland from 1510, and to a lesser extent Sweden, are likely to be more accurate.

Of course, the brief description offered here is a gross simplification of the complexities involved in the truly mammoth task of estimating historical GDP per capita. Generating these estimates includes identifying and pulling together hundreds of data sources and deciding upon historically justifiable assumptions. Great care must be taken in ensuring that prices and baskets of goods are comparable and benchmarked over time (Prados de la Escosura 2000, 2015).

For greater detail, the reader is encouraged to consult the original papers for each country: England/Great Britain (Broadberry et al. 2011, 2015); Italy (Malanima

2011); Holland (van Zanden and van Leeuwen 2012); Sweden (Schön and Krantz 2012); Spain (Alvarez-Nogal and Prados de la Escosura 2013); and Portugal (Reis, Martins, and Costa 2013; Palma and Reis 2014). Data for these countries (and many others) from the nineteenth century to the present are much better known, and information about how they were constructed can be found in a number of sources—probably the most updated discussion of these sources, associated with the Maddison Project, is Bolt and van Zanden (2014).

Conclusion

While the data presented here have notable limitations, they offer the first detailed picture of economic development in Western Europe for the 500 years before the Industrial Revolution. Clearly, the received wisdom that preindustrial economies more than two centuries ago were stagnant is not true. These economies had major and minor phases of economic growth before the nineteenth century, some lasting more than 50 years, which often led to substantial long-run improvements in per capita income—even if these growth rates were not ultimately sustained.

Subsequent research in this area will continue to check and re-estimate the very long-run data on per capita GDP for these economies and others. However, it will also move on to the challenges of explaining the patterns in the data theoretically and econometrically. For example, it will be useful to consider the pre-1800 growth episodes and reversals in European economies, changes in the likelihood of phases of growth and of decline over time, and periods and cycles of divergence and convergence. In addition to GDP per capita, there are more very long-run data available on a number of traditional explanatory variables, such as institutions, human capital, and population changes, as well as shocks like plagues and wars.

Although research using this very long-run data is still in its early stages, it is already offering some insight and challenges for how we think about the processes of economic growth. For example, an economy in which per capita income stagnated for 500 years would have been very different from the preindustrial European economies that experienced multiple peaks and troughs. Each substantial peak and trough in per capita income implied a process of change—of new technologies, institutions, beliefs, and behavior. Each step in history sets the stage for the next step. Thus, the six preindustrial European economies studied here were changing, with agents adjusting to new incentives and constraints and in some ways adopting a substantially new economic system roughly every 50 to 100 years. It seems very possible (even probable) that economies in other regions of the world experienced major peaks and troughs—as China did in the eleventh century (Broadberry, Guan, and Li 2014). However, the dynamism of the rises and falls in European economies from the fourteenth century may offer a clue to the Great Divergence between Europe and China during this time period (Broadberry 2013), which suggests a possible avenue of future research.

Preindustrial Europe also showed patterns of divergence and convergence. Divergence was associated with a new leading economy. Convergence was associated

with phases of economic stagnation or decline amongst leading economies. World economic leaders at one time often seem to struggle to grow beyond a certain range of economic development: for example, this described China in the eleventh century, Italy in the fifteenth century, Holland in the eighteenth century, and even England in the late nineteenth century. In time, a few economies converged on the leader and then, when these catch-up economies developed new technologies and institutions, one of them overtook the leader. It is intriguing to speculate as to whether England, the world leader in per capita GDP in the late nineteenth century, might have stagnated had other economies—like the United States and later Germany—not overtaken it and had England been unable to import new technologies, modes of management, and institutions.

The very long-run historical evidence presented here resolves what had previously appeared to be a major difference between recent developing economy growth patterns and the received wisdom on preindustrial patterns. Received wisdom held that European countries at low levels of economic development before the Industrial Revolution were stagnant. However, empirical studies of developing economies during the last century or so have indicated that GDP per capita has tended to be characterized by spurts of high growth, with periods of stagnation and decline, especially in sub-Saharan Africa (Easterly and Levine 1997; Pritchett 2000; Durlauf, Johnson, and Temple 2006; Easterly 2006; Pinkovskiy and Sala-i-Martin 2014). The findings here suggest that historical patterns of economic growth and decline in preindustrial Europe may have been broadly similar to those of present-day developing economies—another area of ongoing and future research.

Finally, many contending theoretical explanations for past GDP per capita start from the assumption of stagnant economies followed by an economic take-off. Such theories need adjustment to take account of the new evidence. For all of these questions, and many others, the next few years promise exciting advances in our understanding of very long-run economic growth.

■ *We would like to thank Simon Dietz, Baran Doda, Gordon Hanson, Ralph Hippe, Alexander Klein, Enrico Moretti, Suresh Naidu, Leandro Prados de la Escosura, Debraj Ray, Martin Stuermer, Timothy Taylor, and John van Reenen, as well as participants at the AEA annual meeting, for discussions related to this paper, and Bas van Leeuwen, Leandro Prados de la Escosura, Nuno Palma, Jaime Reis, and Nico Voigtländer for their data. Support for this project from the Global Green Growth Institute (GGGI), and the ESRC is gratefully acknowledged.*

References

Abramovitz, Moses. 1986. "Catching Up, Forging Ahead, and Falling Behind." *Journal of Economic History* 46(2): 385–406.

Acemoglu, Daron, and James Robinson. 2012. *Why Nations Fail? The Origins of Power, Prosperity, and Poverty.* Crown Publishers.

Allen, Robert C. 2000. "Economic Structure and Agricultural Productivity in Europe, 1300–1800." *European Review of Economic History* 4(1): 1–26.

Allen, Robert C., Jean-Pascal Bassino, Debin Ma, Christine Moll-Murata, and Jan Luiten van Zanden. 2011. "Wages, Prices, and Living Standards in China, 1738–1925: In Comparison with Europe, Japan, and India." *Economic History Review* 64(s1): 8–38.

Álvarez-Nogal, Carlos, and Leandro Prados de la Escosura. 2013. "The Rise and Fall of Spain (1270–1850)." *Economic History Review* 66(1): 1–37.

Angeles, Luis. 2008. "GDP per Capita or Real Wages? Making Sense of Conflicting Views on Pre-industrial Europe." *Explorations in Economic History* 45(2): 147–63.

Bairoch, Paul. 1988. *Cities and Economic Development: From the Dawn of History to the Present.* University of Chicago Press.

Baumol, William J. 1986. "Productivity Growth, Convergence and Welfare: What the Long-Run Data Show." *American Economic Review* 76(5): 1072–85.

Beveridge, William. 1939. *Prices and Wages in England from the Twelfth to the Nineteenth Century,* Vol. I: *Price Tables: Mercantile Era*. London: Longmans, Green.

Bolt, Jutte, and Jan Luiten van Zanden. 2014. "The Maddison Project: Collaborative Research on Historical National Accounts." *Economic History Review* 67(3): 627–51.

Broadberry, Stephen. 2013. "Accounting for the Great Divergence." Economic History Working Paper 184/13, London School of Economics and Political Science, Department of Economic History.

Broadberry, Stephen, Bruce Campbell, Alexander Klein, Mark Overton, and Bas van Leeuwen. 2011. "British Economic Growth, 1270–1870: An Output-Based Approach." Available at: http://www2.lse.ac.uk/economicHistory/whosWho/profiles/sbroadberry.aspx.

Broadberry, Stephen, Bruce Campbell, Alexander Klein, Mark Overton, and Bas van Leeuwen. 2015. *British Economic Growth, 1270–1870.* Cambridge University Press.

Broadberry, Stephen, Hanhui Guan, and David Daokui Li. 2014. "China, Europe and the Great Divergence: A Study in Historical National Accounting, 980–1850." Economic History Department Paper, London School of Economics.

Broadberry, Stephen, and Bishnupriya Gupta. 2006. "The Early Modern Great Divergence: Wages, Prices and Economic Development in Europe and Asia, 1500–1800." *Economic History Review* 59(1): 2–31.

Buyst, Eric. 2011. "Towards Estimates of Long Term Growth in the Southern Low Countries, ca.1500–1846." Paper for the "Quantifying Long Run Economic Development" conference at the University of Warwick in Venice, March 22–24. http://www2.warwick.ac.uk/fac/soc/economics/events/seminars-workshops-conferences/conferences/venice3/programme/buyst.pdf.

Campbell, Bruce M. S. 2000. *English Seigniorial Agriculture, 1250–1450.* Cambridge University Press.

Campbell, Bruce M. S. 2007. *Three Centuries of English Crop Yields, 1211–1491.* A database. http://www.cropyields.ac.uk.

Carus-Wilson, Eleanora M., and Olive Coleman. 1963. *England's Export Trade, 1275–1547.* Oxford University Press.

Clark, Gregory. 2004. "The Price History of English Agriculture, 1209–1914." In *Research in Economic History*, vol. 22, edited by Susan Wolcott and Christopher Hanes, 41–125. Emerald Group Publishing.

Clark, Gregory. 2005. "The Condition of the Working Class in England, 1209–2004." *Journal of Political Economy* 113(6): 1307–40.

Clark, Gregory. 2006. "England, Prices and Wages since 13th [Century]." Global Price and Income History Group, University of California, Davis. Available at: http://gpih.ucdavis.edu/Datafilelist.htm.

Clark, Gregory. 2007. "The Long March of History: Farm Wages, Population, and Economic Growth, England 1209–1869." *Economic History Review* 60(1): 97–135.

Clark, Gregory. 2010. "The Macroeconomic Aggregates for England, 1209–2008." In *Research in Economic History* vol. 27, edited by Susan Wolcott and Christopher Hanes, 51–140. Emerald Group Publishing.

Crafts, Nicholas F. R. 1980. "Income Elasticities of Demand and the Release of Labour by Agriculture during the British Industrial Revolution." *Journal of European Economic History* 9(1): 153–68.

Crafts, Nicholas F. R., and C. Knick Harley. 1992. "Output Growth and the Industrial Revolution: A Restatement of the Crafts-Harley View." *Economic History Review* 45(4): 703–30.

Daniele, Vittorio, and Paolo Malanima. 2007. "Il prodotto delle regioni e il divario Nord-Sud in Italia (1861–2004)." *Rivista di Politica Economica* 97(2): 267–316.

Deane, Phyllis, and William A. Cole. 1962. *British Economic Growth 1688–1959.* Cambridge University Press.

Deaton, Angus, and John Muellbauer. 1980. *Economics and Consumer Behaviour.* Cambridge University Press.

DeLong, J. Bradford. 1988. "Productivity Growth, Convergence, and Welfare: Comment." *American Economic Review* 78(5): 1138–54.

de Vries, Jan. 1994. "The Industrial Revolution and the Industrious Revolution." *Journal of Economic History* 54(2): 249–70.

de Vries, Jan, and Ad van der Woude. 1997. "The First Modern Economy: Success, Failure and Perseverance of the Dutch Economy, 1500–1815." Cambridge University Press.

Drelichman, Mauricio. 2005. "The Curse of Moctezuma: American Silver and the Dutch Disease." *Explorations in Economic History* 42(3): 349–80.

Durlauf, Steven N., and Paul A. Johnson. 1995. "Multiple Regimes and Cross-Country Growth Behaviour." *Journal of Applied Econometrics* 10(4): 365–84.

Durlauf, Steven N., Paul A. Johnson, and Jonathan R. W. Temple. 2006. "Growth Econometrics." In *Handbook of Economic Growth*, Vol. 1, edited by P. Aghion, and S. N. Durlauf, 555–677. Elsevier.

Easterlin, Richard. 1981. "Why Isn't the Whole World Developed." *Journal of Economic History* 41(1): 1–19.

Easterly, William. 2006. "Reliving the 50s: The Big Push, Poverty Traps, and Takeoffs in Economic Development." *Journal of Economic Growth* 11(4): 289–318.

Easterly, William, and Ross Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *Quarterly Journal of Economics* 112(4): 1203–50.

Federico, Giovanni. 2003. "Le nuove stime della produzione agricola italiana, 1860–1910: Primi risultati e implicazioni." *Rivista di Storia Economica*, issue 3, pp. 359–82.

Feinstein, Charles H. 1972. *National Income, Expenditure and Output of the United Kingdom, 1855–1965*. Cambridge University Press.

Galor, Oded. 2005. "From Stagnation to Growth: Unified Growth Theory." In *Handbook of Economic Growth*, Vol. 1A, edited by P. Aghion and S. N. Durlauf, 171–293. Amsterdam: North Holland.

Goldthwaite, Richard A. 2009. *The Economy of Renaissance Florence.* Johns Hopkins University Press.

Hallam, Henry E. 1988. "Population Movements in England, 1086–1350." In *The Agrarian History of England and Wales,* Vol. 2: *1042–1350*, edited by H. E. Hallam, pp. 508–93. Cambridge University Press.

Hansen, Gary D., and Edward C. Prescott. 2002. "Malthus to Solow." *American Economic Review* 92(5): 1205–17.

Hatcher, John. 1973. *English Tin Production and Trade before 1550.* Oxford: Clarendon Press.

Hausmann, Ricardo, Lant Pritchett, and Dani Rodrik. 2005. "Growth Accelerations." *Journal of Economic Growth* 10(4): 303–29.

Hodgett, Gerald A. J. 1972 [2006]. *A Social and Economic History of Medieval Europe.* London: Methuen. (2006 edition published by Routledge).

Horrell, Sara, Jane Humphries, and Martin Weale. 1994. "An Input-Output Table for 1841." *Economic History Review* 47(3): 545–66.

Keynes, John Maynard. 1931. "Economic Possibilities for Our Grandchildren." In *Essays in Persuasion*, by John Maynard Keynes. London: Macmillan.

King, Peter. 2005. "The Production and Consumption of Bar Iron in Early Modern England and Wales." *Economic History Review* 58(1): 1–33.

Lucas, Robert E. 2000. "Some Macroeconomics for the 21st Century." *Journal of Economic Perspectives* 14(1): 159–68.

Maddison, Angus. 1982. *Phases of Capitalist Development; Booms, Crises, and Globalizations.* Oxford University Press.

Maddison, Angus. 1995. *Monitoring the World Economy, 1820–1992.* OECD Development Center.

Maddison, Angus. 2003. *The World Economy: Historical Statistics.* OECD.

Maddison, Angus. 2007. *Contours of the World Economy 1–2030 AD: Essays in Macro-Economic History.* Oxford University Press.

Malanima, Paolo. 2011. "The Long Decline of a Leading Economy: GDP in Central and Northern Italy, 1300–1913." *European Review of Economic History* 15(2): 169–219.

Mankiw, N. Greg, David Romer, and David N. Weil. 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107(2): 407–37.

North, Douglas C., and Barry R. Weingast. 1989. "Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England." *Journal of Economic History* 49(4): 803–32.

O'Brien, Patrick K., and Philip A. Hunt. 1999. "England, 1485–1815." In *The Rise of the Fiscal State in Europe, c.1200–1850*, edited by R. Bonney, pp. 53–100. Oxford University Press.

Overton, Mark. 1991. "The Determinants of Crop Yields in Early Modern England." In *Land, Labour and Livestock: Historical Studies in European Agricultural Productivity*, edited by B. M. S. Campbell and M. Overton, 284–322. Manchester University Press.

Overton, Mark. 2000. "Prices from Probate Inventories." In *When Death Do Us Part: Understanding and Interpreting the Probate Records of Early Modern England,* edited by T. Arkell, N. Evans, and N. Goose, 120–43. Oxford: Leopard's Head Press.

Overton, Mark, Jane Whittle, Darron Dean, and Andrew Hann. 2004. *Production and Consumption in English Households, 1600–1750.* London: Routledge.

Palma, Nuno, and Jaime Reis. 2014. "Portuguese Demography and Economic Growth, 1500–1850." Paper for the "Accounting for the Great Divergence" conference at the University of Warwick in Venice, May 22–24, 2014. http://www2.warwick.ac.uk/fac/soc/economics/research/centres/cage/events/conferences/greatdivergence14/portuguese_demography_and_economic_growth_1500-1850_-_may_19th_2014.pdf.

Pfister, Ulrich. 2011. "Economic Growth in Germany, 1500–1850." Paper for the "Quantifying Long Run Economic Development" conference at the University of Warwick in Venice, March, 22–24, 2011. https://www.wiwi.uni-muenster.de/wisoge/organisation/personen/pfister/forschung/Growth-Venice-2011.pdf.

Pinkovsky, Maxim, and Xavier Sala-i-Martin. 2014. "Lights, Camera . . . Income! Estimating Poverty Using National Accounts, Survey Means, and Lights." Federal Reserve Bank of New York Staff Report 669.

Pomeranz, Kenneth. 2000. *The Great Divergence: China, Europe, and the Making of the Modern World Economy*. Princeton University Press.

Prados de la Escosura, Leandro. 2000. "International Comparisons of Real Product, 1820–1990: An Alternative Data Set." *Explorations in Economic History* 37(1): 1–41.

Prados de la Escosura, Leandro. 2015. "Mismeasuring Long-Run Growh: The Bias from Splicing National Accounts—The Case of Spain." *Cliometrica*. Published online May 28.

Pritchett, Lant. 1997. "Divergence, Big Time." *Journal of Economic Perspectives* 11(3): 3–17.

Pritchett, Lant. 2000. "Understanding Patterns of Economic Growth: Searching for Hills among Plateaus, Mountains, and Plains." *World Bank Economic Review* 14(2): 221–50.

Quah, Danny T. 1996. "Empirics for Economic Growth and Convergence." *European Economic Review* 40(6): 1353–75.

Quah, Danny T. 1997. "Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs." *Journal of Economic Growth* 2(1): 27–59.

Reinert, Sophus A. 2011. *Translating Empire: Emulation and the Origins of Political Economy*. Harvard University Press.

Reis, Jaime, Conceição A. Martins, and Leonor F. Costa. 2013. "From Major Power to Economic Backwater: Portugal's GDP 1500–1850." Paper presented at "Accounting for the Great Divergence" conference at the University of Warwick, May 28, 2013. http://www2.warwick.ac.uk/fac/soc/economics/research/centres/cage/events/conferences/mayconf/tuesday28may/reis.pdf.

Rodrik, Dani. 1999. "Where Did All the Growth Go? External Shocks, Social Conflict and Growth Collapses." *Journal of Economic Growth* 4(4): 385–412.

Saito, Osamu. 2015. "Growth and Inequality in the Great and Little Divergence Debate: A Japanese Perspective." *Economic History Review* 68(2): 399–419.

Schön, Lennart, and Olle Krantz. 2012. "The Swedish Economy in the Early Modern Period: Constructing Historical National Accounts 1560–2000." *European Review of Economic History* 16(4): 529–49.

Smith, Adam. 1776. *The Wealth of Nations*. London: Methuen & Co.

Solow, Robert M. 1956. "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics* 70(1): 65–94.

Squicciarini, Mara P., and Nico Voigtländer. Forthcoming. "Human Capital and Industrialization: Evidence from the Age of Enlightenment." *Quarterly Journal of Economics*.

Thirsk, Joan. 1978. *Economic Policy and Projects: The Development of a Consumer Society in Early Modern England*. Clarendon Press. Oxford.

Thorold, Rogers, J. E. 1866–1902. *A History of Agriculture and Prices in England*, vols. 1–6. Oxford: Clarendon Press.

Turner, Michael E., John V. Beckett, and Bethanie Afton. 2001. *Farm Production in England 1700–1914*. Oxford University Press.

van Zanden, Jan Luiten, and Bas van Leeuwen. 2012. "Persistent But Not Consistent: The Growth of National Income in Holland 1347–1807." *Explorations in Economic History* 49(2): 119–30.

Wrigley, E. Anthony, Ros S. Davies, James E. Oeppen, and Roger S. Schofield. 1997. *English Population History from Family Reconstitution, 1580–1837*. Cambridge University Press.

Wrigley, E. Anthony, and Roger S. Schofield. 1989. *The Population History of England, 1541–1871: A Reconstruction*. Cambridge University Press.

Journal of Economic Perspectives—Volume 29, Number 4—Fall 2015—Pages 245–252

Recommendations for Further Reading

Timothy Taylor

This section will list readings that may be especially useful to teachers of undergraduate economics, as well as other articles that are of broader cultural interest. In general, with occasional exceptions, the articles chosen will be expository or integrative and not focus on original research. If you write or read an appropriate article, please send a copy of the article (and possibly a few sentences describing it) to Timothy Taylor, preferably by email at taylort@macalester.edu, or c/o *Journal of Economic Perspectives*, Macalester College, 1600 Grand Ave., Saint Paul, Minnesota, 55105.

Smorgasbord

Luigi Zingales delivered the Presidential Address to the American Finance Association: "Does Finance Benefit Society?" "From Libor fixing to exchange rate manipulation, from gold price rigging to outright financial fraud in subprime mortgages, not a day passes without a news of a fresh financial scandal. After the financial crisis, Americans' trust towards bankers has dropped tremendously and has not yet fully recovered. It is very tempting for us academics to dismiss all these feelings as the expression of ignorant populism. After all, we are the priests of an esoteric religion, only we understand the academic scriptures and can appreciate the truths therein revealed. For this reason, we almost wallow in public disdain

■ *Timothy Taylor is Managing Editor,* Journal of Economic Perspectives*, based at Macalester College, Saint Paul, Minnesota. He blogs at http://conversableeconomist.blogspot.com.*

http://dx.doi.org/10.1257/jep.29.4.245 doi=10.1257/jep.29.4.245

and refuse to engage, rather than wonder whether there is any reason for these feelings. This is a huge mistake. . . . First of all, acknowledge that our view of the benefits of finance is inflated. While there is no doubt that a developed economy needs a sophisticated financial sector, at the current state of knowledge there is no theoretical reason or empirical evidence to support the notion that all the growth of the financial sector in the last forty years has been beneficial to society. In fact, we have both theoretical reasons and empirical evidence to claim that a component has been pure rent seeking. By defending all forms of finance, by being unwilling to separate the wheat from the chaff, we have lost credibility in defending the real contribution of finance. Our second task is to use our research and our teaching to curb the rent-seeking dimension of finance. We should use our research to challenge the existing practices in finance and blow the whistle on what does not work. We should be the watchdogs of the financial industry, not its lapdogs." Published in the *Journal of Finance*, August 2015, vol. 70, no. 4, pp. 1327–63. The version prepared for the address is available at http://faculty.chicagobooth.edu/luigi.zingales/papers/research/Finance.pdf.

W. Kip Viscusi and Ted Gayer discuss "Behavioral Public Choice: The Behavioral Paradox of Government Policy." "[T]he behavioral economics literature . . . frequently recommends 'soft paternalism' policies that seek to change the structure of the choices available to individuals in order to encourage a more desirable outcome. But, as behavioral agents themselves, policymakers and regulators are subject to the same psychological biases and limitations as all individuals. Many, although certainly not all, behavioral economics papers focus on the biases and heuristics of ordinary individuals, while seemingly ignoring that regulators are people too and thus subject to the same psychological forces. One study finds that, of the behavioral economics articles proposing paternalistic policy responses, 95.5% do not contain any analysis of the cognitive abilities of policymakers . . . Professor Cass Sunstein observes, 'For every bias identified for individuals, there is an accompanying bias in the public sphere.'" *Harvard Journal of Law & Public Policy*, 2015, vol. 38, no. 3, pp. 973–1007, http://www.harvard-jlpp.com/wp-content/uploads/2010/01/ViscusiGayer_4.pdf.

Leonard E. Burman, William G. Gale, Sarah Gault, Bryan Kim, Jim Nunns, and Steve Rosenthal take stock of the issues concerning "Financial Transaction Taxes in Theory and Practice." "Proponents advocate the FTT [financial transactions tax] on several grounds. The tax could raise substantial revenue at low rates because the base—the value of financial transactions—is enormous. An FTT would curb speculative short-term and high-frequency trading, which in turn would reduce the diversion of valuable human capital into pure rent-seeking activities of little or no social value. They argue that an FTT would reduce asset price volatility and bubbles, which hurt the economy by creating unnecessary risk and distorting investment decisions. It would encourage patient capital and longer-term investment. The tax could help recoup the costs of the financial-sector bailout as well as the costs the financial crisis imposed on the rest of the country. The FTT—called the 'Robin Hood Tax' by some advocates—would primarily fall on the rich, and the revenues could be used to

benefit the poor, finance future financial bailouts, cut other taxes, or reduce public debt. Opponents counter that an FTT is an 'answer in search of a question'." In the conclusion, they write: "An FTT at the rates being proposed and adopted elsewhere would discourage all trading, not just speculation and rent seeking. It appears as likely to increase market volatility as to curb it. It would create new distortions among asset classes and across industries. As a tax on gross rather than net activity, and as an input tax that is not creditable and thus cascades, the FTT clearly can most optimistically be considered a second-best solution. Over the long term, it appears poorly targeted at the kinds of financial-sector excesses that led to the Great Recession." Tax Policy Center, June 2015, http://www.taxpolicycenter.org/UploadedPDF/2000287-Financial-Transaction-Taxes-in-Theory-and-Practice.pdf.

Roc Armenter discusses "A Bit of a Miracle No More: The Decline of the Labor Share." "Until recently, the division between labor and capital income had not received much attention. The reason was quite simple: Labor's share never ventured far from 62 percent of total U.S. income for almost 50 years—through expansions, recessions, high and low inflation, and the long transition from an economy primarily based on manufacturing to one mainly centered on services. As it happened, the overall labor share remained stable as large forces pulling it in opposite directions canceled each other out—a coincidence that John Maynard Keynes famously called 'a bit of a miracle.' But the new millennium marked a turning point: Labor's share began a pronounced fall that continues today. . . . [U]ntil 2001, the BLS's methodology assigned most of proprietor's income to the labor share, a bit more than four-fifths of it. Since then, less than half of proprietor's income has been classified as labor income. . . . [A]t least one-third and possibly closer to half of the drop in the headline labor share is due to how the BLS treats proprietor's income." *Business Review,* Federal Reserve Bank of Philadelphia, Third Quarter 2015, pp. 1–7, https://www.philadelphiafed.org/-/media/research-and-data/publications/business-review/2015/q3/brq315_a_bit_of_a_miracle_no_more.pdf.

Discussing Health

Ronald Lee and Peter R. Orzag chaired a recent committee on behalf of the National Academies of Sciences, Engineering, and Medicine that discusses "The Growing Gap in Life Expectancy by Income: Implications for Federal Programs and Policy Responses." "The implication of these differential trends is that the gap in life expectancies is expanding rapidly. For males born in the 1930 cohort, the highest quintile's life expectancy at age 50 is 5.1 years longer than the lowest quintile's. For males born in the 1960 cohort, the projected gap widens to 12.7 years. For females, the results appear even more pronounced, although the estimates are less reliable. . . . The result is that the gap in life expectancy between high-earning females and low-earning females is projected to expand from 4 years to 13.6 years. . . . Actual and projected population aging raise the costs of government programs for the elderly, leading to fiscal pressures

and a likely policy response. . . . [T]this report evaluates a number of policy changes from the perspective of the impact of differential trends in mortality on lifetime benefits by earnings quintile, a perspective that has rarely been applied in previous policy analyses." 2015, http://www.nap.edu/catalog/19015/the-growing-gap-in-life-expectancy-by-income-implications-for.

The World Health Organization, in its *WHO Report on the Global Tobacco Epidemic 2015,* focuses on the theme "Raising taxes on tobacco." "Despite the fact that raising tobacco taxes to more than 75% of the retail price is among the most effective and cost-effective tobacco control interventions (it costs little to implement and increases government revenues), only a few countries have increased tobacco taxes to best practice level. . . . with only 10% of the world's people living in countries with sufficiently high taxes . . ." "Research from high-income countries generally finds that a 10% price increase will reduce overall tobacco use by between 2.5% and 5% (4% on average). . . . Most estimates from low- and middle-income countries show that a 10% price increase will reduce tobacco use by between 2% and 8% (5% on average). Studies from a number of countries typically show that half of the decline in tobacco use associated with higher taxes and prices results from reduced prevalence (i.e. from users quitting). The remaining half comes from reduced intensity of use (i.e. users consuming less by switching from daily to occasional smoking, or reducing the number of cigarettes smoked each day). . . . Tax and price increases in Brazil explain nearly half of the 46% reduction in adult smoking prevalence between 1989 and 2010. . . . In Thailand, the Asian Development Bank estimates that 60% of the deaths averted by a 50% tobacco price increase would be concentrated in the poorest third of the population, who would pay only 6% of the increased taxes. . . . In China, research suggests that raising taxes on cigarettes so that they account for 75% of retail prices—up from 40% of the share of price in 2010—would avert nearly 3.5 million deaths that would otherwise be caused by cigarette smoking." 2015. http://apps.who.int/iris/bitstream/10665/178574/1/9789240694606_eng.pdf.

Melissa D. Aldridge and Amy S. Kelley discuss the "Epidemiology of Serious Illness and High Utilization of Health Care." "As of 2011, the top 5 percent of health care spenders (18.2 million people) accounted for an estimated 60 percent of all health care costs ($976 billion) . . . In this high-cost subgroup, total annual costs ranged from approximately $17,500 to more than $2,000,000 per person . . ." The essay appears as Appendix E in a 2015 National Academy of Sciences report from a committee chaired by Philip A Pizzo and David M. Walker called *Dying in America: Improving Quality and Honoring Individual Preferences Near the End of Life.* 2015, http://www.nap.edu/openbook.php?record_id=18748.

Long-Term Interest Rates

The White House Council of Economic Advisers has published "Long-Term Interest Rates: A Survey." "Since the early 1980s, both nominal and real

long-term interest rates have declined. Currently, the nominal interest rate on the 10-year Treasury note hovers well below 3 percent, and the comparable interest rates for Germany and Japan are below one percent. Even at today's low inflation rates, the ex post real interest rate (the nominal rate less realized inflation) on long-term bonds has dipped into negative territory for some advanced countries. . . . The potential explanations are high global saving, demographic changes, low productivity growth, and falling term premiums connected with the aftermath of the global financial crisis. An important question, closely related to the global saving glut and secular stagnation hypotheses, is the extent to which these low interest rates will persist. There is no definitive answer to the question, but many hypotheses point to the possibility that while long-term real and nominal interest rates will likely rise off their current lows, their long-run levels have fallen relative to those that prevailed before the financial crisis." July 2015, https://www.whitehouse.gov/sites/default/files/docs/interest_rate_report_final_v2.pdf.

The Bank of International Settlements writes in its *85th Annual Report:* "Globally, interest rates have been extraordinarily low for an exceptionally long time, in nominal and inflation-adjusted terms, against any benchmark. Such low rates are the most remarkable symptom of a broader malaise in the global economy: the economic expansion is unbalanced, debt burdens and financial risks are still too high, productivity growth too low, and the room for manoeuvre in macroeconomic policy too limited. The unthinkable risks becoming routine and being perceived as the new normal." "Our lens suggests that the very low interest rates that have prevailed for so long may not be 'equilibrium' ones, which would be conducive to sustainable and balanced global expansion. Rather than just reflecting the current weakness, low rates may in part have contributed to it by fuelling costly financial booms and busts. The result is too much debt, too little growth and excessively low interest rates." June 28, 2015, http://www.bis.org/publ/arpdf/ar2015e.pdf.

Collections

João Amador and Filippo di Mauro have edited *The Age of Global Value Chains: Maps and Policy Issues*, a VoxEU.org eBook from the Centre for Economic Policy Research. The volume includes an introduction, 13 essays, and a data appendix. From the introduction by Amador and di Mauro: "Due to coordination costs, proximity was very important up until the mid-1980s. It was only then that the information and communication technology (ICT) revolution made it possible to reduce those costs by enabling complexity to be coordinated at a distance. . . . The possibility of relocating the different stages of production theoretically enabled different tasks within a production process to be performed by geographically dispersed production units. This was termed the 'second unbundling' in international trade, leading to the sharing of production between developed and

developing economies from the mid-1980s onwards. . . . The relocation of these stages of manufacturing to developing countries fostered high growth rates in emerging markets and was further enhanced by domestic policies aimed at attracting foreign capital. As a consequence, the 'second unbundling' reversed the previous industrialisation/non-industrialisation pattern prevalent in developed and developing countries. This change of fortunes represents one of the biggest economic transformations of the last decades and it reshaped, and will continue to shape, the balance of power in both international and economic relations. . . . [A]bout 60% of global trade consists of trade in intermediate goods and services, which are then incorporated at different stages of production." 2015, http://www.voxeu.org/sites/default/files/GVCs-ebook.pdf.

Finance & Development (published by the International Monetary Fund) has a seven-paper symposium titled "Latin America: Finding Its Footing." The tone of the discussion concerns "the many challenges facing the region today" and how to go about "avoiding a prolonged slowdown." The lead essay by José Antonio Ocampo sets the stage: "But this positive picture has changed dramatically. Growth per capita ground to a halt in 2014 and much of the region is again viewed with a sense of forgone promise. . . . In contrast to the halcyon decade that ended in 2013, the recent economic performance of Latin America has been poor. Growth fell sharply in 2014 to just 1.1 percent—barely above the region's current low population growth of 1.0 percent—and will continue at a similar or even lower rate in 2015 . . . Investment also declined in 2014, and will continue to do so in 2015. Poverty ratios have stagnated at 2012 levels . . . and, although no hard data are yet available, this seems also true of income distribution." September 2015, http://www.imf.org/external/pubs/ft/fandd/2015/09/pdf/fd0915.pdf.

Julie P. Smith discusses "Market, Breastfeeding and Trade in Mother's Milk" as part of a symposium on the economics of breastfeeding. "Human milk is being bought and sold. Commodifying and marketing human milk and breastfeeding risk reinforcing social and gender economic inequities. Yet there are potential benefits for breastfeeding, and some of the world's poorest women might profit. How can we improve on the present situation where everyone except the woman who donates her milk benefits?" *International Breastfeeding Journal*, 2015, 10:9, http://www.internationalbreastfeedingjournal.com/content/pdf/s13006-015-0034-9.pdf. This and eight other articles on aspects of the economics of breastfeeding are available at http://www.internationalbreastfeedingjournal.com/series/economicaspects.

Interviews

Douglas Clement has an "Interview with Amy Finkelstein." On the topic of whether health insurance will tend to attract the relatively unhealthy: "Suppose you have people—in health insurance we often refer to them as the 'worried well'—who are healthy, so a low-risk type for an insurer, but also risk averse: They're worried that if something happens, they want coverage. . . . As a result,

people who are low risk, but risk averse, will also demand insurance, just as high-risk people will. And it's not obvious whether, on net, those with insurance will be higher risk than those without. . . ." On the question of what Medicaid is worth to recipients: "[O]ur central estimate is that the value of Medicaid to a recipient is about 20 to 40 cents per dollar of government expenditures. . . . The other key finding is that the nominally 'uninsured' are not really completely uninsured. We find that, on average, the uninsured pay only about 20 cents on the dollar for their medical care. This has two important implications. First, it's a huge force working directly to lower the value of Medicaid to recipients; they already have substantial implicit insurance. . . . Second and, crucially, the fact that the uninsured have a large amount of implicit insurance is also a force saying that a lot of spending on Medicaid is not going directly to the recipients; it's going to a set of people who, for want of a better term, we refer to as 'external parties.' They're whoever was paying for that other 80 cents on the dollar." *The Region,* Federal Reserve Bank of Minneapolis, September 2015, https://www.minneapolisfed.org/publications/the-region/interview-with-amy-finkelstein.

David A. Price has an "Interview" with Campbell Harvey. Here are some thoughts on the difference between research on finance inside a company and in academia: "To be published in academic finance or economics, the idea must be unique; it's the same in the practice of finance—you're looking to do something that your competitors haven't thought of. There are differences, though. The actual problems that are worked on by practitioners are more applied than the general problems we work on in financial economics. The second difference is that in academic financial economics, you have the luxury of presenting your paper to colleagues from all over the world. You get feedback, which is really useful. . . . In business, it's different; you cannot share trade secrets. You really have to lean on your company colleagues for feedback. The third thing that's different is access to data for empirical finance. When I was a doctoral student, academia had the best data. For years after that, the pioneering academic research in empirical finance relied on having this leading-edge data. That is no longer the case. The best data available today is unaffordable for any academic institution. It is incredibly expensive and that's a serious limitation in terms of what we can do in our research. . . . The fourth difference is the assistance that's available. Somebody in academia might work on a paper for months with a research assistant who might be able to offer five to 10 hours per week. In the practice of management, you give the task to a junior researcher and he or she will work around the clock until the task is completed. What takes months in academic research could be just a few days. The fifth difference is computing power. Academics once had the best computing power. We have access to supercomputing arrays, but those resources are difficult to access. In the practice of management, companies have massive computer power available at their fingertips. For certain types of studies, those using higher frequency data, companies have a considerable advantage." *Econ Focus,* Federal Reserve Bank of Richmond, First Quarter 2015, 26–30, https://www.richmondfed.org/publications/research/econ_focus/2015/q1/interview.

Discussion Starters

In David Colander's "Economics with Attitude" column, he writes: "Economic Theory Has Nothing to Say about Policy (and Principles Textbooks Should Tell Students That)." "Let me start with a quiz: Question 1: According to economic theory, whenever possible government should avoid tariffs. Question 2: According to economic theory, the minimum wage lowers the welfare of society. Question 3: According to economic theory, if there are no externalities, the market is the preferable way of allocating resources. The correct answer to each of these questions is false; economic theory, on its own, has nothing to say about policy." *Eastern Economic Journal*, Fall 2015, vol. 41, no. 4, pp. 461–65, http://www.palgrave-journals.com/eej/journal/v41/n4/pdf/eej201531a.pdf.

The Task Force on Federal Regulation of Higher Education, made up of high-level college and university administrators, has published its report called *Recalibrating Regulation of Colleges and Universities*. "Another far-reaching analysis was launched by Vanderbilt University in 2014. Initial findings reveal that approximately 11 percent, or $150 million, of Vanderbilt's 2013 expenditures were devoted to compliance with federal mandates. Nearly 70 percent of these costs were absorbed into different offices, affecting a broad swath of faculty, research staff, administrative staff, and trainees in academic departments." As one of many examples in the report of high costs of federal mandates: "Guidance from the Department both in the Handbook for Campus Safety and Security Reporting and subsequent directives indicate that colleges and universities must report crimes that happen in any building or property they rent, lease, or have any written agreement to use (including an informal agreement, such as one that might be found in a letter, email, or hotel confirmation). Even if no payment is involved in the transaction, any written agreement regarding the use of space gives an institution 'control' of the space for the time period specified in the agreement. . . . Department guidance mandates that schools report on study abroad locations when the school rents space for students in a hotel or other facility, and on locations used by an institution's athletic teams in successive years (e.g., the institution uses the same hotel every year for the field hockey team's away games). As a consequence, institutions must attempt to collect crime data from dozens, if not hundreds, of locations . . . One institution has indicated that it requests data from 69 police departments, covering 348 locations in 13 states and five countries, including police at airports and on military bases. The mandate that colleges and universities must collect data from foreign entities is particularly troublesome. . . . In response to one such request, a foreign government accused a U.S. institution of espionage." February 2015, http://www.help.senate.gov/imo/media/Regulations_Task_Force_Report_2015_FINAL.pdf.

The American Economic Association

Correspondence relating to advertising, business matters, permission to quote, or change of address should be sent to the AEA business office: aeainfo@vanderbilt.edu. Street address: American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203. For membership, subscriptions, or complimentary *JEP* for your e-reader, go to the AEA website: http://www.aeaweb.org. Annual dues for regular membership are $20.00, $30.00, or $40.00, depending on income; for an additional fee, you can receive this journal, or any of the Association's journals, in print. Change of address notice must be received at least six weeks prior to the publication month.

The following Statement of Ownership, Management and Circulation is provided in accordance with the requirements, as contained in 39 U.S.C. 3658. *The Journal of Economic Perspectives* is owned, managed and published by the American Economic Association, a nonprofit scientific organization, located at 2014 Broadway, Suite 305, Nashville, Davidson County, TN 37203-2425. The Editor is Professor Enrico Moretti, *Journal of Economic Perspectives*, 2403 Sidney Street, Suite 260, Pittsburgh, PA 15203. During the preceding twelve months the average number of copies printed for each issue was 6,915; the average paid circulation, 5,787; the average free distribution, 233; the average number of copies distributed, 6,220. Corresponding figures for May 13, 2015, the issue published nearest to filing date: total number of copies printed, 6,969; total paid circulation, 5,819; total free distribution, 226; total distribution, 6,045.

The Journal of

Economic Perspectives

Fall 2015, Volume 29, Number 4

Symposia

Overconfidence

The Future of Retail

Online Higher Education

Articles

AMERICAN ECONOMIC ASSOCIATION