The Journal of

Economic Perspectives

A journal of the
American Economic Association

Fall 2012

The Journal of
Economic Perspectives

A journal of the American Economic Association

The Journal of
Economic Perspectives

Contents
Volume 26 • Number 4 • Fall 2012

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.

From Exxon to BP: Has Some Number Become Better than No Number?

Catherine L. Kling, Daniel J. Phaneuf, and Jinhua Zhao

O n March 23, 1989, the *Exxon Valdez* ran aground in Alaska's Prince William Sound and released over 250,000 barrels of crude oil, resulting in 1300 miles of oiled shoreline, the deaths of 250,000 birds, 2800 otters, over 250 seals, and destruction of nearly uncountable salmon and herring eggs (for details, see http://www.evostc.state.ak.us/facts/index.cfm). This event and its aftermath, graphically illustrated to television viewers around the world, ignited debate about the environmental risks of oil usage, the adequacy of regulatory oversight, and the appropriate compensation for damages suffered. The Exxon spill also ignited a debate within the economics profession concerning the adequacy of methods to value public goods, particularly when the good in question has limited direct use, such as the pristine natural environment of the spill region.

Shortly following the *Valdez* grounding, as legal and regulatory processes began, representatives of the state of Alaska, the U.S. government, and Exxon sought expertise in valuing public goods for the purpose of measuring lost economic value from the spill. In turn, a therefore relatively obscure technique referred to as the *contingent valuation* method received considerable attention. In the contingent valuation method, standard measures of economic value such as willingness to pay or willingness to accept are estimated using responses to survey questions. In contemporary lingo, contingent valuation is part of a broader category of approaches known as *stated preference* methods, which rely on peoples' responses

■ *Catherine Kling is Professor of Economics, Iowa State University, Ames, Iowa. Daniel Phaneuf is Associate Professor of Agricultural and Applied Economics, University of Wisconsin, Madison, Wisconsin. Jinhua Zhao is Professor of Economics, Michigan State University, East Lansing, Michigan. Their email addresses are ckling@iastate.edu, dphaneuf @wisc.edu, and jzhao@msu.edu, respectively.*

to questions about researcher-designed—and therefore hypothetical—changes in environmental quality.

The efficacy of stated preference methods generally, and contingent valuation in particular, is no mere academic debate. Billions of dollars are at stake. A contingent valuation study of the damages from the *Exxon Valdez* spill generated an estimate of $4.9 billion (Carson, Mitchell, Hanemann, Kopp, Presser, and Ruud 2003) in lost economic value. In contrast, a recreation demand study of the damages from the spill yielded an estimate of $3.8 million (Hausman, Leonard, and McFadden 1995). The key explanation for the thousand-fold difference is that the estimate from the contingent valuation study is associated almost entirely with *passive-use* or *non-use* value—the value that people place on something simply because it exists, even if they never directly use the good. In contrast, the recreation study only measured economic damages arising from the loss of actual visits to the area of the spill. The authors of the two Exxon studies acknowledged that their methodologies captured distinct values. Carson et al. pointed out that their survey of non-Alaskans meant that the values would be almost exclusively associated with passive use. Hausman, Leonard, and McFadden (p. 29) likewise wrote: "If $3.8 million seems low, the reader must recall that we have estimated only those damages associated with recreational use. Damages associated with commercial use or damages associated with so-called nonuse values are not included in our estimates." Ultimately the *Exxon Valdez* case was settled through a U.S. District Court consent decree in 1991 (*Exxon Valdez* Oil Spill Council, "Settlement") that has paid out approximately $1 billion in damages and over $2 billion in immediate responses and restoration efforts.

While the conceptual basis for passive use value has been clear since John Krutilla's (1967) contribution in the *American Economic Review*, the only available method for measuring it relies on stated preferences, which immediately raises questions for economists. Are stated preference estimates likely to be inaccurate and devoid of useful information, or can a well-constructed survey generate accurate predictions? After all, economists have long favored analysis that is based on what people do rather than what they say. Given the high stakes involved, stated preference methods came under intense scrutiny during the Exxon legal battle. In its wake, the National Oceanic and Atmospheric Administration (NOAA) in 1992 charged a "Blue Ribbon" panel with the task of studying the efficacy of the conginent valuation method (Arrow, Solow, Portney, Leamer, Radner, and Schuman 1993). An influential symposium appearing in this journal in 1994 subsequently provided arguments for and against the credibility of the method, and an extensive research program published in academic journals has continued to this day. The disparity between the estimates of passive use values and direct use values provide ample explanation for this scrutiny, but it is worth emphasizing that for pristine wilderness areas, passive use may be the largest component of value—and stated preference may be the only game in town when it comes to estimation. Thus, if stated preference approaches are deemed unreliable and environmental damage assessment is limited to direct impacts such as lost productivity, health effects, damaged fisheries, displaced recreation, and similar pathways, then the damage from oil spills, toxic releases, and other accidents in remote locations may result in comparatively small monetized losses.

On April 20, 2010, the *Deepwater Horizon* oil rig affiliated with BP suffered an explosion, triggering the release of nearly five million barrels of crude oil into the Gulf of Mexico—a spill 20 times as large as the *Exxon Valdez*. The accident again led to oiled beaches, the death of seabirds and marine wildlife, and the altering of poorly understood and complex ecosystems. As we write, economists and attorneys are at work drawing on existing studies and undertaking new ones to estimate the economic damages from the spill. Much of the work being conducted as part of the legal process is confidential and ongoing, though early evidence from a recreation study (Alvarez, Larkin, Whitehead, and Haab 2012) and a contingent valuation survey (Larkin 2012) is shortly to appear. BP has already set up a $20 billion trust fund for remediation of environmental damages, of which $6 billion was spent as of mid 2012 (Guarino 2012). The large amounts of money involved are once again likely to spur fundamental questions about the veracity of the public goods valuation methods available to economists. This time, however, two decades of research are on the table to guide the work and inform the debate.

The goal of this paper is to assess what occurred in the academic literature between the Exxon spill and the BP disaster in order to shed light on the fundamental question of the validity of contingent valuation and, more generally, stated preference methods. The two oil spills provide useful bookends for our discussion, and the drama surrounding them helps highlight the importance of public goods valuation for policy and litigation purposes. We stress, however, that the issue of stated preference efficacy is much broader than valuing damages from oil spills, and so most of the discussion that follows will be in a more general framework. In particular, we summarize the most salient findings from the now large stated preference literature.[1] The fundamental question is straightforward: are the values elicited from stated preference methods reliable enough to use in policy analysis and/or litigation? We will rely on theoretical developments, neoclassical and behavioral paradigms, empirical and experimental evidence, and a clearer elucidation of validity criteria to provide a framework for readers to ponder this question. Before doing so, however, we first provide a bit of history and then some necessary background on stated preference methods.

Historical Perspective

A search on the Thomson Reuters Web of Science using "contingent valuation" as the topic returns only 49 journal articles as of 1989. These papers, along with important books by Cummings, Brookshire, and Schulze (1986) and Mitchell and

---

[1] We make no attempt to review thoroughly the now extensive stated preference literature. Several reviews trace the literature from the earliest published suggestion of the method (Ciriacy-Wantrap 1947) and its first implementation by Davis (1963), through the refinements and applications studied in the 1970s, 1980s, and 1990s. See for example Randall, Ives, and Eastman (1974), Cummings, Brookshire, and Schulze (1986), Mitchell and Carson (1989), Carson, Flores, and Meade (2001), Carson and Hanemann (2005), and Bennett (2011).

Carson (1989), comprised the bulk of the published literature at that time. Shortly after the *Exxon Valdez* spill, Congress passed the Oil Pollution Act of 1990, which specifically included lost passive use value as a compensable damage. Congress charged the NOAA with identifying methods to value these damages and, facing the fallout from the Exxon debate, NOAA commissioned a panel chaired by Kenneth Arrow and Robert Solow and charged it with answering a deceptively simple question: Is the contingent valuation method capable of providing estimates of lost nonuse values that are reliable enough to be used in natural resource damage assessments? (The panel was also asked to consider whether passive use should be part of damage assessment, but its affirmative answer has not generated the same attention as its other findings, and so we do not consider it further here.) In January 1993, after reviewing the available literature and accepting testimony from researchers, the NOAA Panel provided its answer (Arrow, Solow, Portney, Leamer, Radner, and Schuman 1993, p. 43):

> [W]e identify a number of stringent guidelines for the conduct of CV [contingent valuation] studies. . . . The Panel concludes that under those conditions (and others specified above), CV studies relay useful information. We think it is fair to describe such information as reliable by the standards that seem to be implicit in similar contexts, like market analysis for new and innovative products and the assessment of other damages normally allowed in court proceedings. . . . Thus, the Panel concludes that CV studies can produce estimates reliable enough to be the starting point of a judicial process of damage assessment, including lost passive-use values.

However, the panel left no doubt that its members had very strong reservations with the method and emphasized their concern about several potential biases and problems identified in the literature at that time. They also provided a set of guidelines that effectively established a list of best practices for the design and implementation of contingent valuation surveys.

In reaching their conclusions, the panel cited evidence from only two studies that compared contingent valuation estimates to elicited actual values for public goods and three studies that compared contingent valuation responses with elicited prices for private goods. Based in part on these early tests of the method's accuracy, the panel concluded "that hypothetical markets tend to overstate willingness to pay for private as well as public goods." In the 1994 symposium in this journal, NOAA panel member Paul Portney (1994) provided an introduction to the contingent valuation methodology and traced the key legal and policy developments up through the completion of the panel's report. In two additional papers, W. Michael Hanemann (1994) argued in favor of the method and Peter Diamond and Jerry Hausman (1994) argued against, with the latter authors raising the provocative question "is some number better than no number?"

With the luxury of hindsight, it is now clear that considerable work remained to be done—either to provide convincing evidence of the method's accuracy or its lack

thereof. First, a commonly accepted set of criteria on how to judge whether stated preference studies were adequate for a given task was missing from the vernacular. Second, it was apparent that more theoretical work was needed to understand if and when stated preference studies should be expected to provide unbiased assessments of the underlying economic values. Finally, much empirical work was needed to test the theory and methods in a wide variety of empirical settings.

The economics profession has risen to the challenge. In contrast to the small literature available at the time of the Exxon spill, by 2010 when the BP disaster occurred, at least 25 books and over 2,500 additional journal articles had been published on contingent valuation. This count likely understates the full collection in that newer types of stated preference studies, including choice experiments, may not be flagged under the "contingent valuation" search. In addition, Carson (2011) has amassed a bibliography of over 7,500 studies, which includes many works not published in the peer-reviewed literature.

Stated Preference Methods: A Short Primer

In this section, we frame our discussion of stated preference accuracy by placing it within the larger context of valuing public goods, also referred to as nonmarket valuation, and explaining a few basics on how it works.[2] Two general approaches are available. One makes use of private behavior in related markets to measure the economic value of a nonmarket good such as environmental quality. For example, data on how far people are willing to travel to reach an outdoor recreation destination of a given quality can be used to estimate the tradeoffs people make between money spent on travel and environmental quality at recreation sites. This type of approach is known as *revealed preference.* Hedonic analysis of housing markets is another common type of revealed preference approach routinely applied to environmental goods. Rather than indirectly inferring value from activity in related markets, *stated preference* approaches directly question individuals via surveys to obtain the information needed to value the nonmarket good. In both approaches, the objective is to measure economic value for a change in a nonmarket good by predicting respondents' willingness to pay, or willingness to accept, for the change. For an increase in environmental quality, willingness to pay (more formally, "compensating variation") is the most the individual would be willing to exchange to achieve the improvement. Likewise willingness to accept ("equivalent variation") is the least the individual would accept to forgo the improvement.

There are different types of stated preference approaches. The best-known, and the subject of the Exxon-era debates, is contingent valuation. In a contingent valuation survey, people are asked questions directly related to their willingness to

---

[2] For more complete treatments, see Champ, Boyle, and Brown (2003), Batemen et al. (2002), or Phaneuf and Requate (forthcoming). For a review of the large literature using stated preference method in health economics see de Bekker-Grob, Ryan, and Gerard (2012).

pay for a specific environmental program, commonly in the form of a yes/no answer to a posted price. A second type of stated preference approach is a choice experiment (Louiviere, Hensher, and Swait 2000; Kanninen 2007), in which a person is asked to consider an environmental commodity that is defined by several attributes. The respondent is presented with discrete options that represent different bundles of the attribute levels and asked to select a preferred alternative. A defining characteristic of choice experiments is that a respondent completes multiple choice tasks and selects from three or more options during each task. While contingent valuation and choice experiments share many design elements, the incentives they present to respondents can differ. At the risk of some confusion, we use stated preference and contingent valuation somewhat interchangeably in this essay, both for continuity with the earlier debates and for simplicity. We stress, however, that insights from choice experiments represent an increasingly important component of the literature.

Stated preference surveys typically share similar structures. To value a specific policy change that moves an environmental resource from one well-defined state to another, the survey needs to first describe the environmental good to respondents in a way that is understandable for a lay participant while remaining true to the underlying science. It then needs to communicate the existing level of environmental quality as well as the change being proposed and, finally, the specific policy intervention that will be used to bring about the change. After the commodity has been described, a survey will typically explain the constructed market and method of payment. A best practice for contingent valuation is to describe the market as a referendum in which the respondents are asked whether they would vote for or against the project in a public vote. Since the answer to a question of this type provides only an upper or lower bound on a respondent's value, statistical methods are used to translate this information into an estimate of the distribution of economic value in the population (Haab and McConnell 2002).

A critical part of the referendum question design is the posted price that the respondent is "offered" in considering whether to vote for the project. A careful experimental design is necessary for efficient estimation of mean willingness-to-pay estimates and large sample sizes are generally needed to achieve the desired precision. Other constructed market details include the conditions for provision (for example, whether a majority must vote in favor) and timing of the project. When presenting the posted price, the survey should also describe the method of payment, which can be coercive or voluntary. The former is usually preferred and includes, for example, changes in property tax rates, surcharges on utility bills, or generally assessed fees. The respondent completes the survey by reading the material describing the issue and then deciding, based on personal preferences and her budget constraint, whether to vote "yes" or "no."

Though it is relatively straightforward to describe the components of a contingent valuation study, actual implementation requires attention to many details. Current best practice for survey design involves the iterative use of focus groups, one-on-one interviews, and pre-testing to verify that the commodity description and

constructed market are appropriate for the purposes of the research. A premium is placed on a high level of specificity in the good being valued and the program being evaluated, since vague or abstract descriptions have been shown to lead to unreliable responses. Also, it is generally accepted that the exercise should seek to value the policy package broadly, rather than the change in the commodity narrowly, since context details should matter for how economic value arises. Finally, most surveys include questions designed to gauge how well respondents understood the material, the confidence they have in their responses, and the rationality of their answers.

To help readers who are unfamiliar with stated preference surveys better understand how such surveys are presented to respondents, we provide an abbreviated version of a contingent valuation question from a study published by Loomis, Kent, Strange, Fausch, and Covich (2000) and used as an example in Haab and McConnell's text (2002). The study concerned the valuation of a set of ecosystem services that would be generated by the purchase of water rights from landowners along the South Platte River in Colorado. Detailed information on the proposed plan's effects on wildlife habitat, erosion control, recreational opportunities, and water purification was provided to respondents. An in-person interviewer then asked respondents the following:

> If the majority of households vote in favor of the South Platte River restoration fund, the 45 miles of river would look like (*in-person interviewer points to a figure showing increased water quality and fish and wildlife*). If a majority votes against, these 45 miles of the South Platte River would remain as they are today, as illustrated by (*in-person interviewer points to a figure showing current management*). If the South Platte River restoration fund was on the ballot in the next election and it cost your household $B each month in a higher water bill, would you vote in favor or against?

The dollar amount $B was randomly filled in with one of twelve values ($1, 2, 3, 5, 8, 10, 12, 20, 30, 40, 50, 100). Based on survey responses from 100 local respondents, Loomis et al. estimated an average willingness to pay of $21 per month or over $250 annually per household for the proposal.

Lessons from Theory: When Should Stated Preference Estimates Match Real Payments?

Economists have long believed that observation of actual behavior in which people bear the consequences of their actions is the key to understanding their motives. In turn, this predisposition has given rise to an inclination to doubt the accuracy of answers provided in a survey context, particularly if it involves reporting more than a factual outcome. Recently, however, researchers have developed theories describing how people behave while answering surveys, given the time and cognitive

energy needed for the task. This is our point of departure for understanding what the necessary conditions are for survey answers to reflect real economic values.

The main theoretical tool has been mechanism design, applied to the problem of understanding when it is in a person's best interest to thoughtfully and truthfully report preferences in a stated preference exercise. In response to a critique by Cummings, Harrison, and Rutström (1995) that survey participants do not have the incentive to answer stated preference questions accurately, Carson and Groves (2007) argue that the necessary conditions for truthful reporting involve using an elicitation mechanism that discourages strategic responses, and fielding the survey in a way that encourages respondents to believe that the study's results could ultimately influence their well-being. These conditions are known as *incentive compatibility* and *consequentiality*, respectively. The need for incentive compatibility in eliciting responses from the public is not new: theorists have long known that departures from single-shot, binary, binding outcome choices provide an incentive for self-interested participants to depart from selection of their most preferred option. Indeed, these arguments are part of what led the NOAA Panel to recommend using a binary choice format for contingent valuation elicitation. The last two decades, however, have seen a much more complete investigation into the many nuanced ways that the design features of a stated preference survey can affect choices.

An important example relates to contributions to public goods. Economic theory predicts that, due to the incentive to free ride, a person's voluntary contribution to a collective good will be smaller than that person's true willingness to pay. However, this incentive can play out in a surprising form in hypothetical surveys. If the respondent believes the survey will be used to decide on the ultimate provision of a public good, that person will have incentive to report *more* than true willingness to pay in a voluntary elicitation in which payment is not binding, in order to influence provision so as to have the opportunity to free ride—and contribute less than stated—should the provision become a reality. For example, Champ, Bishop, Brown, and McCollum (1997) find in a field experiment that hypothetical willingness to donate is substantially larger than the donations they actually collected for a public good with mainly nonuse value, though they are not able to compare either to estimates of the true willingness to pay.

Concern about the role of consequentiality in stated preference survey research arose relatively recently (Carson and Groves 2007). Rather than assuming that respondents have the incentive to answer untruthfully (or truthfully), the consequentiality argument suggests that there are no predictable incentives for an inconsequential survey. Specifically, if the respondent has no reason to believe that her answers will influence an outcome that she cares about (either directly, or indirectly by how the survey results are used), there is no reason to expect that the respondent has dedicated effort to the process, and so the meaningfulness of that person's answers cannot be judged. In contrast, if the survey is consequential in the sense that the respondent thinks its conclusions may ultimately influence something that the respondent cares about, she will have incentive to devote effort. In this case, the truthfulness of the respondent's answers hinges on other factors related to the incentive compatibility

and other characteristics of the survey. Continuing the public good example from above, the person has incentive to bid more than her true willingness to pay in the hypothetical voluntary payment survey only if she believes such an act will influence the probability of the good being provided. Absent this condition, there is no prediction we can make about how the respondent will answer the survey question. Herriges, Kling, Liu, and Tobias (2010) show that estimates of economic value from people who received a consequentiality reminder are systematically different from those who did not. However, empirical work on the effect of consequentiality scripts in stated preference surveys is in its infancy.

In short, careful study of the incentives at work when people answer stated preference questions helps us understand when such answers *should* be expected to match the behavior that would occur in a real payment situation. In hypothetical surveys, respondents must be faced with an incentive-compatible instrument and must believe the survey to be consequential, both in terms of affecting the provision of the good and in terms of creating a binding payment commitment. If a stated preference study does not satisfy the conditions under which responses should be expected to match those of a real exchange, then an observed mismatch should not be counted as evidence against the efficacy of stated preference methods. Of course, the corollary is also true: if these conditions are met, then a mismatch provides strong evidence of failings in the method.

Lessons From Behavioral Economics: Are the Challenges Unique to Stated Preference?

Most economists use the neoclassical paradigm of rational, optimizing agents to analyze observed outcomes, including survey responses. The last two decades, however, have seen the emergence of behavioral economics—a competing paradigm that seeks to explain persistent departures from neoclassical predictions. This raises a question for stated preference methods: if behavioral anomalies are observed in stated preference outcomes, is it because of a failure of the stated preference method or a failure of the neoclassical paradigm to supply correct predictions for comparison? In this section, we describe findings from research in behavioral economics that need to be considered when we evaluate the accuracy of stated preference methods.

The findings of behavioral economics can be grouped into two broad categories: 1) individual preferences may not be well-behaved in the neoclassical sense and/or 2) individuals do not always optimize when making choices. Departures from neoclassical preferences come in many guises. One example that is particularly relevant for stated preference is the endowment effect, which predicts that people require more compensation to part with something already in possession than they would give up to newly acquire it. This can explain the large divergence in willingness to pay and willingness to accept that is often observed in stated preference surveys, and which is sometimes cited as evidence of the method's failings.

A further example concerns "warm glow," which is the name given to the private value a person receives from the action of contributing to a worthy cause beyond the actual value of the good the contribution provides. The role of warm glow has been hotly debated in the stated preference literature, and its existence was cited by Diamond and Hausman (1994) as a major deficiency in the contingent valuation method. Warm glow is now understood to be one of many reasons for pro-social behaviors such as contributing to public goods (Schokkaert 2006). Social norms and other-regarding preferences such as altruism and reciprocity can also lead individuals to value an environmental good more than its private benefits, in hypothetical as well as real settings. Finally, new results on choices under uncertainty, such as over-weighting small probabilities, are almost certainly relevant for understanding how people respond to survey questions about environmental programs since environmental outcomes are generally uncertain.

Departures from optimizing behavior can also occur for several reasons. We highlight two that are particularly relevant for valuing public goods. First, people may make "mistakes" in general due to bounded rationality and bounded self-control. For example, in the theory of mental accounting (Thaler 1990), money is not fungible across all categories of expenses, meaning multiple budgets constrain different types of behavior. Payments for environmental services in this context are not necessarily constrained by the overall budget, but instead by an expense category that may be more or less binding than fully rational optimization would imply. Li, Berrens, Bohara, Jenkins-Smith, Silva, and Weimer (2005) offer a piece of evidence for mental accounting in contingent valuation: They found that respondents had lower willingness to pay for reduction of global warming when they received reminders about their discretionary income and its use for environmental causes, compared with when they received reminders about their household budget only.

Second, rationality may be the result of repeated participation in markets, where mistakes are costly and individuals learn, rather than an intrinsic characteristic of individual decisionmakers. Departures from rationality can therefore be aggravated by complex or unfamiliar decision environments and uncertainties, which often result in rule-of-thumb behaviors (Iyengar and Kamenica 2007). Although such departures are prevalent in experiments and in field studies of individual choices, stated preference surveys might be more prone to anomalies for two reasons: choices in inconsequential surveys might not be salient and not subject to regulation by institutions, and survey respondents might not have much experience with the environmental goods being valued or with the choice circumstances. However, such anomalies can be alleviated by consumer experience (Whitehead, Bloomquiest, Hoban, and Crawford 1995; List 2003), and perhaps by competitive institutions (Slembeck and Tyran 2004). For example, Cherry, Crocker, and Shogren (2003) showed that market-induced rationality spills over to nonmarket valuations: subjects disciplined by real market-like arbitrage showed lower rates of preference reversals, and the reduced rates carried over to hypothetical settings with money as well as wildlife lotteries.

These developments in behavioral economics offer a richer set of testable hypotheses and interpretations of evidence in contingent valuation studies. The alternative paradigm may be useful for explaining the highly heterogeneous and sometimes nonrational individual outcomes observed in stated preference surveys and experiments, even when aggregate outcomes conform to expectations. In this sense, behavioral insights are useful for providing input into the design and evaluation of stated preference surveys (Shogren and Taylor 2008). However, these new theories also raise fundamental questions about validity tests and research design. For instance, if choices are context dependent, preferences formed in exchange institutions might differ from those formed in nonmarket settings (Bowles 1998). This observation casts doubt on the standard practice of comparing estimates from surveys with those from market data, and it challenges the presumption that the latter should automatically be preferred for use in policy analysis and damage assessment, given that some values are not formed from markets. The conundrum is that one must choose a behavioral paradigm first—for example, behavior based on neoclassical preferences or behavior based on reference-dependent preferences—and then design and implement a study to test the accuracy of a stated preference estimate based on that paradigm. If the findings of the accuracy test are negative, this may provide evidence that the stated preference method is inaccurate *or* that an incorrect behavioral paradigm was chosen.

Empirical Evidence on Validity

How can we assess the empirical accuracy of stated preference methods? In most instances there is no observable "true" value against which an estimate can be judged, and so researchers have devised other means of looking at the accuracy of their estimates. Using definitions from the American Psychological Association, Mitchell and Carson (1989) introduced the concept of "validity" in the context of stated preferences. The validity of a method is essentially the degree to which it correctly measures the theoretical construct under consideration. Table 1 contains a summary of the validity concepts that have now become standard in the literature. A generic definition of each type of validity is provided in question form in the second column, and in the third column we present an example of the question in the specific context of assessing the validity of stated preference studies. We consider each type of validity in turn.

Criterion Validity: Do Stated Preferences Estimates Match Real Payments?

Tests for criterion validity compare the prediction from a stated preference exercise to a standard that is thought to be a suitable proxy for the true measurement objective, which typically involves real payments. In many ways, this validity concept is the most central and salient. Criterion validity has mainly been assessed in the literature using experimental methods in the laboratory and field, but there are also a small number of studies that have timed stated preference studies to coincide with an actual binding referendum.

Table 1
Summary of Validity Concepts for Stated Preference Methods

Criterion	Generic question	Specific question
Criterion validity	Does the measure relate favorably to other measures that are considered legitimate criteria (i.e., are believed to be accurate)?	Is the estimate generated by stated preference methods the same as a willingness-to-pay value that would be generated if real payment was made?
Convergent validity	Does the measure correlate well with other measures of the same thing?	Is the estimate generated by a stated preference method the same as the willingness-to-pay value that is estimated from a revealed preference method?
Construct validity	Does the measure correlate as expected to other measures as predicted by theory?	Does the estimate generated by a stated preference method relate to income, prices, and other variables in the way economic theory predicts?
Content validity	Does the measure adequately cover the construct's domain?	Does the estimate arise from the best study design practices—including scenario description, econometric analysis, elicitation format, follow up questions, etc.?

Two types of laboratory experiments have been used to gauge criterion validity. In the first, participants are assigned a value for the experimental good as part of the research design. This design allows the researcher to know with certainty the criterion against which real and hypothetical statements of value are compared. Because the value is assigned to the respondent, as opposed to it having arisen internally from the respondent's own preferences, this is known as an "induced value experiment." An advantage of this protocol is that it allows one to focus on value elicitation, as distinct from value formation. Induced value experiments have primarily been used to examine the accuracy of hypothetical referendum-style elicitation vehicles relative to binding real payment votes (for example, Taylor, McKee, Laury, and Cummings 2001; Vossler and McKee 2006; Murphy, Stevens, and Yadav 2010). The results generally show that the distribution of values from hypothetical votes matches the induced-value criterion in aggregate. These findings suggest that a necessary condition for stated preference criterion validity is met. Specifically, when we abstract from the value formation step, there is robust evidence that individuals can be induced to reveal their private willingness to pay for a public good in a properly designed hypothetical situation.

In the second type of experiment, participants' actual values for a real commodity are used as the criterion. These are known as "homegrown value experiments" because participants' own (or homegrown) preferences are the basis for establishing the standard for comparison. In the typical experiment, the criterion is established by a real payment mechanism. For a public good, this takes a referendum format in which all participants must pay a given amount if a majority

votes in favor. The results from hypothetical elicitation formats are then compared to the real payment mechanism as a test of validity. A consistent finding for this type of experiment is that stated values are higher than their real counterparts; this phenomenon has become known as *hypothetical bias.* Meta-analyses by List and Gallet (2001) and Murphy, Allen, Stevens, and Weatherheard (2005) have examined hypothetical bias quantitatively by looking at nearly 30 different lab and field studies that contain both actual and hypothetical estimates of a good's value. List and Gallet find for their sample that hypothetical values exceed actual values on average by a factor of three, while Murphy et al. find the average to be skewed by a few outliers and therefore present a median bias factor of 1.35. More qualitatively, Harrison and Rutström (2008) report that 34 of the 39 studies they surveyed showed upward bias in the hypothetical values. This robust evidence on the existence of hypothetical bias in homegrown value experiments lends support to the notion of criterion *in*validity. The nonvalidity conclusion is also supported by field experiments that include real and hypothetical elicitations for private goods (for example, List 2001; Blumenschein, Blomquist, Johannesson, Horn, and Freeman 2008).

One difficulty in interpreting this set of findings is that not all the studies used in these assessments satisfy the incentive compatibility and consequentiality requirements identified by Carson and Groves (2007) as the necessary conditions for stated responses to match the actual values. For example, Vossler and Evans (2009) find that hypothetical bias disappears from their homegrown value lab experiments when the stated preference elicitation method makes participants feel that their answers are more consequential. Likewise, Landry and List (2007) find that hypothetical bias disappears from their field experiments when respondents are provided with a script emphasizing the consequentiality of the results before answering the value elicitation question. These results jibe well with nonexperimental evidence suggesting that surveys including explicit discussions on how the results might influence policy produce different estimates than those that do not (as in Herriges, Kling, Liu, and Tobias 2010).

Nonetheless, the persistent divergence identified in homegrown value experiments has spawned a large literature dedicated to understanding its causes and finding ways to mitigate its effects. This literature is important for our assessment in that if research can discover a means of eliminating hypothetical bias or predicting its magnitude, the criterion validity of stated preference methods may ultimately be established. For example, one approach is the "cheap talk" method in which participants are explicitly warned of the tendency among people to inflate hypothetically reported values (for example, Cummings and Taylor 1999; List 2001). Over 30 lab and field experiments find that while "cheap talk" can be moderately effective in some circumstances, its net impact varies with the characteristics of participants and the commodity, and the type of script used. The main other alternative, which seems to show more promise, is to calibrate the answers in some way after they have been collected. In one version of this technique, respondents are asked to rate the confidence they have in their answers after completing the elicitation task, which is usually a response to a posted price. Qualitative ranks (for example, "very certain,"

"certain," "uncertain," and do on) as well as multipoint certainty scales have been used, and in most experiments the distribution of hypothetically obtained values can be made to match the distribution of actual values when the uncertain "yes" responses are recoded to "no" responses. Thus, the evidence suggests that one source of hypothetical bias may be in the form of yea-saying by uncertain respondents. Morrison and Brown (2009) provide a summary and reference list of studies related to both the cheap talk, and certainty scale follow-up, methods. Newer vehicles continue to be proposed for minimizing hypothetical bias (Jacquemet, Joule, Luchini, and Shogren forthcoming; Cameron and DeShazo forthcoming; Bateman, Burgess, Hutchison, and Matthews 2008).

A final piece of evidence regarding criterion validity comes from stated preference studies that were conducted in conjunction with actual binding, local referenda. Of these studies, Johnston (2006) is the purest test of criterion validity (and the role of consequentiality) in that the stated preference exercise was executed prior to a local binding referendum and was fielded in an advisory role as input into deciding whether village in Rhode Island should proceed with the installation of a new water system. Vossler and Kerkvliet (2003) also conduct a survey prior to a binding referendum. Their case study is a 1998 vote over a $9.5 million bond measure, funded by higher property taxes, to pay for improvements to a downtown park in Corvallis, Oregon. In both cases, the researchers find that the stated preference predictions match the outcome of the actual election without any need for calibration. An additional study of this type from Vossler, Kerkvliet, Polasky, and Gainutdinova (2003) found that, if undecided respondents were coded as "no" votes, the stated preference responses were statistically consistent with the referenda results.

How should we interpret the weight of evidence on criterion validity? We have seen that hypothetical bias is commonly found in studies where subjects' personal values form the basis of comparison. On the surface, this provides clear evidence of criterion invalidity for contingent valuation studies. However, a number of steps may be possible to reduce this bias. To the extent that the bias is caused by participants not feeling that their responses matter, stated preference surveys and experiments could be run with designs that provide the proper incentives for subjects to respond thoughtfully. Vossler and Poe (2011) take this a step further when they suggest that criterion validity tests that were conducted without adherence to consequentiality requirements should not be considered when assessing the potential for hypothetical bias. They identify four induced value experiments and one homegrown value experiment that they judge to be consistent with the Carson and Groves (2007) requirements, and note that each of these demonstrates criterion validity. If hypothetical bias remains after appropriate consequentiality conditions are met (or it is not possible to achieve consequentiality), a combination of calibration based on the degree of uncertainty and, to a lesser extent "cheap talk" scripts, might be used to manage hypothetical bias in a way that allows stated preference methods to approach criterion validity status more closely. Finally, the evidence from stated preference surveys and binding referenda supports criterion validity, at least in the

case of people making decisions about local public goods. Based on this string of findings, it is difficult to conclude purely in favor of criterion validity, but also difficult to reject it outright.

For the sake of argument, suppose we find the existing evidence to be insufficient to support a conclusion of criterion validity in the pure sense—that is, statistical equivalence between a stated preference estimate and the criterion. We would still be left with the question as to whether stated preference surveys provide useful (albeit imperfect) information for cost–benefit analysis, policy debates, and/or judicial findings. Indeed, statistical equivalence to one estimate of the truth is a strict standard that many economic analyses used for policy—including most revealed preference estimates, we suspect—would have difficulty passing. More importantly, even limited information may be useful in cost–benefit analysis, policy discussions, and litigation. For example, a simple upper or lower bound on estimates of passive use value can sometimes be sufficient to determine whether a project would pass a cost–benefit analysis. In such a case, a point estimate and knowledge of the direction of bias can be adequate for evaluation. Likewise, even when benefit estimates are uncertain and the sign of any bias is unknown, the magnitude of the point estimate relative to cost estimates (which are also likely to be subject to a range of uncertainties) may provide useful input for policymakers and stakeholders.

Convergent Validity: Are Stated and Revealed Preference Estimates the Same?

Convergent validity refers to how well a stated preference estimate correlates with other measures of the same economic value. The most common type of convergent validity tests compare stated preference estimates to those from other techniques, usually based on revealed preferences. Convergent validity tests of this type are not possible for passive use values, but they can be carried out in other instances, such as when the measurement objective concerns a private or quasi-public good. A good example of this is the value of recreation resources, and many studies have used both stated and revealed preference to examine how the environment conveys value through recreation. If the values match, or diverge in expected directions for expected reasons, the estimates are said to be convergent valid. Of course, both estimates may be wrong! Still, if convergence occurs we might have more confidence in both methods, when they are appropriately applied. In terms of evidence, an older meta-analysis from Carson, Flores, Martin, and Wright (1996) supports the notion of convergent validity. Many individual studies have since been done to study convergent validity between specific types of stated and revealed preference data. In some instances, researchers test for the equivalence of econometric parameters, and in others they test for the statistical equality of economic value estimates. While exceptions exist, our sense is that studies that focus on the equivalence of economic values are generally consistent with the findings from Carson, Flores, Martin, and Wright (1996).

In contemporary research, tests of convergent validity *per se* have given way to a more general focus on econometric methods that allow the two types of data to be combined in the same model to exploit their relative strengths. This literature is

surveyed in a book-length treatment by Whitehead, Haab, and Huang (2011). Here, we merely note that the growth of such methods in environmental and nonenvironmental fields is predicated on the implicit acceptance of convergent validity—or at least a common data-generating process—by a wide spectrum of researchers. Two prominent examples include Berry, Levinsohn, and Pakes (2004), who use both actual purchases and stated intentions to estimate the demand for new car purchases, and Small, Winston, and Yan's (2005) use of both stated and revealed preference data to estimate commuters' demand for travel characteristics. Given this, we interpret the weight of evidence on convergent validity to be generally positive.

Construct Validity: Are Stated Preference Estimates Consistent with Theoretical Predictions?

Prior to the experimental revolution and the advent of research using both stated and revealed preference methods, consideration of construct validity—the extent to which predictions from stated preference experiments are consistent with theory—was the main means by which the efficacy of stated preference was assessed. For example, one issue strongly debated in the 1994 JEP symposium by Diamond and Hausman (1994) and Hanemann (1994) concerns "embedding effects"—that is, whether and to what degree willingness to pay for environmental goods should vary with their size. This has become known as the issue of "scope."

Most of the theory used to evaluate stated preference validity was based on price changes involving private goods, as this was the type of good theretofore most studied by economists. This generated testable predictions and assertions that 1) the proportion of people willing to contribute to an environmental good in a stated preference survey should increase when the requested payment amount falls; 2) people should be willing to pay more to have a higher quantity of the good—that is, estimates should exhibit positive response to scope; 3) the income elasticity of willingness to pay should be larger than one, because environmental quality is best viewed as a luxury good; and 4) willingness to pay and willingness to accept for environmental changes should not be substantially different. While the first of these holds true in almost all stated preference studies, the remaining three were often violated for stated preference data—particularly early studies of sensitivity to scope and most studies comparing estimates of willingness to pay and willingness to accept.

These violations were often cited as evidence of construct invalidity. However, additional work in economic theory since the Exxon spill has shown that predictions 2, 3, and 4 are sensitive to two common features of environmental goods: fixed quantities and limited substitutability with other consumption goods. For example, while the marginal willingness to pay curve for a fixed quantity—like a given level of environmental quality—is downward sloping as expected, its relationship to income imbeds several distinct effects. Flores and Carson (1997) show that the income elasticity of willingness to pay for an environmental good depends on three adjustment margins: the implied income elasticity of demand for the environmental good, the substitutability among all the quantity-constrained goods, and

the share of augmented income allocated to market goods. Numerical examples are used to show that an income elasticity of *willingness to pay* that is less than one is in many plausible circumstances consistent with an income elasticity of *demand* for the fixed quantity that is greater than one. In a similar spirit, Amiran and Hagen (2010) show that bounded substitution between market and environmental goods can result in rational behavior failing to exhibit sensitivity to scope, thereby altering prediction 2 for environmental goods.

Recent empirical results on scope effects deserve mention since the early critiques of stated preference methods were based on findings in some studies that estimates of economic value did not go up when the scale of the environmental good was increased. As sensitivity to scope became a litmus test for the construct validity of stated preference estimates, many post-Exxon studies were specifically designed to include "scope tests." Meta-analyses of these studies from Smith and Osborne (1996), Carson (1997), Brouwer, Langford, Bateman, and Turner (1999), and Ojea and Loureiro (2011) show that scope effects are typically present in well-executed studies.

The persistently observed gap between willingness to pay and willingness to accept estimates in stated preference studies also deserves mention. Although Hanemann (1991) and Zhao and Kling (2009) suggest two different theories that can rationalize such a gap without implying construct invalidity from a neoclassical perspective, the size of the difference in many studies appears implausible. Is the divergence due mainly to the hypothetical nature of stated preference surveys? The evidence suggests no. Horowitz and McConnell (2002) reviewed 45 studies and found no difference in the divergence between hypothetical experiments and real experiments. That is, the divergence is not due to the hypothetical nature of stated preference surveys. Although the divergence has been found to be sensitive to the experimental settings (as in Plott and Zeiler 2005) and experience (as in List 2003), the evidence continues to point to alternative preference structures such as the endowment effect. Thus, the divergence does not automatically translate into violations of construct validity, though it may require reconsideration of what theoretical paradigm is used to analyze behavior.

In sum, advances over the last two decades have shown that a combination of neoclassical and behavioral economic theory can give rise to a wider range of predictions that are consistent with the findings of stated preference studies. Of course, the fact that a wider range of outcomes is theoretically consistent does not validate all possible magnitudes of such outcomes. Even with this caveat, a casual browsing of contemporary state-of-the-art stated preference studies suggests that they are almost always consistent with the predictions noted above. For example, the relationship between the posted price and the probability of a "yes" vote is almost universally negative, income effects are robustly positive, and scope criteria are usually met. The anomalous findings that remain—like the divergence between willingness to pay and accept—arise broadly in other forms of microeconomic data and are therefore of little value in considering the construct validity of stated preference methods.

Nonetheless, as new approaches to stated preference elicitation arise, construct validity concerns can reappear and will need careful attention. For example, the mechanism design framework predicts that ordering effects will be present when individuals respond to multiple choice tasks, as is the case with choice experiments. Ordering effects in choice experiments have indeed been confirmed empirically (Day et al. 2012). Thus, a research challenge is to assess how commonly used departures from incentive compatibility compromise predictions from choice experiments.

Content Validity: Is Best Practice Being Followed?

The final type of validity we consider relates to how effectively a stated preference study adheres to the current state of the art. This topic is relevant for our review insomuch as the notion of state of the art has changed dramatically since the immediate post-Exxon days. The two decades since then have seen an explosion of stated preference work. At a minimum, this means the stock of accumulated wisdom—for example, how people react to a particular payment mechanism, how environmental concepts are best communicated in lay language—is orders of magnitude greater than it was. As mentioned above, there are now several how-to books on stated preference methods that provide survey development steps, numerous examples, and advice on avoiding known pitfalls. Given this, genuine surprises in purely applied studies are now rare; the method has matured and become more standardized, and practitioners now have a much better sense of the important design elements of a stated preference survey.

Evidence for this point is apparent when we look at how the challenges identified in the early debates on the method have been researched and findings incorporated in a new understanding of best practice. We provide three specific examples. First, it is now widely accepted that the environmental good needs to be described with a high level of specificity, and the status quo and changed levels of the good precisely defined in a way that lay respondents can understand and place in context. This information is usually presented via a combination of text, photos, graphics, and numbers that has been deliberately developed using focus groups, interviews, and pretests. The increased use of computer-administered surveys has provided additional flexibility for efficiently explaining the environmental good in multiple ways and checking people's comprehension. A result of this emphasis on specificity (and careful communication) is that contemporary studies almost always satisfy sensitivity to scope and other theoretical predictions. A corollary is that a vague or abstract commodity definition—or inadequate evidence of an effective communication strategy—is considered a failure of content validity. Thus, while the NOAA panel early on stressed the importance of specificity (Arrow, Solow, Portney, Leamer, Radner, and Schuman 1993), its evolution into best practice protocols has occurred incrementally through accumulated experience in numerous subsequent applications.

A second area in which best practice has evolved relates to how the constructed market and payment mechanism are defined and interpreted. It is now widely accepted that the constructed market should represent a realistic mechanism for

bringing about the proposed change, meaning that the size of the change arising from the intervention needs to be seen as physically plausible by respondents. Similarly, the payment mechanism needs to be something that respondents find realistic and familiar—both so they will take the exercise seriously, and so they can envision how an actual payment would occur. The attention given to a survey's policy institutions has also led to a consensus among practitioners that estimated values are for the entire package—that is, the environmental change in the context of the described program, rather than the environmental change in a vacuum. Thus, the expectation among current researchers is not that the estimated values should be independent of context. Instead, differences should arise based on the specifics of the program, and validity hinges on the extent to which the differences are consistent with theory and intuition.

The final example of change in best practice relates to ways that researchers attempt to encourage and/or test for the rationality and truthfulness of respondents' contingent behavior. Understanding of what constitutes an incentive-compatible elicitation mechanism has evolved beyond the NOAA panel's recommendation to use a referendum format (Arrow, Solow, Portney, Leamer, Radner, and Schuman 1993). Researchers now know that design elements related to voluntary versus coercive payment, the actual payment vehicle, and commodity provision details can matter. Likewise, framing the survey to be consequential, the presentation of cheap talk scripts, and the use of certainty follow-up questions have, in various combinations, become common practice. In response to advances in theoretical understanding, researchers are also less likely to draw conclusions about construct validity based on narrowly interpreted tests of scope, income effects, and the sensitivity of value estimates to the details of the constructed market. Instead the criteria used to evaluate construct validity are case-specific and start with questions about the extent to which the specific predictions fit with the specific context.

Content validity is a different concept than the other types of validity in that we cannot summarize general evidence to conclude that stated preference methods are valid or invalid in this dimension. Nonetheless there does seem to be a more complete (and a more nuanced) consensus now than two decades ago on the characteristics of a state-of-the-art study. While this does not say much about the general accuracy of stated preference methods, it does illustrate that the early areas of concern have been well researched and best practice has evolved based on the findings. It is up to the reader to decide if this large volume of work implies we are left with an approach that inspires confidence.

Conclusion

Stated preference techniques are in a much different place in the aftermath of the BP accident in 2010 than they were after the Exxon oil spill in 1989. The past two decades have seen the coming of age of experimental economics, new theoretical developments, accumulating insights from behavioral economics, and a general

Table 2

Summary of Authors' Assessment

Validity concept	Assessment	Comments
Criterion	Some Yes, Some No	• Persistence of hypothetical bias in homegrown value experiments implies invalidity. • Emerging consequentiality paradigm suggests potential for validity. • Difficult to conclude purely in favor of validity, but also difficult to outright reject validity.
Convergent	Likely Yes	• Formal tests often accept revealed and stated preference equality. Even when statistically different estimates occur, they appear to illustrate common economic phenomena. • Practice has migrated towards using revealed and stated preference data as complements rather than substitutes.
Construct	Strongly Yes	• Further development of standard theory suggests a wider range of outcomes can still be considered neoclassically rational. • New behavioral theories suggest alternative paradigms might be needed to assess validity. • Definitive construct validity tests are now more difficult to formulate.
Content	Variable	• Content validity is a study-specific concept, but the stock of accumulated wisdom suggests adherence to best practice is now a stronger validity concept than in the past.

maturing of the nonmarket valuation literature. We now have more tools with which to judge the accuracy of stated preference estimates and an emerging consensus on the criteria we should use to do so. Many of the questions that arose in the post-Exxon days have been acknowledged and investigated. Those who formulated their beliefs about contingent valuation two decades ago, whether positive or negative, should update their beliefs based on the research agenda that has unfolded. To help readers with this we have prepared Table 2 as our own summary of possible answers to the question of whether the stated preference method can provide valid and accurate estimates of underlying economics values. While the summary constitutes our personal judgments, we have tried to convey the range of views that different people might take following an objective reading of the literature.

Before concluding, we note four areas of research that seem especially critical for continuing the research agenda related to the validity of stated preference methods. First, validity tests that explicitly include the consequentiality dimension in their design are relatively young, and more research is needed to determine if the initial evidence holds up to further scrutiny. Second, much could be learned by subjecting other methods of valuation to the same level of scrutiny that stated preference methods have received. For example, what methods should be used to assess the validity of estimates from hedonic housing or wage studies? How well do recreation demand model estimates stand up to comparisons with actual transactions?

Answers to these questions would enhance their usefulness for cost–benefit analysis generally and improve our ability to assess the relative performance of stated preference methods. Third, a lot of work remains to be done on understanding how the common use of incentive-incompatible designs in choice experiments affects the validity of this recently popular approach. Finally, there remains substantial uncertainty as to how researchers should execute and interpret validity tests using alternative behavioral paradigms. If the same behavioral anomalies appear in both stated and actual behavior, should a valid survey mimic real world choices or seek to elicit "true" preferences—neoclassical or otherwise—for use in welfare analysis?

Despite these and other questions, our sense is that the last 20 years of research have shown that some carefully constructed number based on stated preference analysis is now likely to be more useful than no number in most instances for both cost–benefit analysis and damage assessment. Of course this is a weaker conclusion than validity, and it is not to say that all studies are equally reliable or that inference from reliable studies will always be appropriately applied. But it is illustrative of the remarkable progress that stated preference researchers have made, and it serves as a model for the evaluation of other policy-critical techniques.

References

Alvarez, Sergio, Sherry L. Larkin, John C. Whitehead, and Timothy C. Haab. 2012. "Substitution, Damages, and Compensation for Anglers due to Oil Spills: The Case of the Deepwater Horizon." http://purl.umn.edu/124779.

Amiran, Edoh Y., and Daniel A. Hagen. 2010. "The Scope Trials: Variation in Sensitivity to Scope and WTP with Directionally Bounded Utility Functions." *Journal of Environmental Economics and Management* 59(3): 293–301.

Arrow, Kenneth, Robert Solow, Paul R. Portney, Edward E. Leamer, Roy Radner, and Howard Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." January 11. *Federal Register* 58: 4601–14.

Bateman, Ian J., Diane Burgess, W. George Hutchison, and David I. Matthews. 2008. "Contrasting NOAA Guidelines with Learning Design Contingent Valuation (LDCV): Preference Learning versus Coherent Arbitrariness." *Journal of Environmental Economics and Management* 55(2): 127–41.

Bateman, Ian J., Richard T. Carson, Brett Day, Michael Hanemann, Nick Hanley, Tannis Hett,

Michael Jones Lee, Graham Loomes, Susana Mourato, Ece Ozdemiroglu, David W. Pearce, Robert Sugden, and John Swanson. 2002. *Economic Valuation with Stated Preference Techniques: A Manual.* Cheltenham, UK: Edward Elgar.

Bateman, Ian J., and Kenneth G. Willis, eds. 1995. *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Methods in the US, EC, and Developing Countries.* Oxford, UK: Oxford University Press.

Bennett, Jeff, ed. 2011. *The International Handbook on Non-Market Environmental Valuation.* Cheltenham, UK: Edward Elgar.

Berry, Steven, James Levinsohn, and Ariel Pakes. 2004. "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market." *Journal of Political Economy* 112(1): 68–105.

Blumenschein, Karen, Glenn C. Blomquist, Magnus Johannesson, Nancy Horn, and Patricia Freeman. 2008. "Eliciting Willingness to Pay without Bias: Evidence from a Field Experiment." *Economic Journal* 118(1): 114–37.

Bowles, Samuel. 1998. "Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions." *Journal of Economic Literature* 36(1): 75–111.

Brouwer, Roy, Ian H. Langford, Ian J. Bateman, and R. Kerry Turner. 1999. "A Meta-analysis of Wetland Contingent Valuation Studies." *Regional Environmental Change* 1(1): 47–57.

Cameron, Trudy Ann, and J. R. DeShazo. Forthcoming. "Demand for Health Risk Reductions." *Journal of Environmental Economics and Management.*

Carson, Richard T. 1997. "Contingent Valuation Surveys and Tests of Insensitivity to Scope." Chap. 5 in *Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods,* edited by Raymond J. Kopp, Walter Pommerhene, and Norbert Schwartz. Boston: Kluwer.

Carson, Richard T. 2011. *Contingent Valuation: A Comprehensive Bibliography and History.* Cheltenham, UK: Edward Elgar.

Carson, Richard T., Nicholas E. Flores, Kerry M. Martin, and Jennifer L. Wright. 1996. "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-public Goods." *Land Economics* 72(1): 80–99.

Carson, Richard T., Nicholas E. Flores, and Norman F. Meade. 2001. "Contingent Valuation: Controversies and Evidence." *Environmental and Resource Economics* 19(2): 173–210.

Carson, Richard T., and Theodore Groves. 2007. "Incentive and Informational Properties of Preference Questions." *Environmental and Resource Economics* 37(1): 181–210.

Carson, Richard T., and W. Michael Hanemann. 2005. "Contingent Valuation." Chap. 17 in the *Handbook of Environmental Economics,* Vol. 2, edited by Karl-Goran Maler and Jeffrey Vincent. North-Holland.

Carson, Richard, Robert Mitchell, W. Michael Hanemann, Raymond J. Kopp, Stanley Presser, and Paul Ruud. 2003. "Contingent Valuation and Lost Passive Use: Damages from the *Exxon Valdez* Oil Spill." *Environmental and Resource Economics* 25(3): 257–83.

Champ, Patricia A., Richard C. Bishop, Thomas C. Brown, and Daniel W. McCollum. 1997. "Using Donation Mechanisms to Value Nonuse Benefits from Public Goods." *Journal of Environmental Economics and Management* 33(2): 151–62.

Champ, Patricia A., Kevin J. Boyle, and Thomas C. Brown, eds. 2003. *A Primer on Nonmarket Valuation.* Dordrecht, the Netherlands: Kluwer Academic Publishers.

Cherry, Todd L., Thomas D. Crocker, and Jason F. Shogren. 2003. "Rationality Spillovers." *Journal of Environmental Economics and Management* 45(1): 63–84.

Ciriacy-Wantrup, S. V. 1947. "Capital Returns from Soil-Conservation Practices." *Journal of Farm Economics* 29(4, Part 2): 1188–90.

Cummings, Ronald G., David S. Brookshire, and William D. Schulze. 1986. *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method.* Totowa, NJ: Rowman and Allanheld.

Cummings, Ronald G., Glenn W. Harrison, and E. Elisabet Rutström. 1995. "Homegrown Values and Hypothetical Surveys: Is the Dichotomous-Choice Approach Incentive-Compatible?" *American Economic Review* 85(1): 260–66.

Cummings, Ronald G., and Laura O. Taylor. 1999. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *American Economic Review* 89(3): 649–65.

Davis, Robert. 1963. *The Value of Outdoor Recreation: An Economic Study of the Maine Woods.* Doctoral dissertation in economics, Harvard University.

Day, Brett, Ian J. Bateman, Richard T. Carson, Diane Dupont, Jordan J. Louviere, Sanae Morimoto, Riccardo Scarpa, and Paul Wang. 2012. "Ordering Effects and Choice Set Awareness in Repeat-Response Stated Preference Studies." *Journal of Environmental Economics and Management* 63(1): 73–91.

de Bekker-Grob, Esther W., Mandy Ryan, and Karen Gerard. 2012. "Discrete Choice Experiments in Health Economics: A Review of the Literature." *Health Economics* 21(2): 145–72.

Diamond, Peter A., and Jerry A. Hausman. 1994. "Contingent Valuation: Is Some Number

Better than No Number?" *Journal of Economic Perspectives* 8(4): 45–64.

Exxon Valdez Oil Spill Trustee Council. N.d. "Oil Spill Facts." Webpage. http://www.evostc.state .ak.us/facts/index.cfm.

Exxon Valdez Oil Spill Trustee Council. N.d. "Settlement." Webpage about the U.S. District Court Settlement on October 9, 1991 between the State of Alaska, the U.S. government, and Exxon. http://www.evostc.state.ak.us/facts/settlement .cfm.

Flores, Nicholas E., and Richard T. Carson. 1997. "The Relationship between the Income Elasticities of Demand and Willingness to Pay." *Journal of Environmental Economics and Management* 33(2): 287–95.

Guarino, Mark. 2012. "BP Oil Spill Settlement: Justice for 100,000 Gulf Coast Victims?" *Christian Science Monitor,* April 18. http://www .csmonitor.com/USA/Justice/2012/0418 /BP-oil-spill-settlement-Justice-for-100-000-Gulf -Coast-victims.

Haab, Timothy C., and Kenneth E. McConnell. 2002. *Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation.* Cheltenham, UK: Edward Elgar.

Hanemann, W. Michael. 1991. "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" *American Economic Review* 81(3): 635–47.

Hanemann, W. Michael. 1994. "Valuing the Environment through Contingent Valuation." *Journal of Economic Perspectives* 8(4): 19–43.

Harrison, Glenn W., and E. Elisabet Rutström. 2008. "Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods." Chapter 81 in *Handbook of Experimental Economics Results,* Vol 1, Part 5, edited by Charles R. Plott and Vernon L. Smith. New York: North-Holland.

Hausman, Jerry A., Gregory K. Leonard, and Daniel L. McFadden. 1995. "A Utility-Consistent, Combined Discrete Choice and Count Data Model Assessing Recreational Use Losses due to Natural Resource Damage." *Journal of Public Economics* 56(1): 1–30.

Herriges, Joseph A., Catherine L. Kling, Chih-Chen Liu, and Justin Tobias. 2010. "What Are the Consequences of Consequentiality?" *Journal of Environmental Economics and Management* 59(1): 67–81.

Horowitz, John K., and Kenneth E. McConnell. 2002. "A Review of WTA/WTP Studies." *Journal of Environmental Economics and Management* 44(3): 426–47.

Iyengar, Sheena S., and Kamenica, Emir. 2007. "Choice Overload and Simplicity Seeking." http:// www.cbdr.cmu.edu/seminar/Emir2.pdf.

Jacquemet, Nicolas, Robert-Vincent Joule, Stephane Luchini, and Jason F. Shogren. Forthcoming. "Preference Elicitation under Oath." *Journal of Environmental Economics and Management.* Available online at http://www.sciencedirect .com/science/article/pii/S0095069612000587.

Johnston, Robert J. 2006. "Is Hypothetical Bias Universal? Validating Contingent Valuation Responses Using a Binding Public Referendum." *Journal of Environmental Economics and Management* 52(1): 469–81.

Kanninen, Barbara J., ed. 2007. *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice.* Springer.

Krutilla, John. 1967. "Conservation Reconsidered." *American Economic Review* 57(4): 777–86.

Landry, Craig F., and John A. List. 2007. "Using *ex ante* Approaches to Obtain Credible Signals for Value in Contingent Markets: Evidence from the Field." *American Journal of Agricultural Economics* 89(2): 420–29.

Larkin, Sherry. 2012. Personal communication with the authors.

Li, Hui, Robert P. Berrens, Alok K. Bohara, Hank C. Jenkins-Smith, Carol L. Silva, and David L. Weimer. 2005. "Testing for Budget Constraint Effects in a National Advisory Referendum Survey on the Kyoto Protocol." *Journal of Agricultural and Resource Economics* 30(2): 350–66.

List, John A. 2001. "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards." *American Economic Review* 91(5): 1498–1507.

List, John A. 2003. "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics* 118(1): 41–71.

List, John A., and Craig A. Gallet. 2001. "What Experimental Protocol Influence Disparities between Actual and Hypothetical Stated Values?" *Environmental and Resource Economics* 20(3): 241–54.

Loomis, John, Paula Kent, Liz Strange, Kurt Fausch, and Alan Covich. 2000. "Measuring the Total Economic Value of Restoring Ecosystem Services in an Impaired River Basin: Results from a Contingent Valuation Survey." *Ecological Economics* 33(1): 103–117.

Louiviere, Jordan J., David A. Hensher, and Joffre D. Swait. 2000. *Stated Choice Methods: Analysis and Applications.* Cambridge University Press.

Mitchell, Robert Cameron, and Richard T. Carson. 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Methods.* Washington, DC: Resources for the Future.

Morrison, Mark, and Thomas C. Brown. 2009. "Testing the Effectiveness of Certainty Scales,

Cheap Talk, and Dissonance-Minimization in Reducing Hypothetical Bias in Contingent Valuation Studies." *Environmental and Resource Economics* 44(3): 307–326.

Murphy, James J., P. Geoffrey Allen, Thomas H. Stevens, and Darryl Weatherheard. 2005. "A Meta-analysis of Hypothetical Bias in Stated Preference Valuation." *Environmental and Resource Economics* 30(3): 313–25.

Murphy, James J., Thomas H. Stevens, and Lava Yadav. 2010 "A Comparison of Induced Value and Home-grown Value Experiments to Test for Hypothetical Bias in Contingent Valuation." *Environmental and Resource Economics* 47(1): 111–23.

Ojea, Elena, and Maria L. Loureiro. 2011. "Identifying the Scope Effect on a Meta-analysis of Biodiversity Valuation Studies." *Resource and Energy Economics* 33(3): 706–24.

Phaneuf, Daniel J., and Till Requate. Forthcoming. *A Course in Environmental Economics: Theory, Policy, and Practice.* Cambridge University Press.

Plott, Charles R., and Kathryn Zeiler. 2005. "The Willingness to Pay–Willingness to Accept Gap, the 'Endowment Effect,' Subject Misconceptions, and Experimental Procedures for Eliciting Valuations." *American Economic Review* 95(3): 530–45.

Portney, Paul R. 1994. "The Contingent Valuation Debate: Why Economists Should Care." *Journal of Economic Perspectives* 8(4): 3–17.

Randall, Alan, Berry Ives, and Clyde Eastman. 1974. "Bidding Games for Valuation of Aesthetic Environmental Improvements." *Journal of Environmental Economics and Management* 1(2): 132–49.

Shogren, Jason F., and Laura O. Taylor. 2008. "On Behavioral-Environmental Economics." *Review of Environmental Economics and Policy* 2(1): 26–44.

Slembeck, Tilman, and Jean-Robert Tyran. 2004. "Do Institutions Promote Rationality?: An Experimental Study of the Three-Door Anomaly." *Journal of Economic Behavior and Organization* 54(3): 337–50.

Small, Kenneth A., Clifford Winston, and Jia Yan. 2005. "Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability." *Econometrica* 73(4):1367–82.

Smith, V. Kerry, and Laura Osborne. 1996. "Do Contingent Valuation Estimates Pass a Scope Test? A Meta-analysis." *Journal of Environmental Economics and Management* 31(3): 287–301.

Schokkaert, Erik. 2006. "The Empirical Analysis of Transfer Motives." Chap. 2 in *Handbook of the Economics of Giving, Altruism and Reciprocity*, Vol. 1, edited by Serge-Christophe Kolm and Jean Mercier Ythier. North-Holland.

Taylor, Laura O., Michael McKee, Susan K. Laury, and Ronald G. Cummings. 2001. "Induced-Value Tests of the Referendum Voting Mechanism." *Economic Letters* 71(1): 61–65.

Thaler, Richard H. 1990. "Anomalies: Saving, Fungibility, and Mental Accounts." *Journal of Economic Perspectives* 4(1): 193–205.

Vossler, Christian A., and Mary F. Evans. 2009. "Bridging the Gap between the Field and the Lab: Environmental Goods, Policy Maker Input, and Consequentiality." *Journal of Environmental Economics and Management* 58(3): 338–45.

Vossler, Christian A., and Joe Kerkvliet. 2003. "A Criterion Validity Test of the Contingent Valuation Method: Comparing Hypothetical and Actual Voting Behavior for a Public Referendum." *Journal of Environmental Economics and Management* 45(3): 631–49.

Vossler, Christian A., Joe Kerkvliet, Stephen Polasky, and Olesy Gainutdinova. 2003. "Externally Validating Contingent Valuation: An Open-Space Survey and Referendum in Corvallis, Oregon." *Journal of Economic Behavior and Organization* 51(2): 261–77.

Vossler, Christian A., and Michael McKee. 2006. "Induced-Value Tests of Contingent Valuation Elicitation Methods." *Environmental and Resource Economics* 35(2): 137–68.

Vossler, Christian A., and Gregory L. Poe. 2011. "Consequentiality and Contingent Values: An Emerging Paradigm." Chapter 7 in *The International Handbook on Non-Market Environmental Valuation*, edited by Jeff Bennett. Northhampton, MA: Edward Elgar.

Whitehead, John C., Glenn C. Blomquist, Thomas J. Hoban, and William B. Crawford. 1995. "Assessing the Validity and Reliability of Contingent Values: A Comparison of On-site Users, Off-site Users, and Non-users." *Journal of Environmental Economics and Management* 29(2): 238–51.

Whitehead, John, Tim Haab, and Ju-Chin Huang, eds. 2011. *Preference Data for Environmental Valuation: Combining Revealed and Stated Approaches.* Routledge Explorations in Environmental Economics series. Routledge.

Zhao, Jinhua, and Catherine L. Kling. 2009. "Welfare Measures when Agents Can Learn: A Unifying Theory." *Economic Journal* 119(540): 1560–85.

Contingent Valuation: A Practical Alternative when Prices Aren't Available

Richard T. Carson

Aperson may be willing to make an economic tradeoff to assure that a wilderness area or scenic resource is protected even if neither that person nor (perhaps) anyone else will actually visit this area. This tradeoff is commonly labeled "passive use value," although it is also known by other names including "existence value" and "stewardship value." As Krutilla (1967) explained in his classic *American Economic Review* piece, "Conservation Reconsidered," passive use is not generally revealed by choices in the marketplace, unlike many local public goods which are either capitalized into property values or which require the consumption of complementary private goods to enjoy. Passive use represents the quintessential pure public good in that exclusion is not possible, nor even desirable, because enjoyment is nonrivalrous. The concept of passive use has played an increasingly important role in economic thinking concerning the value of public goods, and particularly, those involving environmental and natural resource amenities where passive and direct use values are often thought to coexist (Freeman 2003). In Carson, Flores, and Mitchell (1999), my coauthors and I provide a detailed overview of passive use value.

Without market information, other strategies must be considered to develop measures of economic tradeoffs that involve passive use value. For example, passive values can be captured through a single-issue referendum, but popular votes on ballot propositions that relate to these types of concerns are nonexistent at the national level and infrequent at the state or local level. However, more than a half-century ago, early studies of public goods like Bowen (1943) and Ciriacy-Wantrup (1947) drew the implication that, when the ballot box is not available, demand

■ *Richard T. Carson is Professor of Economics, University of California, San Diego, La Jolla, California. His email address is rcarson@ucsd.edu.*

for public goods might be estimated through an appropriately structured survey of a representative sample of the public—in effect, what we now call a contingent valuation survey. Bowen went so far as to argue: "The polling of a "scientifically" selected sample might produce more accurate results than general voting, unless arrangements were made to insure that every person would actually vote."

Contingent valuation studies ask questions that help to reveal the monetary tradeoff each person would make concerning the value of goods or services. In Carson and Louviere (2011), my coauthor and I provide a common nomenclature for such "stated preference" questions. Such surveys are a practical alternative approach for eliciting the value of public goods, including those with passive use considerations. Thousands of contingent valuation studies have been done in over 130 countries looking at cultural, environmental, health, transportation, and other issues (Carson 2011). Almost 60 percent of the estimates in the very large Environmental Values Reference Inventory (EVRI) database maintained by Environment Canada in conjunction with the U.S. Environmental Protection Agency and the environmental agencies of several other countries come from contingent valuation (at https://www.evri.ca). The U.S. Environmental Protection Agency's (1994) estimates of the benefits of the U.S. Clean Water Act—which largely comprises recreation and passive use—is derived using contingent valuation, as are the benefits of individual regulations targeted at specific industries or water bodies (Griffiths et al. 2012). Results from contingent valuation studies are used for many purposes in benefit–cost studies: recent examples include the willingness to pay of Pennsylvania households for additional incarceration versus a rehabilitation program for serious juvenile offenders (Nagin, Piquero, Scott, and Steinberg 2006); the willingness of Lexington, Kentucky, residents to pay higher taxes to help support the construction of a new baseball stadium and basketball arena (Johnson and Whitehead 2007); the value of developing vaccine policies in Africa (Jeuland, Lucas, Clemens, and Wittington 2009); estimating the hourly value of informal care givers in the Netherlands (de Meijer, Brouwer, Koopman-schap, van den Berg, and van Exel 2010); looking at willingness to incur higher water tariffs for less river pollution in Fuzhou, China (Jiang, Jin, and Lin 2011); and the willingness of the U.S. public to pay for climate change measures (Aldy, Kotchen, and Leiserowitz 2012).

This essay begins by discussing the events set in motion by the *Exxon Valdez* oil spill of March 1989, focusing on why it is important to measure monetary tradeoffs for goods where passive use considerations loom large. Although discussions often seem to put much of their emphasis on whether contingent valuation is sufficiently reliable for use in assessing natural resource damages in lawsuits, it is important to remember that most estimates from contingent valuation studies are used in benefit–cost assessments, not natural resource damage assessments. Those working on benefit–cost analysis have long recognized that goods and impacts that cannot be quantified are valued, implicitly, by giving them a limitless value when government regulations preclude certain activities, or giving them a value of zero by leaving certain consequences out of the analysis. Contingent valuation offers a practical alternative for reducing the use of either of these extreme choices. I put

forward an affirmative case for contingent valuation and address a number of the concerns that have arisen.

Events Set in Motion for Contingent Valuation by the Exxon Valdez Spill

Soon after the *Exxon Valdez* spill in March 1989, the state of Alaska funded a contingent valuation study, contained in Carson, Mitchell, Hanemann, Kopp, Presser, and Ruud (1992), which estimated the American public's willingness to pay to avoid an oil spill similar to the *Exxon Valdez* at about $3 billion. The results of the study were shared with Exxon and a settlement for approximately $3 billion was reached, thus avoiding a long court case.[1] Our Carson et al. (1992) $3 billion estimate based on passive use dwarfed the Hausman, Leonard, and McFadden (1995) $4 million dollar estimate of the direct economic losses from lost recreation days in Prince William Sound, illustrating the importance of compensating the public for lost passive use.

In the aftermath of the *Exxon Valdez* oil spill, the U.S. Coast Guard put into place a version of the comprehensive plan for preventing oil spills put forward in the Carson et al. (1992) study. It was based to a large degree on the original risk assessment for shipping oil out of Alaska that had predicted one major spill every ten years from an accident eerily similar to that of the *Exxon Valdez* in the absence of risk-reducing measures (Moore 1994; Carson, Mitchell, Hanemann, Kopp, Presser, and Ruud 2003). In the years prior to the accident, some of the main safety requirements had been abandoned because they seemed "expensive" and unnecessary. One of these was that tankers have "escort tugs." Soon after key elements of the plan in our Carson et al. (1992) study were put into place, another supertanker lost power in the Straits of Valdez and drifted toward a reef near the one hit by the *Exxon Valdez*. One of the plan's new escort tugs pushed the supertanker away from the reef while the other tug shot it a towline. Since then, escort tugs have had to take control of a tanker in Prince William Sound three other times, with the latest being ExxonMobil's *SeaRiver Kodiak* in 2010.

Moreover, recognizing the potentially large passive use costs from oil-related activities led to other changes. The U.S. Oil Pollution Act enacted in 1990 required that tankers held by shell companies without large financial assets carry a $1 billion dollar insurance policy and required those shipping oil to develop comprehensive plans to respond to potential oil spills. These actions significantly reduced the

---

[1] By law, the government trustees must spend any money received for harm to its resources on restoration or acquisition of like resources. Exxon spent approximately $2 billion on response and restoration and $1 billion on natural resource damages, which the government used acquiring like resources. If the case had gone to trial, major contested issues would have included whether Admiralty law limited Exxon's liability; how much of Exxon's expenditures on response and restoration were on response, which did not count toward Exxon's liability, and on restoration, which did; and whether contingent valuation could be used to establish the loss to the public from the spill. Note that losses to commercial fishing are the subject of private, not government, legal claims.

frequency and severity of oil spills from tankers in the United States relative to the rest of the world (Chapple 2000). In contrast, Congress granted offshore wells like the BP *Deepwater Horizon* well a liability limit of $75 million dollars, although unlimited liability applies in the case of gross negligence, willful misconduct, or violating a federal regulation directly related to the spill. (In the aftermath of the March 2010 oil spill, BP waived the offshore well liability cap.)

In the aftermath of the *Exxon Valdez* settlement, Exxon put on a conference where research it had sponsored in preparation for the case was presented (see the volume edited by Hausman, 1993). Its overall conclusion was that contingent valuation was unreliable. In response, the U.S. Department of Commerce assembled a blue-ribbon panel of experts chaired by Kenneth Arrow and Robert Solow to examine contingent valuation. Their report (Arrow, Solow, Portney, Leamer, Radner, and Schuman 1993), known as the NOAA Panel Report, was cautiously supportive, finding that "well conducted CVM [contingent valuation method] studies can produce estimates reliable enough to be the starting point of a judicial process of damage assessment, including lost passive values." The panel also set forward an influential set of guidelines for conducting contingent valuation studies. The two conflicting views encouraged a large amount of theoretical, econometric, experimental, and empirical research on contingent valuation.

Economists, Survey Data, and Contingent Valuation

Economists are naturally skeptical of data generated from responses to survey questions—and they should be! Many surveys, including contingent valuation surveys, are inadequate. Whittington (2002), one of the pioneers of contingent valuation studies in developing countries, has lamented the tendency to implement quick and cheap studies that are likely to yield flawed results. He notes that "we are still a long way from the point where it is possible to do high-quality CV [contingent valuation] surveys with minimal effort or expense." This situation is no different from many other areas of economics that are heavily dependent on survey-based data—income, consumption, education, employment, health status, and so on—but it is sometimes less obvious because economists are often not actively involved in how their data is collected and often have no formal training in survey research.

A good contingent valuation survey is a very different process than the mental image some readers may have of a researcher walking up to people in a shopping mall and asking how much they would pay to save a sea otter. For an example of a real-world contingent valuation survey, interested readers might start at the bottom of my web-page at http://www.econ.ucsd.edu/~rcarson, where they can download and examine the survey instrument for the *Exxon Valdez* study. It is 39 pages long, plus 14 pages of maps and photos, and 10 pages of show cards and figures.

Overall, a well-designed contingent valuation survey must convey to respondents that the government is considering implementing a policy and that their responses to the questions in the survey will be used to help inform that decision. The survey describes the problem that is the focus of the survey and the plan that the government

is considering to address that problem. For complex policies, and particularly, those issues without a lot of previous survey work on closely related policies, the process of developing a survey can be lengthy: it frequently involves focus groups, cognitive interviews, pretests, and pilot studies. Details matter: the survey should be designed so that the plan is seen as an effective response to the problem. This must be done in a way that respondents without a high school degree can understand. The presentation typically involves graphics intended to help people understand the problem and the government's plan. Survey respondents need to understand that if the plan is implemented, it would be paid for using a coercive payment mechanism, typically some type of tax or utility bill; each respondent must be convinced that a mechanism exists that would ensure they would pay in that case. Because it is impossible to get all members of the public to accept all details of the scenario, it is standard practice to ask a sequence of "debriefing" questions to help gauge the likely impact of scenario rejection.

Much of the usefulness of doing a contingent valuation study has to do with pushing scientists and engineers to summarize what the project would do in terms that the public cares about. Further, the process of developing a contingent valuation survey often encourages earlier involvement by policymakers in thinking more critically about a project's benefits and costs and in considering options with lower costs or greater benefits to the public.

As long as respondents believe that there is a positive probability that the government will take the results of the contingent valuation survey into account, they should use the opportunity to influence the government's decision. In Carson and Groves (2007), my coauthor and I demonstrate that the response to a properly formulated binary discrete choice question represents "consequential" economic behavior; and, that the incentive properties of such survey questions with respect to economic behavior are identical to those of a binding ballot proposition.[2] In this sense, responses to a good contingent valuation study can reasonably be treated as *revealed economic behavior,* akin to that obtained in a vote of a representative population on a ballot proposition.

Neoclassical Economic Theory and Contingent Valuation Results

Predictions from simple versions of neoclassical economic theory can sometimes differ from outcomes found by contingent valuation surveys. Of course, predictions from simple versions of neoclassical theory can also differ quite a bit from observed real-world behavior, as the literature on "behavioral economics" has pointed out (DellaVigna 2009). Contingent valuation surveys are designed so that

---

[2] In Carson and Groves (2007), we show that the auxiliary conditions needed for truthful preference revelation to be a dominant strategy are that people can be compelled to comply with the payment provision of the scenario irrespective of the outcome and that the scenario offers a take-it-or-leave it choice that does not influence future offers. These conditions are the same for binding votes and advisory surveys.

the findings will reflect actual behavior, so it should be no surprise that the same behavioral influences on what people do in markets show up in surveys. Indeed, some of the best-known insights of behavioral economics were first demonstrated in contingent valuation surveys.

In their overview paper in this issue, Kling, Phaneuf, and Zhao detail the ways researchers have addressed objections that the findings of contingent valuation surveys appear incongruous with simple versions of neoclassical economic theory. My summary here can be limited to some key examples. In Carson, Flores, and Meade (2001) and Carson and Hanemann (2005), my colleagues and I offer a more detailed discussion of these issues.

One of the most persistent of the claims that contingent valuation surveys are unreliable points to a discrepancy between willingness to pay and minimum willingness to accept compensation for the same nonmarket good. This finding should actually not be a surprise, either in terms of neoclassical economic theory or in terms of behavioral economics. The predicted properties of welfare measures are often quite different for 1) cases where everyone will experience the same level or quantity of the public good, and 2) cases involving price changes where consumers can determine the amount of the good they wish to consume. Hanemann (1991) shows willingness to pay and willingness to accept for a pure public good are likely to be quite far apart, which stands in stark contrast to Willig's (1976) well-known result that willingness to pay and willingness to accept for *a price change* should typically be close together. Hicks (1943) correctly saw that welfare measurements involving rationed goods (of which pure public goods are a special case), so-called "surplus" measurements, are fundamentally different from the "variation" measures which typically involve price changes. The underlying reason is that in the price-change case, the magnitude of the difference between the two welfare measures is governed by an income elasticity, thought to be of reasonable magnitude; meanwhile for a quantity (or quality) change, the difference is governed by the ratio of this income parameter to a Hicksian composite substitution parameter between the good and marketed goods, often thought to be small in magnitude. Kling, Phaneuf, and Zhao point out that enriching the basic neoclassical framework by adding dynamic considerations also tends to drive willingness to pay and willingness to accept measures apart. From a behavioral economics view, the divergence between willingness to pay and willingness to accept is a core prediction of Kahneman and Tversky's (1979) prospect theory because whether a choice is framed in terms of gain or loss influences behavior. Again, the upshot here is that willingness to pay and willingness to accept are often not equal, whether in contingent value surveys or in market-based tests; indeed, Horowitz and McConnell's (2002) meta-analysis shows that the differences are similar in both settings.

Two other situations where it is often asserted that contingent valuation studies produce anomalous results involve estimates of income elasticities and sequence effects. The first contends that if contingent valuation studies were valid, then the estimates of the income elasticity of willingness to pay for the environment should be greater than one, because the environment is a luxury good. The main difficulty here (ignoring the plausibility of the luxury-good assumption for these goods and

the likelihood of measurement error in income) is that the income elasticity of will-ingness to pay is a very different statistic than the income elasticity of demand, upon which an economist's usual definition of a luxury good is based. In Flores and Carson (1997), my coauthor and I show the two elasticities are functionally related, but under most plausible assumptions, the income elasticity of willingness to pay should be considerably smaller than the corresponding income elasticity of demand.[3]

The second assertion is that large differences in the measured value of a good depending upon the sequence of other goods that were also valued in the same survey indicate that contingent valuation is unreliable. However, the basic theory of income and substitution effects suggests that sequence effects should occur. In a willingness-to-pay sequence of k goods, keeping utility constant requires that the agent give up money at each order in the sequence as a new good is acquired, whereas in a willingness-to-accept compensation sequence, keeping utility constant requires giving the agent money as goods are sequentially taken away. In Carson, Flores, and Hanemann (1998), my coauthors and I show that these sequence-related differences can easily be large: specifically, willingness-to-pay and willingness-to-accept sequences for imposed quantity changes involve a partially inverted demand system in terms of the Hicksian substitution terms, such that, if sequence order differences are small in price space, they will typically be large in terms of differences in welfare measures. Thus, a good valued first in a willingness-to-pay sequence will tend to be worth more than if it is valued "lower" in a sequence of possible projects.[4] In a way, this result should be no surprise: after all, in setting a political agenda, controlling the order in which projects are considered is thought to be extremely important.

A final criticism of the contingent valuation method is that different prefer-ence elicitation techniques often obtain different estimates of value, which has been taken by some critics as an indication that survey respondents do not have well-defined preferences for nonmarket goods. This finding has troubled contingent valuation researchers, although it is not unique to contingent valuation. Marketing researchers and experimental economists find the same phenomenon. Indeed, cognitive psychologists such as Tversky, Slovak, and Kahneman (1990) have argued that the divergence in economic values implied by framing decisions in terms of a choice rather than matching response is perhaps the fundamental problem with economic theory. A natural economic response to this issue is to study how different elicitation techniques should affect the answers given. Using neoclassical

[3] Specifically, the income elasticity of willingness to pay is equal to the income elasticity of demand times a matrix of Hicksian substitution terms scaled by the ratio of ordinary income to the sum of ordinary income and the implicit income from all public goods. By definition, this ratio is less than one and likely to be substantially less than one.

[4] There is some irony, though, that critics of the use of contingent valuation in natural resource damage assessments point to the substantial declines often seen in willingness to pay for a good as it is valued farther and farther out in a sequence as a reason not to use contingent valuation. Willingness to accept compensation is the theoretically correct welfare measure for harm from an oil spill (Arrow et al. 1993). Because willingness to accept is greater than willingness to pay for the same good valued first in a sequence and because the value of a good in a willingness-to-accept sequence is increasing in terms of sequence order, so willingness to pay for a good appearing first in a sequence is smaller than willingness to accept for the same good appearing in any sequence order.

mechanism design theory, in Carson and Groves (2007), we demonstrate that different elicitation formats have different incentive and information properties. Rational economic agents should be responsive to these properties in such a way that commonly used preference elicitation formats should produce different welfare estimates. This framework offers a comprehensive set of predictions concerning the characteristics of data collected using different preference elicitation methods and differences in welfare estimates obtained using them, and it has fundamentally changed how researchers view stated preference data (Poe and Vossler 2011).

In short, there are often divergences between predictions of simple neoclassical economic theory and actual behavior, as well as between that same theory and responses to contingent valuation surveys. In both cases, the most productive response is often to investigate both the theory and the data more carefully. Usually, a more realistic theoretical representation provides a reasonable guide to observed responses. But people are not perfect; their choices can reflect mistakes, which will be apparent if an analyst observes them under the equivalent of a microscope. At some point, a judgment has to be made as to whether to accept consumer sovereignty in the form of respecting choices involving the tradeoffs people say they are willing to make when they are observed in a context designed to facilitate careful decisions.

Sensitivity to Scope

The Arrow et al. (1993) NOAA Panel Report put forward a set of recommendations that largely followed the procedures used in the Carson et al. (1992) *Exxon Valdez* study, with one major exception. The Panel recommended that contingent valuations studies being done for litigation should pass a "split scope test." This test requires asking two separate subsamples of respondents about two different descriptions of a good, where the amount of the good along some quality or quantity dimension should make it clearly "larger."

The underlying concern here, voiced by Kahneman and Knetsch (1992) and Hausman (1993), was that respondents to contingent valuation surveys may have a certain amount that they are willing to spend on, say, environmental protection issues generally, and so they will tend to respond with this amount in mind regardless of the actual characteristics of the good being valued. An often-cited example is a contingent valuation study in which respondents to a self-administered shopping mall survey appeared willing to pay the same amount to save 2,000, 20,000, or 200,000 birds from being killed by oil (Desvousges, Johnson, Dunford, Boyle, Hudson, and Wilson 1993). However, in this study respondents were also told that the population of birds was very large, with the percent of birds being killed in the three split-sample treatments being similar: (a) "much less than 1% of the population", (b) "less than 1% of the population", and (c) "about 2% of the population. In short, the seeming insensitivity to scope shown by the respondents in this study is exactly what is likely to have been shown by many professional ecologists given this information. Hanemann (2008) replicates this study in the same shopping

mall context with two subsamples, one where 1 percent of the bird population is impacted and the other with 10 percent. He finds a sizeable statistically significant difference in willingness to pay between these two treatments.

Given the concern over the scope issue, new explicit scope tests quickly appeared, and several instances were also identified where past studies done for policy purposes had used split samples with goods differing in scope. In Carson (1997), I reviewed this literature and found 30 split sample tests which rejected the scope insensitivity hypothesis. Most of these involved goods where passive use was thought to be important. Two large state-of-the-art in-person surveys of the general public, one in the United States involving DDT deposits off the coast of Los Angeles (Carson et al. 1994) and one in Australia involving preservation of the Kakadu Conservation Zone (Carson, Wilks, and Imber 1994) included explicitly designed scope tests using goods where passive use considerations were thought to be the predominant source of value. Each of these surveys used identical descriptions of the local ecosystems involved, how the goods would be provided, and how they would be paid for, but one subsample was provided a good larger in scope than in the other subsample. In both cases, the scope insensitivity hypothesis is strongly rejected ($p < .001$) and willingness-to-pay estimates for the larger good in both cases are almost double that of the smaller good. The argument that scope insensitivity is a generic, unavoidable characteristic of contingent valuation studies has been shown to be false.

Of course, particular studies may show insensitivity to scope, and research has identified two main areas where this tends to occur. First, low-probability risks are often poorly understood in contingent valuation surveys, as they are by consumers in real-world behavior involving financial planning and insurance decisions. Various graphical representations have been shown to improve understanding in contingent valuation surveys (for example, Corso, Hammitt, and Graham 2001), and researchers are now looking at similar ways to assist consumers in making better financial planning decisions. Second, where a program is seen to provide multiple outputs, such as protecting different endangered species, it can be difficult to get distinct willingness-to-pay estimates for the individual outputs as opposed to the entire program.

While well-designed contingent valuation studies will typically pass a scope test, such tests have several conceptual problems that limit their potential usefulness. First, while contingent valuation critics sometimes contend that willingness to pay should be (almost) linearly increasing along some quantity dimension, declining marginal utility is more likely, which can influence the statistical power of scope tests (Rollins and Lyke 1998). Second, for a substantial fraction of the public, the likelihood of the government delivering on very large projects can be perceived to be much lower than that for smaller projects, in which case values placed on two goods may be entangled with beliefs about how well government functions. Finally, true willingness to pay may not even be monotonic in some instances. One can imagine a case, for example, in which a modest increase in the wolf population may be seen as a good thing, while a substantially larger increase is viewed negatively (Heberlein, Wilson, Bishop, and Schaeffer 2005). The time may have come to listen with an open mind to the message that survey respondents are seeking to convey when their answers suggest that changes in the scope of the good do not matter to them.

Contingent valuation critics sometimes also argue that the values of survey respondents must satisfy a more stringent sequential adding-up test whereby a composite good is broken into two parts and all three valued separately. Such a test is logically correct given its assumptions. But as Smith and Osborne (1996) point out, a key implicit assumption in natural resource damage assessment is that the replacement good is a perfect substitute. This requires, for example, that an agent be indifferent between saving a wild bird from being killed by oil and creating a hatchery program that produces a bird.[5] From a survey perspective, the adding-up test is problematic to implement because the survey for the second sub-components requires respondents to imagine they have received the first good and to imagine they have paid for the first good when asked about willingness to pay for the second good. Even putting moral and practical implementation objections aside, many people do not pass this adding-up test with market goods. Bateman, Munro, Rhodes, Starmer, and Sugden (1997) examined this experimentally, using students as the subjects and the highly familiar and frequently consumed goods of coffee and pizza. They fail the adding-up test. Many stores such as car dealers and cell phone providers are routinely successful selling customers additional goods and services *after* they purchase the car or phone that they were not otherwise going to purchase. Good contingent valuation studies do not engage in the survey equivalent of "upselling"; instead, they offer the complete bundle when the bundle is the relevant good for policy purposes.

Difficulties with the Hypothetical Bias Argument

Many economists instinctively think that the responses to contingent valuation questions will automatically overvalue people's true willingness to pay for public goods. In the context of contingent valuation surveys, this is called "hypothetical bias." Ironically, Samuelson (1954) saw the opposite problem in his classic article on public goods when he noted: "It is in the selfish interest of each person to give false signals, to pretend to have less interest in a given collective consumption activity than he really has," and he predicts that having the public complete "questionnaires" at different prices would fall prey to this strategic behavior. From Samuelson's view,

---

[5] The argument is sometimes put forward that anything that is put back physically cannot result in a loss in passive use value. This assumption is equivalent to denying the validity of a loss in utility from pain and suffering associated with a serious automobile injury, as long as the bones are eventually put back in place. In the context of an oil spill, it says agents cannot suffer a utility loss from knowing that animals suffer from being oiled, as long as the animal population and the ecosystem recovers. Because injuries to ecosystems often cannot be completely restored for any plausible cost, proponents of this assumption sometimes advocate the use of "habitat equivalency," a technique that translates loss in type of habitat into gains in another. This biological measure may be a reasonable proxy for small environmental injuries where the restoration or replacement is done in close proximity (on-site) to the original injury and involves very similar resources (in-kind). However, the approach breaks down for large-scale injuries. As an extreme example, the technique would allow the destruction of all wetlands in San Francisco Bay to be compensated for by restoring some amount of prairie grasslands in Nebraska. More important, perhaps, habitat equivalency has no direct tie to public welfare and, as such, should not be seen as a way of making the public whole.

those answering contingent valuation surveys about a public good should follow a free-rider approach of pretending to be less interested, hoping that the costs of providing the public good will fall on others. Which position does the empirical evidence support: the "hypothetical bias" prediction that surveys will overestimate true willingness to pay, or the Samuelson's prediction that strategic behavior will lead to an underestimate? The answer is "both." Survey exercises presented as purely hypothetical or having incentives encouraging overpledging can overestimate willingness to pay. However, contingent valuation surveys that are designed so that participants perceive them as consequential with a coercive payment mechanism and a reasonable set of auxiliary conditions (as discussed earlier) tend to, if anything, follow Samuelson's prediction of underestimating the true value, when they can be compared to other ways of calculating such values.

Studies of "hypothetical bias" are often done in an experimental context with students in which one group of subjects is told they will have to pay and another group of subjects are repeatedly told that responses are "purely hypothetical" in the sense of not having any effect on anything. The "hypothetical treatment" does typically lead to higher willingness to pay. In a meta-analysis of such studies, Murphy, Allen, Stevens, and Weatherhead (2005) find that the median ratio of estimated willingness to pay for purely hypothetical treatments to estimated willingness to pay in the actual payment treatments is 1.35, with a small number of very large outliers that drive up the mean ratio. Since a good contingent valuation study emphasizes the chance to influence whether the government will provide the good and the payment obligations if it is provided, it is not clear whether these purely hypothetical laboratory comparisons are of much relevance.

Another setting sometimes used to assert that contingent valuation suffers from "hypothetical bias" involves comparing actual contributions to a voluntary program to the propensity to contribute expressed in a survey context: that is, people say that they will contribute more in surveys than is actually contributed. This comparison has long been suspect. As we explain in Carson and Groves (2007), the most likely purpose for doing a survey asking about the likelihood of making a voluntary contribution is to help gauge whether to mount a fundraising effort. If the respondent wants the good, the optimal response is to appear ready to contribute in the survey to encourage the voluntary contribution campaign—and then to free-ride hoping others will contribute enough to provide the good. From this perspective, the economic puzzle then is not why the survey estimate is higher than actual contributions, but rather, why the difference between the two estimates is not larger.[6]

---

[6] Similarly, it has long been known that surveys of purchase intentions for new products in private markets tend to over-forecast actual purchases. In Carson and Groves (2007), we show that this result is theoretically predictable—respondents who potentially want to purchase the good should say "yes" to increase the likelihood that it is offered for sale, at which time they can then decide whether to buy. When people are surveyed about their likelihood of buying a private good that is already being offered for sale, respondents tend to have a lower propensity to buy in the survey than they do in markets. This effect is also in the predicted direction since respondents want to encourage firms to lower prices. For an interesting example involving existing toll roads where surveys underpredict usage, see Small, Winston, and Yan (2005).

With quasi-public goods, it is possible to compare estimates from contingent valuation studies with other ways of estimating values through some form of revealed preference for public goods. For example, the "travel cost method" involves people facing different travel costs for visiting a certain place—like a recreational fishing site. This price (in terms of travel costs) for going to that site can be used in conjunction with the number of trips to that site to estimate a demand curve and, in turn, willingness to pay for a trip to the site. The "hedonic pricing" method can be used with housing prices that incorporate spatially delineated amenities. Statistical methods can be used to control for other attributes of the home, like the number of bedrooms, in such a way that an estimate of the value of the environmental amenity can be obtained. The so-called "averting-behavior approach" looks at what people spend to avoid an adverse effect and allows the researcher to back out a derived demand for reducing it.

In Carson, Flores, Martin, and Wright (1996), we conducted a meta-analysis of 83 studies that included 616 comparisons of contingent valuation estimates to revealed preference estimates using these kinds of methods. We found that the mean ratio of contingent valuation to revealed preference estimates is 0.89 (with a 95 percent confidence interval of [0.81–0.96]), suggesting that contingent valuation estimates in the case of quasi-public goods are on average a bit lower than revealed preference estimates and reasonably correlated (0.78) with them.[7] Since that study, other papers have looked at valuing specific classes of goods using contingent valuation and revealed preference approaches and examined whether the details of the approach make a difference. For instance, the value of statistical life estimates from contingent valuation studies are of the same order of magnitude as those from hedonic wage studies using job risks, but smaller (Kochi, Hubbell, and Kramer 2006). Shrestha, Rosenberger, and Loomis (2007) perform a meta-analysis of studies on a large database of outdoor recreation valuation estimates. They find that contingent valuation estimates are significantly lower on average than comparable estimates based on revealed preference methods.

Yet another approach is to look at the small number of U.S. studies where a contingent valuation survey with (nearly) identical wording to an actual ballot proposition can be compared to the actual vote. Like many environmental goods, these ballot propositions tend to involve a mix of direct and passive use. As Kling, Phaneuf, and Zhao note, the comparisons between the contingent valuation estimates and actual votes are quite favorable, and they are clearly conservative when "don't knows" are treated as "no's" (the standard practice in the contingent valuation literature). This should not be surprising. Public polls taken near an actual vote, when the information set is unlikely to change, are on average quite good predictors of two-candidate races and ballot propositions. Predicting voter turnout

---

[7] The median ratio is somewhat lower at 0.75. There is a clear publication bias in studies comparing contingent valuation to revealed preferences estimates: published studies tend to either find a ratio of the two estimates close to one or a ratio that is very large. This two-humped distribution of published results suggests two very divergent expectations among economists, and that results can be cherry-picked to support a particular position.

is generally a harder task than predicting how people are going to vote conditional on the information they have.

Determining the Quality of a Contingent Valuation Study

A recurring theme of this essay has been that high-quality contingent valuation surveys appear to produce high-quality economic data. How does one separate the wheat from the chaff? Survey researchers point out that the most important thing to look at is the "face validity" of the entire contingent valuation survey instrument. Does the survey credibly pose a well-developed policy proposal to respondents and provide them with the necessary information to make an informed decision about it? Does the survey make respondents comfortable making a decision to either support or oppose the policy proposal and make them aware of the consequences if the policy is implemented? The best contingent valuation surveys are among the best survey instruments currently being administered while the worst are among the worst. In the hands of an expert in questionnaire design, face validity is not hard to judge. Economists are not typically trained with these skills so their judgments may need to be supplemented by those of people who do have the requisite training. Economists can judge whether a choice is consequential and when choices will reveal the desired tradeoffs.

Next, turn to the survey development effort. Ask whether adequate development and testing work was done in a deliberate, not pro forma, manner. Look at the survey administration and sampling. The Arrow et al. (1993) NOAA Panel Report recommended that surveys being done for litigation use in-person interviews with experienced professional interviewers to help motivate respondents to pay close attention to the details of the scenario, and that these surveys also have a rigorous sampling plan that is well executed. This is an enormously expensive undertaking, so it is here that one is most likely to see efforts to reduce cost. What are the implications of the survey implementation choices made?[8]

Now look at the basic results of the completed survey: Taking sampling error into account, does the percent of respondents willing to pay the randomly assigned cost amount fall as that amount increases? Is the estimate of willingness to pay derived using a statistical technique that is robust to assumptions about the far right tail of the distribution? Does the study present a construct validity equation that explains a reasonable amount (in a cross-sectional sense) of the heterogeneity in estimated willingness to pay and a comprehensive set of sensitivity analyses? Does the estimate from the study represent a sensible tradeoff that people might make to implement the policy in question?

---

[8] There is a lively debate in the literature over how to best deliver high-quality valuation estimates at lower costs, which is no surprise since the key question facing an agency doing a benefit–cost analysis is the value of spending a marginal dollar on a particular analysis and in allocating that dollar to one part of that analysis versus another.

Concluding Remarks

Contingent valuation is not perfect. No economic technique is. But the alternative to contingent valuation, especially in cases involving passive use considerations, is to place a zero value on goods that the public cares about—which is never likely to be the right choice.

In the two decades since the *Exxon Valdez* oil spill, the amount of research undertaken on contingent valuation has been substantial, including many thoughtful assessments starting with the Arrow et al. (1993) NOAA Panel Report by government agencies and international organizations (for example, Atkinson, Pearce, and Mourato 2006). The debate inside academic circles has often been acrimonious, but ultimately productive. The big issues concerning the reliability of contingent valuation raised by critics in the early 1990s have been resolved favorably with respect to the use of contingent valuation or have been shown to involve generic behavioral effects that also routinely characterize market data. A considerable body of evidence now supports the view that contingent valuation done appropriately can provide a reliable basis for gauging what the public is willing to trade off to obtain well-defined public goods. The time has come to move beyond endless debates that seek to discredit contingent valuation and to focus instead on making it better.

References

Aldy, Joseph E., Mathew J. Kotchen, and Anthony A. Leiserowitz. 2012. "Willingness to Pay and Political Support for a US National Clean Energy Standard." *Nature: Climate Change* 2(8): 596–99.

Atkinson, Giles, David W. Pearce, and Susana Mourato. 2006. *Cost-Benefit Analysis and the Environment: Recent Developments.* Paris: OECD.

Arrow, Kenneth, Robert Solow, Paul Portney, Edward Leamer, Roy Radner, and Howard Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." *Federal Register,* January 15, 58(10): 4601–14.

Bateman, Ian, Alistar Munro, Bruce Rhodes, Chris Starmer, and Robert Sugden. 1997. "Does Part–Whole Bias Exist? An Experimental Investigation." *Economic Journal* 107(441): 322–32.

Bowen, Howard R. 1943. "The Interpretation of Voting in the Allocation of Resources." *Quarterly Journal of Economics* 58(1): 27–48.

Carson, Richard T. 1997. "Contingent Valuation

and Tests of Insensitivity to Scope." In *Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods*, edited by R. J. Kopp, W. Pommerhene, and N. Schwartz, 127–163. Boston: Kluwer.

Carson, Richard T. 2011. *Contingent Valuation: A Comprehensive Bibliography and History*, Northampton, MA: Edward Elgar.

Carson, Richard T., Nicholas E. Flores, and W. Michael Hanemann. 1998. "Sequencing and Valuing Public Goods." *Journal of Environmental Economics and Management* 36(3): 314–23.

Carson Richard T., Nicholas E. Flores, Kerry M. Martin, and Jennifer L. Wright. 1996. "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods." *Land Economics* 72(1): 80–99.

Carson Richard T., Nicholas E. Flores, and Norman F. Meade. 2001. "Contingent Valuation: Controversies and Evidence." *Environmental Resources Economics* 19(2): 173–210.

Carson, Richard T., Nicholas E. Flores, and Robert C. Mitchell. 1999. "The Theory and Measurement of Passive-Use Value." In *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU and Developing Countries*, edited by I. J. Bateman and K. G. Willis, 97–130. Oxford: Oxford University Press.

Carson, Richard T., and Theodore Groves. 2007. "Incentive and Information Properties of Preference Questions." *Environmental and Resource Economics* 37(1): 181–210.

Carson Richard T., and W. Michael Hanemann. 2005. "Contingent Valuation." In: *Handbook of Environmental Economics*, Vol. 2, edited by K. G. Mäler and J. Vincent, 821–936. Amsterdam: North-Holland.

Carson, Richard T., W. Michael Hanemann, Raymond J. Kopp, Jon A. Krosnick, Robert C. Mitchell, Stanley Presser, Paul A. Ruud, and V. Kerry Smith. 1994. *Prospective Interim Lost Use Value Due to DDT and PCB Contamination in the Southern California Bight*. Report to the National Oceanic and Atmospheric Administration. http://www.econ.ucsd.edu/~rcarson/papers/SCalDDT.pdf.

Carson, Richard T., and Jordan J. Louviere. 2011. "A Common Nomenclature for Stated Preference Approaches." *Environmental and Resource Economics* 49(4): 539–59.

Carson Richard T., Robert C. Mitchell, W. Michael Hanemann, Raymond J. Kopp, Stanley Presser, and Paul A. Ruud. 1992. "A Contingent Valuation Study of Lost Passive Use Values Resulting from the *Exxon Valdez* Oil Spill." Report to the Attorney General of the State of Alaska. http://www.econ.ucsd.edu/~rcarson/papers/AlaskaReport.pdf.

Carson Richard T., Robert C. Mitchell, W. Michael Hanemann, Raymond J. Kopp, Stanley

Presser, and Paul A. Ruud. 2003. "Contingent Valuation and Lost Passive Use: Damages from the *Exxon Valdez* Oil Spill." *Environmental and Resource Economics* 25(3): 257–86.

Carson, Richard T., Leanne Wilks, and David Imber. 1994. "Valuing the Preservation of Australia's Kakadu Conservation Zone." *Oxford Economic Papers*, October, 46 (Special Issue on Environmental Economics): 727–49.

Chapple, Clive. 2000. "The 1990 Oil Pollution Act: Consequences for the Environment." Paper presented at the Association of Environmental and Resource Economists Summer Workshop, La Jolla, CA.

Ciracy-Wantrup, S. V. 1947. "Capital Returns from Soil-Conservation Practices." *Journal of Farm Economics* 29(4): 1181–96.

Corso, Phaedra S., James K. Hammitt, and John D. Graham. 2001. "Valuing Mortality-Risk Reduction: Using Visual Aids to Improve the Validity of Contingent Valuation." *Journal of Risk and Uncertainty* 23(2): 165–84.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature* 47(2): 315–72.

de Meijer, Claudine, Werner Brouwer, Marc Koopmanschap, Bernard van den Berg, and Job van Exel. 2010. "The Value of Informal Care–A Further Investigation of the Feasibility of Contingent Valuation in Informal Caregivers." *Health Economics* 19(7): 755–71.

Desvousges, William H., F. Reed Johnson, Richard W. Dunford, Kevin J. Boyle, Sara P. Hudson, and Nicole Wilson. 1993. "Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability." In *Contingent Valuation: A Critical Assessment*, edited by J. A. Hausman, 91–159. Amsterdam: North Holland.

Flores, Nicholas E., and Richard T. Carson. 1997. "The Relationship between the Income Elasticities of Demand and Willingness to Pay." *Journal of Environmental Economics and Management* 33(3): 287–295.

Freeman, A. Myrick. 2003. *The Measurement of Environmental and Resource Values*, 2nd ed. Washington, DC: Resources for the Future.

Griffiths, Charles, Heather Klemick, Matt Massey, Chris Moore, Steve Newbold, David Simpson, Patrick Walsh, and William Wheeler. 2012. "U.S. Environmental Protection Agency Valuation of Surface Water Quality Improvements." *Review of Environmental Economics and Policy* 6(1): 130–46.

Hanemann, W. Michael. 1991. "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" *American Economic Review* 81(3): 635–47.

Hanemann, W. Michael. 2008. *Embedding in Stated Preference*. Final Report to U.S. Environmental

Protection Agency, Agreement R-82966501, Department of Agricultural and Resource Economics, UC Berkeley.

Hausman, Jerry A., ed. 1993. *Contingent Valuation: A Critical Appraisal.* Amsterdam: North-Holland.

Hausman, Jerry A., Gregory K. Leonard, and Daniel McFadden. 1995. "A Utility-Consistent, Combined Discrete Choice and Count Data Model Assessing Recreational Use Losses Due to Natural Resource Damage." *Journal of Public Economics* 56(1): 1–30.

Heberlein, Thomas A., Mathew Wilson, Richard C. Bishop, and Nora C. Schaeffer. 2005. "Rethinking the Scope Test as a Criterion for Validity in Contingent Valuation." *Journal of Environmental Economics and Management* 50(1): 1–22.

Hicks, John R. 1943. "The Four Consumer Surpluses." *Review of Economic Studies* 11(1): 31–41.

Horowitz, John K., and Kenneth E. McConnell. 2002. "A Review of WTA/WTP Studies." *Journal of Environmental Economics and Management* 44(3): 426–47.

Jeuland, Marc, Marcelino Lucas, John Clemens, and Dale Whittington. 2009. "A Cost–Benefit Analysis of Cholera Vaccination Programs in Beira, Mozambique." *World Bank Economic Review* 23(2): 235–67.

Jiang, Yi, Leshan Jin, and Tun Lin. 2011. "Higher Water Tariffs for Less River Pollution—Evidence from the Min River and Fuzhou City in China." *China Economic Review* 22(2): 183–95.

Johnson, Bruce K., and John C. Whitehead. 2007. "Value of Public Goods from Sports Stadiums: The CVM Approach." *Contemporary Economic Policy* 18(1): 48–58.

Kahneman, Daniel, and Jack L. Knetsch. 1992. "Valuing Public Goods: Purchase of Moral Satisfaction." *Journal of Environmental Economics and Management* 22(1): 57–70.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47(2): 263–91.

Kochi, Ikuho, Bryan Hubbell, and Randall Kramer. 2006. "An Empirical Bayes Approach to Combining and Comparing Estimates of the Value of a Statistical Life for Environmental Policy Analysis." *Environmental and Resource Economics* 34(3): 385–406.

Krutilla, John V. 1967. "Conservation Reconsidered." *American Economic Review* 57(4):777–86.

Moore, William H. 1994. "The Grounding of the *Exxon Valdez*: An Examination of the Human and Organizational Factors." *Marine Technology Society Journal* 31(1): 41–51.

Murphy, James J., Geoffrey Allen, Thomas S. Stevens, and Darryl Weatherhead. 2005. "A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation." *Resource and Environmental Economics* 30(3): 313–25.

Nagin, Daniel S., Alex R. Piquero, Elizabeth S. Scott, and Laurence Steinberg. 2006. "Public Preferences for Rehabilitation versus Incarceration of Juvenile Offenders: Evidence from a Contingent Valuation Study." *Criminology and Public Policy* 5(4): 627–51.

Poe, Gregory L., and Christian A. Vossler. 2011. "Consequentiality and Contingent Values: An Emerging Paradigm." In *The International Handbook on Non-Market Environmental Valuation*, edited by J. Bennett, 122–41. Cheltenham, UK: Edward Elgar Publishers.

Rollins, Kimberly, and Audrey Lyke. 1998. "The Case for Diminishing Marginal Existence Values." *Journal of Environmental Economics and Management* 36(3): 324–44.

Samuelson, Paul A. 1954. "The Pure Theory of Public Expenditure." *Review of Economics and Statistics* 36(4): 387–89.

Shrestha, Ram, Randall S. Rosenberger, and John B. Loomis. 2007. "Benefit-Transfer Using Meta-Analysis in Recreation Economic Valuation." In *Environmental Value Transfer: Issues and Methods*, edited by S. Navrud and R. Ready, 161–77. Dordrecht, Netherlands: Springer.

Small, Kenneth A., Clifford Winston, and Jia Yan. 2005. "Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability." *Econometrica* 73(4): 1367–82.

Smith, V. Kerry, and Laura L. Osborne. 1996. "Do Contingent Valuation Estimates Pass a 'Scope' Test? A Meta-analysis." *Journal of Environmental Economics and Management* 31(3): 287–301.

Tversky, Amos, Paul Slovic, and Daniel Kahneman. 1990. "The Causes of Preference Reversal." *American Economic Review* 80(1): 204–217.

U.S. Environmental Protection Agency. 1994. *President Clinton's Clean Water Initiative: Analysis of Benefits and Costs.* EPA 800-R-94-002 (NTIS Document No. PB94-154101). Washington DC: Office of Water, United States Environmental Protection Agency.

Whittington, Dale. 2002. "Improving the Performance of Contingent Valuation Studies in Developing Countries." *Environmental and Resource Economics* 22(1): 323–67.

Willig, Robert D. 1976. "Consumer's Surplus without Apology." *American Economic Review* 66(4): 589–97.

Contingent Valuation: From Dubious to Hopeless

Jerry Hausman

Approximately 20 years ago, Peter Diamond and I wrote an article for this journal analyzing contingent valuation methods (Diamond and Hausman 1994). At that time Peter's view was that contingent valuation was hopeless, while I was dubious but somewhat more optimistic. But 20 years later, after millions of dollars of largely government-funded research, I have concluded that Peter's earlier position was correct and that contingent valuation is hopeless.

In this paper, I selectively review the contingent valuation literature, focusing on empirical findings. I find that three long-standing problems continue to exist: 1) hypothetical response bias that leads contingent valuation to overstatements of value; 2) large differences between willingness to pay and willingness to accept; and 3) the embedding problem which encompasses scope problems. In their overview essay in this journal, Kling, Phaneuf, and Zhao discuss all three of these issues. On the first two points, I do not find their conclusions differ too much from mine. But I think they underestimate the problems of embedding and scope, which are likely to be the most intractable of the problems. Indeed, I believe that respondents to contingent valuation surveys are often not responding out of stable or well-defined preferences, but are essentially inventing their answers on the fly, in a way which makes the resulting data useless for serious analysis. In this comment, I first discuss these issues. I then offer a case study of a prominent contingent valuation study done by recognized experts in this approach, a study that should be only minimally affected by these concerns but in which the answers of respondents to the survey are implausible and inconsistent.

■ *Jerry Hausman is the John and Jennie S. MacDonald Professor of Economics at the Massachusetts Institute of Technology, Cambridge, Massachusetts. His email address is jhausman@mit.edu.*

I am often asked what should be done given my view that contingent valuation should not be used. Should nonuse value be ignored? My view is that expert government agencies and Congress should make informed decisions and enact regulations that attempt to improve the economic allocation process (see also Diamond and Hausman 1994). To the extent that contingent valuation is interpreted as an opinion poll about the environment in general, rather than a measure of preferences about a specific project, public officials and regulators should recognize this concern.[1] However, public policy will do better if expert opinion is used to evaluate specific projects, including nonuse value, and to set appropriate financial incentives to reduce the risk of accidents such as the *Exxon Valdez* and BP disasters.

Responses to contingent valuation surveys for a single environmental issue are typically based on little information, given the limited time involved for each survey respondent. Thus, the results of such surveys are unlikely to be accurate predictors of informed opinion. Contingent valuation about specific projects does not improve the inputs to the analysis, so it should not be included in the policy analysis. Contingent valuation does not provide a good basis for either informed policymaking or accurate damage assessments in judicial proceedings.

Have the Empirical Problems of Contingent Valuation Been Addressed?

Possible problems of contingent valuation have been discussed in the literature for at least the past 30 years. While I focus on three of those problems, my chosen focus should not be taken to imply that other problems do not exist for individual studies or for the method as a whole.

Hypothetical Bias and Upward-Biased Results

The nature of a survey is that it asks a hypothetical question. Hypothetical bias is the bias that arises in answering a hypothetical question with which the respondent has no market experience; put simply, what people say is different from what they do. When hypothetical questions are asked about willingness to pay, the results tend to be upward-biased. This fact is well-known. For example, Jamieson and Bass (1989) studied people's stated intentions to purchase new products, and found that such measures were overstated. Other studies have affirmed this finding, like

---

[1] Referenda are similar to opinion polls except the results are often binding. The claim is sometimes made that contingent valuation studies, to the extent they can forecast how voters would respond in a binding referendum, should be used to design public policy. But even for referenda, the necessity of calibration to individual preferences to do welfare analysis remains, an issue which I discuss subsequently. Further, the use of actual referenda to obtain economic values is highly questionable. For example, because no immediate obvious budget constraint exists for voters in referenda, the evaluation of individual preferences from referenda is highly problematic. The importance of a budget constraint is fundamental to economic choices and its absence has an important distorting effect in contingent valuation studies, as we discuss in Diamond and Hausman (1994).

Hsiao et al. (2002) and Morwitz et al. (2007). (The latter paper finds that familiarity with the new product leads to more successful forecasts of whether people will buy, but familiarity with the product will not be present in most contingent valuation studies.) The NOAA panel (Arrow, Solow, Portney, Leamer, Radner, and Schuman 1993) found upward bias to be present in willingness-to-pay responses for both private goods and public goods, which they determined would extend to contingent valuation studies. In their overview paper, Kling, Phaneuf, and Zhao find that significant biases exist and their "net impact varies with the characteristics of participants and the commodity, and the type of script used."

A standard response to this problem has long been to apply a "fudge factor"—that is, to deflate the stated willingness to pay by some amount. How much? NOAA proposed dividing contingent valuation results by two (*Federal Register* 1994). But of course, there is no reason that the degree of overstatement should be the same across all survey methods, commodities, and types of survey respondents. I do not see how past studies provide a basis for an appropriate estimate of the needed adjustment. Suppose that a new econometric estimator was proposed that had a mean bias of 300 percent, but the 95 percent confidence interval varied from 150 to 800 percent. Such an estimator would not be used in serious policy formation.

Why the bias and large variation in answers to hypothetical questions? One issue, common in the opinion survey literature, is that those being interviewed often seek to please the interviewer (Tourangeau, Rips, and Rasinski 2000). Other issues may arise because of the specific nature of contingent valuation surveys. As Horowitz (2000) points out, the standard form of a contingent valuation question asks about one's willingness to pay a certain amount for a certain outcome, but doesn't say explicitly how the answer will be used, nor offer a range of options, nor offer a chance for discussion and interaction with others (as does a public voting process). Supporters of such surveys spend considerable time and energy on the precise wording of their questions and they test different wording choices in focus groups in an effort to make respondents feel that their answer really matters and that they should take the task of answering seriously. But as Harrison (2007) writes: "The literature on non-market valuation in environmental economics is littered with assertions that one can somehow trick people into believing something that is not true. . . . The claims tend to take the form, 'if we frame the hypothetical task the same way as some real-world task that is incentive compatible, people will view it as incentive compatible.'"

But despite such efforts, for whatever reason, hypothetical bias persists. For example, Murphy and Stevens (2004) note that the literature shows hypothetical bias across a wide variety of contingent valuation approaches. Johnston (2006, p. 469) concurs: "Most research finds significant divergence between stated and actual behaviors." Kling, Phaneuf, and Zhao agree as well.

Sometimes supporters of contingent valuation surveys compare them to polls about public referenda—and thus seek to give them a presumption of legitimacy because polls are useful at predicting the outcome of the democratic process. But of course, polls predicting the outcome of referenda are sometimes accurate, sometimes

not. Often, the polls about a referenda change considerably over the period of time leading up to an election, suggesting that preferences about the choice were not especially stable at the beginning of the process. Vossler, Kerkvliet, Polasky, and Gainutdinova (2003) look at survey responses and an actual referendum vote on a proposal to protect open space in Corvallis, Oregon. They find that it is necessary to treat the "undecided" vote as "no" if the survey is to avoid hypothetical bias and to reflect the actual outcome of the vote. Of course, one should be cautious about extrapolating from surveys about a vote on a local public good to contingent valuation surveys on other subjects. The empirical basis for weighting undecided or other responses to a contingent valuation survey in some way that is intended to reduce hypothetical bias is scanty at best.

Difference between Willingness to Pay and Willingness to Accept

Contingent valuation questions can be phrased in two broad ways: the willingness-to-pay approach seeks to discern what the respondent would pay to avoid a negative outcome (or to achieve a positive outcome), while the willingness-to-accept approach seeks to discern how large a payment the respondent would need to receive in order to accept the negative outcome (or not to receive a positive outcome). Basic economic theory suggests that these two approaches should give (approximately) the same answer, but both supporters and skeptics of contingent value methods recognize that large and persistent disparities commonly arise in answers to contingent valuation surveys.

Broadly speaking, there have been two approaches to rationalizing the large gaps between willingness to pay and willingness to accept: use a theoretical background rooted in behavioral economics, or relax enough assumptions in the neoclassical model. The difficulty with either approach is that if benefit–cost analysis is to be logically coherent, it requires a theoretical framework. The Hicksian foundations of standard welfare analysis are based on compensated demand curves and potential Pareto improvements, and no substitute foundational framework has received wide acceptance to replace the Hicksian approach. (For a discussion of the Hicksian basis for welfare analysis, see Hausman 1981, and Hausman and Newey 1995.) Rationalizations of the gap between willingness to pay and willingness to accept come at the expense of introducing assumptions that render standard benefit–cost analysis invalid.

For example, suppose that consumers do not have neoclassical preferences, but instead are subject to "loss aversion," and thus they will weight prospective losses more heavily than equivalent gains. Such "behavioral" preferences will indeed drive a wedge between willingness to accept and willingness to pay. But it becomes unclear how to do welfare analysis of gains to some groups and losses to others with these assumptions; necessary compensation for aggrieved losers from any policy may well outstrip gains to the winners.

Various efforts have been made to extend the neoclassical framework in a way that rationalizes the gap. Proponents of contingent valuation have attempted rationalizations of these differences, but have not overcome the findings of Diamond and Hausman (1994) or the results of Milgrom (1993). Both papers demonstrate

that the attempts to rationalize the well-recognized and persistent disparity between willingness to pay and willingness to accept fail as a matter of economic theory and observed empirical outcomes.

Of course, one can claim that consumers do not have neoclassical preferences. But standard cost–benefit analysis, and the underlying logic of being able to sum the willingness to pay of many individuals, requires that individual preferences are being measured (Diamond and Hausman 1994, pp. 55–58). Essentially, use in policy analysis or damage analysis depends on willingness to pay being a measure of the compensating variation for avoiding a negative outcome. If the neoclassical assumptions are relaxed so that willingness to pay includes, say, a component for altruism or for sympathy, then willingness to pay will diverge from willingness to accept—but then addition across the willingness to pay of individuals is no longer appropriate. In aggregation, the neoclassical model requires that preferences be over states of the world and not over acts: for example, preferences must be over the choice between two different states of a wilderness area, not over whether the respondent receives a warm glow from the idea of saving a wilderness area, nor about a general attitude about providing public goods in general. Again, this assumption is required for consistent economic policy choices (see also the NOAA Report, Arrow et al. 1993).

Of course, there are a number of other ways to attempt to rationalize the large gaps between willingness to pay and willingness to accept in contingent valuation surveys. It's possible to create a theory that is consistent with (almost) any given set of facts. But the task of building the foundations of a benefit–cost analysis on top of those alternative theories has not been done. And as I demonstrate throughout this essay, the gaps are likely due to the reality that answers to contingent valuation surveys do not actually reflect stable or well-defined preferences but instead are opinions invented on the fly.

Scope and Embedding

The most fundamental challenge to the contingent value method, and the strongest evidence that the answers to such surveys are invented in response to the questions, comes from concerns that are referred to as "scope" and "embedding." Kahneman and Knetsch (1992) were the first to explore the "embedding effect," which demonstrates the nonexistence of preferences in a contingent valuation setting. As they wrote, "perhaps the most serious shortcoming of CVM [contingent value methods]" is that "the assessed value of a public good is demonstrably arbitrary, because willingness to pay for the same good can vary over a wide range depending on whether the good is assessed on its own or embedded as part of a more inclusive package." In Diamond and Hausman (1994), we provide an example of the embedding effect, where willingness to pay to clean one lake is approximately equal to stated willingness to pay to clean up five lakes—including the one asked about individually. Embedding is related to the scope effect, which is the broader proposition that respondents to contingent valuation surveys should be more willing to pay for a large effect than for a subset of that effect. Proponents of contingent valuation would like to demonstrate a scope affect, but the scope

effects typically found are not nearly large enough to make contingent valuation results credible.

In the earlier version of their overview paper, Kling, Phaneuf, and Zhao acknowledged that "scope effects are typically present and positive, if not always large." But I find this result to be similar to a finding in econometric estimation that demand curves slope downward—a very weak test with almost no power. We do not know how large scope effects should be. Indeed, since contingent valuation surveys are typically pretested, the survey design can be manipulated to ensure that at least minimal scope effects are present.

Thus, in Diamond and Hausman (1994), we proposed a more stringent version of a scope test called an "adding-up test." The test works this way: a first group of respondents is asked their willingness to pay for a public good X; a second group is asked their willingness to pay for public good Y; and a third groups is asked their willingness to pay for X and Y together. The total value of the entire project minus the value of the first project should approximately equal the value of the incremental projects.[2] A specification test then permits one to statistically determine whether embedding is present. If Diamond–Hausman tests are done correctly, they first establish what the willingness to pay is for a given project. They then establish the willingness to pay for a larger project in which the first project is present, which establishes the *incremental amount* of willingness to pay for the additional projects. In this way, the adding-up test overcomes the problem of a more general scope test.

Desvousges, Mathews, and Train (2012a) review 109 contingent valuation studies on environmental goods since 1994 that apply a scope test. They find that most studies do not provide sufficient information to determine whether the difference in survey response to variations in scope is "adequate" (p. 4). Those who conduct contingent valuation surveys have typically not collected their data in way that makes an adding-up test possible. They find only one study that permits a test of adding-up: a study by Chapman et al. (2009). This study passes a scope test, but fails the more stringent adding-up test. Desvousges, Mathews, and Train (2012b) expand the Chapman et al. (2009) survey to measure the value of each increment directly with contingent valuation, and they find that the sum of the estimated values of the incremental parts is three times greater than the estimated value of the whole. They conclude—as do Heberlein, Wilson, Bishop, and Schaeffer (2005) and Bateman (2011)—that "standard scope tests are uninformative."

My view is that until contingent valuation surveys can reliably pass the Diamond–Hausman adding-up test (or a similar test) to demonstrate that embedding is not present, the results do not indicate stable or coherent individual preferences.

---

[2] This statement of the test leaves out income effects, but income effects are typically quite small for projects considered by contingent valuation. In a paper we are still working on (Hausman and Newey, no date), my coauthor and I develop bounds that take account of the share of income spent on a good and its income derivative. Our results demonstrate that for the size of contingent valuation projects, typically less than $100, the upper and lower bounds are almost identical for consumer surplus.

Recent empirical evidence demonstrates that some problems that exist in contingent valuation studies also exist in actual market situations. For example, framing of questions can lead to very different results in contingent valuation studies, as I discuss below in my case study. Evidence demonstrates that framing can also affect consumer choices in the market. Thus, some of the problems in contingent valuation also exist in revealed preference outcomes. However, I expect that consumers do better (even though they still make mistakes) for important decisions and for repeated decisions. And consumers have a budget constraint which has a large effect on their decisions. For public policy purposes, expert analysis, as I discuss above, will hopefully avoid the "mistakes" that would arise with the use of contingent valuation and come to better allocative outcomes than if we were to depend on results from contingent valuation surveys that are not consistent with fundamental economic preferences on which we base economic welfare analysis.

A Case Study: Contingent Valuation and Australian Cable Television

To provide a more concrete illustration of these issues, I will consider a particular contingent valuation study.[3] This particular study is chosen for several reasons. It was implemented by Richard Carson, a participant in this symposium and someone widely recognized as a top expert in contingent valuation studies.[4] The design and implementation of the study was not constrained in any meaningful way by lack of a budget. It had large sample sizes. Unlike some contingent value surveys that deal with issues far-removed from the daily experience of the respondents—like the value to be placed on cleaning up Prince William Sound where the *Exxon Valdez* ran aground—this survey dealt with a product well-known to those being surveyed: cable television. Yet despite these advantages, I will demonstrate that the results from this contingent value survey are unreliable. The results demonstrate that although people responded to the survey, their answers cannot correctly be treated as a meaningful measure of preferences.

Of course, one study does not discredit contingent valuation methodology. And the decisionmakers in this case, on the Australian Copyright Tribunal, do not have the last word on economic methodology. Yet the case study is potentially useful for thinking about the issues of contingent valuation. The contingent valuation study did not solve the well-recognized problems of contingent valuation, even though it

---

[3] As an econometrician, I typically do not rely on case studies in my academic research. However, contingent valuation studies that include both a critique and then a decision from an outside party are limited to "high stakes" proceedings. I am not aware of many such cases. Thus, I cannot do a meta-analysis of many contingent valuation studies, because the cost of analyzing them is typically quite high. The Australian case study seems especially useful since it was a "high stakes" proceeding with significant analysis and a decision by an informed tribunal.

[4] All the participants in this study, including myself, involved in the Australian case study were paid consultants. The contingent valuation study was paid for by Screenrights (the copyright holder) and my report was paid for by the cable TV providers.

had experienced academic experts in contingent valuation who were not subject to a tight budget constraint to finance their analysis. And the Australian contingent valuation study failed the Diamond–Hausman test as I describe below.

The background situation is that cable TV companies in Australia retransmit the free-to-air TV channels' broadcasts, as do cable TV companies in most of the world. In 2001, a change in Australian law defined retransmission as an infringement of copyright, requiring the cable TV companies to pay "equitable remuneration" to the copyright owners via their declared collecting society, Screenrights. The cable TV companies and Screenrights were unable to agree upon what constituted "equitable remuneration," and the matter was brought before the Copyright Tribunal, which is administered by the Federal Court of Australia. Screenrights' primary evidence before the Tribunal was a contingent valuation study designed by Jeff Borland and Richard T. Carson, and conducted by a leading Australian research company. The primary evidence opposing this approach was a study provided by Tim Bock and myself.[5] Thus, I confess that yet another reason for choosing this study is that it was the most recent contingent valuation study that I have analyzed in detail. The 2006 Copyright Tribunal decision has a useful extended overview of all the issues of the case, with several sections focused on contingent valuation, available online at http://www.austlii.edu.au/au/cases/cth/ACopyT/2006/2.html.

The contingent valuation study actually involved two parts. In the first part, 2,622 subscribers participated in a 10-minute personal interview by surveyors who knocked on their doors at home. The survey asked general questions about household structure and behavior, and then respondents were both read and asked to read descriptions about the benefits of retransmission and available substitutes. They were then asked, using a formal script, if they had the choice of 1) paying $X extra per month to continue receiving the regular TV channels through cable TV, or 2) paying the same as before and losing these channels from cable TV (but perhaps getting them through a TV aerial), what would they choose? The respondents were randomly allocated to one of five monthly fees: $1.00, $2.50, $5.00, $7.50, and $10.00. In an unexpected twist, it was discovered that the first part of the study had not correctly implemented procedures for recontacting those households not at home ("call-backs"), therefore the study was repeated with a new sample of 2,369 households, and some minor wording changes were made in the second study.

Specific responses to the first and the second survey appear in Table 1. One oddity jumps out immediately: In both studies, the quantity demanded at $10 is higher than demanded at $7.50. Although it is highly unusual in real markets to find that a 33.3 percent increase in price does not cause an outright decline in quantity demanded, at least the increase in quantity demanded here is not statistically significant. Other questions also come to mind: for example, the Copyright Tribunal questioned why the lowest value surveyed was set at $1, rather than some

[5] More explanation and detail are provided in Hausman and Bock (2007).

Table 1

Binary Choice Data from the Australian Cable Television Contingent Valuation Survey

	Monthly subscription fee				
	$1.00	*$2.50*	*$5.00*	*$7.50*	*$10.00*
Study 1					
Pay	299	224	176	139	140
Not pay/Don't know	224	294	333	403	390
Study 2					
Pay	292	207	152	85	90
Not pay/Don't know	199	288	321	375	360

lower number. But I will focus on two fundamental issues and then mention some other points.

First, preferences in this contingent valuation study appear to be irrationally unstable, in the sense that minor differences in wording—that is, the framing of the questions—led to large differences in response. Study 1 finds a 32 percent higher share of respondents who state they are willing to pay $10 per month for retransmission compared to Study 2. Similarly, 39 percent more respondents in Study 1 said they would pay $7.50 than in Study 2. Yet the questions in the two studies are essentially identical, with only a small amount of additional information in Study 2. Borland and Carson agreed that the discrepancies in the demand curves were attributable to changes in question wording. In my view, the only significant change was that respondents were shown both a monthly and annual fee in Study 2, while in Study 1 only the monthly fee was shown. If relatively minor changes in wording lead to significant differences in results, I would refer to this situation as "irrational preference instability." Such results support a conclusion that consumers did not reveal true preferences in the stated preference questionnaire and are instead, to some unknown extent, "making-up" or "inventing" their answers to a hypothetical situation with which they are unfamiliar. I conclude that consumers do not have well-formed preferences, which is why their responses to the main contingent valuation question were significantly influenced by the survey wording.

Second, the study results fail a Diamond–Hausman (1994) adding-up test, discussed earlier. Specifically, this test checks to see if average willingness to pay (WTP) for divisible good X is equivalent to the sum of average WTP for kX and average WTP for $(1-k)X$ *conditional* upon kX already having been supplied, where $0 < k < 1$. For example, if a consumer is willing to pay $50 for two items together, such as local telephone service and a broadband Internet connection, the consumer should be willing to pay (approximately) this same amount if the consumer first purchases the local telephone service and then buys the broadband service (after purchasing the local telephone service).

The Diamond–Hausman test was administered by another fieldwork company using the protocols employed in Study 1. Some minor wording changes were made to the contingent question to improve its intelligibility, but these changes were not contentious between the parties and their experts involved in the proceeding. Three independent samples ($N = 200$ in each; 600 in total) received different versions of the questionnaire, as follows:

Version 1: Retransmission of only ABC, Channel 9 and SBS (that is, kX).

Version 2: Retransmission of Channel 7 and Channel 10, given that ABC, Channel 9 and SBS are already being retransmitted to the household (that is, $(1 - k)X$ given kX).

Version 3: Retransmission of all the free-to-air channels (that is, X).

To estimate a mean from the kind of data in Table 1, one standard approach is to use the Turnbull Lower Bound estimator. (The nonparametric Turnbull estimator begins by determining the fraction of refusals falling into each dollar interval, and a lower-bound estimate of the mean follows from these fractions.[6]) By this measure, the mean for Version 1 is $2.96, the mean for Version 2 is $1.64, and the sum of these two is $4.60. However, the mean is $2.81 for Version 3. Thus, the sum of Version 1 and 2 is 64 percent greater than Version 3.

One potential objection to this comparison is that something is amiss with the follow-up survey. However, the estimated willingness to pay for in Versions (2) and (3) of the adding-up test survey are quite close to the willingness to pay in Study 2 shown in Table 1. This outcome strongly suggests that respondents reacted in a similar manner to these two surveys. Another objection sometimes raised is that asking a respondent to "pretend" they have already obtained part of a good is problematic because it may be difficult to get respondents to take such an exercise seriously. But if this objection is true, it would invalidate both the potential problem and the original contingent valuation study. After all, contingent valuation questions—including the ones in this study— often set up scenarios that ask respondents to pretend.

As one might expect, there were a number of other points at issue in the discussion before the Copyright Tribunal, which are summarized in its report. For example, we also demonstrated that if the demand curves estimated from the contingent valuation survey were to be taken seriously, the cable television companies would increase

---

[6] The Turnbull estimator is computed using the Pooled Adjacent Violators Algorithm and treating "don't know" as "not pay" (as discussed in Bateman et al. 2002, p. 231). The algorithm for the Turnbull estimator is roughly as follows: For the lowest bid level, calculate the proportion of refusals. Then move to the next-highest bid level and again calculate the number of refusals. Continue this process up through the bid levels, and use this data to calculate a cumulative density function from which you can derive a probability density function. Multiply the probability density function by the bid defining the lower bound, and then sum over all bid levels. The other method often used for this calculation is a Weibull distribution, which is a two-parameter location-scale distribution often used in duration models. In this case, the approach is to estimate the two parameters that would characterize a Weibull distribution that fits the pattern of the data, and then to use the properties of that distribution to calculate the mean. The Weibull distribution will typically give a mean estimate greater than the Turnbull estimator.

their profits by charging an extra $10 per month (or more): that is, the extra revenue they would gain by charging more would more than offset the losses from those who decided not to subscribe to cable television at all. Instead, it turned out that one cable company, Foxtel, was charging the same for digital satellite and digital cable, but was including all the free-to-air stations in the first delivery mechanism but not the second. Apparently, providing the free-to-air stations was not a service for which the company thought it could charge more. But the primary argument that is relevant for thinking about contingent valuation methods as a whole is that the answers from such studies are unstable and inconsistent, invented for the moment of the survey, and cannot be treated as preferences in the sense that economist understand that term.

After reviewing all the arguments, Australia's Copyright Tribunal (2006, par. 510 and 512) chose to disregard completely the evidence from the contingent value survey. It quoted a 1965 case to the effect that "[A] person exercising quasi-judicial functions must . . . not spin a coin or consult an astrologer, but he may take into account any material which, as a matter of reason, has some probative value. . . . If it is capable of having any probative value, the weight attached to it is a matter for the person to whom Parliament has entrusted the responsibility of deciding the issue." Having expressed a willingness to give at least some weight to any evidence that might be relevant, the Tribunal wrote: "Courts and tribunals must proceed on the basis of probative evidence, not speculation. . . . We have such a level of doubt about the Survey that we attach no weight to it."

Although the Copyright Tribunal decided that in this situation—to paraphrase the title of Diamond and Hausman (1994)—no number was better than the contingent valuation number, it did rely on a range of other evidence to decide that the cable companies should pay 22.5 cents per subscriber per month in exchange for transmitting the free-to-air content. This amount was not even in the range of possibilities considered in the contingent valuation study. Thus the Copyright Tribunal chose to use its expert opinion to set the rate per subscriber, and it completely ignored the outcome of the contingent valuation study.

Of course, this particular case study addresses only one contingent value survey—but, again, it is presumably a "high quality" study by the standards of this literature. The study by Chapman et al. (2009) of the aesthetic and ecosystem value of certain water resources in Oklahoma is another "high quality" contingent study designed and implemented by proponents of such studies. It had large sample sizes, a budget constraint that did not bind very tightly, and claimed to meet best-practice guidelines. Yet as discussed earlier, Desvousges, Mathews, and Train (2012a, b) show that the results of this contingent valuation survey are unreliable for various reasons, including failing an adding-up test.

Conclusion

The controversy over contingent valuation studies often follows a predictable pattern. A contingent value study is designed and carried out, with much talk about

how methodology has strengthened over time. When the results are announced, critics point out potentially severe problems, like hypothetical bias and overstatement, disagreements between willingness to pay and willingness to accept, and problems of scope or embedding. Supporters then respond that perhaps this particular study wasn't well-designed, and that there are ways to make adjustments, and that it would be wrong to conclude from one study that the enterprise of contingent valuation is fundamentally flawed. Then the next study arrives and is criticized and defended in the same way. For those of us who have criticized a number of contingent valuation studies, it feels as if proponents of contingent valuation retreat to the position that all studies shown to be inaccurate are examples of poor practice rather than any inherent flaw. But despite all the positive-sounding talk about how great progress has been made in contingent valuation methods, recent studies by top experts continue to fail basic tests of plausibility.

I expect that if contingent value respondents had been asked about Prince William Sound (where the *Exxon Valdez* ran aground) and another group was asked about Prince Andrew Sound (fictitious) after being told that Price William Sound had been saved, and a third group was asked about Price William Sound and Price Andrew Sound together, the combined response would not be much different than the individual responses, so that the sum of the individual responses would be significantly greater than the combined response. When contingent studies can routinely pass Diamond–Hausman adding-up tests I am willing to reconsider my conclusion of little or no progress over the past 20 years in solving the most important problems with contingent valuation. But even if that event occurs, contingent valuation would still face problems like how to address the upward bias in responses and how to build a framework for cost–benefit analysis in a setting where the data show a gulf between willingness to pay and willingness to accept.

I do not expect these problems to be resolved, so in my view "no number" is still better than a contingent valuation estimate. Moreover, as the discussion of Australian Copyright Tribunal (2006) showed, other pieces of evidence can be brought to bear on goods that are not directly valued in the market. For example, in environmental damage situations, the method of "habitat equivalency analysis" relies on a group of trustees appointed through government or the courts to analyze what expenditures are needed to restore the environment (Damage Assessment and Restoration Program 2006). The political process can also provide outcomes. As Diamond and I wrote in our 1994 essay in this journal (pp. 58–59), "the choice is between relying on Congress after doing a contingent valuation study and relying on Congress without doing such a contingent valuation study." My theme is that unless or until contingent value studies resolve their long-standing problems, they should have zero weight in public decision-making.

I do not expect that proponents and opponents of contingent valuation will ever agree. Some bad ideas in economics and econometrics maintain a surprising viability. Numerous branches of the federal government continue to fund contingent valuation research in the hope that it will support their favored policies subject to cost–benefit analyses. In turn, the proposed regulations lead to push-back from

those who would bear the costs. In cases like the *Exxon Valdez* spill or the BP *Deep-water Horizon* spill, vast amounts of money are at stake. I do not find my view that such debates will persist to be at all cynical; rather, it is the expected outcome given the incentives that all parties face.

References

Arrow, Kenneth, Robert Solow, Paul R. Portney, Edward E. Leamer, Roy Radner, and Howard Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." National Oceanic and Atmospheric Administration (U.S. Department of Commerce).

Australian Copyright Tribunal (see Copyright Tribunal of Australia)

Bateman, Ian J. 2011. "Valid Value Estimates and Value Estimate Validation: Better Methods and Better Testing for Stated Preference Research." Chap. 16 in *The International Handbook on Non-Market Environmental Valuation,* Jeff Bennett, editor, Cheltenham UK: Edward Elgar Publishing.

Bateman, Ian J., et al. 2002. *Economic Valuation with Stated Preference Techniques: A Manual.* Cheltenham: Edward Elgar.

Chapman, David J., Richard C. Bishop, W. Michael Hanemann, Barbara J. Kanninen, Jon A. Krosnick, Edward R. Morey, and Roger Tourangeau. 2009. "Natural Resource Damages Associated with Aesthetic and Ecosystem Injuries to Oklahoma's Illinois River System and Tenkiller Lake." Expert Report for State of Oklahoma. Case No. 05-CV-0329-GKF-SAJ. *State of Oklahoma v. Tyson Foods, et al.* In the United States District Court for the Northern District of Oklahoma. Volume I.

Copyright Tribunal of Australia. 2006. *Audio-Visual Copyright Society Ltd v Foxtel Management Pty Ltd (No 4) [2006] ACopyT 2* (Ruling), Date of Judgment: May 3, 2006. http://www.austlii.edu.au/au/cases/cth/ACopyT/2006/2.html.

Damage Assessment and Restoration Program, National Oceanic and Atmospheric Administration, U.S. Department of Commerce. 2006. "Habitat Equivalency Analysis: An Overview." May 23. (Originally published March 21, 1995.) http://www.darrp.noaa.gov/library/pdf/heaoverv.pdf.

Desvousges, William, Kristy Mathews, and Kenneth Train. 2012a. "Adequate Responsiveness to Scope in Contingent Valuation." http://elsa.berkeley.edu/~train/scope.pdf.

Desvousges, William, Kristy Mathews, and Kenneth Train. 2012b. "Rationality Reduc: The Incremental Adding up Test for Contingent Valuation." Unpublished paper, Department of Economics, University of California, Berkeley.

Diamond, Peter A., and Jerry A. Hausman. 1994. "Contingent Valuation: Is Some Number Better Than No Number?" *Journal of Economic Perspectives* 8(4): 45–64.

Federal Register. 1994. May 14, 59(85): 1183.

Harrison, Glenn W. 2007. "Making Choice Studies Incentive Compatible." Chap. 4 in *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice,* edited by Barbara Kanninen. Dordecht, The Netherlands: Springer.

Hausman, Jerry. 1981. "Exact Consumer Surplus and Deadweight Loss." *American Economic Review* 71(4): 662–76.

Hausman, Jerry A., and Timothy Bock. 2007. "The Falsification of Contingent Valuation: A Case Study." Unpublished paper.

Hausman, Jerry A., and Whitney K. Newey. 1995. "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss." *Econometrica* 63(6): 1445–76.

Hausman, Jerry A., and Whitney K. Newey. No date. "Non-separable Estimation of Exact Consumers Surplus and Deadweight Loss with Arbitrary Heterogeneity." Paper not completed.

Heberlein, Thomas A., Matthew A. Wilson, Richard C. Bishop, and Nora Cate Schaeffer. 2005. "Rethinking the Scope Test as a Criterion for Validity in Contingent Valuation." *Journal of Environmental Economics and Management* 50(1): 1–22.

Horowitz, John. 2000. Policy Discussion section of "Stated Preference: What Do We Know? Where Do We Go?" Proceedings of Session One of "Theory and Design of Stated Preference Methods," A Workshop Sponsored by the USEPA's National Center for Environmental Economics and National Center for Environmental Research. October 12–13, Doubletree Hotel, Washington, DC. http://epa.gov/ncer/publications/workshop/pdf/EE-0436-01.pdf.

Hsiao, Cheng, Baohong Sun, Vicki G. Morwitz. 2002. "The Role of Stated Intentions in New Product Purchase Forecasting." In *Advances in Econometrics,* vol. 16, pp. 11–28. Emerald Group Publishing.

Jamieson, Linda F., and Frank M. Bass. 1989. "Adjusting Stated Intention Measures to Predict Trial Purchase of New Products: A Comparison of Models and Methods." *Journal of Marketing Research* 26(3): 336–45.

Johnston, Robert J. 2006. "Is Hypothetical Bias Universal? Validating Contingent Valuation Responses Using a Binding Public Referendum." *Journal of Environmental Economics and Management* 52(1): 469–81.

Kahneman, Daniel, and Jack Knetsch. 1992. "Valuing Public Goods: The Purchase of Moral Satisfaction." *Journal of Environmental Economics and Management* 22(1): 57–70.

Milgrom, Paul. 1993. "Is Sympathy an Economic Value?" In *Contingent Valuation: A Critical Assessment,* edited by J. Hausman, 417–42. Amsterdam: North Holland Press.

Morwitz, Vicki, Joel H. Steckel, and Alok Gupta. 2007. "When Do Purchase Intentions Predict Sales?" *International Journal of Forecasting* 23(3): 347–64.

Murphy, James J., and Thomass H. Stevens. 2004. "Contingent Valuation, Hypothetical Bias, and Experimental Economics." *Agricultural Resource Economics Review* 33(2): 182–192.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response.* Cambridge: Cambridge University Press.

Vossler, Christian A., Joe Kerkvliet, Stephen Polasky, and Olesya Gainutdinova. 2003. "Externally Validating Contingent Valuation: An Open-Space Survey and Referendum in Corvallis, Oregon." *Journal of Economic Behavior and Organization* 51(2): 261–77.

The End of Cheap Chinese Labor

Hongbin Li, Lei Li, Binzhen Wu, and Yanyan Xiong

I n recent decades, cheap labor has played a central role in the Chinese model, which has relied on expanded participation in world trade as a main driver of growth (Lin, Cai, and Li 2003; Bernstein 2004). At the beginning of China's economic reforms in 1978, the annual wage of a Chinese urban worker was only $1,004 in U.S. dollars: that is, 615 renminbi yuan divided by China's official exchange rate of 1.68 yuan/dollar in that year, and then deflated to the 2010 level by the U.S. GDP deflator. (The official exchange rate was overvalued at the time, but it is useful in measuring the price that U.S. consumers pay for Chinese labor embodied in Chinese goods.) Back in 1978, China's wage was only 3 percent of the average U.S. wage at that time, and it was also significantly lower than the wages in neighboring Asian countries such as the Philippines and Thailand. The Chinese wage was also low relative to productivity. According to Ceglowski and Golub (2007), China's "unit labor cost"—wage as a percentage of labor productivity—relative to the same ratio in the United States declined from over 70 percent in the 1980s to about 30 percent in the mid-1990s.

However, wages are now rising in China. In 2010, the annual wage of a Chinese urban worker reached $5,487 in U.S. dollars—that is, 37,147 yuan divided by the

■ *Hongbin Li is C.V. Starr Professor of Economics, Lei Li is a Ph.D. student, and Binzhen Wu is Assistant Professor of the School of Economics and Management, all at Tsinghua University, Beijing, China. Yanyan Xiong is Associate Professor, School of Economics and Management, Southeast University, Nanjing, China. All four authors are also affiliated with the China Data Center, Tsinghua University. Their email addresses are lihongbin@sem.tsinghua.edu.cn, lil6.11@sem.tsinghua.edu.cn, wubzh@sem.tsinghua.edu.cn, and xiongyanyan@seu.edu.cn, respectively. Yanyan Xiong is the corresponding author.*

exchange rate of 6.77 yuan/dollar—which is similar to wages earned by workers in the Philippines and Thailand and significantly higher than those earned by workers in India and Indonesia. China's wages also increased faster than productivity since the late 1990s, suggesting that Chinese labor is becoming more expensive in this sense as well.

The opening section of this paper discusses China's rising wages in more detail. For example, the increase in China's wages is not confined to any sector, as wages have increased for both skilled and unskilled workers, for both coastal and inland areas, and for both exporting and nonexporting firms. We then benchmark wage growth to productivity growth using both national- and industry-level data, showing that Chinese labor was kept cheap until the late 1990s but the relative cost of labor has increased since then. Finally, we discuss the main forces that are pushing wages up. For example, the reforms in the late 1990s re-established a flexible labor market in China, enabling firms to pay workers according to productivity. China's labor force may have already reached its peak in 2011; and China's rural-to-urban migration will also slow down because the rural young are highly mobile; almost all rural youth in the 16–20 age bracket are already working off the farm (Rozelle, Huang, Zhang, and Li 2008). Therefore, future increases in migrant labor must come from those who are older or those who have established families, who will require the prospect of larger wage gains than migrants of the past if they are to find migration worthwhile.

Rising Wages in China

China's urban areas have a dual labor market: one for urban workers and the other for low-skilled migrant workers. Urban workers (both skilled and unskilled) have urban *hukou* (household registration), which means that they are registered with the government as living permanently in cities. Migrant workers have rural *hukou*, and they are very mobile in two aspects. First, they live temporarily in the places where they work during the year and return to their rural homes during the Chinese New Year holidays and the peak agricultural seasons (Wang and Zuo 1999). Although more migrant workers have chosen to stay in cities permanently in recent years, an average migrant worker spent 2.2 months in their rural homes and 9.8 months in off-farm work away from home in 2011 (2011 Survey of Off-farm Laborers by China's National Bureau of Statistics). Second, these workers often change their migration destinations and jobs. After the Chinese New Year holidays, which is a period of one to two months, they return to the cities to work, but they are highly likely to change jobs or migrate to a new city. They also change jobs and locations within a year, which makes tracking them in regular household surveys very difficult. With the difficulty of surveying this particular group, no large-scale survey data that cover all Chinese workers are available. However, China's National Bureau of Statistics does have good data for urban workers.

Figure 1
Real Annual Wages of Chinese Urban Workers
(deflated to 2010 prices)
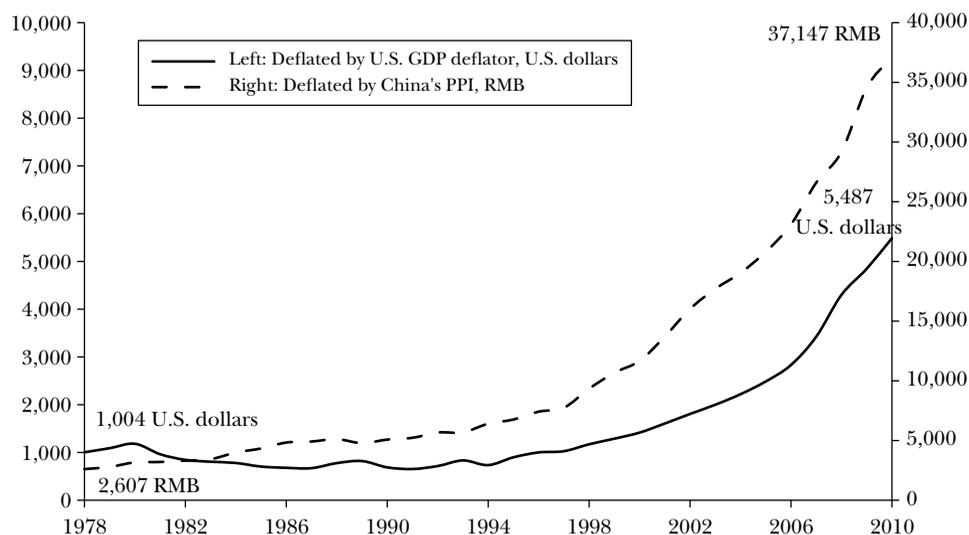
*Source: China Statistical Yearbooks.*
Note: PPI is producer price index.

In the first two decades of the reform period from 1978 up to the later part of the 1990s, the growth of workers' wages in Chinese urban areas was relatively low, as shown in Figure 1. According to the *Statistical Yearbooks* published by China's National Bureau of Statistics, the annual real wage of a Chinese urban worker increased only slightly from $1,004 in 1978 to $1,026 in 1997, at an average annual growth rate of only 0.1 percent (before tax, including pensions, and again converted from yuan to U.S. dollars using the current exchange rate, and to the 2010 level using the U.S. GDP deflator). This growth rate of China's urban wages is significantly lower than China's annual real growth rate of 4.0 percent (in real U.S. dollars) in these two decades.[1]

However, China's wage growth started to pick up steam in the late 1990s. In 1998, the real wage as measured in U.S. dollars grew by over 14.1 percent, marking the start of a new era of fast wage growth. In the period from 1998 to 2010, the average annual growth rate of real wages was 13.8 percent, exceeding the real GDP

---

[1] This seemingly low growth rate for China's economy might be surprising, but remember that, because we are interested in China's wage levels in the context of the world market, we are converting at the official yuan/dollar exchange rate—and that exchange rate was overvalued before the foreign exchange reform in 1994. However, the same qualitative pattern of GDP growing faster than wages holds true if the comparison is done in yuan. The growth rate of China's GDP as measured in real renminbi yuan over these years was 5.9 percent for wages and 9.9 percent for GDP.

Figure 2
Annual Manufacturing Wages of Asian Emerging Economies
(real U.S. dollars in 2010, deflated by the U.S. GDP deflator)
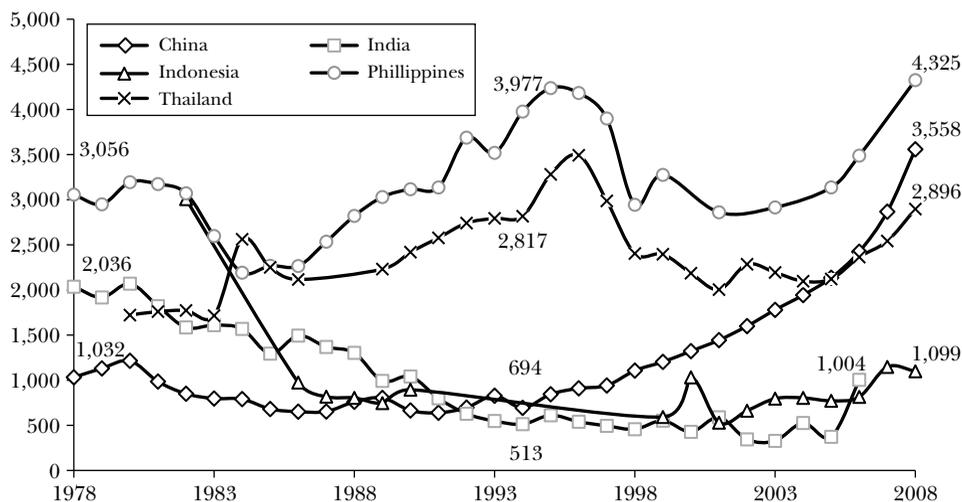

Source: International Labor Organization LABORSTA Database.

growth rate of 12.7 percent. The fast rise of China's urban wages since the late 1990s is due in part to institutional factors such as the privatization of state-owned enterprises in the mid-1990s, the re-establishment of the labor market, and the slowdown of labor force growth and migration, which we will discuss in detail below. As Figure 1 shows, the real wage in yuan grew faster than the real wage converted to dollars at the then-current yuan/dollar exchange rate because China's official exchange rate was overvalued before 1994.

China's wages have also increased compared with the wages of other developing economies. In Figure 2, we compare the manufacturing wages of a group of Asian developing countries. Note that China's manufacturing wage is lower than the overall wage in Figure 1. Among these countries, China had one of the lowest manufacturing wage rates in 1994 at $694 in U.S. dollars, or about 17 percent of the manufacturing wage in the Philippines. By 2008, the last year in which we have data for most of these countries, China's wages are second only to those of the Philippines, marking a wage gap of only 18 percent. The wages of two other populous Asian countries, India and Indonesia, are much lower at only about 41 and 34 percent of China's wages, respectively, in 2006, the last year in which data for India are available.

To explore whether the rise of China's wages is confined to certain sectors, we use micro-level data from the Urban Household Survey, which covers all urban areas in China and uses probabilistic sampling and a stratified, multistage method

Figure 3
Annual Wages of Urban Workers by Education Level
(real U.S. dollars in 2010, deflated by the U.S.GDP deflator)
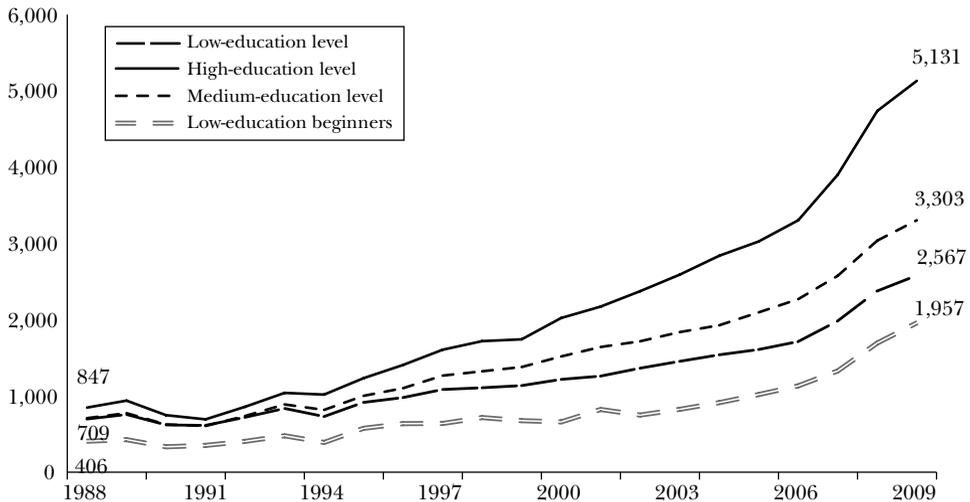
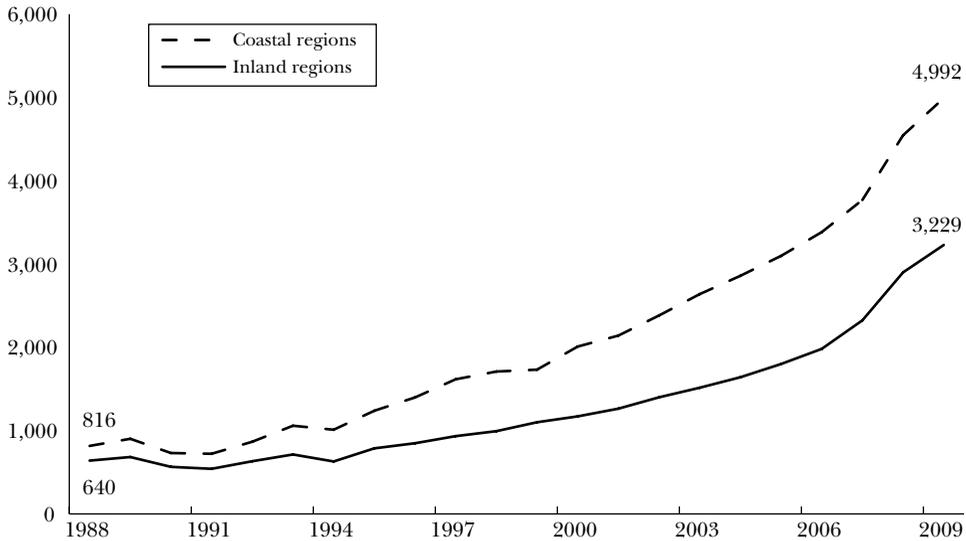*Source:* The Urban Household Survey data in 9 provinces, 1988–2009.
Notes: Education levels: "low" refers to junior high school and below, "medium" refers to academic/technical high school, and "high" refers to college and above. "Low-education beginners" are low-education workers with working experience less than 5 years.

to select households. To be included in the sample, a household must reside in a specific city for at least six months. Therefore, this sample does not include migrant workers who live in a city for less than six months or those who work off the farm in rural areas. The sampled households are asked to keep a detailed record of their incomes and expenditures every day. In the subsequent discussion, we will examine China's wages by exploring this dataset.[2] The wages reported in this study are lower than the aggregate statistics reported in Figures 1 and 2 because firms report aggregate wages, including pensions and taxes.

Wages are increasing for China's workers at all skill levels. Figure 3 shows that growth rates of real wages for those with low education level (junior high school and below), medium education level (academic and technical high school), and high education level (college and above) are all increasing at high speeds—at 6.5, 7.6, and 9.0 percent per year, respectively. Fast wage growth rates even for unskilled workers

[2] Our access to the Urban Household Survey covers the period from 1988 to 2009 for the nine provinces of Beijing, Liaoning, Zhejiang, Anhui, Hubei, Guangdong, Sichuan, Shaanxi, and Gansu. These provinces represent different regions and economic conditions. After excluding the unemployed individuals and individuals younger than 16 or older than 60, we are left with 321,311 individuals in the sample, with an average annual sample size of 14,605. The mean values and trends of most variables in our sample are comparable with those of the national sample.

Figure 4
Annual Wages of Urban Workers by Region
(real U.S. dollars in 2010, deflated by the U.S. GDP deflator)

Source: Wages by regions are from the Urban Household Survey data in nine provinces, 1988–2009.

suggest an overall rise in wages. To explore this phenomenon further, we examine the wages for low-education beginners, or low-education workers with less than five years of experience in the job market. Their wage growth was 7.8 percent per year in the 1988–2009 sample period and was actually higher at 9.8 percent annually from 1997 to 2009, which is a faster increase than for urban workers as a whole.

Wages are also rising in both the more-developed coastal regions and the less-developed inland regions despite the wage gap that exists between the two, as illustrated in Figure 4. The growth rate of wages in the inland regions was 7.7 percent per year from 1988 to 2009, which was one percentage point lower than the growth rate in the coastal regions. As a result, this regional wage gap increased to 54.6 percent in 2009. However, the inland regions have been catching up since 1997, with their annual wage growth rate being one percentage point higher than in coastal regions (10.9 versus 9.9 percent annually). Wages are also rising for both exporting and nonexporting firms. Interestingly, nonexporting firms have higher wages than exporting firms, similar to the findings of Lu (2010). However, the wage gap between these two types of firms is also declining over time.

The Urban Household Survey does not include the most mobile migrant workers, who may be the lowest-paid workers in China. To examine this issue, we use some aggregate statistics of migrant wages released by the National Bureau of Statistics. The annual real wages of migrant workers, who tend to have a junior high school degree

or lower education level, averaged $2,541 (in U.S. dollars) in 2009, is almost the same as the real wages of low-education workers in the Urban Household Survey sample in the same year ($2,567). The wages of the two samples corroborate each other, and their similarity suggests that the wages of low-skilled urban workers may track those of migrant workers. In terms of wage growth, the wages of migrant workers in the sample from the Rural Household Survey also increased at a high annual rate of 9.6 percent from 2003 to 2009. Furthermore, the wages of migrant workers increased even faster in the last two years, reaching $3,535 (in U.S. dollars) in 2011.

Wages versus Productivity

Although China's wages have been rising fast, if their growth rate is lower than that of labor productivity, then labor is effectively becoming cheaper per unit of product. So, to answer whether China's fast wage growth since 1997 implies that Chinese labor is becoming more expensive, we need to compare wage growth to labor productivity growth. In doing so, we find that Chinese labor becomes cheaper before the late 1990s, but not since then.

We calculate the growth of gross labor productivity by deducting the growth rate of the labor force from the growth rate of real GDP, drawing on data published in China's *Statistical Yearbooks*. In the period of 1982 (the first year in which we have data on labor force) to 1997, China's GDP (converted at the official exchange rate and deflated to real 2010 U.S. dollars) increased by 5.5 percent, and the labor force grew by 1.9 percent, implying labor productivity growth of 3.6 percent per year. This figure almost triples the real wage growth of 1.3 percent per year during that period, suggesting that Chinese labor was becoming cheaper relative to productivity during this period. Of course, this is a simple calculation that it does not adjust for changes in the quality and quantity of other inputs such as capital and the human capital of the labor force, but it is nonetheless revealing.

Chinese labor also became cheaper relative to other countries over this time period. Ceglowski and Golub (2007) find that manufacturing "unit labor costs"— the ratio of wages to labor productivity—fell for China relative to that of the United States from over 70 percent in the early 1980s to about 30 percent in the mid-1990s. Moreover, the relative cost of Chinese labor at this time was not only lower than that of developed countries but was also lower than that of developing countries such as India, Malaysia, and Mexico. It appears that wage growth fell far behind productivity growth in China during this period.

However, China's wages have increased at a much faster rate than productivity since 1997. Using our aggregate data, we find that China's GDP in real U.S. dollars increased by 12.7 percent annually in the period from 1997 to 2010, whereas labor force growth decreased to only 1.4 percent, implying an annual growth rate of 11.3 percent for gross labor productivity. Although this rate is much faster than that of the productivity growth before 1997, it is lower than the astonishing annual

Table 1

Characteristics of Selected Manufacturing Sectors in China

Industry	Employment (million) 2010 (1)	Export/ Sales ratio 2010 (2)	Capital/Labor ratio (1,000 US$) 2010 (3)	Value added per worker (1,000 US$) 2010 (4)	Wage/(value-added output per worker) 2007 (5)	Wage/(value-added output per worker) 2010 (6)
Leather, fur, and other textiles	2.76	0.30	20.88	11.60	0.28	0.39
Wearing apparel	4.47	0.28	23.22	11.04	0.31	0.41
Instruments	1.25	0.33	61.14	18.63	0.35	0.25
Electronics	7.73	0.63	72.10	19.32	0.12	0.24
Electrical machinery	6.04	0.19	77.53	23.01	0.22	0.20
Transport equipment	5.74	0.11	123.53	30.51	0.24	0.15
Basic iron and steel	3.46	0.03	196.51	55.13	0.20	0.08

Source: China Statistical Yearbook (various years).
Notes: All values are in 2010 real U.S. dollars. Industry classification follows the GB/T4754 standard released by China's National Statistics Bureau. The 1997 data include all enterprises in the specific industry, while the 2010 data only include enterprises above a designated size in an industry. "Instruments" refers to precision and optical instruments; "Electronics" refers to radio, television, and communication equipment and apparatus; "Electrical machinery" refers to electrical machinery and apparatus.

real wage growth of 13.8 percent in the same period. Therefore, Chinese labor is indeed becoming more expensive.[3] According to Ceglowski and Golub (2007), China's relative unit labor cost was 63 percent that of Malaysia and 70 percent that of Korea by 2002. If the gap between wages and productivity in China continues to close by 2.5 percent per year, China's advantage of lower labor cost relative to Korea will be completely eliminated by the year 2018, and that relative to Malaysia will be gone by 2022.

We draw on recent industry-level data to examine the unit labor cost: in this case, our measurement is the average wage as a proportion of value added per worker. In particular, we select several two-digit industries as cases from China's *Statistical Yearbooks*. As shown in column 1 of Table 1, these industries are large in employment size: they employ more than 1 million workers each and 31 million workers in total in 2010. Most of them are also major exporters, with 29 percent of their sales as exports on average (column 2). To facilitate analysis, we sort these firms by the capital/labor ratio (column 3), which can be viewed as a measure of labor (or capital) intensity or as revealing the level of technology.

---

[3] Ceglowski and Golub (2007) find that relative unit labor cost stabilized or increased slightly from 1995 until 2002, the last year covered by their study.

These variables suggest several interesting associations. First, labor-intensive industries are major exporters with very high export/sales ratios, consistent with the fact that China uses its abundant labor for export industries. Among the seven industries, basic iron and steel has the lowest labor intensity, and its sales are mainly domestic. Second, there is a clear negative association between the capital/labor ratio (column 3) and unit wage cost (column 6), which is defined as the ratio of wage to value added per worker, suggesting that labor-intensive industries have higher unit labor costs.

Finally, the last two columns suggest that labor-intensive industries experienced a sharp rise in unit labor costs. As an example, for the apparel industry, the unit labor cost increased from 0.31 in 1997 (column 5) to 0.41 in 2010 (column 6). The unit labor cost also increased in two other relatively low-tech industries: leather, fur, and other textiles; and electronics (which is mostly assembly). However, industries with higher capital/labor ratios like transport equipment and basic iron and steel experienced a decline in unit labor costs in the same period, suggesting that the labor cost advantage remains for the relatively high-tech industries.

Potential Reasons for Rising Wages

The discussion to this point has argued that China's real wage was more or less stagnant or stable in the 1980s and early 1990s, and grew substantially beginning in the late 1990s. In this section, we discuss three potential reasons for this change: institutional reforms, the disappearing "demographic dividend," and the slowing of rural–urban migration.

Institutional Reforms

China has transformed its economy from a planned labor allocation system to a more market-oriented labor market. In the planned system, workers were allocated by the central planner to the state-owned enterprises, and jobs were permanent with little mobility (Fleisher and Wang 2004). The central planner set the wages of all workers in the country using a simple system of grades, with the grade mainly depending on seniority. Wages were set low, and so was the wage gap between grades. In this system, wages did not reflect productivity, and because of this and the misallocation of workers, productivity was low.

The first major step in China's urban labor market reforms was to establish an internal pay incentive system within state-owned enterprises. Starting in the late 1980s, the financial insolvency of many state-owned enterprises prompted the Chinese government to undertake a series of reforms. The reforms started by allowing profitable firms to pay higher wages and even bonuses to the more productive workers, which increased the pay difference among workers (Park, Song, Zhang, and Zhao 2008). However, because private firms were still not allowed in these areas and job mobility was low, there was essentially no external labor market.

Figure 5
Private Employment as a Proportion of Total Urban Employment
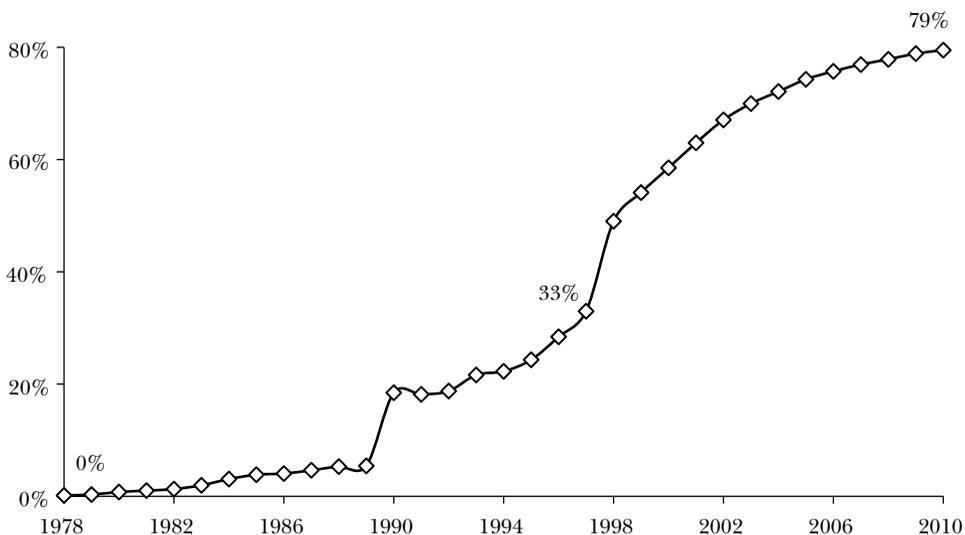
Source: China Statistical Yearbooks.

The second step in urban labor market reforms was to establish an external labor market. The most aggressive enterprise reforms took place in the mid-to-late 1990s, when China started to privatize state-owned enterprises and when the status of private firms was legalized (Cao, Qian, and Weingast 1999; Li 2003). These reforms were dramatic, with millions of state-owned workers being laid off and moving to jobs in the private sector. At the same time, the government started to allow large-scale migration of rural workers to cities (Cai and Wang 2010). Taken as a whole, these reforms established an external labor market that not only helped reallocate workers but also linked wages more closely to productivity (Zhang, Zhao, Park, and Song 2005). With these reforms, the private sector has become a prominent player in the labor market, with private sector employment as a proportion of total urban employment rising from literally nothing in the early 1980s to about 80 percent at present, as shown in Figure 5.

One consequence of the labor market reforms is the increase in the return to education, suggesting that the link between wage and productivity is becoming stronger. We replicate Zhang et al. (2005) in calculating a Mincer-style rate of return to education—that is, using wages as a dependent variable, and level of education and work experience as the key explanatory variables—but use the Urban Household Survey sample, which covers more provinces and a longer time series. The results of this calculation, reported in Figure 6, show that the return to an additional year of schooling is only 2.3 percent in 1988, but it increased to about 9 percent in 2000 and has been stable in the past decade. The return to an additional year of

Figure 6
Returns to Education in Urban China
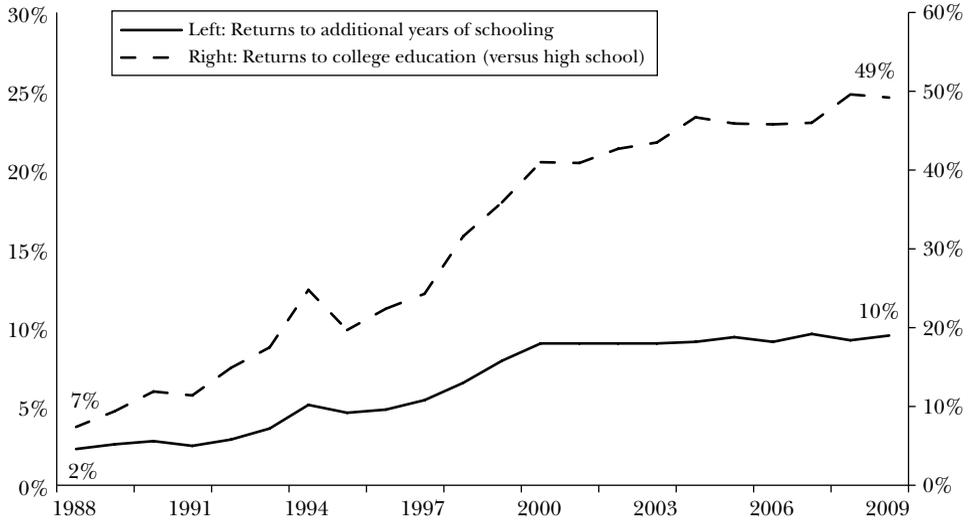
Source: Authors' estimations using data from the Urban Household Survey data in 9 provinces, 1988–2009.

education in the latest year of 2009 was 9.5 percent, similar to the world average rate of 9.7 percent, as reported by Psacharopoulos and Patrinos (2004). The return to college education in China increased the fastest, from only about 7.4 percent in the largely planned labor allocation system in 1988 to 49.2 percent in the much more flexible labor market system of 2009, exceeding the average 40 percent return in developed economies.

Demographic Transition and Labor Shortage

China has experienced a demographic transition from high to low birthrates since the 1970s. China had a baby boom from 1950 to 1978, with the total fertility rate averaging 5.2 births per woman, although a break occurred during the so-called "Great Leap Forward" from 1958 to 1961, during which around 30 million people died (Ashton, Hill, Piazza, and Zeitz 1984). China's total population increased from 552 million in 1950 to 963 million by 1978. In 1979, China started the "one-child policy," the largest and strictest population control policy in human history (Banister 1987). According to this policy, still largely in effect today, each woman is allowed to have only one child, and above-quota births are heavily fined. The one-child policy, together with other social and economic changes, has a significant impact on the fertility rate (Li and Zhang 2007; Li, Zhang, and Zhu 2011). China's total fertility rate fell sharply from 6 births per woman in 1970 to only 1.4 in 2010 (*The Economist* 2011).

Many countries have experienced lower birthrates as their economies develop, but the speed and magnitude of China's demographic transition are unprecedented in world history. The natural growth rate of China's population has decreased to an annual rate of 0.56 percent since 2001, similar to the population growth rate in Japan from 1980 to 1985. The United Kingdom took about 200 years (1750 to 1955) to complete its demographic transition to having low birthrates, and the United States took 140 years (1800 to 1940) to do the same (Livi-Bacci 1997; Greenwood and Seshadri 2002), while China's transition took only about 30 to 40 years.

This rapid shift to lower birthrates creates the "demographic dividend," that is, a situation in which a disproportionate share of the population is in its prime working years, with relatively few children or elderly. As Chinese baby boomers entered the labor market in the past three decades, China's labor force increased from 583 million in 1980 to about one billion in 2011. The sharply declining fertility of the baby boomers then leads to a low "young dependency ratio" or a large proportion of working-age people. Indeed, the working-age population aged 15–64 increased from 59.3 percent of China's population in 1980 to 74.4 percent in 2011. A large proportion in the work force and a low dependency ratio all tend to be accompanied by a high savings rate, abundant labor, and abundant working time for labor, which are beneficial for economic growth (Li and Zhang 2007).

However, the same demographic transition also means that China has entered a period when its labor force will increase much more slowly. Because of the low fertility of the baby boomers, China's labor force growth has been slowing down. As shown in the last section, coinciding with the faster wage growth after 1997, the growth rate of China's labor force dropped from 1.9 percent before 1997 to 1.4 percent since 1997, suggesting that the declining labor force growth could be one of the reasons behind the fast wage growth in the last decade.

Based on the "low variant" estimates by the United Nations (2011), China's population is expected to begin declining by 2015, and the labor force may have already peaked in 2011. According to these projections, which assume no upward shift of the birth rate, China's population will decrease from its current level of 1.34 billion to 1.13 billion by 2050, and the working population will decrease from one billion to 696 million. By then, China's labor force as a proportion of the population will drop to 62 percent, whereas the proportion of the elderly (aged 65 and over) will reach 29 percent. The dropping labor force proportion will cause labor shortages and help push wages up further.[4]

---

[4] The labor force as a proportion of the population is even lower for the medium and high variant estimates because of the assumed higher birth rates. The "medium variant" estimate of the population is 1.30 billion by 2050, with the labor force as a proportion of the population dropping to 61 percent and the proportion of the elderly reaching 26 percent. The "high variant" estimate of the population is 1.48 billion by 2050, with the labor force as a proportion of the population dropping to 60 percent and the proportion elderly reaching 22 percent.

China is already experiencing labor shortages. China's Ministry of Human Resources and Social Security carries out quarterly surveys that cover official job centers in 117 cities. These surveys show that the "position/seeker ratio," which is the ratio of newly created positions over the number of job seekers, has been trending upward from 0.65 at the start of 2001 to above 1.0 in the first quarter of 2010 and 1.08 in the first quarter of 2012. When the position/seeker ratio exceeded one for the first time in China, the number of newly created jobs exceeded the number of job seekers.

The labor shortage phenomenon is especially evident in China's coastal areas, such as the Pearl River Delta and the Yangtze River Delta areas. In the first quarter of 2008, the position/seeker ratio in the Pearl River Delta area reached a historical high of 1.89. Although the ratio dropped during the global financial crisis, it has rebounded with the recovering economy. In the first quarter of 2010, the ratios for the areas of the Pearl River Delta, Southeast Fujian, Yangtze River Delta, and Bohai increased to levels above 1, indicating a labor shortage in most of the coastal regions.

Can population growth increase again if China relaxes the one-child policy? This outcome seems unlikely. China is already relaxing the policy by allowing couples who are both an only child to have a second child. The first cohorts of the "one-child policy" children (born in the 1980s) are already in their childbearing age, but the birth rate has shown no sign of recovery so far. The one-child policy, together with many other social and economic reforms, may have caused a lasting change in preferences about fertility in China. There have also been calls for a total removal of all birth control policies, but the government seems hesitant to do so because of political resistance from both the multitude of birth control agencies and the people who were penalized by the one-child policy.

Slowing Down of Structural Changes

The migration of rural workers to urban areas in recent decades has helped keep the wages low in urban areas, but this pattern seems ripe for change as well. According to various *Statistical Yearbooks* of China, the number of migrant workers increased from 25 million in 1985 to 159 million in 2011. However, the growth rate of migrant workers is slowing down. Using 1997 as the breaking point, the growth rate declined from 10.8 percent per year before 1997 to only about 4.6 percent annually since then.

The major barriers to increasing migration from rural areas are a combination of the remaining institutions, such as the *hukou* system, and the fact that many of those who can migrate at the lowest cost are already doing so. Chinese households are managed by the *hukou* (household registration) system, which was established in the early 1950s to consolidate socialist governance, control domestic migration, and administer the planned economy. Every person is required to be registered at his/her place of birth and then acquire a *hukou* certificate that specifies rural/urban status and location from there. All administrative activities, such as land distribution,

Table 2
Proportion of Chinese Rural Residents Working Off-Farm

Age cohorts	1995	2004	2007
16–20	23.7	74.3	97.7
21–25	33.6	80.7	86.5
26–30	28.8	70.5	77.1
31–35	26.9	62.0	65.6
36–40	20.5	53.7	73.5
41–50	20.8	41.5	54.3

Sources: De Brauw, Huang, Rozelle, Zhang, and Zhang (2002), Rozelle, Huang, Zhang, and Li (2008).

issuance of identity cards, registration of a child in school, and medical insurance, are based on *hukou* status. Until the early 1990s, the *hukou* was also used to distribute food, cooking oil, and clothing coupons, thus restricting internal mobility in both urban and rural areas. Although the *hukou* system has been gradually relaxed since the mid-1990s, it still restricts migration in many ways. Farmers have been allowed to migrate to cities to work since the mid-1990s, but they cannot change their *hukou* status and thus cannot enjoy public services in the cities such as education, medical insurance, housing, and pensions. Unfortunately, most of these public services are of much lower quality or do not even exist in rural areas because most of the government spending is in urban areas. The huge gap in public welfare provision prevents the government from removing the *hukou* policy, as entrenched urban residents do not want to share their welfare benefits with the migrants, even though they need migrant workers to provide services. A recent proposal to allow migrant children to take the college entrance exams in cities has received heavy resistance from the local residents in Beijing and Shanghai.[5]

Various pieces of evidence suggest that the marginal cost of migration is rising, and the distance and need for migration are being shaped by the *hukou* system. According to de Brauw, Huang, Rozelle, Zhang, and Zhang (2002) and Rozelle, Huang, Zhang, and Li (2008), who conducted three rounds of surveys in China in 1995, 2004, and 2007, most young rural residents no longer worked in agriculture by 2007. The probability of doing off-farm work for all rural laborers increased from 31 percent in 1995 to 60 percent in 2007. The sharpest increase was observed among the youngest age group of 16–20, for which the probability of working off farm increased from 24 percent in 1995 to 98 percent in 2007, as shown in Table 2. Now, the youngest cohorts of laborers have almost no farmers. Even for older cohorts, the

---

[5] NetEase, a popular website, reported on September 6, 2012, that 95 percent of the local residents oppose the proposal. The Chinese link is at http://edu.163.com/12/0906/11/8ANETDNL00294JD0 .html, and a translation by Google is at http://translate.google.com.hk/translate?hl=zh-CN&sl=auto&tl =en&u=http%3A%2F%2Fedu.163.com%2F12%2F0906%2F11%2F8ANETDNL00294JD0.html.

probability of working off farm was very high by 2007: the probability was 87 percent for cohorts aged 21–25 and 77 percent for cohorts aged 26 to 30. As fewer young rural laborers are available, the marginal migrant worker will become older and have a higher marginal cost of migration. Indeed, the average age of migrants is increasing. According to a survey by China's National Bureau of Statistics, the average age of off-farm laborers increased from 34 years in 2008 to 36 years in 2011. The proportion of older off-farm laborers (aged 41 or above) increased from 30 to 38 percent in the same period. Thus, the potential supply is largely tapped out.

Migrants are staying closer to home, which helps reduce their migration costs. Again, according to the National Bureau of Statistics, the proportion of migrants working out of the province decreased from 53 percent in 2008 to 47 percent in 2011. The proportion of migrant workers working away from their home provinces decreased most significantly in China's poorest western areas, from 63 to 57 percent in the three-year period.

One reason why migrants can remain closer to home is that some factories and laborers are also moving from the coastal to the inland areas because of rising production costs such as labor and land costs, a phenomenon termed "flying geese" by Akamatsu (1962). For instance, Foxconn Technology Group is the largest contract supplier (assembling electronic products by Sony, Apple, Hewlett-Packard, Nokia, and other brand names) in China, employing over a million workers. This company moved its major plants away from Shenzhen to inland provinces such as Hebei and Henan (Luk 2010), mainly because of the hefty wage increases in Shenzhen.

However, the migration of firms and workers back to the inland areas is likely to be limited. In fact, Chinese manufacturing industries have been highly concentrated in the coastal areas since the central government established five special economic zones in the coastal areas in the 1980s (Wen 2004), and they have become even more geographically agglomerated in the last decade (Long and Zhang 2012). The agglomeration of industries in the coastal or border areas has occurred in other East Asian countries as well as in the rest of the world (Hanson 1996). One explanation for this phenomenon is the Krugman (1991) lock-in hypothesis, which argues that once industries are concentrated in the coastal regions for whatever reason, and then increasing returns to scale in production occur, more firms are encouraged to locate in these regions to benefit from the backward and forward linkages and from the external economies created. The reduction in China's trade barriers can also facilitate agglomeration in the coastal and border areas (Hanson 2001).

Conclusion

In the 1980s and through the late 1990s, Chinese workers were cheap in the sense that their labor cost relative to productivity was much lower than that of most other countries. Therefore, firms earned rents by sourcing in China, which triggered fast employment growth and rural–urban migration. But because of changes

since the late 1990s, including institutional reforms in China's labor markets and demographic transition that have reduced what used to be a huge amount of slack in labor supply, the "underpricing" of Chinese labor appears to be coming to an end. Wages are rising faster than labor productivity, particularly in labor-intensive exporting industries such as apparel and electronics, which will likely move out of China and probably go to countries like India and Vietnam.

China is becoming a middle-wage country. If the annual rate of wage growth of 13.8 percent over the past decade is maintained, the average real wage in urban China would reach $20,000 in U.S. dollars by 2020. As benchmarks for comparison, U.S. compensation for manufacturing workers reached $20,000 in 1980, Japanese compensation for manufacturing workers reached this level in 1986, and the annual wage in Korea reached $20,000 in 1995, according to the U.S. Bureau of Labor Statistics. Even if China's wage only grows at par with productivity at 11.3 percent per year, the Chinese average real wage will reach $20,000 by 2022. Similarly to other middle-wage countries, China needs to make a transition toward higher value-added industries, whether export or domestic.

Is China ready to move up the technological ladder? We believe the answer is "yes," for several reasons. First, China's aggregate labor productivity has been increasing at 11.3 percent per year for over a decade, partly because of manufacturing firms' heavy investment in research and development (with expenditures on research and development per worker increasing at an annual rate of 16.9 percent in the past 20 years) and capital deepening (total assets per worker in China increased to $94,240 in 2010). Second, human capital, at least measured in quantity, has risen dramatically. In 1999, the Chinese government started an aggressive "College Expansion" movement that increased the college entry class enrollment from 1.1 million in 1998 to 6.6 million in 2011. By 2050, Ma (2010) predicts that 40 percent of China's labor force can be expected to hold a college degree, similar to the level of Japan's labor force today.

In short, the end of cheap labor in China does not mean the end of Chinese economic growth. Rising productivity and education mean that China's comparative advantage is shifting. If China can improve the quality of education and develop institutions that help to foster innovation and entrepreneurship, then going forward, China may join a place alongside Korea and Japan (with a lag of two decades) as a formidable force in high value-added manufacturing and innovation.

References

Akamatsu, Kaname. 1962. "A Historical Pattern of Economic Growth in Developing Countries." *The Developing Economies* 1(1): 3–25.

Ashton, Basil, Kenneth Hill, Alan Piazza, and Robin Zeitz. 1984. "Famine in China, 1958–61." *Population and Development Review* 10(4): 613–45.

Banister, Judith. 1987. *China's Changing Population*, pp. 85–118. Stanford, California: Stanford University Press.

Bernstein, Aaron. 2004. "Shaking Up Trade Theory." *Business Week*, December 6. http://www.businessweek.com/magazine/content/04_49/b3911408.htm.

Cai, Fang, and Meiyan Wang. 2010. "Growth and Structural Changes in Employment in Transition China." *Journal of Comparative Economics* 38(1): 71–81.

Cao, Yuanzheng, Yingyi Qian, and Barry R. Weingast. 1999. "From Federalism, Chinese Style to Privatization, Chinese Style." *Economics of Transition* 7(1): 103–131.

Ceglowski, Janet, and Stephen Golub. 2007. "Just How Low Are China's Labour Costs?" *World Economy* 30(4): 597–617.

de Brauw, Alan, Jikun Huang, Scott Rozelle, Linxiu Zhang, and Yigang Zhang. 2002. "The Evolution of China's Rural Labor Markets during the Reforms." *Journal of Comparative Economics* 30(2): 329–53.

Economist, The. 2011. "China's Population: The Most Surprising Demographic Crisis: A New Census Raises Questions about the Future of China's One-Child Policy." May 5. http://www.economist.com/node/18651512.

Fleisher, Belton M., and Xiaojun Wang. 2004. "Skill Differentials, Return to Schooling, and Market Segmentation in a Transition Economy: The Case of Mainland China." *Journal of Development Economics* 73(2004): 315–28.

Greenwood, Jeremy, and Ananth Seshadri. 2002. "The U.S. Demographic Transition." *American Economic Review* 92(2): 153–59.

Hanson, Gordon H. 1996. "Localization Economies, Vertical Organization, and Trade." *American Economic Review* 86(5): 1266–78.

Hanson, Gordon H. 2001. "U.S.–Mexico Integration and Regional Economies: Evidence from Border-City Pairs." *Journal of Urban Economics* 50(2): 259–87.

Krugman, Paul. 1991. "History and Industry Location: The Case of the Manufacturing Belt." *American Economic Review* 81(2): 80–83.

Long, Cheryl, and Xiaobo Zhang. 2012. "Patterns of China's Industrialization: Concentration, Specialization, and Clustering." *China Economic Review* 23(3): 593–612.

Li, Hongbin. 2003. "Government's Budget Constraint, Competition, and Privatization: Evidence from China's Rural Industry." *Journal of Comparative Economics* 31(3): 486–502.

Li, Hongbin, and Junsen Zhang. 2007. "Do High Birth Rates Hamper Economic Growth?" *Review of Economics and Statistics* 89(1): 110–17.

Li, Hongbin, Junsen Zhang, and Yi Zhu. 2011. "The Effect of the One-Child Policy on Fertility in China: Identification Based on the Differences-in-Differences." Discussion Paper 00019, Chinese University of Hong Kong.

Lin, Justin Yifu, Fang Cai, and Zhou Li. 2003. *The China Miracle: Development Strategy and Economic Reform* (Revised Edition), pp.103–202. Hong Kong: The Chinese University Press.

Livi-Bacci, Massimo. 1997. *A Concise History of World Population*. Oxford: Blackwell.

Lu, Dan. 2010. "Exceptional Exporter Performance? Evidence from Chinese Manufacturing Firms." Unpublished paper, University of Chicago.

Luk, Lorraine. 2010. "Foxconn Relocates Production in China." *Wall Street Journal*, June 30. http://online.wsj.com/article/SB10001424052748703426004575338201918020996.html.

Ma, Li. 2010. "Trends and Outlooks of Chinese Labor Force (*Zhong Guo Lao Dong Li Bian Hua Qv Shi Ji Pan Duan*)." Unpublished report, presented at the 2010 Conference of Population and Development in China, May 18, Beijing.

Ministry of Human Resources and Social Security of the People's Republic of China. 2001–2012. *Quarterly Report on Labor Supply and Demand in Public Employment Service Organizations* (in Chinese). http://www.chinajob.gov.cn.

Park, Albert, Xiaoqing Song, Junsen Zhang, and Yaohui Zhao. 2008. "Rising Returns to Skill, Labor Market Transition, and the Growth of Wage Inequality in China." http://www.ceauk.org.uk/2008-conference-papers/wageineq-Park.doc.

Psacharopoulos, George, and Harry Anthony Patrinos. 2004. "Returns to Investment in Education: A Further Update." *Education Economics* 12(2): 111–34.

Rozelle, Scott D., Jikun Huang, Linxiu Zhang, and Xiaofei Li. 2008. "China's Labor Transition and Future Educational Challenges." Unpublished paper, Stanford University. http://scid.stanford.edu/group/siepr/cgi-bin/scid/?q=system/files/shared/Rozelle_10-20-08.pdf.

United Nations (Population Division). 2011. "*World Population Prospects: The 2010 Revision.*" http://esa.un.org/wpp/Documentation/publications.htm.

Wang, Feng, and Xuejin Zuo. 1999. "Inside China's Cities: Institutional Barriers and Opportunities for Urban Migrants." *American Economic Review* 89(2): 276–80.

Wen Mei. 2004. "Relocation and Agglomeration of Chinese Industry." *Journal of Development Economics* 73(1): 329–47.

Zhang, Junsen, Yaohui Zhao, Albert Park, and Xiaoqing Song. 2005. "Economic Returns to Schooling in Urban China, 1988 to 2001." *Journal of Comparative Economics* 33(4):730–52.

Labor Market Outcomes and Reforms in China[†]

Xin Meng

O ver the past few decades of economic reform, China's labor markets have been transformed to an increasingly market-driven system.
China has two segregated economies: the rural and urban. Understanding the shifting nature of this divide is probably the key to understanding the most important labor market reform issues of the last decades and the decades ahead. From 1949, when China's Communist Party rose to power, the Chinese economy allowed virtually no labor mobility between the rural and urban sectors. Rural–urban segregation was enforced by a household registration system called "*hukou*." Individuals born in rural areas receive "agriculture *hukou*" while those born in cities are designated as "nonagricultural *hukou*." For simplicity, the two groups are referred to as rural and urban *hukou*. During the first 30 years of the Communist regime, more than 80 percent of the population lived in the countryside.[1] The rationale for keeping most of the population on farms was based on the low levels of agricultural productivity and the need to ensure food provision for cities, which were deemed essential for industrialization (Perkins and Yusuf 1984; Meng 2000). In the countryside, employment and income were linked to the commune-based production system. Collectively owned communes provided very basic coverage for health, education, and pensions. In cities, state-assigned

■ *Xin Meng is a Professor, Research School of Economics, College of Business and Economics, Australian National University, Canberra, Australia. Her email address is xin.meng@anu .edu.au.*

[1] The only ways one could change *hukou* status before economic reform started were by 1) obtaining a tertiary degree; or 2) joining the People's Liberation Army and being promoted to first lieutenant or above. Recently some cities have allowed rural *hukou* holders to obtain urban *hukou* by purchasing housing in the city, but cases are rare.

life-time employment, centrally determined wages, and a cradle-to-grave social welfare system were implemented.

In the late 1970s China's economic reforms began, but the timing and pattern of the changes were quite different across rural and urban labor markets.

In rural areas, the early reforms of the late 1970s and early 1980s dramatically increased agricultural productivity and, by the mid-1980s, rural underemployment became a serious problem. At first, those living in rural areas were encouraged to set up rural township and village enterprises (TVEs) to absorb the surplus labor; this generated economic growth and employment for a time, but the effect soon reached its peak. During the 1980s, limited rural–urban migration began as a response to the demand for services in cities as well as demand for unskilled labor in the limited Special Economic Zones, where imports were duty-free and exports by foreign investors enjoyed significant tax concessions. Nevertheless, during the 1980s and into the early 1990s, city governments continued to push migrants back to the countryside and rural–urban migration was extremely restrictive (Wang and Wang 1995; Xiang 1996; Zhao 2000; West and Zhao 2000; Meng 2000). In the mid to late 1990s, economic growth in the cities began to accelerate and the demand for unskilled labor rose substantially. In particular, after China became a member of the World Trade Organization in November 2001, China's labor-intensive, export-led growth generated major demand for unskilled labor. Migration restrictions then relaxed considerably. Between 1990 and 1997, rural migrants working in cities increased slightly from 25 million to 37 million, but by 2009 the number of rural migrants to China's cities almost quadrupled to reach 145 million.

Urban economic reforms began later than rural reforms and proceeded at a slower pace. Before 1978, virtually all urban employment was in the state or collective sectors, with just 0.02 percent of China's urban *hukou* labor force being self-employed in 1978 (NBS 2010). Individuals were assigned to jobs; employers were not allowed to hire or fire; and wages were determined by the Central Planning Commission. Lifetime employment and centrally determined wages reduced mobility and incentives, which, in turn, led to overstaffing, shirking, and low productivity (Meng 2000).

Mild urban labor market reforms began in the 1980s, but labor mobility and incentives weren't much affected until two large events. Back in the early years of the Cultural Revolution in the late 1960s, tens of millions of urban high school–aged children had no school and no job, so Mao Zedong sent many of them to the countryside to work. Then, in the early 1980s, a majority of these "sent-down youth" returned to the cities, but few jobs were available, and for the first time under China's communist regime, the urban economy experienced large-scale open unemployment. The government responded by encouraging self-employment for the first time (Feng 2003). This was the first event.

The second was state-sector restructuring. China began this restructuring in the mid-1990s, at a time when more than 40 percent of state-owned enterprises were making losses. In 1997, the government introduced a new policy—"Hold on to the Large, Let Go of the Small"—which aimed at maintaining the largest

1,000 state-owned enterprises and pushing smaller ones into the marketplace to compete or go bankrupt. Official statistics on the state sector share of China's whole economy are not available, but the state/collective sector share in industrial output value fell from over 90 percent in 1990 to 70 percent by 1997 and to 30 percent in 2008 (National Bureau of Statistics (NBS) 2009).

Within three to four years of the restructuring, tens of millions of state sector urban workers were made redundant and the urban private sector began to expand quickly (Meng 1997; Fan 2000; Appleton, Knight, Song, and Xia 2002; Meng 2004; and Giles, Park, and Cai 2006). The labor market for those with urban *hukou* began a transformation.

This paper focuses on employment and wages in the urban labor markets, the interaction between the urban and rural labor markets through migration, and future labor market challenges. The next section looks at the evolution of urban labor market outcomes in China and discusses remaining tensions within this market. The following section discusses the most important change in China's labor market over the past two decades: large-scale rural-to-urban migration. I then discuss future changes in the quantity and quality of the Chinese labor force and the challenges that these pose for sustaining China's future economic growth. One main thesis is that despite the remarkable changes that have occurred, inherited institutional impediments still play an important role in the allocation of labor. The *hukou* system continues to restrict labor mobility, though to a lesser extent, and 72 percent of China's population is still identified as rural *hukou* holders (according to data from the NBS *Comprehensive Statistical Data and Material on 50 Years of New China* (1999) and the 1% Population Survey of 2005). I will argue that China must continue to ease its restrictions on rural–urban migration, and must adopt policies to close the widening rural–urban gap in education, or it risks suffering both a shortage of workers in the growing urban areas and a deepening urban–rural economic divide.

The principal data sources on labor markets used in this paper are from the Urban Household Survey (UHS) conducted annually by China's National Bureau of Statistics for the years 1988 to 2009 and Rural Urban Migration in China and Indonesia (RUMiCI) Project data for the years 2008 to 2010. I also use the 1 percent sample of the Population Census of 2000; the 20 percent sample of the 1% Population Survey of 2005; and aggregated data from *China Statistical Yearbooks*. For a brief overview of these and other sources of labor market data for China, see Appendix A, available with this paper at http://e-jep.org.

Urban Labor Market Reform

Employment and Unemployment

Communist orthodoxy holds that all people of workforce age should contribute to the common good by being employed, and in pre-reform China almost all able-bodied adults were employed. Economic reform changed incentive systems and encouraged labor mobility, allowed workers to choose where and whether they

wanted to work, and allowed firms to choose who to hire and fire. This flexibility also gave rise to the possibility of unemployment (Meng 2000; Cai, Park, and Zhao 2009; Maurer-Fazio, Connelly, Chen, and Tan 2010).

The employment-to-population ratio for urban *hukou* population can illustrate some effects of these changes. In 1988, the employment rate for urban *hukou* holders, aged 16 to 64, was 83 percent for men and 75 percent for women. The averages for OECD countries in 1988 were significantly below these levels at 77.3 percent for men and 52.4 percent for women (OECD Labor Force Statistics, various years).[2] But by 2002, the employment rates had fallen to 75 percent for Chinese men and 59 percent for women. Since 2002, the employment rate for men has hovered around the 2002 levels, while for women it fell further but at a slower pace than during the previous period, reaching 57 percent for women by 2009— much closer to OECD levels. The timing of the significant employment reduction coincided closely with state-sector restructuring.

The unemployment rates derived from data from China's Urban Household Survey lie between 3 and 5 percent for both urban men and women from the late 1980s until 1998. By 2002, the unemployment rate rises to about 6 percent for men and 11 percent for women, and remains there through 2009. The timing of this rise in unemployment is probably not captured accurately by these estimates because the significant restructuring of the state sector starting in the mid-1990s involved large-scale job losses well before 2002 (Appleton et al. 2002; Giles et al. 2006; Cai et al. 2009). One reason why *reported* urban unemployment did not rise sooner is that in the beginning of the state-sector restructuring, most laid-off workers received support from their original state enterprises, and many regarded their layoff as temporary. Thus, when the Urban Household Survey was conducted during those years, many might not have reported themselves as being unemployed.[3] China's first Unemployment Insurance Act was issued in 1999 (State Council of the Peoples

[2] The data used in this subsection mainly come from the Urban Household Survey (UHS). I use data for 16 of 31 provinces. Using the full sample for the 1988 to 2001 period does not change the results presented in the paper. The definition of employment and unemployment used here differ somewhat from the Western standard. For employed people, the survey asks for current occupation as well as annual earnings. People who are not working need to provide reasons from the following choices: 1) unemployed; 2) waiting to be assigned to a job; 3) disabled; 4) retired; 5) students; 6) waiting for further education; 7) housekeeping; and 8) not otherwise specified. The employed are defined as those who are currently working with positive annual earnings, while the unemployed include categories 1 and 2 in the "not working" choice set, as well as those who reported as working but without positive annual earnings. The latter group account for 0.4–1.4 percent of the total labor force across different years.

[3] Two alternative data sources—the China Income Project Surveys and the China Urban Labor Survey, both with more accurate definitions of unemployment—find much higher unemployment rates in the mid and late 1990s than the official figures. Indeed, the China Income Project Survey asked two questions in 1995 regarding unemployment. When individuals were asked their current labor force status, which is the same question in the official Urban Household Survey, only 3.2 percent reported being unemployed. However, in a different place the workers were first asked whether their enterprise was making losses and then were asked whether they were unemployed. With this sequence of questions, an additional 7.7 percent of people who answered as being employed in the first place now reported as being unemployed. This finding suggests that total unemployment was much higher in the mid to late 1990s than the UHS data suggest.

Republic of China 1999), but a centralized unemployment support system was not formally established until the early 2000s. Since then, laid off workers collect their unemployment payment from central offices—and this is the period when the *surveyed* unemployment statistics rise.

It is important to note that during this period of rising urban unemployment more than 100 million rural *hukou* workers moved to cities to work and their unemployment rate is extremely low. Obviously jobs were available in cities and are growing quickly. However, the jobs rural migrants take are normally regarded as 3D (Dirty, Dangerous, and Demeaning) jobs, and most urban *hukou* workers are unwilling to accept them. Thus, urban unemployment during this period may be regarded as "voluntary." Nevertheless, the differences between average skill levels (measured in formal education) of the unemployed urban workers and employed rural-migrant workers in cities is large. For example, in 2009, 22 percent of unemployed urban workers had three-year college or above education and 63 percent had senior high school or above education, whereas in the same year, the proportion of employed migrant workers with college or above education was 5.7 percent and the proportion with senior high school or above education was 33 percent. This difference suggests that the unemployment of the urban *hukou* workers during this period is also "structural." There exists a mismatch between skills and available jobs.

To gain more insight into employment and unemployment patterns, it is useful to look more closely at the effects by age, gender, and ownership structure. Younger age groups have experienced large employment reductions. For example, in 1988, 32 and 87 percent of 18 and 22 year-old men (senior high school and university graduation ages), respectively, were employed. By 2002, these percentages had fallen to 0.6 and 40 percent, respectively; by 2009, the percentages had fallen to 0.5 and 31 percent for 18 and 22 year-old men, respectively. The main factor behind these employment declines has been rapid expansion of education. Since the late 1990s, China has been expanding its tertiary (three-year college and four-year university) enrollment by roughly 40 percent per year. In 1988 around 30 percent of 16 to 24 year-olds were at school or waiting to continue their schooling; by 2009, the ratio had increased to 70 percent. China's rise in tertiary enrollment was not driven by the demand for highly skilled workers, but instead was a way to respond to sluggish aggregate demand and slow employment growth in the late 1990s (Meng, Shen, and Xue 2012). In addition, the sharp increase in college-university enrollment may have reduced the quality of the education. The issue of unemployment among new college-university graduates has become a hot issue attracting considerable attention in Chinese and in Western media (Park, Cai, and Du 2010).

Older workers have also seen a decline in employment. For men, this decline mainly occurred in their 50s, with the employment/population ratio of men aged 55 to 65 falling from 62 percent in 1988 to 48 percent in 2009. For women, the change occurred mainly in their late 40s. Fifty-three percent of women aged 45–65 years were employed in 1988, and by 2009 the ratio dropped to 39 percent. The unemployment rates for these groups are quite low, though. This change in employment among older workers probably resulted from workers who lost jobs

because of economic restructuring and then became discouraged and left the labor force (Appleton, Knight, Song, and Xia 2002; Giles, Park, and Cai 2006; Maurer-Fazio, Connelly, Chen, and Tang 2010).

Women on average have seen a larger decline in employment rates and a larger rise in unemployment than men. While these effects have been larger at younger and older ages, they have occurred across the board. For prime-aged women, age 30–40, for example, 98 percent were working in 1988, but by 2002, the employment/population ratio for the group fell to about 85 percent and more or less remained there to 2009. The unemployment rate for prime-aged women has stayed at near 10 percent throughout the 2000s.

Some of the reduction in female employment is related to an increase in household income and housing availability, both of which have reduced co-residing with parents, which, in turn, has increased women's responsibilities for their nuclear families and hence reduced their labor supply (Maurer-Fazio, Connelly, Chen, and Tang 2010). Women now also have more discretion to choose whether to work, and the social stigma associated with not working has gradually diminished. Lack of job flexibility in the workplace may also be important. Part-time employment is rare in China, and the proportion of employed women working less than 35 hours per week has hovered in the range of 6.5–7.2 percent from the mid-1990s into the late 2000s (China Income Project Surveys (CHIPs) 1995, 2002, and Rural Urban Migration in China and Indonesia (RUMiCI) urban sample 2008). However, the reasons why unemployment for prime-aged women has stayed at near 10 percent throughout the 2000s have not been sufficiently examined in the existing literature.

The change in ownership structure has also affected employment patterns. As late as 1991, more than 97 percent of urban *hukou* workers were still state sector employees.[4] In the mid to late 1990s, the share of workers in state employment started to decline, falling from 93 percent in 1995 to 82 percent in 2001, and falling further to 50 percent by 2008–09.

How do China's state and nonstate sectors differ? Table 1 compares the state and nonstate shares of employment in China's urban economy. In 2008, the state and collective sectors together employed 56.8 percent of workers (with the collective sector accounting for 6 percentage points). Domestic privately owned workplaces employ 35 percent of total workers, and foreign-owned companies employ 4.5 percent. If we look only at production-related employment—that is, excluding public servants and people working for not-for-profit institutions—the state sector hires 30 percent of the workforce while domestic private firms hire almost 49 percent.

---

[4] The Urban Household Survey categorizes ownership into five groups: 1) state, 2) collective, 3) private (including self-employment), 4) other ownership, including joint venture, foreign-owned, and shareholding companies, and 5) those who are not otherwise specified, including domestic maids, childcare workers, and others. The categories 1 and 2 are grouped as the "state"; 3 and 4 as "private"; and 5 as "other."

Table 1
State versus Nonstate Employment

	Number of workers	% of total	% of production firm workers	Average number of workers in the workplace
Government/party/state not-for-profit institution	2,011	31.66		636
Private not-for-profit institution	295	4.64		386
Production firms:				
State firms	1,218	19.18	30.10	1,472
Collective firms	318	6.00	9.42	507
Domestic private firms/ self-employment	1,978	31.14	48.89	143
Foreign-owned firms	285	4.49	7.04	740
Other type firms	184	2.90	4.55	102

Distribution of each ownership category among different industries

	Manuf., mining, agri/fish, construction, transport	Utility, water, environment	Communication, IT, financial, real estate, science/tech	Retail/ wholesale trade and services	Education, health, social welfare, culture, sports, public service
Government/party/state not-for-profit institution	19.57	6.08	10.26	13.60	50.50
Private not-for-profit institution	36.61	2.37	13.56	34.24	13.22
State firms	56.30	8.89	15.47	14.90	4.44
Collective firms	41.73	3.94	16.27	32.28	5.77
Domestic private firms/ self-employment	29.89	2.18	10.67	53.57	3.69
Foreign-owned firms	44.56	2.81	20.70	28.77	3.16
Other type firms	20.11	4.89	9.24	48.37	17.39
Total	33.08	4.92	12.34	30.07	19.59

Distribution of each industrial sector among different ownership categories

	Manuf., mining, agri/fish, construction, transport	Utility, water, environment	Communication, IT, financial, real estate, science/tech	Retail/ wholesale trade and services	Education, health, social welfare, culture, sports, public service
Government/party/state not-for-profit institution	18.72	39.10	26.31	14.31	81.58
Private not-for-profit institution	5.15	2.24	5.11	5.29	3.14
State firms	32.59	34.62	24.01	9.49	4.34
Collective firms	7.58	4.81	7.92	6.45	1.77
Domestic private firms/ self-employment	28.16	13.78	26.95	55.50	5.87
Foreign-owned firms	6.05	2.56	7.54	4.30	0.72
Other type firms	1.76	2.88	2.17	4.66	2.57

Source: RUMiCI Migrant and Urban Surveys, 2008.
Note: Table 1 compares the state and nonstate shares of employment in China's urban economy.

The state has held on to the large firms with high capital intensity and operated mainly in the production sector rather than the service sector, where the private firms are smaller and less capital intensive. Firms in the state sector average about 1,472 employees; foreign-owned firms average 740 people; and the domestically owned private sector firms average 142 people. By industry, 56 percent of workers in the state sector firms work in manufacturing, mining, construction, or transport industry, while 53 percent of the workers in the domestic private sector work in retail/wholesale trade, hotel/restaurants, rental/commercial services, or household services industries. This difference helps to explain why the state sector also has much higher capital intensity, as measured by fixed assets per employee: 59 percent higher than the private sector in 1998; 184 percent higher, in 2003; and 283 percent higher, in 2007 (OECD Labor Force Statistics 2010). The state sector is less productive than the private sector, although it has improved in the 2000s (OECD Labor Force Statistics 2010).

Who lost state sector jobs? I estimated a probit model for each year between 1988 and 2009 using the sample of employed individuals to answer this question. The dependent variable is whether the worker has a job with the state sector (= 1) or the nonstate sector (= 0). The independent variables are age and its squared term, gender, education level, and dummy variables for each province. The detailed regression results are available in Appendix B, Table B1, available with this paper at http://e-jep.org. Figure 1 shows the marginal effects by year and indicates that in the early years, when almost everyone was in the state sector, there was no effect of education, age, or gender. However, after the state-sector reforms, those with higher levels of education are more likely to end up in the state sector (Figure 1A), females are significantly less likely to work in the state sector (Figure 1B), and older workers are more likely to be working in the state sector (Figure 1C). At the end of the period, 62 percent of 50 year-olds were employed in the state sector, but only 20 percent of 20 year-olds. Thus, the general pattern of the state-sector reforms has been to shift those who are younger, less educated, and female out of the state to the private sectors (or has limited them from entering the state sector).

Wage Structure and Inequality

Between 1988 and 2009, real annual earnings for urban *hukou* workers increased from 3,880 yuan to 19,674 yuan, a fivefold increase (Urban Household Survey data, excluding the top and bottom 1 percent of observations). This dramatic increase in real earnings was accompanied by an equally dramatic change in wage structure.

In China's pre-reform era, there were two separate wage ranks: one for production workers and one for managerial and professional workers. Apart from slight regional variations, such as heating subsidies, the whole nation followed the same wage system (Meng 2000; Huang 2004). This system offered a low return to education, but a high return for additional years of experience that did not taper off at higher ages. Experience, education, and occupation variables explained a far larger proportion of the variation in individual-level earnings than in Western country

Marginal Effects of Individual Characteristics on Probability of State (versus Nonstate) Sector Employment

A: Education

B: Dummy for females

C: Age effect relative to age 35
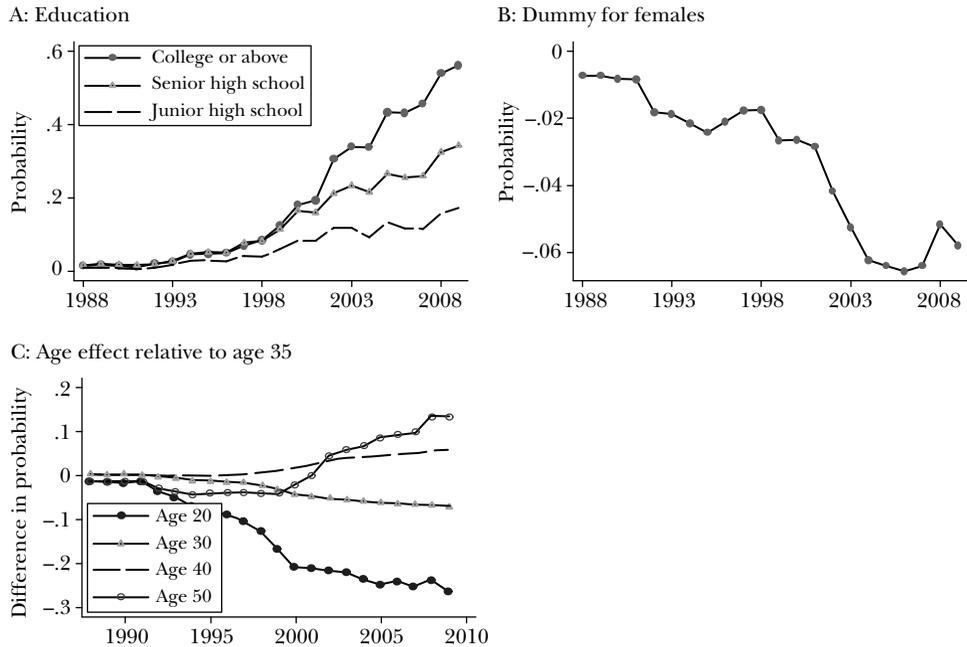
Notes: Based on the estimation of a probit model for each year between 1988 and 2009 using the sample of employed individuals. The dependent variable is whether the worker has a job with the state sector ($= 1$) or the nonstate sector ($= 0$). See the text for details.
Source: Author's own estimation results based on data from the 1988–2009 Urban Household Survey conducted by China's National Bureau of Statistics.

earnings equations, reflecting the dominance of the administratively determined wage structure (Meng and Kidd 1997). As the administrative wage system weakened, the returns to education have risen significantly, while the returns to experience have fallen (Appleton, Song, and Xia 2005; Zhang, Zhao, Park, and Song 2005).

Figure 2 presents the changing effect of observable characteristics on the urban wage structure between 1988 and 2009. The figures are based on regressions in which (log) real annual earnings are the dependent variable. The independent variables are three categories of education (three-year college and above, senior high school, and junior high school, with the omitted category as primary school); nine categories of work experience (0–4, 5–9, . . . ≥40); a dummy variable indicating employment in the state sector; seven occupational dummy variables; a female dummy variable; and dummy variables for each province. These regressions are run on data for each year, and then the coefficients for each year are presented in four panels. (Again, detailed regression results on which this figure is based are presented in Appendix B, Table B2, available with this paper at http://e-jep.org.)

Figure 2
Effect of Characteristics on Earnings

A: Education

B: Females

C: State-collective employment

D: Work experience
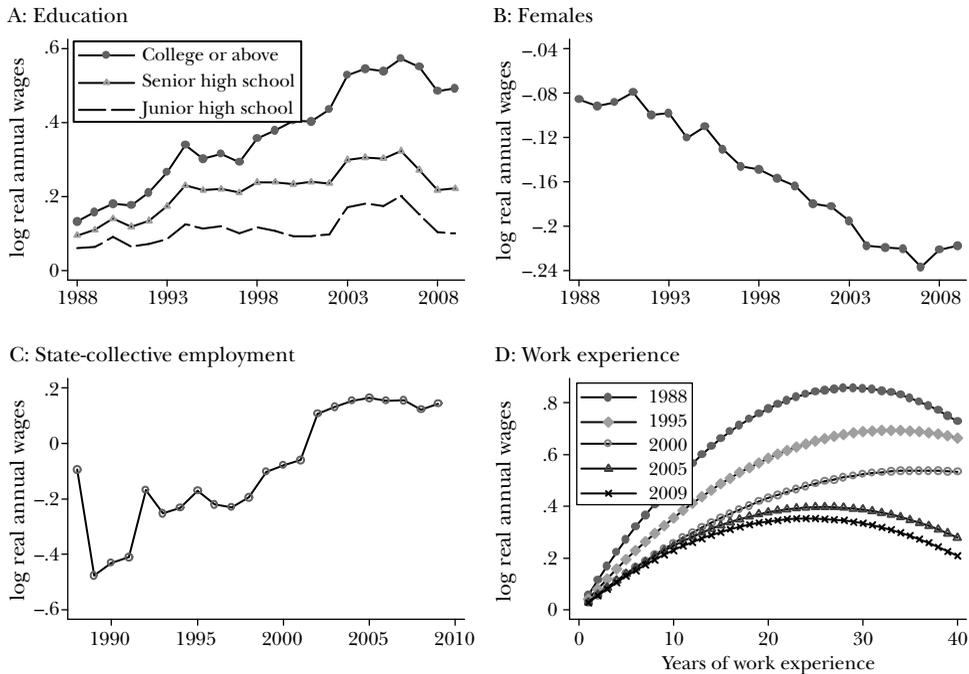
Notes: Figure 2 presents the changing effect of observable characteristics on the urban wage structure between 1988 and 2009. The figures are based on regressions in which (log) real annual earnings are the dependent variable. See text for details.
Source: Author's own estimation results based on data from the 1988–2009 Urban Household Survey conducted by China's National Bureau of Statistics.

Figure 2A shows that returns to college-and-above education have risen from around 16 percent in the late 1980s to over 50 percent by 2003, but since then returns have slipped back slightly, which may be related to the large influx of graduates due to the 1999 university expansion and an associated decline in quality. Figure 2B shows the gender earnings differential has been widening. Women used to earn around 8 percent less than men; by 2009, the gap had widened to around 23 percent. Figure 2C shows that the state-collective sector paid significantly less than the private sector in the 1990s, but this pattern has reversed in the 2000s.[5] In 2002, as part of a fight against corruption, China introduced a public servant examination system and increased public sector pay substantially. Since then, state sector pay has been about 20 percent higher than the private sector, along

---

[5] If the collective sector is not combined with the state sector, the state sector (alone) dummy variable always has a positive and significant coefficient, and the magnitude is always above 10 percent over the entire period.

Figure 3
The Adjusted R^2 from Separate Earnings Regressions
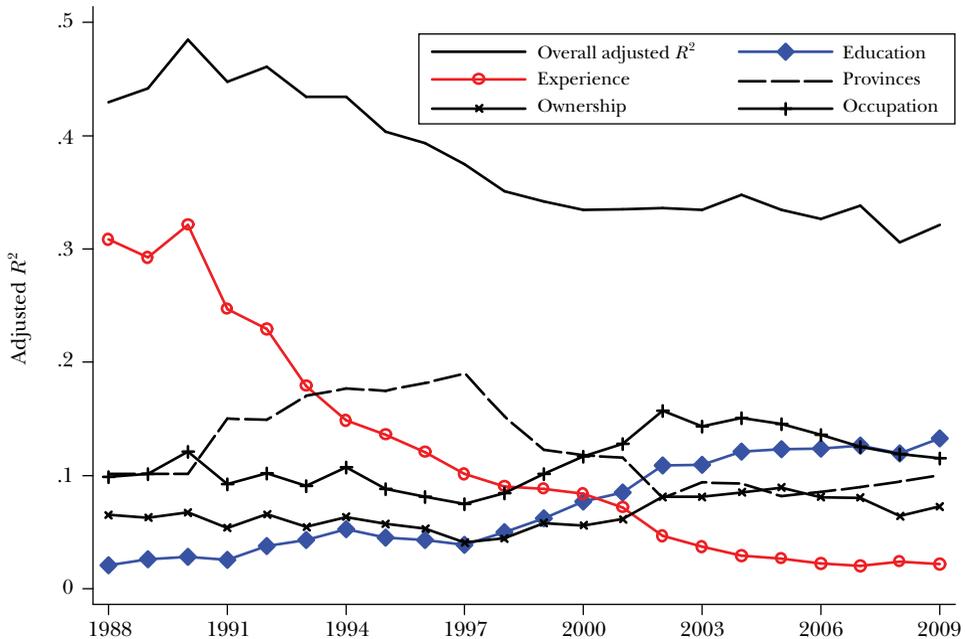
Note: Figure 3 is based on a regression of the log of real annual earnings on each of the explanatory characteristics alone, and shows the share of the wage variance explained by that factor alone, as measured by adjusted R^2.
Source: Author's own estimation results based on data from the 1988–2009 Urban Household Survey conducted by China's National Bureau of Statistics.

with offering significantly higher social insurance and other benefits. Figure 2D shows that experience–earnings profiles have continued to flatten over the period, although more slowly in recent years.

As the wage structure changes, so does the relative importance of different contributing factors. To illustrate, I regressed the log of real annual earnings on each of the explanatory characteristics alone, and then look at the share of the wage variance explained by that factor alone as measured by the adjusted R^2.[6] Figure 3 shows the results. In 1988, work experience alone explained 30 percent of wage variation, but by the end of the data period it explained only around

---

[6] This approach follows the lead of Dickens and Katz (1987). The idea is to derive a bounded range for the contribution of each characteristic to wage determination. The lower bound is derived by examining the increase in explanatory power by adding each of the characteristics into a regression that already includes the other set of regressors. The upper bound of the range is found by regressing log real annual earnings on each of the characteristics alone. The upper bound is illustrated in Figure 3. The pattern for the lower bound adjusted R^2s is very similar to that presented in Figure 3. It is available in the online Appendix B with this paper at http://e-jep.org.

3 percent. Education, on the other hand, explained only 2 percent of the variation in wages in 1988, but 13.3 percent of total variation in 2009. Occupation and ownership of employment, each taken alone, explain a slightly higher share of variation in wages from the mid-1990s forward.

Interestingly, the dummy variables on provinces, taken alone, explain more of the variation in wages than any other factor for several years in the mid-1990s, but have now shifted back to a similar level of contribution to that observed during the early reform era. China has always had significant regional price and wage variations, particularly in the 1990s when regional protectionism was at its peak (Young 2000; Jiang and Li 2005; Fan and Wei 2006; Brandt and Holz 2006; Gong and Meng 2009).

Finally, if one takes all the variables together, the proportion of wage variation that they can explain has fallen steadily over time (see the solid line in Figure 3). At the beginning of the period, a regression that includes all the independent variables explained about 43 percent of the total variance, while by the end of the period the adjusted R^2 had fallen to 30 percent, similar to the amount of wage variation that can be explained by individual characteristics in most Western countries. This change is to be expected when shifting from a national administrative system to a system in which wages are set in a more market-oriented environment.

Another aspect of these wage changes is that inequality has increased dramatically in the urban population. Between 1988 and 2009, the Gini coefficient for annual wages increased from 0.26 to 0.38.[7] The most significant increase in earnings inequality occurred during the 1990s, when state-sector restructuring was prominent, regional earnings variation was at its peak, returns to experience were falling, and returns to education began to increase significantly. In the 2000s, when returns to education stopped growing and regional earnings variation stabilized, the increase in earnings inequality ceased. The substantial increase in inequality during the 1990s was mainly driven by the disproportional increase in the earnings of the top of the distribution (Park, Song, Zhang, and Zhao 2004; Li, Zhao, and Lu 2007; Meng, Shen, and Xue 2012).[8] Figure 4 shows that the 90th to the 10th decile ratio increased from 3 to 6; in contrast, the 50th to 10th decile ratio increases only modestly.

Studies on wage inequality among urban *hukou* workers exclude the rural-to-urban migrants who live in urban areas. Using RUMiCI data, the Gini coefficient of

[7] The Gini coefficient is a measure of inequality ranging from 0 to 1, where zero implies everybody has the same income and one indicates that one person in the society has all of the income. The Gini coefficients presented here are for urban *hukou* workers only. OECD (2011) indicates that the Gini coefficient for household income for China as a whole is 0.33 in 1993 and 0.41 in 2008, whereas for OECD countries as a whole, the Gini coefficients in the same period changed from 0.30 to 0.31.

[8] Li, Zhao, and Lu (2007) find that education was an equalizing force during the period 1988 to 1997, but became a driving force for earnings inequality between 1997 and 2003. In Meng, Shen, and Xue (2012), my coauthors and I decompose earnings inequality between 1988 and 2009 and find that the dominant factor driving the significant increase in the earnings variance in the 1990s was an increase in the price of unobserved skills. As an economy shifts from an administratively determined wage system to a market-oriented one, rewards to both observed and unobserved skills increase. In the mid-2000s, because the expansion in college attendance that started in 1999 has increased the supply of educated workers sharply, the increase in returns to both observed and unobserved skills has plateaued.

Figure 4
Urban Wage Inequality
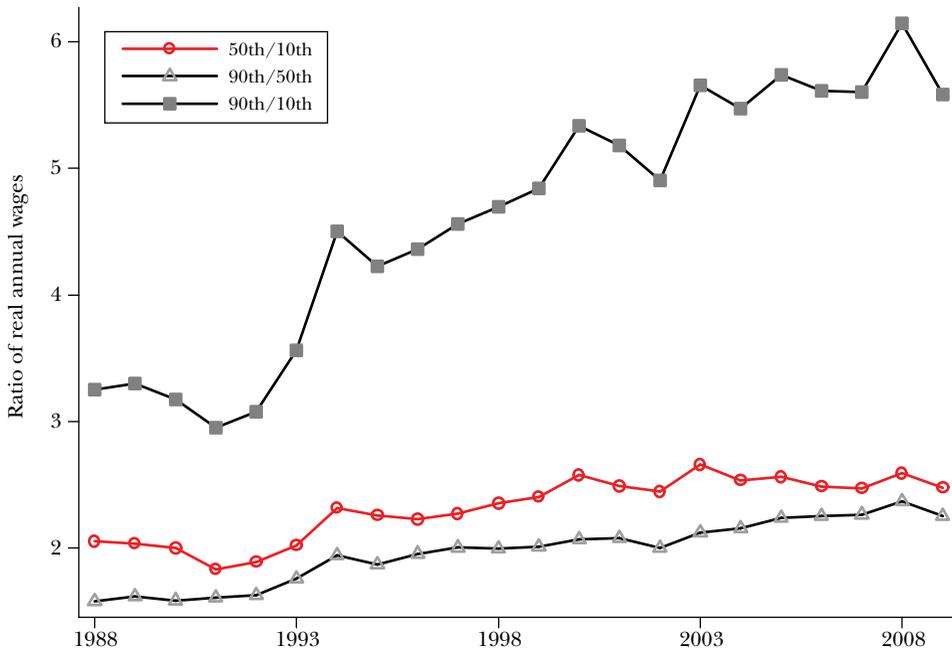
Note: Figure 4 shows the evolution of the wage structure in terms of ratios of real annual wages for the 50th to 10th deciles, the 90th to 50th deciles, and the 90th to 10th deciles.
Source: Author's own estimation results based on data from the 1988–2009 Urban Household Survey conducted by China's National Bureau of Statistics.

monthly earnings in 2008 for urban workers alone is 0.38, while for the predominantly low-wage migrant workers it was just 0.23. Combining migrants with urban workers gives an overall Gini coefficient of 0.34. Although individual earnings inequality overall is reduced when migrant workers are added to the sample, inequality between the two groups is large and the sparse evidence available suggests that it is increasing. It is also important to note that earnings inequality discussed here does not include "gray" income—due to underreporting of income at higher income levels. Using data from a special survey, Wang and Woo (2011) found that without including gray income, the income of the highest 10 percentile is 23 times that of the lowest 10 percentile. If "gray" income is included, the multiple increases to 65.

Rural–Urban Migration

China's single most important labor market change over the past two decades is probably the rapid growth in rural-to-urban migration. From the late 1990s up

to the present, the number of rural migrants increased by more than 100 million to 145 million. In the next few decades, more than 300 million rural *hukou* workers may move to cities to work. The world has never seen such a large-scale human movement within such a short time. This large-scale movement of workers from the low-productivity agriculture sector to the high-productivity urban sector is one of the forces driving China's unprecedented economic growth (Bosworth and Collins 2008; Gong, Kong, Li, and Meng 2008).

However, rural–urban migration in China remains restricted. China follows a "guest worker" system with controls over the type of jobs rural migrants are allowed to have and the social welfare and social services to which migrants are entitled. China's rural migrants often take jobs which urban workers are unwilling to take (Zhao 2000; West and Zhao 2000; Meng 2000; Meng and Manning 2010). When in the cities, migrants have little access to unemployment supports, health care, retirement pensions, or the Minimum Living Allowance scheme available to urban *hukou* holders as a last resort for poverty alleviation in urban areas. In addition, migrant children are often denied access to urban public schools (Meng and Manning 2010). From 2008 to 2010, the proportion of migrant workers with access to unemployment insurance increased from 12 to 13.5 percent, while the proportion for workers with urban *hukou* increased from 60 to 66 percent. The proportion of migrants with access to urban health insurance is 20 percent in 2010, while for urban *hukou* workers it is 87 percent (Frijters, Gregory, and Meng forthcoming).

As a result of this institutionalized discrimination, most rural–urban migrants do not see their long–run future in cities. Instead, they leave their families behind and migrate to the cities hoping to earn as much as possible before returning home (Meng and Manning 2010). Migrants normally go to cities in their late teens. Women typically begin to return home in large numbers between age 25 and 35 to marry and have children, while men start returning in their mid-30s. On average, migrants stay in cities for only about seven years.[9] At the peak of migration—an age of 25 for males and 20 for females—55 percent of the male and 50 percent of the female rural labor force has migrated, as shown in Figure 5. In total, 22 percent of the rural labor force was working in cities in 2009.[10]

When migrants lose a job or fall ill, their fallback position is to return to their home village. The global financial crisis hit China in mid-2008, causing 20 to 45 million migrant workers to return to their home villages during the end of 2008, which is a substantial part of the reason the unemployment rate increase

[9] The length of stay is calculated using information on the year the migrant first moved to the city and the calendar year of the survey. As we have no information on churning (the number of spells in between the first migration and the survey date), this observed length of stay to the survey date may be an overestimate. In addition, this observed duration suffers from two potential biases. First, individuals who have completed their durations have left the city and cannot be observed. Second, the remaining individuals are those who have not completed their durations and hence are right-censored. Assuming a constant exit rate and a steady state, the two biases may largely offset each other.

[10] These data are from the RUMiCI rural survey. The 145 million migrants number indicated earlier is an aggregated figure published by China's National Bureau of Statistics using Rural Household Survey data, which gives a slightly higher proportion, at around 27 percent of rural *hukou* labor force.

Figure 5
Rural Labor Force Migration by Age and Gender, 2009
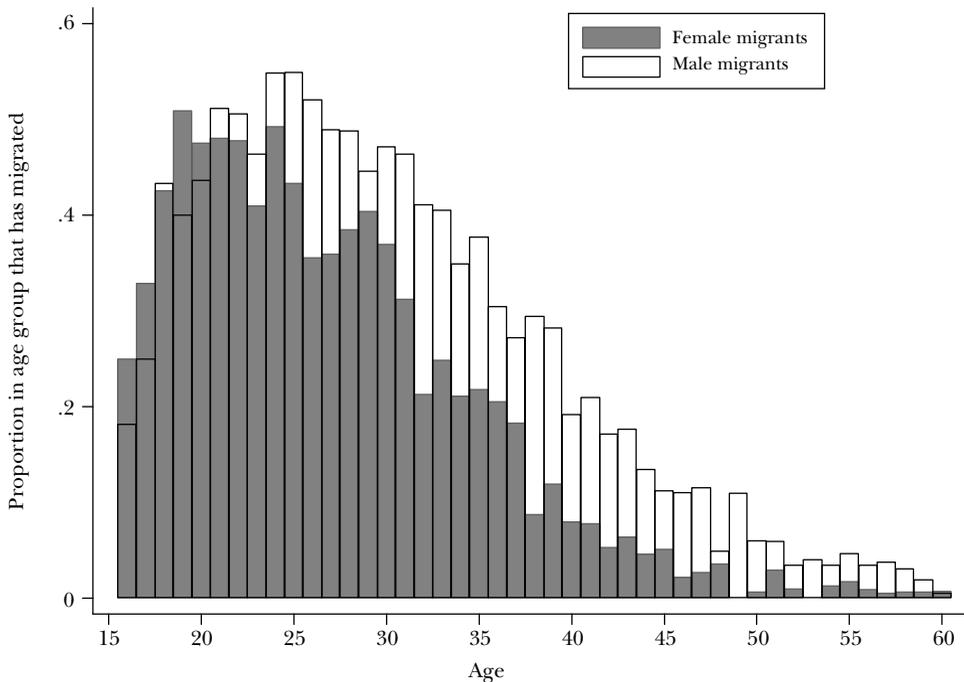
Source: 2009 Rural Urban Migration in China and Indonesia (RUMiCI) Project data for China; author's own calculations.

in cities was muted (Huang, Zhi, Huang, Rozelle, and Giles 2011; MoHRSS 2010; Kong, Meng, and Zhang 2010).

Migrant Employment and Occupation

Because institutional restrictions discourage migrants from staying in cities when unemployed, it is no surprise that migrant employment rates are extremely high. The employment rate for migrants in 2009 was 94 percent, while for their urban *hukou* counterparts in the same cities it was 63 percent (as shown in Table 2). These striking differences extend to other labor market attributes. Migrants work an average of 63 hours per week, while their urban *hukou* counterparts work 44 hours. Wage-earning migrants work 15 hours more on a weekly basis than their urban *hukou* counterparts, while for self-employed the difference is 20 hours.

Few migrants working in cities are employed in the state sector: 7.3 percent in 2009, compared with 49.4 percent of urban *hukou* workers. Furthermore, migrants are more likely to be self-employed. In 2009, around 27 percent of employed migrants fall into the self-employed category, compared with 8.4 percent of urban workers. This gap, is, in part, a response to the discrimination against migrants in salaried sectors in

Table 2

Individual Characteristics by Migration Status

	Migrants			Urban hukou holders		
	Females	*Males*	*Total*	*Females*	*Males*	*Total*
Labor force participation rate (%)	93.36	97.73	**95.87**	60.78	78.55	**69.45**
Labor force employment rate (%)	91.66	96.26	**94.30**	54.18	72.08	**62.91**
% of employed work as self-employed	29.68	26.28	**27.69**	7.48	9.11	**8.40**
Ownership of employment:						
State	5.26	8.74	**7.31**	45.75	52.25	**49.43**
Collective	2.72	4.04	**3.50**	5.15	5.24	**5.20**
Private	92.02	87.21	**89.19**	49.10	42.51	**45.37**
Occupation						
Professional	0.66	0.58	**0.61**	20.42	24.74	**22.88**
Managers including shop owners	4.15	5.61	**5.01**	3.43	7.29	**5.63**
Clerks	6.68	3.67	**4.90**	27.54	22.05	**24.41**
Sales/Service workers	60.96	52.42	**55.91**	32.79	18.59	**24.69**
Production workers	26.20	37.15	**32.67**	9.04	20.29	**15.46**
Others	1.35	0.58	**0.89**	6.78	7.03	**6.93**
Age	31.43	32.31	**31.95**	38.13	41.81	**40.20**
% males			**58.75**			**56.47**
Years of schooling	9.04	9.38	**9.24**	12.29	12.31	**12.30**
Education level:						
Primary	19.94	11.90	**15.24**	3.60	3.00	**3.26**
Junior high	50.15	52.98	**51.80**	18.99	19.14	**19.07**
Senior high	24.04	29.45	**27.21**	36.11	36.01	**36.06**
3-year college	5.24	4.75	**4.95**	25.63	23.05	**24.18**
University or above	0.63	0.92	**0.80**	15.67	18.81	**17.43**
	Wage earners	*Self-employed*	*Total*	*Wage earners*	*Self-employed*	*Total*
Average weekly hours worked (hours)	57.65	77.47	**63.21**	42.49	57.23	**43.77**

Source: RUMiCI Migrant and Urban Surveys, 2009.

terms of job attainment and earnings (Frijters, Kong, and Meng 2011). For example, over 89 percent of migrant workers are employed as unskilled workers in sales/service or production jobs, while only 40 percent of urban *hukou* workers are in this category.[11]

[11] What proportion of these large differences in occupational attainment is due to the difference in observable characteristics, such as age, gender, and education of the migrant workers? In the online Appendix B, Table B3, available with this paper at http://e-jep.org, a linear probability regression is estimated to predict whether an individual has a skilled (defined as professional, managerial, or clerical) or unskilled job (including retail/wholesale trade, services, production laborers, and not otherwise specified). Education and age are all significantly associated with being a skilled worker. Controlling for all observable characteristics, including city where employed, migrants are around 16 to 24 percent less likely to have a skilled job. This is true for both men and women, and indicates that over and above their attribute differences there is still a high level of disadvantage in the occupational choices that migrants face.

Migrants, on average, are younger than their urban local counterparts by around eight years. They are also more likely to be male and less educated. Although it is the better-educated rural workers who tend to migrate, they still have considerably less schooling (9.2 years) relative to their urban *hukou* counterparts (12.3 years), as shown in Table 2. In particular, 42 percent of urban *hukou* workers have a three-year college degree or above, while the proportion for migrant workers is around 6 percent, which is still higher than for those who did not migrate.

Migrant Wage Growth and Labor Market Discrimination

Migrants have always been at the lower end of the wage distribution, earning on average only 45 percent of the average urban *hukou* workers' hourly wage in 2009 (Frijters, Gregory, and Meng forthcoming). There are no representative and consistent migrant surveys extending through the last decades to allow a comparison of changes in migrant earnings with those of their urban *hukou* counterparts.

However, the RUMiCI survey does ask migrants to report information on their first job in cities, including the year, the duration, the first and last month pay of the first job, as well as other employment details. Combining these data with those from the Urban Household Survey, the monthly earnings of migrant and urban workers for the 15 cities between 2002 and 2008 can be constructed, although the sample sizes are not large.[12] During this period, real earnings during the first month of the first job for migrant workers rose about 3.8 percent per year, while pay on the last month of the first job for migrants rose about 2.1 percent per year in real terms. This increase is about one-quarter of the increase for an urban *hukou* new entrant (with up to one year of experience)—their real earnings increased by 16 percent per annum during this period. If we look at total urban *hukou* blue collar workers (the majority of migrant workers are blue collar workers), the change is 11.4 percent per annum. The raw difference in wages and in the growth rate of wages roughly remains if one adjusts for various observable characteristics. Using the log of monthly real earnings as the dependent variable and controls for age, city work experience, education levels, gender, a migrant indicator, regional dummies, and year dummy variables, and interacting each of the independent variables with a dummy variable for whether the worker is a migrant, the implied growth rate for urban blue-collar workers adjusted for these other factors is around 6.8 percent per annum, and for migrants is 2.3 percent per annum (see also Golley and Meng 2011). Thus, migrant wages are increasing in real terms but falling behind in relative terms. It is important to note that the data used here is not ideal. Lack of large-scale official earnings data which cover both urban *hukou* and migrant workers extending

---

[12] The reason only data for 2002 and after are used is to minimize recall errors and to increase the sample size for each year to around 300. Even so, there must be considerable doubt as to the precision of the migrant earnings data. To check whether the migrant earnings information is in the ballpark, I examined monthly earnings data from the CHIPs 2002 migrant survey. The data show that in 2002 an average migrant in the CHIPs sample was paid 783 yuan per month, while RUMiCI data for the first and last month pay of the first job is 685 yuan and 822 yuan in that year, respectively. Both are on the right-hand side of the distribution and are quite close to the mean, suggesting that recall errors may not be large.

back to the late 1990s or early 2000s is a serious impediment to our understanding of this important issue.

Several conjectures might explain the differential earnings growth between urban and migrant workers. First, labor supply conditions for the two types of workers are very different. Currently more than 70 percent of the population of China has rural *hukou,* and of the rural *hukou* workforce only 22 percent are working in urban areas. However, restrictions placed on rural–urban migrants in terms of job access prevent them from becoming perfect substitutes for urban *hukou* workers. The labor markets for urban *hukou* workers and migrant workers are segregated. The potential supply of migrant workers is significantly larger than for urban *hukou* workers, which should suppress wage growth for migrant workers. Second, if the economy adopts a skill-biased technology, demand for labor should be biased towards highly educated, skilled workers and hence favor predominantly urban *hukou* workers (see education levels in Table 2). These elements are all related to the long-standing policy of a rural–urban divide: urban workers are protected from competition from rural labor supply; they are directly protected so they may obtain good jobs; and indirectly, they receive better-quality education and hence are able to reap the rewards from skill-biased technologies.

A number of studies find that labor market discrimination suppresses migrant wages and wage growth (Meng and Zhang 2001; Zhang 2009; Frijter, Lee, and Meng 2011; Frijter, Gregory, and Meng forthcoming). In particular, in Frijter, Gregory, and Meng (forthcoming), my coauthors and I show that in 1995 migrant workers in Shanghai earned 50 percent of the hourly earnings of urban *hukou* workers, and that 47 percent of this gap could not be explained by differences in observed characteristics. By 2009, the hourly earnings gap between the two groups in the same city increased to 60 percent, and 53 percent of the gap could not be explained by differences in observable characteristics. The wage gaps between rural migrants and those with urban *hukou* do not seem to be narrowing.

If migrants stayed longer in cities could they narrow the gap or "catch up" to urban workers? Zhang (2009) finds that migrants who stay longer do experience a statistically significant and weak narrowing of the gap, but not a complete catch-up. She also finds that the migrants' experience-earnings profile peaks around the 15-year mark. But remember that migrants on average only stay in cities for seven years. Thus, the policies discouraging migrants from staying in cities longer not only lower migrant earnings but also disadvantage the economy by not allowing migrants to reach their potential productivity peak.

The Future of China's Labor Force: Quantity and Quality

The One-Child Policy and Institutional Impediments to Future Labor Supply

At the end of the 1960s, low agricultural productivity was raising concerns as to whether China could feed its ever-growing population. In response, the government began discussing a series of policies to discourage fertility. For example,

Figure 6
Population Pyramid, 2000

A: Rural

B: Urban
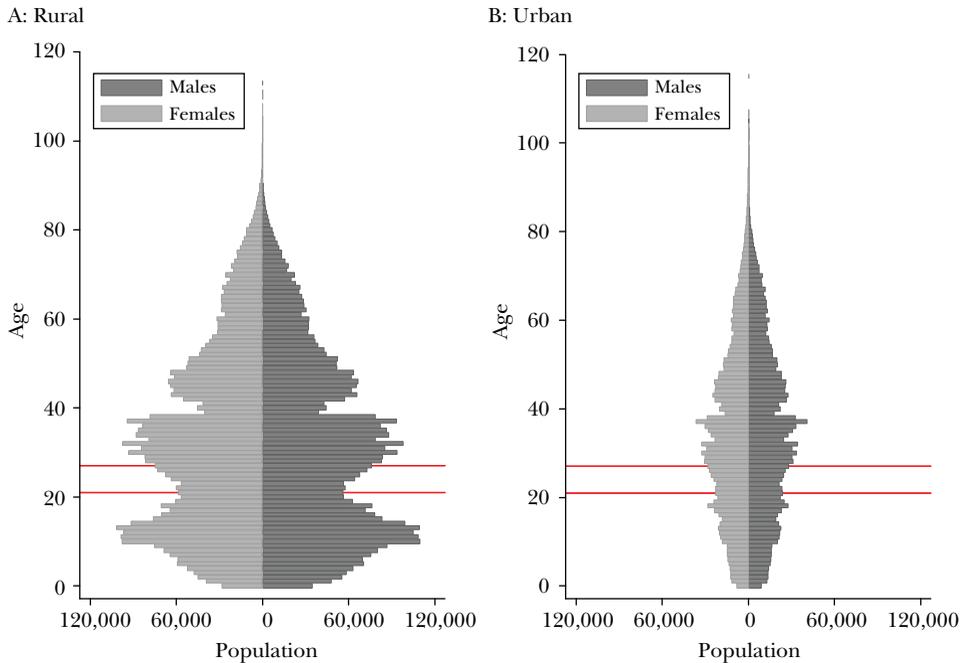
Source: One percent sample of the 2000 Population Census data, author's own calculations.

the policy of "Later, Longer, and Fewer"—have children later, have a longer time between births, and have fewer children—was introduced in 1973. In 1978, the policy slogan was "One is the Best and Two is the Most," and "Reward Having One Child and Punish Having Three." In 1979 the "One Child per Couple" policy was implemented (Center for Population Studies et al. 1986; Peng 1991; Feeney and Wang 1993).

The one-child policy was strictly enforced in many urban areas, but in rural areas a second child, or even a third, has always been allowed if the previous births were girls (Peng 1991). The one-child policy was associated with a total fertility rate decline from 2.8 children per woman in 1979 to 1.8 in 2000. The total population reduction in the first 25 years of implementation relative to what the population otherwise would have been is 250 to 300 million people (Festini and de Martino 2004).

Figure 6 presents the population pyramids for the urban and rural *hukou* populations separately for the year 2000. The two horizontal lines indicate those who were born in 1973 (the introduction of the "Later, Longer, and Fewer" policy) and 1979 (the introduction of the "One Child per Couple" policy). The pyramid for the rural *hukou* population shows a low number of births between the early 1970s and mid-1980s. However, this is less about the one-child policy than it is an "echo" effect

of the Great Famine of the late 1950s.[13] In fact, the cohorts after the echo effect are about the same size, if not larger than, the cohorts born immediately after the Great Famine, suggesting that the one-child policy did not significantly reduce the reproduction of the population for rural *hukou* population. Figure 6 also suggests that a new round of the echo effect may be underway. This can be observed in the small size of the population under 10 years of age. But this population reduction will not be huge. In 2000, the size of the 0–10 year-old rural *hukou* population was 85 percent that of the rural population aged 11–20, 98 percent that of the 20–30 age group, and 90 percent that of the rural population aged 31–40.

The urban pyramid depicts a different story. Below the horizontal lines, population size fell almost year after year so that the "pyramid" is diamond-shaped, indicating a shrinking urban *hukou* population size that started about the time of the one-child policy. The size of the urban *hukou* population aged 0–10, for example, is 58 and 55 percent that of the population aged 21–30 and 31–40, respectively.

Nevertheless, as more than 70 percent of the population has rural *hukou*, the limited effect of the one-child policy on the rural population dominates the total population story. The Chinese population is projected to grow to 1.46 billion in 2030 and then to decline to 1.41 billion by 2050 if there is only a very slight increase in fertility (United Nations, Department of Social and Economic Affairs, Population Division, 2007).[14] The population pyramids illustrate that the new entrants to China's labor force in the next decades will be predominantly drawn from the rural *hukou* population.

Since 2004, there have been reports of migrant labor shortages in coastal Chinese cities, along with debate as to whether China has reached the (Arthur) Lewis-style "turning point," whereby rural labor supply is exhausted and unskilled wages in urban areas begin to rise substantially in response (for example, Cai 2007, 2010; Golley and Meng 2011; Knight, Deng, and Li 2011; Yang, Chen, and Monarch 2010). The discussion in the last section has shown that the earnings of migrants in real terms have been increasing, but this is unlikely to be the result of an absolute "labor shortage" but rather an effect of institutional restrictions on migration. As discussed earlier, as a result of the institutional restrictions on migrant access to social welfare and social services in cities, only 22 percent of the rural *hukou* labor force has migrated to cities so far, and they often stay for a relatively short time. It is unlikely that at the aggregate level there is a "shortage" of unskilled labor. Imagine if the restrictions on migrant access to social welfare and social services in cities were abolished and the duration of migration, say, doubles. Any conceivable "labor shortages" would disappear. Many more rural workers would also become available to migrate. When answering the question "if the policy restrictions on

---

[13] The Chinese Great Famine occurred between 1959 and 1961. In 2000, those who were born during the famine were 39 to 41 years old.

[14] Based on the same projection, if fertility were to increase by 0.5 child per woman, the population will continue to increase until 2050 to 1.65 billion. In fact, an increase in fertility by 0.5 per woman seems fairly likely to happen, because many provinces have already allowed couples who are both a single child to have two children.

migration were to be relaxed, how long would you be willing to stay in the city?" 62 percent of the migrants in the 2010 RUMiCI survey indicated that they would stay in the city forever. Another 24 percent answered that they do not know how long they would stay. The "labor shortages" currently observed in coastal regions are not generated by the lack of absolute numbers of potential workers but by the institutional restrictions.

Thus, one policy response to city labor shortages would be to change the institutional restrictions to increase both the inflow of migrants and their length of stay in cities. However, such a reform would not be a trivial undertaking: it raises many financial and political complications. For example, in 2008 the central government introduced a new law that requires all employers to pay the health, unemployment, work injury, and pension insurances for migrant workers. The policy should be enforced by the local governments, but they have limited incentive to do so. Enforcing the policy implies a significant increase in labor cost, which has direct implications for local economic growth, upon which local governments are evaluated. Employers, in turn, are unlikely to pay if they can avoid it (Meng and Manning 2010). As a result, although the new law has been in place for four years, very few improvements were made with regard to migrants' access to social insurance. In addition, the portability of insurance is also an issue. Local governments are unwilling to let their social insurance funds move to somewhere else. Furthermore, when it comes to paying for social services, local governments only care about the welfare of local constituents and are unwilling to pay for rural migrants from other provinces. Policy issues also arise in rural areas. Should migrants who leave the countryside still be allowed to keep their land usage rights in rural areas once they are given the same welfare entitlement as urban *hukou* holders?[15] To find answers and solutions for all these issues takes time.

Meanwhile, without the institutional changes, the unskilled labor "shortage" in cities may worsen and migrant workers' wages in cities will rise further. This may induce capital outflow to other low-cost countries and Chinese industries will increasingly move to the capital and technology-intensive end of the spectrum—which may sound like a good outcome, but it does imply that China will not follow an optimal resource allocation growth path. In the meantime, maintaining the restrictions on rural migrants means that there is no obvious mechanism for the 50 percent of unskilled and underemployed farmers still in the rural areas to share in China's economic growth.

Might it be possible that most of the current rural areas will be urbanized, so that rural workers do not have to move to cities to become "urbanized"? It has always been the Chinese government's policy intention to develop small and medium-sized cities, precisely for the purpose of reducing large-scale rural–urban migration. The policy slogan is "leaving agricultural work, but not leaving rural areas." As a result of this policy, China has many small cities. Recent studies on urbanization in China,

---

[15] Currently, migrants who move to cities keep land usage rights as a de facto welfare insurance to support them if they lose their city job, fall ill, or become disabled.

however, have found that the majority of Chinese cities are too small to be economically viable (Wang and Xia 1999; Au and Henderson 2006; Wang 2011). Henderson (2009) estimates that if the average size of the Chinese prefecture-level cities can be doubled, their value added per worker can be increased by 20 to 35 percent.

Might it be possible to change the education system, so that the skill set of the future labor force from rural areas can fit China's new industrial structure? This leads to the question of the quality of China's future labor force.

Polarization in Education

During the Communist regime, China did well in reducing illiteracy. The literacy rate increased from less than 30 percent in the 1940s to 48 percent in 1964, 66 percent in 1978, and 96 percent by the early 2000s (Dreze and Loh 1995; Zhang 1997; Zhang and Kanbur 2005). The average years of schooling for the rural *hukou* population have risen steadily from about three years for those born in 1935 to eight years for those born in 1980. Meanwhile, the years of schooling for the urban *hukou* population have risen from seven years for those born in 1935 to 13 years for those born in 1980 (based on data from the 1% Population Survey of 2005).

Despite the rising literacy rate, the rural–urban divide in education has been growing. The average schooling gap between urban and rural *hukou* holders fell from 4.8 years for the cohort born before the communists took power (the before-1940 cohorts) to around 3.6 years for the cohort born in the 1950s and early 1960s, and then gradually increased to over 4.0 years again for those who were born in the 1970s. Wu (2011) also finds an increase in the senior high school enrollment gap between rural and urban *hukou* holders for cohorts born during the 1970s and 1980s. In 2008, the proportion of rural *hukou* holders with three years of college or above was less than 1 percent, and the proportion with senior high school or above was 12 percent. The corresponding ratios for urban *hukou* holders were 17 percent and 51 percent, respectively.

A number of factors may have contributed to the resurgence of the widening of the rural–urban education divide. First, when economic reform abolished the rural commune system, most rural services financed by communes deteriorated, including education.

Second, greater economic opportunities for rural young people at different stages of economic reform—working in the rural township and village enterprises during the early reform years and moving to cities to work in the later reform period—increased the opportunity cost of going to school (Unger 2002; Brown 2006; de Brauw and Giles 2006).

Third, as reform deepened, the Chinese government did not adequately replace education provision in rural areas, but instead, moved towards an education system that relied more on local and private funds. The central government spending on education as a proportion of GDP hovered between 2 to 3 percent (Tsang and Ding 2005; Hannum, Behrman, Wang, and Liu 2008; Robertson and Xu 2008; Hannum and Park 2007). The urban/rural income ratio increased from two-fold in the mid-1980s to 3.5-fold in the mid 2000s (NBS, various years). As this

income gap widened, access to education became less equal across the urban/rural divide. Hannum and Park (2007) find that even in a very poor rural setting, the education achievement of children is significantly affected by the wealth of their families. In Frijters, Luo, and Meng (2010), my coauthors and I also find a large gap in school performance between rural and urban children and a strong association between school performance and income within rural areas.

Fourth, the policy to expand university enrollment since 1999 benefited urban areas more than rural areas. From 1998 to 2007, university enrollment increased from 108 million to 565 million (NBS, various years). The proportion of urban *hukou* workforce aged 23 to 60 with three-years of college and above increased from 12 percent in 1988 to just below 40 percent in 2009. The differential impact of this large expansion for the rural and urban divide has not been fully documented. However, looking across provinces, in Luo and Meng (2010) we find that every 1 percentage point increase in the share of urban *hukou* population in a province is associated with an additional 0.49 percentage points higher university enrollment and this variable alone can explain 40 percent of the variation in university enrollment across different provinces. Li (2010) also finds that the probability of having a university education increases 6.5-fold if the father has urban *hukou*.

The increase in the urban–rural education divide has important implications for future Chinese economic growth patterns. As indicated above, the future labor supply will mainly come from the rural *hukou* population. As the economy grows, it will inevitably improve its technology, which will require a more educated labor force. If rural education does not catch up, it will place significant pressure on the quality of China's future labor supply and generate a mismatch between demand and supply for labor. This situation may be worsened if institutional restrictions on rural–urban migration are not relaxed fast enough. These restrictions generate unskilled labor shortages in cities, which in turn will shift China towards a capital- and technology-intensive industrial structure further and faster than if development were based on optimal resource allocation. Thus, if rural education investment increasingly lags behind so that rural workers cannot fill the more-skilled urban job openings, we may see a slower urbanization process, an increase in rural underemployment, and a further widening of the income gap.

The increase in the supply of college graduates has slowed down the increase in the returns to education, which may indicate an oversupply of college graduates in the short run. However, in the long run, this increase in the supply of skilled workers might be preparing China for the upgrading of industries along the value-added chain. Nonetheless, the shrinking urban *hukou* population indicates that the key to China's future growth will be how fast China narrows the rural–urban education divide.

Concluding Remarks

China has experienced exceptional labor market changes in recent decades, which will echo into the future. First, China has a shrinking urban *hukou* population,

and so new entrants to the labor market will come primarily from the rural *hukou* population. Second, the rural population is significantly less educated. Third, considerable institutional restrictions remain on access of rural migrants to jobs and to social welfare and social services in cities. As a result, Chinese cities have begun to experience some unskilled labor shortages even though more than 50 percent of the labor force is still in the rural sector. The clear policy implication is that China should adopt labor market reforms on all fronts to enable a better utilization of its abundant resources and a smooth transition to a more skill-intensive economy: it should reduce restrictions on access of rural migrants to cities and also adjust its educational investment in rural areas to increase the skill level of rural workers. Without substantial policy changes on these fronts, China may observe a continued or even faster increase in earnings for urban *hukou* holders in the next 10 to 20 years, but the rural–urban income divide—with most of the population on the rural side—could become very large.

Some adjustments on these fronts can be observed, but progress is slow. One reason is that while China's central government passes laws and states policy intensions, resources and implementation are local government responsibilities. Thus, any real change to eliminate migration restrictions and the rural–urban education divide will have to be driven and paid for by the central government, or a significant reform on the current public finance system has to be implemented either to centralize the financing of China's social welfare system or to increase significantly the resources of local governments.

References

Appleton, Simon, John Knight, Lina Song, and Qingjie Xia. 2002. "Labor Retrenchment in China: Determinants and Consequences." *China Economic Review* 13(2–3): 252–75.

Appleton, Simon Mark, Lina Song, and Qingjie Xia. 2005. "Has China Crossed the River? The Evolution of Wage Structure in Urban China during Reform and Retrenchment." *Journal of Comparative Economics* 33(4): 644–63.

Au, Chun-Chung, and J. Vernon Henderson. 2006. "Are Chinese Cities Too Small?" *Review of Economic Studies* 73(3): 549–76.

Bosworth, Barry, and Susan M. Collins. 2008. "Accounting for Growth: Comparing China and India." *Journal of Economic Perspectives* 22(1): 45–66.

Brandt, Loren, and Carsten A. Holz. 2006. "Spatial Price Differences in China: Estimates and Implications." *Economic Development and Cultural Change* 55(1): 43–86.

Brown, Philip H. 2006. "Parental Education and Investment in Children's Human Capital in Rural China." *Economic Development and Cultural Change* 54(4): 759–89.

Cai, Fang. 2007. "Lewisian Turning Point during China's Economic Growth" (in Chinese). In *Green Book of Population and Labor: Lewisian Turning Point and Its Policy Challenges* (in Chinese), edited by Cai Fang and Du Yang. Beijing, China: Social Science Academic Press.

Cai, Fang. 2010. "Demographic Transition, Demographic Dividend, and Lewis Turning Point in China." *China Economic Journal* 3(2):107–19.

Cai, Fang, Albert Park, and Yao-Hui Zhao. 2009. "The Chinese Labor Market in the Reform Era." In *China's Great Economic Transformation,* edited by Loren Brandt and Thomas Rawski, pp 167–214. Cambridge University Press.

Center for Population Studies, Chinese Academy of Social Sciences and Editorial Board of the China's Population Yearbook. 1986. "The Major Events of China's Population Activities" (in Chinese). In *China's Population Yearbook 1985* (in Chinese), pp. 1263–88. Beijing: Chinese Social Sciences Press.

de Brauw, Alan, and John Giles. 2006. "Migrant Opportunity and the Educational Attainment of Youth in Rural China." IZA Discussion Papers 2326, Institute for the Study of Labor (IZA).

Dickens, William T., and Lawrence F. Katz. 1987. "Inter-industry Wage Differences and Industry Characteristics." In *Unemployment and the Structure of Labor Markets,* edited by K. Lang and J. S. Leonard, 48–49. Oxford: Basil Blackwell Inc.

Dreze Jean, and Jackie Loh. 1995. "Literacy in India and China." *Economic and Political Weekly,* November 11, pp. 2868–78.

Fan, Gang. 2000. "The Dynamics of Transition in China: Change of Ownership Structure and Sustainability of Growth." Paper presented at the International Conference of China Growth Sustainability in the 21st Century, Canberra, Australia.

Fan, Simon C., and Xiangdong Wei. 2006. "The Law of One Price: Evidence from the Transitional Economy of China." *Review of Economics and Statistics* 88(4): 682–97.

Feeney, Griffith, and Feng Wang. 1993. "Parity Progression and Birth Intervals in China: The influence of Policy in Hastening Fertility Decline." *Population and Development Review* 19(1): 61–101.

Feng, Lanrui. 2003. "Comments on Change in Employment in China" (in Chinese). In *Employment and Social Security during the Transition* (in Chinese), edited by Zheng Gongcheng, Zeng Xiangquan, Zheng Yushuo, and Zhang Qiheng. Beijing: China Labor and Social Security Publishing House.

Festini, Filippo, and Maurizio de Martino. 2004. "Twenty-five Years of the One-Child-Family Policy in China: Problems and Future Prospects." *Journal of Epidemiology and Community Health* 58(5): 358–60.

Frijters, Paul, Robert G. Gregory, and Xin Meng. Forthcoming. "The Role of Rural Migrants in the Chinese Urban Economy." In *Migration-Economic Change, Social Challenge,* edited by Christian Dustmann. Oxford: Oxford University Press.

Frijters, Paul, Tao Kong, and Xin Meng. 2011. "Migrant Entrepreneurs and Credit Constraints under Labour Market Discrimination." IZA Discussion Paper 5967.

Frijters, Paul, Leng Lee, and Xin Meng. 2010. "Jobs, Working Hours, and Remuneration Packages for Migrants and Urban Residents." Chap. 2 in *The Great Migration: Rural-Urban Migration in China and Indonesia,* edited by X. Meng, C. Manning, with S. Li, and T. N. Effendi. Edward Elgar Publishing.

Frijters, Paul, Chuliang Luo, and Xin Meng. 2010. "Child Education and the Family Income Gradient in China." Working Paper, University of Queensland.

Giles, John, Albert Park, and Fang Cai. 2006. "How has Economic Restructuring Affected China's Urban Workers." *China Quarterly* 185 (March): 61–95.

Golley, Jane, and Xin Meng. 2011. "Has China Run out of Surplus Labor?" *China Economic Review* 22(4): 555–72.

Gong, Xiaodong, Sherry Tao Kong, Shi Li, and Xin Meng. 2008. "Rural–Urban Migrants: A Driving Force for Growth." Chapter 6 in *China's Dilemma,* edited by Ligang Song and Wing Thye Woo. Canberra, Australia: ANU E Press and Asia Pacific Press.

Gong, Cathy Honger, and Xin Meng. 2009. "Regional Price Differences in Urban China 1986–2001: Estimation and Implication." IZA Discussion Papers 3621, Institute for the Study of Labor.

Hannum, Emily, Jere Behrman, Meiyan Wang, and Jihong Liu. 2008. "Education in the Reform Era." Chapter 7 in *China's Great Economic Transformation,* edited by Loren Brandt and Thomas Rawski. Cambridge University Press.

Hannum, Emily, and Albert Park. 2007. "Academic Achievement and Engagement in Rural China." In *Education and Reform in China,* edited by Emily Hannum and Albert Park, pp. 154–72. Routledge.

Henderson, J. Vernon. 2009. "Urbanization in China: Policy Issues and Options." Unpublished manuscript, Brown University. At the author's webpage, http://www.econ.brown.edu/faculty /henderson/ under the link "Reports for the China Economic Research and Advisory Program."

Huang, Xinyuan. 2004. "Wage Ranking in 1956" ("1956 Nian De Ding Ji"). In *People's Political Consultative Times (Ren Min Zheng Xie Bao),* October, 28.

Huang, Jikun, Huayong Zhi, Zhurong Huang, Scott Rozelle, and John Giles. 2011. "The Impact of the Global Financial Crisis on Off-Farm Employment and Earnings in Rural China." *World Development* 39(5): 797–807.

Jiang, Xiaojuan, and Hui Li. 2005. "Regional Income Inequality Using Real Income is Smaller than That Using Nominal Income" (in Chinese). *Economic Research Journal (Jinji Yanjiu),* 2005(9), pp.11–18.

Knight, John, Quheng Deng, and Shi Li. 2011. "The Puzzle of Migrant Labour Shortage and the Rural Labour Surplus in China." *China Economic Review* 22(4): 585–600.

Kong, Sherry Tao, Xin Meng, and Dandan Zhang. 2010. "Impact of Economic Slowdown on Migrant Workers." Chap. 12 in *China's New Place in a World in Crisis: Economic, Geopolitical and Environmental Dimensions,* edited by Ross Garnaut, Ligang Song and Wing Thye Woo. Canberrra, Australia: ANU E Press.

Li, Chunling. 2010. "Tertiary Education Expansion and Inequality in Education Opportunity" (in Chinese). *Sociology Research (She Hui Xue Yian Jiu),* 2010(3), pp. 1–34.

Li, Xiaohua, Yaohui Zhao, and Lili Lu. 2007. "Effects of Education on Wage Inequality in Urban China: 1988–2003." Paper presented at the 6th PEP Research Network General Meeting in Lima, Peru, June 14–16, 2007.

Luo, Chuliang, and Xin Meng. 2010. "Inequality in Tertiary Education, Decision to Enroll in Senior High School, and Rural–Urban Divide" (in Chinese). Unpublished manuscript, Beijing Normal University.

Maurer-Fazio, Margaret, Rachel Connelly, Lan Chen, and Lixin Tang. 2010. "Childcare, Eldercare, and Labor Force Participation of Married Women in Urban China, 1982–2000." *Journal of Human Resources* 46(2): 261–94.

Meng, Xin. 1997. "China's Labour Market Reform." Report for the East Asian Analytical Unit at the Department of Foreign Affairs and Trade, Australia. This report formed the basis for Chapter 12 in *China Embraces the Market.*

Meng, Xin. 2000. *Labour Market Reform in China.* Cambridge University Press, Cambridge.

Meng, Xin. 2004. "Economic Restructuring and Income Inequality in Urban China." *Review of Income and Wealth* 50(3): 357–79.

Meng, Xin, and Michael Kidd. 1997. "Wage Determination in China's State Sector in the 1980s." *Journal of Comparative Economics* 25(3): 403–21.

Meng, Xin, Kailing Shen, and Sen Xue. 2012. "Economic Reform, Education Expansion, and Earnings Inequality for Urban Males in China, 1988–2009." *Journal of Comparative Economics,* available online. http://www.sciencedirect.com /science/article/pii/S0147596712000376.

Meng, Xin, and Chris Manning. 2010. "The Great Migration in China and Indonesia—Trend and Institutions." Chap. 1 in *The Great Migration: Rural–Urban Migration in China and Indonesia,* edited by X. Meng and C. Manning, with S. Li, and T. Effendi. Edward Elgar Publishing.

Meng, Xin, and Junsen Zhang. 2001. "Two-Tier Labor Market in Urban China: Occupational Segregation and Wage Differentials between Urban Residents and Rural Migrants in Shanghai." *Journal of Comparative Economics* 29(3): 485–504.

MoHRSS. 2010. "Special Report Number Three: Employment Demand of Enterprises in Spring 2010 and Employment Situation of Rural Migrants in 2009" (in Chinese). In *The Green Book of Population and Labor* (in Chinese), edited by Fang Cai. Beijing: Academy of Social Sciences Press.

National Bureau of Statistics (NBS). 1999. *Comprehensive Statistical Data and Materials on 50 Years of New China.* Beijing: China Statistical Press.

National Bureau of Statistics (NBS). Various years. *China Statistical Yearbook.* Beijing: China Statistical Press.

OECD. Various years. Labour Force Statistics, OECD.StatExtracts from http://www.oecd-ilibrary .org/statistics.

OECD. 2010. *OECD Economic Surveys: China 2010.* Organisation for Economic Co-operation and Development.

OECD. 2011. *Divided We Stand: Why Inequality Keeps Rising.* Organisation for Economic Co-operation and Development.

Park, Albert, Fang Cai, and Yang Du. 2010. "Can China Meet Her Employment Challenges?" In *Growing Pains: Tensions and Opportunities in China's Transformation,* edited by Jean Oi, Scott Rozelle, and Xuegang Zhou. Stanford: Stanford Asia-Pacific Research Center.

Park, Albert, Xiaohua Li, Xiaoqing Song, Junsen Zhang, and Yaohui Zhao. 2004. "Rising Returns to Skill, Labor Market Transition, and the Growth of Wage Inequality in China." Unpublished paper, University of Michigan.

Peng, Xizhe. 1991. *Demographic Transition in China.* Oxford: Clarendon Press.

Perkins, Dwight, and Shahid Yusuf. 1984. *Rural Development in China.* Johns Hopkins University Press.

Robertson, Peter E., and Jessica Y. Xu. 2008. "Education Investment in China and Its Implications for China and East Asia." http://wwwdocs.fce .unsw.edu.au/economics/Staff/PROBERTSON /Papers/Robertson_Xu.pdf.

State Council of People's Republic of China. 1999. "The Unemployment Insurance Legislation," Number 258 Regulation of the State Council (In Chinese). http://www.gov.cn/banshi/2005-08 /04/content_20258.htm.

Tsang, Mun C., and Yanqing Ding. 2005. "Resource Utilization and Disparities in Compulsory Education in China." *China Review—An Interdisciplinary Journal on Greater China* 5(1): 1–31.

Unger, Jon. 2002. *The Transformation of Rural China.* Armonk: M. E. Sharpe.

United Nations, Department of Social and Economic Affairs, Population Division. 2007. Graph from World Population Prospects, The 2006 Revision titled "Analyses: Tables, Figures &Maps: China's Population, 1950–2050" (http:// www.china-profile.com/data/printing/pr_fig _pop_wpp-2006.htm). For more information, see Country Profiles at http://www.china-profile.com /data/printing/pr_fig_pop_wpp-2006.htm.

Wang, Hansheng, and Xiaoqiang Wang. 1995. "Migration and Migration Selection" ("Zai Liu Dong Zhong Shi Xian Jing Ying Yi Min"). *Strategy and Management* (in Chinese), 1995(5): 112–20.

Wang, Xiaolu. 2011. "The Route of Urbanisation in China from an International Perspective." Chapter 18 in *Rising China: Global Challenges and Opportunities,* edited by J. Golley and L. G. Song. Canberra: ANU E Press.

Wang, Xiaolu, and Xiaolin Xia. 1999. "Promoting Economic Growth by Optimizing City Size" (in Chinese). *Economic Research Journal* (in Chinese), 1999(9): 22–29.

Wang, Xiaolu, and Wing Thye Woo. 2011. "The Size and Distribution of Hidden Household Income in China." *Asian Economic Papers* 10(1): 1–26.

West, Loraine A., and Yaohui Zhao. 2000. *Rural Labor Flows in China.* Berkeley, CA: University of California Press.

Wu, Xiaogang. 2011. "The Household Registration System and Rural–Urban Educational Inequality in Contemporary China." Report 11-735, Population Studies Center, University of Michigan.

Xiang, Biao. 1996. "Mobility, Traditional Networking, Marketization and the Development of a 'Non-Government' Controlled Space," ("Chuan Tong He Xin Shihui Kong Jin De Sheng Cheng.") *Strategy and Management* (in Chinese), 1998(6).

Yang, Dennis T., Vivian Chen, and Ryan Monarch. 2010. "Rising Wages: Has China Lost Its Global Labor Advantage?" *Pacific Economic Review* 15(4): 482–504.

Young, Alwyn. 2000. "The Razor's Edge: Distortions and Incremental Reform in the People's Republic of China." *Quarterly Journal of Economics* 115(4): 1091–1135.

Zhang, Dandan. 2009. "Assimilation or Dissimilation?—The Labour Market Performance of Rural Migrants in Chinese Cities." Ph.D. Dissertation, Australian National University.

Zhang, Junsen, Yaohui Zhao, Albert Park, and Xiaoqing Song. 2005. "Economic Returns to Schooling in Urban China, 1988 to 2001." *Journal of Comparative Economics* 33(4): 730–52.

Zhang, Xiaobo, and Ravi Kanbur. 2005. "Spatial Inequality in Education and Health Care in China." *China Economic Review* 33(1): 189–204.

Zhang, Zhaowen. 1997. "A Report on the Current Situation and Trends of Development of Literacy Education in China." Asia Literacy Regional Forum, Manila, Philippines. http://www .glp.net/c/document_library/get_file?folderId =12858&name=DLFE-1134.pdf.

Zhao, Yaohui. 2000. "Rural-to-Urban Labor Migration in China: The Past and the Present." In *Rural Labor Flows in China,* edited by L. West and Y. Zhao, 15–33. Berkeley, CA: University of California Press.

Understanding China's Growth: Past, Present, and Future

Xiaodong Zhu

The pace and scale of China's economic transformation have no historical precedent. In 1978, China was one of the poorest countries in the world. The real per capita GDP in China was only one-fortieth of the U.S. level and one-tenth the Brazilian level. Since then, China's real per capita GDP has grown at an average rate exceeding 8 percent per year. As a result, China's real per capita GDP is now almost one-fifth the U.S. level and at the same level as Brazil. This rapid and sustained improvement in average living standard has occurred in a country with more than 20 percent of the world's population so that China is now the second-largest economy in the world.

To set the stage in this paper, I will begin by discussing briefly China's historical growth performance: that is, how China went from the world's leading economic power about 900 years ago to a situation in which it essentially missed the Industrial Revolution and had close-to-zero growth in per capita GDP from 1800 to 1950. I then present growth accounting results for the period from 1952 to 1978 and the period since 1978, using as my starting point a standard growth accounting exercise that decomposes the sources of growth into capital deepening, labor deepening, and productivity growth. For the period from 1952 to 1978, China's per capita GDP did rise by about 3 percent per year, but all of the growth was due to forced increases in government investment as well as a rise in education levels. Productivity actually regressed during this period, as China's economy went through the enormous disruptions of the famine in the late 1950s and the Cultural Revolution starting in the late 1960s. But the main focus of this paper will be to examine the sources of growth since 1978, the year when China started economic reform. Perhaps

■ *Xiaodong Zhu is Professor of Economics, University of Toronto, Toronto, Ontario, Canada. His email address is xzhu@chass.utoronto.ca.*

surprisingly, given China's well-documented sky-high rates of saving and investment, I will argue that China's rapid growth over the last three decades has been driven by productivity growth rather than by capital investment. The growth contributions made by human capital accumulation and an increase in labor participation are positive but modest. I will also examine the contributions of sector-level productivity growth, and of resource reallocation across sectors and across firms within a sector, to aggregate productivity growth. Overall, gradual and persistent institutional change and policy reforms that have reduced distortions and improved economic incentives are the main reasons for the productivity growth.

Despite the rapid growth of the last three decades, China's productivity is still only 13 percent of the U.S. level, which suggests that China still has plenty of room for productivity growth through further economic reforms. Even if China can replicate its extraordinary growth performance for another two decades, its productivity would still be only around 40 percent of the frontier productivity level.

Before delving into the analysis, let me first mention the three main data sources that I use for this paper. For examining China's historical performance, I use the data constructed by Madison (2007); for comparing China with other countries, I use the purchasing power parity data from Penn World Table (PWT7.0); and for detailed growth accounting exercises, I mainly use the data series my coauthor and I constructed for Brandt and Zhu (2010), in which we made adjustments to China's official statistics by using alternative deflators and information from household surveys.

China's Historical Economic Performance

China was a world economic and technological leader in the "premodern" era. Many historians think that China's premodern economic performance reached a peak in the Song Dynasty (circa 1200) when China is though to have had the most advanced technologies (Needham and Ronan 1978), the highest iron output (Hartwell 1962), the highest urbanization rate (Chao 1986), and the largest national economy (Madison 2007) in the world. However, sometime between 1500 and 1800, China lost its leadership position to Western Europe. Figure 1 plots Angus Madison's estimates of per capita GDP for China and Western Europe. According to his estimates, China's per capita GDP stagnated between 1500 and 1800 while Western Europe's per capita GDP increased steadily during the same period. These estimates suggest that, by the end of the fifteenth century, China had already started to fall behind Western Europe, well *before* the Industrial Revolution occurred in England. Some historians and economists attribute China's falling behind during this period to the more centralized and inward-looking political systems of the Ming (1368–1644) and Qing (1644–1911) dynasties that stifled innovation and commercial activities in China.

Not all economic historians agree with this explanation. Kenneth Pomeranz (2000) argues in *The Great Divergence* that in the eighteenth century, living standards and the degree of commercialization in China's Lower Yangzi region were comparable to those in the richest parts of Europe and that China only started

*Figure 1*
Per capita GDP of China and Western Europe
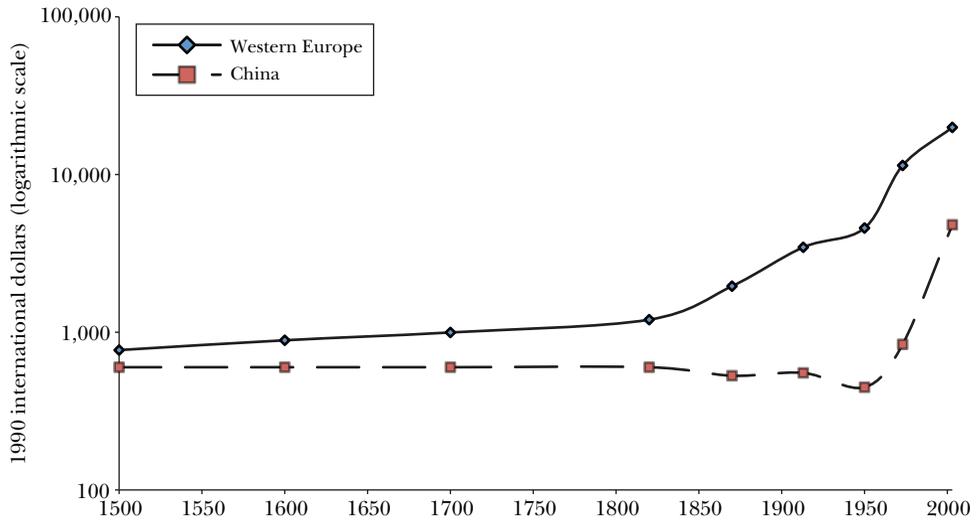
to fall behind Western Europe *after* the Industrial Revolution in England. Shiue and Keller (2007) provide evidence that in the late eighteenth century, the degree of market integration was higher in the Lower Yangzi region than in continental Europe and only slightly lower than that in England. Instead of asking what went wrong in China, Pomeranz attributes the success of the Industrial Revolution to two lucky breaks for England: accesses to coal and colonies.

The questions of why China was not able to maintain its technological lead and the exact time when China started to fall behind Western Europe remain unresolved. There is no doubt, however, about the great divergence in economic performance between China and Western Europe in the nineteenth century and the first half of the twentieth century. Brandt, Ma, and Rawski (2012) review the debates over possible causes and the related literature. They argue that China's economic failure during this time period was due to an imperial political-institutional system that protected vested interests of elite groups—like imperial households, members of bureaucracy, and local gentry—who in turn were resistant to adoptions of new technologies. This imperial system was significantly weakened and eventually collapsed after two Opium Wars between China and Great Britain in the 1840s and 1850s and the Sino-Japanese War of 1894–95. The series of Chinese defeats was in effect a forced opening of China's borders, and it led to territories and treaty ports being conceded to the West and to Japan. These changes brought to China industrial technologies and factories, but continuous civil wars and World War II prevented the industrialization process from gaining much momentum in China until the 1950s. Indeed, industrialization had so little effect during this time that China's per capita GDP *declined* between 1800 and 1950.

Figure 2
GDP per capita of China and US: 1952–2009
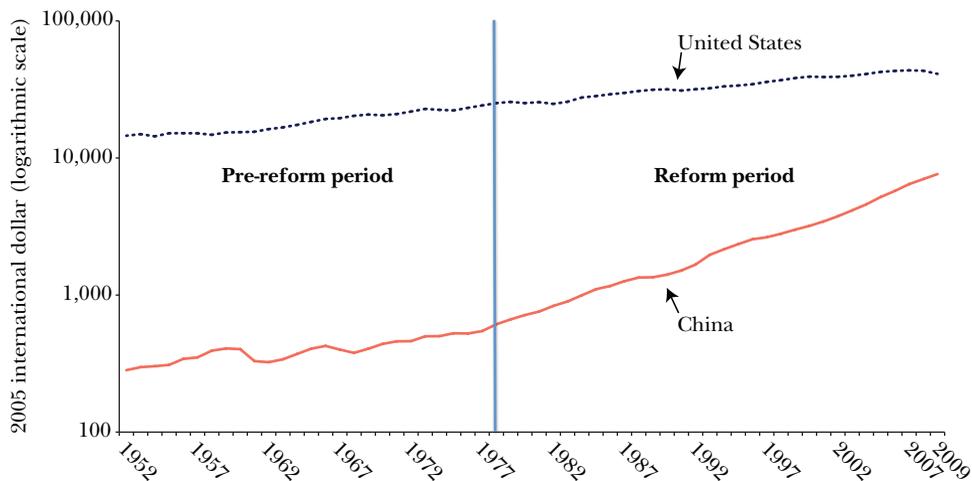
A Growth Accounting Decomposition for Modern China

After the establishment of the People's Republic in October 1949, China finally started its industrialization process in the early 1950s. However, growth performance before and after 1978 differs significantly. Prior to 1978, the average growth rate of real per capita GDP was a modest 3 percent a year, not much different from the growth rate in the United States though starting from a much lower base. Since 1978, China's growth in per capita GDP has accelerated to a rate in excess of 8 percent per year, and Figure 2 shows (on a log scale) how China's per capita GDP has begun to close the gap with U.S. per capita GDP.

Why did China's growth performance differ so much before and after 1978? To answer this question, I begin in this section by using the standard growth accounting method to take a look at the sources of China's growth in both periods, which shows that capital accumulation was the main source of economic growth in the 1952–1978 period while productivity growth has been the main source of growth since then. In the next two sections, I offer more details on these two periods, including why the capital-investment-led growth of the 1952–1978 period was unsustainable and came at such a high cost to the country, and what has been underlying the rapid productivity growth since 1978.

Let the relationship between production inputs (physical capital, human capital, and labor) and GDP be represented by a standard Cobb–Douglas production function:

$$Y = AK^{\alpha}(hL)^{1-\alpha}.$$

Here Y is GDP, K is physical capital stock, L is labor (number of workers), h is the average level of human capital, A is total factor productivity (TFP), and α is the output elasticity of physical capital, which is usually measured by capital's share of national income. Hall and Jones (1999) show how to use this framework to calculate per capita GDP,[1] while Kehoe and Prescott (2002) note that in this framework the growth rate of per capita GDP can be decomposed as the sum of four terms:

Growth rate of per capita GDP = *growth rate of labor participation rate*

$+ \alpha/(1-\alpha)$ *growth rate of the capital/output ratio*

$+$ *growth rate of average human capital*

$+ 1/(1-\alpha)$ *growth rate of total factor productivity.*

Note that in this decomposition the contribution of total factor productivity growth is weighted by $1/(1-\alpha)$, taking into account both the direct contribution of total factor productivity and the indirect contribution through its impact on capital accumulation.

For Table 1, I will set the value of α to $1/2$ (as in Brandt, Hsieh, and Zhu 2008) to match China's average capital income share as reported in China's national accounts. With this assumption in place, Table 1 presents a decomposition of China's per capita GDP growth into contributions from growth of the labor participation rate, the capital/output ratio, average human capital, and total factor productivity.[2]

This decomposition reveals very different patterns of growth in the two periods. In the pre-1978 period, growth was mainly coming from increases in both physical and human capital rather than increases in productive efficiency. Total factor productivity actually deteriorated during this period, declining by 1.07 percent per year. Due to the increases in average schooling years, average human capital grew at 1.55 percent a year, partially offseting the reduction in total factor productivity.

---

[1] Specifically, Hall and Jones (1999) show that in this Cobb–Douglas framework one can express the GDP per capita in the following way:

$$\frac{Y}{Pop} = \frac{L}{Pop}\left(\frac{K}{Y}\right)^{\frac{\alpha}{1-\alpha}} hA^{\frac{1}{1-\alpha}}.$$

In this formulation, *Pop* is the population. GDP per capita can thus be calculated as the product of four terms: the labor participation rate, the capital/output ratio raised to the power of $\alpha/(1-\alpha)$, the average level of human capital, and total factor productivity raised to the power of $1/(1-\alpha)$. The other variables are defined in the text.

[2] The data on GDP per capita, GDP per worker, and labor participation rate are taken from the Penn World Table (PWT7.0). The Penn World Table contains two versions of data for China. I use version 1 because it is more consistent with the series we constructed for Brandt and Zhu (2010) using China's national accounts data, with adjustments made to deflators in a way that is similar to what Alwyn Young (2003) did for the data over a shorter period of time. The physical capital stock data are constructed using the real investment data from the PWT7.0 and the perpetual inventory method with a depreciation rate of 0.06. The initial capital stock in 1952 was set to $I_{52}/(0.06 + \ln(I_{57}/I_{52})/5)$, where I_t is the real investment in year t. The average level of human capital is constructed using the average schooling years reported in the Barro and Lee (2010) dataset and the method of Hall and Jones (1999).

Table 1
Decomposing China's Growth: 1952–2007

Period	Average annual growth rates (%)				
	GDP per capita	Labor partication rate	Capital/output ratio	Average human capital	TFP
1952–1978	2.97	0.11	3.45	1.55	−1.07
1978–2007	8.12	0.57	0.04	1.18	3.16

Period	Contributions to per capita GDP growth				
	GDP per capita	Labor partication rate	Capital/output ratio	Average human capital	TFP
1952–1978	100	3.63	116.15	52.25	−72.03
1978–2007	100	7.05	0.51	14.55	77.89

Source: Authors calculations. The data on GDP per capita, GDP per worker, and labor participation rate are taken from the Penn World Table (PWT7.0). The average level of human capital is constructed using the average schooling years reported in the Barro and Lee (2010) dataset. See footnote 2 for details.
Notes: Table 1 presents a decomposition of China's per capita GDP growth into contributions from growth of labor participation rate, capital/output ratio, average human capital, and total factor productivity. "TFP" is total factor productivity. See text for details.

The labor participation rate increased slightly, growing at 0.11 percent a year. The most important source of growth was increases in the physical capital/output ratio, which on average grew 3.45 percentage points a year and accounted for 116 percent of the per capita GDP growth.

After 1978, capital accumulation and total factor productivity growth reversed their roles. Between 1978 and 2007, the physical capital/output ratio remained roughly constant and the average human capital growth rate was lower than the growth rate in the pre-1978 period. The two sources combined contributed to around 15 percent of the growth in per capita GDP. Demographic factors played a very limited role. Partly due to the one child policy, the labor participation rate grew at 0.57 percent a year during this period, faster than in the pre-1978 period. But the contribution of the increases in labor participation rate was still modest, accounting for only about 7 percent of the growth. In contrast, total factor productivity grew rapidly at 3.16 percent a year. (Bosworth and Collins, 2008, in this journal, and Perkins and Rawski, 2008, report similar results in their growth accounting exercises.) Since the contribution of total factor productivity growth is weighted by $1/(1-\alpha)$ and α is 0.5, the growth contribution of total factor productivity growth is $2 \times 3.16 = 6.32$ percentage points, or 78 percent of the growth in GDP per capita.

The finding that aggregate productivity growth has been the most important source of China's growth since 1978 may seem surprising because it runs in the face of a popular view that China has followed an *investment-driven* growth model that relied heavily on capital-deepening for growth over the last three decades (for

example, Wolf 2011). But, although the share of annual GDP that flows to real fixed capital investment in China increased from 33 percent to 39 percent between 1978 and 2007, China's capital-to-output ratio barely increased during this time. China's capital investment since 1978 has been keeping up with its rapid rate of output growth but not leading it. Examining the data between 1978 and 1998, Young (2003) also comes to the conclusion that capital deepening was *not* the source of China's growth. As Solow (1956) taught us: persistent economic growth can only come from growth in total factor productivity. More than three decades of rapid economic growth in China would not have been possible without significant growth in aggregate total factor productivity.

Government-led Industrialization between 1952 and 1978

After the People's Republic of China was established in 1949, the Chinese Communist Party government, like governments of many other countries at the time, thought the most effective way to speed up the industrialization process was by increasing investment in heavy industries such as steel, concrete, and heavy machinery. China's government mobilized the resources for investment by limiting household consumption and setting low prices for agricultural goods so that forced savings and surpluses extracted from the agricultural sector could be used for investment in such industries.

This strategy of extensive growth based so heavily on capital accumulation was not sustainable and had grave welfare consequences. The big push towards industrialization during the Great Leap Forward years (1958–1960) not only failed to raise the GDP growth rate, it also had such disruptive effects on agricultural production that a severe famine occurred when China was hit by adverse weather shocks in 1959 (Li and Yang 2005). The Great Leap Forward became the Great Leap Famine of 1959–1961, when the official statistics admit to 15 million deaths and unofficial estimates suggest double that number or more.

Despite these disastrous results, the Chinese government continued its unbalanced growth strategy with only minor adjustments after the famine. Unfavorable terms of trade were set on farm products, which effectively imposed heavy taxes on farmers. The *hukou* or household registration system was implemented to keep heavily taxed farmers from leaving rural areas. Furthermore, farmers were prohibited from engaging in any nonfarm activities. These policies initially helped to ensure that the government could extract surpluses from the agricultural sector to support the capital accumulation in the industrial sector. However, they also created incentive problems that significantly reduced the productivity of farmers. As a result, agricultural output grew slowly. In the late 1970s, the agricultural sector included more than 70 percent of China's labor force but was not even able to provide China's population with 2,300 calories per capita per day (near the UN-established minimum). Emergency grain imports were frequently needed to meet food deficits (Huang, Otsuka, and Rozelle 2008). China's nonagricultural sector was little better. It was dominated by the state-owned enterprises in which resource allocation and production activities

were carried out according to government plan rather than market signals. Most of the state-owned enterprises at that time were inefficient, overflowing with redundant workers, and often producing output for which there was no market demand. At the same time, there were very few firms in the light industries like home appliances, furniture, and clothing, and there were constant shortages of consumer products.

Given this background, it may seem paradoxical that China's economy managed an average per capita GDP growth rate of even 3 percent from 1952 to 1978. The main reason for such a gain, as earlier emphasized earlier in Table 1, was the increases in physical and human capital, both of which were at very low levels in 1952. The capital/output ratio rose by about 140 percent during this time, from 0.91 in 1952 to 2.22 in 1978. In addition, average years of education rose from 0.74 in 1952 to 3.75 in 1978. Even with the substantial decline in aggregate productivity, these factors were sufficient to increase China's per capita GDP over this time.

In summary, the industrialization policies pursued by the Chinese government during this period from 1952 to 1978 created adverse incentives and gross misallocation of resources that resulted in declining aggregate productivity, recurring food crises, and relatively little improvement in living standards.

Sectoral Shifts and Productivity Growth Since 1978

When the Cultural Revolution ended after the death of the Communist Party chairman Mao Zedong in 1976, the Chinese government under the leadership of Deng Xiaoping sought to increase its legitimacy by improving aggregate economic performance and raising living standards. In December 1978, the government decided on a general policy of *Gaige Kaifang* or "reform and opening up." Xu (2011) reviews the institutional changes during the reform period in China. There was no grand design of systematic reform policies; instead, economic reforms have taken place in a gradual, experimental, and decentralized fashion. How did the reforms generate such impressive growth? Is the growth sustainable? As a starting point to answering this question, in this section, I look at productivity growth in different sectors and the reallocation of labor across sectors. In the following two sections, I then discuss the key economic reforms and institutional changes that were behind the sector-level productivity growth in agriculture and in the nonstate sector.

Table 2 presents total factor productivity growth rates of the aggregate economy, the agricultural sector, and the nonagricultural sector. Because of the importance of the state sector in the Chinese economy, the nonagricultural sector is divided into the state and the nonstate sectors. The "state sector" includes both state-owned enterprises and shareholding companies; and the "nonstate sector" includes domestic private firms, foreign-invested firms,[3] and collective

---

[3] In China, "foreign-invested" firm is a term used for any one of a number of legal entities with foreign stakeholders, including equity joint ventures, cooperative joint ventures, wholly-owned foreign enterprises, and foreign-invested companies limited by shares.

Table 2

Employment Share, GDP Share, and Total Factor Productivity Growth by Sector

| | | *Average annual total factor productivity growth (%)* | | |
| | | *Nonagricultural sector* | | |
Period	*Agriculture*	*Nonstate*	*State*	*Aggregate*
1978–2007	**4.01**	**3.91**	**1.68**	**3.61**
1978–1988	2.79	5.87	−0.36	3.83
1988–1998	5.10	2.17	0.27	2.45
1998–2007	4.13	3.67	5.50	4.68
Year		*Employment share (%)*		
1978	69	15	16	100
2007	26	62	12	100
Year		*GDP share (%)*		
1978	28	27	45	100
2007	10	70	20	100

Source: Brandt and Zhu (2010).

Notes: Table 2 presents total factor productivity (TFP) growth rates of the aggregate economy, the agricultural sector, and the nonagricultural sector, with the nonagricultural sector divided into state and the nonstate sectors. See text for details on the categorization of firms and enterprises into sectors. Because the TFP growth rates reported in this table are based on China's national accounts data that use domestic prices, they are different from the TFP growth rates reported in Table 1, which are calculated from the Penn World Table data that use international prices.

firms in the nonagricultural sector. We include the shareholding companies in the state sector because many of them are former state-owned enterprises that were restructured into shareholding companies after the mid-1990s but are still controlled by the state. They continue to receive favorable treatment by the state, have easy access to bank credit, and are concentrated in protected industries such as energy and telecommunication. In contrast, the collective firms, including those that are controlled by lower-level governments, receive little support from the state and, like domestic private firms, have difficulties getting bank credit and entering into protected industries. Thus, we include them in the nonstate sector.

The growth rates are reported for the entire period of 1978–2007 and three subperiods. The productivity growth rates are calculated using China's official national accounts data on nominal output and fixed investment, the revised GDP and fixed investment deflators, the revised employment series that is consistent with

China's census data, and the schooling year data of Barro and Lee (2010). In Brandt and Zhu (2010), we offer details on the construction of the data series.[4]

Total factor productivity grew rapidly in both the agricultural and the nonstate sectors. For the overall period from 1978 to 2007, the average annual growth rates of total factor productivity in these two sectors is 4.01 and 3.91 percent, respectively. In contrast, the average growth rate of total factor productivity in the state sector is only 1.68 percent per year. Prior to 1998, in particular, the state sector had very low productivity growth rates. After 1998, though, total factor productivity in the state sector grew rapidly, averaging 5.5 percent annually.

The similarity of productivity growth rates in agriculture and in the nonstate sector are associated with very different movements of these two sectors' employment shares. As reported in Table 2, agriculture's share of total employment declined from 69 percent in 1978 to 26 percent in 2007. The high rate of productivity growth in agriculture helped to push workers away from jobs in agriculture. Conversely, the nonstate sector's share of employment increased from 15 percent in 1978 to 62 percent by 2007. The extraordinary increase in the number of workers in this sector was not sufficient to drive down their productivity. Instead, the growth of the nonstate secotor represents the productivity benefits of a sectoral shift away from the agricultural sector to a sector of the economy that could absorb this labor and still generate rapid productivity growth.

The state sector's share of total employment remained remarkably constant at around 16–17 percent of the total labor force from 1978 until 1997. The restructuring of state enterprises circa 1998 led both to a rise in the rate of productivity growth for this sector and also to a decline in its share of China's labor to 12 percent in 2001—a level where it has remained since.

In the next section, I'll discuss the transformation in agriculture in more depth. In the following section, I'll delve more deeply into productivity growth for the nonstate and state producers in the nonagricultural sector.

Productivity Growth in Agriculture and Structural Transformation

Since China had experienced recurring food crises before 1978, it is not surprising that its economic reform started in the agricultural sector. There were two important reforms. First, the government increased prices for agricultural goods. Second, the previous "collective farming system" was shifted to the "household-responsibility system." Under the new system, each farm household was assigned a fixed quota of grains that the household had to sell to the government at

---

[4] Because the national accounts use domestic prices rather than international prices, these growth rates are not the same as the growth rates calculated from the Penn World Tables. However, the differences are small. For the entire period of 1978–2007, the annual growth rates of GDP per worker and total factor productivity calculated using the Penn World Tables are 7.55 and 3.16 percent, respectively. The corresponding growth rates calculated using China's national accounts data are 7.58 and 3.61 percent.

official prices. However, any extra grain the household produced could be sold at market prices. The reforms were implemented gradually and completed in 1984. Between 1978 and 1984, total factor productivity in the agricultural sector grew 5.62 percent per year. Several studies argue that most of the productivity growth during this period can be attributed to the price and institutional reforms that generated strong positive incentive effects on farmers' efforts and input choices (for example, McMillan, Walley, and Zhu 1989; Lin 1992).

As a result of the productivity growth, China's agricultural output increased by 47 percent during this period. The increase in food availability alleviated China's subsistence food constraint and started a structural transformation that reallocated a large amount of labor from agriculture to industry. From 1978 to 1984, agriculture's share of total employment fell from 69 percent to 50 percent: that is, in just six years, 19 percent of China's labor force—more than 49 million workers—reallocated out of the agricultural sector. Most of the 49 million reallocated workers did not move to urban centers. Instead, they went to work in the rural industrial enterprises set up by township and village-level governments that are called "township and village enterprises" (TVEs).

For the first few years, the price and institutional reforms increased agricultural output mainly by improving incentives without much change in the production technologies being used. However, by about 1984 these static efficiency gains, from workers using the same technology with a much more rewarding set of incentives, were largely exhausted. Both agricultural productivity and structural transformation stagnated in the second half of the 1980s. Starting around 1990, markets for agricultural inputs and outputs were gradually liberalized and government interventions were significantly reduced. Huang, Otsuka, and Rozelle (2008) document extensive market liberalization in China's agricultural sector and state: "aside from restrictions on land ownership, China today may have one of the least distorted domestic agricultural economies in the World." As this market liberalization provided farmers with strong incentives to adopt new technologies, the average annual growth rate of total factor productivity in agriculture reached 5.10 percent between 1988 and 1998, and remained at 4.13 percent between 1998 and 2007. Most of agriculture's growth in total factor productivity after 1990 came from technological progress (Jin, Ma, Huang, Hu, and Rozelle 2010). Structural transformation also resumed after 1990. By 2007, agriculture's share of total employment had been reduced from 46 percent in 1991 to 26 percent in 2007.

How did productivity growth in agriculture contribute to the overall economic growth in China? Since T. W. Schultz (1953)'s pioneering work, economists have long emphasized the role of agriculture in economic development.[5] The standard argument is that productivity growth in agriculture not only contributes to aggregate productivity growth directly, but also indirectly through structural transformation. When agricultural productivity increases, food demand can be met with a smaller

[5] See, for example, Johnston and Mellor (1961), Jorgenson (1961), Schultz (1964), and Timmer (1988). See also Gollin, Parente, and Rogerson (2002), Restuccia, Yang, and Zhu (2008), and Yang and Zhu (2010) for some recent analyses.

number of workers in the agricultural sector than before. As a result, some workers can be reallocated to the nonagricultural sector. Because average labor productivity is generally higher in the nonagricultural sector than in the agricultural sector, the real-location of workers from agriculture contributes positively to aggregate productivity growth. Indeed, in 1978, the average labor productivity in China's nonagricultural sector was six times as high as in the agricultural sector, and therefore one would expect a significant contribution from the labor reallocation. For this reason, Young (2003) suggests that the reforms in the agricultural sector may have been the most important source of China's growth during the first two decades of economic reform.

In Brandt and Zhu (2010), my coauthor and I use a multisector model to quantify this contribution during the period of 1978–2007. We find that, out of the 43 percentage points of the reduction in agriculture's share of employment between 1978 and 2007, total factor productivity growth in agriculture accounts for 39 percentage points or 91 percent of the total reduction. Taking both the direct and indirect effects into account, we find that the contribution of total factor productivity growth in agriculture to aggregate productivity growth is 1.5 percentage points a year for the entire period of 1978 and 2007. However, we also find that the role of agriculture's productivity growth diminishes over time, from a contribution of 2.1 percentage points per year between 1978 and 1988, to a contribution of only 0.6 percentage points per year for the period between 1998 and 2007. There are two reasons for the decline. First, as the economy grew, agriculture's share of value-added decreased, and therefore its direct contribution also diminished. Second, the marginal contribution of reallocation is a decreasing function of the agricultural productivity level. After 20 years of productivity growth, the gain from reallocation naturally declined in the later years. As agriculture's share of employment and value-added continue to decline, the contribution of productivity growth in agriculture to aggregate productivity growth will be even smaller in the future.

Growth outside Agriculture: A Tale of Two Sectors

Before economic reform started in 1978, resource allocation was centrally determined by the government's plan rather than by the market. The state sector dominated nonagricultural activity, accounting for 80 percent of the total urban employment and more than three-quarters of industrial output. The nonstate sector at that time mainly consisted of collective firms. Urban collectives were confined to producing a small number of consumer goods and providing neighborhood services. Rural collectives were only allowed to manufacture producer goods for the agricultural sector.

1978–1988: Rise of the Nonstate Sector
In the early 1980s, encouraged by the success of the rural reforms, the Chinese government started two market reforms in the nonagricultural sector. First, a dual-track system was introduced. State-owned enterprises were still given quotas

on both production inputs and output that transacted at official prices, but they were also allowed to buy inputs and sell output beyond quotas at market prices. Moreover, the non-state-owned enterprises, including collectives (as we discussed earlier), small-scale individual businesses, and foreign-invested firms in the special economic zones, were allowed to enter previously forbidden industries, buying and selling their inputs and outputs at market prices. Second, the central government also devolved economic decision-making powers to lower-level governments and provided them with fiscal incentives. Starting in 1980, a "fiscal contracting system" was implemented that effectively made local governments the "residual claimants" of the enterprises under their control (Qian 1999). As a result, provincial-, city-, and county-level governments controlled most of the state-owned enterprises while the township- and village-level governments controlled the group of rural collective enterprises that became known as the "township and village enterprises."

Under these reforms, the township and village enterprises based on the old rural collectives flourished and led the way to an expansion of the nonstate sector, while the state-owned enterprises did not. The number of township and village enterprises increased from 1,520,000 in 1978 to 18,880,000 in 1988 (National Bureau of Statistics of China 1999). The success of the agricultural reforms made available to these enterprises a large number of local workers, and the dual-track system allowed them to gain access to capital and raw materials from the markets. Between 1978 and 1988, the share of total employment in nonstate enterprises increased from 15 percent to 39 percent. The expansion of employment in the nonstate sector was also accompanied by total factor productivity growth averaging 5.87 percent a year during this period.

The reforms did less for state-owned enterprises. Local governments at county level and above sought to improve the economic performance of the state-owned enterprises under their control by implementing a "managerial responsibility system" that linked managers and workers' income to financial outcomes of the enterprises. The reforms did have some positive effect on productivity. Using a panel data set of 272 industrial state-owned enterprises collected by the Chinese Academy of Social Sciences, Li (1997) estimates that their total factor productivity on average grew at 4.68 percent per year between 1980 and 1989, and that most of the productivity growth could be attributed to stronger incentives, increased market competition, and better allocation of production inputs. Using the same data set, Groves, Naughton, Hong, and McMillan (1994) also report positive incentive effects of the managerial responsibility system on productivity.

While enterprise reforms made industrial state-owned enterprises more efficient, their productivity growth was slower than that of the nonstate enterprises and not fast enough to offset the rising real cost of material inputs. Using more aggregate data on industrial enterprises reported by China's National Statistical Bureau, Jefferson, Rawski, and Zheng (1996) estimate that between 1980 and 1988, the average annual growth rate of total factor productivity was 2.96 percent for state-owned enterprises and 3.66 percent for the nonstate collective enterprises. However, these estimated rates of productivity growth are based on a production function that uses gross output rather than value-added. If the costs of real material

inputs are rising, using gross output rather than value-added may be misleading.[6] In the pre-reform period, prices of material inputs were kept artificially low, and so during the reform period, market prices of material inputs rose significantly faster than output prices. Using the information reported in Jefferson, Rawski, and Zheng, I calculated the growth of total factor productivity for the state and collective industrial enterprises between 1980 and 1988 using value-added, rather than gross output. By this metric, the state-owned firms had annual productivity growth of –1.33 percent, while the nonstate collective enterprises had a growth rate of positive 3.11 percent per year. (There has been no comparable study for the specific state and nonstate enterprises in services because data for such a study are not available.)

In short, the basic lesson is that productivity growth of the nonagricultural sector during this period was mainly due to the rise of the nonstate sector. As Table 2 showed earlier for the 1978–1988 period, the state sector had an average annual growth rate of total factor productivity during this time of –0.36 percent, while the nonstate sectors had annual productivity growth of 5.87 percent.

1988–1998: From Reform without Losers to Inevitable Tradeoffs

The drastic difference in economic performances between the township and village enterprises and the state-owned enterprises may seem implausible; after all, both are enterprises under the control of local governments, albeit at different levels. One reason for the difference is that state-owned enterprises remained under the constraints of government planning for a longer time, unable to sell their products at market prices, although these restrictions diminished over time (Naughton 1995).

But the more important difference is the commitment made by the central government to support employment in the state sector. Remember that employment in the state-owned sector remained essentially constant at about 16 percent of the workforce from 1978 up through 1997. This stability reflected the central government strategy of letting the nonstate sector grow without downsizing the state sector. The strategy had the political benefit of minimizing social instability and reducing resistance to reform. Lau, Qian, and Roland (2000) call it "reform without losers." To avoid laying off workers or shutting down factories, the government usually asked the state-owned banks to bail out loss-making state-owned enterprises. The possibility of bailout created a "soft budget constraint," to use a term common in the literature on centrally planned economies, that further reduced the economic incentives of the state-owned enterprises (Kornai 1980; Qian and Roland 1998; Brandt and Zhu 2001). The lack of exit also eliminated market selection as an important mechanism for improving aggregate productivity in the state sector. In contrast, the central government had no commitment to support employment in the township and village enterprises. While the local governments that ran the township and village enterprises did have political incentives to minimize unemployment and maintain social stability in their communities, these local governments had only weak influence on

---

[6] Specifically, let s_m be the share of material inputs in gross output, then $\Delta \ln(TFP_{\text{value-added}}) = [\Delta \ln(TFP_{\text{gross output}}) - s_m \Delta \ln(\text{real material input cost})]/(1 - s_m)$.

banks. For example, millions of township and village enterprises went bankrupt when there was a general tightening of credit in 1989 (Qian and Xu 1993). Thus, township and village enterprises faced a much tighter budget constraint and stronger market discipline than did the state-owned enterprises.

Unsurprisingly, at least to economists, a "reform without losers" strategy still poses tradeoffs. In the absence of hard budget constraints and market discipline, the state-owned enterprises continued to be outperformed by the nonstate sector. Between 1988 and 1998, the average annual growth rate of total factor productivity in the state sector was only 0.27 percent, while the comparable growth rate of the nonstate sector was 2.17 percent (as shown earlier in Table 2). Faced with increasing competition from the more efficient nonstate firms and without significant productivity growth, the financial condition of the state-owned firms deteriorated. The resources needed to support the state-owned enterprises increased steadily between 1986 and 1993. Nonperforming loans in the state banking system increased rapidly, and the creation of money to make these loans was leading to chronic high inflation (Brandt and Zhu 2000).

By 1994, it had become clear that the strategy of "reform without losers" could no longer be sustained. In 1995, the Chinese government reduced its commitment to stable employment in the state sector. Many small-scale state-owned enterprises were allowed to go bankrupt or be privatized through management buyouts. Between 1995 and 2001, the state sector's share of total employment declined from 17 percent to 12 percent.[7] More diversified ownership forms were also introduced within the state sector. Some of the large-scale state-owned enterprises were converted into shareholding companies, with a majority of shares controlled by the state.

1998–2007: Privatization and Trade Liberalization

The 15th Congress of the Chinese Communist Party held in 1997 was a milestone in China's economic policies. The Congress formally sanctioned ownership reforms of the state-owned firms and also legalized the development of private enterprises. With the reduction of legal barriers, private enterprises grew rapidly. Collective enterprises such as township and village enterprises lost their edge, some were closed and many of them were privatized, also in the form of management buyouts. As part of the lead-up to China's joining the World Trade Organization in 2001, China's government also started to cut tariffs, broadened trade rights, and liberalized its regime for foreign direct investment (Branstetter and Lardy 2008). Between 1998 and 2007, the share of total urban employment in domestic private enterprises and foreign-invested enterprises increased from 8 to 24 percent. The increase in the manufacturing sector was even more pronounced. By 2007, domestic private enterprises alone accounted for 51 percent of total urban employment in the manufacturing sector (National Bureau of Statistics of China, 2008, tables 4–2,

---

[7] China's official employment statistics did not record a reduction in the employment of state-owned enterprises until 1998. The state-owned sector actually started to downsize and lay off workers a few years earlier in 1995.

4–6, 4–13). Song, Storesletten, and Ziliboti (2011) present a model that describes the transformation during this period.

The combination of privatization and trade liberalization had strong effects on productivity growth in both the state and nonstate sectors. Between 1998 and 2007, the average annual total factor productivity growth rates of the state and nonstate sectors were 5.50 percent and 3.67 percent, respectively (as shown in Table 2). After stagnating for much of the first two decades of reform, the state sector finally experienced productivity growth in the last decade.

In the manufacturing sector, productivity growth during this period is even higher. Using data of the China Annual Survey of Industries,[8] Brandt, Van Biesebroeck, and Zhang (2012) estimate that, for the manufacturing sector, the total factor productivity growth rate is 13.4 percent a year. Because even state-owned enterprises were allowed to go bankrupt and exit during this period, reallocation through the process of entry and exit contributed significantly to productivity growth, accounting for 72 percent of the aggregate growth of total factor productivity in the manufacturing sector. Jefferson, Rawski, and Zhang (2008) report similar results. Using the same data, Hsieh and Klenow (2009) examine the contribution of capital and labor reallocation among existing firms to the aggregate total factor productivity growth in the manufacturing sector. They find that between 1998 and 2005, a more efficient allocation within four-digit-level manufacturing industries contributed 2 percentage points per year to aggregate total factor productivity growth in the manufacturing sector, with a significant portion of it coming from the reallocation from state-owned to nonstate enterprises. In short, privatization and trade liberalization reduced barriers to entry and exit, and increased competition, which in turn led to rapid productivity growth in the manufacturing sector by raising within-firm productivity and through reallocation along both the extensive and intensive margins.

However, China's nontradable sectors—primarily construction and services—have faced much less international competition. There have also been significant barriers to entry of private and foreign-invested firms into service industries, and significant barriers to exit of state-owned enterprises in services. In 2007, the state sector still accounted for 77 percent of total urban employment in services, in contrast to 15 percent in manufacturing. It is perhaps not surprising, then, that researchers have found that productivity growth in the nontradable sector lagged behind growth in the tradable sector (for example, He, Zhang, Han, and Wu 2012).

Sources of Aggregate Productivity Growth in China: A Summary

From 1978 to 2007, China's annual growth rate of total factor productivity was 3.61 percent per year. We can summarize the sources of aggregate productivity growth in China during the reform period as follows. In the agricultural sector, productivity

---

[8] This firm-level survey has been conducted annually by China's National Bureau of Statistics. It covers all Chinese industrial firms (manufacturing, mining, and construction) with sales over 5 million renminbi yuan.

growth contributed 1.5 percentage points a year to aggregate productivity growth over the 1978–2007 period, both directly and indirectly through structural transformation. However, this source of growth diminished over time as agriculture's share of GDP diminished, and its contribution to China's future growth will be small. In the nonstate sector, productivity growth contributed 2.27 percentage points per year to aggregate productivity growth over the 1978–2007 period.[9] This source of growth will continue to drive China's future growth as the nonstate sector's share of total nonagricultural employmentiy has risen from 48 percent in 1978 to 84 percent in 2007 (more than 60 percent of total employment). Productivity stagnated in the state sector until the late 1990s, and for the 1978–2007 period as a whole this sector contributed essentially zero to aggregate growth in total factor productivity. However, since 1998, the state sector also experienced rapid productivity growth as a result of restructuring.

The proximate sources of productivity growth have shifted over time. For example, productivity growth in agriculture under the dual-track system led the way from 1978 up to about 1984; starting in the mid-1980s, the nonstate sector in the form of township and village enterprises under its own dual-track system led the way through much of the 1980s and 1990s; and from the late 1990s and into the 2000s, the nonstate sector in the form of privately-owned firms and a restructured state-owned sector led the way in an economic climate much friendlier to the private sector and with lots of entry, exit, and competitive pressures. Whenever the effect of one set of reforms on productivity seemed to be exhausted, the Chinese government found a way to initiate new reforms that reignite growth.

The Future of China's Economic Growth

Experiences from other economies, especially the East Asian economies such as Japan, Korea, and Taiwan, suggest that periods of extremely rapid growth eventually slow down and China's more than 8 percent a year per capita GDP growth rate will not last. China's per capita GDP is now around 20 percent of the U.S. level. Will China's per capita GDP level out at 40 percent of the U.S. level, or 80 percent, or 120 percent? Of course, any answer to this question will contain a large dose of speculation. But I will attempt to address this question by discussing what would be the key sources of China's growth in the future based on what we know about the sources of China's growth in the last three decades.

Following the earlier decomposition of the sources of economic growth, we can decompose China's GDP per capita relative to that of the United States into

---

[9] In Brandt and Zhu (2010), we estimate that, if there were no total factor productivity growth in the nonstate sector, the producutvitiy growth rate for the nonagricultural sector would have been close to zero for the entire period between 1978 and 2007 and during each of the three sub-periods. I should also note that productivity growth in different sectors may interact so that one cannot infer the contribution of productivity growth in the state sector by simply substracting the contributions of agriculture and nonstate sector from the aggregate productivity growth.

Figure 3
China Relative to the Uniteed States
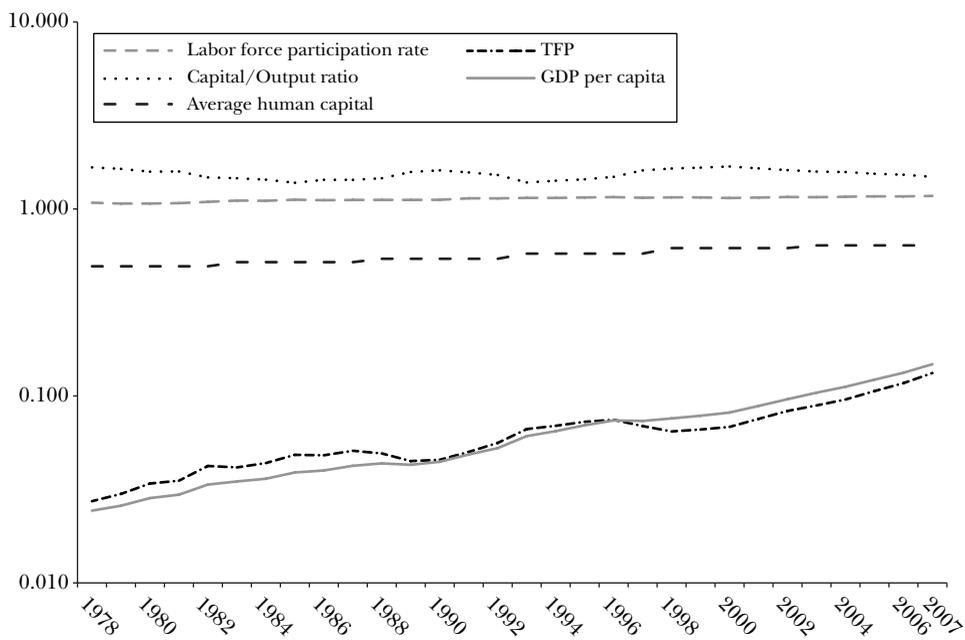
Note: "TFP" is total factor productivity.

four ingredients: relative labor participation rate, relative average human capital, relative capital/output ratio, and relative total factor productivity.[10] Figure 3 plots these ratios for the period between 1978 and 2007. China's labor force participation and capital/output ratios are above U.S. levels, while China's relative level of human capital has risen somewhat over time (notice that Figure 3 is on a log scale). But clearly, the growth of China's relative GDP per capita is mainly driven by the growth of China's relative total factor productivity. To answer the question about China's future growth, then, one has to assess the future of China's relative productivity growth.

Although economic reforms have been crucial in generating productivity growth in China over the last three decades, many other economies in Eastern Europe and Latin America also had economic reforms, but their growth performances are nowhere near the performance achieved by China. What is special about China? One potential explanation is simply China's backwardness at the start of economic reform in 1978, which increased China's potential for catch-up growth.

---

[10] To be precise, the last two ratios should be the relative capital/output ratio raised to the power of $\alpha/(1-\alpha)$, and relative total factor productivity raised to the power of $1/(1-\alpha)$. For simplicity, however, I will simply refer them as "relative capital/output ratio" and "relative total factor productivity."

When China started economic reform in 1978, its aggregate total factor productivity was less than 3 percent of the U.S. level, much lower than Mexico and the economies in Eastern Europe and South America. Because China was far away from the frontier, the impact of reforms in closing the productivity gap has been particularly large. Parente and Prescott (1994) present a model along these lines, and Kehoe and Ruhl (2010) suggest that this argument may explain why economic reforms have produced rapid growth in China, but less growth in Mexico.

Has China's productivity gap now been narrowed enough so that China will find it difficult to generate further productivity growth? Compare China's growth experience with three other East Asian economies that also had rapid and sustained reductions in their productivity gaps with the U.S. economy but eventually experienced significant slowdown in relative productivity growth: Japan, Korea, and Taiwan. In 1950, Japan's total factor productivity was 56 percent of the U.S. level; by 1975, Japan's was at 83 percent of the U.S. level. But since then, Japan's relative total factor productivity has somewhat fallen back. In 1965, Korea's total factor productivity was 43 percent of the U.S. level; by 1990, it had reached 63 percent of the U.S. level. After 1990, Korea's relative productivity has continued to converge with the U.S. level, but at a much slower rate of about 0.24 percent per year. In 1965, Taiwan's total factor productivity was 50 percent of the U.S. level; by 1990, it had reached 80 percent of that in the United States. Since then, Taiwan's relative total factor productivity has continued to converge, but (like Korea) at a much slower rate.

Back in 1978, China was starting at a far lower level of productivity than these comparison countries: indeed, from 1978 to 2007, after three decades of rapid productivity growth, China's total factor productivity had risen from 3 percent to 13 percent of the U.S. level. Even if China can replicate this extraordinary growth performance for another two decades, its productivity level would still be only 40 percent of the frontier U.S. level—still below the level of Japan in the 1950s or South Korea and Taiwan in the 1960s. In Japan, South Korea, and Taiwan, relative total factor productivity grew rapidly for a sustained period of time and did not slow down until after the relative productivity had reached 60 percent or higher.

China's economy still has large opportunities for raising productivity growth through reducing the still-existing distortions and inefficiencies in its production. For example, Hsieh and Klenow (2009) use firm-level data to estimate within-industry misallocation of capital and labor across existing firms in China's manufacturing industries. They find a reduction in distortions between 1998 and 2005, but they still estimate a potential total factor productivity gain of 30 percent for China's manufacturing sector if the distortions are reduced to the U.S. level. Song and Wu (2011) find a very similar gain using a different dataset and method. In Brandt, Tombe, and Zhu (2012), my coauthors and I take sector-level total factor productivity in each province as given and measure the potential productivity gain from eliminating factor market distortions across provinces and between the state and the nonstate sectors in China. We find the potential total factor productivity gain in China's nonagricultural economy to be

at least 20 percent in our estimates, in which half the gain comes from eliminating cross-province dispersion in returns to labor and the other half comes from eliminating within-province difference in returns to capital between the state and the nonstate sectors.

While these potential efficiency gains are substantial, many obstacles exist that may prevent these gains from being realized. Despite many years of financial sector reforms, China's banking sector is still dominated by the state-controlled banks that lend disproportionately to local government projects and to firms in the state sector. Protected by barriers to entry of private and foreign firms, state-controlled firms continue to enjoy substantial monitoring rights and profits in industries ranging from energy, transportation, and telecomunication to banking, entertainment, education, and health care. Further institutional change and policy reforms will be needed if China is to maintain its productivity growth by reducing these distortions.

In the last three and half decades, China's leaders have chosen to carry out economic reform without political reform or the establishment of rule of law. Instead, they have implemented institutional changes and policy reforms in a piecemeal fashion that usually provided benefits to key interest groups within the state sector. Giving monopoly rights to state-controlled or politically connected firms is one example. While this approach has helped to reduce political resistence to economic reform, it has also resulted in corruption and income inequality in addition to economic distortions. If reducing the state sector's monopoly rights in various industries is important for reducing distortions and solving associated social-political problems of corruption and income inequality, it remains to be seen if China's leadership will be flexible enough and strong enough to do so.

Finally, I conclude by noting that I have only considered the direct contributions of human capital accumulation and demographic factors on GDP growth. It is possible that the increases in average years of education and the decreases in the dependence ratio due to the one-child policy have also reduced the cost of migration, facilitated the reallocation of labor away from agriculture, and therefore contributed positively to aggregate total factor productivity growth. If that is the case, I may have underestimated the growth contribution of the demographic factors and human capital accumulation. However, given that the marginal gains from labor reallocation have been decreasing over time, the contributions of these factors to productivity growth should also decline in the future.

References

Barro, Robert J., and Jong-Wha Lee. 2010. "A New Data Set of Educational Attainment in the World, 1950–2010." NBER Working Paper 15902.

Bosworth, Barry, and Susan M. Collins. 2008. "Accounting for Growth: Comparing China and India." *Journal of Economic Perspectives* 22(1): 45–66.

Brandt, Loren, Chang-Tai Hsieh, and Xiaodong Zhu. 2008. "Growth and Structural Transformation in China." In *China's Great Economic Transformation,* edited by Loren Brandt and Thomas G. Rawski, 683–728. Cambridge University Press.

Brandt, Loren, Debin Ma, and Thomas Rawski. 2012. "From Divergence to Convergence: Re-evaluating the History behind China's Economic Boom." Economic History Working Papers 41660, London School of Economics and Political Science, Department of Economic History.

Brandt, Loren, Trevor Tombe, and Xiaodong Zhu. 2012. "Factor Market Distortions across Time, Space and Sectors in China." http://www.economics.utoronto.ca/xzhu/paper/BTZ0912.pdf.

Brandt, Loren, Johannes Van Biesebroeck, and Yifan Zhang. 2012. "Creative Accounting or Creative Destruction? Firm-level Productivity Growth in Chinese Manufacturing." *Journal of Development Economics* 97(2): 339–51.

Brandt, Loren, and Xiaodong Zhu. 2000. "Redistribution in a Decentralized Economy: Growth and Inflation in China under Reform." *Journal of Political Economy* 108(2): 422–39.

Brandt, Loren, and Xiaodong Zhu. 2001. "Soft Budget Constraint and Inflation Cycles: A Positive Model of Macro-Dynamics in China during Transition." *Journal of Development Economics* 64(2): 437–57.

Brandt, Loren, and Xiaodong Zhu. 2010. "Accounting for China's Growth." Working Papers tecipa-394, University of Toronto, Department of Economics.

Branstetter, Lee, and Nicholas R. Lardy. 2008. "China's Embrace of Globalization." In *China's Great Economic Transformation,* edited by Loren Brandt and Thomas G. Rawski, pp. 633–82. Cambridge University Press.

Chao, Kang. 1986. *Man and Land in Chinese History: An Economic Analysis.* Stanford University Press.

Gollin, Douglas, Stephen L. Parente, and Richard Rogerson. 2002. "The Role of Agriculture in Development." *American Economic Review* 92(2): 160–64.

Groves, Theodore, Barry Naughton, Yongmiao Hong, and John McMillan. 1994. "Autonomy and Incentives in Chinese State Enterprises." *Quarterly Journal of Economics* 109(1): 183–209.

Hall, Robert E., and Charles I. Jones. 1999. "Why Do Some Countries Produce So Much More Output per Worker Than Others?" *Quarterly Journal of Economics* 114(1): 83–116.

Hartwell, Robert. 1962. "A Revolution in the Chinese Iron and Coal Industries during the Northern Sung, 960–1126 A.D." *Journal of Asian Studies* 21(2): 153–62.

He, Dong, Wenlang Zhang, Gaofeng Han, and Tommy Wu. 2012. "Productivity Growth of the Non-Tradable Sectors in China." Working Papers 082012, Hong Kong Institute for Monetary Research.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics* 74(4): 1403–48.

Huang, Jikun, Keijiro Otsuka, and Scott Rozelle. 2008. "Agriculture in China's Development: Past Disappointment, Recent Successes, and Future Challenges." In *China's Great Economic Transformation,* edited by Loren Brandt and Thomas G. Rawski, 467–505. Cambridge University Press.

Johnston, Bruce F., and John W. Mellor. 1961. "The Role of Agriculture in Economic Development." *American Economic Review* 51(4): 566–93.

Jorgenson, Dale W. 1961. "The Development of a Dual Economy." *Economic Journal* 71(282): 309–34.

Jefferson, Gary H., Thomas G. Rawski, and Yifan Zhang. 2008. "Productivity Growth and Convergence across China's Industrial Economy." *Journal of Chinese Economic and Business Studies* 6(2): 121–40.

Jefferson, Gary H., Thomas G. Rawski, and Yuxin Zheng. 1996. "Chinese Industrial Productivity: Trends, Measurement Issues, and Recent Developments." *Journal of Comparative Economics* 23(2): 146–80.

Jin, Songqing, Hengyun Ma, Jikun Huang, Ruifa Hu, and Scott Rozelle. 2010. "Productivity, Efficiency and Technical Change: Measuring the Performance of China's Transforming Agriculture." *Journal of Productivity Analysis* 33(3): 191–207.

Kehoe, Timothy J., and Edward C. Prescott. 2002. "Great Depressions of the 20th Century." *Review of Economic Dynamics* 5(1): 1–18.

Kehoe, Timothy J., and Kim Ruhl. 2010. "Why Have Economic Reforms in Mexico Not Generated Growth?" *Journal of Economic Literature* 48(4): 1005–27.

Kornai, Janos. 1980. *The Economics of Shortage.* London and New York: Elsevier Science.

Lau, Lawrence J., Yingyi Qian, and Gerard Roland. 2000. "Reform without Losers: An Interpretation of China's Dual-Track Approach to Transition." *Journal of Political Economy* 108(1): 120–43.

Li, Wei. 1997. "The Impact of Economic Reform on the Performance of Chinese State Enterprises, 1980–1989." *Journal of Political Economy* 105(5): 1080–1106.

Li, Wei, and Dennis Tao Yang. 2005. "The Great Leap Forward: Anatomy of a Central Planning Disaster." *Journal of Political Economy* 113(4): 840–77.

Lin, Justin Yifu. 1992. "Rural Reforms and Agricultural Growth in China." *American Economic Review* 82(1): 34–51.

Madison, Angus. 2007. *Contours of the World Economy, 1–2030 AD.* Oxford and New York: Oxford University Press.

McMillan, John, John Whalley, and Lijing Zhu. 1989. "The Impact of China's Economic Reforms on Agricultural Productivity Growth." *Journal of Political Economy* 97(4): 781–807.

Naughton, Barry. 1995. *Growing out of the Plan: Chinese Economic Reform 1978–93.* Cambridge: MIT Press.

National Bureau of Statistics of China. 1999. *Rural Statistical Yearbook of China.* Beijing: China Statistics Press.

National Bureau of Statistics of China. 2008. *China Statistical Yearbook.* Bejing: China Statistics Press.

Needham, Joseph, and Colin A. Ronan. 1978. *The Shorter Science and Civilisation in China: An Abridgement of Joseph Needham's Original Text.* Cambridge and New York: Cambridge University Press.

Parente, Stephen L., and Edward C. Prescott. 1994. "Barriers to Technology Adoption and Development." *Journal of Political Economy* 102(2): 298–321.

Parente, Stephen L., and Edward C. Prescott. 1999. "Monopoly Rights: A Barrier to Riches," *American Economic Review* 89(5): 1216–33.

Perkins, Dwight H., and Thomas G. Rawski. 2008. "Forecasting China's Economic Growth to 2025." In *China's Great Economic Transformation,* edited by Loren Brandt and Thomas G. Rawski, 829–86. Cambridge University Press.

Pomeranz, Kenneth. 2000. *The Great Divergence: China, Europe, and the Making of the Modern World Economy.* Princeton University Press.

Qian, Yingyi. 1999. "The Institutional Foundations of China's Market Transition." Paper presented at the World Bank's Annual Conference on Development Economics, Washington, DC.

Qian, Yingyi, and Gerard Roland. 1998. "Federalism and the Soft Budget Constraint." *American Economic Review* 88(5): 1143–62.

Qian, Yingyi, and Chenggang Xu. 1993. "Why China's Economic Reforms Differ: The M-Form Hierarchy and Entry/Expansion of the Non-state Sector." *Economics of Transition* 1(2): 135–70.

Restuccia, Diego, Dennis Tao Yang, and Xiaodong Zhu. 2008. "Agriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis." *Journal of Monetary Economics* 55(2): 234–50.

Schultz, Theodore W. 1953. *The Economic Organization of Agriculture.* New York: McGraw-Hill Book Company, Inc.

Schultz, Theodore W. 1964. *Transforming Traditional Agriculture.* Chicago: University of Chicago Press.

Shiue, Carol H., and Wolfgang Keller. 2007. "Markets in China and Europe on the Eve of the Industrial Revolution." *American Economic Review* 97(4): 1189–1216.

Solow, Robert. 1956. "A Contribution to the Theory of Growth." *Quarterly Journal of Economics* 71(1): 65–94.

Song, Michael Zheng, Kjetil Storesletten, and Fabrizio Zilibotti. 2011. "Growing Like China." *American Economy Review* 101(1): 202–241.

Song, Michael Zheng, and Laura Wu. 2011. "A Structural Estimation on Capital Market Distortions."

Timmer, Peter C. 1988. "The Agricultural Transformation." Chap. 8 *Handbook of Development Economics,* Vol.1, Part II, edited by Hollis Chenery and T. N Srinivasan, 276–331. New York: North-Holland.

Wolf, Martin. 2011. "How China Could Yet Fail Like Japan." *Financial Times,* June 14.

Xu, Chenggang. 2011. "The Fundamental Institutions of China's Reforms and Development." *Journal of Economic Literature* 49(4): 1076–1151.

Yang, Dennis Tao, and Xiaodong Zhu. 2010. "Modernization of Agriculture and Long-Term Growth." University of Toronto Working Paper 369.

Young, Alwyn. 2003. "Gold into Base Metals: Productivity Growth in the People's Republic of China during the Reform Period." *Journal of Political Economy* 111(1): 1220–61.

Aggregate Savings and External Imbalances in China

Dennis Tao Yang

The high savings and investment rates in China have been a major driving force behind its rapid economic growth. During the 1980s and 1990s, China's high savings rates in the range of 35–40 percent of GDP were not accompanied by external imbalances; its current account balance fluctuated within 2 percent of GDP in most of the years. However, starting around 2001, China's already high savings rate soared further, and the current account surplus also rose along a steep trajectory. In 2008, China's aggregate savings rate reached 53 percent of GDP, whereas the current account surplus exceeded 9 percent of GDP. Although the current account surplus moderated during the financial crisis, it remained at a lofty 5.2 percent of GDP in 2010. With the accumulation of the annual current account surplus and net inflows of capital, the foreign exchange reserves of China climbed to an unprecedented level, topping $3 trillion in March 2011. This total is nearly triple the amount held by Japan, the second largest holder of foreign reserves in the world.

This paper presents an explanation for the evolving macroeconomic imbalances in China. I argue that the extraordinarily high savings rate and current account surpluses are primarily attributable to a set of policies, institutions, and structural distortions embedded in the Chinese economy. When China joined the World Trade Organization in 2001, its business climate improved and trade barriers fell dramatically, increasing the profitability of firms. However, due to a set of institutional rules that centered on export promotion and that favored firms and government over the household sector, a high percentage of this windfall gain of profits was either saved in the corporate sector or was collected by the government,

■ *Dennis Tao Yang is Professor of Business Administration, Darden School of Business, University of Virginia, Virginia. His email address is YangD@darden.virginia.edu.*

which has not accordingly adjusted its social welfare spending upward. The result was an extraordinary upsurge in aggregate savings, along with weak domestic consumption and anemic demand for imported goods. The imbalance was made even worse by a rise in household savings, which was due to structural shifts in the labor market, incomplete social welfare reforms, and demographic changes resulting from population control policies—as well as by stringent pursuit of policies of export promotion. When the ill-functioning financial system of China failed to channel the increased savings to high-return production investment or consumption loans, the excess savings ended up as huge foreign exchange reserves invested in low-yielding overseas government bonds.

The macroeconomic imbalances in China pose a number of risks. The country's extraordinarily low consumption-to-GDP ratio implies that the rapid economic growth in China over the past decade has been mainly propelled by domestic investment and foreign demand. Thus, sustaining growth amid declining investment efficiency or negative shocks in external demand is challenging. With foreign exchange reserves heavily invested in U.S. dollar- and euro-denominated bonds, China faces potentially enormous capital losses if the dollar and euro depreciate.[1] Moreover, as a fast-growing developing country, standard economic theory suggests that China should post a net inflow of foreign investment; instead, its huge savings are invested abroad earning low returns. From a global perspective, the enormous trade surplus of China is mirrored by equally large trade deficits elsewhere. Several major trading partners of China are upset, and their politicians and analysts have blamed China for contributing to the failure of domestic firms and job losses. Ultra-large trade deficits also are associated with economic and financial disruptions, which present threats to global macroeconomic stability.

This paper begins by documenting the trends in the balance of payments of China, including dramatic changes in the current account balance and net foreign asset position, and the significant buildup of foreign exchange reserves. I then present the corresponding changes in national savings and investment. With these facts in evidence, I propose a unified framework for understanding the joint causes of the high savings rate and external imbalances in China.[2] My explanations first focus on an array of factors that encouraged saving across the corporate, government, and household sectors, such as policies that affected sectoral income distribution, along with factors like incomplete social welfare reforms, and population control policies. I then turn to policies that limited investment in China, thus preventing the high savings from being used domestically. Finally, I will examine how trade policies, such

---

[1] China held about two-thirds of its foreign exchange reserves in the U.S. dollar and more than one-fifth in the euro in 2007. These investments earned an average 3 percent annual rate of return (Sheng 2011).
[2] Despite a general awareness of the internal and external linkages, academic and policy research often focus on either the high savings in China or on the trade surplus and exchange rate policies. For example, Ma and Wang (2010), and Yang, Zhang, and Zhou (2012) recently conducted two surveys on the high savings rate in China. See Goldstein and Lardy (2009), and Corden (2009) for analyses of the current account surplus and exchange rate policies of China. Wen (2011) and Du and Wei (2012) are two recent exceptions that study the linkages between internal and external imbalances in China.

as export tax rebates, special economic zones, and exchange rate policies, strongly promote exports. In conclusion, I recommend some policy reforms for rebalancing the Chinese economy.

Trends in Trade Accounts and National Savings

Balance of payments statistics report all cross-border flows of value between a country and the rest of the world over a period of time. To document the external imbalances of China, the categories of flows are classified as current account, foreign direct investment, portfolio-and-other investment, official foreign exchange reserves, and a remaining statistical discrepancy. The sum of these components is necessarily zero.

Figure 1A shows the current account balance of China from 1985 to 2010. As noted earlier, the trade imbalance was insignificant through the mid 1990s. In fact, trade deficits were recorded in some years. However, from 2001 onwards, the surplus rose along a steep trajectory, accelerating further in 2005 and reaching 10.1 percent of GDP in 2007. The other line in the figure shows that the surplus in trade of goods and services is almost identical to the current account surplus. Therefore, these two terms are used interchangeably in subsequent discussions.

Figure 1B shows that the net capital and financial account, which consists of assets and liabilities of foreign direct investment as well as portfolio and other investment, exhibits similar patterns as the trade account. Although a rise in the surplus occurred in the mid 1990s, the capital and financial account was more or less balanced before China joined the World Trade Organization in 2001. After that, however, the financial and capital account registered a sharp rise in the surplus. Figure 1B also shows that China has experienced a continued net inflow of foreign direct investment since the mid 1990s, and in fact has become the second largest recipient of foreign direct investment after the United States. In the aftermath of the financial crisis of 2008, both the net accounts of foreign direct investment and portfolio/other investment for China were in positive territory, summing to a surplus of 4 percent of GDP.

The persistent "twin surpluses" in the current and capital accounts in the past decade have resulted in an explosion in foreign exchange reserves. In 2000, China only had $10.9 billion in reserves, equivalent to 0.9 percent of GDP. The subsequent rise in currency reserves was astonishing. In 2004 alone, as Figure 1C shows, the yearly accumulation jumped to 10.7 percent of GDP. After reaching a peak of 13.2 percent of GDP in 2008, it hovered at around 8 percent of GDP in 2010. As a result, China's foreign exchange reserves topped $1 trillion in 2006. By June 2011, China's total foreign exchange reserves topped $3.2 trillion, which was approximately triple the amount held by Japan. The reserve assets of China are mostly invested in low-return U.S. government bonds.

While balance-of-payment statistics capture the cross-border flows of value in trade and capital, net foreign asset statistics provide the stock position of the

Figure 1
International Balance of Payments of China: 1985–2010

A: Current account
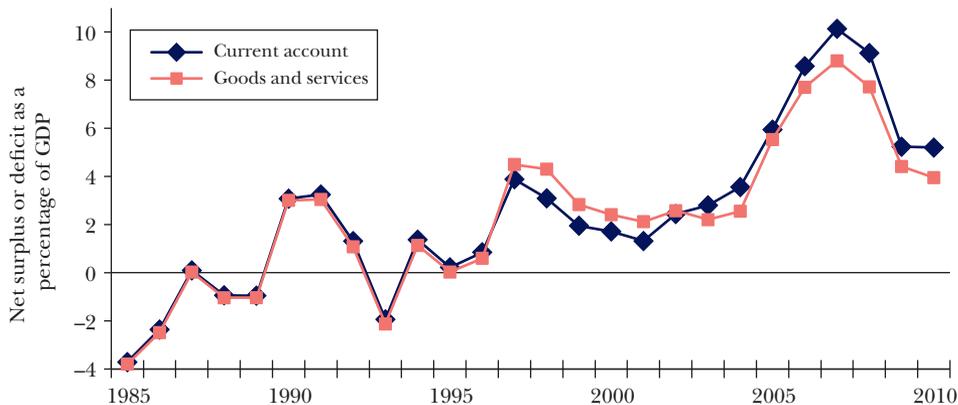
B: Capital and financial account
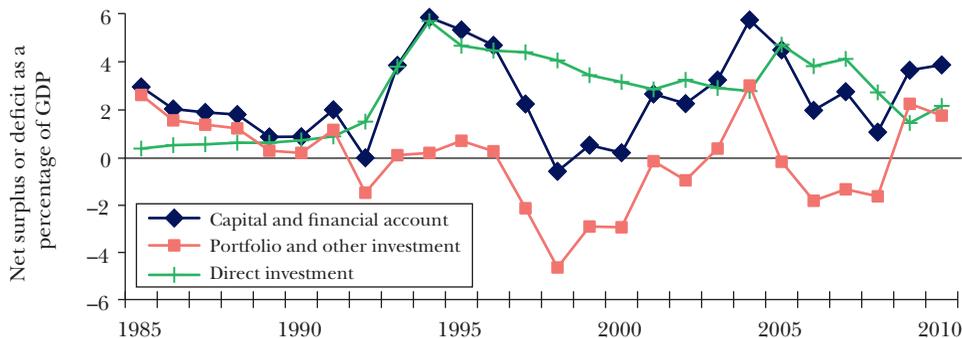
C: Foreign reserve and statistical discrepancy
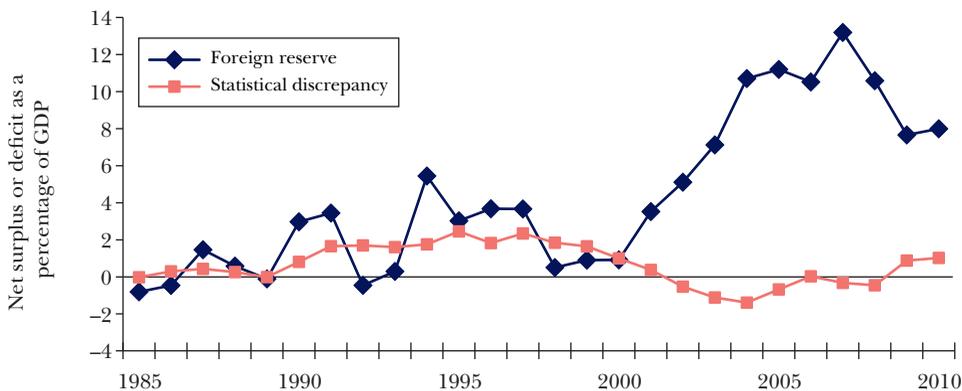
Source: State Administration of Foreign Exchange of China (SAFE 2011).

Table 1
Global Current Account Balances
(billions of U.S. dollars)

Country or region	1995	2000	2005	2008	2010
Advanced economies:	29.8	−270.6	−411.2	−471.8	−95.5
Japan	111.4	119.6	165.7	157.1	194.8
United States	−113.6	−416.4	−747.6	−668.9	−470.2
Euro area	70.5	−39.4	41.1	−86.7	11.6
Germany	−29.6	−32.6	142.8	245.7	176.1
Spain	−1.8	−23.1	−83.3	−156.0	−63.3
Other	−38.5	65.6	129.7	126.7	168.4
Norway	5.3	25.3	49.1	79.9	53.3
Australia	−18.4	−15.3	−41.7	−47.2	−31.7
Emerging and developing economies	−92.2	95.2	443.0	704.2	378.1
Asia	−36.9	41.7	167.5	435.9	308.1
China	1.6	20.5	160.8	436.1	306.2
India	−5.6	−4.6	−10.3	−24.9	−49.0
Middle East and North Africa	−1.2	80.4	212.7	343.1	152.8
Sub-Saharan Africa	−9.9	2.1	−3.4	0.0	−24.9
Latin America and the Caribbean	−37.9	−48.4	36.3	−31.2	−56.9
Central and Eastern Europe	−10.2	−28.9	−57.7	−151.3	−76.0
Former Soviet Union	3.8	48.3	87.6	107.7	75.0
Statistical discrepancy	−62.4	−175.4	31.8	232.4	282.6

Source: International Monetary Fund, *World Economic Outlook Database* (IMF 2011).

economy in external financial assets and liabilities. A current account surplus translates into an increase in net foreign assets, while a current account deficit translates into a decrease in the net foreign asset position. Adopting an approach similar to that of Lane and Milesi–Ferretti (2007), Ma and Zhou (2009) document the emergence of China as a large and rising creditor in the world. In only 10 years, the net foreign asset position of China rose so substantially that the country swung from being a net debtor of approximately 6.2 percent of GDP in 2000 to a net creditor of approximately 30.5 percent of GDP in 2010 (SAFE 2011). Foreign exchange reserves account for a lion's share of the net foreign assets, reaching 69 percent in 2010.

Economists recognize that both trade deficits and trade surpluses can bring economic gains depending on the situation of an economy. However, large and sustained external imbalances can also be a prelude to economic adjustments that may be wrenching. Table 1 offers a global perspective on the evolution of these imbalances. For example, back in 1995, the "advanced economies" as a group were running trade surpluses, while the "emerging and developing economies" as a group had trade deficits. By 2000, these positions had reversed. From 2000 up through the years prior to the financial crisis in 2008, the current account deficit of the advanced economies grew enormously, as did the current account surpluses

of the emerging and developing economies. By 2008, China emerged as the largest net lender with a surplus of $436.1 billion, which is equivalent to 24.3 percent of the total trade surplus for all countries running surpluses. Germany followed a similar path, swinging from having a current account deficit of $32.6 billion in 2000 to having a surplus of $245.7 billion in 2008. The sum of the current account surpluses of these two countries roughly equals the huge current account deficit of the U.S. economy at $689 billion in 2008. In the aftermath of the financial crisis, China continued to have the largest current account surplus among all countries as of 2010.

The well-known national income identity helps us explore potential connections from domestic savings to the external trade balance by stating the accounting relationship between national public and private savings (S), domestic capital formation (I), and the current account balance ($X-M$):

$$S - I = X - M.$$

This identity clarifies a straightforward interpretation of the situation in China: the amount earned by the trade surplus that is not consumed or invested must end up being saved. Moreover, the gap between savings and investment equals the net flow of foreign investment over time: that is, national savings not invested at home are invested abroad. This equation can help us understand how internal and external imbalances have evolved in China.

Figure 2 presents the trends in aggregate savings and investment in China from 1992 to 2008, complementing the balance of payments statistics presented in Figure 1.[3] The data on flow of funds accounts contain the composition of gross domestic savings and investment by household, business, and government, and also information on income and expenditures within each of these sectors.

Figure 2A shows that the national savings of China moved in near-lockstep with aggregate investment in the late 1990s. However, from 2000 onwards, the aggregate savings rate increased along a steep trajectory, moving above the rate of investment. After 2004, the national savings rate exhibited a strong upsurge at approximately two additional percentage points of GDP per year for four consecutive years, reaching a new height of 53.2 percent in 2008, whereas the investment rate plateaued to 42 to 44 percent of GDP. The result was a huge gap between savings and investments—and a corresponding increase in the current account surplus—from 2005 to 2008.

Figures 2B and 2C present disaggregated information on savings and investment by corporate, household, and government sectors. On the investment side, of the 8.9 total percentage point increase in the domestic investment rate

---

[3] In 1995, the National Bureau of Statistics (NBS) of China began to publish the Flow of Funds Accounts based on the physical transactions of national income accounting, covering the government, corporate, and household sectors. Because of a three-year-lag policy, the most recent data available for analysis cover 1992 to 2008.

Figure 2
Savings and Investment in China: 1992–2008

A: Aggregate savings and investment rate
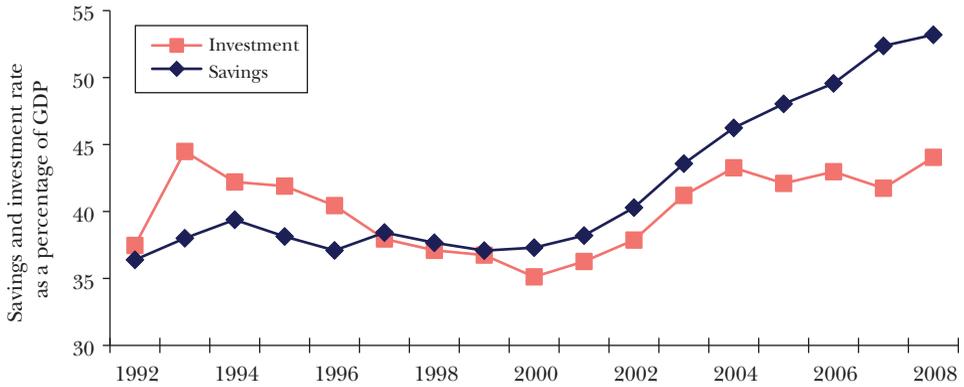
B: Investment rate by sector
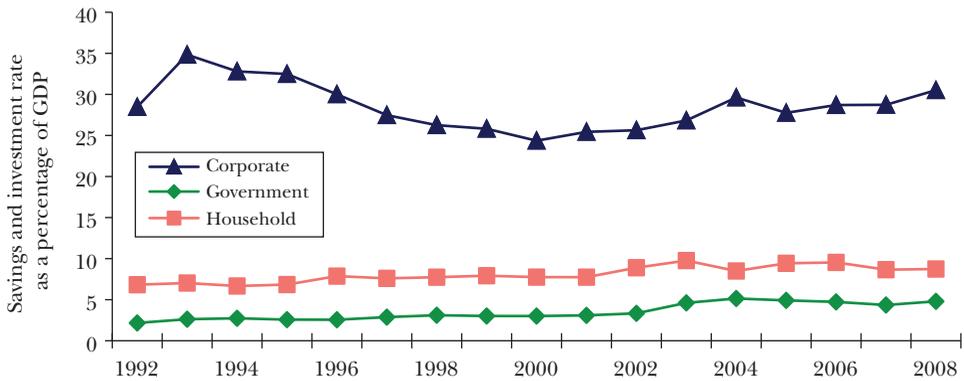
C: Savings rate by sector
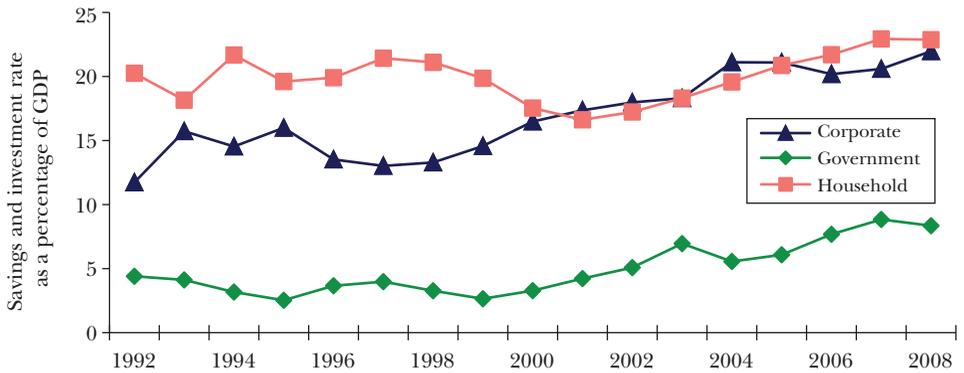
Source: Statistical Yearbook of China (NBS 2009).

(as a percentage of GDP) from 2000 to 2008, the corporate sector contributed 6.1 percentage points. On the savings side, all three sectors contributed significantly, and rather evenly, to the 15.9 percentage-point total increase in the national savings rate during the period.

The explosion in external imbalances and the widening of the savings–investment gap starting in 2000 caught the Chinese government off guard. The policy target of the 11th Five-Year Plan was to achieve a balanced current account from 2006 to 2010, which differs drastically from the realized outcome. Because much is at stake with China's enormous trade surpluses, a number of studies have attempted to understand their causes. One common view is that exchange rate intervention by the Chinese government is the culprit for the severe trade surplus. Other possible causal factors include financial market imperfections, the migration of processing trade into China because of the global division of labor, and the pursuit of export-led development strategies (for example, Yu 2007; Goldstein and Lardy 2009; Song, Storesletten, and Zilibotti 2011). In his "saving glut" hypothesis, Federal Reserve Chairman Bernanke (2005) emphasizes how the changes in desired savings and investment in a region, like higher desired rates of savings in China, affect the external balances of this region and those of other countries around the world. Governor Zhou (2009) of the People's Bank of China highlights the role of a high savings rate in affecting the current account surplus of China. He has expressed a clear policy intention to reduce the savings ratio of China. Although these studies recognize the relevance of savings to the current account balance, they do not investigate the reasons behind the high savings rates. An even more difficult question is whether the high savings rate is a cause or an effect of the current account surplus in China.

What factors sharply increased the macroeconomic imbalances in China after 2000? The next three sections consider 1) upward pressures on an already high domestic savings rate; 2) limits on investment in productive capacity; and 3) trade and exchange rate policies that promoted exports. Some of these policies and structural factors are historical legacies that were inherited from China's central planning system, while others are more recent government policies and regulations imposed on the household and corporate sectors.

The Aggregate Savings Rate in China

The aggregate national savings rate is comprised of the savings rates of three sectors: households, firms, and government. Thus, the national savings rate can rise either because the savings rate within one or more sectors rises, or because a high-saving sector expands while a low-saving sector contracts. Based on the flow of funds data, the rising saving rate of China from 2000 to 2008 can be decomposed into three changes: 1) a sharp rise in the disposable income of the high-saving enterprise sector contributed an increase of 5.5 percentage points to China's aggregate savings as a percentage of GDP, 2) a rise in the rate of government savings contributed another 4.1 percentage points, and 3) an increase in the rate of household savings

contributed another 7.6 percentage points. The other three elements of the decomposition, namely, changes in the savings rate in the corporate sector, changes in the share of government in GDP, and changes in the share of household consumption in GDP, played a limited or nonexistent role in the change in aggregate savings. Let us consider the three sectors in turn.

Enterprise Savings

In the flow of funds data for China, "enterprise savings" equals the value added for both financial and nonfinancial companies minus labor compensation, production taxes, net asset payments, and net transfer payments.[4] Therefore, by definition, the corporate sector has a unitary propensity to save because total corporate savings are equivalent to the "total disposable income" of the business sector, where final consumption does not take place (Ma and Wang 2010; Yang, Zhang, and Zhou 2012). From 2000 to 2008, the share of corporate income in the GDP of China rose by 5.5 percentage points, absorbing almost all the 5.7 percentage-point decline in the share of household income in GDP.

Several structural reasons contributed to the soaring profitability of enterprises in the early 2000s. By the late 1990s, China completed a series of economic reforms, including the use of labor-incentive schemes and the relaxation of worker mobility restrictions. Moreover, China implemented massive privatization of state-owned enterprises in the late 1990s with the objectives of improving corporate governance and maintaining the competitiveness of the state sector in the national economy. As a result, the employment share of the state sector fell, its labor productivity rose, and competitive pressures spread, increasing the efficiency of nonstate firms (Meng, in this issue of the journal).

However, these reforms were incomplete in a number of important ways. China continued to maintain the high-accumulation strategy that characterized the central planning era. Policies involved suppression of wage increases, low-interest payments on loans, and low land rentals to subsidize enterprises. For example, state-owned enterprises financed their loans and paid their debts at interest rates significantly lower than the prevailing market rates. If the state-owned enterprises had paid market interest rates, their existing profits, and thus their savings, would have been greatly reduced (Ferri and Liu 2010; Huang and Tao 2010). On top of these input market distortions, segmentation of rural and urban markets in the past implies the availability of massive amounts of rural unskilled labor that can migrate into cities as needed to meet industrial demand, thus decelerating urban wage growth. As a result, China in the late 1990s experienced a substantial rise in productivity, but the costs of production did not rise in a proportionate manner. The combination of these productivity-enhancing but cost-constraining reforms and policies thus increased the disposable income of enterprises despite a gradual reduction in market distortions over time.

---

[4] More specifically, asset payments include interest payments, dividends, and land rentals, whereas transfers include corporate income tax, social insurance fees, social subsidies, and social welfare payments.

When China joined the World Trade Organization in 2001, the resulting decline in trade barriers and tariffs allowed China to experience a dramatic expansion in external demand, which became a significant factor increasing firm productivity and profits. The continued inflows of foreign direct investment as well as the importation of sophisticated intermediate inputs also boosted Chinese exports. Between 2000 and 2008, the growth in China's exports reached an incredible 24.8 percent per annum (NBS 2009). The ratio of profits to industrial value added rose from an average rate of 22.6 percent over 1995–1999 to 34.4 percent in 2008. The share of enterprise income in the GDP rose from 14.2 percent in the second half of the 1990s to 22.9 percent in 2008.

An increase in corporate profits does not necessarily imply an increase in the aggregate savings rate, especially if the profits are distributed to workers or owners of the firms. In China, however, the corporate sector retained a significant amount of the increase in firm profits. In our study on the long-term wage trends in China using a national representative sample of urban households (Ge and Yang 2012), we report that average real wages increased by approximately 8 percent per annum, which is approximately 2 percentage points below the real annual growth of GDP, from 2000 to 2007. Dividend payments rose but were still quite small, with the ratio of dividend to value added staying at less than 0.5 percent by 2007 (Yang, Zhang, and Zhou 2012). One reason for low dividends is that the Chinese government did not ask state-owned enterprises to pay dividends until 2008, although they have enjoyed improved profits since state sector restructuring in the late 1990s.[5] Moreover, private enterprises had an extra incentive to save: they had to meet their funding needs largely through their internal savings (Song, Storesletten, and Zilibotti 2011). This is because credit creation in China is mostly controlled by state banks, which have an intrinsic bias in favor of state-owned enterprises.

If enterprises would send a larger share of their profits to the household sector, China's aggregate savings rate would decline. Compared with firms, households have a significantly higher propensity to consume and to import. Hence, changes in policies and institutions to encourage a reallocation of enterprise income to households would reduce both the internal imbalance of extremely high national savings and China's external imbalance of enormous current account surpluses.

Government Savings

Government savings in China rose from 3.3 percent of GDP in 2000 to 8.4 percent in 2008. In China, government income comes from several sources: the value added from the production of the state sector, income from properties, taxes on production, income taxes, and revenue from social insurance funds (minus spending out of those funds). Taking these factors together, revenues of the central government nearly quadrupled from 1.891 trillion yuan in 2000 to 6.797 trillion yuan in 2008

---

[5] These aggregate statistics appear to be consistent with firm-level data reported by Zhang (2008), who cites that for a large sample of Chinese firms in the period 1999 to 2003, the average and median dividends-to-earnings ratios were 0.35 and 0.16, respectively.

(NBS 2011). However, the pace of the increase in government spending failed to catch up with the pace of the increase in revenues.

The rise in tax revenues from production was the largest contributor to the growth in the government's disposable income during the period, accounting for 70 percent of the increase. This growth in production was spurred partly by the improving business environment and the expanding external demand that resulted from the entry of China to the World Trade Organization. However, the institutional foundation behind the rise in tax revenues linked to production can be traced to the 1994 fiscal reform. At that time, the Chinese economy had been growing for more than a decade, but much of the growth was accounted for by the nonstate sector. The tax code was not designed to gather revenues from that sector, thus state revenues actually were on a declining trend starting in the mid 1980s. The 1994 reform aimed to boost revenue collections by the central government (Wong and Bird 2008). From having a low net-revenue-to-GDP ratio in the early 1990s, the government posted much higher ratios from 2000 to 2008 due to implementation of an effective tax system and a robust annual GDP growth of approximately 10.4 percent.

The other major factor contributing to the government's growing surplus was that the increase in the collection of income taxes and social insurance fees outpaced the government's increase in spending. For example, in 2008 the Chinese government collected 1.5 trillion yuan worth of income taxes and 1.4 trillion yuan worth of social insurance fees, but it spent only 1.6 trillion yuan on social welfare payments, social insurance provisions, and other transfers. Back in 2000, revenues from income taxes and social insurance fees were only slightly higher than spending. This rise in what may be called "net current transfers"—the excess of income and social insurance revenues over and above the amount paid out in benefits—accounted for 22 percent of the growth in government disposable income during the same period. The accumulation of social insurance funds could reflect the effort of the Chinese government to build surpluses in anticipation of the rise in elderly dependency looming in the next decades, although it is difficult to prove, or disprove, this possibility.

The pattern of substantially growing government revenues, with spending lagging behind, is consistent with the "Nation Rich, People Poor" view that has been widely discussed in the public media in China. A piece of corroborative evidence is that the share of household income in GDP declined from an average of 68 percent over 1995–1999 to 57 percent in 2008, whereas the share of government income rose from 17 percent to 21 percent in the corresponding periods (NBS 2009; Bai and Qian 2010). However, even after the recent increases in tax revenues, tax revenue as a percentage of GDP in China is still lower than that of major developed economies such as Japan, Germany, and the United States.

Household Savings

The household savings rate in China has risen substantially in the past three decades against the backdrop of fast income growth. In the late 1970s, household

Figure 3
Average Household Savings Rate by Age of Household Head
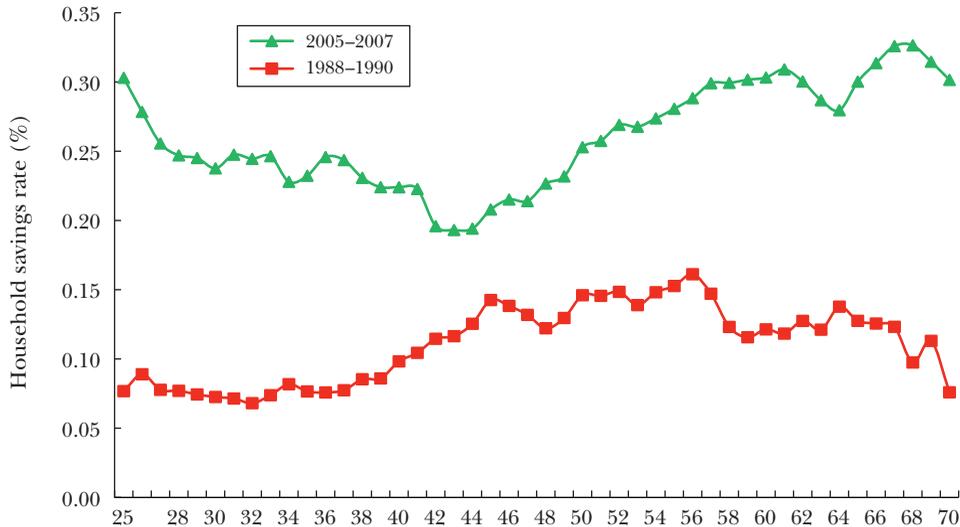
Source: Data are drawn from China's Urban Household Surveys covering five representative provinces (Liaoning, Zhejiang, Guangdong, Sichuan, and Shannxi) and a municipality (Beijing). I compute three-age moving average savings rates covering data of three-year intervals because some data cells of individual age and year have limited observations.

savings only accounted for 6 to 7 percent of GDP, but it grew consistently until reaching 23 percent in 2008. Economists have studied savings in China from many perspectives, including classic lifecycle theory, permanent-income approaches, and the significance of habit formation and cultural-based explanations for saving behavior. Some of the potential explanations include the change in age–earnings profiles for younger Chinese workers relative to older generations, the underdevelopment of the public and private pension systems in China; the effects of demographic changes like an aging population, the one-child policy, and the gender imbalance; and the effects of China's transition from public to private provision of education, health care, and housing.

As a starting point, one striking feature of Chinese household saving behavior is change in the age–savings profile. In the late 1980s, the age–savings profile reveals a relatively flat "hump shape," resembling the typical lifecycle saving profiles observed in other advanced economies (Modigliani 1970). However, as Figure 3 shows, the age–savings profile for the 2005–2007 period exhibits dramatic change: 1) there is a substantial increase in savings rates for households of all ages, and 2) the household's lifecycle savings pattern turns "U-shaped"—that is, the young and the old save relatively more than the middle aged. These patterns were first documented by Chamon and Prasad (2010) for selected Chinese provinces from 1995 to 2005 and later reconfirmed in Song and Yang (2010) using the national

sample of the Chinese Urban Household Surveys. These features challenge us in understanding the determination of household savings in China.

In Song and Yang (2010), we present a household model and show quantitatively that the dramatic rise in household savings and the corresponding changes in age–savings profiles are outcomes of two structural changes in China: First, there are large upward shifts in the earnings of successive younger cohorts of workers. Meanwhile, the age–earnings profile in China for any point in time has flattened over the past two decades. These changes reflect labor market transitions from a centrally planned economy, where seniority was highly regarded, to a market system, where earnings reward the productive human capital of the younger generations. Second, the aggregate pension replacement rate, which is the ratio of average pension per retiree to average wages per worker in a given year, declined from approximately 80 percent in the early 1990s to just above 50 percent in 2007. Incorporating these features of the Chinese economy into a dynamic optimization model of heterogeneous agents, we show that these factors can explain both the recent surge in household savings and the U-shaped age–savings profiles over the lifecycle.

Population control policies and the resulting demographic structural changes in China are another substantial influence on household savings. Because the younger population of an economy consumes without generating income, a fall in their share in the population tends to increase the household savings rate. Moreover, because China still lacks a mature social security system, adult children often provide old-age support to their parents, in effect acting as a substitute for lifecycle savings. In an analysis rooted in these insights and using aggregate time series data, Modigliani and Cao (2004) find that demographic changes in China raised Chinese household savings through the two effects of "less mouths to feed" and old-age security. In Ge, Yang, and Zhang (2012), we provide corroborative evidence through a cohort-specific analysis based on data from the Census of Population and Urban Household Surveys. We find that household savings rates increase in older families because a reduction in the number of adult children in these families induces them to save more to provide for their old-age security themselves rather than relying on their children. For households of younger generations, savings rates increase because of the rise in the burden of parental support as a result of the reduced number of siblings.[6]

The imbalanced sex ratio in China also results in a competitive motive for saving. Wei and Zhang (2011) begin by noting that that traditional preference for a son is widespread in China. Moreover, with restrictive population control policies, many families use inexpensive ultrasonic technology to detect the gender of fetuses and engage in sex-selective abortion, resulting in a severe imbalance in the sex ratio,

[6] These findings are controversial: for example, they are not confirmed by aggregate panel data studies. Neither the aggregate dependency ratio (Kraay 2000) nor separate accounts of the young and the old dependency ratios (Horioka and Wan 2007) are found to have a significant effect on the household saving rates across Chinese provinces. Applying a cohort analysis to the data from the Urban Household Survey, Chamon and Prasad (2010) reach a similar conclusion that demographic structural shifts do not go very far in explaining saving behavior in China.

which is defined as the ratio of males to females at specific ages. The intensified competition among men for potential wives stimulates households with a son to save and to accumulate wealth to gain a competitive edge in the marriage market. Wei and Zhang use provincial panel data (1978 to 2006) to test the effect of sex ratio imbalance on household savings. They show that the imbalanced sex ratio significantly increases household savings, with approximately 68 percent of the increase in rural savings rate and 18 percent in the urban rate attributed to the rise in the sex ratio.

Finally, the incomplete transition from public to private provision of education, health care, and housing also contributes to the rising household savings. Several authors point to the backwardness of China's financial institutions, arguing that China fails to pool risks by providing adequate medical insurance and unemployment insurance and also fails to transform savings into loans for education, housing, and other investments (for example, Woo 2008; Chamon and Prasad 2010). These factors might become less important over time but matter considerably during the transition period. Lin, Dinh, and Im (2010) investigate the implications of the financial structure on household savings. They argue that Chinese institutions impose a dampening effect on wage growth because the labor-intensive, small and medium-sized enterprises cannot receive adequate loans from state-dominated banks. In addition, ordinary people are not given a share in the high profits of state-monopolized industries and the natural resource sectors. The resulting higher levels of income disparity and a greater concentration of wealth for the rich tend to increase household savings. While some of the factors analyzed above do not necessarily have a distinctive time effect on household savings after 2000, they have contributed to the rise in the aggregate savings rate in recent years.

Constraints on Investment Growth

In 2000, on the eve of its entry into the World Trade Organization, the aggregate investment rate in China was at a trough of 35 percent of GDP (as shown earlier in Figure 2). The level of investment in China had moderated in the late 1990s when its economy experienced deflation and overcapacity production. However, between 2000 and 2005, the investment rate began to climb rapidly along with the savings rate, before the investment rate settled into a narrow range of 42 to 44 percent from 2005 to 2008. As noted earlier, the savings rate kept climbing higher as the investment rate plateaued. Why didn't investment keep rising?

The Chinese government exercises more effective control over investment through the state-dominated banking system than it does over savings decisions, which are more decentralized. Improvements in the investment climate, which resulted from entry to the World Trade Organization in 2001, boosted both foreign direct investment and domestic investments in China. According to Anderson (2008), these domestic investments were made mostly by large state-owned enterprises and concentrated on heavy industries, such as metals, materials, machinery,

automobiles, and chemical products. These investments increased production capacity, displaced imports of related products, and subsequently led to exports of surplus production.

However, in 2005, when the central government felt it should avoid overheating the economy, the National Development and Reform Commission issued a directive to impose controls against overinvestment with a list of "prohibited industries"—industries that should avoid further expansion. The heavy industries that had undergone dramatic expansions in capacity topped that list. Since then, with continued fear of an overheating economy, the Chinese government has sought to keep the aggregate investment rate at a steady level.

Moreover, the inefficient financial system did not help the country find ways to funnel excess savings to profitable investment opportunities. As Song, Storesletten, and Zilibotti (2011) explain, the state-owned banks are essentially incapable of providing effective loans to the growing and more-efficient private firms because of various legal and political problems. The immaturity of the financial system also hinders the channeling of the excess savings to education, housing, and other family-based investment loans (Woo 2008). Chinese banks are awash with cash, but rather than investing in high-return projects of nonstate enterprises or in consumer loans, the banks mostly use the funds to invest in low-yielding U.S. government bonds.

Trade Policies

China has vigorously pursued export promotion policies since the start of its economic reform in the late 1970s. In the years leading to its accession to the World Trade Organization in 2001, China practiced a combination of export-promoting and import-restricting policies through tariffs, quotas, and import licenses. In the early years of economic reforms, the primary concern of the government was to limit imports to avoid balance-of-payments problems resulting from excessive borrowing and trade deficits. In compliance with a membership requirement by the World Trade Organization, China phased out many of its import barriers by the late 1990s. However, a number of export-promoting strategies that were already being practiced since the 1980s remained in place. These include the "self-balancing regulation" on the export content of foreign firms, special trade zones, liberalization of ownership restrictions on foreign direct investment, export tax rebates, and exchange rate policy. The entry of China to the World Trade Organization was a catalyst that amplified the effect of export-promoting policies and helped to push China's trade surplus extraordinarily high.

The "self-balancing regulation," which is part of the law governing multinational companies, requires that foreign direct investment be oriented toward export industries (Yu 2007). A 1990 version of the implementation guideline sets an explicit rule that exports must exceed 50 percent of the total annual output of foreign firms. Although the law stating an explicit floor on exports was relaxed in

2001, the regulation remains in place. The share of foreign-invested enterprises in Chinese exports rose from approximately 20 percent in the early 1990s to 56 percent in 2009.

In the early 1980s, China established special economic zones for export in coastal cities. Owing to their initial success, special zones were expanded into inland cities. Multinational companies in these zones enjoy better protection of intellectual property rights, a lower corporate tax rate of 15 percent, duty-free treatment of imported inputs, cheap land, and incentives of zero property tax in the first five years. Additional benefits were also given to foreign firms if they exported most of their products (Wang 2010). The first boom period for these zones was from 1990 to 1993, when the cumulative number of zones jumped from 18 to 130. The second boom was from 1999 to 2003, when the number increased from 139 to 196 (Sheng and Yang 2012). A total of 221 policy zones had been established in China as of 2006. Wang (2010) finds that these special economic zones attract foreign investment in export-oriented industrial enterprises. The Chinese government also gradually lifted various ownership restrictions on foreign direct investment by expanding a list of encouraged industries while reducing the categories of restricted or prohibited industries in these zones (Sheng and Yang 2012). Major jumps in the list of "industries" occurred in 2002 and in 2007. These nationwide initiatives on ownership liberalization raised the volume of processing export and the product varieties of multinational firms.

Export tax rebates are another policy tool for promoting exports. This program entails the refund of tariffs on imported inputs and value-added tax already paid on exported goods. These policies discriminate against goods sold domestically, especially goods using imported inputs, and create an incentive for firms to sell products abroad. Under conceivable circumstances, goods are sold to foreign buyers at cheaper prices than domestic sales. After the Asian financial crisis in 1997, China lifted the rebate rates several times, reaching an average of 15 percent in 1999, to raise the competiveness of Chinese exports before joining the World Trade Organization. The total value of the rebate payment increased substantially after China joined the World Trade Organization, quintupling in value from 2002 to 2008. These tax rebates are substantial: In 2006, the total tax rebates for exports received by exporting firms were equivalent to 10 percent of aggregate corporate savings and approximately 14 percent of government tax revenue in the same year (Yang, Zhang, and Zhou 2012). Empirical studies show that duty drawbacks and value-added tax rebates are important in promoting exports in China (Chao, Yu, and Yu 2006). Although export tax rebates generally are permitted under the framework of the World Trade Organization, China uses this tool in a far more widespread and uniform way than most other countries. In a survey covering 55 developing countries, fewer than half of the countries had a legal framework or implementation regulations for their duty drawback schemes, which of course limits their implementation (Ianchovichina 2007).

Limits on certain imports to China also play a role in its enormous trade surpluses. China, the second largest economy in the world, is a developing country in need of

Figure 4
Index of Renminbi Exchange Rate against the U.S. Dollar, 1990–2011
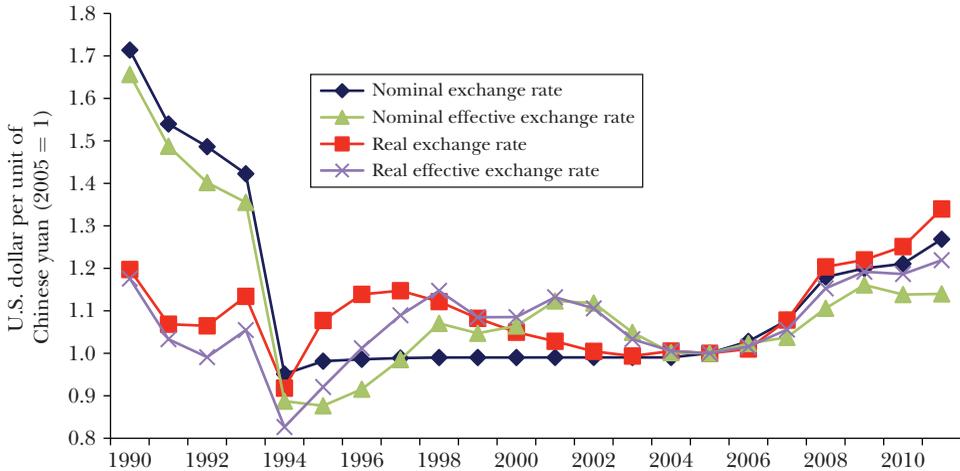
Source: Nominal exchange rate, nominal effective exchange rate, and CPI-based real effective exchange rate are from International Monetary Fund, *International Financial Statistics* (IMF 2012). The inflation-adjusted real exchange rate is based on author's calculations.

advanced technology. However, U.S. exports to China in 15 categories of goods with the highest technology content, are far below U.S. exports of the same goods to countries such as Canada, Japan, and Holland. In fact, among the same categories of high-tech goods, the exports from the United States to China are below the levels to India and Mexico (Ju, Ma, and Wei 2011). China's limited imports in these high-tech areas result in large part from the export restrictions imposed by the U.S. government or the complicated application and approval procedures.

Finally, the role of the exchange rate policy is frequently cited in public debates as causing the current account surplus of China. Some critics and researchers argue that the pegging of the renminbi to the dollar at a low value is the reason behind the large trade surplus of China, although disagreements over the issue abound. Systematic research using the latest data has not yet established robust evidence either that exchange rate regimes are connected with current account imbalances or that the changes in exchange rates have led to significant adjustments in imports and exports (for example, Corden 2009; Cheung, Chinn, and Fujii 2010).

Figure 4 presents indices of the renminbi exchange rate against the U.S. dollar for the 1990–2011 period, including nominal, real, and effective exchange rates. The trends in the exchange rate suggest that the sharp increase in the trade surplus of China from 2005 was not the result of a deliberate shift in exchange rate policy. From 1994 to 2005, China kept its nominal exchange rate stable at about 8.28 yuan to the dollar. During that period, China only had moderate trade surpluses of less than 2 percent of GDP in most years. In July 2005, after registering a jump in its foreign exchange reserves in the previous year, China allowed the renminbi to

appreciate by 2.1 percent and embarked on a process of steady appreciation of its currency. By January 2012, the nominal exchange rate of the renminbi against the U.S. dollar had risen by more than 30 percent. The real and effective exchange rates, which make adjustments for differential inflation in the two countries and the weighting of trade volumes, followed a similar pattern.[7] Apparently, this sizable appreciation of the Chinese currency was not enough to reverse the trend in the trade balance. Instead, the aforementioned institutional factors and other potential forces must have played a dominating role in creating the continued upsurge in the trade surplus.[8] Given the analyses presented in the present paper, the exchange rate policy can hardly be the only factor driving the external imbalance, and it may well not be among the most important factors. Of course, this argument over causes does not rule out that a weaker currency might be part of a policy solution to address the external imbalances of China.

Options for Policy Reforms

The Chinese economy has been strongly influenced by institutions and policies favoring the corporate and government sectors over households and promoting exports. Many of these individual policies appear relatively innocuous, in the sense that they seem too small to affect macroeconomic performance in any significant way under normal circumstances. However, with the huge external shock of the entry of China to the World Trade Organization in 2001, the effect of the individual policies was amplified, and the joint effects of these policies caused internal and external imbalances of China to become gigantic.

Mounting pressure exists for China to "rebalance" its economy, which in broad terms means to adjust the sectoral distribution of income away from the corporate and government sectors and toward the household sector. The level of consumption in China in the last few years, at roughly 47 to 50 percent of GDP, constitutes the lowest fraction of GDP recorded in any major economy. Several looming structural changes, such as slower economic growth and population aging, will likely help reduce the national savings rate of China in the future. However, these changes will likely be gradual and modest. What are some of the more direct reforms that could correct the imbalances of China? The coordination and timing of such reforms are complex. In what follows, I sketch several broad areas where I believe reforms are warranted.

---

[7] See Cheung, Chinn, and Fujii (2010) and Knight and Wang (2011) for additional analyses and alternative measures of the exchange rate.

[8] Although some critics have not considered the realized renminbi appreciation as being sufficient, market forces have begun to operate in the opposite direction. On several occasions this year, the renminbi has fallen to the bottom of the trading band set by the Chinese government, indicating the pressure to reverse the appreciation process since 2005. In the first half of 2012, the Chinese currency weakened by 1 percent against the dollar.

The conventional methods of rebalancing the Chinese economy often focus on government spending and the exchange rate. The Chinese government should shift the composition of its spending from investment to research and education, as its public expenditure on education as a share of GDP is still less than 4 percent, which is below the average for developing countries. Other major areas of expenditure increase should include the healthcare system, pensions, social security, and other selective social programs, especially for the vast rural population. These programs can directly reduce government savings and encourage households to consume more because of reduced incentives for precautionary savings. Meanwhile, China should increase flexibility in the exchange rate of the renminbi and allow its appreciation to continue to the extent such a direction is consistent with market forces.

While these changes to government budgets and to exchange rates are a plausible part of the overall picture, a solution to the macroeconomic imbalances in China will require a more sophisticated approach than these steps alone. It will require a shift in the policies and institutions that have favored production at the expense of consumption. Here are seven examples.

First, the state-dominated financial system should channel more domestic savings toward high-return investment by private, small and medium-sized enterprises, which are more labor intensive. Preferential access to credit and heavily subsidized capital financing for giant corporations and the state sector has lead to capital misallocation with the side effect of increased corporate savings.[9] Complementary to improving investment efficiency in production, it is imperative to develop the still-immature consumer loan system to help households finance education, housing, and durable goods consumption.

Second, despite more than three decades of economic reforms, China still faces the challenge of reducing input-market distortions and removing various impediments in market structures. Desirable policy reforms include restoring land prices to market values; breaking up state monopolies in industries such as natural resources, telecommunication, and financial services; and strengthening corporate governance and dividend policies for both state-owned and private enterprises. These policies can facilitate the determination of enterprise profitability based on sound economic principles and lower aggregate savings through the increased consumption of households and the government.

Third, the new Labor Contract Law, which took effect in 2008, should be effectively implemented. The law seeks to protect basic workers' rights, which are often abused because of asymmetric information and uneven bargaining power between

---

[9] Using enterprise-level data on interest payments, finance costs, and total debts outstanding at the end of the year, Ferri and Liu (2010) calculated that interest rate differentials paid by enterprises of different ownership categories were substantial over the 2001 to 2005 period. The yearly average interest rates paid by state-owned enterprises fell into a narrow range of 2.2–2.9 percent, whereas the range for cooperative enterprises was 4.6–12.4 percent and for private firms 3.8–13.4 percent. Although the contribution of state-owned enterprises to China's GDP was around 25 percent, they received about 65 percent of total bank loans.

employers and low-skilled workers as China has become the workshop of the world. The enforcement of the law is particularly relevant for the approximately 150 million rural migrant workers, who, due to the legacy of the urban household registration (*hukou*) system in China, still lack the full legal rights of those urban households.

Fourth, a gradual reduction in import duty drawbacks and export tax rebates should be planned. The moderation of export promotion policy will help curtail the transfer of tax revenue to exporting firms and insert downward pressure on the prices of goods sold in the domestic market. Both of these outcomes will help stimulate consumption.

Fifth, the removal of preferential policies towards export-oriented foreign direct investment in special policy zones—policies such as credit access, reduced tax rates, subsidized land prices, and lower environmental requirements—should be considered to set all firms in the market on an equal competitive footing. As a result, more domestic savings can be channeled to investment projects under competitive conditions, which will help to close the savings–investment gap.

Sixth, Chinese enterprises, especially those from the private sector, should be encouraged to invest abroad. Offshore investments by Chinese firms remain very small relative to the size of the Chinese economy and its foreign exchange reserves. Diversifying the overseas asset portfolio and raising the rate of investment returns closer to the level of the domestic market of China remains an important challenge.

Seventh, China should review its population control policies in the context of the anticipated rise in the elderly dependency ratio over the next four decades, which will have serious implications for savings and economic growth. The return to a more normal demographic structure can help alleviate unintended stress on the economy due to abrupt changes in population policies.

Each of these seven reforms in its own way would reduce the savings–investment gap and, together with the conventional recipes, would naturally ease China's huge current account surplus. These suggested reforms would also help improve the efficiency of resource allocation in the Chinese economy. For interested economists, there is a rich agenda for research that could deepen our understanding of the role of these policies and institutions in the determination of savings, investment, and current account imbalances. The spatial variations across the Chinese provinces and potential international data with variations in policy intervention and current account statistics may well provide a basis for further empirical investigation into relationships between key variables.

The development, over the last decade or so, of an enormous gap between savings and investments, along with mammoth current account surpluses and a growing net foreign asset position, was by and large not intended or desired by China. As the imbalances rapidly developed, policymakers failed to understand the complexity of the phenomena or devise effective remedies. Also, the hands of the Chinese government are, if not quite tied, highly restricted by past policies and institutions that often favored production and government at the expense of household consumption. In this sense, China is both the culprit and the victim of its own macroeconomic imbalances.

References

Anderson, Jonathan. 2008. "All about Rebalancing." In *How to Think About China*, 1–44. Asian Economic Perspectives, UBS Investment Research.

Bai, Chong-En, and Zhenjie Qian. 2010. "The Factor Income Distribution in China: 1978–2007." *China Economic Review* 21(4): 650–70.

Bernanke, Ben. 2005. "The Global Saving Glut and the U.S. Current Account Deficit." Sandridge Lecture, Virginia Association of Economics, Richmond, Virginia, March 10.

Chamon, Marcos D., and Eswar S. Prasad. 2010. "Why Are Saving Rates of Urban Households in China Rising?" *American Economic Journal: Macroeconomics* 2(1): 93–130.

Chao, Chi-Chur, Eden S. H. Yu, and Wusheng Yu. 2006. "China's Import Duty Drawback and VAT Rebate Policies: A General Equilibrium Analysis." *China Economic Review* 17(4): 432–448.

Cheung, Yin-Wong, Menzie D. Chinn, and Eiji Fujii. 2010. "China's Current Account and Exchange Rate." In *China's Growing Role in World Trade*, edited by Robert C. Feenstra and Shang-Jin Wei. Chicago: National Bureau of Economic Research (NBER), University of Chicago Press.

Corden, W. Max. 2009. "China's Exchange Rate Policy, Its Current Account Surplus and the Global Imbalances." *Economic Journal* 119(541): F430–F441.

Du, Qingyuan, and Shang-Jin Wei. 2012. "A Darwinian Perspective on 'Exchange Rate Undervaluation.'" Working Paper 16788, National Burueau of Economic Research.

Ferri, Giovanni, and Li-Gang Liu. 2010. "Honor Thy Creditors Beforan Thy Shareholders: Are the Profits of Chinese State-Owned Enterprises Real?" *Asian Economic Papers* 9(3): 50–71.

Ge, Suqin, and Dennis Tao Yang. 2012. "Changes in China's Wage Structure." IZA Discussion Papers 6492, Institute for the Study of Labor (IZA).

Ge, Suqin, Dennis Tao Yang, and Junsen Zhang. 2012. "Population Control Policies and the Chinese Household Saving Puzzle: A Cohort Analysis." (Forthcoming as an IZA Discussion Papers, Institute for the Study of Labor).

Goldstein, Morris, and Nicholas Lardy. 2009. *The Future of China's Exchange Rate Policy*. Washington, DC: Peterson Institute for International Studies.

Horioka, Charles Yuji, and Junmin Wan. 2007. "The Determinants of Household Saving in China: A Dynamic Panel Analysis of Provincial Data." *Journal of Money, Credit and Banking* 39(8): 2077–96.

Huang, Yiping, and Kunyu Tao. 2010. "Factor Market Distortion and the Current Account Surplus in China." *Asian Economic Papers* 9(3): 1–36.

Ianchovichina, Elena. 2007. "Are Duty Drawbacks on Exports Worth the Hassle?" *Canadian Journal of Economics* 40(3): 881–913.

IMF. 2011. *World Economic Outlook Database*. Online database. International Monetary Fund.

IMF. 2012. *International Financial Statistics*. Online database. International Monetary Fund.

Ju, Jiandong, Hong Ma, and Ziru Wei. 2011. "Anti-Comparative Advantage: A Puzzle in US-China Bilateral Trade." Working paper, School of Economics and Management, Tsinghua University.

Knight, John, and Wei Wang. 2011. "China's Macroeconomic Imbalances: Causes and Consequences." *World Economy* 34(9): 1476–1506.

Kraay, Aart. 2000. "Household Saving in China." *World Bank Economic Review* 14(3): 545–70.

Lane, Philip, and Gian Maria Milesi-Ferretti. 2007. "The External Wealth of Nations Mark II: Revised and Extended Estimates of Foreign Assets and Liabilities, 1970–2004." *Journal of International Economics* 73(2): 223–50.

Lin, Yifu Justin, Hinh T. Dinh, and Fernando Im. 2010. "US–China External Imbalance and the Global Financial Crisis." *China Economic Journal* 3(1): 1–24.

Ma, Guonan, and Yi Wang. 2010. "China's High Saving Rate: Myth and Reality." *International Economics* 122(2): 5–40.

Ma, Guonan, and Haiwen Zhou. 2009. "China's Large and Rising Net Foreign Asset Position." *China & World Economy* 17(5): 1–21.

Modigliani, Franco. 1970. "The Life Cycle Hypothesis of Saving and Intercountry Differences in the Saving Ratio." In *Induction, Growth and Trade*, edited by W. A. Eltis, M. F. Scott, and J. N. Wolfe, 197–225. Oxford: Clarendon Press.

Modigliani, Franco, and Shi Larry Cao. 2004. "The Chinese Saving Puzzle and the Life-Cycle Hypothesis." *Journal of Economic Literature* 42(1): 145–70.

National Bureau of Statistics of China (NBS). 2009. *China Statistical Yearbook*. Beijing: China Statistics Press.

National Bureau of Statistics of China (NBS). 2011. *China Statistical Yearbook*. Beijing: China Statistics Press.

SAFE, State Administration of Foreign Exchange. 2011. Data and Statistics: Balance of Payments. Available at: http://www.safe.gov.cn.

Sheng, Liugang. 2011. "Did China Diversify Its Foreign Reserves?" *Journal of Applied Econometrics*. Published online in WileyOnline Library (wileyonlinelibrary.com) doi: 10.1002/jae.1268.

Sheng, Liugang, and Dennis Tao Yang. 2012. "Expanding Export Variety: The Role of Institutional Reforms in Developing Countries." (This paper was previously titled "Speeding up the Product Cycle: the Role of Host Country Reforms," IZA Discussion Paper 6054.)

Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti. 2011. "Growing Like China." *American Economic Review* 101(1): 196–233.

Song, Zheng, and Dennis Tao Yang. 2010. "Life Cycle Earnings and Savings in a Fast-Growing Economy." Unpublished paper, Chinese University of Hong Kong.

Wang, Jing. 2010. "The Economic Impact of Special Economic Zones: Evidence from Chinese Municipalities." https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=NASM2011&paper_id=239.

Wei, Shang-Jin, and Xiaobo Zhang. 2011. "The Competitive Saving Motive: Evidence from Rising Sex Ratios and Savings Rates in China." *Journal of Political Economy* 119(3): 511–64.

Wen, Yi. 2011. "Explaining China's Trade Imbalance Puzzle." Working Paper 2011-018A, Federal Reserve Bank of St. Louis.

Wong, Christine, and Richard Bird. 2008. "China's Fiscal System: A Work in Progress." In *China's Great Economic Transformation*, edited by Loren Brandt and Thomas Rawski, 429–66. Cambridge University Press.

Woo, Wing Thye. 2008. "Understanding the Sources of Friction in U.S.–China Trade Relations: The Exchange Rate Debate Diverts Attention from Optimum Adjustment." *Asian Economic Papers* 7(3): 61–95.

Yang, Dennis Tao, Junsen Zhang, and Shaojie Zhou. 2012. "Why Are Saving Rates So High in China?" In *Capitalizing China: Translating Market Socialism with Chinese Characteristics into Sustained Prosperity*, edited by Joseph Fan and Randall Morck. The National Bureau of Economic Research (NBER), University of Chicago Press.

Yu, Yongding. 2007. "Global Imbalances and China." *Australian Economic Review* 40(1): 1–33.

Zhang, Haiyan. 2008. "Corporate Governance and Dividend Policy: A Comparison of Chinese Firms Listed in Hong Kong and in the Mainland." *China Economic Review* 19(3): 437–59.

Zhou, Xiaochuan. 2009. "Some Observations and Analysis of the Saving Rate Problem." Speech at the Bank Negara Malaysia's High-Level Conference, 10 February.

How Did China Take Off?

Yasheng Huang

There are two prevailing explanations of what caused China's rate of economic growth to take off. The first view gives the pride of place to globalization. According to this view, Chinese growth started when Deng Xiaoping liberalized trade and foreign investments by setting up special economic zones in the coastal provinces. In this view, China's export-oriented manufacturing, largely foreign-funded, employed millions of rural migrants, boosted their income, and reduced poverty far and wide. The second perspective emphasizes the importance of internal reforms—especially in rural, interior regions—of the agricultural pricing system; land contracting; and the entry of rural businesses known as township and village enterprises.

China's early external reforms are politically important. Special economic zones were ideologically controversial at the time they were introduced and their establishment signaled a triumph of the reformist leaders over conservatives. Also, the inflows of foreign investments were not spontaneous; they required an explicit shift in policies and legal practices. Politically, the 1979 passage of the Law on Chinese–Foreign Equity Joint Ventures, only three years after the Cultural Revolution and committing the Chinese government to the protection of foreign property rights, was pathbreaking.

But the economic contributions of foreign investments do not remotely match those of China's rural industry. At their peak, firms funded by foreign capital employed 18 million people (in 2010). By contrast, at their trough in 1978, township and village enterprises employed 28 million people. Between 1978 and 1988,

■ *Yasheng Huang is Professor of International Management, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts. His email address is yshuang@mit.edu.*

China's poverty headcount declined by 154 million, by far the most impressive record during China's three decades of reforms. The contributions of foreign capital toward China's initial poverty reduction during this period are miniscule. Employment by firms funded by foreign capital was 60,000 in 1985 and 660,000 in 1990. The same two figures for township and village enterprises are 69.8 million and 92.7 million, respectively (National Bureau of Statistics 2011). China's take-off in economic growth starting in the late 1970s and its poverty reduction for the next couple of decades was completely a function of its rural developments and its internal reforms in general.

During the golden era of rural industry in the 1980s, China had none of what are often thought of as the requisite features of the China growth model, like massive state-controlled infrastructural investments and mercantilism. In the 1980s, China had an overvalued exchange rate. Between 1980 and 1990, it had trade deficits every year except 1982, 1983, and 1990. (By contrast, since 1989 China has had trade surpluses every year except 1993.) In the 1980s, the household consumption to GDP ratio stood at over 50 percent, compared with 35 percent in recent years.

To understand how China's economy took off requires an accurate and detailed understanding of its rural development, especially rural industry spearheaded by the rise of township and village enterprises. Many China scholars believe that township and village enterprises have a distinct ownership structure—that they are owned and operated by local governments rather than by private entrepreneurs. That these firms could be so dynamic and efficient, yet government-owned, is often treated as a paradox in the economics literature. This statist view of township and village enterprises, together with the widespread belief that Chinese government has retained tight control of finance, led many scholars to conclude that Chinese growth has defied the conventional wisdom on the importance of private entrepreneurship and financial liberalization for growth.

But my own historical narrative—formulated on the basis of voluminous government and bank documents and data from the 1980s—directly contradicts this heterodox interpretation of Chinese reforms.[1] I will show that township and village enterprises from the inception have been private and that China undertook significant and meaningful financial liberalization at the very start of reforms. Rural private entrepreneurship and financial reforms correlate strongly with some of China's best-known achievements—poverty reduction, fast GDP growth driven by personal consumption (rather than by corporate investments and government spending), and an initial decline of income inequality.

The conventional view of China scholars is right about one point—that today's Chinese financial sector is completely state-controlled. How does one reconcile my reading of the historical evidence on financial reforms with the well-established

---

[1] The documents include a 22-volume compilation of documents of the central bank, all major state-owned commercial banks, and the rural credit cooperatives between 1982 and 2004. While they are available at libraries at Chinese University in Hong Kong and Harvard, they have not previously been examined by researchers. (Of these, the specific documents I cite are in the reference list.)

current fact of the statist financial controls? The same documentary and data sources provide the answer: China reversed almost all of its financial liberalization sometime around the early to mid 1990s. This financial reversal, despite its monumental effect on the welfare of hundreds of millions of rural Chinese, is almost completely unknown in the West.

My discussion is heavily tilted toward rural China because this is where Chinese growth took off. Developments in rural China also affected China's overall transition to a market economy, not only because in the 1980s the vast majority of the population was rural but also because Chinese capitalism is rural in origin. Reforms in rural China determine the pace and the nature of China's overall transition to a market economy. In the 1980s, it was the rural entrepreneurs who responded quickly to the incipient political and policy flexibility and who started businesses that competed directly with the urban state-owned enterprises. By the same token, reversing rural reforms not only suppressed rural entrepreneurship but also had the effect of slowing down China's overall transition to a market economy. In the 1990s and 2000s, China had no shortage of urban reforms, such as opening to foreign trade and investments; privatization of loss-making small state firms; and housing reforms. Yet China remains one of the most statist economies in the world. This is because the government reversed rural reforms in the 1990s.

The first section of this paper provides an account of how both the initial conditions and the specific reforms led to the rapid emergence of a market economy in rural China. The second section discusses an important institution in rural China—the township and village enterprises. A careful reading of original government documents suggests that contrary to the widespread belief, township and village enterprises may be history's most successful private sector story. The third section discusses financial liberalization and the subsequent reversals. The fourth section offers some speculative comments about how the policy reversals may have affected Chinese growth and the composition of growth.

"Nothing Other than Revolutionary Reforms"

China's rural reforms started in 1978, and their success was huge and instantaneous. Rural per capita income more than doubled between 1978 and 1984, and real rural per capita consumption increased by 51 percent between 1978 and 1983. Rural poverty declined sharply within the first decade of reforms (Riskin 1987).

The success of rural reforms is striking, considering that many economists thought that China's rural reforms were mere "modest" departures from the status quo (Lau, Qian, and Roland 2000). China did not dismantle all planned prices, but moved to a dual-track system in which farmers sold their crops at the market prices after they fulfilled their obligations to the state at state-fixed prices. Land was not privatized; it was contracted out to farmers on long-term leases. State firms were not privatized, but entrepreneurs were allowed to start their own businesses. Hausmann, Pritchett, and Rodrik (2004) emphasize the virtues of relaxing existing

restrictions compared with institutional reforms. Deng Xiaoping's agricultural reforms, which they characterized as "humble" in origin, are a prominent example in their framework—evolution, not revolution, in other words.

Deng himself seemed to disagree. In 1984, Deng stated (as quoted in Rural Economy Research Team 1998), "The rural reforms that were carried out in the past few years are nothing other than revolutionary reforms." Which perspective is correct? It depends on what benchmark you use. Western economists benchmark China to what is sometimes called a "Washington consensus" template that includes privatization, deregulation, financial liberalization, rule of law, and democratization. By that standard, Chinese rural reforms were modest. But the benchmark Deng had in mind was the Cultural Revolution, when the life of Chinese private entrepreneurs was, to quote Hobbes, "nasty, brutish, and short." From this dynamic perspective, allowing partial market pricing of agricultural crops, land contracting, and millions of startups by rural entrepreneurs was not modest at all.

The rural reforms did not happen in isolation. They were implemented in conjunction with some broad changes in Chinese politics. The Chinese leadership took deliberate and well-publicized gestures to instill confidence in policy credibility and political stability. In 1979, the Chinese government returned confiscated assets—bank deposits, bonds, gold, and private homes—to former capitalists and landlords. A large number of people imprisoned during the Cultural Revolution for engaging in private commerce were released from jail. (One survey showed that some 10 percent of China's private entrepreneurs were former prisoners.) In 1980, two vice premiers personally brought New Year's greetings to a woman who received the first license to open a business in Beijing. In 1981, the Politburo passed a resolution that made private entrepreneurs—code-named "individual laborers"—eligible for Party membership. In 1984, Wenzhou—a backwater rural county then—released eight peasants-turned-entrepreneurs (arrested in 1982 for graft) and made restitution of their assets. It also published an open apology in local newspapers. In the 1980s, China also introduced village elections (in Huang 2008, I provide more details).

Economists typically think of constraints on government as a source of credibility. The effect of a marginal change—moving from the capricious Cultural Revolution to the more predictable era of Deng Xiaoping—is therefore underestimated. This directional change from complete opposition to private business during the Cultural Revolution to a supportive stance during the Deng era helps resolve a paradox: Why millions of entrepreneurs were willing to invest even though the power of the state was, and still is, absolute. One reason is that in the 1980s entrepreneurs no longer faced automatic risks of being arrested and executed. In the 1980s, China went a long way in establishing security of proprietors even though security of property was far from being achieved. One should never underestimate the incentive effect of not being executed! (By the same token, as the Cultural Revolution fades in memory and the safety of proprietor is taken for granted, the safety of property and the rule of law will increase in their importance for China's future growth.)

Policy and political stability also mattered for the workings of the dual-track price system. Economists praise the dual-track price system because its mechanics are straightforward, but they ignore a precondition for the system to work: peasants had to be confident that the planned targets would not be ratcheted up later. The dual-track price system was neither novel nor Chinese. The Soviet Union tried it many times but failed. The difference is the Chinese leadership committed itself to not ratcheting up the targets and did so credibly. It was the politics of the Deng Xiaoping era, not just the mechanics of the reform, that accounted for the success.

There is another difference with the Soviet Union. Chinese reforms started in the rural areas and China was far more agrarian than the Soviet Union when that country began its own reform program. Rural China has always been more predisposed toward entrepreneurship and capitalism than urban China. Today many of the best private manufacturing firms are not located in Beijing or Shanghai but originated in then-agrarian provinces such as Zhejiang, Hunan, Anhui, and Sichuan.

Rural China's capitalist predispositions arise in part from the nature of agriculture, which is less conducive to planning than industry, and in part from history. The Cultural Revolution was primarily an urban political shock that cleansed urban China of any vestiges of capitalism, but there were still some free market activities in rural China. For the land contracting reforms to work, economic agents need to have some knowledge of residual claims. That knowledge was still extant in rural China but absent in the urban area. The rural readiness for capitalism acted as a multiplier that amplified the effects of policy reforms. This is why many of the reforms that elicited a huge supply response in China and Vietnam completely failed in the more industrialized Soviet Union (and for that matter, in urban China). Contexts and initial conditions provide the necessary complements to policy.

Private Entrepreneurship in Rural China

China's economy took off not just because peasants became more productive in producing grain, but also because reforms created conditions for Chinese peasants to switch to higher value-added activities such as industrial production and service provision. According to one study, four-fifths of the income gains came from improving allocative efficiency (Riskin 1987).

Township and village enterprises played a vital role in this process. They raised rural income, absorbed rural surplus labor, and contributed to a decline in the rural–urban income gap in the 1980s. The value-added produced by these rural businesses increased from 6 percent of GDP in 1978 to 26 percent of GDP in 1996 (Naughton 2007, p. 274). They also injected competition into the Chinese economy. In the 1980s, these rural businesses were the only source of competition to the incumbent state-owned enterprises at a time when foreign firms were still restricted and urban private firms were small. They undermined the monopoly of state-owned enterprises in both product markets and factor markets (in labor and capital). They played "a catalytic role" in China's economic transformation (Naughton 2007, p. 271).

The conventional view on township and village enterprises—laid out in a textbook on the Chinese economy—is that "township and village enterprises had a special distinction during this period [1978–1996] because of their unusual ownership and corporate governance setup. Originating under the rural communes, most township and village enterprises were collectively-owned. . . ." (Naughton 2007, p. 271). Collective ownership means ownership at the lower level of the Chinese political system, such as township and village. This special feature of township and village enterprises, according to Roland (2000), poses a challenge to researchers because, given their public ownership, they are not supposed to perform well. The strong theoretical priors of mainstream economists are that private ownership rights motivate entrepreneurs to invest and to take risks. The stellar performance of the township and village enterprises without this incentive device is puzzling.

Elaborate theories—some backed up by formal mathematical proofs—have been proposed to explain the performance of township and village enterprises as public-sector businesses. For example, Chang and Wang (1994) and Li (1996) argue that township and village enterprises had the advantage of political protection provided by the local governments and that in a biased financial system they had preferential access to capital. Che and Qian (1998) and Roland (2000) develop models of township and village enterprises as an efficient substitute for an economic environment with weak legal institutions. Stiglitz (2006) goes one step further. According to him, the biggest problem in transitional economies is not underdevelopment of the private sector but stealing on the part of the private sector. The hybrid nature of township and village enterprises aligns the interests of the central government with those of the local governments and effectively prevents private stealing of public assets.

These theoretical conceptualizations are all predicated on one important empirical detail—that township and village enterprises are public. Is that claim true?

Many Western scholars believe that township and village enterprises owed their origins to the rural "commune and brigade enterprises" created during the Great Leap Forward (1958–1961) and thus should be viewed as collective institutions. This belief is only partially correct. In 1978, there were about 1.5 million commune and brigade enterprises, but by 1985 there were already 12 million businesses labeled as township and village enterprises (Ministry of Agriculture 2003). Clearly, the vast majority of township and village enterprises had nothing to do with the Great Leap Forward. They were, instead, a product of the rural reforms that began in 1978.

In the records that I studied, the term "township and village enterprise" first appeared in a policy document issued by the State Council—the Chinese cabinet—on March 1, 1984. This landmark document broke new policy grounds. It officially replaced the previous term "commune and brigade enterprise" with "township and village enterprises" precisely because, as the document pointed out, the old term was no longer an accurate description of many of the new enterprises spawned by rural reforms. The second paragraph of this historic document gave the following definition of township and village enterprises (Ministry of Agriculture 1985, p. 450): "Township and village enterprises include enterprises sponsored by townships and

villages, the alliance enterprises formed by peasants, other alliance enterprises and individual businesses."

The enterprises sponsored by townships and villages are collectively owned—the kind of firms Western economists assume to represent the universe of all township and village enterprises. The other two categories under the label of "township and village enterprises" are straightforward private businesses or entities. The private township and village enterprises are divided by size. Individual businesses are equivalent to single proprietorships in the West and they typically have an employment of seven people or under. The alliance enterprises (in Chinese, *lianying*) are a 1980s reference to larger private-sector enterprises. They have multiple founders/shareholders (who are usually unrelated by family ties). They employ more than seven people. Over time, references to alliance enterprises were replaced by the term "private-run enterprises" (*siying qiye*) after a major 1987 Politburo document began to explicitly use the term "private-sector firms" (*China Township and Village Enterprise Yearbook* 1989a, p. 138).

This definition of township and village enterprises as including private ownership is consistently applied across the official documents. Here are four excerpts from a sampling of official statements, policy documents, and references. First, a manual prepared by the Shanxi Township and Village Enterprise Management Bureau (1985, p. 1) says that a township and village enterprise "belongs to collective ownership or individual ownership." Second, a 1989 Ministry of Agriculture report to the State Council on the state of township and village enterprise development summarized: "Nowadays a large portion of township and village enterprises comprise individual businesses and alliance enterprises. . . . Currently, individual businesses and alliance enterprises account for a large share of the township and village enterprises in the northwest, southwest, and other economically backward regions" (*China Township and Village Enterprise Yearbook* 1990, p. 4). Third, a 1987 document by Agricultural Bank of China instructed its regional branches not only to lend to enterprises at the township and village level but also to lend to alliance enterprises and household businesses (*China Township and Village Enterprise Yearbook* 1989b, p. 524). Fourth, an analysis in the *China Township and Village Enterprise Yearbook (1978–1987)* states: "Compared with a state-owned enterprise, a township and village enterprise…is a collective-ownership or individual-ownership enterprise with a lot of autonomy and able to make decisions concerning its own fate" (*China Township and Village Enterprise Yearbook* 1989b, p. 3).

The critical distinction between the Chinese official definition of township and village enterprises and the definition widely accepted in the Western economics literature is that the Chinese definition refers to *geography*—enterprises *located* in townships and villages regardless of their ownership. Western economists, mistakenly, assumed that the definition referred to enterprises *owned* by townships and villages.[2] Interestingly, some Chinese bureaucrats have committed the same mistake.

[2] One early study published in English that got this question right is by Bryd and Lin (1990), a joint research project by the World Bank and Chinese academics. It identified township and village enterprises

Wan Li, a foremost reformer and the vice premier in charge of agriculture in the 1980s, remarked in 1984, "[Some officials] only include the original collectively owned enterprises of townships and villages started by the masses as township and village enterprises, but do not include those businesses later established by peasants on their own or those alliance enterprises financed from pooled capital as township and village enterprises" in the definition of township and village enterprises (as quoted in *China Township and Village Enterprise Yearbook* 1991 p. 128).

The absolute majority of township and village enterprises, from the very beginning, were in fact private rather than public. In 1985, according to Ministry of Agriculture data, there were over 12 million township and village enterprises, of which 10.5 million were private. By contrast, there were only 1.57 million collective township and village enterprises in the same year. In 1978, the number of legally registered private township and village enterprises was zero; by 1985, this number had grown to 10 million strong. China's township and village enterprises are probably one of history's most remarkable private sector success stories.

Figure 1 is a graphical representation of the development of township and village enterprises over time based on Ministry of Agriculture data. (The Ministry of Agriculture data are more detailed than those provided by the National Bureau of Statistics because they are broken down by more detailed ownership categories. The two sources are consistent with each other.) In 1985, this data began to divide township and village enterprises into three categories: 1) collective, 2) privately-run, and 3) household businesses. The latter two categories are private, but differ in the size of their employment.

The growth of township and village enterprises occurred almost entirely on the private spectrum of township and village enterprises. The number of collective township and village enterprises was highest in 1986 and 1993, but then declined substantially throughout the rest of the 1990s. Meanwhile, the number of household businesses rose sharply throughout the 1980s as did the number of privately-run enterprises in the 1990s.

It is almost certain that these official figures of collective firms overstate the size of the public ownership of township and village enterprises, especially for the 1980s. The reason is that in the 1980s the Chinese government supported private sector development through political statements and financial opening, but it had not developed a legal framework for private entrepreneurs to register their businesses as explicitly private. Many private entrepreneurs therefore registered their businesses under the collective township and village enterprises. (One of the most famous examples of these "red-hat firms" is Wanxiang, which was founded by Lu Guanqiao. Wanxiang is now China's leading automobile component supplier, and it has an operation in Illinois.) This changed in 1994 when China's Company Law went into effect. Newly established private firms began to be explicitly registered as private and those previously registered as collective were converted into privately

---

as covering both public and private sectors. But this study had no apparent effect on the modeling work of economists that treated township and village enterprises as run by local government.

Figure 1
Number of Township and Village Enterprises by Three Ownership Categories:
Collective, Private-run, and Household Businesses, 1985–2002
(millions of units)
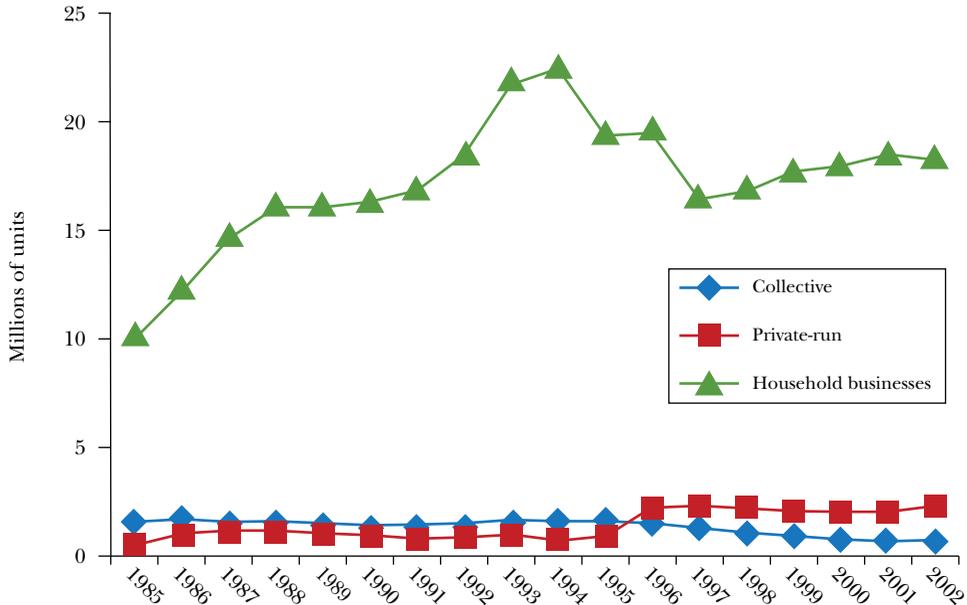
Notes: Data are from Ministry of Agriculture (2003).

registered entities. This change appears in Figures 1 and 2. After 1994, there was a sharp increase in the number and employment of privately-run township and village enterprises and a corresponding decrease in collective township and village enterprises. Some scholars mistook this development as privatization of collective township and village enterprises. In reality, it was a clarification of titles.

As shown in Figure 2, collective township and village enterprises were larger in terms of employment in the mid 1980s. In 1985, the collective enterprises accounted for 59 percent of overall employment in township and village enterprises. Keep in mind that the collective enterprises were founded in the late 1950s and had had 30 years of operation whereas the private township and village enterprises were new entrants in the 1980s. Even with their late start, by 1989, the share of employment by private township and village enterprises matched that of collective township and village enterprises, and the employment share of collective township and village enterprises declined throughout the reform era (except for a brief interval between 1992 and 1994). Other measures show the same dynamism among private township and village enterprises. In 1989, private township and village enterprises claimed 58 percent of the after-tax profits and 45 percent of the total wage bill of all township and village enterprises. In short, the growth miracle

Figure 2

Employment Shares of Township and Village Enterprises by Ownership Categories: Collective, Private-run, and Household Businesses, 1987–2002
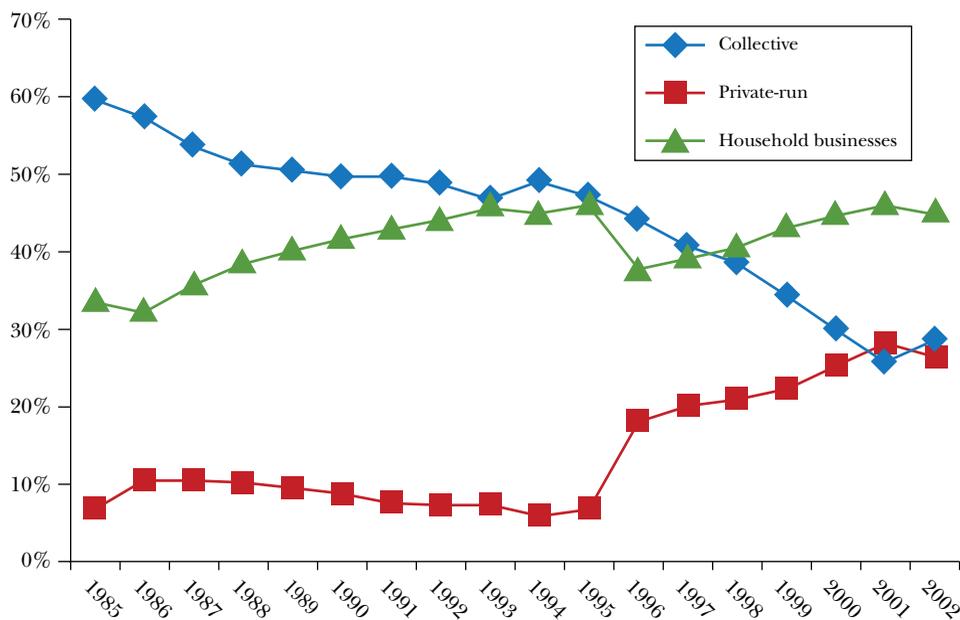
Notes: Data are from Ministry of Agriculture (2003).

of the township and village enterprises took place almost entirely in the private sector, not in the collective sector.

Data on output tells a similar story about the importance of private township and village enterprises. In 1987, private township and village enterprises produced 32 percent of the gross output value of the township and village enterprise sector.[3] However, this aggregate measure misses an important detail that has substantial policy implications—that private township and village enterprises prospered in poor provinces whereas collective ones tended to prevail in richer and less-industrialized provinces.

Figures 3 and 4 divide China's 29 provinces into three groups based on their per capita GDP and rural shares of population, respectively (with the middle group comprising nine provinces). The graphs are based on data for 1987. In Figure 1, the ten provinces in the top tier of per capita GDP—averaging 1,473 yuan per person—have the lowest private share of the gross output value of the township and village enterprises compared with the provinces in the lower income tiers. (Gross output value includes both industry and service sectors.) On average for the provinces in

---

[3] The unweighted average of the private township and village enterprises' share of the value of gross output across the provinces of China in 1987 is 40 percent.

Private Shares of the Gross Output Value of Township and Village Enterprises for Three Groups of Provinces Based on their per Capita GDP, 1987
(provincial averages)
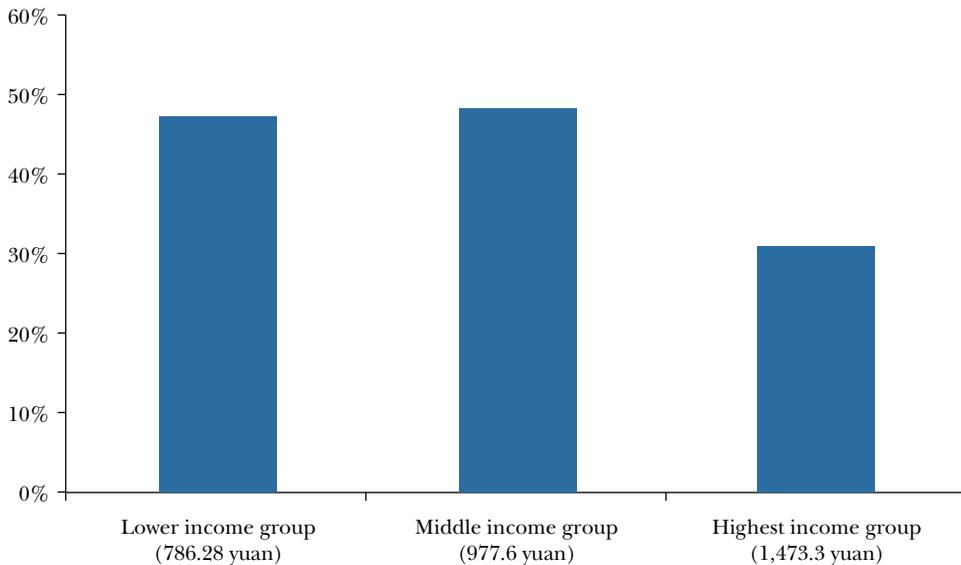
Notes: Data are from Ministry of Agriculture (2003). Twenty-nine provinces are divided into three groups based on their per capita GDP. The middle group has nine provinces. The figures in parentheses on the horizontal axis refer to the average per capita GDP in that income group.

the top income tier, private township and village enterprises produced 31 percent of the gross output value for the entire sector, compared with 47 and 48 percent for the other two income tiers.

The highest private share of township and village enterprise output is found in the province of Hebei, at 70.4 percent. In 1987, 85 percent of Hebei's population was rural. By contrast, Shanghai had the lowest private share of the township and village enterprise output, only 6 percent, meaning there is an extraordinary range between provinces in private shares of township and village enterprise output. The three regions with the lowest private shares of the township and village enterprise output are all cities: Shanghai (6 percent), Beijing (10.9 percent), and Tianjin (12.2 percent).[4] The contrast between Hebei on the one hand and Shanghai, Beijing, and Tianjin on the other illustrates the two sides in the development of Chinese capitalism: capitalism is rural and socialism is urban. Figure 4 further illustrates this divide. The ten provinces with the highest share of rural population (86.5 percent) averaged 49 percent in private share of total township and village enterprise output. For the ten provinces

[4] Chinese cities are an administrative concept, not an economic one. Cities also have rural population under their administration. In 1987, 34 percent of Shanghai's population was rural.

Figure 4

Private Shares of the Gross Output Value of Township and Village Enterprise for Three Groups of Provinces based on their Rural Shares of Population, 1987
(provincial averages)
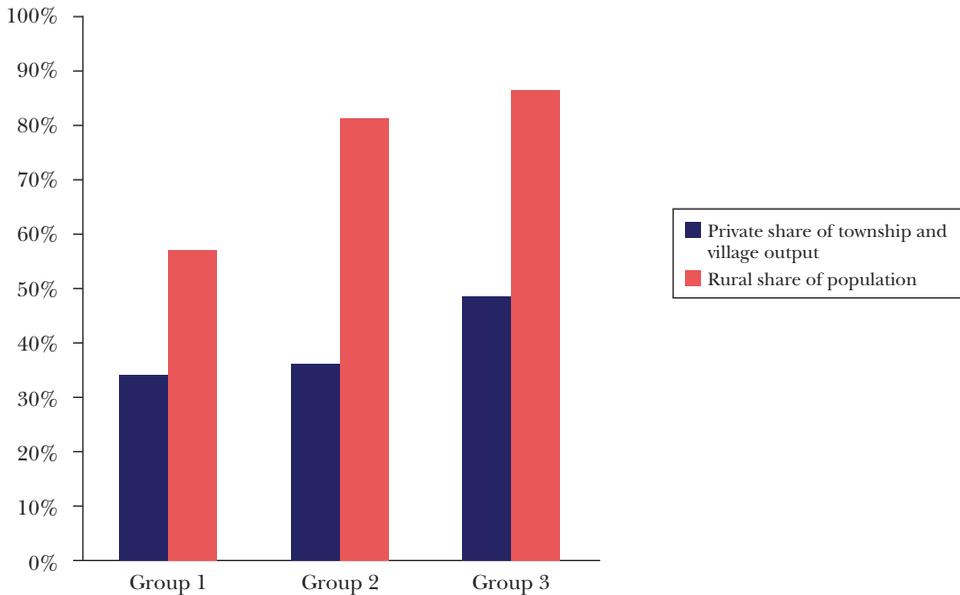
Notes: Data are from Ministry of Agriculture (2003). Twenty-nine provinces are divided into three groups based on their shares of rural population. The middle group has nine provinces.

with the lowest share of rural population (57.2 percent), the figure is 34.6 percent. There are other factors influencing private shares of township and village enterprise output, and the correlations are not perfect but fairly substantial. Across 29 provinces, the two-way correlation between per capita GDP and private shares of township and village enterprise output is about −0.71; for rural population shares it is 0.49.

One of the undisputed achievements of Chinese reforms is a sharp reduction of poverty. In China, as in all developing countries, poverty is concentrated in underdeveloped rural regions. For this reason, it is important to focus on the economic and business developments in the poor provinces in order to understand how China reduced its poverty. In the poor provinces, it was private entrepreneurship, not government-run township and village enterprises, that contributed to the bulk of output production. In aggregate, the scale of population affected by private township and village enterprises was massive.

As of 1987, private township and village enterprises already contributed about half of the output of the township and village enterprise sector in eight provinces, home to 260 million rural Chinese (30 percent of China's rural population). In another 15 provinces, the rural private sector produced between 30 to 50 percent

of the township and village enterprise output value. These 15 provinces accounted for an additional 427.8 million rural Chinese (about 50 percent of the rural population). Within a single decade of reforms, private township and village enterprises went from essentially nonexistent to contributing a substantial share of nonagricultural output in provinces with close to 80 percent of the rural Chinese population. It would not be an overstatement to say that rural private entrepreneurship played an instrumental role in China's impressive record of poverty reduction.

Financial Reforms and Reversals

An influential finance paper sought to explain why China can grow without financial liberalization by identifying informal finance as a key source of capital for private entrepreneurs (Allen, Qian, and Qian 2005). Informal finance and formal finance are, according to this view, substitutes. To evaluate this claim requires some contextual knowledge. Unlike other developing countries, China and other centrally planned economies categorically banned underground finance. Unlicensed deposit-taking or loan-making could be considered a capital crime. Thus, in China, that informal finance is allowed to operate at all is, *ipso facto*, evidence of financial liberalization rather than evidence of financial controls. It is unsurprising that informal finance is most vibrant during the most liberal period of Chinese reforms (such as in the 1980s) rather than in the Maoist period of the 1970s, and it is most vibrant in those regions of China with liberal economic policies (such as Wenzhou or Zhejiang province). After all, informal financiers are themselves private entrepreneurs. Formal finance and informal finance are substitutes only if government policies suppress or discriminate against the private sector. Otherwise they are complements.

Based on the data of the early 2000s, Allen, Qian, and Qian (2005) are correct that the private sector lacked access to formal finance, but they are wrong in believing that this was a permanent feature of Chinese reforms. In fact, China implemented far-reaching financial liberalization at the very start of rural reforms, a fact that is basically unknown outside of China despite its monumental significance. The reason for this lack of knowledge is that in the early 1990s, China reversed these reforms. The view that China challenges the standard view of economics by growing without financial liberalization is due to an observation bias: much of the economic research was done during the reversal period since the early 1990s rather than during the liberal period of the 1980s.

Rural financial reforms in the 1980s had three main components: First, the authorities began to delegate control rights of an important rural financial institution—the rural credit cooperatives—to its depositor-members, who, nominally, were the true shareholders. Second, entry of entrepreneurs into financial intermediation was permitted, even encouraged. Third, as a cumulative result of these aforementioned reforms, access to credit for rural private entrepreneurs was eased considerably.

In 1985, rural credit cooperatives accounted for 76.8 percent of all agricultural loans and 47.8 percent of all loans extended to township and village enterprises. They also intermediated and disbursed many of the loans originating from Agricultural Bank of China (China Finance Association 1986, p. II–19). Rural credit cooperatives were first established in 1951 as genuinely private financial institutions. Their members elected the officers and determined the lending priorities and criteria of their own branches. But in the 1960s and 1970s, as central planning took hold in China, a state bank, Agricultural Bank of China, and local governments took control of rural credit cooperatives.

Reform of rural credit cooperatives was one of the first acts of the reformist leaders. And unlike the much better-known household responsibility reforms, which were initiated by a group of Chinese peasants on their own in Anhui province, the financial reforms were launched by the central government. As early as 1980, the Politburo tasked a finance leadership group specifically to reform the rural credit cooperative system. The plan formulated by this group was to restore three founding principles of rural credit cooperatives. These are "organizational reliance on the members, managerial democracy, and operational flexibility."

This 1980 policy document, cited numerous times by later policy documents, is remarkable in several ways. First, it clarifies an issue of timing. Financial reforms occurred at the very outset of China's growth process as a deliberate political and policy decision. They were exogenous to—or at least concurrent with—China's economic takeoff. This detail on the timing of financial reforms directly contradicts a common claim among economists who study China—that, to the extent that China has implemented institutional reforms, they are endogenous with growth. Second, only four years after the Cultural Revolution, this document directly and explicitly singled out "government control" as the most serious problem afflicting rural credit cooperatives. This framing of government control as a problem set the tone for the reforms of rural finance for the rest of the decade.

In 1983, Agricultural Bank of China shifted from micro controls of rural credit cooperatives to indirect macro management. Agricultural Bank of China imposed a reserve requirement equivalent to 30 percent of the rural credit cooperative's deposit base and rural credit cooperatives could lend the rest on their own (Agricultural Bank of China [1983] 1985). By 1985, 80 percent of the rural credit cooperatives in the country had adopted reforms along this line (Agricultural Bank of China [1985] 1986, p. 34). In 1988, Agricultural Bank of China began to implement governance reforms. Article 11 of the employment regulations of rural credit cooperatives drafted by Agricultural Bank of China called for selection of rural credit cooperative branch managers through "democratic election" at shareholder meetings (Agricultural Bank of China 1988b, p. 200). To safeguard the function of these elections, the same document explicitly banned a widespread practice in Chinese bureaucracy—rotating the heads of departments across regions.

The authorities also liberalized entry. Informal finance emerged not because of porous bans but because of deliberate and proactive policy encouragements.

These two statements from the 1980s are quite telling. First, here is Chen Muhua (1987): "In addition to the capital provided by the state banks and rural credit cooperatives, there are now various kinds of businesses with deposit-taking and lending operations. Non-governmental capital mobilization and non-governmental rural cooperatives have emerged. The various methods of financial mobilization have made a positive contribution to local economic development." Second, here is Han Lei (1984, p. 51): "Rural areas need state-owned banks and credit cooperatives for finance but at the same time, under bank supervision, we need to allow the existence of private free lending and borrowing."

These are not two liberal academic economists advocating reforms. Chen Muhua and Han Lei were, respectively, the Governor of the People's Bank of China (PBoC)—China's central bank—and the Chairman of Agricultural Bank of China, arguably the most important state-owned bank in China in the 1980s, given the pioneering rural reforms. Several bank documents from that era justified informal finance on competitive grounds—that they competed with and therefore helped improve the state-owned banks. (The bank documents in the 1990s made the same factual observation on competition but drew exactly the opposite normative and policy implications.)

Western academics who did research in the 1990s record informal finance primarily in liberal and richer regions of the country, such as Wenzhou and Fujian (Tsai 2002). Bank documents in the 1980s record informal finance activities across many economically heterogeneous regions, such as Guizhou (China's poorest province), Guangxi (poor and populated by ethnic minorities), and Jilin (a conservative stronghold of state-owned enterprises). A 1987 report by the Jilin's Branch of the People's Bank of China (p. 151) shows that 69 percent of rural households had access to informal finance and 81 percent of informal loans were used to finance production. A better predictor of informal finance is not geography, but liberal policies. That informal finance was present in many parts of the country is evidence that policies were liberal nationwide.

The nongovernment financial institutions that Governor Chen and Chairman Han referred to are the "rural cooperative foundations." Rural cooperative foundations were local savings and loan institutions similar to rural credit cooperatives except that their shareholders exercised real control rights. Although the People's Bank of China never formally acknowledged rural cooperative foundations as a legitimate financial institution, it tacitly permitted their operation. The Ministry of Agriculture provided the political cover by giving rural cooperative foundations a formal status. Rural cooperative foundations competed directly with rural credit cooperatives and Agricultural Bank of China on both deposit and lending businesses. The scale of the rural cooperative foundations was massive. As of 1990, the rural cooperative foundations operated in 38 percent of Chinese rural townships (Rural Work Leadership Team of Fujian Communist Party Committee 1997). In Wenzhou, by the end of the 1980s, rural cooperative foundations began to approach Agricultural Bank of China in both loan size and network reach (*Wenzhou Financial History* 1995, p. 152, 225).

In the 1980s, rural households also had substantial access to formal credit in addition to the new availability of informal finance. Bank documents from the era consistently called for easing of credit access for private entrepreneurs. Fixed interest rates are a barrier to private access to capital because they lead to scarcity and credit rationing. Another barrier is requiring fixed assets as collateral. In 1984, Agricultural Bank of China permitted floating interest rates and waived collateral requirements for household businesses, and then in 1988 extended the same policies to larger private-run enterprises (Agricultural Bank of China [1984] 1986, p. 364; Agricultural Bank of China 1988a).

On all three fronts—reforms of rural credit cooperatives, entry liberalization, and credit access, bank documents reveal evidence of explicit, direct, and complete policy reversals in the early 1990s. The 1988 ban on rotating rural credit cooperative managers was rescinded in 1992 (Agricultural Bank of China 1992b). Rotation is a mockery of democracy as it nullifies any elections in the rotated regions. (Imagine rotating the governor of Massachusetts to Maine.) The control rights of rural credit cooperatives were recentralized. The three principles cited often in the 1980s documents—"organizational reliance on the members, managerial democracy, and operational flexibility"—completely disappeared in the 1990s. The bank documents now emphasized the "cooperative" nature of rural credit cooperatives, which in the Chinese parlance means in a state of transition from private to public ownership. After 1996, even the implied autonomy associated with "cooperative" was dropped. In 1999, Shi Jiliang (1999), a vice governor of the central bank defined rural credit cooperatives explicitly as "local government financial institutions." In March 1998, the People's Bank of China ([1998] 1999) formally assumed operational controls of rural credit cooperatives, including personnel appointments, screening of candidates, account examinations, and the termination of appointments. The micromanagement of rural credit cooperatives, still nominally owned by their members, was numbingly detailed. A 1995 Agricultural Bank of China document instructed its officers to examine the filing systems and to inspect the computer software used in rural credit cooperatives (Agricultural Bank of China 1995). The management of rural credit cooperatives had completely returned to their *status quo* during the central planning era.

In the 1990s, the authorities mounted a ferocious attempt to wipe out informal finance. Several private financial entrepreneurs were arrested and punished severely. One, an illiterate woman in Zhejiang, was executed. (She committed the alleged fraud in 1986 but was executed only when the macro policy changed, in 1991. Incidentally, another Zhejiang woman, Wu Ying, was given a commuted death sentence in 2012 for "illegal capital mobilization.")

The most direct target was rural cooperative foundations. In 1993, rural cooperative foundations were stripped of their deposit-taking businesses and were ordered to transfer their deposits to rural credit cooperatives (State Council [1993] 1994, p. 7). The following year, the authorities restricted the lending operations of rural cooperative foundations to low-margin agricultural production and forbade establishment of new branches and lending to urban residents (Rural Work Leadership Team of Fujian Communist Party Committee 1997).

The official language on competition changed from laudatory to derogatory. A 1996 document described rural cooperative foundations as "competing viciously" with state-owned banks. Instead of welcoming the competition as Governor Chen Muhua did in 1986, this decree ordered a complete takeover of all rural cooperative foundations by rural credit cooperatives, which, as pointed out before, were themselves being recentralized. Then in 1998, in a decree issued by the State Council and signed by none other than the premier himself, rural cooperative foundations and all forms of informal finance were categorically declared illegal. The decree criminalized not only the informal finance itself but also any failures by officials to refer the informal financial operations to public security bureaus (State Council 1998).

The private sector's access to formal finance was sharply curtailed. In the 1980s, rural credit cooperatives and rural cooperative foundations were important sources of funding for nonfarm entrepreneurship in rural China. In the 1990s, rural cooperative foundations were abolished and rural credit cooperatives were ordered to redirect their lending focus to agriculture. High floors were set for agricultural lending in the total loan program of rural credit cooperatives (usually above 40 percent). This is an implicit discrimination against rural private entrepreneurs who started businesses *to get out of agriculture.* Another implicit discrimination was the order to direct lending to production rather than investment. In a 1996 State Council document, fixed-asset loans were capped at 30 percent of all rural credit cooperative loans. Private entrepreneurs, by definition, were new entrants and needed loans to finance construction of new facilities.

The lending criteria were tightened substantially for private borrowers (despite clear evidence that it was urban state-owned enterprises, not rural entrepreneurs, that defaulted on loans). In 1992, Agricultural Bank of China required rural credit cooperatives to collect deposits from private entrepreneurs for a "risk guarantee fund" as an eligibility criterion (Agricultural Bank of China 1992a). In the 1980s, rural credit cooperatives used floating and higher interest rates to mitigate against default risks. Both evidence and theory suggest that this approach can be productive because it is able to distinguish between good and bad borrowers. Potentially productive borrowers can generate returns to pay for the higher interest costs. The 1992 policy required an upfront payment, and it biased lending to the incumbent businesses at the expense of new entrepreneurial entrants.

In the 1980s, the Agricultural Bank of China and rural credit cooperatives waived collateral requirements under certain conditions. In the 1990s, collateral requirements were reinstated and tightened. The most common form of assets in rural China—land for private farming and private housing structures—was purposely excluded as collateralizable assets. This raised the threshold for loan eligibility substantially (see Rural Work Leadership Team 1997). Even all these measures were deemed inadequate. In 1994, Agricultural Bank of China required rural credit cooperatives to issue loans to private entrepreneurs only after two signatures were secured, one from the loan officer and the other from the head of the rural credit cooperative at the next higher level. For example, a loan in a township had to be approved at the county level (Agricultural Bank of China 1994).

Along with these financial controls, the governance of rural credit cooperatives deteriorated as rural credit cooperatives essentially became cashiers of local governments rather than serving the needs of their nominal shareholders. Here is a catalog of egregious lending practices compiled by the People's Bank of China (2001): making loans to peasants in the form of goods rather than money and forcing peasants to sell the goods to buyers designated by rural credit cooperatives; expropriating the share capital contributions of the members of the rural credit cooperatives when extending loans; collecting taxes and fees from peasants when making loans; making loans to township and village governments to finance their tax bills to higher-level governments; forcing peasants to purchase shares of rural credit cooperatives and deducting their share contributions from their loans; financing government office construction and purchasing luxurious sedans while operating at a loss.

The health of rural finance deteriorated massively. The bank documents reveal that while rural credit cooperatives carried some nonperforming loans on their books in the 1980s, almost all of these nonperforming loans resulted from lending during the Cultural Revolution when rural credit cooperatives extended consumption loans to indigent households. Nonperforming loans increased in the 1990s *after* financial controls were instituted. In 1994, 31.4 percent of the loan assets of the rural credit cooperatives were nonperforming, and in 1996 nonperforming loans increased to 38 percent, according to Dai Xianglong (1997), the Governor of the People's Bank of China. The shareholder equity of the rural credit cooperatives was reported to be 63.2 billion yuan in 1995, 54.8 billion yuan in 1996, 31 billion yuan in 1997, 15.1 billion yuan in 1998, and –8.5 billion yuan in 1999 (China Finance Association 1997, p. 452; 2000). Within a single decade of policy reversals, an institution that had played an important role in China's economic takeoff became technically insolvent.

Conclusion

This concluding section addresses four issues. First, why did China reverse its rural financial reforms? Second, is the claim that China initially opened and then subsequently closed its rural finance supported by quantitative as well as documentary evidence? Third, did the policy reversals in rural China matter for real economic outcomes? Fourth, did the policy reversals affect China's pace of transition to a market economy?

The rural policy reversals coincided closely in timing with the assumption of power by a new group of leaders in the aftermath of the 1989 Tiananmen crackdown. In the 1980s, the leader in charge of the economy—Zhao Ziyang—pioneered rural reforms in Sichuan, but he fell from power in 1989. Between 1989 and 2002, China was led by a group of urban technocrats who made their political career in the most urban and statist region of China—Shanghai (Huang and Qian 2010). The urban technocrats launched massive infrastructure projects that required classic

mobilization of financial resources. They reduced the credit allocation to rural China and stamped out informal finance—mostly in rural China—in order to fund these big-push projects. The instability of 1989 also prompted the leadership in the 1990s to strengthen the controls by the Communist Party across the board and to discontinue the incipient political reforms introduced by Zhao Ziyang. Rural China bore the brunt of this reversal because political reforms went farthest in rural China.

There is quantitative—not just documentary—evidence for the financial reform/reversal story as laid out in this paper. Examining a large rural household dataset conducted by the Ministry of Agriculture between 1986 and 2002, in Qian and Huang (2012), my coauthor and I found that credit access by rural households declined sharply between the 1980s and the 1990s. Specifically, during the 1986–1991 period, close to 30 percent of rural households reported receiving either formal or informal credit. This ratio declined to 10 percent between 1995 and 2002. We found that, controlling for a variety of household characteristics, the political status of households—such as having a Communist Party member—had no effect on credit access in the 1980s but a substantial and positive effect in the 1990s. As credit access was reduced, political power became more important in credit allocation. Another finding is that in the 1990s fewer loans went to households who ran nonfarm businesses as compared with the 1980s. The policy changes shown in the documentary examination had a real and substantial effect.

One can easily counter the claim that rural financial policy reversals have inhibited economic growth by pointing out that China's GDP has continued to grow strongly since the early 1990s. But it seems the rural financial policy reversals affect the composition of the growth rather than the growth itself. One clear difference between the 1980s and the 1990s lies in the growth rates of rural household income. Although there are data complications (mostly having to do with how rural migrant labor income is measured), the overall trend is clear. Between 1978 and 1988, growth of real rural household income per capita averaged more than 10 percent a year, exceeding the GDP growth. Between 1989 and 2002, the growth slowed down to 4 percent a year, less than half of the GDP growth. In fact for several years in the late 1990s, there was no nominal growth of rural household income at all. All the growth came from deflation. (Since 2003, rural household income growth recovered to about 7 percent per year.)

It would be surprising if this slowdown in rural household income on the part of some 70 percent of the population did not affect China's growth patterns. In the 1980s, household consumption to GDP ratio was around 50 percent. This ratio began to decline in the early 1990s, and now it is only around 35 percent, probably the lowest for any major economy. In the 1990s, Chinese growth became increasingly driven by the state-controlled investments and then since 2000 by net exports. This is in sharp contrast to the 1980s, when China was investing at a level comparable to other East Asian economies and its trade accounts consistently produced a deficit. A reasonable hypothesis is that reduced growth of rural household income led to a slowdown of rural household consumption and this income effect is in addition to and possibly more important than the changes in household savings behavior.

My account is heavily tilted toward developments in rural China. One may argue that this rural focus, while important to a story of how China initially took off, is less relevant to China today. Rapid industrialization has reduced the rural share of population from 80 percent in the 1980s to around 50 percent today. Also one may argue that the negative effects of rural policy reversals are more than offset by the urban reforms since the early 1990s, such as opening to foreign direct investments, privatization of urban state-owned enterprises, and restructuring of the banking system. On the basis of these reforms, some economists have pronounced that China has already succeeded in transitioning to a market economy and its major challenge today is economic development rather than reforms (Qian 1999; Naughton 2007).

My account does not support this optimistic assessment of Chinese reforms. A key insight here rests on some fundamental differences between rural and urban China. Chinese capitalism is rural in origin, and rural capitalism is highly entrepreneurial in a Schumpeterian sense. Urban China, by contrast, is far more state-controlled. The urban reforms since the early 1990s have led to the rise of a politically-connected, rent-seeking private sector (the most prominent example of which is the real estate private firms), whereas the reversals of rural reforms have been at the expense of a more arm's-length, entrepreneurial type of private sector. The effects of rural policy reversals go beyond rural China; they may have slowed down the overall pace of market transition.

Those who argued that China has completed its market transition cite rising and high shares of private output as evidence. For example, an OECD study shows that private firms accounted for 52.3 percent of industrial value-added in 2003, up from 27.9 percent in 1998 and, most likely, from zero percent in 1978 (Dougherty and Herd 2005). Estimates like this tend to overstate the true size of China's private sector. First, the definition of private sector is complex. The OECD study includes "legal-person" shareholding firms in its definition of private firms. In reality, many of these firms are affiliates or subsidiaries of state-owned enterprises and should be classified as state-owned enterprises. A more realistic estimate is that the private share of industrial value added in 1998 was 18 percent in 1998 and 28 percent in 2003 (Huang 2008).

The fact that the output share of the private sector has grown is not disputed. The issue is whether the size of private sector output is a reliable indicator of economic transition. Here an example from the Soviet Union is illustrative. No one accuses Leonid Brezhnev of being pro-market, but under his leadership, private plots contributed as much as half of agricultural household income in the Soviet Union (Gregory and Stuart 1981, p. 230). This occurred because private farming was so much more efficient than state farming, so its contribution to income was disproportionate to the inputs allocated to it. Private plots only accounted for 1.4 percent of cultivable land in the Soviet Union (Hewett 1988, p. 117).

This Soviet example shows that input, not output, is a more reliable measure of private-sector policies. Output measures of transition conflate two effects: an efficiency effect and a policy effect. Because private firms are more efficient, their

output share can rise even if the policy environment is fixed or adverse. In 1990, the post-Tiananmen leadership cracked down on the private sector, and yet the output share of the private sector still increased. By one input-based measure, fixed asset investments, China's private sector development was not linear. Between 1981 and 1989, the private share of fixed asset investments averaged around 21 percent. Between 1993 and 2001, the share declined sharply to 13 percent (for details, see Huang 2008).

The rural policy reversals directly contributed to this development. In the 1980s, almost all of the private fixed asset investments were rural, and the net effect of rural policy reversals was to reduce the growth of the private sector from a much larger, initial rural base of capitalism. The private sector did grow, especially since 2000, but that growth was off a far smaller urban base.

To understand the two sides of China's growth story—its steep poverty reduction on account of fast personal income growth, and the statist orientation of its economy today—it is important to get the rural story right. China's takeoff began with rural entrepreneurship and substantial liberalization, but for political reasons, its leadership in the 1990s chose to reverse or discontinue much of the reform package that proved so promising. Financial reforms were completely reversed and land contracting reforms did not deepen. One reform that survived—village elections—was weakened considerably by the fact that the Chinese Communist Party tightened its direct controls of villages. Only village directors are subject to elections; village Party secretaries, who have real power, are not.

Between 1978 and 2012, it has taken China more years to "reform" the central planning system than it took for it to establish and operate that system from 1949 to 1978. Under a broad definition of private sector, China's private fixed-asset investment share was around 34 percent in 2005. This is less than the same ratio in India—of Indira Gandhi, not of Manmohan Singh, that is. In 1983, the comparable Indian ratio was 58 percent (World Bank 1989). China has moved from central-planning to what might be called a commanding-heights economy. This is progress, but China is far from completing its transition to a market economy.

References

Agricultural Bank of China. [1983] 1985. *Guanyu Gaige Xingyongshe Guanli Dizhi De Shidian Yijian [Opinion on Pilot Programs to Reform the Supervision System of Credit Cooperatives]*, edited by the General Office of the People's Bank of China. Tianjin: Zhongguo Jingrong Chubanshe.

Agricultural Bank of China. 1984. "Guanyu 1984 Nian Shangbang Nian Nongcun Xindai Gongzuo Qingkuang [Situation of Rural Credit Work in the First Half of the Year]." In *Nongcun Jingrong Guizhang Ahidu Suanbian [Selection of Rural Financial Regulations]*, edited by the General Office of Agricultural Bank of China. Beijing: Zhongguo Jingrong Chubanshe.

Agricultural Bank of China. [1984] 1986. "Nongcun Geti Gongshangye Daikuan Shixing Banfa [Provisional Regulations on Loans to Rural Individual Industrial and Commercial Businesses]." In *1984 Nian Nongcun Jingrong Guizhang Ahidu Xuanbian [Selection of Rural Financial Regulations in 1984]*, edited by the General Office of Agricultural Bank of China. Tianjin: Zhongguo Jingrong Chubanshe.

Agricultural Bank of China. [1985] 1986. *1985 Nongcun Jingrong Gongzuo Zongjian [Summary of Rural Financial Work in 1985]*, edited by the General Office of the People's Bank of China. Tianjin: Zhongguo Jingrong Chubanshe.

Agricultural Bank of China. 1988a. "Geti Gongshanghu, Siying Qiye Daikuan Guanli Zhanxing Bangfa [Provisional Regulations on Loans to Individual Businesses and Private Enterprises]." In *Jingrong Guizhang Zhidu Xuanbian 1988 [Selection of Financial Regulations 1988]*, edited by the General Office of the People's Bank of China. Beijing: Zhongguo Jingrong Chubanshe.

Agricultural Bank of China. 1988b. *Nongcun Xinyong Hezuoshe Zhigong Guanli Zhanxing Guiding [Provisional Regulations on the Employment Practices of Rural Credit Cooperatives]*, edited by the General Office of the People's Bank of China. Beijing: Zhongguo Jingrong Chubanshe.

Agricultural Bank of China. 1992a. *Guanyu Jiakuai Nongcun Jingrong Gaige Kaifang De Yijian [An Opinion to Speed up the Rural Financial Reforms and Opening]*, edited by the General Office of the People's Bank of China. Beijing: Zhongguo Jingrong Chubanshe.

Agricultural Bank of China. 1992b. *Guanyu Jiaqiang Nongcun Xinyongshe Zhigong Duiwu Jianshe De Yijian [Strengthening the Employment System at Rural Credit Cooperatives]*, edited by the General Office of the People's Bank of China. Beijing: Zhongguo Jingrong Chubanshe.

Agricultural Bank of China. 1994. *Guanyu Jinyibu Jiaqiang Nongcun Xinyongshe Xindai Guanli De Yijian [Opinion to Strengthen the Supervision of Loans in Rural Credit Cooperatives]*, edited by the General Office of the People's Bank of China. Beijing: Zhongguo Jingrong Chubanshe.

Agricultural Bank of China. 1995. *Nongcun Xinyong Hezuoshe Jihe Chufa Zhanxing Guiding [Provisional Regulation on Auditing of Rural Credit Cooperatives and Punishment Measures]*, edited by the General Office of the People's Bank of China: Zhongguo Jingrong Chubanshe.

Allen, Franklin, Jun Qian, and Meijun Qian. 2005. "Law, Finance and Economic Growth in China." *Journal of Financial Economics* 77(1): 57–116.

Byrd, William A., and Qingsong Lin. 1990. "China's Rural Industry: An Introduction." In *China's Rural Industry*, edited by William A. Byrd and Lin Qingsong. New York: Oxford University Press.

Chang, Chun, and Yijiang Wang. 1994. "The Nature of the Township Enterprise." *Journal of Comparative Economics* 19(3): 434–52.

Che, Jiahua, and Yingyi Qian. 1998. "Insecure Property Rights and Government Ownership of Firms." *Quarterly Journal of Economics* 113(2): 467–96.

Chen Muhua. 1987. *Zhongguo Muqian Jinrong Gongzuo [Current Chinese Monetary Work]*. Beijing: Zhongguo Jinrong Chubanshe.

China Finance Association. 1986. *Almanac of China's Finance and Banking*. Beijing: China Finance Press.

China Finance Association. 1997. *Zhongguo Jinrong Nianjian 1997 [Chinese Financial Yearbook 1997]*. Beijing: Zhongguo Jinrong Nianjian Bianjibu.

China Finance Association. 2000. *Almanac of China's Finance and Banking 2000*. Beijing: China Finance Press.

Dai Xianglong. 1997. "Ba Nongcun Xinyongshe Banzheng Hezuo Jingrong Zuzhi Genghao De Zhichi Nongye He Nongcun Jingji Fanzhan [Convert Rural Credit Unions into Cooperative Financial Organizations; Further Support Agriculture and Rural Economic Development]." In *Nongcun Xinyong Hezuoshe Wenjian Xuanbian 1995–1997 [Selection of Documents on Rural Credit Cooperatives 1995–1997]*, edited by the People's Bank of China. Beijing: People's Bank of China.

Dougherty, Sean, and Richard Herd. 2005. "Fast-Falling Barriers and Growing Concentration: The Emergence of a Private Economy in China."

Paris: Organisation for Economic Co-operation and Development.

Gregory, Paul R., and Robert C. Stuart. 1981. *Soviet Economic Structure and Performance.* New York: Harper & Collins.

Han Lei. 1984. "Zai Zhongguo Nongye Yinghang Quanguo Fenhang Hangzhang Zuotanhui Shang De Kaimuci [Opening Speech at the Conference of Branch Presidents of the Agricultural Bank of China]." In *Nongcun Jingrong Guizhang Zhidu Xuanbian [Selections of Regulations on Rural Finance]*, edited by the General Office of Agricultural Bank of China. Beijing: Zhongguo Jingrong Chubanshe.

Hausmann, Ricardo, Lant Pritchett, and Dani Rodrik. 2004. "Growth Accelerations." NBER Working Paper 10566.

Hewett, Edward A. 1988. *Reforming the Soviet Economy: Equality versus Efficiency.* Washington DC: Brookings Institution Press.

Huang, Yasheng. 2008. *Capitalism with Chinese Characteristics: Entrepreneurship and State during the Reform Era.* New York: Cambridge University Press.

Huang, Yasheng, and Yi Qian. 2010. "Is Entrepreneurship Missing in Shanghai?" Chapter 10 in *International Differences in Entrepreneurship*, edited by Joshua Lerner and Antoinette Shoar. Chicago: University of Chicago Press.

Jilin Branch of the People's Bank of China. 1987. *Jingrong Gaige Ziliao Huibian [Compilation of Materials on Financial Reforms]*. Changchun.

Lau, Lawrence J., Yingyi Qian, and Gerard Roland. 2000. "Reform without Losers: An Interpretation of China's Dual-Track Approach to Transition." *Journal of Political Economy* 108(1): 120–43.

Li, David. 1996. "A Theory of Ambiguous Property Rights in Transition Economies." *Journal of Comparative Economics* 23(1): 1–19.

Ministry of Agriculture. 1985. *ZhongGuo Nongye Nianjian 1985 [China Agricultural Statistical Yearbook 1985]*. Beijing: Nongye Chubanshe.

Ministry of Agriculture. 2003. *Zhongguo Xiangzhen Qiye Tongji Ziliao (1978–2002) [Statistical Materials of Township and Village Enterprises in China 1978–2002]*. Beijing: Zhongguo Nongye Chubanshe.

National Bureau of Statistics. 2007. *China Yearbook of Rural Household Survey 2007.* Beijing: China Statistics Press.

National Bureau of Statistics. 2011. *Statistical Yearbook of China 2011.* Beijing: China Statistics Press.

Naughton, Barry. 2007. *The Chinese Economy: Transitions and Growth.* Cambridge: MIT Press.

People's Bank of China. [1998] 1999. "Nongcun Xinyong Hezuoshe He Nongcun Xinyong Hezuoshe Lianshe Zhuyao Fuzeren Renzhi Zige Guanli Shixing Banfa [Provisional Methods on Managing Appointment Qualifications of the Principals of the Rural Credit Cooperatives and the Associations of Rural Credit Cooperatives]." In *Nongcun Xinyong Hezuoshe Wenjian Xuanbian [Selection of Documents on Rural Credit Unions]*, edited by the General Office of the People's Bank of China. Beijing: People's Bank of China.

People's Bank of China. 2001. *Guanyu Jiaqiang Dui Nongcun Xinyongshe Jianguan Youguan Wenti De Tongzhi [Circular on Strengthening the Supervision of Rural Credit Cooperatives]*, edited by the General Office of the People's Bank of China. Beijing: Zhongguo Jingrong Chubanshe.

Qian, Meijun, and Yasheng Huang. 2012. "How Gradualist are Chinese Reforms? Evidence from Rural Finance." Unpublished paper, National University of Singapore Business School.

Qian, Yingyi. 1999. "The Institutional Foundations of China's Market Transition." Paper presented at The World Bank Annual Conference on Development Economics, April 28–30, in Washington, DC.

Riskin, Carl. 1987. *China's Political Economy.* New York: Oxford University Press.

Roland, Gerard. 2000. *Transition and Economics: Politics, Markets, and Firms.* Cambridge: MIT Press.

Rural Economy Research Team. 1998. "Zhongguo Nongcun Gaige Yu Shichang Jingji [China's Rural Reforms and Market Economy]." *State Council Development Research Center Research Report*, December 2, 1–16.

Rural Work Leadership Team of Fujian Communist Party Committee. 1997. *Nongcun Zhengce Fagui Huibian [Selections of Rural Policy Documents]*. Fuzhou: Fuzhou Provincial Government Agricultural Office.

Shanxi Township and Village Enterprise Management Bureau. 1985. *Xiangzhen Qiye Guanli Jianmin Cidian [A Concise Dictionary of Township and Village Enterprise Management]*. Taiyuan: Shanxi Renmin Chubanshe.

Shi Jiliang. 1999. "Zai Quanguo Nongcun Xinyongshe Gongzuo Huiyi Shang De Zongjie Jianghua [Concluding Speech at the Work Conference of Rural Credit Unions]." In *Nongcun Xinyong Hezuoshe Wenjian Xuanbian [Selection of Documents on Rural Credit Unions]*, edited by the General Office of the People's Bank of China. Beijing: People's Bank of China.

State Council. [1993] 1994. *Guanyu Jingrong Tizhi Gaige De Jueding [Decision on Financial System Reform]*. Beijing: Zhongguo Jingrong Chubanshe.

State Council. 1998. *Feifa Jingrong Jigou He Feifa Jingrong Yewu Huodong Qudi Banfa [Procedures to Ban Illegal Financial Institutions and Illegal Financial*

Operations and Practices]. Edited by the General Office of the People's Bank of China. Beijing: Zhongguo Jingrong Chubanshe.

Stiglitz, Joseph E. 2006. "The Transition from Communism to Market: A Reappraisal after 15 Years." In *European Bank for Reconstruction and Development Annual Meeting.* London.

***Township and Village Enterprise Yearbook* Editorial Committee.** *1989a. Zhongguo Xiangzhen Qiye Nianjian 1989 [China Township and Village Enterprise Yearbook 1989].* Beijing: Zhongguo Nongye Chubanshe.

***Township and Village Enterprise Yearbook* Editorial Committee.** 1989b. *Zhongguo Xiangzhen Qiye Nianjian (1978–1987) [China Township and Village Enterprise Yearbook (1978–1987)].* Beijing: Zhongguo Nongye Chubanshe.

***Township and Village Enterprise Yearbook* Editorial Committee.** 1990. *Zhongguo Xiangzhen Qiye Nianjian 1990 [China Township and Village Enterprise Yearbook 1990].* Beijing: Zhongguo Nongye Chubanshe.

***Township and Village Enterprise Yearbook* Editorial Committee.** 1991. *Zhongguo Xiangzhen Qiye Nianjian 1991 [China Township and Village Enterprise Yearbook 1991].* Beijing: Zhongguo Nongye Chubanshe.

Tsai, Kellee S. 2002. *Back Alley Banking: Private Entrepreneurs in China.* Ithaca: Cornell University Press.

***Wenshou Financial History* Editorial Committee.** 1995. *Wenzhou Shi Jingrong Zhi [Wenzhou Financial History].* Shanghai: Shanghai Kexun Jishu Wenxuan Chubanshe.

World Bank. 1989. *India: An Industrializing Economy in Transition.* Washington, DC: World Bank.

Amy Finkelstein: 2012 John Bates Clark Medalist

Jonathan Levin and James Poterba

Amy Finkelstein is the 2012 recipient of the John Bates Clark Medal from the American Economic Association. A notable feature of Amy's research program is the dedication with which she has pursued her core interests in insurance markets and health care. She has addressed, with new insights at each turn, central issues such as whether asymmetric information leads to inefficiencies in insurance markets, how large social insurance programs affect healthcare markets, and the determinants of innovation incentives in health care.

Amy's work is primarily empirical, but it is usually motivated by a deep connection to economic theory. She has designed empirical tests that shed light on competing theories and uncovered data patterns that provide the starting point for theoretical developments. By blending theory, problem-appropriate empirical methods, and high-quality administrative data, Amy has been able to make progress on long-standing questions that have defied easy resolution and to advance substantially our understanding of the economics of insurance markets and the markets for healthcare services more generally.

Amy was born and grew up in New York City. Her parents, Alan and Joan Finkelstein, are both biologists with Ph.D.s from Rockefeller University. Amy's mother left Poland and immigrated to the United States with her parents in 1940 at the age of three. Amy's maternal grandmother received a Ph.D. in comparative literature from the University of Warsaw in 1928, which makes three generations of doctorate-holding women in Amy's family.

■ *Jonathan Levin is Holbrook Working Professor of Economics, Stanford University, Stanford, California. James Poterba is Mitsui Professor of Economics, Massachusetts Institute of Technology, and President of the National Bureau of Economic Research, both in Cambridge, Massachusetts. Their email addresses are jdlevin@stanford.edu and poterba@mit.edu.*

Amy began her stellar academic career at the Brearley School in New York City. Perhaps anticipating a future of entertaining lectures and seminar presentations, she earned pocket money as a high school student by working as a juggler and clown at birthday parties. Later, as an undergraduate at Harvard College, Amy majored in government. She also took a number of economics classes, however, and courses by Larry Katz and Alberto Alesina particularly attracted her to the field. When Amy received a Marshall Scholarship for postgraduate study in the United Kingdom, she enrolled in Oxford's two-year M.Phil. program in Economics. She received her degree in 1997. She then returned to the United States and spent a year in Washington, D.C., working at the Council on Economic Advisors. This experience proved formative, as Amy became interested in insurance markets. She also became convinced of the value of economics after seeing how the staff members at the Council were able to articulate clear frameworks for analyzing policy decisions.

Amy started graduate school at the Massachusetts Institute of Technology in the fall of 1998. She completed her dissertation in three years, working under the guidance of James Poterba, Jonathan Gruber, and Jerry Hausman. During her time as a graduate student, she overlapped with the last five winners of the Clark Medal: Emmanuel Saez, Esther Duflo, and Jonathan Levin were fellow Ph.D. students, while Daron Acemoglu and Susan Athey were junior faculty members.

A particularly notable event in Amy's graduate school career occurred toward the end of her second year, in the spring of 2000, when she met Harvard graduate student Ben Olken at a seminar given by Steve Levitt. After a spirited debate about Levitt's conclusions—the seminar was on Levitt's well-known work linking the legalization of abortion in the early 1970s to falling crime rates two decades later—Amy proposed that they continue their discussion over dinner that Friday. Ben counteroffered with the less-romantic suggestion of Thursday lunch, but, happily, he recovered from this early courtship mistake. Following in family tradition—her parents also met while in graduate school—Amy married Ben in 2005.

After receiving her Ph.D. from MIT in 2001, Amy spent a year as a postdoctoral fellow at the National Bureau of Economic Research (NBER) in Cambridge, with the support of the National Institutes of Health. She then was elected to the Harvard Society of Fellows, where she spent three years as a Junior Fellow. During this time, she wrote several influential papers on adverse selection in insurance markets and on innovation, and quickly established herself as rising star in applied economics. Amy returned to MIT as a faculty member in 2005, and was promoted to tenure two years later. Her husband, Ben, who was also a member of the Harvard Society of Fellows, joined the MIT economics faculty in 2008. Amy is currently the Ford Professor of Economics at MIT, and she is also the codirector, along with Raj Chetty of Harvard, of the NBER's Public Economics Program. In these roles, Amy has distinguished herself not just for her research but for her advising and professional leadership. Her colleagues and coauthors note that Amy stands out for her abundant energy, her enviable organizational skills, and her upbeat and warm sense of humor. Her students at MIT have repeatedly recognized her for her commitment to teaching and advising.

Amy Finkelstein

In the balance of this essay, we describe a number of Amy's key research contributions, with particular emphasis on those that were identified by the Honors and Awards Committee of the American Economic Association in her Clark Medal citation, as well as her broader contributions to the field of economics. Table 1 provides numerical references to Amy's papers cited in this essay.

Asymmetric Information in Insurance Markets

One recurring theme in Amy's research is how asymmetric information in insurance markets can affect the behavior of insurance buyers, insurance companies, and market equilibrium. Her analyses of long-term care insurance, annuities, and health insurance have helped to establish the empirical importance of asymmetric information in insurance markets, and also have explored the interaction between public policies and private market outcomes.

Simple theoretical models predict that the losses of those with greater insurance coverage should be greater than those with less coverage or no coverage at all. This outcome can arise either from selection, if individuals who are at high risk purchase more insurance, or from moral hazard, if individuals who purchase more insurance take fewer precautions against loss. Perhaps surprisingly, however, empirical studies of insurance markets have not always shown a positive correlation between insurance coverage and subsequent loss. Indeed, early studies of auto and

Table 1
Selected Papers by Amy Finkelstein

1. "The Effect of Tax Subsidies to Employer-Provided Supplementary Health Insurance: Evidence from Canada." 2002. *Journal of Public Economics* 84(3): 305–340.
2. "Selection Effects in the Market for Individual Annuities: New Evidence from the United Kingdom," (with James Poterba). 2002. *Economic Journal* 112(476): 28–50.
3. "Minimum Standards, Insurance Regulation and Adverse Selection: Evidence from the Medigap Market." 2004. *Journal of Public Economics* 88(12): 2515–47.
4. "Static and Dynamic Effects of Health Policy: Evidence from the Vaccine Industry." 2004. *Quarterly Journal of Economics* 119(2): 527–64.
5. "Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market," (with James Poterba). 2004. *Journal of Political Economy* 112(1): 183–208.
6. "The Interaction of Partial Public Insurance Programs and Residual Private Insurance Markets: Evidence from the U.S. Medicare Program." 2004. *Journal of Health Economics* 23(1): 1–24.
7. "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market," (with Kathleen McGarry). 2006. *American Economic Review* 96(4): 938–58.
8. "Why is the Market for Long-Term Care Insurance So Small?" (with Jeffrey R. Brown). 2007. *Journal of Public Economics* 91(10): 1967–91.
9. "The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare." 2007. *Quarterly Journal of Economics* 122(3): 1–37.
10. "Input and Technology Choices in Regulated Industries: Evidence from the Health Care Sector," (with Daron Acemoglu). 2008. *Journal of Political Economy* 116(5): 837–80.
11. "The Interaction of Public and Private Insurance: Medicaid and the Long-Term Care Insurance Market," (with Jeffrey R. Brown). 2008. *American Economic Review* 98(3): 1083–1102.
12. "What Did Medicare Do? The Initial Impact of Medicare on Mortality and Out of Pocket Medical Spending," (with Robin McKnight). 2008. *Journal of Public Economics* 92(7): 1644–69.
13. "EZ-Tax: Tax Salience and Tax Rates." 2009. *Quarterly Journal of Economics* 124(3): 969–1010.
14. "Redistribution by Insurance Market Regulation: Analyzing a Ban on Gender-Based Retirement Annuities," (with James Poterba and Casey Rothschild). 2009. *Journal of Financial Economics* 91(1): 35–58.

life insurance showed little relationship, raising the question of whether asymmetric information was an important factor in these markets, and more generally whether the classic models of insurance market equilibrium under asymmetric information were empirically relevant.

Amy tackled this issue in joint work with Kathleen McGarry [7] on the long-term care insurance market. Their analysis postulates that potential insurance buyers differ on two dimensions: their risk of experiencing an insured event, and their risk aversion. For a given risk aversion level, individuals with greater risk of loss—that is, with a higher probability of needing long-term care—should be more likely to purchase policies. At the same time, for a given probability of loss, individuals who are more risk averse should be more likely to purchase policies. When the two dimensions of heterogeneity are both present, however, whether or not policyholders will experience a higher rate of claims is ambiguous: it depends on the relative importance of, and the correlation between, risk aversion and risk type.

Amy and Kathleen explored this theory using data from the Asset and Health Dynamics Among the Oldest Old (AHEAD) survey. Their analysis showed that

Table 1—continued

15. "The Private Market for Long-Term Care Insurance in the U.S: A Review of the Evidence," (with Jeffrey R. Brown). 2009. *Journal of Risk and Insurance* 76(1): 5–29.
16. "Estimating Welfare in Insurance Markets Using Variation in Prices," (with Liran Einav and Mark Cullen). 2010. *Quarterly Journal of Economics* 123(3): 877–921.
17. "Optimal Mandates and the Welfare Cost of Asymmetric Information: Evidence from the U.K. Annuity Market," (with Liran Einav and Paul Schrimpf). 2010. *Econometrica* 78(3): 1031–92.
18. "Insuring Long-Term Care in the United States," (with Jeffrey R. Brown). 2011. *Journal of Economic Perspectives* 25(4): 119–42.
19. "The Effects of Medicaid Coverage–Learning from the Oregon Experiment," (with Katherine Baicker). 2011. *New England Journal of Medicine,* published July 20 at NEJM.org. doi: 10.1056/nejmp1108222.
20. "Selection in Insurance Markets: Theory and Empirics in Pictures," (with Liran Einav). 2011. *Journal of Economic Perspectives* 25(1): 115–38.
21. "Beyond Testing: Empirical Models of Insurance Markets," (with Liran Einav and Jonathan Levin). 2011. *Annual Review of Economics* 2: 311–36.
22. "Selection on Moral Hazard in Health Insurance," (with Liran Einav, Paul Schrimpf, Stephen Ryan, and Mark Cullen). *American Economic Review,* forthcoming.
23. "The Oregon Health Insurance Experiment: Evidence from the First Year," (with Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, and the Oregon Health Study Group). 2012. *Quarterly Journal of Economics* 127(3): 1057–1106.
24. "How General are Risk Preferences? Choices under Uncertainty in Different Domains," (with Liran Einav, Iuliana Pascu, and Mark Cullen). *American Economic Review* 102:(6): 2606–38.
25. "Moral Hazard in Health Insurance: How Important is Forward Looking Behavior?" (with Aviva Aron-Dine, Liran Einav, and Mark Cullen). 2012. NBER Working Paper 17802.

individuals' subjective assessments of future nursing home use are predictive of subsequent utilization, even after controlling for observable attributes that the insurance company uses in setting the prices of its policies. Thus, individuals seem to have some private information about their potential long-term care needs. At the same time, individuals in their sample who buy insurance also are more likely to engage in health-promoting "preventive behaviors," such as self-care and seat belt use, that may be correlated with risk aversion. Thus, individuals who are relatively cautious about their personal care ("belt and suspenders" types) are also more likely to purchase long-term care.

These findings provide a way to understand one of the main findings in Amy and Kathleen's paper: in contrast to the basic theoretical prediction, the unconditional correlation between insurance coverage and loss is not statistically significantly different from zero. The presence of multiple dimensions of private information provides an explanation for this pattern because some types of private information, such as beliefs about the likelihood of nursing home use, are positively correlated with the demand for long-term care insurance and with the likelihood of receiving payment from the policy, while other types of private information, such as seat belt use, are positively correlated with the demand for insurance but negatively correlated with the likelihood of policy payouts. This paper, and particularly Amy and Kathleen's observations about the relationship between risk-of-loss and insurance

purchases, has inspired a growing literature that focuses on broader forms of consumer heterogeneity in insurance markets and other contracting environments.

Annuities are another insurance market characterized by asymmetric information and adverse selection. Voluntary annuity markets in most countries are quite small, even though standard lifecycle theories suggest that many households might benefit from annuitizing at least some of their wealth, thereby increasing their insurance against living unexpectedly long. Adverse selection is one potential explanation for the limited size of the annuity market. If the only buyers of annuities are those who expect to live for many years, and if these expectations are correlated with actual life length, then an annuity policy that is actuarially fair for annuity buyers will be unattractive from the perspective of an individual with the population-average mortality rate. Amy has examined the annuity market in detail, generating both new insights about this specific market and about the welfare economics of insurance markets more generally.

Amy's first two studies of annuity markets, written with James Poterba, compared the degree of adverse selection in various segments of the U.K. annuity market. Until 2011, participants in many U.K. retirement plans were required to purchase annuities, and so that nation has one of the largest markets for annuity contracts. One study [2] contrasts the voluntary U.K. annuity market, in which individuals purchase annuities with wealth that has been accumulated outside retirement accounts, with the compulsory market, in which individuals who have accumulated resources in defined contribution pension plans are required to annuitize a fraction of these assets. The participants in such pension plans—the "compulsory annuitants"—are not a random sample of the population, but because the decision to work for a particular employer is typically made many years before retirement, these participants are less likely to be self-selected on the basis of mortality prospects than are those who purchase annuities in the voluntary market. The data on mortality rates support this proposition. The age-specific mortality rates for the population as a whole are greater than those for compulsory annuitants, which are in turn greater than those for voluntary annuitants. This finding is consistent with a greater degree of adverse selection in the voluntary than the compulsory annuity market.

The other Finkelstein and Poterba study [5] of annuities examines a rich data set on those who purchased annuity policies from a large U.K. insurance company over a two-decade span. The insurance company offered policies with different time profiles of payouts, including a level nominal payment policy as well as one with payouts that rose at the rate of inflation. A central contribution of this study is its focus on the choice of insurance policy attributes, rather than the purchase of the policy per se, as a manifestation of private information. The empirical findings suggest that individuals who chose back-loaded annuity policies where the payouts rose over time, such as inflation-indexed policies, on average lived longer than those who bought annuities that offered most of their benefits "up front." In addition, those who purchased annuities that promised survivor benefits to the annuitant's heirs on average predeceased those who bought annuities without this clause. The findings provide further support for the role of private information in

the annuity market, and they suggest the possibility of designing richer tests than discrete comparisons between those who have and do not have insurance.

Amy's subsequent collaboration with Liran Einav and Paul Schrimpf [17] built on these insights about annuity purchasing by developing and estimating a structural model of insurance demand that allowed for what may be the first empirical estimates of the welfare consequences of asymmetric information. It also illustrates the complexity of policy design: for example, it shows that in certain conditions, a requirement that everyone participate in the insurance market may reduce welfare relative to the equilibrium of an imperfect private market with adverse selection. The structural modeling approach taken in the paper presents an interesting contrast to Amy's earlier work. It requires some strong assumptions about how individuals form expectations and make choices. But in turn, these assumptions make it possible to analyze a far wider range of actual and hypothetical policies than in the earlier studies that limited attention to the correlation between insurance choices and subsequent mortality (given existing policies).

One lesson from Amy's work on insurance market equilibrium is that the degree of adverse selection and the extent to which risks are pooled depends a great deal on whether insurers can condition prices on individual characteristics. In [14], with Poterba and Casey Rothschild, Amy investigates how restricting the information on which insurers can condition prices affects market outcomes. As an example of its more general findings, the paper rejects the simple notion that banning the use of gender in pricing annuities will inevitably redistribute from men—who on average die sooner and therefore receive higher annuity payouts in a gender-separating equilibrium—to women. Instead it suggests that insurers may find a range of ways to induce self-selection in the insurance market, for instance by offering back-loaded and front-loaded policies. If the policies are priced so that women find the back-loaded policies more attractive, and men the front-loaded, then the selection generated by these voluntary choices will reduce the degree of redistribution associated with a ban on gender-based pricing.

Adverse Selection and Moral Hazard in Health Insurance

Another component of Amy's work on insurance markets has been a fruitful collaboration with Einav and Mark Cullen, focusing primarily on problems of adverse selection and moral hazard in health insurance.

In [16], they showed how variation in insurance prices can be used to obtain estimates of insurance demand and coverage costs, which in turn allow for estimates of the consumer and producer surplus that might result from different pricing arrangements. Amy and her coauthors exploit a rich data set on the health insurance offerings that are available to employees at a large multidivision U.S. company. The "price" variation that underlies their empirical work arises from the fact that different employees within the company are charged different premiums for purchasing more comprehensive health insurance. Employees face a number

of choices with regard to copay rates, deductibles, and corresponding monthly premium costs.

The paper explores the empirical implications of the well-known principle that when an insurance company offers a suite of policies to a set of potential buyers, changes in the price of a given policy can affect the cost of delivering that policy through the induced effect on the set of consumers who purchase the policy. Like some other studies of employer-sponsored health plan choice, this project yields clear evidence of adverse selection, with the costliest individuals to cover also being the most eager to enroll in insurance plans with generous coverage. At the same time, however, the results suggest that the conditions for large welfare distortions are not met.

This result—that the presence of adverse selection does not itself imply large welfare losses—deserves elaboration. In a competitive insurance market, the potential for welfare loss arises because at a zero-profit equilibrium the market price for insurance is equal to the cost of covering the average enrollee. With adverse selection, this price is above the cost of covering the marginal enrollee, so that too few individuals purchase insurance. Practically speaking, welfare losses are likely to be large only when adverse selection results in a substantial price distortion and there is enough price-sensitivity on the part of consumers that the overly high price deters many from purchasing. Amy's paper was among the first to show empirically that the adverse selection could be substantial, yet the resulting welfare distortion relatively modest.

One especially nice feature of Amy's [16] paper with Einav and Cullen is that it explains welfare distortions under adverse selection using a demand and supply approach that connects to more standard partial equilibrium welfare analyses of taxes or subsidies. Amy's paper with Einav in this journal, [20], walks through this approach graphically, and makes for a useful and relatively light introduction to the recent empirical literature on adverse selection.

In a follow-up to this work using data from the same firm [22], the same coauthors, together with Schrimpf and Stephen Ryan, explore the relationship between adverse selection and moral hazard in health insurance. They begin with the observations that individuals differ in how they respond to more- and less-generous coverage and that these differences can generate a novel form of selection in insurance purchasing. For instance, moving someone into a high-deductible plan might lead to a considerable reduction in healthcare utilization. But if a company introduces a high-deductible plan as one of several options, the effect may be muted if the individuals who select it are precisely those whose healthcare utilization is relatively insensitive to out-of-pocket costs. This study exploits the rich data on insurance offerings and worker choices not only to document differences in individual price responsiveness, but also to estimate a structural model that incorporates both plan selection and healthcare utilization. The model suggests that "selection on moral hazard" indeed may mitigate some of the anticipated healthcare spending reductions from offering plans with high cost sharing, as price-responsive individuals tend to avoid these plans.

The degree to which individual healthcare utilization responds to changes in out-of-pocket costs is a much-debated empirical question, and it lies at the center of many proposals for insurance market reform. The question is difficult to answer for several reasons. First, an appropriate research design would approximate random assignment of individuals into different cost sharing arrangements—but such an arrangement is difficult to find in practice. Second, most insurance plans are highly nonlinear, often involving a deductible, a coinsurance rate, and an out-of-pocket maximum, so it often is not clear what "price" consumers face. In [25], Amy with Avia Aron-Dine, Einav, and Cullen develop a clever strategy for analyzing how consumers respond to nonlinear insurance coverage. They rely on the fact that coverage is not prorated for individuals hired at different points during a calendar year. A worker who is hired in November, and who participates in a plan with a $500 annual per-worker deductible, is much less likely to have medical expenses in excess of the deductible than is a worker who is hired in January and who participates in the same plan for the whole calendar year. By comparing individuals hired at different points in time, some of whom will later exceed their deductible, it is possible to assess whether individuals respond to their expected out-of-pocket price given the level of total medical expenditures that they may experience before year-end. The paper finds that individual behavior is somewhat, but not fully, forward looking.

A related paper with Einav, Iuliana Pascu, and Cullen [24] returns to the subject of heterogeneity in insurance demand, and studies the extent to which individuals display similar degrees of risk aversion in various choice domains. It finds that product choice in one insurance market can be quite informative about product choice in other insurance markets but that there is less association between choices in insurance markets and other financial decisions, such as the stock-bond mix in a retirement plan.

A common feature of Amy's health insurance papers as well as some of her work on annuities is that they involve analysis of large datasets consisting of administrative records that include information about insurance plan choices and subsequent insurance claims. This type of data has become an important resource for economists working on insurance markets. Amy's survey paper with Einav and Levin [21] discusses some of the issues that arise in modeling and estimating empirical models of insurance demand and costs with this type of data. The paper explains that there are two general strategies for assessing pricing incentives or welfare losses under asymmetric information. One involves estimating how changes in price affect consumer purchases and insurer costs, while the other involves estimating parameters such as individual risk-aversion in a modeling framework that specifies exactly how and why consumers derive utility from insurance.

Government Provision and Regulation of Health Insurance

Issues concerning the economic effects of government provision of health insurance arise in many settings, including the Medicare and Medicaid programs,

and the ongoing national debate about health insurance coverage and market structure. Amy has made several important contributions in this area.

One of Amy's best-known papers, [9], examines how the introduction of Medicare affected the demand for health care in the United States. Many noneconomists expect that the answer to this question is well-known, since Medicare represented a dramatic change in health insurance provision. But because the change affected the whole nation, empirical researchers have found it difficult to construct a counterfactual that can be used to understand how the presence of Medicare has affected the healthcare marketplace.

Amy devised an insightful identification strategy to circumvent this difficulty by comparing the effect of the introduction of Medicare on hospital expenditures by elderly households in different regions of the United States. Prior to the introduction of Medicare, the incidence of hospitalization insurance policies varied widely across regions, so the fraction of elderly households who experienced changes in their ability to pay for hospital-based care when Medicare was introduced also varied across regions. Although the policy change was national, the effect of the policy relative to the prior situation varied substantially across regions.

Amy's study finds that in places where Medicare had an especially large effect in expanding insurance coverage, there was an especially large increase in health expenditures. A related study with Robin McKnight [12] finds that the introduction of Medicare affected the distribution of out-of-pocket medical expenditures for elderly households in the United States, although it had no discernible impact on mortality rates. This work on the introduction of Medicare is arguably the most convincing empirical analysis to date of how the most dramatic change in health insurance policy in U.S. history affected the consumption and financing of health care.

Another important and related project evaluates the effect of expanding Medicaid coverage. This project, which Amy has carried out in conjunction with a large team of collaborators including Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph Newhouse, Heidi Allen, Katherine Baicker, and the Oregon Health Study Group [19, 23], is remarkable on many dimensions. One is the ingenuity of the research approach. In 2008, Oregon expanded the coverage of its Medicaid program but did not have enough resources to allow universal access to this program. As a result, access to coverage was allocated by lottery, thereby creating a large-scale randomized experiment. After learning about this lottery system in the news media, Amy and her fellow researchers rushed to obtain funding and to field a large-scale mail survey collecting data on insurance status, healthcare utilization, financial strain, and health conditions among those who would potentially be eligible for Medicaid coverage. They subsequently combined this data with information from credit reports, hospital discharges, and mortality records. For a subset of the population, they have also conducted in-person interviews and physical health exams, and collected biometric data as well. In a remarkably short period of time, the research team has amassed an astonishing array of information about the individuals who were potentially affected by the Oregon health insurance lottery.

The early findings from the Oregon study suggest that the insurance expansion increased utilization of medical care and also that it reduced out-of-pocket health costs and led to higher self-reported physical and mental health status. Subsequent analysis using the biometric data are likely to shed light on whether increased care leads to measureable short-term clinical health benefits. Results from analyzing the Oregon experiment have just begun to appear. But because of the randomized access to health insurance that underlies this study, and the carefully planned and executed statistical analysis, the results are likely to be viewed as the "gold standard" for evaluating how health insurance availability affects household behavior and welfare.

In addition to her work on Medicare and Medicaid, Amy has tackled a number of other issues related to government's role in health insurance markets. Her earliest work on health insurance, [1], explored the effect of the tax exclusion for health insurance premiums paid by employers. Existing studies of this topic have been hamstrung by the empirical difficulty of distinguishing differences in marginal tax rates across households from differences in other factors that may affect the demand for health insurance. Amy examined a provincial tax reform in Quebec in 1993 that substantially reduced the tax subsidy to employer-provided health insurance. By comparing the corresponding change in health insurance demand in Quebec with the change in health insurance demand in other provinces, she concluded that tax subsidies have significant effects on health insurance demand. In the case of Quebec, the estimates suggest that insurance coverage fell 20 percent as a result of the tax change. This finding implied a price elasticity of employer demand for insurance coverage with respect to the after-tax price of such coverage of about −0.5.

A related paper [3] studies another government policy that affects insurance markets: minimum coverage requirements. There is a long-standing debate in insurance markets and other settings about whether government minimum quality standards raise welfare. On one side, minimum quality standards may raise product cost and lead some consumers to drop out of markets in which they would otherwise have purchased lower-quality and less-expensive goods. This effect tends to reduce welfare. However, by providing consumers who continue to purchase the product with a form of quality assurance, such regulations might raise welfare. Amy evaluates this trade-off in the case of the health insurance market for elderly Americans. In the early 1980s, the federal government encouraged states to require that all insurers offering policies designed to supplement the basic Medicare insurance package—so-called "Medigap policies"—had to provide a specified level of minimum coverage. Comparing insurance purchases by elderly households before and after the imposition of these minimum standards reveals a substantial decline in the overall rate of coverage, suggesting that the negative effect of minimum standards on quantity demanded operated for at least a significant share of Medicare beneficiaries.

Amy's focus on the detailed provisions of insurance markets and on how they affect market participants and market equilibrium is also evident in [6], a study that explores how the provision of partial public insurance may affect insurance market outcomes. Amy explores this issue by considering how the availability of Medicare,

which is partial public insurance for healthcare outlays, affects the private market for insurance for residual, Medicare-uninsured, risks. She focuses on insurance for prescription drugs, which were not covered by Medicare during the time period of her study. Amy notes that prior to age 65, many individuals are covered by health insurance policies that provide both hospitalization coverage and prescription drug coverage. At age 65, however, when they become eligible for Medicare, they may lose their prescription drug coverage. While this study finds little change in the overall probability of individuals having prescription drug insurance before and after age 65, it does uncover a shift in the set of individuals who have this coverage. After age 65, the coverage rate among the least healthy members of the population rises, while that among healthier individuals remains stable or declines. In theoretical models of private market insurance, it is possible that the introduction of partial coverage by the government can upset a pooling equilibrium in which different risk types purchase the same insurance coverage, and lead to a separating equilibrium in which low-risk individuals will opt not to purchase coverage. The findings in Amy's study are consistent with the emergence of a separating equilibrium for the post-65 population with the highest risk group, those individuals in poor health, receiving full insurance while the better risks do not have prescription drug insurance.

Long-Term Care Insurance

Amy has devoted substantial research effort to understanding the market for long-term care insurance. She studied the role of multidimensional heterogeneity in this market in her joint work with McGarry; in a related sequence of papers with Jeffrey Brown, [8, 11, 15], she has studied other aspects of this insurance market. Some of her work with Brown, [18], appeared in this journal.

While aging individuals face substantial risk that they will require costly long-term care, relatively few purchase long-term care insurance. In [8], Brown and Finkelstein estimate the load factors on long-term care insurance policies. While the expected payouts on these policies are typically lower than their premiums, making them less than actuarially fair, the divergence from the benchmark of actuarial fairness is not substantially different from the divergence in a number of other insurance markets in which there is more active consumer interest.

A related study, [11], suggests that Medicaid, which covers the cost of long-term care for households that exhaust their assets, may play a key role in crowding out the private insurance market. For many households, Medicaid is a substitute for private long-term care insurance, and its presence reduces the willingness to pay for private insurance. In particular, private insurance policies must continue to pay even if someone becomes Medicaid-eligible, which makes the cost of private long-term care insurance high relative to the net-of-Medicaid benefits received by an insurance buyer.

The study undertakes a quantitative analysis to see whether the implicit tax that Medicaid imposes on private insurance benefits can explain why the market is so

small. The substantive importance of Medicaid depends on the distribution of wealth and income among elderly households because some households with high levels of financial wealth, who are very unlikely to draw down their assets completely, are also unlikely to be affected by the presence of Medicaid. For low-wealth households, in contrast, the implicit tax effects can be much more important. Using a carefully calibrated model along with household-level data on income and wealth, this study finds that a substantial fraction of elderly households face very high implicit taxes, which result in private long-term care insurance having relatively low value once Medicaid is taken into account. The results in these papers suggest a straightforward demand-side explanation for the limited interest in private long-term care insurance.

Innovation in Health Care

Explaining the rising cost of health care relative to other goods is one of the central questions in both health economics and public finance because health costs are a crucial component of government spending. One frequently mentioned hypothesis holds that technological change in the healthcare sector has produced increasingly expensive treatments for various diseases. To evaluate this possibility, and how public policy might affect the future course of innovation, it is important to understand the extent to which new product introductions in health care are "supply driven"—that is, determined by the current state of technical knowledge—or "demand driven," and thus determined by the market returns from a new product. Amy has studied this issue in several contexts.

In [4], Amy explores how changes in federal health insurance rules with regard to vaccine coverage, and other changes that affected the demand for specific vaccines, affected the level of private research and development spending on new vaccines. Amy identifies three policy changes that increased the potential benefits to pharmaceutical firms producing vaccines, two of which are related to insurance policy reforms. One is the 1993 expansion of Medicare to cover the cost of flu vaccines. The second is the 1986 introduction of the federal Vaccine Injury Compensation Fund, which created a trust fund to compensate those injured by receiving a vaccination—and thus reduced the potential legal liability of firms producing such vaccines. The third reform she examines, which did not operate through insurance markets but is likely to have resulted in an outward shift in the demand for vaccinations, is the 1991 recommendation by the Centers for Disease Control that all infants should receive Hepatitis B vaccinations. In the years following each of these reforms, Amy finds an increase in the number of new clinical trials for vaccines that are directed at the specific diseases covered by each reform. These results offer a novel and interesting source of evidence for the link between government policies and the nature of innovation and technical change in the healthcare sector.

Another creative paper on innovation, joint with Daron Acemoglu [10], examines how a change in Medicare that increased the share of capital relative to labor costs for which hospitals could receive reimbursement from Medicare

affected hospitals' technology choices and input mix. The paper suggests that hospitals shifted away from labor inputs after this policy change. The skill mix of hospital employees also shifted, as various models of capital–skill complementarity would predict, and technology adoption increased, suggesting a high degree of substitution between labor and technology. This paper provides further evidence for the significance of government policies in affecting innovation and technology adoption in the health sector.

Final Remarks

Our discussion of Amy's research has highlighted Amy's contributions to the economics of insurance markets and health care and also her remarkable productivity. Amy has managed to carry out several large-scale research programs while supervising a steady flow of graduate students, coparenting two small children with her husband Ben, and still finding time to send multipage, single-spaced emails to colleagues whose papers she has read. Perhaps reflecting her early training as a juggler, Amy manages these multiple roles with good cheer and good humor.

We started this essay by noting the dedication with which Amy has pursued her primary research interests. Her research has occasionally ranged farther afield, for instance in her clever study [13] showing how the introduction of automated "EZ-Pass" systems to collect bridge and toll fees has led to more frequent fee increases. However, Amy generally has targeted her research toward a well-defined set of important and challenging questions: how insurance markets are affected by problems of asymmetric information; how large social insurance programs affect healthcare markets; and how different incentives affect the rate and direction of healthcare innovation. Amy has made progress on these classic issues both by connecting economic theory to empirical application and by finding rich new data sets. These include provincial data on health insurance utilization in Canada; administrative health claims data from large employers; heretofore unavailable data from British insurance companies; data from American Hospital Association archives that retained paper records that could be digitized; and, in her recent work on Medicaid, credit histories, hospital records, and biometric data from health exams. With careful and rigorous econometric analysis, Amy's work has enriched the scholarly analysis of insurance and healthcare markets, and has provided important insights for policymakers who are considering reforms to these markets.

Retrospectives
Irving Fisher's *Appreciation and Interest* (1896) and the Fisher Relation

Robert W. Dimand and Rebeca Gomez Betancourt

This feature addresses the history of economic terms and ideas. The hope is to deepen the workaday dialogue of economists, while perhaps also casting new light on ongoing questions. If you have suggestions for future topics or authors, please write to Joseph Persky of the University of Illinois at Chicago at jpersky@uic.edu.

Introduction

From 1886 up until the start of the *American Economic Review* in 1911, the American Economic Association published the proceedings of its annual meetings along with occasional monographs; these *Publications of the American Economic Association* are now freely available at JSTOR (http://jstor.org). In an outstanding contribution to that First Series of AEA publications, Irving Fisher's monograph *Appreciation and Interest* (1896) proposed his famous equation showing expected inflation as the difference between nominal interest and real interest rates. In addition, he drew attention to insightful remarks and numerical examples scattered through the earlier literature, and he derived results ranging from the uncovered interest arbitrage parity condition between currencies to the expectations theory of the term structure of interest rates. As J. Bradford DeLong (2000, pp. 83, 85) wrote in this journal, "The story of 20th century macroeconomics begins with Irving Fisher" and

■ *Robert W. Dimand is Professor of Economics at Brock University, St. Catharines, Ontario, Canada. Rebeca Gomez Betancourt is Associate Professor of Economics, Université Lumière Lyon 2, Lyon, France. Their email addresses are rdimand@brocku.ca and Rebeca .GomezBetancourt@univ-lyon2.fr, respectively.*

specifically with *Appreciation and Interest* because "the transformation of the quantity theory of money into a tool for making quantitative analyses and predictions of the price level, inflation, and interest rates was the creation of Irving Fisher."

In 1896, Irving Fisher was an assistant professor in his 20s, just five years out of graduate school, who had been teaching mathematics rather than economics for the first four of those years. Fisher was not trained as a monetary specialist. His 1891 doctoral dissertation in mathematics and political economy (Yale's first Ph.D. in political economy or economics, see Barber 1986), which was published as *Mathematical Investigations in the Theory of Value and Prices* (1892), brought general equilibrium analysis to North America; it was supervised jointly by the physicist and engineer J. Willard Gibb and the economist and sociologist William Graham Sumner.[1] Paul Samuelson once described Fisher (1892) as the "greatest doctoral dissertation in economics ever written" (quoted by Barber, in Fisher 1997, Vol. 1, p. 4). However, Robert Dorfman (1995, footnote on p. 23) made the point that "[i]f Fisher's examiners had been better versed in European economic literature than they were, a promising career might have been blighted at its inception," because Fisher invented general equilibrium analysis for himself before his last-minute discovery of the writings of Léon Walras and Francis Ysidro Edgeworth. Fisher's thesis went beyond these writings in one striking respect: influenced by Gibbs's work in mechanics, Fisher not only imagined but actually built a hydraulic mechanism to simulate the determination of equilibrium prices and quantities—in effect, a hydraulic computer in the days before electronic computers (Brainard and Scarf 2005; Dimand and Ben-El-Mechaiekh forthcoming). His first academic appointment was in Yale's Department of Mathematics in 1891 (Fisher coauthored an elementary geometry textbook in 1896 and published a brief introduction to calculus the next year), and he did not transfer to the Department of Political Economy until the summer of 1895. Fisher's course in the mathematics department on "The Mathematical Theory of Prices," based on his dissertation, was far ahead of its time in the 1890s. A typical American course in political economy in the 1890s such as that taught by J. Laurence Laughlin at the University of Chicago still used John Stuart Mill's *Principles of Political Economy* (1848 [1871]) as the textbook; it did not use Alfred Marshall's *Principles of Economics* (1890 [1920]), let alone mention general equilibrium.

Edgeworth invited Fisher to apply a simplified version of his hydraulic model to "The Mechanics of Bimetallism" for presentation to the Economics Section of the British Association for the Advancement of Science and then publication, in 1894, in the *Economic Journal,* which Edgeworth edited. Bimetallism was a hot topic at the time. Prices tended to decline under the gold standard from 1873 to 1896 as real demand for money rose faster than the world supply of gold (a situation

---

[1] Sumner was also one of the two supervisors of Thorstein Veblen's 1884 Yale Ph.D. dissertation on Kant's ethics and so was a mentor to two outstanding young economists, with Veblen's institutionalism and Fisher's formal theorizing and pioneering econometrics offering very different paths for American political economy (Dimand 1998).

that changed after 1896 with the Witwatersrand and Klondike gold rushes and the cyanide process for extracting gold from low-grade ore). As prices fell, the real value of nominal debts increased. In the United States, Midwestern populists denounced the rising real burden of farm mortgages held by Eastern banks and, together with Western silver miners, demanded "bimetallism"—that is, increasing the money supply by free coinage of silver as well as gold. During the 1896 presidential election campaign, the Democratic and Populist nominee William Jennings Bryan in his speech accepting the Democratic nomination famously condemned the gold standard for crucifying mankind on a "cross of gold." The leading bimetallist tract, William Harvey's *Coin's Financial School* (1894 [1963]), sold perhaps a million copies, vastly exceeding the circulation of any mainstream economics book of the time (see Hofstadter's introduction to the reprint, and Willard Fisher 1896). In the book, Harvey's fictional financier Coin soundly defeated Professor J. Laurence Laughlin of the University of Chicago, a hard-money stalwart, in public debate—although when Laughlin was able to speak for himself in a real public debate with Harvey in 1895, he fared much better (Skaggs 1995; Willard Fisher 1896). The bimetallists followed the quantity theory of money in holding that an increase in the quantity of money would raise prices, but went beyond the quantity theory in insisting that a higher price level would have lasting real benefits. Laughlin and some other academic defenders of the gold standard met such populist use of the quantity theory not just by insisting on the long-run neutrality of money, but also by rejecting the quantity theory's explanation of changes in the purchasing power of money (Skaggs 1995; Gomez Betancourt 2010).

By July 1895, Fisher was writing to a friend that he was "working on an essay which will either be a long article or a short book on bimetallism *against* its expediency or necessity . . . I never was so morally aroused I think as against the 'silver craze'" (Fisher 1997, Vol. 1, p. 7, Fisher's emphasis). Fisher's *Appreciation and Interest* was presented to the American Economic Association in Indianapolis in December 1895 and then published by the association in August 1896.[2] The title featured the appreciation of the purchasing power of money during deflation rather than its depreciation in a period of price inflation. Fisher had two goals in his 1896 monograph: to show the fallacy of bimetallist claims of permanent gains from inflation while rescuing the quantity theory of money from its populist misuses. During his long and productive career, Fisher attempted statistical verification of the relation (making use of correlation analysis and introducing distributed lags into economics) and developed a monetary theory of fluctuations in economic activity based on slow adjustment of inflationary expectations (that is, modeling expected inflation with a form of adaptive expectations). In the 1930s, when propounding his

---

[2] The other three issues of that year's volume of *Publications of the American Economic Association* were, regrettably, devoted to Frederick L. Hoffman (1896), a work of racist pseudo-science by the statistician to the Prudential Insurance Company of America. Irving Fisher was also a strident eugenicist (Fisher 1997, Vol. 13, pp. 160–207; Cot 2005; Dimand 2005), and Fisher's *The Rate of Interest* (1907) expressed strong views on racial and ethnic differences in rates of time preference, which he considerably toned down in *The Theory of Interest* (1930).

debt-deflation theory of the Great Depression, Fisher came to regret his earlier lack of sympathy with the bimetallists. By then, he viewed them as having raised a real problem—the short-run non-neutrality of deflation—while he continued to reject the soundness of their proposed remedy, which would have required the monetary authority to fix the relative price of two commodities, gold and silver.

The Message of *Appreciation and Interest*

Fisher (1896) stressed that an appreciating value of money redistributed wealth from debtors to creditors only to the extent that the appreciation was a surprise. If the appreciation was expected, it would have been taken into account when the debts were incurred and the interest rates negotiated. A high rate of interest need not harm trade, nor need a low rate of interest encourage activity. What matters is whether the interest rate is high or low relative to the rate of appreciation of some standard. If i is the interest rate expressed in some standard I, and j is the interest rate expressed in some other standard J, and a the rate at which standard I (say, money) is expected appreciate in terms of standard J (say, commodities) over the relevant time period, then the equilibrium condition is $(1 + j) = (1 + i)\ (1 + a)$, which offers no possibility for profitable arbitrage. Falling prices need not harm farmers who owe mortgages as long as expectations of the falling prices were reflected in the interest rates on the mortgages: "It is clear that if the unit of length were changed and its change were foreknown, contracts would be modified accordingly . . . To alter the mode of measurement does not alter the actual quantities involved, but merely the numbers by which they are represented" (Fisher 1896, p. 1). "We thus see that the farmer who contracts a mortgage in gold is, *if the interest is properly adjusted*, no worse off and no better off than if his contract were in a 'wheat' standard or a 'multiple' standard" (Fisher 1896, p. 16, his italics). Appreciation or depreciation of the purchasing power of money only matters if expectations are wrong, and they won't be wrong in the long run, because people learn from experience, gather and process information, and adjust their expectations.[3]

If that was all Fisher (1896) had to say, it would have undermined the bimetallist argument for long-run non-neutrality and drawn attention to a crucial factor overlooked in monetary discussions by many leading economists. Fisher (1896, pp. 67–70) gleefully cited unsound passages written by luminaries of that time like William Stanley Jevons, Thomas Tooke, William Newmarch, and by Oxford

---

[3] A number of modern authors have cited Fisher's (1930) *The Theory of Interest* as the source for the Fisher relationship (for example, Crowder 1997, pp. 1124, 1127). But the Fisher equation does not appear in that work. Fisher presented the equation in *Appreciation and Interest* (1896) and in an appendix to *The Rate of Interest* (1907), but his discussion of the relation in chapters 2 and 19 of *The Theory of Interest* (1930) is verbal, supplemented with diagrams but no equation (Dimand 1999). Similarly, the famous Fisher two-period diagram of optimal consumption-smoothing, often attributed to *The Theory of Interest* (1930), does not appear in that book, where the discussion of intertemporal optimization is largely verbal. Instead, the diagram is in Fisher (1907, p. 409), as discussed in Humphrey (2010).

professor Bonamy Price, and could have provided many more examples. He noted (p. ix): "The views here put forward . . . differ radically from those expressed by Mr. Giffen and many other eminent economists." But, except for writing the relation as an equation, he would simply have been drawing attention to a relation already understood by such well-known figures as John Stuart Mill, Alfred Marshall, and John Bates Clark, as Fisher acknowledged.

However, Fisher did much more. Viewing Marshall's terms "real" and "nominal" interest as inadequate, Fisher applied his formula to any two standards: gold and silver; money and goods; two national currencies; or two commodities (like wheat and barley). From the principle that asset prices and returns will move to eliminate any profitable opportunity for arbitrage, he derived what is now called the uncovered interest parity condition: that is, the difference between interest in any two currencies (say, dollar interest rates in New York and pound sterling interest rates in London) is due to the expected rate of change of the exchange rate between the two currencies.[4] To show this empirically, and to show that money interest reflects the rise or fall of prices, Fisher (1896) assembled and published a wide variety of tables: on interest rates on India silver and gold bonds; Berlin, Paris, Calcutta, Tokyo, and Shanghai interest rates in relation to falling and rising prices; New York interest rates in relation to rising and falling prices and wages; London interest rates in relation to rising and falling prices, wages, and incomes; and U.S. interest rates on "coin" bonds (payable in gold coin) and "currency" bonds (payable in greenbacks) before the U.S. economy returned to the gold standard. He also examined interest rates in the same standard for loans of differing duration, explaining the term structure of interest rates by expectations of what would happen to the purchasing power of money.

Having shown that, with perfect foresight, appreciation or depreciation of the purchasing power of money would not affect real interest rates, Fisher based his monetary theory of economic fluctuations on the slow adjustment of expectations and money interest to monetary shocks in a world of imperfect foresight, which implied that monetary shocks would affect real interest in the short run. To be explicit, Irving Fisher did not believe that the Fisher relation held fully in the short run. Alfred Marshall had mentioned this insight in three sentences in his "Note on the Purchasing Power of Money in Relation to the Real Rate of Interest" in the first edition of his *Principles of Economics* in 1890 (quoted by Irving Fisher 1896, p. 79), but Fisher developed it into a full-blown theory of fluctuations (Fisher 1896, 1907, 1926; Fisher with Brown 1911, chap. 4), declaring the "so-called 'business cycle'" to be a "dance of the dollar."

In the 1920s, Fisher began to use distributed lags of past price level changes as a proxy for expectations of future price changes in his correlation analyses (Rutledge 1977). To carry out these empirical studies, Fisher (1922) proposed the Fisher ideal

[4] More than a quarter of a century later, John Maynard Keynes (1923) added the covered interest parity condition: that the spread between forward and spot exchange rates equals the difference between interest rates in two currencies.

index (the geometric mean of the Paasche and Laspeyres price indexes), and, in the absence of a government-produced price index, Fisher created and published his own weekly price index. While Fisher (1896) used statistical tables to show that money interest rates were high in periods of rising prices, Fisher (1930) correlated money interest rates with distributed lags of price changes to demonstrate both that higher expected inflation raises money interest rates and that the adjustment is slow and incomplete. In a series of articles, Fisher correlated distributed lags of price level changes with economic activity and unemployment. His article "A Statistical Relationship between Unemployment and Price Level Changes" (1926 [1973]), little noticed when first published by the International Labour Office, attracted rather more attention when reprinted almost 50 years later in the *Journal of Political Economy* as "Lost and Found: I Discovered the Phillips Curve—Irving Fisher."

How Original was Fisher (1896)?

It is a commonplace observation among researchers in many disciplines that as statistics professor Stephen Stigler (1999, p. 277) put it: "No scientific discovery is named after its original discoverer." (Naturally, Stigler attributed the insight to Robert K. Merton.) In this spirit, Humphrey (1983 [1986], p. 158) wrote: "The real/nominal rate distinction is of 18th rather than 20th century vintage. Irving Fisher, now generally regarded as the father of real/nominal interest rate analysis, originated none of the concepts now bearing his name. Neither the so-called *Fisher relationship* (according to which the nominal rate equals the real rate plus expected inflation), nor the *Fisher effect* (according to which the nominal rate fully adjusts for inflation leaving the real rate intact), nor the *Fisher neutrality proposition* (according to which equilibrium nominal rate adjustments entail no real effects) originated with him. Rather they long predate him, having been enunciated by earlier generations of writers."

But on the subject of appreciation and interest, there appear to be just three brief, isolated insights in this area that predate the work of Jacob de Haas (1889) (discussed below), and Fisher (1896) managed to uncover two of those examples. Fisher (p. 3) wrote, "It is an astonishing fact that the connection between the rate of interest and appreciation has been almost completely overlooked, both in economic theory and in its bearing upon the bimetallist controversy. Of the few writers who have conceived this connection, apparently the earliest was the anonymous author of the remarkable pamphlet entitled: 'A Discourse Concerning the Currencies of the British Plantations in America.' Boston, 1740 (Reprinted in the 'Overstone Tracts' 1857)." Following up on Fisher, Charles J. Bullock of Harvard identified the Scottish-born physician William Douglass as the author of the "remarkable pamphlet" (Douglass 1740 [1897]), which Bullock republished in the *Publications of the American Economic Association* the year after Fisher's monograph (see also Bumsted 1964). Although, as Fisher stated, Lord Overstone had included Douglass's discourse in a collection of reprints of early monetary tracts, Fisher was the first to

notice and quote the relevant paragraphs, including the argument that "*the larger the Emissions* [of colonial paper currency], *natural Interest becomes the higher;* therefore the Advocates for Paper Money (who are generally indigent Men, and Borrowers) ought not to complain, when they hire Money at a dear nominal Rate" (quoted by Fisher 1896, p. 4, Douglass's italics). A later work by Douglass (1760) was cited by Adam Smith (*Wealth of Nations*) for information about the British colonies in North America, but not the 1740 pamphlet.

A second example, discussed by Humphrey (1983 [1986]) but unknown to Fisher, is from a speech in the British House of Commons speech in 1811 about the Bullion Report, in which Henry Thornton remarked that "in countries in which the currency was in a rapid course of depreciation . . . the current rate of interest was often . . . proportionately augmented" as "partly compensation for an expected increase of depreciation of the currency" (Thornton 1811 [1939], p. 336). Humphrey (1983 [1986], p. 153) notes that this passage of Thornton's speech went beyond Douglass in explicitly stating that the premium refers to expected future inflation, not actual past inflation, but observes that "it conflicts with that part of [Thornton's] analysis that ignores anticipated inflation." Thornton's speech was overlooked until 1939 when Friedrich Hayek reprinted it in an appendix to his edition of Thornton's 1802 *Paper Credit.*

The third prior example is from John Stuart Mill's *Principles of Political Economy* (1848 [1871], Book 3, Chapter 23, Section 4, p. 656), where he remarked in a single sentence: "We thus see that depreciation, merely as such, while in the process of taking place, tends to raise the rate of interest: and the expectation of further depreciation adds to this effect; because lenders who expect that their interest will be paid, and the principal perhaps redeemed, in a less valuable currency than they lent, of course require a rate of interest sufficient to cover this contingent loss" (quoted by de Haas 1889, pp. 115–116).

As Fisher (1896, footnote on p. 5) noted, Mill's devoted only a single paragraph (and that of only one sentence) to the subject. De Haas (1889) doubted that Mill understood the full significance of the point, since he made no other mention or use of the insight. Mill's paragraph was overlooked before de Haas and Fisher. Of the three discussions that constitute these "earlier generations of writers," only that in Douglass's long forgotten tract is longer than a paragraph, and Humphrey (1983 [1986], p. 153, 158) acknowledges that Thornton's remark was inconsistent with other parts of Thornton's analysis and that Douglass and Mill did not distinguish between complete and incomplete adjustment of the nominal rate to inflation. Douglass's work became known only because of Fisher's experience with his dissertation: he had been shocked to discover Walras and Edgeworth when his own thesis on general equilibrium was almost finished, and this taught him to search carefully for forerunners before publishing ideas that he had developed independently.

Much more substantial contributions were made by three of Fisher's contemporaries, and were, together with Douglass (1740 [1897]), warmly acknowledged by Fisher: "The idea on which this theory is founded appears to have occurred independently to several writers, of whom Mr. Jacob de Haas, Jr., of Amsterdam, seems

most fully to have realized its importance . . . A principle which apparently has been independently discovered by each of these economists and quite possibly by others, is likely to be of some importance. It is the object of the present essay to develop the theory in a quantitative form, to bring it to a statistical test, and to apply it to current problems, and to the theory of interest" (1896, pp. ix, 5).[5] The Dutch economist Jacob de Haas, Jr. (1889) devoted an article in the *Journal of the Royal Statistical Society* to arguing that "the expected rate of change in the purchasing power of money" is "A Third Element in the Rate of Interest," the other two being "the remuneration for abstinence, i.e., the hire of capital" and "the insurance against loss or remuneration for risk" (de Haas 1889, pp. 107, 110–111; Fisher 1896, pp. ix, 5). Ever-skeptical Humphrey (1983 [1986], p. 154) writes: "All in all, de Haas contributed little new to the analysis of real and nominal interest rates. His work, despite its apparent originality, contains nothing that cannot be found in Thornton, although Fisher, being unaware of this, thought highly of him. Marshall too knew of his work and cited it in the first edition of the *Principles*."[6] This seems too generous a reading of a passing remark in Thornton's speech (acknowledged to be inconsistent with other writings by Thornton) and too severe a critique of de Haas, who recognized the importance of the topic sufficiently to make it the subject of his article, which caught the attention of Marshall and Fisher. John Bates Clark (1895) was the first to bring the relationship between nominal interest and expected deflation into the debates over bimetallism. Reviewing Fisher (1896) in the *Economic Journal*, Clark (1896, p. 568) held: "The reader who attaches to Dr. Fisher's statistics and theories their true significance will probably conclude that, in a time of such steady and prolonged appreciation of money, the rate of interest on loans would be so reduced as fully to neutralise the increasing costliness of the money."

Given the real but limited contributions of his predecessors in this area, Fisher's originality is highlighted rather than eclipsed. He stated the relation between interest in two standards as $(1 + j) = (1 + i)\,(1 + a)$ or, equivalently, $j = i + a + ia$. The other writers discussed the relation verbally without writing out the equation. Fisher (1896, footnote on p. 9) pointed out that, except for Marshall, they failed to compound, omitting the cross-product term and equating j to $(i + a)$, so that the numerical examples in Douglass (1740 [1897]) and Clark (1895) were wrong, or at least only approximately correct.[7] Fisher (1896, footnotes on pages 78, 79,

---

[5] But Fisher (1896, p. 56, his italics) also pointed out some misstatements in the de Haas essay: "The relation of high or low prices to the rate of interest must not be confused with the relation of *rising* or *falling* prices to the rate of interest . . . de Haas appears to have fallen into this confusion both in his criticism of Jevons and in his treatment of statistics."

[6] In later editions of Marshall's *Principles*, references to Fisher (1896, 1907) replaced mention of de Haas (Marshall 1890 [1920], footnote on p. 493).

[7] Although Fisher (1896) reproved Clark (1895) and others for "erroneous" results due to omitting the cross-product term, in *The Rate of Interest* (1907) and later works Fisher used $j = i + a$ as an acceptable continuous-time approximation, and it is in that additive form that the Fisher relation is now usually written. Reviewing Fisher (1896), Fabian Franklin (1897, p. 341) remarked: "The formula reduces approximately to $i = j - a$, which is quite accurate enough for most purposes; and Professor Fisher lays too much stress on the deviation from this simple equation."

86, and 90) also cited Marshall's evidence in the 1886 *Report on Depression of Trade* and 1888 *Report of the Gold and Silver Commission* (both reprinted in Marshall 1926), and Marshall (1926) referred favorably to Fisher (1896) in testimony to the Indian Currency Committee in 1899.

Fisher (1896, p. ix) complained: "The connection between monetary appreciation and the rate of interest has received very scant attention from economists. The writer has been led to believe that this neglect has somewhat retarded the progress of economic science and the successful interpretation of economic history—in particular the monetary history of the last twenty years." He warmly acknowledged de Haas (1889), Marshall (1890 [1920]), and Clark (1895) as contemporary exceptions to this neglect, and hailed Douglass's 1740 tract and Mill's paragraph as overlooked forerunners. But it was Fisher, not these contemporaries and forerunners, who ended the neglect. He expressed what is now called the Fisher relation as an equation (including the cross-product term), undertook a substantial statistical verification of the theory, and extended the analysis from real and nominal interest to interest rates in two currencies (uncovered interest parity), interest rates over different durations of loans (the expectations theory of the term structure of interest rates), and interest rates in pairs of commodities (own rates of interest). While upholding the long-run neutrality of money against the populist advocates of bimetallism, Fisher argued that money interest and expected inflation or deflation adjust slowly and, in the short run, incompletely to monetary shocks, so that fluctuations in real economic activity and employment are a "dance of dollar" driven by fluctuations in real interest. A year after *Appreciation and Interest*, Fisher (1897) first presented his version of the equation of exchange $MV + M'V' = PT$, extending Simon Newcomb's version of that equation to allow currency (M) and bank deposits (M') to have different velocities of circulation (V and V', respectively), where P is the price level and T the volume of transactions. This approach was to be central to Edwin Kemmerer's *Money and Credit Instruments in Their Relation to General Prices* (1907, a revision of his 1903 Cornell dissertation[8]) and to Fisher's *The Purchasing Power of Money* (Fisher with Brown 1911).

In 1898, Fisher was promoted from assistant professor to full professor of political economy at Yale, and shortly afterwards was told that he had tuberculosis and only six months to live. If he had died then, he would have been known primarily to the very small community of mathematical economists. *Appreciation and Interest* was received warmly by its handful of reviewers within the profession (Clark 1896, Powers 1897, Franklin 1897), but even they lamented "the use of complicated mathematical formulae. . . they deter the uninitiated. The readers who will labor through this part of the work can be counted on one's fingers" (Powers 1897, p. 124).[9] Fisher

---

[8] See Kemmerer (1907, pp. 11, 75, 115, 133, 153) for citations of Fisher, primarily of Fisher (1897), and Fisher with Brown (1911, pp. 14, 25, 45, 139–40, 213, 226, 276–79, 282, 331, 430–32, 487) for citations of Kemmerer.

[9] In this context "complicated mathematical formulae" means $(1 + a)(1 + i) = 1 + j$, not the system of equations for general equilibrium in Fisher (1892).

became widely known through the economics profession as a whole after *The Rate of Interest* (Fisher 1907). With *The Purchasing Power of Money* (Fisher with Brown 1911), he became not only a leader within the discipline but also a prominent public intellectual, consulted by policymakers and writing extensively for the popular press. But Fisher (1896, 1897) had already laid the foundations for the approach to monetary economics that he was to pursue after his recovery from tuberculosis: an operational, quantitatively-grounded revival of the quantity theory of money, combining long-run neutrality of money with a monetary theory of economic fluctuations driven by incomplete short-run adjustment of nominal interest to monetary shocks. His contributions in *Appreciation and Interest* concerning expected appreciation or depreciation as the wedge between interest rates in different standards remain fundamental to financial, monetary, and international economics.

References

Barber, William J. 1986. "Yale: The Fortunes of Political Economy in an Environment of Academic Conservatism." In *Breaking the Academic Mold: Economists and American Higher Learning in the Nineteenth Century*, edited by William J. Barber, 132–68. Middletown, CT: Wesleyan University Press.

Brainard, William C., and Herbert E. Scarf. 2005. "How to Compute Equilibrium Prices in 1891." In *Celebrating Irving Fisher: The Legacy of a Great Economist*, edited by Robert W. Dimand and John Geanakoplos, pp. 57–83, with comments by Donald Brown and Felix Kubler, pp. 85–87, and K. R. Sreenivasan, pp. 89–92. Malden, MA, and Oxford, UK: Blackwell.

Bumsted, John M. 1964. "Doctor Douglass's Summary: A Polemic for Reform." *New England Quarterly* 37(2): 242–50.

Clark, John Bates. 1895. "The Gold Standard of Currency in the Light of Recent Theory." *Political Science Quarterly* 10(3): 389–403.

Clark, John Bates. 1896. Review of "Appreciation and Interest" by Irving Fisher (1896). *Economic Journal* 6(December): 567–570; reprinted in *Irving Fisher: Critical Responses*, edited by Robert W. Dimand (2007), Vol. 1.

Cot, Annie L. 2005. "'Breed Out the Unfit and Breed in the Fit': Irving Fisher, Economics, and the Science of Heredity." *American Journal of Economics and Sociology* 64(3): 793–826.

Crowder, William J. 1997. "The Long-Run Fisher Relation in Canada." *Canadian Journal of Economics* 30(4): 1124–42.

de Haas, Jacob A., Jr. 1889. "A Third Element in the Rate of Interest." *Journal of the Royal Statistical Society*, Series A, 52(1): 99–116.

DeLong, J. Bradford. 2000. "The Triumph of Monetarism?" *Journal of Economic Perspectives* 14(1): 83–94.

Dimand, Robert W. 1998. "Fisher and Veblen: Two Paths for American Economics." *Journal of the History of Economic Thought* 20(4): 449–66.

Dimand, Robert W. 1999. "Irving Fisher and the Fisher Relation: Setting the Record Straight." *Canadian Journal of Economics* 32(3): 744–50.

Dimand, Robert W. 2005. "Economists and the Shadow of 'The Other' Before 1914." *American Journal of Economics and Sociology* 64(3): 827–50.

Dimand, Robert W., ed. 2007. *Irving Fisher: Critical Responses*, 3 vols. London and New York: Routlege.

Dimand, Robert W., and Hichem Ben-El-Mechaiekh. Forthcoming. "General Equilibrium Reaches North America: The Hydraulic Simulation Model in Irving Fisher's *Mathematical Investigations in Value and Prices* (1891)." *Journal of Economic and Social Measurement.*

Dorfman, Robert. 1995. "Austrian and American Capital Theories: A Conflict of Cultures." *Journal of the History of Economic Thought* 17(1): 21–34.

Douglass, William. 1740 [1897]. "A Discourse Concerning the Currencies of the British Plantations in America," as reprinted with introduction by Charles J. Bullock, *Publications of the American Economic Association*, Second Series 2(5): 1–228.

Douglass, William. 1760. *A Summary, Historical and Political, of the British Settlements in North America*, 2 vols. London.

Fisher, Irving. 1892. "Mathematical Investigations in the Theory of Value and Prices." *Transactions of the Connecticut Academy of Arts and Sciences* 9(July): 1–124, as reprinted in *The Works of Irving Fisher* (1997), Vol. 1.

Fisher, Irving. 1894. "The Mechanics of Bimetallism" *Economic Journal* 4(September): 527–537. Reprinted in *The Works of Irving Fisher* (Fisher, 1997) Vol. 1.

Fisher, Irving. 1896. "Appreciation and Interest." *Publications of the American Economic Association*, First Series, 11(4): 1–110 [331–442], and as *Appreciation and Interest*, New York: Macmillan, 1896; reprinted in *The Works of Irving Fisher* (Fisher, 1997), Vol. 1.

Fisher, Irving. 1897. "The Rôle of Capital in Economic Theory." *Economic Journal* 7(December): 511–37. Reprinted in *The Works of Irving Fisher* (Fisher, 1997), Vol. 1.

Fisher, Irving. 1907. *The Rate of Interest.* New York: Macmillan. Reprinted in *The Works of Irving Fisher,* (Fisher, 1997), Vol. 3.

Fisher, Irving, with Harry G. Brown. 1911. *The Purchasing Power of Money.* New York: Macmillan. Reprinted in *The Works of Irving Fisher,* (Fisher, 1997), Vol. 4.

Fisher, Irving. 1922. *The Making of Index Numbers.* Boston: Houghton Mifflin. Reprinted in *The Works of Irving Fisher,* (Fisher, 1997), Vol. 7.

Fisher, Irving. 1926 [1973]. "A Statistical Relationship between Unemployment and Price Level Changes." *International Labour Review* 13(6): 785–92. Reprinted in 1973 as "Lost and Found: I Discovered the Phillips Curve—Irving Fisher," *Journal of Political Economy* 82(2, Part II): 496–502, and in *The Works of Irving Fisher,* edited by William J. Barber (1997), Vol. 8.

Fisher, Irving. 1930. *The Theory of Interest.* New York: Macmillan. Reprinted in *The Works of Irving Fisher* (Fisher, 1997) Vol. 9.

Fisher, Irving. 1997. *The Works of Irving Fisher,* 14 volumes, edited by William J. Barber, assisted by Robert W. Dimand and Kevin Foster; consulting editor James Tobin. London: Pickering and Chatto.

Fisher, Willard. 1896. "'Coin' and his Critics." *Quarterly Journal of Economics* 10(January): 187–208.

Franklin, Fabian. 1897. Review of "Appreciation and Interest" by Irving Fisher (1896). *Political Science Quarterly* 12(2): 340–45. Reprinted in *Irving Fisher: Critical Responses*, edited by Robert W. Dimand (2007), Vol. 1.

Gomez Betancourt, Rebeca. 2010. "Edwin Walter Kemmerer's Contribution to the Quantity Theory of Money." *European Journal of the History of Economic Thought* 17(1): 115–40.

Harvey, William H. 1894 [1963]. *Coin's Financial School.* Reprinted with introduction by Richard Hofstadter. Cambridge, MA: Harvard University Press.

Hayek, Friedrich A. 1939. Introduction to *An Enquiry into the Nature and Effects of the Paper Credit of Great Britain* [1802], by Henry Thornton, edited by Friedrich A. Hayek. New York: Rinehart and Company.

Hoffman, Frederick L. 1896. "Race Traits and Tendencies of the American Negro." *Publications of the American Economic Association*, First Series, 11(1–3): 1–330.

Humphrey, Thomas M. 1983 [1986]. "The Early History of the Real/Nominal Interest Rate Relationship." *Federal Reserve Bank of Richmond Economic Review* 69(3), as reprinted in *Essays on Inflation*, 5th ed, by Thomas M. Humphrey, published by Federal Reserve Bank of Richmond (1986).

Humphrey, Thomas M. 2010. "Intertemporal Utility Maximization—The Fisher Diagram." In *Famous Figures and Diagrams in Economics*, edited by Mark Blaug and Peter Lloyd, 421–25. Cheltenham, UK, and Northampton, MA: Edward Elgar Publishing.

Kemmerer, Edwin W. 1907. *Money and Credit Instruments in their Relation to General Prices.* New York: Henry Holt & Company. Reprinted by Garland Publishing, New York, 1983.

Keynes, John Maynard. 1923. *A Tract on Monetary Reform.* London: Macmillan.

Marshall, Alfred. 1890 [1920]. *Principles of Economics*, 8th edition. London: Macmillan. Variorum (9th) edition, edited by C. W. Guillebaud, published by Macmillan for the Royal Economic Society, 1961.

Marshall, Alfred. 1926. *Official Papers*, edited by John Maynard Keynes. London: Macmillan.

Mill, John Stuart. 1848 [1871]. *Principles of Political Economy*, 7th edition. London: Longmans, Green. Reprinted with an introduction by Vincent

W. Bladen in *Collected Works of John Stuart Mill*, edited by John M. Robson, Vols. 2 and 3. Toronto: University of Toronto Press, 1965.

Overstone, Samuel Jones Lloyd, Lord. 1857. *Tracts and Other Publications on Metallic and Paper Currency.* London: privately printed. Reprinted in 1972 by Augustus M. Kelley, Clifton, NJ.

Powers, H. H. 1897. Review of "Appreciation and Interest" by Irving Fisher (1896). *Annals of the American Academy of Political and Social Science* 9(1): 122–26.

Rutledge, John. 1977. "Irving Fisher and Autoregressive Expectations." *American Economic Review* 67(1): 200–205.

Skaggs, Neil T. 1995. "The Methodological Roots of J. Laurence Laughlin's Anti-Quantity Theory of Money and Prices." *Journal of the History of Economic Thought* 17(1): 1–20.

Stigler, Stephen M. 1999. *Statistics on the Table: The History of Statistical Concepts and Methods.* Cambridge, MA: Harvard University Press.

Thornton, Henry. 1811 [1939]. "Two Speeches of Henry Thornton on the Bullion Report, May 1811," as reprinted in 1939 as Appendix III to Henry Thornton, *An Enquiry into the Nature and Effects of the Paper Credit of Great Britain, 1802*, edited with an introduction by Friedrich A. Hayek, New York: Rinehart and Company.

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

The Institute of Medicine has published *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America.* "If banking were like health care, automated teller machine (ATM) transactions would take not seconds but perhaps days or longer as a result of unavailable or misplaced records. If home building were like health care, carpenters, electricians, and plumbers each would work with different blueprints, with very little coordination. If shopping were like health care, product prices would not be posted, and the price charged would vary widely within the same store, depending on the source of payment. If automobile manufacturing were like health care, warranties for cars that require manufacturers to pay for defects would not exist. As a result, few factories would seek to monitor and improve production

■ *Timothy Taylor is Managing Editor, Journal of Economic Perspectives, based at Macalester College, Saint Paul, Minnesota. He blogs at http://conversableeconomist.blogspot.com.*

line performance and product quality. If airline travel were like health care, each pilot would be free to design his or her own preflight safety check, or not to perform one at all." The report discusses evidence that "[i]f all states had provided care of the quality delivered by the highest-performing state, 75,000 fewer deaths would have occurred across the country in 2005," and that one-quarter of the $2.8 trillion in annual U.S. health expenditures is excess cost. Consensus Report, released September 6, 2012. A prepublication copy of uncorrected proofs is available with free registration at http://www.iom.edu/Reports/2012/Best-Care-at-Lower-Cost -The-Path-to-Continuously-Learning-Health-Care-in-America.aspx.

For the last five years, Sallie Mae has published an annual study based on survey data called "How America Pays for College." What do families actually pay for college? On average, the answer was $20,902 in 2011–2012, which is down from $24,097 in 2009–2010. How did families hold down college costs? "The most common cost-saving measures include living at home (51% in 2012 compared with 44% in 2011 and 43% in 2010) or adding a roommate (55% in 2012), reducing spending by parents (50% in 2012), reducing spending by students (66% in 2012), students working more hours (50% in 2012) . . . In 2012, families maintained the shift toward lower-cost two-year public schools that emerged in the 2011 responses. Twenty-nine percent attend two-year public schools and 45 percent attend four-year public schools (compared with 2010's 23% and 52% respectively)." Sallie Mae, 2012. At https://www1.salliemae.com/NR/rdonlyres/75C6F178-9B25-48F5-8982-41F9B3 F35BF6/0/HowAmericaPays2012.pdf.

The newly-created Consumer Financial Protection Bureau (CFPB) has published a Report to Congress on *Reverse Mortgages*. The number of new reverse mortgages in the U.S. was below 10,000 per year in the 1990s. But it rose to more than 110,000 in both 2008 and 2009, before dropping back to a range of 70,000– 80,000 in 2010 and 2011. Some facts: "Reverse mortgage borrowers are taking out loans at younger ages than in the past. In FY2011, nearly half of borrowers were under age 70. . . . By tapping their home equity early, these borrowers may find themselves without the financial resources to finance a future move—whether due to health or other reasons." "Reverse mortgage borrowers are withdrawing more of their money upfront than in the past. In FY2011, 73 percent of borrowers took all or almost all of their available funds upfront at closing. This proportion has increased by 30 percentage points since 2008." "A surprisingly large proportion of reverse mortgage borrowers (9.4 percent as of February 2012) are at risk of foreclosure due to nonpayment of taxes and insurance. This proportion is continuing to increase." June 28, 2012. At http://files.consumerfinance.gov/a/assets/documents/201206 _cfpb_Reverse_Mortgage_Report.pdf.

Anthony P. Carnevale, Stephen J. Rose, and Andrew R. Hanson discuss the evidence concerning *Certificates: Gateway to Gainful Employment and College Degrees*. From the Executive Summary: "Over 1 million certificates were awarded in 2010; up from 300,000 in 1994. Certificates have grown from 6 percent of postsecondary awards in 1980 to 22 percent of awards today. . . . Only 2 percent of workers reported a vocational certificate as their highest educational attainment in 1984. Today, that

figure stands at 11 percent." "On average, college degree holders earn more than workers with certificates; but many certificate holders earn more than workers with Associate's degrees and some earn more than workers with Bachelor's degrees." Georgetown University Center on Labor and the Workforce, June 2012. At http://cew.georgetown.edu/certificates.

Steve Kaplan delivered the 2012 Martin Feldstein Lecture on the subject of "Executive Compensation and Corporate Governance in the U.S.: Perceptions, Facts and Challenges." He argues: "[W]hile average CEO pay increased substantially through the 1990s, it has since declined. Indeed, CEO pay levels relative to other highly paid groups today are comparable both to their average level in the early 1990s and to their average level since the 1930s. And, the ratio of large company CEO pay to firm market value has remained roughly constant since 1960. Furthermore, CEOs are typically paid for performance and penalized for poor performance. Finally, boards do monitor CEOs, and that monitoring appears to have increased over time. CEO tenures in the 2000s are lower than in the 1980s and 1990s, and CEO turnover is tied to poor stock performance. . . . There have been corporate governance failures and pay outliers where managerial power surely has been exercised. . . . At the same time, a meaningful part of CEO pay appears to have been driven by the market for talent." *NBER Reporter,* 2012, No. 3. pp. 1–6 at http://www.nber.org/reporter/2012number3/2012no3.pdf. You can watch a video of the lecture at http://www.nber.org/feldstein_lecture_2012/feldsteinlecture_2012.html. A longer version of the presentation is available as Chicago Booth Research Paper No. 12-42, which can be downloaded at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2134208.

The Federal Reserve Bank of Chicago has devoted an issue of its own *Economic Perspectives* journal to six articles about the Dodd–Frank Wall Street Reform and Consumer Protection Act that was signed into law on July 21, 2010. From the opening essay by Douglas D. Evanoff and William F. Moeller: "The stated goals of the act were to provide for financial regulatory reform, to protect consumers and investors, to put an end to too-big-to-fail, to regulate the over-the-counter (OTC) derivatives markets, to prevent another financial crisis, and for other purposes. . . . Implementation of Dodd–Frank requires the development of some 250 new regulatory rules and various mandated studies. There is also the need to introduce and staff a number of new entities (bureaus, offices, and councils) with responsibility to study, evaluate, and promote consumer protection and financial stability. Additionally, there is a mandate for regulators to identify and increase regulatory scrutiny of systemically important institutions. . . . Two years into the implementation of the act, much has been done, but much remains to be done." Third Quarter 2012. At http://www.chicagofed.org/webpages/publications/economic_perspectives/index.cfm.

The Federal Deposit Insurance Corporation has released its *2011 FDIC National Survey of Unbanked and Underbanked Households.* The "unbanked" lack any deposit account at a banking institution, while the "underbanked" have a bank account but also rely on providers of "alternative financial services" like payday loans,

pawnshops, nonbank check cashing and money orders, and the like. "8.2 percent of US households are unbanked. This represents 1 in 12 households in the nation, or nearly 10 million in total. Approximately 17 million adults live in unbanked households. . . . 20.1 percent of US households are underbanked. This represents one in five households, or 24 million households with 51 million adults." September 2012. At http://www.fdic.gov/householdsurvey/2012_unbanked report.pdf.

Anders Åslund makes the case as to "Why a Breakup of the Euro Area Must Be Avoided: Lessons from Previous Breakups." Åslund focuses on six breakups of monetary unions in Europe in the last century and draws some lessons. First, "the logistical and legal problems of reintroducing national currencies, while transitional, would be severe and protracted." Second, "capital flight and distress in the financial system would disrupt trade and investment." Third, a "plunge in business and consumer confidence would likely be accompanied by a renewed dive in asset prices inside and outside the Eurozone." Fourth, the "challenge of maintaining fiscal credibility and securing government funding would be intensified. This would call for yet more fiscal tightening measures, particularly for the weaker peripheral Eurozone countries." Fifth, non–euro area countries would suffer from sharp appreciation of their currencies, "compounding the damage to their export growth." Åslund also emphasizes that the payments system across the euro countries could be sharply disrupted. These issues lead him to conclude: "The Economic and Monetary Union must be maintained at almost any cost. . . . The exit of any single country from the EMU, at the present time when large imbalances have been accumulated, would likely lead to a bank run, which would cause the EMU payments system to break down and with it the EMU itself." Peterson Institute for International Economics Policy Brief PB 12-20, August 2012. At http://www.piie .com/publications/pb/pb12-20.pdf.

From International Organizations

The *Inclusive Wealth Report 2012*, sponsored by a number of UN agencies, is the first of what is intended to be an annual report looking at a broad measure of wealth: "Wealth is the social worth of an economy's assets: reproducible capital; human capital; knowledge; natural capital; population; institutions; and time." The theme of this first report is "Measuring progress toward sustainability." Thus, one chapter seeks to measure "changes in the inclusive wealth index (IWI) and its components for 20 countries for the period from 1990 to 2008. Wealth is primarily assessed here as the value of manufactured, human, and natural capital stocks. The Index is additionally adjusted for population changes by presenting per capita measures. 6 out of the 20 countries analyzed decreased their IWI per capita in the last 19 years. In 5 out of 20 countries, population increased at a faster rate than inclusive wealth, resulting in negative changes in the IWI per capita." United Nations University International Human Dimensions Programme on Global Environmental Change (UNU-IHDP)

and the United Nations Environment Programme, 2012. At http://www.ihdp.unu
.edu/file/download/9927.pdf.

The *82nd Annual Report of the Bank of International Settlements* includes an inter-
esting Chapter 4 titled "The Limits of Monetary Policy": "In the major advanced
economies, policy rates remain very low and central bank balance sheets continue
to expand in the wake of new rounds of balance sheet policy measures. These
extraordinarily accommodative monetary conditions are being transmitted to
emerging market economies in the form of undesirable exchange rate and capital
flow volatility. . . . Central banks' decisive actions to contain the crisis have played a
crucial role in preventing a financial meltdown and in supporting faltering econo-
mies. But there are limits to what monetary policy can do. It can provide liquidity,
but it cannot solve underlying solvency problems. Failing to appreciate the limits
of monetary policy can lead to central banks being overburdened, with potentially
serious adverse consequences. Prolonged and aggressive monetary accommoda-
tion has side effects that may delay the return to a self-sustaining recovery and may
create risks for financial and price stability globally. The growing gap between what
central banks are expected to deliver and what they can actually deliver could in the
longer term undermine their credibility and operational autonomy." June 24, 2012.
At http://www.bis.org/publ/arpdf/ar2012e.pdf.

The IMF discusses "The Good, the Bad and the Ugly: 100 Years of Dealing
with Public Debt Overhangs," in Chapter 3 of its *World Economic Outlook*. "Public
debt levels above 100 percent of GDP are not uncommon. Of the 22 advanced
economies for which there is good data coverage, more than half experienced at
least one high-debt episode between 1875 and 1997. Furthermore, several countries
had multiple episodes: three for Belgium and Italy and two for Canada, France,
Greece, the Netherlands, and New Zealand." After looking at a range of data and
some case studies, the report suggests lessons for reducing a debt overhang: "The
first key lesson is that a supportive monetary environment is a necessary condition
for successful fiscal consolidation." "Second, debt reduction is larger when fiscal
measures are permanent or structural and buttressed by a fiscal framework that
supports the measures implemented." Third, strong external demand is a huge
help. Finally, it takes years to address a massive debt overhang. The IMF authors
look at annual deficit/GDP ratios, and find that "sustained improvements of more
than 1 percentage point a year are rare." October 2012. At http://www.imf.org
/external/pubs/ft/weo/2012/02/pdf/c3.pdf.

Interviews

Douglas Clement has an in-depth "Interview with Darrell Duffie," which
ranges from the Volcker rule and the Financial Stability Oversight Council, to
repo markets and clearing banks for tri-party repos, to the euro-zone and issues of
systemic risk. "And there has been a lot of progress made, but I *do* feel that we're
looking at years of work to improve the plumbing, the infrastructure. And what

I mean by that are institutional features of how our financial markets work that can't be adjusted in the short run by discretionary behavior. They're just there or they're not. It's a pipe that exists or it's a pipe that's not there. And if those pipes are too small or too fragile and therefore break, the ability of the financial system to serve its function in the macroeconomy—to provide ultimate borrowers with cash from ultimate lenders, to transfer risk through the financial system from those least equipped to bear it to those most equipped to bear it, to get capital to corporations—those basic functions which allow and promote economic growth could be harmed if that plumbing is broken. If not well designed, the plumbing can get broken in any kind of financial crisis if the shocks are big enough. It doesn't matter if it's a subprime mortgage crisis or a eurozone sovereign debt crisis. If you get a big pulse of risk that has to go through the financial system and it can't make it through one of these pipes or valves without breaking it, then the financial system will no longer function as it's supposed to and we'll have recession or possibly worse." *The Region,* Federal Reserve Bank of Minneapolis, June 2012, 12–27. At http://www.minneapolisfed.org/publications_papers/pub_display .cfm?id=4900.

Aaron Steelman conducted an "Interview" with John List. Here's List making the case for conducting randomized economic experiments in natural or "field" settings. "So I come along, and I say we really need to use the tool of randomization, but we need to use it in the field. Here's where the skepticism arose using that approach: People would say, 'You can't do that, because the world is really, really messy, and there are a lot of things that you don't observe or control. . . .' That reasoning stems from the natural sciences. Consider the example with the chemist: If she has dirty test tubes her data are flawed. The rub is that chemists do not use randomization to measure treatment effects. When you do, you can balance the unobservables—the 'dirt'—and make clean inference. As such, I think that economists' reasoning on field experiments has been flawed for decades, and I believe it is an important reason why people have not used field experiments until the last 10 or 15 years. . . . When I look at the real world, I want it to be messy. I want there to be many, many variables that we don't observe and I want those variables to frustrate inference. The reason why the field experiments are so valuable is because you randomize people into treatment and control, and those unobservable variables are then balanced. I'm not getting rid of the unobservables—you can never get rid of unobservables—but I can balance them across treatment and control cells. Experimentation should be used in environments that are messy; and I think the profession has had it exactly backwards for decades. They have always thought if the test tube is not clean, then you can't experiment. That's exactly wrong. When the test tube is dirty, it means that it's harder to make proper causal inference by using our typical empirical approaches that model mounds and mounds of data." *Region Focus,* Federal Reserve Bank of Richmond, Second/Third Quarter 2012, pp. 32–38). (Full disclosure: John List is one of the coeditors of this journal.) At http://www.richmondfed.org/publications/research/region_focus/2012/q2-3 /pdf/interview.pdf.

Discussion Starters

Steve H. Hanke and Nicholas Krus have compiled an historical list of "World Hyperinflations." The infamous German hyperinflation of 1922–23 is near the top of the list, but ranks only fifth for highest monthly rate of inflation. Hungary holds the record for highest monthly hyperinflation rate in July 1946, when the price level doubled every 15 hours. The Zimbabwe hyperinflation of November 2008 is a close second, when prices doubled every 25 hours. Many of the hyperinflations on the list occur either in the aftermath of World War II, or in the aftermath of the break-up of the Soviet Union in the early 1990s. Cato Working Paper, August 15, 2012. At http://www.cato.org/pubs/researchnotes/WorkingPaper-8.pdf.

Ramanan Laxminarayan discusses "A Matter of Life and Death: The Economics of Antibiotic Resistance." "In the United States, for example, resistance to the bacterium methicillin-resistant Staphylococcus aureus (MRSA), has reached 60 percent. This means six out of 10 patients with this virulent staph infection can no longer be treated with oxacillin, a relatively low cost drug. But what still amounts to a cost problem in rich countries is becoming a serious threat to public health in the developing world: lower-income countries face a growing toll of death and morbidity from curable infections because the generally available antibiotics no longer work." "The logic of regulating antibiotics differently from other drugs arises from the fact that one person's use contributes to lower effectiveness for everyone else. The spread of resistance by overuse of antibiotics is like other shared resource problems, such as global warming or overfishing—a phenomenon dubbed 'the tragedy of the commons.' Approaching antibiotic resistance as a resource problem is not just a convenient metaphor; it can help shape strategies to use antibiotics in ways that provide the greatest benefit to society, both today and in the future. . . . However, no such regulations have been forthcoming in the case of antibiotics for at least two reasons. First, medical practitioners resist any form of regulation that would limit their discretion to prescribe antibiotics. Second, no single federal agency has the authority to intervene." *Milken Institute Review*, Third Quarter 2012, pp. 13–21. Available with free registration at http://www.milkeninstitute.org/publications /review/2012_7/12-21MR55.pdf.

Martin West summarizes "Global Lessons for Improving U.S. Education." "Among the 34 developed democracies that are members of the Organization for Economic Cooperation and Development (OECD), 15-year-olds in the United States ranked 14th in reading, 17th in science, and no better than 25th in mathematics." "[T]here are three broad areas in which the consistency of findings across studies using different international tests and country samples bears attention. *Exit exams.* Perhaps the best-documented factor is that students perform at higher levels in countries (and in regions within countries) with externally administered, curriculum-based exams at the completion of secondary schooling that carry significant consequences for students of all ability levels. . . . *Private-school competition.* . . . Rigorous studies confirm that students in countries that for historical reasons have a larger share of students in private schools perform at higher levels on

international assessments while spending less on primary and secondary education. . . . In addition, the achievement gap between socioeconomically disadvantaged and advantaged students is reduced in countries in which private schools receive more government funds. . . . *High-ability teachers.* Much attention has recently been devoted to the fact that several of the highest-performing countries internationally draw their teachers disproportionately from the top third of all students completing college degrees. This contrasts sharply with recruitment patterns in the United States." Spring 2012 edition of *Issues in Science and Technology.* At http://www.issues .org/28.3/index.html.

Pete Klenow reports some calculations about greater equality of opportunity and economic output in "The Allocation of Talent and U.S. Economic Growth." "In 1960, 94 percent of doctors were white men, as were 96 percent of lawyers and 86 percent of managers. By 2008, these numbers had fallen to 63, 61, and 57 percent respectively. Skilled occupations have become more equally distributed across race and gender, as have earnings within occupations. The result is arguably better allocation of talent and human capital investment. . . . How much of overall growth in income per worker between 1960 and 2008 in the U.S. can be explained by women and African Americans investing more in human capital and working more in high-skill occupations? Our answer is 15% to 20% . . . White men arguably lost around 5% of their earnings, as a result, because they moved into lower skilled occupations than they otherwise would have. But their losses were swamped by the income gains reaped by women and blacks." Stanford Institute for Economic Policy Research Policy Brief. July 2012. At http://siepr.stanford.edu/?q=/system/files /shared/pubs/papers/briefs/July_2012_brief.pdf.

Notes

For additional announcements, check out the continuously updated JEP online Bulletin Board, http://www .aeaweb.org/bulletinboard.php. Calls for papers, notices of professional meetings, and other announcements of interest to economists should be submitted to Ann Norman at jep@jepjournal.org in one or two paragraphs containing the relevant information. These will be posted at the JEP online Bulletin Board. Given sufficient lead time, we will also print a shorter, one-paragraph version of your notice in the "Notes" section of the Journal of Economic Perspectives. *It is best to* **send announcements for JEP "Notes" before** *March 20 for the JEP Spring issue, which mails the end of May; before June 20 for the JEP Summer issue, which mails the end of August; before September 20 for the JEP Fall issue, which mails the end of November; and before December 10 for the JEP Winter issue, which mails the end of February. We reserve the right to edit material received.*

Call for Sessions and Papers for the 2014 American Economic Association Annual Meeting to be held January 3–5, 2014, in Philadelphia, PA. Members wishing to give papers or organize complete sessions for the program for the meetings in Philadelphia are invited to submit proposals electronically to Professor William Nordhaus via the AEA website. (The submission portal for the 2014 annual AEA meeting will open on March 1, 2013.) While papers covering a wide array of topics in economics will be included on the 2014 program, Professor Nordhaus especially encourages interdisciplinary proposals.

To be considered, **individual paper proposals** (with abstracts) and up to *two Journal of Economic Literature bibliographic codes* in rank order **should be submitted by April 1, 2013; complete session proposals by April 15, 2013.** *At least one author of each paper must be an AEA member.* All authors of papers on a complete session must join the AEA if the session is selected for the program.

Proposals for complete sessions have historically had a higher probability of inclusion (35–40%) than papers submitted individually (10–15%). Individual paper contributors are strongly encouraged to use the AEA's Econ-Harmony website (http://www .aeaweb.org/econ-harmony/) to form integrated sessions. Proposals for a complete session should be submitted only by the session organizer. Sessions normally contain three or four papers.

Please make certain your information is complete before submission. No changes will be accepted until a decision is made about inclusion on the program (usually in July). Econometric studies or highly mathematical papers are not appropriate for sessions sponsored by the AEA; such papers should be submitted to the Econometric Society. Do not send a complete paper. The Association discourages multiple proposals from the same person, and under no circumstances should the same person submit more than two proposals.

Some of the papers presented at the annual meeting are published in the May *American Economic Review* (the *Papers & Proceedings*). The President-elect includes at least three contributed sessions (12 papers) from among those submitted in response to this Call for Sessions and Papers.

The Annual Meeting of the American Economic Association will be held in San Diego, California, **January 4–6, 2013**. The headquarters will be the Manchester Grand Hyatt. Registration and exhibits will also be in the Manchester Grand Hyatt. Information and procedures for employers and job seekers are in the registration material at the AEA website. There is no on-site interview arrangement service; all correspondence, including interview scheduling, should take place over the Internet prior to arrival. The interview tables will be the San Diego Marriott Marquis & Marina. For additional information or to register, please go to the AEA website: www.vanderbilt.edu/AEA.
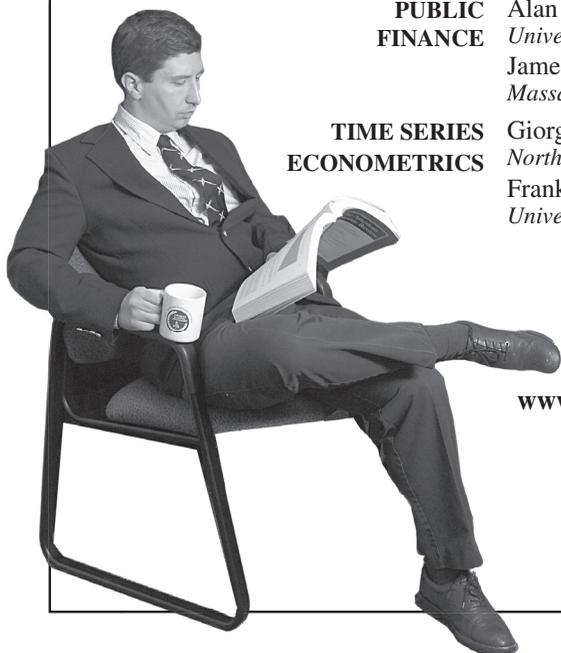
AEA Continuing Education Program is held immediately after the Annual Meeting in January. The program aims to help mid-career economists and others maintain the value of their human capital. It is tailored primarily to faculty at liberal arts colleges and teaching-oriented state universities that may have fewer research opportunities than colleagues at

universities with Ph.D. programs. The lecturers are leading scholars who also are excellent expositors. The focus is on content to help improve teaching and research. The three topics for January 2013 in San Diego are: Labor Economics (Jesse Rothstein, University of California-Berkeley, and Lowell Taylor, Carnegie Mellon University), Public Finance (Alan Auerbach, University of California–Berkeley, and James Poterba, MIT), Time-series Econometrics (Giorgio Primiceri, Northwestern, and Frank Schorfheide, University of Pennsylvania). For more information go to http://www.aeaweb.org/cont_education/index.php.

Election Results. The American Economic Association is pleased to announce the election results for the 2012 Election for Officers of the American Economic Association for 2013: William D. Nordhaus (President-elect), Raquel Fernandez and Paul R. Milgrom (Vice-presidents), Amy M. Finkelstein and Jonathan D. Levin (members).

The Third Annual AEA Conference on Teaching (at the undergraduate and graduate levels) **and Research in Economic Education** (all levels, including precollege) will be held May 29 to May 31, 2013, in Chicago, hosted by the Committee on Economic Education, in cooperation with the *Journal of Economic Education* and the Federal Reserve Bank of Chicago. The conference is at the Renaissance Chicago Downtown, with a dinner at the Federal Reserve Bank of Chicago. Plenary talks will be given by John List, Steven Levitt, and Derek Neal. Submissions will be accepted via the online submission system of the AEA: http://www.aeaweb.org/committees/AEACEE/Conference/submissions/. Submissions may be of individual papers, complete sessions of three or four papers, workshops, or panels. We especially encourage submissions of completed sessions. The **submission deadline is December 15, 2012**. To see past CEE conference presentations, program and schedule of events go to: http://www.aeaweb.org/committees/AEACEE/Conference/past_conf.php. Questions about submissions should be sent to Tisha Emerson at Tisha_Nakao@baylor.edu.

Call for Proposals for Poster Session for the ASSA Meetings, held January 3–5, 2014, in Philadelphia, PA. The Committee on Economic Education will sponsor a poster session at the 2014 ASSA Meetings in Philadelphia devoted to active learning strategies across the economics curriculum. Instead of papers, session presenters will prepare large visual poster summaries of their work, which will be mounted in an exhibition room to allow presenters to talk directly with session participants. Although we encourage presenters to include evidence that their strategy enhances learning, we do not require quantifiable evidence. Presenters should emphasize the originality of their strategy and provide sufficient information so that session participants may apply the technique in their own classrooms. Proposals should describe the teaching strategy and explain how it will be described in the poster. Posters marketing textbooks, commercial software, or similar materials will not be considered for the session. Proposals are limited to two pages and are **due by April 1, 2013**. Proposals should include full contact information for all authors. Please send to: Steven Cobb at scobb@unt.edu.

The AEA Committee for the Status of Women in the Economics Profession (CSWEP) will organize three sessions on gender-related topics and three sessions on structural econometrics at the January 2014 Meeting of the American Economic Association in Philadelphia, PA. Accepted papers will be considered for publication in the 2014 Papers and Proceedings of the *American Economic Review*. Email a cover letter (specifying to which set of sessions the abstract is being submitted) and a copy of a one or two page abstract (250–1000 words) to cswep@econ.duke **by March 1, 2013**. Be sure to clearly label all authors' names, affiliations, and contact information and to designate a corresponding author.

ASA/NSF/BLS Fellowship Opportunity Are you interested in conducting research at the Bureau of Labor Statistics? If so, apply to our Senior Research Fellow Program. **Proposals are due January 20, 2013.** The program facilitates collaboration between academic scholars and government researchers in fields such as statistics, economics, survey methodology, and social science. Fellows are funded to conduct research at the BLS headquarters in Washington, DC, use BLS data and facilities, and work closely with BLS staff. For more information, see our website at http://www.bls.gov/osmr/asa_nsf_bls_fellowship_info.htm or our brochure at http://www.amstat.org/careers/pdfs/ASANSFBLSFellowshipProgram.pdf.

Call for Papers: A Special Issue of the *International Review of Economics & Finance* on The Dynamics of International Migration. Manuscripts, should be submitted electronically on or **before December 15, 2012**, to Wilhelm Kohler, Professor of International Economics, University of Tuebingen. Email: wilhelm.kohler@uni-tuebingen.de.

Call for Papers. IJEB and **Serials Publications, New Delhi**, announce the **International Conference on Economic and Business Issues**, to be held December 19 and 20, 2013, in Hotel Grand Ramee, Apte Road, Pune 411004, India. The paper/abstract submissions **deadline is May 31, 2013**; registration deadline is October 31, 2013. For more information contact Dr. Kishore G. Kulkarni, Editor of the *Indian Journal of Economics and Business* (see www.ijeb.com), at Metropolitan State University of Denver. Phone: 303-556-2675; fax: 303-556-3966; e-mail: kulkarnk@msudenver.edu.

The American Economic Association

MIX
Paper from responsible sources
FSC
www.fsc.org
FSC® C101537

The Journal of

Economic Perspectives

Fall 2012, Volume 26, Number 4

Symposia

Contingent Valuation

Catherine L. Kling, Daniel J. Phaneuf, and Jinhua Zhao, "From Exxon to BP: Has Some Number Become Better than No Number?"

Richard T. Carson, "Contingent Valuation: A Practical Alternative when Prices Aren't Available"

Jerry Hausman, "Contingent Valuation: From Dubious to Hopeless"

China's Economy

Hongbin Li, Lei Li, Binzhen Wu, and Yanyan Xiong, "The End of Cheap Chinese Labor"

Xin Meng, "Labor Market Outcomes and Reforms in China"

Xiaodong Zhu, "Understanding China's Growth: Past, Present, and Future"

Dennis Tao Yang, "Aggregate Savings and External Imbalances in China"

Yasheng Huang, "How Did China Take Off?"

Articles

Jonathan Levin and James Poterba,
"Amy Finkelstein: 2012 John Bates Clark Medalist"

Features

Robert W. Dimand and Rebeca Gomez Betancourt, "Retrospectives: Irving Fisher's *Appreciation and Interest* (1896) and the Fisher Relation"

Recommendations for Further Reading • Notes