Estimating the Tradeoff Between Risk Protection and Moral Hazard with a Nonlinear Budget Set Model of Health Insurance*

Amanda E. Kowalski†

September 19, 2012

Abstract

Insurance induces a tradeoff between the welfare gains from risk protection and the welfare losses from moral hazard. Empirical work traditionally estimates each side of the tradeoff separately, potentially yielding mutually inconsistent results. I develop a nonlinear budget set model of health insurance that allows for both simultaneously. Nonlinearities in the budget set arise from deductibles, coinsurance rates, and stoplosses that alter moral hazard as well as risk protection. I illustrate the properties of my model by estimating it using data on employer sponsored health insurance from a large firm. Within my empirical context, the average deadweight losses from moral hazard substantially outweigh the average welfare gains from risk protection. However, the welfare impact of moral hazard and risk protection are both small relative to transfers from the government through the tax preference for employer sponsored health insurance and transfers from some agents to other agents through a common premium.

Keywords: risk protection, moral hazard, nonlinear budget set, health insurance.

*I thank the following individuals for helpful comments: Joseph Altonji, C. Lanier Benkard, Steve Berry, John Beshears, Tom Chang, Victor Chernozhukov, Raj Chetty, Jesse Edgerton, Randy Ellis, Hanming Fang, Amy Finkelstein, Jason Fletcher, Michael Grossman, Jonathan Gruber, Ben Handel, Justine Hastings, Phil Haile, Jerry Hausman, Naomi Hausman, Kate Ho, Panle Jia, Jonathan Kolstad, Kory Kroft, Fabian Lange, Claudio Lucarelli, Will Manning, Costas Meghir, Whitney Newey, Chris Nosko, Matt Notowidigdo, Ariel Pakes, Mark Pauly, Stephen Ryan, Bernard Salanie, Paul Schrimpf, Larry Seidman, Hui Shan, Ebonya Washington, and Heidi Williams. I thank participants at the public finance seminar and the econometrics lunch at MIT and the junior faculty and summer faculty lunches at Yale for feedback on very early versions of this paper. I am also grateful for feedback from Columbia, the Cowles Summer Structural Micro Conference, the Econometric Society Summer and Winter Meetings, Harvard, Hunter College, the NBER Summer Institute, NYU IO Day, the University of North Carolina, the University of Virginia, the University of Illinois at Chicago, and the AEA annual meeting. Jean Roth and Mohan Ramanujan provided invaluable support at the NBER, and Brian Dobbins and Andrew Sherman provided invaluable support with the Yale University Faculty of Arts and Sciences High Performance Computing Center. Toby Chaiken provided excellent research assistance. The National Institute on Aging, Grant Number T32-AG00186, provided support for this project.

†Department of Economics, Yale University and NBER: amanda.kowalski@yale.edu

1 Introduction

Standard theoretical models of insurance emphasize the tradeoff between welfare losses from moral hazard and offsetting welfare gains from risk protection (Arrow (1963), Pauly (1968), Zeckhauser (1970), and Ehrlich and Becker (1972)). The sign and magnitude of this tradeoff is an empirical question, but empirical evidence traditionally focuses on only one side or estimates moral hazard and risk protection using separate techniques. For example, with regard to health insurance, a long literature focuses exclusively on estimating the magnitude of moral hazard,[1] including Manning et al. (1987), Newhouse (1993), Eichner (1997), Eichner (1998), Kowalski (2009), and Duarte (2010). A more limited set of studies, including Feldstein (1973), Feldman and Dowd (1991), Feldstein and Gruber (1995), Manning and Marquis (1996), Finkelstein and McKnight (2008), and Engelhardt and Gruber (2010), examine risk protection as well as moral hazard associated with health insurance. These studies generally impose a functional form for utility to estimate risk protection and a separate functional form for demand to estimate moral hazard and then compare the estimates to get a sense of the tradeoff. They acknowledge that because both functional forms are likely mutually inconsistent, the estimated tradeoff can be subject to bias of unknown sign and magnitude.[2] In this paper, I improve upon the literature by developing and estimating a model of health insurance that includes welfare losses from moral hazard and welfare gains from risk protection, allowing for estimation of the tradeoff between them.

An important feature of my approach is that it models nonlinear cost sharing structures in the form of deductibles, coinsurance rates, and stoplosses. These cost sharing structures are designed to increase social welfare by exposing patients to higher marginal prices than they would face under full insurance, mitigating moral hazard. However, they can also decrease social welfare by exposing patients to higher expenditure risk, mitigating risk protection. Depending on the entire nonlinear budget set induced by the specific magnitude of each cost sharing parameter, health insurance could theoretically result in a net welfare gain or loss. The magnitude of the net gain or loss in any given plan is an empirical question.

My model allows me to estimate the welfare impact of existing and counterfactual nonlinear plans. For example, I can address a particular policy question set forth by Feldstein (2006) - which is better for welfare: a plan where a consumer pays 100% of medical expenditures up to a high deductible, or a plan where a consumer pays 50% of expenditures up to twice the same high deductible? While the first type of plan is very prevalent, I am not aware that the second type of "Feldstein plan" exists, making policy analysis difficult without a model. Using my model, I compare welfare in a "Feldstein plan" counterfactual plan to welfare in an observed high deductible plan. I also compare welfare across several other existing and counterfactual plans. By allowing for counterfactual simulations and welfare analysis, my paper contributes to the growing literature on the welfare cost of asymmetric information in health insurance markets, of which Einav et al.

[1] I provide a definition of moral hazard in the context of my model in Section 2.1.3. I also describe how my estimates of moral hazard relate to estimated "price elasticities of expenditure on medical care" from the literature.

[2] Feldstein and Gruber (1995) state, "We simplify the welfare calculations by assuming that the two welfare effects can be evaluated separately and added together" (page 116).

(2010b) provide a recent review.

The welfare impact of nonlinear cost sharing structures for health insurance is an important policy consideration because nonlinear cost sharing structures are ubiquitous in public and private health insurance plans, and several government policies affect the purchase of health insurance. Some government policies explicitly provide or induce the purchase of health insurance policies with specific nonlinear cost-sharing structures. For example, public prescription drug insurance for seniors established by the Medicare Modernization Act (MMA) of 2003 follows a nonlinear cost-sharing schedule with a well-known "doughnut hole," in which seniors with intermediate drug expenditures face the full cost of those drugs until they reach a higher amount. The recent national health reform Affordable Care Act (ACA) of 2010 attempts to close the "doughnut hole" by requiring drug manufacturers to give discounts to seniors in the relevant expenditure range, but nonlinearities persist. Furthermore, the MMA encouraged the purchase of high-deductible private health insurance plans by establishing health savings accounts that could only be held by high deductible policyholders.

Welfare in private plans is also important for policy because some policies explicitly encourage the purchase of private health insurance plans. For example, the ACA requires most individuals to have health insurance from a private or public source. Given that much of this health insurance will likely be purchased through employers or on the private market through exchanges,[3] welfare associated with nonlinear cost sharing in private plans is also important for policy. Furthermore, perhaps the most expensive policy that affects the purchase of health insurance is the tax advantage for employer sponsored health insurance. Employer sponsored health insurance is the leading source of health insurance for the nonelderly.

I illustrate my model by estimating it using data on employer sponsored health insurance from a large firm. Within my empirical context, results suggest that on average, the deadweight losses from moral hazard far outweigh the welfare gains from risk protection in the existing plans. However, I find that the welfare impact of moral hazard and risk protection are both small relative to the welfare impact of transfers from other agents through the premium and from the government through the tax preference for employer sponsored health insurance. An important contribution of my approach over the existing literature is that I can estimate the tradeoff separately for each agent in my data. The ability to calculate welfare separately for each agent allows me to move beyond average welfare to make statements about the distribution of welfare and welfare for agents with specific observable characteristics. The results suggest that there is considerable variation in the net welfare gain from insurance across agents. Ranked by valuation, the top 1% of agents have a net *gain* from insurance that is 100 times smaller than the loss for agents at the mean, and the bottom 1% of agents have a net loss from insurance that is ten times larger than the loss for the individuals at the mean. I also find considerable variation in the net welfare gain by observable characteristics. Some individuals receive large transfers from others because predicted spending varies widely based on observable characteristics, but the premium does not vary with predicted

[3] In work that examines the Massachusetts health reform of 2006, largely considered to be a model for national health reform, I find that private and public sources of insurance coverage increased in roughly equal proportions (Kolstad and Kowalski (2010)).

spending.

Beyond the existing plans in my data, counterfactual simulations from my model allow me to consider the the welfare impact of alternative nonlinear structure of health insurance plans, given the tradeoff between moral hazard and risk protection. By performing a counterfactual simulation, I find that a hypothetical Feldstein plan improves welfare relative to a similar high deductible plan. Although the counterfactual simulations that I conduct illustrate the value of my approach, I discuss the difficulty of extrapolating my findings beyond my empirical context.

In next the section, I present the model and develop a simulated minimum distance estimator that is tied very closely to the model. In Section 3, I discuss my empirical context and data. In Section 4, I present the estimates and perform counterfactual simulations using the estimates. I conclude in Section 5.

2 The Model

2.1 Overview of the Model

Methodologically, my model builds on the literature developed to estimate labor supply elasticities using nonlinearities in the budget set induced by taxes, summarized by Hausman (1985). I extend that literature in several ways, but most notably by incorporating risk protection. Two papers, Keeler et al. (1977), and Eichner (1998) have applied similar models to the medical care context, but their models only allow them to consider moral hazard. Ellis (1986) develops a nonlinear budget set model of medical care that allows for moral hazard and risk protection, but he does not incorporate risk protection into his empirical specification. Manning and Marquis (1996) consider moral hazard and risk protection in simple plans, but their model cannot capture the full nonlinear budget set implied by plans with more than two segments. As I discuss below, plans with more than two segments are empirically ubiquitous, and they introduce substantial complexity into the modeling and estimation of the tradeoff. Other papers by Marsh (2009) and Bajari et al. (2010) exploit nonlinear cost sharing structures in medical care for identification in the spirit of a regression discontinuity design, but they also do not take the entire structure of the budget set into account, and their models do not allow them to measure risk protection.

My model allows me to examine the tradeoff between moral hazard and risk protection for individuals enrolled in health insurance plans with varying nonlinear cost sharing schedules. To focus on this tradeoff using minimal structure, the model abstracts away from several aspects of the agent's decision problem. The model does not include dynamics within or across years. Furthermore, the model does not distinguish between consumer decisions and doctor decisions. It largely abstracts away from supply side (insurer) considerations, such as those examined in Lustig (2010) and Starc (2010).

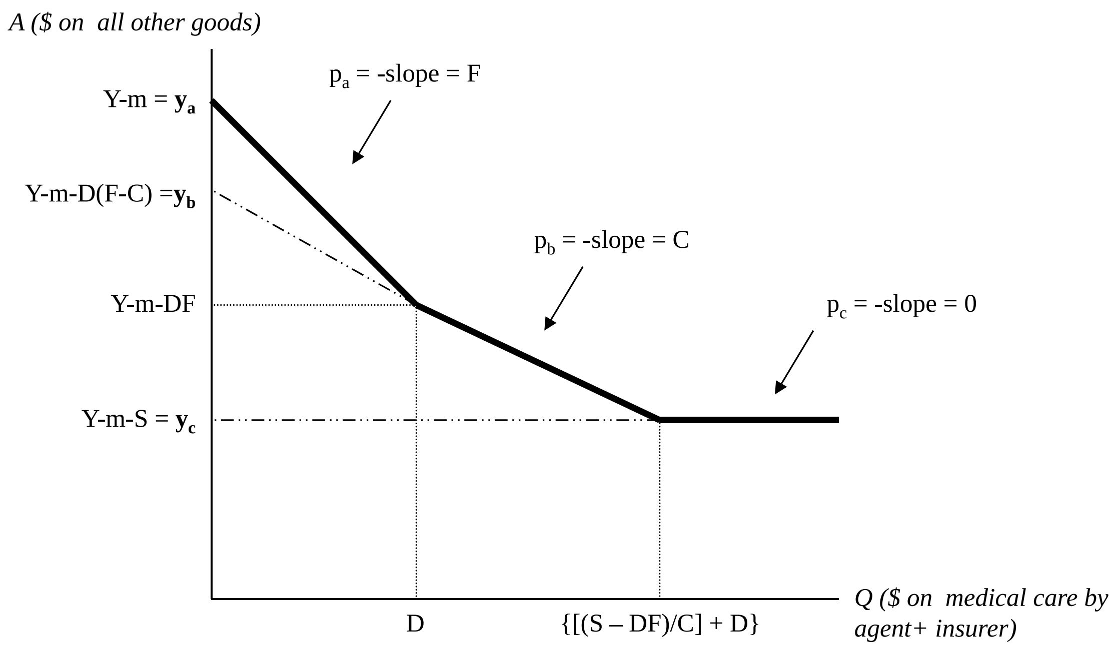
Despite these simplifications, the model takes several aspects of the agent's decision problem very seriously, aiming to capture the aspects most likely to affect the tradeoff between moral hazard and risk protection. The model gives a new, unified, framework for measuring the tradeoff between moral hazard and risk protection. Empirical estimates, which are tied closely to the model, illustrate the tradeoff in a specific context.

In Section 2.1.1, I introduce the nonlinear budget set induced by a general health insurance plan. Next, in Section 2.1.2, I develop a general model of utility maximization subject to a nonlinear constraint. I then provide a framework for calculating the tradeoff between moral hazard and risk protection within the general model in Section 2.1.3. I specify a functional form so I can use the model within my empirical context in Section 2.1.4.

2.1.1 Nonlinear Budget Set from Health Insurance

A traditional health insurance plan has three basic components that induce nonlinearities in the consumer budget set: a deductible, a coinsurance rate, and a stoploss.[4] The "deductible" is defined as the yearly amount that the beneficiary must pay before the plan covers any expenses. The percentage of expenses that the beneficiary pays after the deductible is met is known as the "coinsurance rate." The insurer pays the remaining fraction of expenses until the beneficiary meets the "stoploss," (also known as the "maximum out-of-pocket"), and the insurer pays all expenses for the rest of the year. Figure 1 illustrates how these three parameters generate nonlinearities

Figure 1: Nonlinear Budget Set Model of Health Insurance



in the consumer budget set. This partial equilibrium diagram relates medical care expenditure in dollars by the beneficiary and insurer, Q, to expenditure on all other goods, A. Medical care, Q, is measured in terms of dollars of expenditure on all types of medical care rather than in terms of

[4]In practice, there are many other possible health insurance plan provisions. For example, some plans restrict care to a certain provider network, require a per-visit "copayment," or impose lifetime limits on plan payments. However, for many policies, including those that I study, the three basic components provide a relatively complete description of plan attributes.

specific services because in most health insurance policies, the marginal price that the consumer pays for a dollar of medical care does not vary with the type of care consumed.[5]

In this diagram, D denotes the deductible, C denotes the coinsurance rate, and S denotes the stoploss. The budget set has three linear segments, denoted by a, b, and c. The consumer's marginal price associated with each segment s is p_s. Specifically the three marginal prices are: $p_a = F$, $p_b = C$, and $p_c = 0$. In all of the plans that I observe in my data, the first marginal price is one ($p_a = 1$) but I model it more generally as the fraction F, which allows me to examine counterfactual plans such as the Feldstein plan with $F = 0.5$ before the deductible.

A central issue in nonlinear budget set models is that it is difficult to control for income because nonlinearities in the budget set create a disparity between marginal income and actual income. One approach to deal with this difficulty is to control for what Burtless and Hausman (1978) call "virtual income." Virtual income is the income that the consumer would have if each segment of the budget set were extended to the vertical axis. It represents the "marginal income" that is traded off against a marginal unit of expenditure. In the figure, actual income is denoted by Y, and virtual income on each segment is denoted by y_s. In terms of income and plan characteristics, virtual income on each segment can be expressed as follows:

$$\begin{aligned} y_a &= Y - m \\ y_b &= Y - m - D(F - C) \\ y_c &= Y - m - S. \end{aligned}$$

As shown in the figure, the premium that the individual pays to be part of the plan, m, shifts income and virtual income vertically.

In Appendix A, I compare the nonlinear budget set to other nonlinear budget sets from the literature, of which the leading example is the nonlinear budget set induced by progressive taxation. While the budget set induced by progressive taxation is convex, the budget set induced by health insurance is inherently nonconvex. Nonconvexities make utility maximization more complicated because it is possible to have multiple tangencies between an indifference curve and a nonconvex budget set. While convex budget sets imply "bunching" at the kinks, nonconvex imply dispersion at the kinks. Although progressive taxes generally lead to convex budget sets, more complex budget sets, especially those that result from public assistance programs, can be nonconvex. Several papers, including Burtless and Hausman (1978), Hausman (1980), and Hausman (1981) estimate models that incorporate nonconvex segments. However, I am not aware of any other papers that incorporate two or more nonconvex segments, as I do in my model, which makes estimation much

[5]In traditional demand theory, expenditure is equal to the quantity of units demanded multiplied by the per-unit price. In my model, I make some slight modifications to the standard notation from demand theory to incorporate expenditure on behalf of the consumer by another party, the insurer. To do so, I measure the quantity of units demanded, Q, in dollars of medical care, and I measure the per-unit price, p, in terms of the marginal price that the *consumer* pays for a dollar of medical care. The marginal price that the *insurer* pays for a dollar of medical care is given by $(1 - p)$. Since the marginal price paid by the consumer and the insurer always sums to unity, the number of units of medical care demanded by the consumer, Q, is equal to total expenditure on behalf of the consumer, $Q \times 1 = Q$. Thus, unlike in standard demand models, Q measures demand as well as total expenditure. To fit this model into traditional demand theory, I model Q as a function of p, as I discuss below.

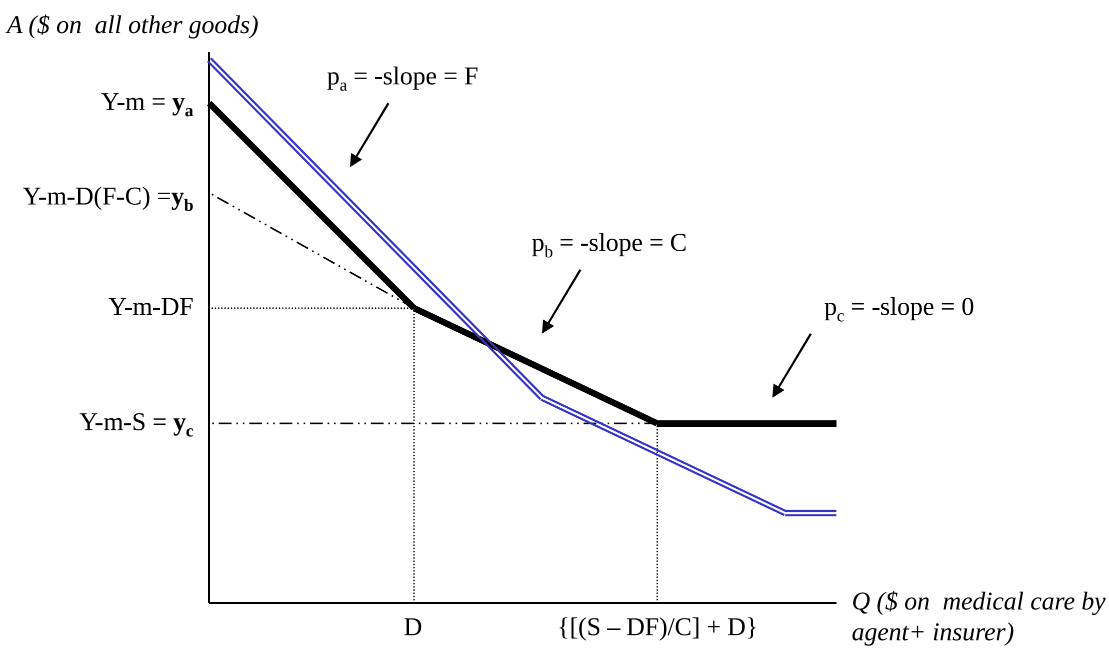
more difficult.

2.1.2 The Agent's Problem

Agents make decisions in two periods, in the spirit of Cardon and Hendel (2001). Agents make both choices by maximizing expected or actual utility over the dollars of medical care, Q, and dollars of all other goods, A. For simplicity, I define utility over Q and A, but the model could be extended in the spirit of Grossman (1972) and Phelps and Newhouse (1974) so that agents derive utility from health instead of medical care.

In the first period, agents choose a health insurance plan from the menu of available nonlinear cost sharing options. In this period, insurance offers agents protection from risk as well as access to lower marginal prices than they would face under no insurance. When choosing a plan, the agents know their observable characteristics Z_i and the distribution of medical expenditure shocks that they will face $f(r_i)$. Agents also know how they will respond to marginal prices in each plan, which allows moral hazard to affect the risk protection of a particular plan.[6] Agents maximize expected utility over alternative budget sets such as those depicted in Figure 2. As shown, the solid budget set has a higher premium, but it offers lower prices, so it will provide higher expected utility to individuals who expect to have high expenditures. The other budget set will provide higher expected utility to individuals who expect to have low expenditures.

Figure 2: Two Alternative Nonlinear Budget Sets



In the second period, agents choose how much medical care to consume given the nonlinear cost

[6]In this model, because agents can choose plans knowing their magnitude of moral hazard, there can be "selection on moral hazard" as estimated in Einav et al. (forthcoming) and discussed in Karlan and Zinman (2009).

sharing schedule. In this period, there is no more uncertainty, so insurance only offers value insofar as it provides lower prices. Given their chosen plan, their individual characteristics, and the private information of their realized medical shock, agents choose how much medical care to consume. I solve this problem backwards, starting with the second period.

Suppose that agent i is enrolled in a health insurance plan with general nonlinear cost sharing schedule j. In the second period, there is no uncertainty, as agent i has already realized a shock r from the distribution $f(r_i)$. The agent maximizes utility on each segment s of the nonlinear budget set j following the general constrained optimization problem:

$$v_{ijsr}(y_{ijsr}, p_{js}) = \max_{Q_{ijsr}} U_{ijsr}(Q_{ijsr}, A_{ijsr}) : p_{js}Q_{ijsr} \leq y_{ijsr}, \underline{Q_{js}} \leq Q_{ijsr} \leq \overline{Q_{js}},$$

where v is indirect utility, U is direct utility, y_s is virtual income, and p_s is the marginal price of medical care on each linear segment s of plan j. $\underline{Q_s}$ and $\overline{Q_s}$ represent the lower and upper bound on Q_s imposed by each linear segment. After maximizing utility on all segments, the agent chooses the segment and corresponding Q_{ijr} that give the highest utility. To determine plan choice in the first period, the agent chooses the plan that yields the highest expected utility.

To fully specify the agent's problem, we must impose one and only one functional form for direct utility, indirect utility, or demand. Roy's Identity,

$$-\frac{\partial v(y_{isjr}, p_{isjr})/\partial p_{isjr}}{\partial v(y_{isjr}, p_{isjr})/\partial y_{isjr}} = Q(y_{isjr}, p_{isjr}),$$

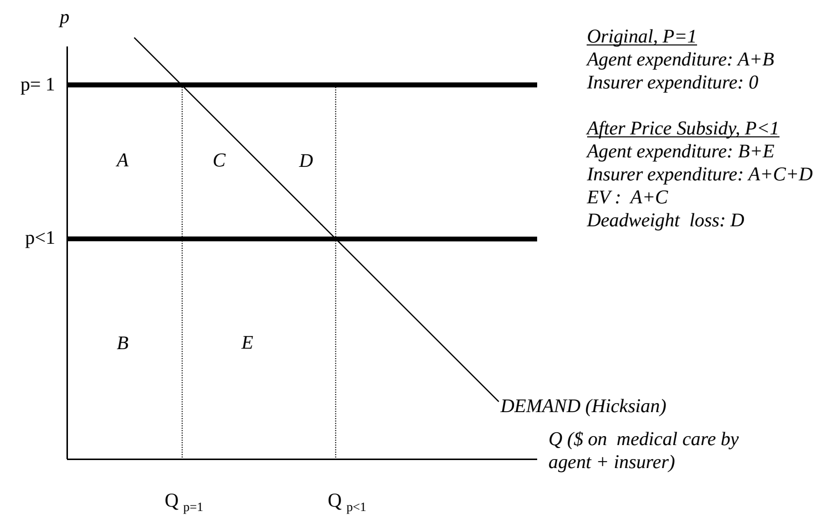
relates indirect utility to demand when the maximum utility occurs on the interior of a budget segment. Therefore, given the budget set and conditions for integrability discussed in Appendix B, this model requires a single functional form which I refer to as "demand/utility." Before specifying a functional form, I demonstrate how to calculate the tradeoff between moral hazard and risk protection in the general model.

2.1.3 Calculating the Tradeoff Between Moral Hazard and Risk Protection

First consider the deadweight loss from moral hazard. Define a plan and individual-specific measure of "moral hazard" as the dollars of extra spending incurred by agent i, induced by the substitution effect of the price change from no insurance to plan j. By definition, the no insurance case has no moral hazard and hence no deadweight loss from moral hazard. In the second period, assume that we observe the true values of the parameters as well as the agent's realization of unobserved heterogeneity r. There is no longer any value of *insurance* associated with either plan, but plan j offers lower out-of-pocket prices than the no insurance plan. Deadweight loss arises in plan j if the agent does not value the price reduction at its social cost. Even though deadweight loss results in a social welfare loss, the lower prices that the agents face that lead to the deadweight loss lead to an individual welfare gain.

Figure 3 demonstrates the deadweight loss calculation using compensated (Hicksian) demand curve for agents facing a simple linear budget set. The deadweight loss from moving from no insurance ($p = 1$) to some insurance ($p < 1$) is insurer spending ($A + C + D$) minus the equivalent

Figure 3: Calculation of DWL - Linear Budget Set



variation $(A + C)$. In a nonlinear plan, it is harder to depict deadweight loss graphically, but the relevant quantities are the same as those required in a linear plan. Building on Equation 2, we can calculate the deadweight loss of moral hazard in a general linear or nonlinear plan j for individual i with shock r as follows:

$$DWL_{ijr} = INS_{ijr} - \omega_{ijr}, \tag{1}$$

where the deadweight loss, DWL_{ijr}, is equal to insurer spending on behalf of the individual, INS_{ijr}, minus the individual's valuation of that spending (the equivalent variation), ω_{ijr}.[7] The amount of insurer spending is obtained by applying plan cost sharing rules to to the total amount of agent plus insurer spending, Q_{ijr}. If plan j offers full insurance, $INS_{i,full,r} = Q_{i,full,r}$. To obtain the equivalent variation ω_{ijr} in a general nonlinear plan j, we construct a simple indifference condition as follows:

$$U(Q_{ijr},\ y_{ijr} - p_{ijr}Q_{ijr} - \omega_{ijr}) = U(Q_{i,noins,r},\ Y - Q_{i,noins,r}), \tag{2}$$

where the left side of the equation gives utility in plan j. The first argument of the utility function, Q_{ijr}, represents medical spending, and the second argument, $y_{ijr} - p_{ijr}Q_{ijr} - \omega_{ijr}$, reflects spending on all other goods, determined by the virtual income and price on the relevant segment. The right side of the equation gives utility under no insurance, where the superscript $noins$ reflects the zero insurance budget set with zero premium. For this calculation, we do not include the premium for

[7] Note that this deadweight loss calculation is based on Hicksian demand instead of Marshallian demand. This differs from the deadweight loss calculation of Feldstein and Gruber (1995), who use Marshallian demand for simplicity. I define equivalent variation following Mas-Collell et al. (1995). As an alternative, I could base my welfare analysis on compensating variation. In practice, both measures will differ insofar as there are wealth effects, but both will provide correct welfare rankings Mas-Collell et al. (1995).

either plan, but we consider it in subsequent welfare calculations.[8]

Next, we turn to measuring the welfare gain from risk protection. For this calculation, we construct an indifference condition in the first period as follows:

$$\int U(Q_{ijr},\ y_{ijr} - p_{ijr}Q_{ijr} - \pi_{ij}) f(r_i) dr_i = \int U(Q_{i,noins,r},\ Y - Q_{i,noins,r}) f(r_i) dr_i,$$

where the left side of the equation gives expected utility over all possible values of r_i, in plan j, where utility is determined for each realization r as it is on the left side of Equation 2. The right side of the equation gives expected utility under no insurance for all possible values of r_i. The term π_{ij} captures the utility gain from insurance (the risk protection premium) as well as the utility gain from lower prices. To isolate the risk protection premium, we need to subtract the expected gains from lower prices over all r_i. We calculate RPP_{ij}, the risk protection premium for individual i under plan i, as follows:

$$RPP_{ij} = \pi_{ij} - \int (\omega_{ijr}) f(r_i) dr_i.$$

We have calculated the welfare gain from risk protection using an indifference condition in the first period, and we have calculated the welfare loss from moral hazard using an indifference condition in the second period. To examine the tradeoff between risk protection and moral hazard, we calculate the *expected* deadweight loss for agent i in the first period as follows:

$$DWL_{ij} = \int (INS_{ijr} - \omega_{ijr}) f(r_i) dr_i.$$

The tradeoff between moral hazard and risk protection, expressed as the net social benefit of insurance for agent i, is given by

$$RPP_{ij} - DWL_{ij} = \pi_{ij} - \int (INS_{ijr}) f(r_i) dr_i.$$

Thus far, we have considered the net social benefit of health insurance for a single agent i. This tradeoff can vary across individuals because individuals differ in their observable characteristics. To examine variation in the tradeoff across the population, we can calculate quantiles of DWL_{ij}, RPP_{ij}, and $RPP_{ij} - DWL_{ij}$. As another approach to examine variation across the population, we can calculate the mean tradeoff within each demographic group determined by an observable characteristic. To aggregate the welfare analysis across all individuals according to a utilitarian social welfare function that weights all agents equally, we can calculate the mean tradeoff across all individuals:

[8]Note that the price change from no insurance to plan j induces a price effect that consists of a substitution effect and an income effect, and our calculation of deadweight loss excludes the income effect. This calculation of the deadweight loss from moral hazard conforms to the recommendation of Nyman (1999), who emphasizes that in health insurance, the income effect results from a transfer of resources from the healthy to the ill through the insurer, so it should not be included in the calculation of moral hazard. In Equation 1, because ω_{ijr} measures the equivalent variation, it captures only the income effect of a price change, and it is subtracted from insurer spending in the calculation of DWL.

$$\overline{RPP_j} - \overline{DWL_j} = \frac{1}{N}\sum\nolimits_{i=1}^{N}\pi_{ij} - \frac{1}{N}\sum\nolimits_{i=1}^{N}\int(INS_{ijr})f(r_i)dr_i \tag{3a}$$

$$= \frac{1}{N}\sum\nolimits_{i=1}^{N}\pi_{ij} - m_j/\zeta, \tag{3b}$$

where $\overline{RPP_j}$ and $\overline{DWL_j}$ denote the mean risk protection premium and deadweight loss, respectively in plan j. Equation 3b gives another interpretation of the social tradeoff: it is equal to the average gains from risk protection and moral hazard minus the premium before loading. The premium for plan j, m_j, is equal to average insurer spending multiplied by the loading factor ζ. Only the premium before the loading, m_j/ζ, is included in the social tradeoff because the loading is a transfer from the agents to the insurer.[9]

To aid in assessing whether the mean calculated welfare cost is large or small, we scale it by the expected amount of money at stake for the population, $\overline{MAS}$, which we define as expected spending under no insurance:[10]

$$\overline{MAS} = \frac{1}{N}\sum\nolimits_{i=1}^{N}\int(Q_{i,noins,r})f(r_i)dr_i.$$

2.1.4 Specification of Functional Form

I choose to specify demand/utility starting with utility. I specify the following utility function in in health insurance plan j on a given linear segment s, where a denotes the first segment:

$$U(Q_{is}, A_{is}) = \left\{ \begin{array}{cc} -\exp(-\gamma A_{is}) + \frac{Q_{is}[\ln(Q_{is}/\alpha_i)-1]}{\ln\beta} & \text{if } (Q_{is} > 0 \text{ and } \alpha_i > 0) \\ -\exp(-\gamma y_{ia}) & \text{otherwise} \end{array} \right\} \tag{4}$$

where

$$\alpha_i = Z_i'\delta + r_i, r_i \sim N(\mu, \sigma^2).$$

The budget set is given by:

$$A_{is} = y_{is} - p_s Q_{is},\ 0 \leq \underline{Q_s} \leq Q_{is} \leq \overline{Q_s}. \tag{5}$$

By straightforward constrained utility maximization, Marshallian demand within segment s is given by:

$$Q_{is} = \max(\min(\alpha_i\beta^{\lambda_i p_s}, \overline{Q_s})\underline{Q_s}), \tag{6}$$

where λ denotes the marginal utility of spending on all other goods, $\gamma\exp(-\gamma(y_{is} - p_s Q_{is})) = \gamma\exp(-\gamma A_{is})$. Z_i is a vector of observable characteristics of individual i with associated vector of coefficients δ, including an indicator for male, an indicator that the employee is salaried instead

[9]Therefore, we do not need to know the loading to calculate the tradeoff. However, we do model the loading for estimation, as discussed in Section 3.1.

[10]We could use an alternative definition of money at stake, such as the portion of the premium nominally paid by the agent. Such a definition would make all of the welfare magnitudes appear larger.

of hourly, indicators for Census divisions, indicators for family size, an indicator for whether the employee was not enrolled in the a plan in the previous year, and spending in the previous year for those agents enrolled in the previous year.[11] Z_i does not include a constant. γ, $\beta > 0$, μ, σ^2, and δ are parameters to be estimated. We expect $\gamma > 0$ if agents are risk averse, with a larger value of γ indicating greater risk aversion, and we expect $0 < \beta < 1$ if demand is downward sloping, with a larger β indicating greater spending and less price sensitivity. Given parameters in the expected ranges, greater medical spending brings higher utility, and the second order condition is satisfied.[12]

This functional form builds on that of Ellis (1986).[13] It has several attractive features for studying the tradeoff between moral hazard and risk protection. First, the separability between A_s and Q_s gives a simple specification of risk aversion over A_s but not over Q_s. This specification seems realistic because health insurance can fully insure an agent against fluctuations in consumption of all other goods, but it cannot fully insure an agent against consumption of medical care.[14] I specify that agents have constant absolute risk aversion (CARA) preferences over spending on all other goods, A_s. The distinguishing feature of the CARA functional form is that it does not allow income to affect risk aversion over its argument. Although income will not affect risk aversion over A_s, income will still affect utility over medical care and hence the demand for medical care.[15] In my empirical context, all agents work for the same large employer, so income variation is not as large as it is in the population, making the CARA form attractive.

A second advantage of this functional form is that relative to the class of utility functions that imply infinite utility and demand when the price of one good is zero, this utility function implies finite demand when the price of medical care is zero, and that demand has an intuitive interpretation. When medical care is not free, our expected values of the parameters imply that spending will be lower than it is when care is free. When medical care is free ($p_s = 0$), medical spending is equal to what we would predict the agent will spend given his observable characteristics

[11] Spending from the previous year is a strong predictor of spending in the current year, motivating its inclusion. However, its inclusion generates a modeling inconsistency because while we model spending in the current year as a function of plan structure, we do not model previous year spending as a function of previous plan structure. To address this issue, it would be possible to impose the parameter estimates from the current year to calculate a measure of previous spending that would be common across plans. In turn, we could use those estimates for previous spending to estimate new parameters in the current year. Because of computational limitations, the results presented here take previous year spending as given. Alternatively, we could remove spending from the previous year from the model, sacrificing predictive power.

[12] Marginal utility of spending on medical care is given by $(\ln(Q_{is}/\alpha_i)/\ln\beta)$ if $Q_{is} > 0$ and $\alpha_i > 0$. It follows from Equation 6 that if $0 < \beta < 1$, $0 < p_s < 1$, $\lambda > 0$ (which is implied by $\gamma > 0$), then $0 < (Q_{is}/\alpha_i) < 1$. Because $\ln x \leq 0$ for $0 < x < 1$, marginal utility of spending on medical care is positive. The second derivative of the utility function with respect to $Q_{is} > 0$ is $1/(Q_{is}\ln\beta)$ if $Q_{is} > 0$ and $\alpha_i > 0$, which is negative for $0 < \beta < 1$, so the second order condition is satisfied. In practice, I check but do not impose that the estimated parameters are in the expected ranges.

[13] Relative to Ellis (1986), one main difference is that my utility function incorporates constant absolute risk aversion (CARA) preferences instead of constant relative risk aversion (CRRA) preferences over all other goods. Furthermore, Ellis (1986) does not use the proposed utility function for estimation. To make it estimable, I define utility in a single period, and I specify an α that varies across individuals.

[14] My model does not allow for health insurance to affect the distribution of medical shocks through ex ante moral hazard as discussed in Fang and Gavazza (2007). My model also does not allow for deficient provision of medical care as discussed in Ma and Riordan (2002).

[15] Income appears in the demand for medical care in Equation 6 through the marginal utility of spending on all other goods, λ.

$Z_i'\delta$, plus his realized medical care shock, r_i. If the insurer charges the same price to all agents regardless of observable characteristics, as is common in employer-sponsored plans, $Z_i'\delta$ and r_i are sources of private information that can lead to adverse selection. The units of the shock r_i can be interpreted as the dollars of care that an agent would consume in response to the shock if care were free, which is easy to conceptualize relative to specifications that involve additive shocks to utility. In theory, we could allow for a more flexible distribution of the medical shock, and we could also allow the distribution of the health medical shock as well as other parameters to vary across individuals, but we do not do so to reduce the requirements for identification.

A third advantage of this functional form is that it allows for consumption of zero care through a corner solution decision.[16] Given that in most empirical settings, a large fraction of agents consume zero care, models of the demand for medical care must incorporate a mechanism through which agents can consume zero care. One traditional method to model agents who consume zero care is through the use of a two-part model. The two-part model uses one estimating equation for the extensive margin decision to consume any care and another estimating equation for the intensive margin decision of how much care to consumer.[17] Although the two part model is a convenient and simple model, I am not aware of any exposition that shows that it is consistent with utility maximization.

One disadvantage of the functional form of the demand function is that Q_s appears on both sides of the equation. This creates a computational disadvantage because predicted demand must be obtained through maximization techniques instead of through a closed form. It also prohibits direct reduced form estimation of the demand equation, which makes it harder use the model developed here to inform reduced form techniques. However, among the functional forms that I have considered for utility, very few lead to a closed form expression for demand with the properties that I desire. For example, estimates the linear demand specification described in Kowalski (2008) does not allow for a parsimonious representation of risk protection. Since the linear demand specification does not allow for a simple representation of risk protection, estimates that use linear demand to examine moral hazard and a simple specification to examine risk protection are likely to produce results that are mutually inconsistent, motivating the use of the methods developed here.

2.2 Identification

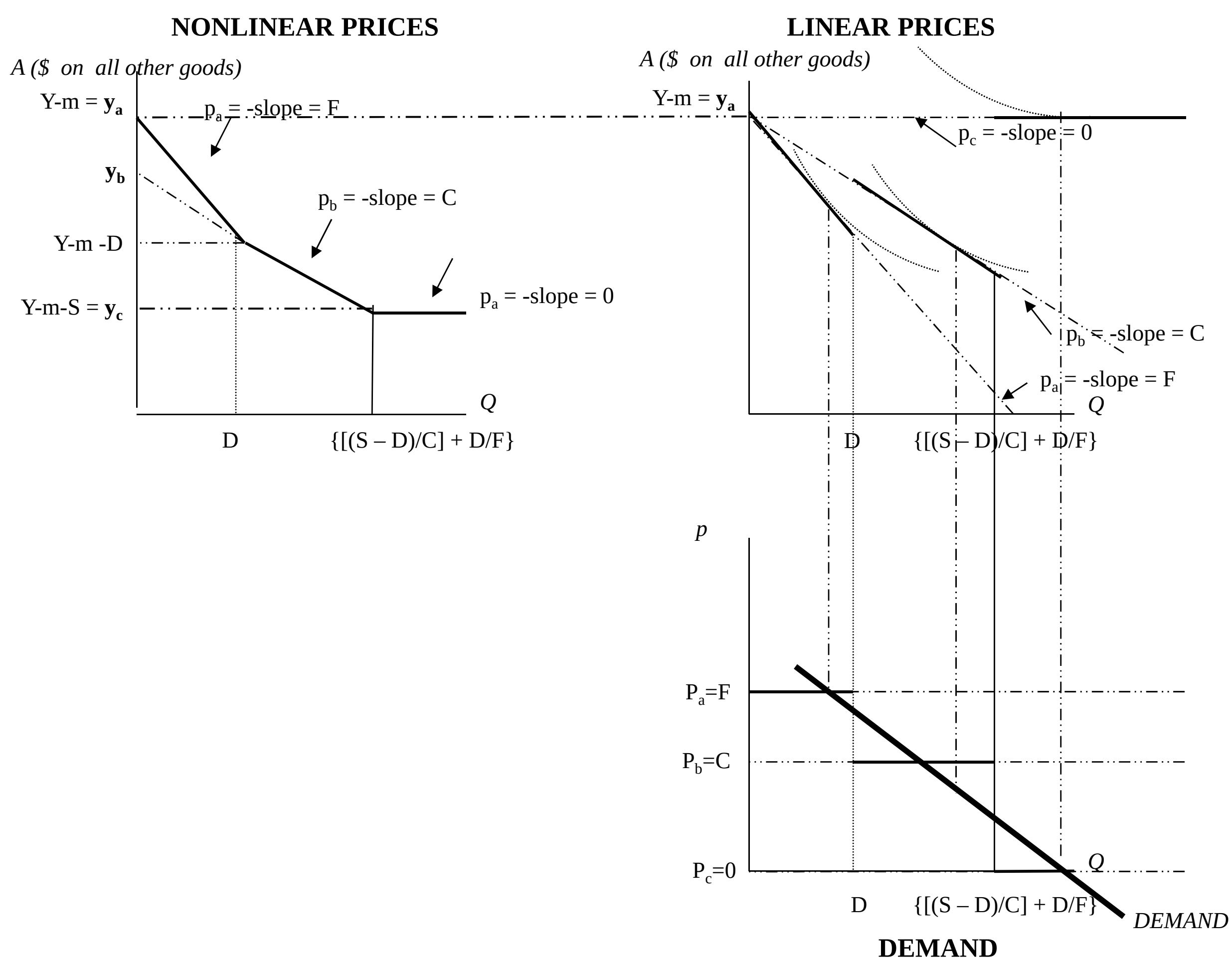
Identification of the tradeoff between moral hazard and risk protection comes partially from agents' choices of plan and spending conditional on plan, and partially from functional form. Although identification by functional form is generally undesirable, particularly if it is inaccurate, in this

[16] The lowest possible value of medical care is $\underline{Q_a} = 0$. Agents will choose this corner solution when the tangency of the indifference curve and the budget set occurs at $Q < 0$. When $Q = 0$, we impose that utility is equal to the limit of utility as $Q = 0$, as shown in the second line of Equation 4. When $\alpha < 0$, Equation 6 shows that the interior solution will occur at a negative value of Q, so the agent will consume $Q = 0$. Because utility would be undefined when when $(Q_s/\alpha < 0)$, we impose utility at zero in the second line of Equation 4.

[17] For a prominent example, see the generalization of the two part model in Manning et al. (1987). Another method to model agents who consume zero care is through the use of censored estimators, such as the censored quantile instrumental variable estimator used in Kowalski (2009). Relative to the censored quantile instrumental variable framework, the modeling approach used here requires more structure, but to the extent that the structure is correct, the modeling approach used here is more efficient.

context, the functional form of the budget set is likely to be accurate. Additional identification comes from the functional form of demand/utility and from observable heterogeneity across agents in covariates and expenditure. We consider identification of the welfare loss from moral hazard and the welfare gain from risk protection in turn.

Figure 4: Graphical Depiction of Identification



First consider identification of the welfare loss from moral hazard in the first period. The upper left of Figure 4 shows the agent's problem of utility maximization subject to a nonlinear budget set in the second period. The nonlinear budget set has three segments, with three different marginal prices. In the upper right of Figure 4, I have re-specified the agent's problem, holding virtual income constant as in the estimation. Agents choose expenditure to maximize utility subject to three different prices, and we can translate maximization subject to these three prices into a Marshallian demand curve in the lower left of the figure. In practice, additional identification comes from the bounds on each segment, but the bounds are not binding as depicted because there is a utility maximum on the interior of each bolded segment. In the data, a larger degree of moral hazard will result in more dispersion around the kinks in the budget set.

Next consider identification of the welfare gain from risk protection in the first period. In theory, I can identify the welfare gain from risk protection using the same variation that I use to identify the welfare loss from moral hazard because any demand/utilty function also implies a value of risk protection.[18] However, I can also incorporate variation in plan choice to aid in identification. In the data, a larger value of risk protection will result in the choice of more generous plans.

2.3 Estimation

As with previous nonlinear budget set models, the estimation follows directly from the model. In nonlinear budget set models with only convex kinks, it is possible to specify a closed form likelihood expression where the parameter values create an ordered choice of budget segments as in Burtless and Hausman (1978). However, because my application has more than one nonconvex kink, the utility ordering of each segment can vary across individuals, making it harder to specify a closed form likelihood. Given these limitations, I implement a simulated minimum distance estimator instead of a maximum likelihood estimator.

The simulated minimum distance estimator finds the parameter values that minimize the distance between actual spending and spending predicted by the model over all agents. Let θ denote the vector of all parameters. Given starting values of θ and the data matrix, which includes actual spending Q_i, the algorithm for the simulated distance estimator is as follows:

1. For each individual i of N, for each plan j of J, for each repetition r of R, draw $\eta_{ir} \sim N(\mu, \sigma^2)$. For each segment $s \in \{a, b, c\}$, predict

$$\widehat{Q_{ijrs}} = \arg\max_{Q_s} U_{ijrs}(Q_s, A_s) : p_{sj} Q_{ijrs} \leq y_{ijs}, \underline{Q_{sj}} \leq Q_{sj} \leq \overline{Q_{sj}}$$

   and the associated $U_{ijrs}\widehat{(Q_s, A_s)}$. Calculate the segment that yields the maximum utility for each i, j, r combination. Retain as $\widehat{Q_{ijr}}$.

2. Calculate the plan j that yields the maximum expected utility over r. Retain as $\widehat{Q_i}$.

3. Solve
$$\widehat{\theta} = \arg\min_{\theta} \sum\nolimits_{i=1}^{N} \left(\min(Q_i, \psi) - \min(\widehat{Q_i}, \psi) \right)^2 .$$

where we censor the predicted and actual values of spending at ψ so that extreme values do not drive the results.

Given estimated values of the parameters, we can predict spending in any counterfactual plan j, and we can use R simulated draws from the estimated distribution of unobserved heterogeneity $\widehat{r_{ir}} \sim N(\widehat{\mu}, \widehat{\sigma}^2)$ to calculate the deadweight loss from moral hazard and the welfare gain from risk protection as follows:

[18] Other papers in the literature also use the same variation to identify moral hazard and risk protection, but they use different functional forms for demand and utility, which are likely mutually inconsistent. For example, Engelhardt and Gruber (2010) use the establishment of Medicare Part D to identify moral hazard and risk protection, but they use different functional forms for demand and utility.

$$\widehat{DWL}_{ij} = \frac{1}{R}\sum\nolimits_{r=1}^{R}(\widehat{INS}_{ijr} - \widehat{\omega_{ijr}})$$
$$\widehat{RPP}_{ij} = \widehat{\pi_{ij}} - \frac{1}{R}\sum\nolimits_{r=1}^{R}\widehat{\omega_{ijr}}$$

where we obtain $\widehat{\pi_{ij}}$ and $\widehat{\omega_{ijr}}$ numerically as the solution to a fixed point problem.

3 Empirical Context

3.1 Data

I use 2003 and 2004 data from a firm in the retail trade industry that insures over 500,000 employees plus their enrolled family members.[19] I selected this firm because it has more than one year of available data, it has a large size, and because the four plans that it offered differed only in their nonlinear cost sharing schedules. Because the firm that I study offers only plans of the type that I can model with nonlinear budget set, my sample is not selected on the plan type dimension within the firm, which offers an advantage over other papers in this literature such as Handel (2009) and Einav et al. (2010a) that exclude Health Maintenance Organization (HMO) plans. Restricting analysis to a single firm allows me to better control for unobservable characteristics across agents, but it limits the external validity of the empirical results here as it does in other papers in this literature such as Einav et al. (2010a), Einav et al. (forthcoming), and Handel (2009).

My data include information on plan structure, claims, and enrollment compiled by Medstat (2004). The Medstat data offer several advantages over stand-alone claims data because they allow me to observe individuals that are enrolled but consume zero care in the course of the entire year, which in my estimation sample is 31% of individuals. Another advantage of the Medstat data is that they provide detailed information on plan structure, which is crucial to my nonlinear budget set analysis.

The top panel of Table 1 depicts the characteristics of the four plans offered by the firm. The firm offered only these plans in 2003 and 2004. As Table 1 shows, the deductible varies from \$350 to \$1,000; the coinsurance rate is always 0.2; and the stoploss (or maximum out-of-pocket) varies from \$2,100 to \$6,000. Agents are exposed to cost sharing until the total amount paid by the agent plus the insurer equals \$9,100 in the most generous plan to \$26,000 in the least generous plan ([(S - DF)/C] + D as shown in Figure 1). The generosity of these plans spans the range of plans typically offered in the market at the time of the data.[20]

One complicating factor is that the deductibles depicted in the table apply to individuals, and family plans also feature a family deductible and stoploss. The family deductible is three times the individual deductible and the family stoploss is two times the individual stoploss net of the

[19] As is common in other studies using claims and enrollment data, I do not observe anything about unenrolled family members or employees. According to the Bureau of Labor Statistics, the retail trade industry accounted for about 11.7 percent of all employment and about 12.9 percent of all establishments in 2004.

[20] A deductible of \$1,000 was set as the minimum deductible required by the Medicare Modernization Act of 2003 for classification as a "high deductible" plan.

Table 1: Plan Characteristics

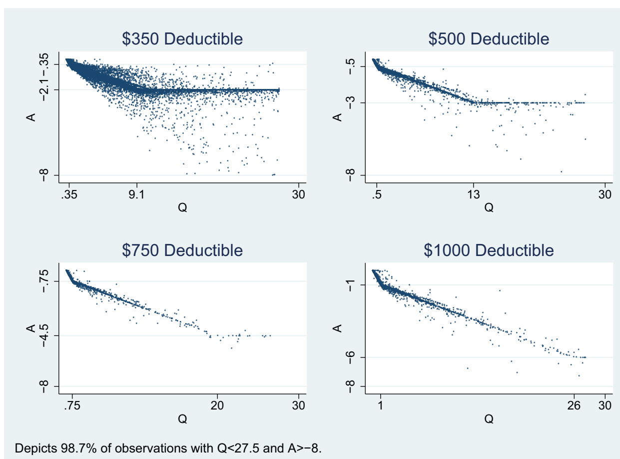
Plans	Fraction before Deductible *F*	Deductible *D*	Coinsurance *C*	Stoploss *S*
Offered				
$350 Deductible	1	350	0.2	2,100
$500 Deductible	1	500	0.2	3,000
$750 Deductible	1	750	0.2	4,500
$1,000 Deductible	1	1,000	0.2	6,000
Hypothetical				
50% Frac to $2,000 Deduct	0.5	2,000	0.2	6,000
0% Frac (Full Insurance)	0	NA	NA	NA
20% Frac	0.2	NA	NA	NA
40% Frac	0.4	NA	NA	NA
50% Frac	0.5	NA	NA	NA
60% Frac	0.6	NA	NA	NA
80% Frac	0.8	NA	NA	NA
100% Frac (No Insurance)	1	NA	NA	NA

individual deductible. The family budget set is not simply the budget set depicted in Figure 1 with the family values of the deductible and the stoploss. The budget set for someone in a family starts out as the individual budget set, as his family members spend more on medical care during the course of the year, his individual deductible and stoploss become weakly lower because of the presence of the family deductible and stoploss. Because the family deductible is *three* times the individual deductible, when *three* or more other family members have each separately met their individual deductibles, the next family member pays automatically according to the coinsurance rate.

Without some assumption about which family member's spending occurs first, I cannot model the budget sets of individual family members (or of the family). I know which family member's spending occurs first ex post, but it seems unlikely that individual family members would know whose spending will occur first ex ante. To address this issue, I limit my sample to individuals enrolled in families of three or fewer. For individuals in families of three, the family interaction occurs only at the stoploss. Since it is very unlikely that more than one individual in a family meets the stoploss, I assume that individuals in families of three maximize utility as if they face the individual stoploss. Although this assumption might introduce some measurement error, it should offer an improvement in external validity because it allows me to consider members of families. However, it will still not take into account correlated risks among family members that could lead to higher values of risk protection for individuals in families. I limit the estimation sample to the employee from each family to better control for unobservable characteristics and because all family members must choose the same plan.

Another complication that arises when I apply the nonlinear budget set model to my empirical context is that the plans offered by this firm are preferred provider organization (PPO) plans that offer incentives for beneficiaries to go to providers that are part of a network. These plans are very common: according to the Kaiser Family Foundation (2010), of workers covered by employer-sponsored health insurance, 67% of all workers and 96% of workers at large firms are covered by PPO plans. In the plans that I study, the general coinsurance rate is 20%, and the out-of-network coinsurance rate is 40%. The network itself does not vary across plans. In the data, there are no identifiers for out-of-network expenses, but the data allow me to observe beneficiary expenses as well as total expenses. As shown in Figure 5, beneficiary expenses follow the in-network schedule with a high degree of accuracy, indicating that out-of-network expenses are very rare.[21] Accordingly, in my analysis, I assume that everyone faces the in-network budget set. I use the observed value of Q and calculate the value of A that is consistent with the in-network budget set.

Figure 5: Empirical Budget Set by Plan



Additional limitations arise because of data availability. The two main data limitations are that I do not observe the premium, and I do not observe income. In the place of data on the premium, I use average insurer payments by plan, multiplied by a loading factor $\zeta = 1.25$.[22] The advantage of modeling the premium is that I can also predict the premium for counterfactual plans. I also do

[21]Out-of-network expenses cause observations to fall outside of the statutory budget set line. Even though the figure gives more visual weight to the points that fall outside of the line, there appears to be strong concentration exactly on the line. The $350 deductible plan appears to have more noise than the other plans, but it has over eight times as many enrollees, creating the appearance of more noise even if the fraction of observations that fall on distinct points away from the statutory line is the same across plans.

[22]This loading factor is motivated by Handel (2009) and Phelps (2010), page 350. Firms do not generally charge different premium amounts to different workers. However, at some firms, per-person premia can be different for individuals and families of different sizes. In the absence of premium information, I assume that per-person premia are the same regardless of family size. As shown in the fourth row of Table 2, the calculated premium is $2,498 for the $350 deductible plan, $1,496 for the $500 deductible plan, $1,032 for the $750 deductible plan, and $773 for the $1000 deductible plan. According to the Kaiser Family Foundation (2004), the average premium for individual PPO coverage at a firm with over 200 employees was $3,782 in 2004, but premia could be lower at this firm because it is especially large and self-insured.

not observe the portion of the premium paid for by the employer, so I follow the empirical evidence in assuming that the full incidence of the premium is on the worker, regardless of the statutory incidence (Gruber (2000)). Measurement error in the employee premium will have the same effect in the model as measurement error in income because both shift the entire budget set vertically. In the place of data on actual income, I use median income by zip code of residence from the 2000 census.[23] Given that this measure is likely to contain a great deal of measurement error, I do not further adjust it for taxes or for the tax-advantage of employer health insurance.

3.2 Summary Statistics and Evidence of Moral Hazard and Adverse Selection

My selected sample consists of 101,343 employees enrolled in 2004. The selected sample reflects about one fifth of the employees that I ever observe in the 2004 data. I describe sample selection in detail in the Data Appendix, available upon request. In general, I lose approximately 30% of observations because I require everyone in the family to be continuously enrolled for all of 2004, I lose approximately 20% of the remaining sample to other data issues, I lose a further 10% of the remaining employees when I restrict the sample to families of three or fewer, and then I lose approximately 50% that cannot be matched to income information.[24]

Table 2 provides descriptive statistics on the estimation sample. The first column presents statistics for the entire sample, and the other columns present statistics by plan. The most generous plan (the $350 deductible plan) is the most common plan in the data, selected by 74% of the sample. In order of decreasing generosity, the other plans enroll 12%, 4%, and 10% of employees.

The final column compares my sample to individuals aged 18 to 64 in the 2004 Medical Expenditure Panel Survey (MEPS), which is designed to be nationally representative. I restrict the MEPS sample to the reference person in each family. In general, demographic characteristics are very similar in my sample to those in the MEPS. In my sample, average median income by zip code is $40,824. At the end of the year, agents are on budget segments with an average marginal price for medical care of 0.65. The majority of the sample are women (63%), and the vast majority are hourly instead of salaried employees (92%). Workers are located in every Census Division, with the largest fractions in the South Atlantic (26.4%) and West South Central (21%), which are over-represented relative to the MEPS. The average age is 42. Of the employees enrolled in 2004, 72% can be matched to plan and expenditure information from 2003, which is summarized in the second panel of the table. The vast majority of employees in the sample are enrolled as individuals; 19% have one dependent and 8.5% have two dependents.

In my estimation sample, average annual spending by the beneficiary and insurer is $2,335, and average beneficiary spending is $619. Average spending in the MEPS on inpatient care, outpatient care, and office based visits is much lower, at $952, perhaps reflecting that individuals at my firm

[23]I censor median income from below at $10,000 on the grounds that income is likely to be higher among people with health insurance through an employer. With this restriction, in the actual policies, it is not possible for me to observe someone in the data who spends more than his income on medical care. The largest possible amount of spending on medical care is the premium plus the stoploss, which is at most $6,771.

[24]The income match cannot be conducted for all observations because some of the zip codes are missing in the Medstat data and because the 2000 Census ZTCAs do not correspond exactly with zip codes.

Table 2: Summary Statistics

	Estimation Sample					MEPS
		By Deductible				
Full Sample	**All Plans**	**$350**	**$500**	**$750**	**$1,000**	**Various**
Total Spending/1,000	2.335	2.637	1.779	1.412	1.147	0.952
Consumer Spending/1,000	0.619	0.639	0.582	0.586	0.529	0.031
Insurer Spending/1,000	1.716	1.998	1.197	0.826	0.618	0.921
Implied Premium/1,000	2.145	2.498	1.496	1.032	0.773	
Income/1,000	40.824	40.876	40.836	40.545	40.538	43.396
Virtual Income/1,000	38.440	38.120	39.137	39.331	39.601	
Price	0.650	0.598	0.731	0.815	0.872	
Male	0.373	0.336	0.443	0.464	0.532	0.515
Salary	0.077	0.072	0.101	0.089	0.087	
Census Division 1 - New England	0.017	0.017	0.014	0.018	0.021	0.191
Census Division 2 - Middle Atlantic	0.032	0.031	0.028	0.033	0.038	
Census Division 3 - East North Central	0.151	0.144	0.176	0.176	0.164	0.240
Census Division 4 - West North Central	0.101	0.089	0.143	0.138	0.128	
Census Division 5 - South Atlantic	0.264	0.281	0.215	0.222	0.215	
Census Division 6 - East South Central	0.139	0.147	0.124	0.117	0.107	0.342
Census Division 7 - West South Central	0.206	0.206	0.210	0.196	0.202	
Census Division 8 - Mountain	0.067	0.062	0.070	0.080	0.093	0.227
Census Division 9 - Pacific	0.023	0.023	0.020	0.019	0.033	
Age	42.187	42.943	41.072	39.327	39.110	42.580
Missing 2003	0.281	0.259	0.313	0.350	0.371	
2003 Spending*Nonmissing 2003	1.356	1.569	0.976	0.641	0.527	
$350 Deductible in 2003*Nonmissing 2003	0.562	0.727	0.079	0.114	0.103	
$500 Deductible in 2003*Nonmissing 2003	0.082	0.007	0.589	0.064	0.034	
$750 Deductible in 2003*Nonmissing 2003	0.023	0.002	0.008	0.455	0.018	
$1,000 Deductible in 2003*Nonmissing 200	0.053	0.004	0.011	0.017	0.475	
In Family of 2	0.189	0.170	0.240	0.247	0.244	0.298
In Family of 3	0.085	0.070	0.119	0.130	0.131	0.166
N	101,343	74,933	12,095	4,140	10,175	
Share of N	1.000	0.739	0.119	0.041	0.100	

Nonmissing in 2003 only					
2003 Spending*Nonmissing 2003	1.885	2.119	1.422	0.987	0.837
$350 Deductible in 2003*Nonmissing 2003	0.781	0.982	0.116	0.175	0.164
$500 Deductible in 2003*Nonmissing 2003	0.114	0.010	0.857	0.099	0.054
$750 Deductible in 2003*Nonmissing 2003	0.032	0.003	0.011	0.700	0.028
$1,000 Deductible in 2003*Nonmissing 200	0.073	0.006	0.016	0.026	0.754
N	72,898	56,964	8,286	2,310	5,338
Share of N	0.719	0.562	0.082	0.023	0.053

All values are for 2004 unless otherwise noted.

Census Division 1 - New England omitted.

In Family of 1 omitted.

MEPS sample consists of reference persons in each family who work and have private insurance.

spend a lot more than the national average. However, the MEPS requires individuals to report spending in a variety of categories, and categorization recall could affect the results. Respondents report an average of only $30 of spending out of pocket on inpatient care, outpatient care, and office visits, but they report total spending out of pocket of $622.63, which is much closer to beneficiary spending in my estimation sample.

In the plans that I study, average spending decreases as plan generosity decreases, providing evidence of either moral hazard or adverse selection. We can formalize this evidence using the "bivariate probit" or "positive correlation" test proposed by Chiappori and Salanie (2000), which tests for a positive correlation between the amount of insurance purchased and the amount of realized insurable spending. If such a correlation exists, it provides evidence of moral hazard and/or adverse selection. As is apparent from the summary statistics in Table 2, agents in more generous plans (those with lower deductibles) spend more on medical care. I formalize this comparison by running a regression of spending on the deductible. As reported in Table 3, I find a statistically significant negative coefficient, which indicates that there is a positive correlation between generosity and claims, implying the presence of moral hazard and/or adverse selection. This test only tests for the presence of moral hazard and/or adverse selection; it does not give any evidence on the magnitude.

Table 3: Positive Correlation Test

Positive Correlation Test (Null Hypothesis: No Moral Hazard or Adverse Selection)

Dependent Variable:	Spending			
Variable	**Estimate**		**95% confidence**	
Deductible	-2.46	***	-2.73	-2.18

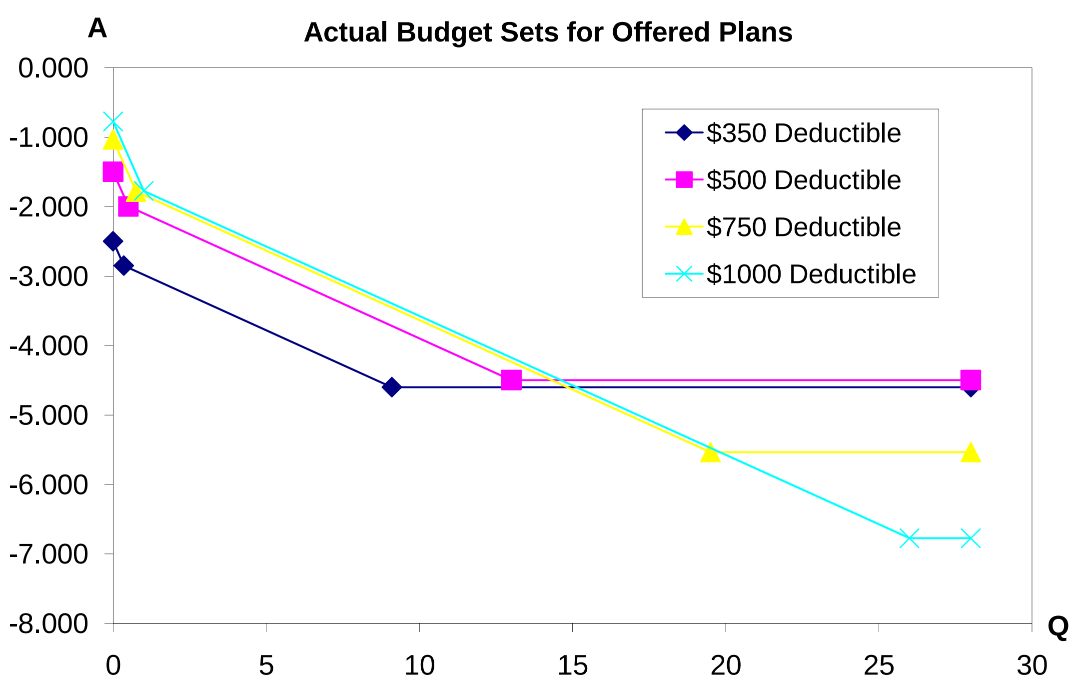
Regression includes constant (coefficient not reported).
N=101,343 R Squared = 0.0030.
***p<0.01, **p<0.05,*p<0.1

To further examine adverse selection, I implement the "unused observables" test for adverse selection proposed by Finkelstein and Poterba (2006). This approach tests for the presence of adverse selection in a context with or without moral hazard. While moral hazard results from hidden actions, adverse selection results from hidden characteristics. The premise of the unused observables test is that if the econometrician can observe any characteristics that are unpriced ("hidden" or "unused"), that are correlated with coverage generosity as well as realized insurable spending, there is evidence of adverse selection. I assume that all observable characteristics are unpriced, as is common in employer sponsored plans, so all of them can generate adverse selection. I run two regressions to implement the test - one of spending on characteristics, and another of insurance generosity (the deductible) on characteristics. I present the results in Table 4. Whether I run two separate regressions for each characteristic or two regressions that include all characteristics, I find evidence of adverse selection. Seventeen characteristics show a statistically significant relationship with the deductible and with spending, and the regressions including all characteristics have several statistically significant coefficients. Although these tests provide some evidence of adverse selection and moral hazard, they do not provide any assessment of their magnitudes or welfare costs, motivating the use of my model.

3.3 Evidence on Empirical and Predicted Plan Choice

In Figure 6, I depict the four offered plans, using the calculated premium. As shown, the \$350 deductible plan is completely dominated by the other plans - for any amount of spending, the agent would be better off in another plan.[25] Nonetheless, the vast majority of agents enroll in this plan. It is not surprising that many agents choose a dominated plan given the growing literature that suggests that health insurance plan choices exhibit biases (see Abaluck and Gruber (2009)). One reason that individuals might choose dominated plans is that switching costs might be high from one year to the next. Indeed, Handel (2009) finds that switching costs are so large that some agents select dominated plans.

Figure 6: Actual Budget Sets for Offered Plans



Because the \$350 deductible plan can never yield the highest expected utility, my model will never predict that an agent will choose it, which is problematic given that it is the most popular plan. Although my model could potentially effectively predict plan choice in other settings where dominated plans are not offered or where agents must make active choices, I alter it for use in

[25]Even though the premiums depicted are likely to be calculated with error, it is difficult to find a menu of premiums in which at least one plan would never be chosen.

Table 4: Unused Observables Test

Null Hypothesis: No Adverse Selection, With or Without Moral Hazard

	Dependent Variable: Spending							
	Separate Regressions				Single Regression			
Variable	**Estimate**		**95% confidence**		**Estimate**		**95% confidence**	
Income/1,000	0.0007		-0.0026	0.0041	0.0005		-0.0028	0.0038
Male	-1.0535	***	-1.1702	-0.9369	-0.4319	***	-0.5499	-0.3139
Salary	-0.5603	***	-0.7717	-0.3488	-0.1150		-0.3246	0.0947
Census Division 2 - Middle Atlantic	-0.7758	***	-1.0993	-0.4524	-0.0699		-0.5936	0.4538
Census Division 3 - East North Central	0.3405	***	0.1827	0.4983	0.5394	**	0.0938	0.9851
Census Division 4 - West North Central	0.0888		-0.0983	0.2760	0.4334	*	-0.0232	0.8899
Census Division 5 - South Atlantic	0.1035		-0.0246	0.2317	0.4099	*	-0.0259	0.8457
Census Division 6 - East South Central	-0.1932	**	-0.3567	-0.0297	0.2113		-0.2364	0.6591
Census Division 7 - West South Central	0.0151		-0.1246	0.1549	0.3526		-0.0871	0.7924
Census Division 8 - Mountain	-0.3339	***	-0.5599	-0.1080	0.1037		-0.3690	0.5764
Census Division 9 - Pacific	0.0325		-0.3432	0.4082	0.3454		-0.2101	0.9010
Age	0.0801	***	0.0755	0.0846	0.2332	**	0.0514	0.4151
Age Squared/100	0.0971	***	0.0918	0.1025	-0.5551	**	-1.0037	-0.1064
Age Cubed/1,000	0.1445	***	0.1366	0.1524	0.5240	***	0.1711	0.8768
Missing 2003	-0.1399	**	-0.2657	-0.0141	0.8205	***	0.6880	0.9529
2003 Spending*Nonmissing 2003	0.3058	***	0.2963	0.3153	0.4656	***	0.4478	0.4834
2003 Spending*Nonmissing 2003 Squared/1,000	0.6806	***	0.6233	0.7379	-1.9927	***	-2.2359	-1.7494
2003 Spending*Nonmissing 2003 Cubed/1,000,00	0.6628	***	0.5055	0.8200	2.2463	***	1.7068	2.7858
$500 Deductible in 2003*Nonmissing 2003	-0.4816	***	-0.6878	-0.2753	-0.3138	***	-0.5211	-0.1066
$750 Deductible in 2003*Nonmissing 2003	-0.8903	***	-1.2689	-0.5117	-0.4715	**	-0.8442	-0.0989
$1,000 Deductible in 2003*Nonmissing 2003	-1.0796	***	-1.3325	-0.8267	-0.5233	***	-0.7762	-0.2704
In Family of 2	0.3858	***	0.2414	0.5302	0.0820		-0.0627	0.2268
In Family of 3	-0.5806	***	-0.7835	-0.3778	-0.1444		-0.3489	0.0602

	Dependent Variable: Deductible							
	Separate Regressions				Single Regression			
Variable	**Estimate**		**95% confidence**		**Estimate**		**95% confidence**	
Income/1,000	-0.0001	**	-0.0002	0.0000	0.0000	**	-0.0001	0.0000
Male	0.0559	***	0.0533	0.0585	0.0010	***	0.0197	0.0236
Salary	0.0179	***	0.0132	0.0226	0.0018	***	-0.0126	-0.0056
Census Division 2 - Middle Atlantic	0.0129	***	0.0057	0.0201	0.0045		-0.0147	0.0028
Census Division 3 - East North Central	0.0132	***	0.0097	0.0167	0.0038		-0.0092	0.0057
Census Division 4 - West North Central	0.0340	***	0.0299	0.0382	0.0039		-0.0015	0.0137
Census Division 5 - South Atlantic	-0.0247	***	-0.0276	-0.0219	0.0037	***	-0.0212	-0.0067
Census Division 6 - East South Central	-0.0228	***	-0.0264	-0.0191	0.0038	***	-0.0223	-0.0073
Census Division 7 - West South Central	-0.0021		-0.0053	0.0010	0.0037		-0.0134	0.0013
Census Division 8 - Mountain	0.0312	***	0.0262	0.0362	0.0040		-0.0022	0.0136
Census Division 9 - Pacific	0.0227	***	0.0143	0.0310	0.0047		-0.0076	0.0109
Age	-0.0017	***	-0.0018	-0.0016	0.0015	*	-0.0001	0.0060
Age Squared/100	-0.0021	***	-0.0022	-0.0020	0.0038	**	-0.0150	0.0000
Age Cubed/1,000	-0.0032	***	-0.0034	-0.0030	0.0030		-0.0013	0.0105
Missing 2003	0.0376	***	0.0348	0.0404	0.0011	***	0.0989	0.1033
2003 Spending*Nonmissing 2003	-0.0021	***	-0.0023	-0.0019	0.0002	***	-0.0012	-0.0006
2003 Spending*Nonmissing 2003 Squared/1,000	-0.0019	***	-0.0032	-0.0007	0.0021	***	0.0043	0.0124
2003 Spending*Nonmissing 2003 Cubed/1,000,00	-0.0013		-0.0048	0.0022	0.0046	***	-0.0238	-0.0059
$500 Deductible in 2003*Nonmissing 2003	0.0754	***	0.0708	0.0799	0.0018	***	0.1389	0.1458
$750 Deductible in 2003*Nonmissing 2003	0.2898	***	0.2816	0.2980	0.0032	***	0.3469	0.3593
$1,000 Deductible in 2003*Nonmissing 2003	0.5254	***	0.5207	0.5300	0.0022	***	0.5626	0.5710
In Family of 2	0.0357	***	0.0325	0.0389	0.0012	***	0.0212	0.0261
In Family of 3	0.0566	***	0.0521	0.0611	0.0017	***	0.0219	0.0287

All regressions include constants (coefficients not reported).
N=101,343 for all regressions.
R squared =0.0444 in spending single regression. R squared=0.3217 in deductible single regression.
***p<0.01, **p<0.05,*p<0.1

my empirical context. Specifically, I alter my estimation algorithm to incorporate the empirical regularity that many agents remain in the same plan in both years to aid in identification of plan choice. I provide the updated algorithm for my simulated minimum distance estimator as well as other estimation and inference details in Appendix C.

In the updated algorithm, agents choose each plan with the predicted probability estimated by a multinomial logit model over all available plans. All of the characteristics in Z enter the multinomial logit model. I also include income and dummy variables for all plans that were available last year, which are populated if the agent was enrolled in the previous year. Incorporating last year's plan is advantageous because the multinomial choice model does a much better job of correctly predicting plans when it is included.[26] Furthermore, it allows for identification of the demand function in the second period through an exclusion restriction: conditional on Z, which includes spending last year, the plan from last year only affects spending this year through choice of plan. A violation would be possible in the scenario that an agent picked a generous plan last year because he expected large expenditures, but he actually had low expenditures last year and the high expenditures did not start until the current year.

One disadvantage of specifying plan choice in this way is that it has limited usefulness in predicting selection into a set of counterfactual plans.[27] Because of this limitation, I only consider counterfactual simulations in which all agents are in the same plan.

4 Results

4.1 Estimation Results

Table 5 reports the estimated coefficients. The coefficients δ_1 to δ_{17} should be positive for demographic groups with larger spending and negative for demographic groups with smaller spending. The coefficients are generally of the expected sign, with men spending less than women (as is the case among the nonelderly because of pregnancy), and individuals with higher ages spending more. We will examine the magnitude of the spending differences between demographic groups in the counterfactual simulations reported in Tables 8, 11, and 12.

The estimated coefficient of absolute risk aversion is 0.0769. Because other papers in the literature generally define CARA utility over one argument in a single argument utility function, but I define CARA utility only over A in a utility function that also includes Q, this coefficient is not directly comparable to those in the literature. The estimated price coefficient, β, is 0.33. This coefficient will affect the plan-specific measures of moral hazard that we present in the counterfactual simulations below. As discussed above, β should be between 0 and 1, with greater price sensitivity approaching 0. To produce a price elasticity for comparison to the literature, I conduct

[26]The model correctly predicts 74% of plans when last year's plan is not included and 81.9% of plans when last year's plan is included.

[27]If I wanted to model selection into plans, I could follow Handel (2009), who shows that agents make rational health insurance decisions when they are forced to make a change. In this way, if I assumed that my counterfactual simulation required all agents to change plans, I could assume that agents are rational utility maximizers across all plans. However, I prefer to eliminate the influence of selection in my counterfactual simulations so that I can focus on the tradeoff between moral hazard and risk protection.

Table 5: Estimated Coefficients

Interpretation	Parameter	Simulated Minimum Distance Estimate		95% confidence	
Mean of medical spending shock	μ	-1.0005	***	-1.2283	-0.7727
Male	δ_1	-0.5568	***	-0.6088	-0.5048
Salary/1,000	δ_2	-0.1129	***	-0.1854	-0.0403
Census Division 2 - Middle Atlantic	δ_3	-0.1290	*	-0.2699	0.0120
Census Division 3 - East North Central	δ_4	0.4612	***	0.3675	0.5549
Census Division 4 - West North Central	δ_5	0.2246	***	0.1120	0.3371
Census Division 5 - South Atlantic	δ_6	0.2912	***	0.2080	0.3745
Census Division 6 - East South Central	δ_7	0.2277	***	0.1308	0.3247
Census Division 7 - West South Central	δ_8	0.2511	***	0.1609	0.3412
Census Division 8 - Mountain	δ_9	0.0389	**	0.0056	0.0722
Census Division 9 - Pacific	δ_{10}	-0.0456		-0.1004	0.0093
Age	δ_{11}	0.1049	***	0.0931	0.1167
Age Squared/100	δ_{12}	-0.2102	***	-0.2375	-0.1829
Age Cubed/1,000	δ_{13}	0.2066	***	0.1773	0.2359
Missing 2003	δ_{14}	0.7034	***	0.6414	0.7653
2003 Spending*Nonmissing 2003	δ_{15}	0.3661	***	0.3428	0.3894
2003 Spending*Nonmissing 2003 Squared/1,000	δ_{16}	-3.0281	***	-3.7135	-2.3427
2003 Spending*Nonmissing 2003 Cubed/1,000,000	δ_{17}	1.4863		-1.3739	4.3465
In Family of 2	δ_{18}	0.0873	**	0.0166	0.1581
In Family of 3	δ_{19}	-0.0551	***	-0.0963	-0.0138
Standard deviation of medical spending shock	σ	0.0371	**	0.0067	0.0676
Coefficient of absolute risk aversion	γ	0.0769	***	0.0223	0.1314
Price sensitivity parameter	β	0.3319	***	0.1485	0.5153
N (observations)		101,343			
R (draws of ind. het.)		5			
stepsize (in thousands)		0.001			

***p<0.01, **p<0.05,*p<0.1

Confidence intervals obtained by subsampling. See text for details.

Census Division 1 - New England and In Family of 1 omitted.

a counterfactual exercise. The arc elasticity of -0.22 reported from the Rand Health Insurance Experiment comes from a counterfactual exercise that places individuals in plans with either 25% cost sharing and no stoploss or 95% cost sharing and no stoploss (Manning et al. (1987), Keeler and Rolph (1988)). In their framework, the counterfactual exercise is not as straightforward as it is here because they do not model the nonlinear budget set. Here, I can simply place agents in two counterfactual plans with constant prices p_I and p_{II}, predict associated spending, Q_I and Q_{II}, and compute a midpoint arc elasticity as follows:

$$arc = \frac{Q_I - Q_{II}}{Q_I + Q_{II}} \div \frac{p_I - p_{II}}{p_I + p_{II}}. \tag{7}$$

With this calculation, I obtain an arc elasticity in my data of -0.0015, which is much smaller than the Rand elasticity. Also, for comparison to Kowalski (2009), which computes an arc elasticity from the range of 0.2 to 1, I conduct another counterfactual simulation, and I compute an arc elasticity of -0.0021.[28]

4.2 Model Fit

Table 6 presents statistics on the fit of the model. The first panel shows the results from a regression of predicted spending on actual spending. The results indicate a good fit, with a coefficient on mean predicted spending of near one and a coefficient on the constant of zero, with both precisely estimated. The second panel shows the actual and predicted expenditure distribution by budget segment. Because the model matches expenditure but does not match segment explicitly, the match by segment provides a stricter test of model fit. The first column presents results for all plans, and other columns disaggregate the results by actual plan. In each cell, the first number shows the actual proportion of observations on each segment, the second second number shows the mean predicted proportion of observations over all draws of heterogeneity, and the third number shows the proportion of observations with just one draw of heterogeneity. The results from one draw of heterogeneity are more likely to show dispersion around budget set kinks because taking the mean over all draws smooths over the kinks.

From utility theory, we expect that no individual should locate exactly on a kink because the budget set is nonconvex. Indeed, as shown in the rows "At Deductible" and "At Stoploss," no agents are predicted to locate exactly at the kinks. To examine whether agents locate *near* the kinks, Figure 7 graphs the distribution of actual and predicted spending in the overall in the top row, around the deductible in the second row, and around the stoploss in the third row. This figure uses only the prediction from one draw of heterogeneity. It only includes agents enrolled in the $350 deductible plan. In this plan, as shown in Figure 6, the second kink occurs at $9,100 of total spending. Figures for the other plans look similar. The exercise of comparing the actual distribution to the predicted distribution is similar in spirit to those in Liebman and Saez (2006), Saez (2010), and Chetty et al. (2011), which examine "bunching" around kinks, except that my empirical context implies dispersion around the kinks. However, the distribution of predicted spending follows directly from my model, so I do not need to employ techniques to examine excess

[28] Even though this estimate is on data from the same firm as Kowalski (2009), which finds a much larger arc elasticity of -2.3 from the 0.65 to the 0.95 conditional quantiles of the expenditure distribution, this estimate is not directly comparable for several reasons. First, the interaction between individual and family deductibles causes me to limit my sample to families of four or more in Kowalski (2009) and to families of three or fewer here, so I cannot compare results from the same estimation sample in both papers. Second, the methods that I use in both studies are very different. I use a censored quantile instrumental variable estimator in Kowalski (2009) and a nonlinear budget set simulated minimum distance estimator here. Third, and perhaps most importantly, both papers rely on different sources of variation - Kowalski (2009) relies on price variation induced by the injury of a family member, and this paper relies on price variation induced by the nonlinear cost sharing rules. As shown below, the response of the distribution of spending to the cost sharing rules does not appear to be very pronounced, making the small estimated price elasticity unsurprising. I caution against too much emphasis on the comparison of the results across papers because their focus is on different questions. Here, I go beyond Kowalski (2009) by estimating the welfare implications of moral hazard and risk protection. The *relative* magnitudes of moral hazard and risk protection are important for welfare.

Table 6: Model Fit

Regression of Actual Spending on Mean Predicted Spending Over All Draws

Variable	**Estimate**	**95% confidence**	
Mean predicted spending	0.99	0.98	1.01
Constant	0.02	-0.03	0.06
N	101,343		
R Squared	0.09		

Percent of Sample by Actual and Predicted Budget Segment in Actual Plan

Actual *Mean Predicted* *One Draw Predicted*	**All**	**By Deductible** **$350**	**$500**	**$750**	**$1,000**
Zero Spending	30.88	27.39	35.92	41.30	46.37
	0.21	*0.20*	*0.19*	*0.17*	*0.30*
	0.30	*0.29*	*0.30*	*0.19*	*0.48*
Before Deductible	26.73	24.01	31.17	35.87	37.78
	6.29	*2.69*	*6.99*	*15.89*	*28.06*
	6.25	*2.65*	*6.82*	*15.97*	*28.12*
At Deductible	0.01	0.01	0.01	0.00	0.00
	0.00	0.00	0.00	0.00	0.00
	0.00	*0.00*	*0.00*	*0.00*	*0.00*
Between Deductible and Stoploss	36.99	41.90	30.18	21.69	15.15
	92.80	*96.17*	*92.76*	*83.94*	*71.64*
	92.75	*96.13*	*92.82*	*83.84*	*71.40*
At Stoploss	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.00
	0.00	*0.00*	*0.00*	*0.00*	*0.00*
After Stoploss	5.39	6.70	2.72	1.14	0.71
	0.70	*0.94*	*0.06*	*0.00*	*0.00*
	0.70	*0.93*	*0.07*	*0.00*	*0.00*
N	101,343	74,933	12,095	4,140	10,175

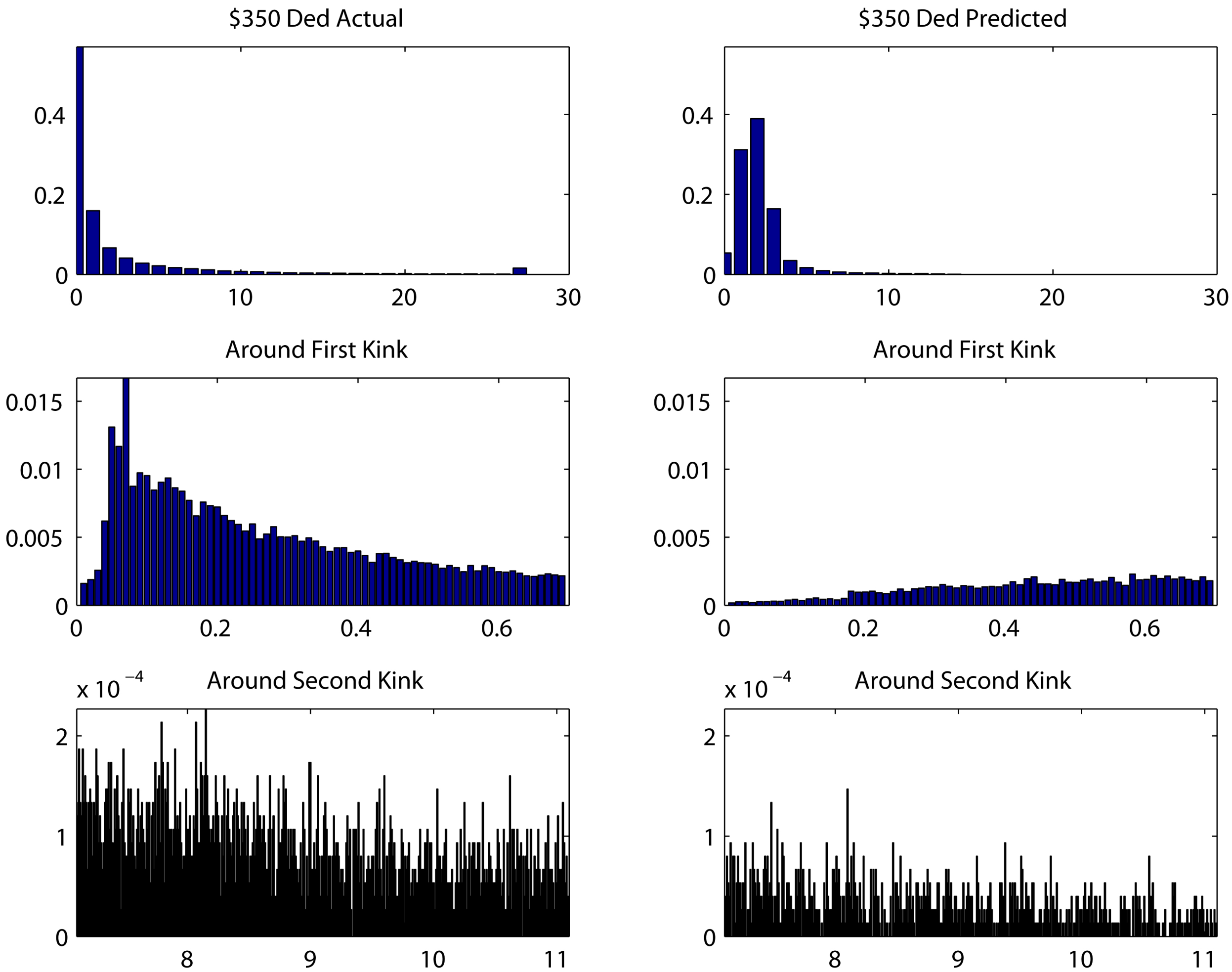
mass in the distribution.

In general, the model predicts skewness in the data. However, the first row of Figure 7 and the first row of Table 6 show that the model predicts fewer individuals with zero expenditures than we observe in the actual data. Although the model predicts that some individuals in each plan consume zero care, it underestimates the number that consume zero care.

The second and third rows of Figure 7 do not show any dispersion around the deductible or stoploss. However, the actual distribution also does not show any dispersion around the kinks. The lack of dispersion in the actual distribution previews the small estimated price elasticity.

Overall, the model fit is good in terms of average predicted expenditure, but it does not capture some aspects of the overall distribution of expenditure. Though the predicted distribution captures some skewness, the largest predicted value of spending of $14,500 is much smaller than the

Figure 7: Actual and Predicted Spending in Thousands



largest actual value. Furthermore, though the model predicts some expenditures on all segments, it overestimates the distribution in the middle and underestimates it at the extremes.

4.3 The Estimated Tradeoff Between Moral Hazard and Risk Protection

To understand the predictions of the model and to calculate the tradeoff between moral hazard and risk protection, I conduct counterfactual simulations using my model and estimates. In each simulation, I place all agents in a single plan and compare it to the no insurance case. I then compare the results across plans. In practice, I could compare any plan to any other plan directly, but the no insurance case provides a useful benchmark. I caution against interpreting the no insurance case as more than a benchmark because I do not observe any agents with no insurance in my data.

I consider the four offered plans as well as eight hypothetical plans, which I describe in Table 1. The first hypothetical plan has 50% cost sharing up to a $2,000 deductible and a $6,000 stoploss (labeled as "50% Frac to $2,000 Deduct"). I construct this "Feldstein plan" for direct comparison to the existing plan that has 100% cost sharing up to a $1,000 deductible and a $6,000 stoploss. The next seven hypothetical plans have linear cost sharing schedules with marginal prices of 1 (full insurance), 0.8, 0.6, 0.5, 0.4, 0.2, and 0 (no insurance), respectively. I label these plans as "0% Frac" to "100% Frac" in the tables. These linear plans provide simple benchmarks that vary a single marginal price.

To demonstrate the value of my model, I first perform simple counterfactual simulations without the model. In these simple simulations, I place all agents into each counterfactual plan in turn, and I predict expenditure, assuming no moral hazard. These simulations assume that each agent will spend the same amount regardless of the plan. The total amount of agent and insurer spending is only determined by the cost sharing rules in each plan. The first panel of Table 7 presents total spending, spending by the insurer, and spending by the agent in each plan. Plans are more generous if the insurer spends more. In these simple simulations, the $1,000 deductible plan appears more generous than the Feldstein plan given the distribution of total expenditure observed in the data.

In the bottom half of Table 7, I present results from counterfactual simulations from the model. Because my model allows for moral hazard, total agent plus insurer spending varies across plans.[29] Comparing total spending across plans gives a plan-specific measure of moral hazard. As the estimated price elasticity is so small, variation in spending across plans is also small, on the order of approximately $16 from full insurance to no insurance. With a larger estimated price elasticity, variation in total spending across plans would be more pronounced. The second and third columns show predicted insurer and agent spending. For these columns, comparison of the first panel of the table to the second reflects the actual *distribution* of spending vs. the predicted *distribution* of spending as well as moral hazard. Although the $1,000 deductible plan appears more generous than the Feldstein plan in the simulations without the model, here, insurer spending is higher under the Feldstein plan.

While Table 7 focuses on average spending, Table 8 shows how predicted spending varies with covariates. The first panel shows actual spending, and the second panel shows predicted spending by plan for different demographic groups. Groups with higher actual spending, such as women, individuals over median age, and hourly employees, also have higher predicted spending. We do not see a clear monotonic pattern in actual spending by income quartile, and we also do not see a clear pattern in predicted spending by income quartile. Among all demographic groups shown, insurer spending is higher under the Feldstein plan than it is under the $1,000 deductible plan. The descriptive comparison of generosity across plans and by covariates does not allow us to make statements about consumer welfare across plans because consumers must pay for extra plan generosity, and this simple exercise does not tell us how much they value extra plan generosity.

In Table 9, I present results that move beyond analysis of spending to analysis of welfare. Calculated as discussed in Section 2.1.3, the first panel of Table 9 shows the distribution of the welfare gain from insurance in the first period, π_{ij}, and the second panel shows the distribution of

[29] For all plans, predicted spending is slightly lower than it is in the top half of the table, reflecting that predicted spending is slightly lower on average than actual spending, as reflected in Table 6.

Table 7: Counterfactual Simulation Results: Spending

*Counterfactual Without Model**	Agent + Insurer Q_{ij} Mean	Insurer INS_{ij} Mean	Agent INS_{ij}-Q_{ij} Mean
Offered			
$350 Deductible	1,963.20	1,383.19	580.01
$500 Deductible	1,963.20	1,259.05	704.16
$750 Deductible	1,963.20	1,106.00	857.21
$1,000 Deductible	1,963.20	998.54	964.66
Hypothetical			
50% Frac to $2,000 Deduct	1,963.20	854.10	1,109.10
0% Frac (Full Insurance)	1,963.20	1,963.20	0.00
20% Frac	1,963.20	1,570.56	392.64
40% Frac	1,963.20	1,177.92	785.28
50% Frac	1,963.20	981.60	981.60
60% Frac	1,963.20	785.28	1,177.92
80% Frac	1,963.20	392.64	1,570.56
100% Frac (No Insurance)	1,963.20	0.00	1,963.20
Counterfactual Using Model			
Offered			
$350 Deductible	1,956.20	1,291.80	664.40
$500 Deductible	1,956.00	1,174.10	781.90
$750 Deductible	1,955.70	991.30	964.40
$1,000 Deductible	1,955.30	821.90	1,133.40
Hypothetical			
50% Frac to $2,000 Deduct	1,954.50	1,105.90	848.60
0% Frac (Full Insurance)	1,958.70	1,958.70	0.00
20% Frac	1,956.10	1,564.90	391.20
40% Frac	1,953.30	1,172.00	781.30
50% Frac	1,951.80	975.90	975.90
60% Frac	1,950.20	780.10	1,170.10
80% Frac	1,946.90	389.40	1,557.50
100% Frac (No Insurance)	1,943.10	0.00	1,943.10

Values in dollars.

*Agent+Insurer censored above $27,500 for each agent for comparison to model.
Censoring affects 1,311 agents (approximately 1.3% of sample).

the welfare gain from insurance in the second period, ω_{ij}. In the last column, I divide all means by the money at stake measure (MAS) described above as average predicted spending under no insurance. Here, MAS=$1,943.

The welfare gain in the first period varies considerably across individuals, from a minimum of 0 to a maximum of $12,162, with a median of $1,627 in the $350 deductible plan. Among the offered and hypothetical plans, the welfare gain is larger for the more generous plans. The welfare gain in the second period also varies considerably across individuals, but it is remarkably similar to the welfare gain in the first period. Because the welfare gain in the first period includes risk protection

Table 8: Counterfactual Simulation Results: Spending By Covariates

		Mean By Sex		Mean By Income Quartile				Mean By Age		Mean By Type	
				(Low)			(High)	Age<	Age>		
	Mean	Male	Female	1	2	3	4	med	med	Salary	Hourly
Actual Spending											
Offered											
All plans	2,335	1,675	2,728	2,272	2,417	2,303	2,353	1,510	3,205	1,818	2,378
$350 Deductible	2,637	1,981	2,968	2,584	2,721	2,630	2,614	1,710	3,516	2,103	2,678
$500 Deductible	1,779	1,315	2,148	1,701	1,920	1,610	1,908	1,236	2,431	1,302	1,833
$750 Deductible	1,412	943	1,818	1,285	1,451	1,317	1,626	1,142	1,847	1,065	1,446
$1,000 Deductible	1,147	868	1,465	1,128	1,186	1,104	1,180	795	1,737	1,128	1,149
Predicted Spending, Censored at 27,500											
Offered											
$350 Deductible	1,956	1,330	2,329	1,951	1,964	1,945	1,966	1,335	2,611	1,528	1,992
$500 Deductible	1,956	1,330	2,329	1,951	1,963	1,945	1,966	1,334	2,611	1,528	1,992
$750 Deductible	1,956	1,329	2,329	1,950	1,963	1,945	1,966	1,334	2,611	1,527	1,992
$1,000 Deductible	1,955	1,329	2,329	1,949	1,963	1,945	1,966	1,333	2,611	1,527	1,991
Hypothetical											
50% Frac to $2,000 Deduct	1,955	1,328	2,328	1,948	1,962	1,944	1,966	1,333	2,610	1,526	1,991
0% Frac (Full Insurance)	1,959	1,332	2,332	1,957	1,966	1,947	1,967	1,336	2,615	1,530	1,995
20% Frac	1,956	1,330	2,329	1,951	1,963	1,945	1,966	1,335	2,611	1,528	1,992
40% Frac	1,953	1,328	2,326	1,945	1,961	1,944	1,966	1,333	2,607	1,526	1,989
50% Frac	1,952	1,327	2,324	1,942	1,959	1,943	1,966	1,332	2,605	1,524	1,988
60% Frac	1,950	1,326	2,322	1,938	1,957	1,942	1,966	1,331	2,603	1,523	1,986
80% Frac	1,947	1,324	2,318	1,930	1,954	1,940	1,965	1,329	2,598	1,521	1,983
100% Frac (No Insurance)	1,943	1,322	2,314	1,922	1,950	1,938	1,965	1,327	2,593	1,518	1,979
Predicted Insurer Spending, INS_{ij}											
Offered											
$350 Deductible	1,292	796	1,587	1,288	1,298	1,283	1,299	795	1,815	950	1,321
$500 Deductible	1,174	688	1,464	1,170	1,180	1,166	1,181	685	1,689	838	1,202
$750 Deductible	991	533	1,264	987	998	984	998	519	1,489	669	1,018
$1,000 Deductible	822	406	1,070	818	828	815	828	378	1,290	520	847
Hypothetical											
50% Frac to $2,000 Deduct	1,106	719	1,336	1,101	1,111	1,100	1,113	707	1,527	829	1,129
0% Frac (Full Insurance)	1,959	1,332	2,332	1,957	1,966	1,947	1,967	1,336	2,615	1,530	1,995
20% Frac	1,565	1,064	1,863	1,561	1,571	1,556	1,573	1,068	2,089	1,222	1,594
40% Frac	1,172	797	1,396	1,167	1,176	1,166	1,180	800	1,564	915	1,194
50% Frac	976	664	1,162	971	980	971	983	666	1,303	762	994
60% Frac	780	531	929	775	783	777	786	532	1,041	609	794
80% Frac	389	265	464	386	391	388	393	266	520	304	397
100% Frac (No Insurance)	0	0	0	0	0	0	0	0	0	0	0

Values in dollars.
Median age is 43. Income first quartile: $30,208; median: $37,222; third quartile: $49,113.
Median actual spending: $242; 75th percentile: $1,420; 99th percentile: $32,193.

as well as lower prices, but the welfare gain in the second period includes lower prices only, we can see from the table that agents do not derive much value from risk protection.

Indeed, the first panel of Table 10 shows the distribution of the welfare gain from risk protection RPP_{ij} is really small - on the order of pennies. In contrast, the deadweight loss from moral hazard DWL_{ij}, shown in the second panel is much larger. Although some individuals have no deadweight loss from moral hazard, others have as much as $601 of deadweight loss.

The third panel of Table 10 gives the tradeoff between the welfare gain from risk protection and

Table 9: Counterfactual Simulation Results: Welfare in First and Second Period

		Quantiles									
	Min	**1**	**5**	**25**	**50**	**75**	**95**	**99**	**Max**	**Mean**	**Mean as % of *MAS***
Welfare Gain From Insurance in First Period Before Premium, π_{ij}											
Offered											
$350 Deductible	0	0	79	625	1,099	1,627	2,959	6,348	12,162	1,286	66
$500 Deductible	0	0	0	505	979	1,507	2,839	6,228	11,262	1,169	60
$750 Deductible	0	0	0	305	779	1,307	2,639	6,028	10,796	986	51
$1,000 Deductible	0	0	0	105	579	1,107	2,439	5,828	10,596	817	42
Hypothetical											
50% Frac to $2,000 Deduct	0	88	224	565	861	1,307	2,639	6,028	10,796	1,102	57
0% Frac (Full Insurance)	0	176	449	1,132	1,725	2,386	4,052	8,290	14,262	1,951	100
20% Frac	0	140	359	905	1,379	1,907	3,239	6,628	11,396	1,559	80
40% Frac	0	105	269	679	1,034	1,429	2,428	4,966	8,536	1,169	60
50% Frac	0	88	224	565	861	1,191	2,022	4,137	7,109	974	50
60% Frac	0	70	179	452	689	952	1,617	3,307	5,684	779	40
80% Frac	0	35	89	226	344	476	808	1,652	2,839	389	20
100% Frac (No Insurance)	0	0	0	0	0	0	0	0	0	0	0
Welfare Gain From Insurance in Second Period Before Premium, ω_{ij}											
Offered											
$350 Deductible	0	0	79	625	1,099	1,627	2,959	6,348	12,162	1,286	66
$500 Deductible	0	0	0	505	979	1,507	2,839	6,228	11,262	1,169	60
$750 Deductible	0	0	0	305	779	1,307	2,639	6,028	10,796	986	51
$1,000 Deductible	0	0	0	105	579	1,107	2,439	5,828	10,596	817	42
Hypothetical											
50% Frac to $2,000 Deduct	0	88	224	565	861	1,307	2,639	6,028	10,796	1,102	57
0% Frac (Full Insurance)	0	176	449	1,132	1,725	2,386	4,052	8,290	14,262	1,951	100
20% Frac	0	140	359	905	1,379	1,907	3,239	6,628	11,396	1,559	80
40% Frac	0	105	269	679	1,034	1,429	2,428	4,966	8,536	1,169	60
50% Frac	0	88	224	565	861	1,191	2,022	4,137	7,109	974	50
60% Frac	0	70	179	452	689	952	1,617	3,307	5,684	778	40
80% Frac	0	35	89	226	344	476	808	1,652	2,839	389	20
100% Frac (No Insurance)	0	0	0	0	0	0	0	0	0	0	0

Values in dollars. Money At Stake (MAS) is $1,943.

the deadweight loss from moral hazard. The distribution of the tradeoff at any quantile generally is not equal to the difference between DWL and RPP at those quantiles. However, as shown in the penultimate column, the mean tradeoff is equal to the mean DWL minus the mean RPP. For all offered and hypothetical plans considered, the results show that the average deadweight loss exceeds the gain from risk protection. The average net welfare loss in each of the offered plans is around $5, or 0.25% of money at stake. On average, the net welfare loss in the Feldstein plan is smaller than the welfare loss of the comparable $1,000 deductible plan. However, there is considerable variation across agents. In the offered plans, the top 1% of agents have a net *gain* from insurance that is 100 times smaller than the loss for agents at the mean. The bottom 1% of agents have a net loss from insurance that is ten times larger than the loss for the individuals at the mean.

Tables 11 and 12 provide more insight into the nature of heterogeneity in the welfare impact of insurance by presenting welfare analysis by covariates. These tables show that observed hetero-

Table 10: Counterfactual Simulation Results: DWL and RPP

	Min	1	5	25	50	75	95	99	Max	Mean	Mean as % of *MAS*
	Quantiles										
RPP $_{ij}$											
Offered											
$350 Deductible	0.00	0.00	0.00	0.02	0.03	0.05	0.10	0.13	0.27	0.04	0.002
$500 Deductible	0.00	0.00	0.00	0.02	0.03	0.05	0.09	0.13	0.27	0.04	0.002
$750 Deductible	0.00	0.00	0.00	0.01	0.03	0.05	0.09	0.13	0.27	0.04	0.002
$1,000 Deductible	0.00	0.00	0.00	0.01	0.03	0.05	0.09	0.13	0.27	0.03	0.002
Hypothetical											
50% Frac to $2,000 Deduct	0.00	0.00	0.01	0.02	0.03	0.05	0.08	0.12	0.29	0.03	0.002
0% Frac (Full Insurance)	0.00	0.00	0.01	0.02	0.03	0.06	0.10	0.14	0.39	0.04	0.002
20% Frac	0.00	0.00	0.01	0.02	0.03	0.05	0.10	0.13	0.38	0.04	0.002
40% Frac	0.00	0.00	0.01	0.02	0.03	0.05	0.08	0.12	0.33	0.04	0.002
50% Frac	0.00	0.00	0.00	0.02	0.03	0.04	0.08	0.10	0.29	0.03	0.002
60% Frac	0.00	0.00	0.00	0.01	0.02	0.04	0.06	0.09	0.25	0.03	0.001
80% Frac	0.00	0.00	0.00	0.01	0.01	0.02	0.04	0.05	0.14	0.02	0.001
100% Frac (No Insurance)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.000
DWL $_{ij}$											
Offered											
$350 Deductible	0.00	0.00	0.00	1.04	2.81	6.08	17.49	45.15	600.82	5.52	0.284
$500 Deductible	0.00	0.00	0.00	0.98	2.79	6.06	17.37	43.18	476.15	5.36	0.276
$750 Deductible	0.00	0.00	0.00	0.80	2.71	5.99	17.25	42.78	474.55	5.23	0.269
$1,000 Deductible	0.00	0.00	0.00	0.38	2.48	5.87	17.15	42.59	472.95	5.04	0.259
Hypothetical											
50% Frac to $2,000 Deduct	0.00	0.00	0.04	0.50	1.46	4.33	16.04	42.78	474.55	4.35	0.224
0% Frac (Full Insurance)	0.00	0.00	0.18	1.61	4.23	9.04	25.33	59.61	600.82	7.82	0.403
20% Frac	0.00	0.00	0.12	1.07	2.84	6.14	17.59	43.51	479.19	5.44	0.280
40% Frac	0.00	0.00	0.06	0.63	1.67	3.65	10.66	27.76	327.22	3.33	0.171
50% Frac	0.00	0.00	0.03	0.44	1.19	2.60	7.68	20.56	248.31	2.41	0.124
60% Frac	0.00	0.00	0.01	0.29	0.78	1.71	5.11	13.92	172.52	1.60	0.083
80% Frac	0.00	0.00	0.00	0.08	0.20	0.45	1.37	3.94	49.99	0.43	0.022
100% Frac (No Insurance)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.000
(RPP-DWL) $_{ij}$											
Offered											
$350 Deductible	-600.81	-45.11	-17.45	-6.04	-2.77	-1.00	0.00	0.04	0.23	-5.48	-0.282
$500 Deductible	-476.14	-43.16	-17.32	-6.02	-2.75	-0.94	0.00	0.04	0.23	-5.32	-0.274
$750 Deductible	-474.54	-42.70	-17.19	-5.95	-2.67	-0.75	0.00	0.04	0.23	-5.20	-0.268
$1,000 Deductible	-472.94	-42.54	-17.12	-5.83	-2.44	-0.34	0.00	0.03	0.23	-5.01	-0.258
Hypothetical											
50% Frac to $2,000 Deduct	-474.54	-42.70	-16.00	-4.29	-1.42	-0.47	0.00	0.05	0.25	-4.32	-0.222
0% Frac (Full Insurance)	-600.81	-59.59	-25.29	-9.00	-4.19	-1.57	-0.14	0.04	0.24	-7.78	-0.400
20% Frac	-479.18	-43.46	-17.55	-6.10	-2.80	-1.03	-0.07	0.05	0.23	-5.40	-0.278
40% Frac	-327.20	-27.70	-10.63	-3.61	-1.64	-0.59	-0.02	0.05	0.29	-3.29	-0.169
50% Frac	-248.30	-20.52	-7.65	-2.57	-1.16	-0.41	0.00	0.05	0.25	-2.37	-0.122
60% Frac	-172.51	-13.89	-5.08	-1.69	-0.75	-0.26	0.01	0.04	0.21	-1.58	-0.081
80% Frac	-49.98	-3.92	-1.36	-0.44	-0.19	-0.06	0.01	0.03	0.14	-0.42	-0.022
100% Frac (No Insurance)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.000

Values in dollars. Money At Stake (MAS) is $1,943.

Table 11: Counterfactual Simulation Results: Welfare in First and Second Period by Covariates

		Mean By Sex		Mean By Income Quartile				Mean By Age		Mean By Type	
	Mean	Male	Female	(Low) 1	2	3	(High) 4	Age< med	Age> med	Salary	Hourly
Welfare Gain From Insurance in First Period Before Premium, π_{ij}											
Offered											
$350 Deductible	1,286	792	1,581	1,276	1,292	1,281	1,298	792	1,807	946	1,315
$500 Deductible	1,169	685	1,457	1,158	1,175	1,163	1,181	682	1,682	834	1,197
$750 Deductible	986	530	1,258	975	992	981	997	516	1,482	665	1,013
$1,000 Deductible	817	403	1,064	806	823	813	827	375	1,283	517	842
Hypothetical											
50% Frac to $2,000 Deduct	1,102	717	1,331	1,091	1,107	1,097	1,112	705	1,520	826	1,125
0% Frac (Full Insurance)	1,951	1,327	2,323	1,940	1,958	1,942	1,966	1,331	2,604	1,524	1,987
20% Frac	1,560	1,061	1,857	1,549	1,565	1,553	1,572	1,064	2,081	1,218	1,588
40% Frac	1,169	795	1,392	1,160	1,173	1,164	1,179	798	1,560	913	1,190
50% Frac	974	662	1,159	965	977	970	983	665	1,299	761	991
60% Frac	779	529	927	772	781	776	786	531	1,039	608	793
80% Frac	389	265	463	385	390	388	393	266	519	304	396
100% Frac (No Insurance)	0	0	0	0	0	0	0	0	0	0	0
Welfare Gain From Insurance in Second Period Before Premium, ω_{ij}											
Offered											
$350 Deductible	1,286	792	1,581	1,276	1,292	1,280	1,298	792	1,807	946	1,315
$500 Deductible	1,169	685	1,457	1,158	1,175	1,163	1,180	682	1,682	834	1,197
$750 Deductible	986	530	1,258	975	992	981	997	516	1,482	665	1,013
$1,000 Deductible	817	403	1,064	806	823	813	827	375	1,283	517	842
Hypothetical											
50% Frac to $2,000 Deduct	1,102	717	1,331	1,091	1,107	1,097	1,112	705	1,520	826	1,125
0% Frac (Full Insurance)	1,951	1,327	2,323	1,940	1,958	1,942	1,966	1,331	2,604	1,524	1,987
20% Frac	1,560	1,061	1,857	1,549	1,565	1,553	1,572	1,064	2,081	1,218	1,588
40% Frac	1,169	795	1,392	1,160	1,173	1,164	1,179	798	1,560	913	1,190
50% Frac	974	662	1,159	965	977	970	983	665	1,299	760	991
60% Frac	779	529	927	772	781	776	786	531	1,039	608	793
80% Frac	389	265	463	385	390	388	393	266	519	304	396
100% Frac (No Insurance)	0	0	0	0	0	0	0	0	0	0	0

Values in dollars.
Median age is 43. Income first quartile: $30,208; median: $37,222; third quartile: $49,113.
Median actual spending: $242; 75th percentile: $1,420; 99th percentile: $32,193.

geneity goes a long way in explaining the variation across the valuation quantiles. Table 11 shows the welfare gain from insurance in the first and second period for distinct demographic groups. We see that the welfare gains that we estimate from insurance in Table 11 are very similar to predicted insurer spending reported in Table 8; demographic groups with higher predicted insurer spending derive a larger welfare gain from insurance than demographic groups with lower predicted insurer spending. As we see in Table 12, demographic groups with larger predicted insurer spending also have larger deadweight losses. For example, women have predicted insurer spending that is almost twice as large as that for men. The welfare gain from insurance in each period and the deadweight loss from moral hazard is also roughly twice as large for women as it is for men. Risk protection does not appear to vary with predicted insurer spending because the magnitudes of the risk protection premium are so small, but there is also some variation in the risk protection premium across demographic groups.

Table 12: Counterfactual Simulation Results: DWL and RPP By Covariates

		Mean By Sex		Mean By Income Quartile				Mean By Age		Mean By Type	
				(Low)			(High)	Age<	Age>		
	Mean	**Male**	**Female**	**1**	**2**	**3**	**4**	**med**	**med**	**Salary**	**Hourly**
***RPP** $_{ij}$*											
Offered											
$350 Deductible	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
$500 Deductible	0.04	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
$750 Deductible	0.04	0.03	0.04	0.04	0.04	0.04	0.04	0.03	0.04	0.03	0.04
$1,000 Deductible	0.03	0.02	0.04	0.03	0.03	0.03	0.03	0.03	0.04	0.03	0.03
Hypothetical											
50% Frac to $2,000 Deduct	0.03	0.03	0.04	0.03	0.03	0.04	0.04	0.03	0.04	0.03	0.04
0% Frac (Full Insurance)	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
20% Frac	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
40% Frac	0.04	0.03	0.04	0.03	0.04	0.04	0.04	0.04	0.04	0.03	0.04
50% Frac	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
60% Frac	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
80% Frac	0.02	0.01	0.02	0.01	0.02	0.02	0.02	0.02	0.02	0.01	0.02
100% Frac (No Insurance)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
***DWL** $_{ij}$*											
Offered											
$350 Deductible	5.52	3.58	6.68	12.43	5.70	2.96	0.80	3.32	7.84	4.08	5.64
$500 Deductible	5.36	3.42	6.52	12.07	5.54	2.88	0.78	3.23	7.61	3.96	5.48
$750 Deductible	5.23	3.21	6.44	11.78	5.41	2.81	0.76	3.07	7.52	3.80	5.36
$1,000 Deductible	5.04	2.90	6.32	11.34	5.22	2.70	0.74	2.75	7.45	3.53	5.17
Hypothetical											
50% Frac to $2,000 Deduct	4.35	2.40	5.51	9.77	4.51	2.34	0.63	2.03	6.80	2.81	4.48
0% Frac (Full Insurance)	7.82	5.16	9.41	17.64	8.05	4.18	1.14	4.93	10.87	5.95	7.98
20% Frac	5.44	3.55	6.57	12.26	5.62	2.92	0.79	3.35	7.65	4.08	5.56
40% Frac	3.33	2.15	4.03	7.47	3.44	1.79	0.49	2.00	4.73	2.46	3.40
50% Frac	2.41	1.55	2.92	5.40	2.49	1.30	0.35	1.43	3.44	1.76	2.46
60% Frac	1.60	1.03	1.95	3.59	1.66	0.87	0.23	0.94	2.30	1.17	1.64
80% Frac	0.43	0.28	0.53	0.97	0.45	0.23	0.06	0.25	0.63	0.31	0.44
100% Frac (No Insurance)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
***(RPP-DWL)** $_{ij}$*											
Offered											
$350 Deductible	-5.48	-3.55	-6.63	-12.39	-5.66	-2.92	-0.76	-3.29	-7.80	-4.04	-5.60
$500 Deductible	-5.32	-3.38	-6.48	-12.03	-5.50	-2.84	-0.74	-3.19	-7.57	-3.93	-5.44
$750 Deductible	-5.20	-3.18	-6.40	-11.75	-5.37	-2.77	-0.73	-3.03	-7.48	-3.77	-5.32
$1,000 Deductible	-5.01	-2.87	-6.28	-11.31	-5.18	-2.67	-0.70	-2.73	-7.41	-3.51	-5.13
Hypothetical											
50% Frac to $2,000 Deduct	-4.32	-2.37	-5.48	-9.74	-4.48	-2.31	-0.60	-2.00	-6.76	-2.77	-4.45
0% Frac (Full Insurance)	-7.78	-5.12	-9.36	-17.60	-8.01	-4.14	-1.09	-4.89	-10.82	-5.91	-7.94
20% Frac	-5.40	-3.51	-6.53	-12.22	-5.58	-2.88	-0.75	-3.31	-7.61	-4.04	-5.52
40% Frac	-3.29	-2.11	-3.99	-7.44	-3.41	-1.75	-0.45	-1.96	-4.69	-2.42	-3.36
50% Frac	-2.37	-1.52	-2.89	-5.37	-2.46	-1.26	-0.32	-1.40	-3.41	-1.73	-2.43
60% Frac	-1.58	-1.00	-1.92	-3.57	-1.64	-0.84	-0.21	-0.91	-2.27	-1.14	-1.61
80% Frac	-0.42	-0.26	-0.51	-0.95	-0.44	-0.22	-0.05	-0.23	-0.61	-0.30	-0.43
100% Frac (No Insurance)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Values in dollars.
Median age is 43. Income first quartile: $30,208; median: $37,222; third quartile: $49,113.
Median actual spending: $242; 75th percentile: $1,420; 99th percentile: $32,193.

4.4 Discussion

Within my empirical context, the average deadweight losses from moral hazard substantially outweigh the average welfare gains from risk protection. However, both quantities are empirically very small relative to money at stake. Closer consideration of the institutional features of health insurance markets and closer examination of my empirical context relative to the broader market can help to explain the the results.

Two institutional features of the market for employer sponsored health insurance have thus far been outside of the scope of the analysis. First, there is a tax advantage for employer sponsored health insurance, which means that firm and employee portions of health insurance premiums can generally be paid before taxes. Second, employees at the same firm generally pay the same premium regardless of their observable characteristics because of regulations that govern Section 125 cafeteria plans. Thus far, the welfare calculations have not taken either of these institutional features into account, as they have reported welfare gains and losses before consideration of the premium. As shown in Table 7, average predicted insurer spending in the \$350 deductible plan is \$1,292, so if there were no cost of running the insurance plan, the premium would be equal to that amount. More likely, as discussed above, the premium after loading would be equal to 1.25 times insurer spending, \$1,615. As shown in Table 9, the mean welfare gain from insurance in the first period is \$1,286. Individuals in the 75th percentile and above have a welfare gain that is larger than the predicted premium after loading, but other individuals do not, calling into question why they would enroll in the plan at all.

The tax advantage for employer sponsored health insurance and the de facto "community rating" (the insurance practice of charging all individuals a common community price) within employer sponsored plans can help to explain why individuals enroll, even though my welfare calculations would predict that enrollment would lead to a welfare loss. First, the tax advantage serves to substantially reduce the premium. Assuming an average tax rate of 35%, the effective premium would be only \$1,050, which is smaller than the welfare gain from insurance for the median individual. Therefore, the tax advantage alone explains the enrollment for approximately another 25% of the population. Second, because individuals pay the same price regardless of their predicted spending, it is rational for individuals with high predicted spending to enroll because they receive transfers from other individuals as well as from the government. For example, as shown in Table 8, individuals above median age enrolled in the \$350 deductible plan have expected insurer spending of \$1,815, which is much higher than the premium they face.

However, even after considering these two institutional features, it remains a puzzle why individuals with low predicted spending, such as individuals below median age, whose expected insurer spending in the \$350 deductible plan is \$795 enroll. One answer is that they do not actually enroll in the in the \$350 deductible plan - they enroll in a less generous plan such as the \$1,000 deductible plan. Indeed, Table 2 shows that individuals in the less generous plans are younger on average. Our positive correlation test showed evidence of moral hazard and or adverse selection. Given that the model shows a small magnitude of moral hazard with a small associated welfare impact, adverse selection must play a role. As we formalized with the unused observables test in Table 4, younger individuals enroll in less generous plans and have lower spending than their older counterparts. Our welfare calculations did not take into account the welfare impact of this adverse selection because they held individual characteristics constant on both sides of the indifference conditions. Taking into account that individuals with lower predicted spending sorted into the least generous plans and that those plans had lower premiums explains the enrollment of a greater fraction of the population. Our counterfactual simulation that places all individuals into the same plan predicts insurer spending of $822 in the least generous $1,000 deductible plan, but actual insurer spending would be much lower given adverse selection. If the premium is set based on average spending in each plan, the premium for the least generous plan should be much less than $822. However, it still remains a mystery why some individuals with very low predicted welfare gains from insurance would enroll.

To explain the enrollment of the remaining individuals with the smallest predicted welfare gains from insurance, it is helpful to take into account some additional features of my empirical context. Most importantly, even the least generous $1,000 deductible plan only leaves individuals subject the risk of spending up to $6,000 out of pocket in a given year, which occurs when the individual incurs $26,000 or more in spending. Given that the value of risk protection derives from protection from large expenditures, but none of the agents in my data actually face large expenditure risk by virtue of their enrollment, my results likely understate the welfare gains from risk protection. The reported welfare gains from risk protection are based on small variation in maximum actual risk from $2,100 to $6,000. Therefore, it is not surprising and likely true that risk protection does not vary much across the plans that I study, but protection from catastrophic risk of the type not observed in my data could explain the enrollment of the remaining agents.

In some sense, it is a limitation of my paper that my empirical context focuses on individuals in employer sponsored plans. If my analysis included uninsured individuals, I might find a larger value for risk protection, and the tradeoff between moral hazard and risk protection would be more empirically relevant. In Appendix D, I describe how to calculate calculate optimal insurance generosity using counterfactual simulations from my results, and I find that optimal insurance in my context should have very limited generosity because incremental moral hazard is so large relative to incremental risk protection. Observation of a wider range of generosities in the data would give the results more external validity.

In another sense, however, my focus on individuals in employer sponsored plans is not a limitation of my analysis because employer sponsored health insurance is so ubiquitous. Within this context, my results imply that because medical spending can be predicted so readily by a small vector of observable characteristics, the welfare gain from risk protection is small relative to transfers from other agents through the premium and from the government through the tax preference for employer sponsored health insurance. In the plans that I study, because individuals are already protected from catastrophic risk, incremental generosity yields a larger deadweight loss from moral hazard than it does a welfare gain from risk protection.

5 Conclusion

Using the theory of utility maximization subject to a nonlinear constraint, I develop a model to estimate the tradeoff between the welfare gain from risk protection and the welfare loss from moral hazard in health insurance plans. Relative to the literature on the tradeoff between moral hazard and risk protection, my model allows for estimation of both sides of the tradeoff using the same framework. Relative to the literature on moral hazard, I incorporate the choice of zero care as a corner solution decision within my model.

One important aspect of my methodology is that I take into account the agent's response to the entire nonlinear budget set induced by the price schedule for medical care under insurance. Recent work suggests that it is important to consider the nonlinear pricing schedule for medical care because agents are not strictly myopic. (See Aron-Dine et al. (2012)). I advance the nonlinear budget set literature by allowing for risk protection and by developing a simulated minimum distance estimator that allows for estimation when there is more than one nonconvex kink in the budget set. Relative to other nonlinear budget set applications, the medical care application allows for a particularly tight link between the agent's actual budget set, the model, and the estimation strategy. However, my model could potentially be applied in other contexts to estimate the tradeoff between risk protection and insurance in social programs with benefits that are nonlinear in income.

I estimate my model using data on employees with health insurance in a specific empirical context. My empirical context focuses on individuals who purchase health insurance through a single large firm. Focusing on a specific empirical context places some limitations on external validity. However, the welfare implications of nonlinear health insurance policies offered by firms is relevant for policy because recent national health reform legislation will require most individuals to purchase health insurance, and it will collect penalties from firms that do not provide coverage. In counterfactual simulations that require agents to purchase a single plan offered by a large firm, I find that the average deadweight loss from moral hazard outweighs the average welfare gain from risk protection. However, the welfare impact of moral hazard and risk protection are both small relative to the welfare impact of transfers from other agents through the premium and from the government through the tax preference for employer sponsored health insurance.

References

Jason Abaluck and Jonathan Gruber. Choice inconsistencies among the elderly: Evidence from plan choice in the medicare part d program. *NBER Working Paper 14759*, February 2009.

Aviva Aron-Dine, Liran Einav, and Amy Finkelstein adn Mark Cullen. Moral hazard in health insurance: How important is forward looking behavior? *NBER Working Paper 17802*, February 2012.

Kenneth J. Arrow. Uncertainty and the welfare economics of medical care. *The American Economic Review*, 53(5):941–973, 1963.

Martin Neil Baily. Some aspects of optimal unemployment insurance. *Journal of Public Economics*, 10:379–402, December 1978.

Patrick Bajari, Han Hong, Minjung Park, and Robert Town. Regression discontinuity designs with an endogenous forcing variable and an application to contracting in health care. *Working Paper*, 2010.

Soren Blomquist. Restriction in labor supply estimation: Is the macurdy critique correct? *Economics Letters*, 47:229–235, 1995.

Soren Blomquist and Whitney Newey. Nonparametric estimation with nonlinear budget sets. *Econometrica*, 70(6):2455–80, 2002.

Gary Burtless and Jerry A. Hausman. The effect of taxation on labor supply: Evaluating the gary negative income tax experiment. *The Journal of Political Economy*, 86(6):1103–30, 1978.

Gary Burtless and Robert A. Moffit. The joint choice of retirement age and postretirement hours of work. *Journal of Labor Economics*, 3(2):209–236, April 1985.

James H. Cardon and Igal Hendel. Symmetric information in health insurance: Evidence from the national medical expenditure survey. *Rand Journal of Economics*, 32(3):408–427, 2001.

Raj Chetty. A general formula for the optimal level of social insurance. *Journal of Public Economics*, 90(10-11):1879–1901, November 2006.

Raj Chetty. A new method of estimating risk aversion. *American Economic Review*, 96(5):1821–1834, December 2006b.

Raj Chetty, John N. Friedman, Tore Olsen, and Luigi Pistaferri. Adjustment costs, firm responses, and labor supply elasticities: Evidence from danish tax records. *Quarterly Journal of Economics*, 126(2):749–804, 2011.

Pierre Andre Chiappori and Bernard Salanie. Testing for asymmetric information in insurance markets. *The Journal of Poltical Economy*, 108(1):56–58, February 2000.

Fabian Duarte. Price elasticity of expenditure across health care services. *Yale University Dissertation*, Chapter 2, 2010.

Isaac Ehrlich and Gary S. Becker. Market insurance, self-insurance, and self-protection. *Journal of Politcal Economy*, 80(4):623–48, July/August 1972.

Matthew J. Eichner. Medical expenditures and major risk health insurance. *Massachusettes Institute of Technology Dissertation*, pages 1–66, 1997.

Matthew J. Eichner. The demand for medical care: What people pay does matter. *The American Economic Review. Papers and Proceedings of the Hundred and Tenth Annual Meeting of the American Economic Association*, 88(2):117–121, May 1998.

Liran Einav, Amy Finkelstein, and Mark Cullen. Estimating welfare in insurance markets using variation in prices. *Quarterly Journal of Economics*, 125(3):877–921, August 2010a.

Liran Einav, Amy Finkelstein, and Jonathan Levin. Beyond testing: Empirical models of insurance market. *Annual Review of Economics*, 2(1):311–336, September 2010b.

Liran Einav, Amy Finkelstein, Stephen Ryan, Paul Schrimpf, and Mark Cullen. Selection on moral hazard in health insurance. *American Economic Review*, forthcoming.

Randall Ellis. Rational behavior in the presence of coverage ceilings and deductibles. *The RAND Journal of Economics*, 17(2):158–175, 1986.

Gary V. Engelhardt and Jonathan Gruber. Medicare part d and the financial protection of the elderly. *NBER Working Paper 1615*, 2010.

Gary V. Engelhardt and Anil Kumar. Employer matching and 401(k) saving:evidence from the health and retirement study. Mimeo, 2006.

Hanming Fang and Alessandro Gavazza. Dynamic inefficiencies in employment-based health insurance system theory and evidence. *NBER Working Paper 13371*, September 2007.

Roger Feldman and Bryan Dowd. A new estimate of the welfare loss of excess health insurance. *The American Economic Review*, 81(1):297–301, March 1991.

Martin Feldstein. The welfare loss of excess health insurance. *The Journal of Political Economy*, 81(2, Part 1):251–280, April 1973.

Martin Feldstein. Balancing the goals of health care provision and financing. *Health Affairs*, 25(6): 1603–1611, 2006.

Martin Feldstein and Jonathan Gruber. A major risk approach to health insurance reform. *Tax Policy and the Economy*, 9(1):103–130, 1995.

Amy Finkelstein and Robin McKnight. What did medicare do? the initial impact of medicare on mortality and out of pocket medical spending. *Journal of Public Economics*, 92(7):1644–1668, 2008.

Amy Finkelstein and James Poterba. Testing for adverse selection with unused 'observables. *NBER Working Paper 12112*, March 2006.

Leora Friedberg. The labor supply effects of the social security earnings test. *The Review of Economics and Statistics*, 82(1):48–63, February 2000.

Michael Grossman. *The Demand for Health.* Columbia University Press, 1972.

Jonathan Gruber. Chapter 12: Health insurance and the labor market. *Handbook of Health Economics*, 1(1):645–706, 2000.

Janice Halpern and Jerry Hausman. Choice under uncertainty: A model of applications for the social security disability insurance program. *Journal of Public Economics*, 31:131–161, 1986.

Benjamin Handel. Adverse selection and switching costs in health insurance markets: When nudging hurts. Northwestern University. Mimeo, 2009.

Jerry A. Hausman. Individual discount rates and the purchase and utilization of energy-using durables. *The Bell Journal of Economics*, 10(1):33–54, 1979.

Jerry A. Hausman. The effect of wages, taxes, and fixed costs on women's labor force participation. *Journal of Public Economics*, 14(2):161–94, October 1980.

Jerry A. Hausman. *The Effect of Taxes on Labor Supply," H. Aaron and J. Pechman, How Taxes Affect Economics Behavior.* Washington, D.C.: Brookings, 1981.

Jerry A. Hausman. The econometrics of nonlinear budget sets. *Econometrica*, 53(6):1255–82, 1985.

Bradley T. Heim and Bruce D. Meyer. Structural labor supply models when constraints are nonlinear. *Mimeo*, December 2003.

Michael Hurd. Estimation of nonlinear labor supply functions with taxes from a truncated sample. *Stanford Research Institute. Research Memorandum 36*, November 1976.

Kaiser Family Foundation. Employer health benefits 2004 annual survey, 2004. http://www.kff.org/insurance/7148/upload/2004-Employer-Health-Benefits-Survey-Full-Report.pdf.

Kaiser Family Foundation. Employer health benefits 2010 annual survey, 2010. http://ehbs.kff.org/pdf/2010/8085.pdf.

Dean Karlan and Jonathan Zinman. Observing unobservables: Identifying information asymmetries with a consumer credit field experiment. *Econometrica*, 77(6):1993–2008, November 2009.

E.B. Keeler, J.P. Newhouse, and C.E. Phelps. Deductibles and the demand for medical care services: The theory of a consumer facing a variable price schedule under uncertainty. *Econometrica*, 45 (3):641–656, April 1977.

Emmett Keeler and John E. Rolph. The demand for episodes of treatment in the health insurance experiment. *Journal of Health Economics*, 7(4):337–367, December 1988.

Jonathan T. Kolstad and Amanda E. Kowalski. The impact of an individual health insurance mandate on hospital and preventive care: Evidence from massachusetts. *NBER Working Paper 16012*, 2010.

Amanda E. Kowalski. Nonlinear budget sets and medical care. *Massachusetts Institute of Technology. Dissertation. Chapter 3*, pages 125–155, 2008.

Amanda E. Kowalski. Censored quantile instrumental variable estimates of the price elasticity of expenditure on medical care. *NBER Working Paper 15085*, 2009.

Anil Kumar. Nonparametric estimation of the impact of taxes on female labor supply. *Federal Reserve Bank of Dallas*, July 2004.

Jeffrey Liebman and Emmanuel Saez. Earnings responses to increases in payroll taxes. *Mimeo*, September 2006.

Jeffrey Liebman and Richard Zeckhauser. Schmeduling. Mimeo, October 2004.

Hamish Low and Luigi Pistaferri. Disability risk and the value of disability insurance. *Stanford University Working Paper*, October 2010.

Josh Lustig. Measuring welfare losses from adverse selection and imperfect competition in privatized medicare. *Manuscript. Boston University Department of Economics*, March 2010.

C.-T. Albert Ma and M. H. Riordan. Health insurance, moral hazard, and managed care. *Journal of Economics and Management Strategy*, 11:81–107, 2002.

Thomas MaCurdy. Work disincentive effects of taxes: A reexamination of some evidence. *American Economic Review*, 82(2):243–49, 1992.

Thomas MaCurdy, David Green, and Harry Paarsh. Empirical approaches for analyzing taxes and labor supply. *The Journal of Human Resources. Special Issue on Taxation and Labor Supply in Industrial Countries*, 25(3):4415–490, 1990.

Willard G. Manning and M. Susan Marquis. Health insurance: The tradeoff between risk pooling and moral hazard. *Journal of Health Economics*, 15(5):609 – 639, 1996.

Willard G. Manning, Joseph P. Newhouse, Naihua Duan, Emmet B. Keller, and Arleen Leibowitz. Health insurance and the demand for medical care: Evidence from a randomized experiment. *The American Economic Review*, 77(3):251–277, June 1987.

Christina Marsh. Estimating health expenditure elasticites using nonlinear reimbursement. University of Minnesota. Mimeo, November 2009.

Andrew Mas-Collell, Michael D. Whinston, and Jerry R. Green. *Microeconomic Theory*. Oxford University Press, 1995.

Medstat. *MarketScan Database*. Ann Arbor,MI, 2004.

Joseph P. Newhouse. *Free for All? Lessons from the RAND Health Insurance Experiment*. Harvard University Press, Cambridge, 1993.

John A. Nyman. The economics of moral hazard revisited. *Journal of Health Economics*, 18:811–24, 1999.

Mark V. Pauly. The economics of moral hazard. *American Economic Review*, 58(3):531–37, June 1968.

Charles E. Phelps. *Health Economics, Fourth Edition.* Prentice Hall, 2010.

Charles E. Phelps and Joseph P. Newhouse. Coinsurance, the price of time, and the demand for medical services. *The Review of Economics and Statistics*, 56(3):334–342, August 1974.

Harvey Rosen. Taxes in a labor supply model with joint wage-hours determination. *Journal of Public Economics*, 11:1–23, 1979.

Emmanuel Saez. Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 2(3):180–212, 2010.

Amanda Starc. Insurer pricing and consumer welfare: Evidence from medigap. *Manuscript. Harvard University*, November 2010.

Richard Zeckhauser. Medical insurance: A case study of the tradeoff between risk spreading and appropriate incentives. *Journal of Economic Theory*, 2(1):10–26, 1970.

A Appendix: Comparison to Nonlinear Budget Set Literature

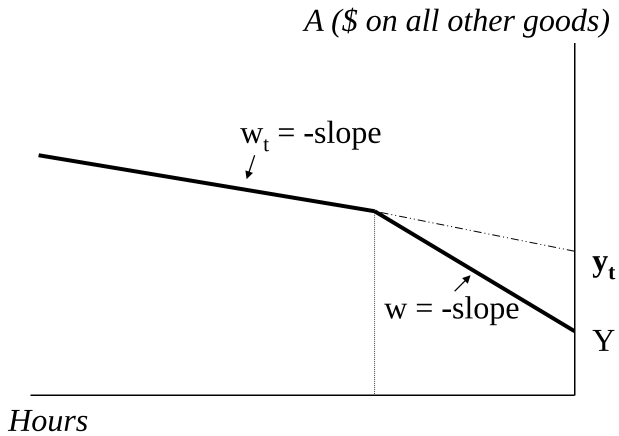
My paper builds on the nonlinear budget set literature that estimated the labor supply elasticity using nonlinear budget sets induced by progressive taxes. Hausman (1985) provides a survey of the early literature.[30] To facilitate comparison of the nonlinear budget set in my application to the nonlinear budget set in the labor supply application, Figure 8 depicts a nonlinear budget set induced by a simple progressive tax. The after-tax wage, w, that a worker faces varies with the tax rate, t.[31] The labor supply application examines the effect of the after-tax wage (the slope) on hours (the horizontal axis) controlling for income (the vertical axis). Similarly, I examine the effect the marginal price (the slope) on quantity of medical care consumed in dollars (the horizontal axis) controlling for income (the vertical axis). As is apparent from the comparison of Figure 1 to Figure 8, the budget set induced by health insurance is inherently nonconvex, but the budget set induced by progressive taxes is generally convex.

Some difficulties that are present in the labor supply application are not present in my application. For example, in the labor supply application, one important issue is that several individuals work zero hours, and the potential wage for these individuals is unknown. The medical care application does not suffer from this difficulty, however. Although several individuals do not consume

[30]Some early estimates of the labor supply elasticity using nonlinear budget set models include Hurd (1976), Rosen (1979), and Burtless and Moffit (1985). Other applications of the nonlinear budget set model include the demand for air conditioners in Hausman (1979), the disability insurance program in Halpern and Hausman (1986), the Social Security earnings test in Friedberg (2000), and 401(k) saving in Engelhardt and Kumar (2006). However, the labor supply elasticity remains the most prevalent application of the nonlinear budget set model.

[31]Comparison with Figure 1 is slightly difficult because hours are a "bad," but both figures are drawn so that the hypothetical arrow of increasing preference points to the upper right.

Figure 8: Reference Case: Nonlinear Budget Set Under Simple Progressive Tax



any medical care, the price that they would face is observable because it is determined by the insurance policy. This transparency is possible because, unlike the wage, the price does not vary at the individual level.

One advantage of the transparency of the price schedule in the budget set for medical care is that the agent and the econometrician are likely to be aware of the agent's current segment on the budget set. Liebman and Zeckhauser (2004) hypothesize that individuals respond suboptimally to complex schedules - a phenomena they call "schmeduling." While "schmeduling" may be very likely with respect to the complex tax rules addressed by the labor supply elasticity estimates, it is arguably less likely with respect to health insurance because the price schedule is so simple. In the labor supply application, since the slope of each segment varies with the underlying marginal wage, the exact segment is often unknown to the econometrician and possibly to the agent.

The transparency of the price schedule in the medical care application comes at the cost of reduced underlying variation for identification. Blomquist and Newey (2002) have developed nonparametric techniques to estimate nonlinear budget set models which have been applied by Kumar (2004) and others. These nonparametric techniques would likely have less power in this application because the slopes of the segments of the budget set do not vary across individuals. More importantly, the Blomquist and Newey (2002) approach requires that the budget set be convex.

The main contribution that I make to the class of nonlinear budget set models is that I can use my model to estimate the tradeoff between moral hazard and risk protection for a given individual, within a nonlinear health insurance plan. In the labor supply context, my model could be applied to study the tradeoff between labor supply disincentives and risk protection associated with the nonlinear subsidy structure of disability insurance or the earned income tax credit. This problem differs from the tradeoff between moral hazard and risk protection considered by Baily (1978) and generalized by Chetty (2006), because risk protection in those studies is measured by *consumption smoothing* over time instead of the generosity of a particular benefit schedule before the risk is realized.

I am aware of at least three other studies that have incorporated risk aversion into models of labor supply, but their models do not allow for measurement of the analog to the tradeoff that I consider in the medical care context. Halpern and Hausman (1986) model the decision to apply for the social security disability insurance program in the face of uncertainty about acceptance, and risk aversion informs the decision of whether or not to apply, but the authors do not consider the tradeoff between moral hazard and risk protection that the benefit schedule induces for those who are accepted. Relatedly, recent work by Low and Pistaferri (2010) considers the tradeoff between the cost of giving disability insurance to individuals who are not disabled and discouraging those who are disabled from applying, but it abstracts away from the tradeoff between moral hazard and risk protection for enrollees. Finally, Chetty (2006b) demonstrates that there is a fundamental relationship between labor supply elasticities and risk aversion, and he uses elasticity estimates from other studies to estimate risk aversion. In the health insurance context, I demonstrate a related fundamental relationship between price elasticities (moral hazard) and risk aversion, which could in turn have implications for the labor literature. In addition to developing the theoretical framework, I estimate my model using administrative data.

B Appendix: Discussion of Conditions for Integrability

Symmetry and negativity of the Slutsky matrix is necessary to recover preferences from demand. (See Mas-Collell et al. (1995)) In a partial equilibrium model, the Slutsky matrix is necessarily symmetric. From the Slutsky equation, the Slutsky matrix S is defined as

$$S = \frac{\partial Q(y_s, p_s)}{\partial p_s} + \frac{\partial Q(y_s, p_s)}{\partial y_s} Q(y_s, p_s). \tag{8}$$

In the nonlinear budget set literature, Slutsky conditions have received a great deal of attention. In the labor supply literature, the Slutsky condition can be satisfied globally if the labor supply elasticity is positive and the income elasticity is negative, but it is not automatically satisfied. MaCurdy et al. (1990) and MaCurdy (1992) brought attention to the role of Slutsky condition in the labor supply literature and proposed an alternative local linearization method to smooth around the kinks in the budget set and relax the Slutsky condition. However, Blomquist (1995) shows that even under local linearization, the Slutsky condition must be satisfied for the estimated parameters to be interpreted as labor supply parameters. He also shows that neither method automatically produces parameter estimates that satisfy the Slutsky condition. More recently, Heim and Meyer (2003) emphasize that though the MaCurdy work is valuable because it demonstrates where the Slutsky condition matters, it does not provide an alternative method.

C Simulated Minimum Distance Estimator With Multinomial Plan Choice

Let $\widehat{prob_{ij}}$ denote the predicted probability that agent i chooses plan j, obtained from a multinomial logit model that predicts plan choice using Z and indicators for all plans available in the previous year, which are populated if the agent was enrolled in the previous year. Given starting values of θ and the data matrix, which includes actual spending Q_i, the algorithm for the simulated distance estimator is as follows:

1. For each individual i of N, for each plan j of J, for each repetition r of R, draw $\eta_{ir} \sim N(\mu, \sigma^2)$. For each segment $s \in \{a, b, c\}$, predict

$$\widehat{Q_{ijrs}} = \arg\max_{Q_s} U_{ijrs}(Q_s, A_s) : p_{sj}Q_{ijrs} \leq y_{ijs}, \underline{Q_{sj}} \leq Q_{sj} \leq \overline{Q_{sj}}$$

   and the associated $U_{ijrs}\widehat{(Q_s, A_s)}$. Calculate the segment that yields the maximum utility for each i, j, r combination. Retain as $\widehat{Q_{ijr}}$.

2. Solve

$$\widehat{\theta} = \arg\min_{\theta} \sum_{i=1}^{N} \left(\min(Q_i, \psi) - \min(\sum_{r=1}^{R} \sum_{j=1}^{J} \widehat{prob_{ij}} \widehat{Q_{ijr}}, \psi) \right)^2 .$$

I set ψ to \$27,500, which is \$1,500 larger than agent plus insurer spending at the stoploss in the least generous plan. In practice, less than 1.3% of agents in my sample have spending higher than this amount.

I use $R = 5$. Because I do not have a closed form expression for demand, I estimate demand at each evaluation of the objective function using a grid with step size of \$1. I estimate this simulated minimum distance estimator on the full sample using an optimization algorithm in Matlab, and I parallelize the objective function to allow for faster computing. To obtain confidence intervals, I use subsampling instead of boostrapping for computational efficiency.[32]

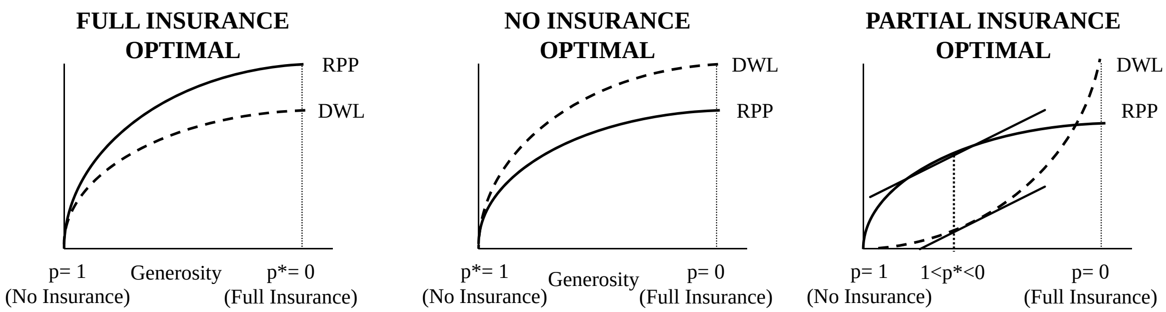
D Appendix: Implications for Optimal Insurance

Counterfactual simulations from my model allow me to consider the optimal nonlinear structure of health insurance plans, given the tradeoff between moral hazard and risk protection. If there is no moral hazard and agents are risk averse, there will be a net welfare gain from any insurance, with the highest gain for full insurance, so full insurance will be optimal. Conversely, when there is moral hazard and agents are not risk averse, there will be a net welfare loss from any insurance, with

[32] I subsample 10% of observations (10,134 observations) without replacement from the full sample and run the multinomial logit and simulated minimum distance estimation on each of 100 subsamples. Using estimates from the 186 subsamples that converge in the time allotted, I construct the empirical standard deviation of each parameter estimate. I then scale the empirical standard deviation by $\sqrt{10}$ to correct for the sample size difference between the subsamples and the full samples. I construct confidence intervals using critical values from the standard normal, such that the 95% confidence interval is equal to the point estimate plus or minus 1.96 times the scaled standard deviation.

the largest loss for full insurance, so no insurance will be optimal. In the presence of nonzero risk aversion and nonzero moral hazard, it might seem to follow that partial insurance will be optimal, but partial insurance need not be optimal.

Figure 9: Optimal Insurance



In Figure 9, I depict optimal insurance under three scenarios. All three scenarios assume that generosity can be represented as a single index. For example, consider a succession of linear plans in which the price to the consumer decreases from one to zero. In the left scenario, as generosity increases, RPP always grows at a faster rate than DWL, implying that full insurance is optimal (marginal RPP exceeds marginal DWL for every level of generosity, so we have a corner solution at full insurance). In the middle scenario, as generosity increases, DWL always grows at a faster rate than RPP, implying that zero insurance is optimal. In the third scenario, DWL and RPP grow at different rates as generosity increases, and the optimum occurs where marginal DWL is equal to marginal RPP. Partial insurance is only optimal in the case where risk protection increases at a decreasing rate and moral hazard increases at a decreasing rate as generosity increases.

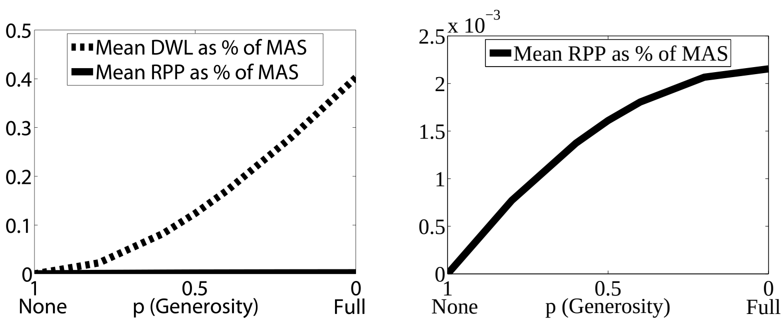
The key in recognizing that partial insurance need not be optimal is that deadweight loss and risk protection always move in the same direction, as drawn in Figure 9. To see why, conduct a thought experiment: think of a linear or nonlinear plan that will decrease moral hazard but will increase risk protection for the same individual. It is not possible to construct such a plan for a single individual. This example gives a concrete example of why it is beneficial to consider moral hazard and and risk protection jointly: in my model, they must always move in the same direction.[33]

In Figure 10, I construct the empirical analog of Figure 9 using the results from my counterfactual simulations, which vary the marginal price in plans with linear cost sharing from 1 to 2. These results are reported in Table 10. In the left subfigure, it is difficult to see RPP because it coincides with the horizontal axis, but the right subfigure graphs RPP separately on a different scale. The figure shows that in my empirical context, DWL grows at increasing rate, while risk protection grows at a decreasing rate, suggesting that partial insurance will be optimal. However, the level of

[33]This observation does not seem to have been obvious in the literature. For example, in Feldstein and Gruber (1995), the authors consider a counterfactual exercise in which they move agents into new plans. Although they always estimate a reduction in DWL, they sometimes find reductions and sometimes find increases in risk protection. Such a finding is not possible in my model, which considers both sides of the tradeoff simultaneously.

deadweight loss is so much larger than the level of risk protection that the optimal level of partial insurance will be very close to zero.

Figure 10: Estimates of Optimal Insurance with Varying Linear Price



(a) DWL and RPP (b) RPP Only (Different Scale)

This result stands in contrast to the result reported by Manning and Marquis (1996) from a similar exercise. They also find that as generosity increases, DWL increases at an increasing rate, and RPP increases at a decreasing rate, implying that partial insurance is optimal. But they find that much more generous insurance, insurance with a 45% coinsurance rate is optimal. There could be several reasons for why my result differs from theirs, including differences in modeling assumptions and differences in the underlying data.[34]

I caution against taking the implications of my results for optimal insurance too literally because, as discussed above, I do not observe any agents with full insurance or zero insurance in my data. However, the counterfactual simulations demonstrate how such an analysis could be undertaken using counterfactual simulations derived from an empirical context with a wider range of observed insurance generosities. Furthermore, another reason not to take the results literally is that restricting insurance contracts to be linear likely imposes severe restrictions on welfare gains, which is why empirically, many plans have stoplosses. The consideration of optimal insurance structures should allow for the nonlinear cost sharing schedules used in practice. My model allows for counterfactual simulations using complex nonlinear plans.

Although the simple simulations in Figure 10 show smaller welfare in more generous plans, when generosity is measured in only one dimension at a time (the coinsurance rate), this result does not hold more generally. If we increase generosity in one dimension and decrease it in another, even if we can calculate the net impact on how much the insurer will pay, we need the model to calculate the net impact on welfare. Returning to the comparison of the $1,000 deductible plan to the Feldstein plan, the Feldstein plan results in higher insurer spending for individuals with total spending below $1,000 because the insurer now pays 50% of spending before the deductible as opposed to 0%. However, for individuals with over $1,000 of spending, the Feldstein plan results in lower insurer spending because the insurer now pays 50% of spending as opposed to 80%. Whether the Feldstein plan is more or less generous on net depends on the empirical distribution of agents. As shown

[34]Manning and Marquis (1996) are limited to simple simulations in which plans only have two segments. The ability of my model to handle an unlimited number of segments allows me to apply it to more recent data and to conduct richer counterfactual plans such as the Feldstein plan.

in Table 7, the counterfactual simulation without the model suggests that the Feldstein plan is less generous than the $1,000 deductible plan, and the counterfactual simulation with the model suggests that the Feldstein plan is more generous than the $1,000 deductible plan. Despite the increase in modeled generosity from the $1,000 deductible plan to the Feldstein plan, the welfare calculations in Table 10 show an increase in average welfare from $1,000 deductible plan to the Feldstein plan. This exercise demonstrates the need for a model that considers all segments of the health insurance plans.

Knowledge of how DWL and RPP change as plan structure changes is relevant for insurance design. Though the empirical results from the simple simulation presented here suggests that very limited *linear* insurance is optimal, allowing for nonlinearities could improve the optimality of insurance. Furthermore, society might weight other factors against the net welfare gain from moral hazard and risk protection. For example, agents and society might decide to insure for other reasons such as externalities, paternalism, or behavioral factors. If these other factors are present, optimal insurance trades off the welfare implications of addressing these factors against the net welfare gain from moral hazard and risk protection calculated here.