

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

Amanda E. Kowalski[†]

April 17, 2011

Abstract

Insurance induces a well-known tradeoff between the welfare gains from risk protection and the welfare losses from moral hazard. Empirical work traditionally estimates each side of this tradeoff separately, potentially yielding mutually inconsistent results. I develop a nonlinear budget set model of health insurance that allows me to estimate both sides of this tradeoff jointly, allowing for a relationship between moral hazard and risk protection. An important feature of this model is that it considers nonlinearities in the consumer budget set that arise from deductibles, coinsurance rates, and stoplosses that lessen moral hazard as well as risk protection relative to full insurance. Within my empirical context of health insurance plans offered by a large firm, results suggest that on average, the deadweight losses from moral hazard substantially outweigh the welfare gains from risk protection. There is considerable variation in the estimated tradeoff across individuals.

*Comments are especially welcome. The following individuals provided helpful comments: Joseph Altonji, C. Lanier Benkard, Steve Berry, John Beshears, Tom Chang, Victor Chernozhukov, Jesse Edgerton, Amy Finkelstein, Michael Grossman, Jonathan Gruber, Justine Hastings, Phil Haile, Jerry Hausman, Naomi Hausman, Kate Ho, Panle Jia, Jonathan Kolstad, Kory Kroft, Fabian Lange, Whitney Newey, Chris Nosko, Matt Notowidigdo, Stephen Ryan, 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, the junior faculty and summer faculty lunches at Yale, the Cowles Summer Structural Micro Conference, Hunter College, 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. 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.

In this paper, I examine behavior in health insurance plans with nonlinear cost sharing structures to determine the net effect of moral hazard and risk protection on social welfare. 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

¹I provide a definition of moral hazard in the context of my model in Section 4.1. I also describe how my estimates of moral hazard relate to estimated “price elasticities of expenditure on medical care” from the literature.

²Feldstein 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).

to a high deductible, or a plan where a consumer pays 50% or 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. (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. 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. Based on the widespread belief that the gains from risk protection outweigh the losses from moral hazard for seniors in the doughnut hole, 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.

The welfare impact of the cost sharing structure in *private* plans is also an important policy consideration because of government policies that induce the purchase of health insurance policies with specific cost-sharing structures. For example, 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. More dramatically, 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.

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 discussed below, 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

³In 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)).

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.

Within my empirical context of plans offered by a large firm, I estimate my model. Results suggest that on average, the deadweight losses from moral hazard far outweigh the welfare gains from risk protection in the existing plans. 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.

Beyond the existing plans in my data, 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 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 my model demonstrates that this is not necessarily the case. 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 risk protection jointly: we learn that they always move in the same direction.⁴

⁴This 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

When I consider counterfactual plans with linear cost sharing structures so that I can summarize generosity with a single partial insurance rate between zero and one, or a single deductible from \$0 to \$20,000, I find that deadweight losses always increase faster than risk protection as generosity increases. This result suggests that the conventional wisdom that some level of partial insurance will achieve the optimal balance between the deadweight loss from moral hazard and the welfare gains from risk protection is misguided. As generosity increases, if the deadweight loss always grows faster than the welfare gain from risk protection, zero insurance will be optimal. Conversely, if the gain from risk protection always grows faster than the deadweight loss, full insurance will be optimal. Partial insurance will only be optimal in cases where the marginal welfare gains and welfare losses are equal as generosity increases. In either of the extreme cases, the relevant welfare question is not what level of partial insurance yields the optimal balance; the relevant welfare question is how the magnitude of the net welfare gain or loss will change as generosity changes. As a specific example, when I move agents from a high deductible plan to a similar hypothetical Feldstein plan, I find that the net welfare loss from insurance decreases.

In the next section, I motivate the use of a nonlinear budget set model of medical care, and I place my model in the context of the nonlinear budget set literature. In Sections 3 and 4, I present the model and I develop a simulated minimum distance estimator that is tied very closely to the model. In Section 5, I discuss my empirical context and data. In Section 6, I present the estimates, I perform counterfactual simulations using the estimates, and I consider the implications for optimal insurance. I conclude in Section 7.

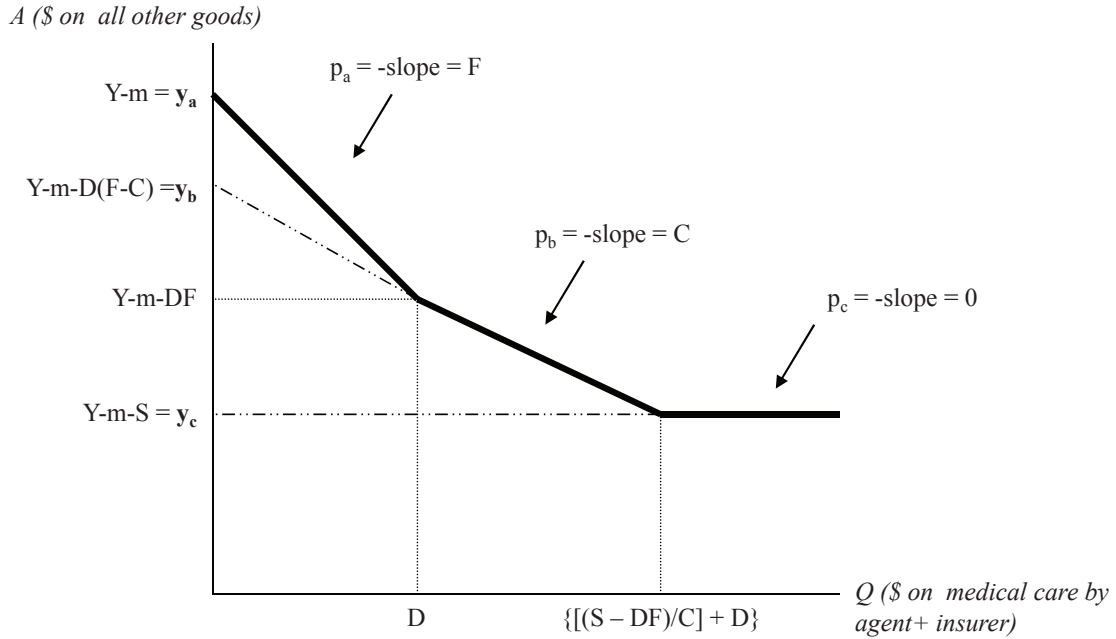
2 Nonlinear Budget Set Models

2.1 Nonlinear Budget Set from Health Insurance

A traditional health insurance plan has three basic components: a deductible, a coinsurance rate, and a stoploss. 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 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 mea-

increases in risk protection. Such a finding is not possible in my model, which considers both sides of the tradeoff simultaneously.

Figure 1: Nonlinear Budget Set Model of Health Insurance



sured in terms of dollars of expenditure on all types of medical care rather than in terms of 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

⁵In 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.

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 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,” and impose lifetime limits on plan payments. However, for many policies, including those that I study, the parameters discussed above provide a relatively complete description of plan attributes.

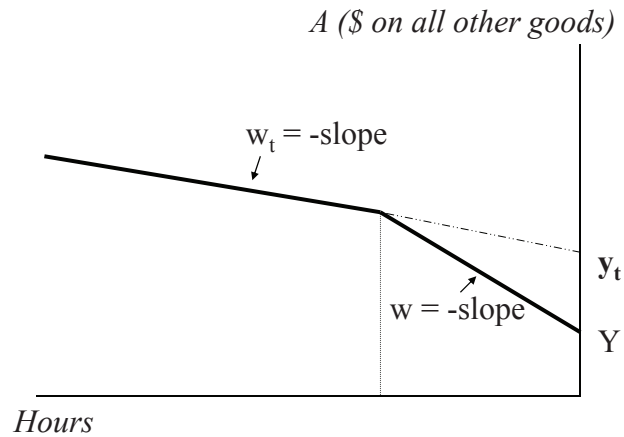
2.2 Comparison to Nonlinear Budget Set Literature

My paper builds on the original 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.⁶ To facilitate comparison of the nonlinear budget set in my application to the nonlinear budget set in the labor supply application, Figure 2 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 .⁷ 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).

⁶Some 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.

⁷Comparison 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 2: Reference Case: Nonlinear Budget Set Under Simple Progressive Tax



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 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 that 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.

As is apparent from the comparison of Figure 1 to Figure 2, the budget set induced by health insurance is inherently nonconvex, but the budget set induced by progressive taxes is generally convex. 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. Estimation is more difficult in the case with two or more nonconvex budget sets because it is no longer possible to construct an ordered likelihood function, which motivates me to develop a simulated minimum distance estimator.

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.

3 The Model

3.1 Scope of the Model

My model allows me to examine the tradeoff between moral hazard and risk protection for individuals enrolled in health insurance plans with different 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.

3.2 The Agent’s Problem

Agents make decisions in two periods, in the spirit of Cardon and Hendel (2001). In the first period, agents choose a health insurance plan from the menu of available nonlinear cost sharing options. When choosing a plan, the agents know their observable characteristics and the distribution of medical expenditure shocks that they will face. 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.⁸ In the second period, 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. At first, for simplicity of exposition, I abstract away from all heterogeneity across individuals.

In the second period, an entire plan-year, agents consume dollars of medical care, Q , and dollars of all other goods, A , and they face a nonlinear budget set of the form depicted in

⁸In this model, because agents can choose plans knowing their magnitude of moral hazard, there can be adverse “selection on moral hazard” as estimated in Einav et al. (2010c) and discussed in Karlan and Zinman (2009).

Figure 1. 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. An agent maximizes utility on each segment s of the nonlinear a budget set following the general constrained optimization problem:

$$v_s(y_s, p_s) = \max_{Q_s} U_s(Q_s, A_s) : p_s Q_s \leq y_s, \underline{Q}_s \leq Q_s \leq \overline{Q}_s$$

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 . \underline{Q}_s and \overline{Q}_s represent the lower and upper bound on Q_s imposed by each linear segment. Then, across all segments, the agent chooses the segment and corresponding Q that give the highest utility.

This modeling approach has important advantages over other approaches used to model the demand for medical care. First, it incorporates the decision to consume zero care within the model: \underline{Q}_s on the first segment is zero. Given that in most empirical settings, a large fraction of agents consume zero care, models of the demand for medical care must incorporate this feature. One traditional method to model agents who consume zero care is through the use of a two-part model.⁹ 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. 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.

Second, the structure required by this model is transparent and parsimonious. When the maximum utility occurs on the interior of a budget segment, the Q that achieves the maximum can be expressed in terms of the demand function $Q(y_s, p_s)$. From standard utility theory, Roy's Identity relates indirect utility to demand:

$$-\frac{\partial v(y_s, p_s)/\partial p}{\partial v(y_s, p_s)/\partial y} = Q(y_s, p_s).$$

Therefore, given the budget set, and conditions for integrability discussed in Appendix A, this model requires one and only one functional form for direct utility, indirect utility, or demand because the functional form for the other functions are implied by the specified functional form. Because of this I refer to the single “demand/utility” functional form that

⁹The 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. For a prominent example, see the generalization of the two part model in Manning et al. (1987).

I specify. Even though specifying demand or utility is equivalent in terms of the model, I choose to specify demand/utility starting with utility. In an earlier version of this paper (Kowalski (2008)), I began with a demand specification that allowed for a parsimonious specification of moral hazard but not of risk protection. In the following section, I describe and motivate my choice of functional form, which builds on that of Ellis (1986).¹⁰ I choose this utility function because it simple to understand where moral hazard and risk protection enter this function.

3.3 Specification of Functional Form

Given a functional form for direct utility and a budget set, an agent's demand function is fully specified.

Proposition:

On a given segment s , given the following specification of the utility function:

$$U(Q_s, A_s) = \left\{ \begin{array}{ll} -\exp(-\gamma A_s) + \frac{Q_s[\ln(Q_s/\alpha)-1]}{\ln \beta} & \text{if } (Q_s > 0 \text{ and } \alpha > 0) \\ -\exp(-\gamma y_a) & \text{otherwise} \end{array} \right\} \quad (1)$$

and the budget set:

$$A_s = y_s - p_s Q_s, \quad 0 \leq \underline{Q}_s \leq Q_s \leq \overline{Q}_s \quad (2)$$

where $\gamma, \alpha,$ and $\beta > 0$ are parameters of the model. Marshallian demand within segment s is given by:

$$Q_s = \max(\min(\alpha\beta^{\lambda p_s}, \overline{Q}_s)\underline{Q}_s) \quad (3)$$

where λ denotes the marginal utility of spending on all other goods, $\gamma \exp(-\gamma(y_s - p_s Q_s)) = \gamma \exp(-\gamma A_s)$.

Proof:

The proof is straightforward from utility maximization subject to the budget set.

This functional form has several attractive features. 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

¹⁰Relative to Ellis (1986), one main difference in my utility function is that it 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 heterogeneity across individuals, as I discuss below.

medical care.¹¹ I specify that agents have constant absolute risk aversion (CARA) preferences over spending on all other goods, A_s . We expect $\gamma > 0$ if agents are risk averse, with a larger value of γ indicating greater risk aversion. 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.¹² 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 when medical care is free ($p_s = 0$), medical care expenditure is fixed at $Q_s = \alpha$ dollars. Relative to the class of utility functions that imply infinite utility and demand when the price of one good is zero, this utility function is attractive because empirically, agents face a zero price and consume a finite amount of medical care. This specification also allows for parsimonious incorporation of unobserved heterogeneity through α , discussed in the following subsection.

When medical care is not free ($p_s > 0$), given what we expect about the ranges and signs of the parameters, this functional form implies that medical spending is less than when it is free, and spending is governed by the price, the price coefficient β , and the marginal utility of spending on all other goods λ .¹³ β acts as a price sensitivity parameter. If demand is downward sloping, β will be between 0 and 1, with a larger β indicating greater spending and less price sensitivity.¹⁴ If demand is upward sloping, $\beta > 1$. If $\beta \leq 0$, utility is undefined because $\ln(\beta)$ appears in the denominator of the utility function. In practice, I do not impose $\beta > 0$ in the estimation, but I check that it is satisfied.

A third advantage of this functional form is that it allows for consumption of zero care through a corner solution decision. The lowest possible value of medical care is $Q_a = 0$. 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 $Q \rightarrow 0$, as shown in the second line of Equation 1. When $\alpha < 0$, Equation 3 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 $(Q_s/\alpha < 0)$, we impose utility at zero in the second line of Equation 1.

One disadvantage of the functional form of the demand function is that Q_s appears on

¹¹My 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).

¹²Income appears in the demand for medical care in Equation 3 through the marginal utility of spending on all other goods, λ . A positive value of γ yields a positive λ .

¹³Given $0 < \beta < 1$, $0 < p_s < 1$, and $\lambda > 0$ (which is implied by $\gamma > 0$), Equation 3 shows ($Q < \alpha$).

¹⁴Because $\ln \beta \leq 0$ for $0 < \beta < 1$, marginal utility of spending on medical care ($\ln(Q/\alpha)/\ln \beta$) will be positive when $(Q/\alpha) < 1$.

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.

3.4 Incorporating Individual Heterogeneity

Thus far, I have not specified any variation across agents. I specify observable and unobservable heterogeneity across agents as follows:

$$\alpha_i = Z_i' \delta + \eta_i.$$

The above derivation holds with i subscripts on all components of the above model that vary with α_i , which implies the following demand function,

$$Q_{is} = \max(\min(\alpha_i \beta^{\lambda_i p_s}, \overline{Q_s}) \underline{Q_s}) \quad (4)$$

where Z_i is a vector of observable characteristics of individual i , including an indicator for male, an indicator that the employee is salaried instead of hourly, indicators for Census divisions, an indicator for whether the employee was not enrolled in the a plan in the previous year, spending in the previous year for those agents enrolled in the previous year, and indicators for family size.¹⁵ δ is the vector of coefficients on the characteristics observable to the econometrician. Z_i does not include a constant. All unobservable individual heterogeneity enters through η_i , which is distributed $N(\mu, \sigma^2)$, where μ and σ^2 are parameters to be estimated. There is nothing in the model or estimation that requires unobserved heterogeneity

¹⁵Spending from the previous year is a strong predictor of spending in the current year, motivating its inclusion in the model. 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.

to be normally distributed, but I specify the normal distribution because it is a convenient distribution that can be summarized with only two parameters.

One advantage of this specification of individual heterogeneity is that when medical care is free ($p_s = 0$), medical expenditure is equal to α_i , which is completely determined by the agent's observable characteristics Z_i , the sensitivity of medical expenditure to those characteristics δ , and the agent's realized medical shock η_i . As discussed above, when medical care is not free, the amount of spending α_i is scaled down by a function of the price sensitivity parameter β . Another advantage of this specification is that the units of the shock η_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.

This specification of individual heterogeneity provides the basis for the tradeoff between moral hazard and risk protection in the model. We formalize the welfare calculations in Section 4.1 but provide intuition here. First consider moral hazard. Individual heterogeneity provides the basis for moral hazard because individual heterogeneity is private information. Because we assume that the insurer charges the same price to all agents regardless of observable characteristics, as is common in employer-sponsored plans, the unobserved *and* observed components of individual heterogeneity ($Z_i'\delta$ and η_i , respectively) are sources of private information that can lead to moral hazard. Next consider risk protection. In the first period, the agent knows $Z_i'\delta$ and the parameters of the distribution of η_i , μ and σ^2 , but he faces risk because he does not know what his realized value of η_i will be. It is in the first period of the model, before uncertainty is realized, that the agent derives value from risk protection.

3.5 Plan Selection

The only aspect of the model that remains to be specified is plan selection in the first period. There is a growing literature on health insurance plan selection that suggests that choices might not be rational (Abaluck and Gruber (2009)) and that one reason for health insurance choices that do not minimize expenditure is that plan switching costs might be high from one year to the next (Handel (2009)). Handel (2009) finds that switching costs are so large that some agents select dominated plans. Similarly, in my empirical context, agents choose among four health insurance plans, but two appear to be completely dominated by other plans - they yield lower spending on all other goods for every amount of spending on medical care. The most internally consistent way for agents to choose plans within my model would be for them to maximize expected utility over all plans. However, informed by the literature and empirical evidence in my context, I do not model plan selection using expected utility because without specifying additional heterogeneity, agents would never choose the two dominated

plans. Instead, I allow last year’s plan choice to affect the current plan choice.

In the first period, 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. In my empirical context, all plans available this year were also available last year, making it very likely that last year’s plan is informative about this year’s plan. 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.¹⁶ 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 if the 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. In that way, last year’s plan would be related to this year’s expenditure in a way that is not captured by last year’s expenditure. If this type of violation occurs, the exclusion restriction would be violated.

One disadvantage of specifying plan choice in this way is that it has limited usefulness in predicting selection into a set of counterfactual plans. Because of this limitation, I do not consider counterfactual simulations that allow agents to sort across different plans. I only consider counterfactual simulations in which all agents are in the same plan, so there is no adverse selection.¹⁷

4 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.

¹⁶The 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.

¹⁷If 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.

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. As an input into the simulated minimum distance estimator, I first estimate the multinomial logit model described above. Formally, \widehat{prob}_{ij} denotes the predicted probability that agent i chooses plan j , obtained from a multinomial logit model that predicts plan choice using X and indicators for all plans available in the previous year, which are populated if the agent was enrolled in the previous year. 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. Solve

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

I censor the predicted and actual values of spending at ψ so that extreme values do not drive my results. 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 bootstrapping for computational efficiency.¹⁸

¹⁸I subsample 10%(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.

4.1 Calculating the Tradeoff Between Moral Hazard and Risk Protection

The techniques developed here to measure the deadweight loss from moral hazard and the gains from risk protection within the same framework are an innovation of this paper. Counterfactual simulations are integral to the welfare calculations. With true or estimated parameters of the demand/utility function, we can change the budget set, and we know what agents will spend and what their utility will be in any actual or counterfactual plan. Counterfactual simulations have more external validity if the counterfactual plans are “close” to the actual plans. I discuss the actual and counterfactual plans that I consider below. Here we consider a general plan j , and we place all agents into the same plan. Using the methodology that we describe here for calculating the tradeoff in the general plan j , we can calculate the tradeoff in any existing or counterfactual plan and compare the results across plans.

First, consider the deadweight loss from moral hazard. In this context, we 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. Assume that we observe the true values of the parameters as well as the agent’s realization of unobserved heterogeneity, η_{ir} , which we represent with the subscript r as above. In the second period, the unobserved shock r has been realized. 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. First, consider the agent’s utility gain from the lower price schedule that he faces in any linear or nonlinear plan j relative to no insurance.¹⁹ We can calculate ω_{ijr} , the dollar value of the price reductions in plan j relative to no insurance for individual i with shock r , as the dollar amount that we would need to take away from the agent to make him indifferent to moving to no insurance. We calculate ω_{ijr} quantity using the standard technique for calculating equivalent variation (EV)²⁰ as follows:

$$U(Q_{ijr}, y_{ijr} - p_{ijr}Q_{ijr} - \omega_{ijr}) = U(Q_{i,noins,r}, Y - Q_{i,noins,r}) \quad (5)$$

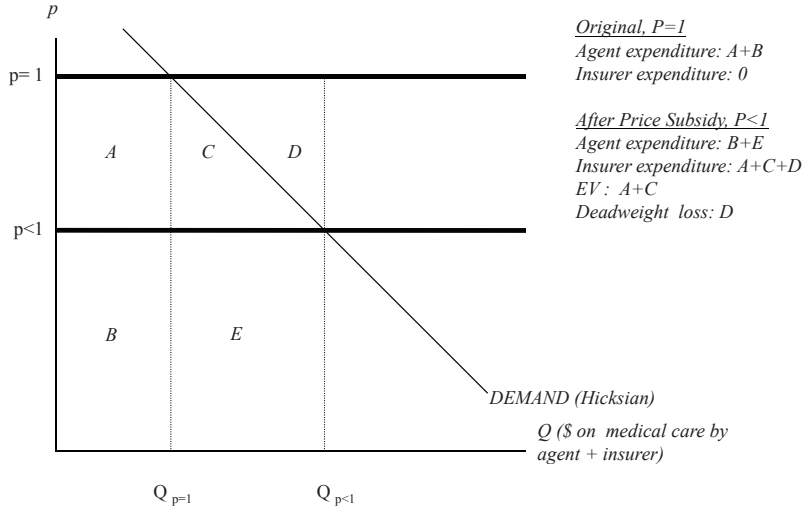
where the left side of the equation gives utility in plan j . The first argument of the utility

¹⁹In practice, I could use any plan for comparison to plan j , but I choose no insurance because it provides a useful benchmark. As stated above, 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.

²⁰This definition conforms to the definition of equivalent variation in Mas-Collell et al. (1995). As an alternative, we could base our 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).

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. For this calculation, the premium is zero.²¹ Q_{ijr} is determined by Equation 4 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. As

Figure 3: Calculation of DWL - Linear Budget Set



discussed above, moral hazard creates a wedge between the individual's value of insurance and the insurer's cost of providing insurance, creating deadweight loss. Here, we can calculate deadweight loss using standard theory with the insurer in the role that the government traditionally holds. As depicted in Figure 3, which shows a simple deadweight loss calculation using the compensated (Hicksian) demand curve for agents facing a linear budget set, the deadweight loss from moving from no subsidy ($p = 1$) to a subsidy ($p < 1$) is insurer spending ($A+C+D$) minus the equivalent 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 5, 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:

$$DWL_{ijr} = INS_{ijr} - \omega_{ijr} \quad (6)$$

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, ω_{ijr} .²² The amount of

²¹We introduce the premium into subsequent welfare calculations below.

²²Note that this deadweight loss calculation is based on Hicksian demand instead of Marshallian demand.

insurer spending is obtained by applying plan cost sharing rules to the total amount of agent plus insurer spending, Q_{ijr} . If plan j offers full insurance, $INS_{i,full,r} = Q_{i,full,r}$.

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 6, 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.

Next, we turn to measuring the welfare gain from risk protection. In the first period, the agent does not know his realized value of η_i , he only knows the distribution, $f(\eta_i)$, which is governed by the true parameters μ and σ . In this period, we can calculate the dollar value of insurance in plan j relative to no insurance for individual i , π_{ij} , as the dollar amount that we would need to take away from agent i under plan j to make him indifferent between plan j and no insurance. We calculate this quantity using the standard technique for calculating a risk premium as follows:

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

where the left side of the equation gives expected utility over all possible values of η_i , in plan j , where utility is determined for each realization r as it is on the left side of Equation 5. The right side of the equation gives expected utility under no insurance for all possible values of η_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 η_i . We calculate RPP_{ij} , the risk protection premium for individual i under plan i , as follows:

$$RPP_{ij} = \pi_{ij} - \int (\omega_{ijr})f(\eta_i)d\eta_i$$

To examine the tradeoff between risk protection and moral hazard, we calculate the *expected* deadweight loss for agent i in the first period²³ as follows:

This differs from the deadweight loss calculation of Feldstein and Gruber (1995), who use Marshallian demand for simplicity.

²³The expected tradeoff in the first period is the relevant tradeoff for welfare analysis because there is no tradeoff in the second period; there are no gains to risk protection once uncertainty has been realized.

$$DWL_{ij} = \int (INS_{ijr} - \omega_{ijr})f(\eta_i)d\eta_i$$

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

$$RPP_{ij} - DWL_{ij} = \pi_{ij} - \int (INS_{ijr})f(\eta_i)d\eta_i$$

Thus far, we have considered the net social benefit of health insurance for a single individual. The tradeoff will vary across individuals because individuals differ in their observable characteristics. Even though the parameters of the distribution of unobserved heterogeneity do not differ across individuals, the different observable characteristics interact with shocks from the same distribution to produce welfare consequences that vary across individuals. 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. 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:

$$\overline{RPP}_j - \overline{DWL}_j = \frac{1}{N} \sum_{i=1}^N \pi_{ij} - \frac{1}{N} \sum_{i=1}^N \int (INS_{ijr})f(\eta_i)d\eta_i \quad (7a)$$

$$= \frac{1}{N} \sum_{i=1}^N \pi_{ij} - m_j/\zeta \quad (7b)$$

where \overline{RPP}_j and \overline{DWL}_j denote the mean risk protection premium and deadweight loss, respectively, in plan j . Equation 7b 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.²⁴

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:²⁵

²⁴Therefore, we do not need to know the loading calculate the tradeoff. However, we do model the loading for estimation, as discussed in Section 5.1.

²⁵We 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.

$$\overline{MAS} = \frac{1}{N} \sum_{i=1}^N \int (Q_{i,noins,r}) f(\eta_i) d\eta_i$$

When we do not observe the true values of the parameters, it is straightforward to estimate these welfare quantities using estimated values of the parameters and R simulated draws from the estimated distribution of unobserved heterogeneity $\widehat{\eta}_{ir} \sim N(\widehat{\mu}, \widehat{\sigma}^2)$ as follows:

$$\begin{aligned} \widehat{DWL}_{ij} &= \frac{1}{R} \sum_{r=1}^R (\widehat{INS}_{ijr} - \widehat{\omega}_{ijr}) \\ \widehat{RPP}_{ij} &= \widehat{\pi}_{ij} - \frac{1}{R} \sum_{r=1}^R \widehat{\omega}_{ijr} \end{aligned}$$

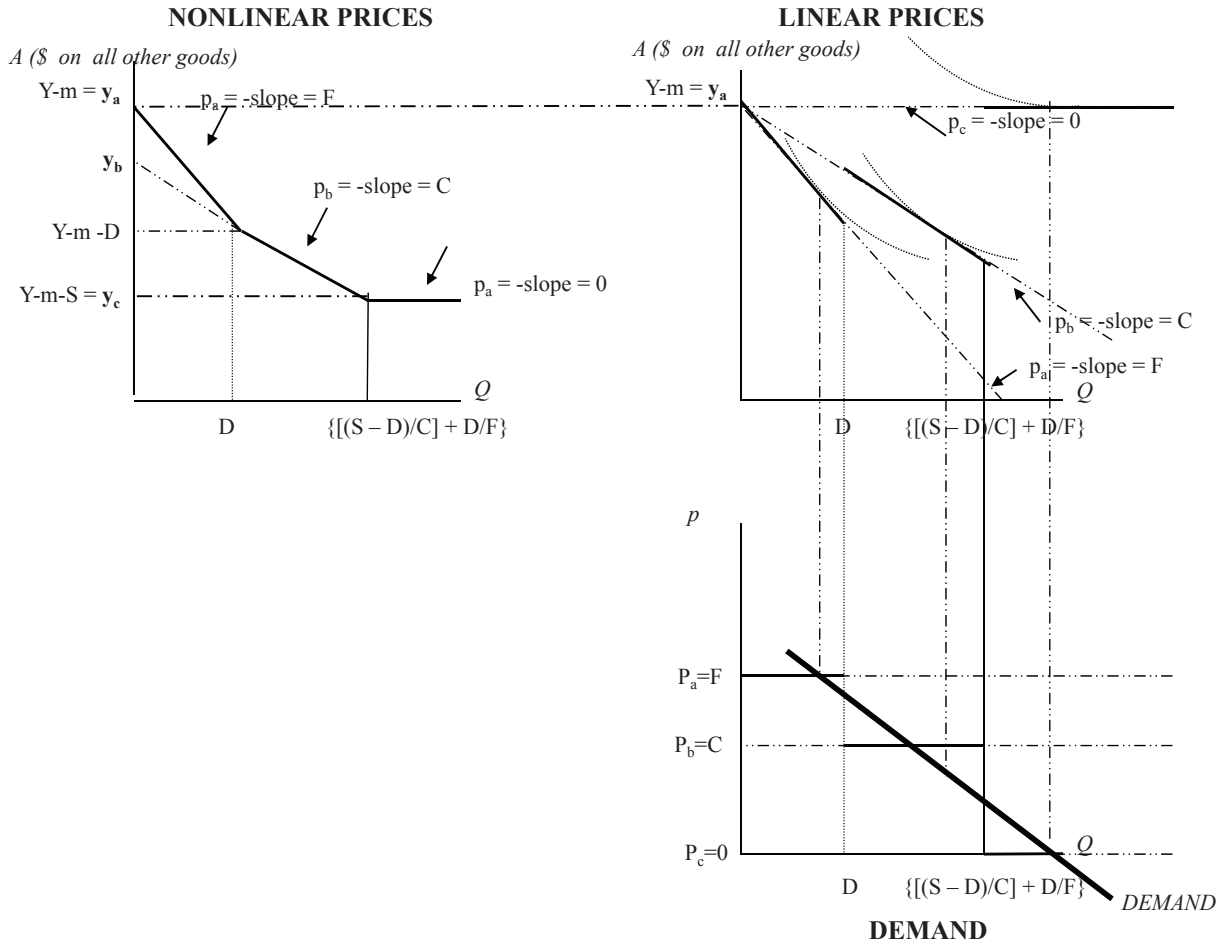
Solving for $\widehat{\pi}_{ij}$ and $\widehat{\omega}_{ijr}$ involves a fixed point problem because, for example, the starting value of $\widehat{\pi}_{ij}$ implies a value \widehat{Q}_{ir} of that in turn implies a new starting value of $\widehat{\pi}_{ij}$. We solve this problem numerically.

4.2 Identification

Identification of the tradeoff between moral hazard and risk protection comes partially from agent choices of plan and spending conditional on plan, and partially from functional form. One functional form that provides identification is the functional form of the budget set. Although identification by functional form is generally undesirable, particularly if it is inaccurate, in this context, the functional form of the budget set is likely to be accurate, as described above. In general, additional identification comes from the functional form of demand/utility and from observable heterogeneity across agents in covariates in expenditure. We consider identification of the welfare loss from moral hazard and the welfare gain from risk protection in turn.

In the second period, after plan choice has been made, identification of moral hazard is similar to identification of the labor supply elasticity in the nonlinear budget set context. To aid in understanding how the parameters are identified, I developed Figure 4. The upper left of Figure 4 shows the agent's problem of utility maximization subject to a nonlinear budget set. The nonlinear budget set has three segments, with three different marginal prices. In the upper right of Figure 4, I have respecified the agent's problem, holding virtual income constant as I do in the estimation. In this context, agents choose expenditure to maximize utility subject to three different prices, and now it is standard to 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 here because there is a utility maximum on the interior of each bolded

Figure 4: Graphical Depiction of Identification



segment.

In the first period, identification of risk protection comes from the same variation used to identify moral hazard as well as from the choice of plan. One key innovation of my approach is that I demonstrate that any demand/utility function used to examine moral hazard in this framework also implies a value of risk protection. 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.

5 Empirical Context

5.1 Data

My empirical context focuses on individuals employed by a firm in the retail trade industry that insures over 500,000 employees plus their enrolled family members. 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. Restricting analysis to a single firm limits 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. (2010c), and Handel (2009). However, as in the literature, my contribution is not limited to my empirical findings.

My data are proprietary data on plan structure, claims, and enrollment compiled by Medstat (2004). I model demand in 2004 using information on demand and plan choice in 2003. 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 selected 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

Table 1: Plan Characteristics

Plans		Fraction			
		Deductible <i>F</i>	Deductible <i>D</i>	Coinsurance <i>C</i>	Stoploss <i>S</i>
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
	\$1,000 Deductible/Stoploss	1	1,000	NA	1,000
	\$5,000 Deductible/Stoploss	1	5,000	NA	5,000
	\$10,000 Deductible/Stoploss	1	10,000	NA	10,000
	\$20,000 Deductible/Stoploss	1	20,000	NA	20,000

of the four plans offered at the firm that I study. The firm offered only these plans in 2003 and 2004. I model demand in 2004 using information on demand and plan choice in 2003. 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 structures. Several other firms in the data allowed individuals to choose between plans with nonlinear cost sharing structures and Health Maintenance Organization (HMO) plans without nonlinear cost sharing structures. Because the firm that I study offers only plans of the type that I can model, 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). Although my model could be used to model demand at other firms with nonlinear cost sharing structures, I restrict my analysis to a single firm to better control for unobservable characteristics across agents.

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 ($[(S - DF)/C] + D$). The generosity of these plans spans the range of plans typically offered in the market at the time of the data.²⁶

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 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.

In another paper, I use the interaction between the individual and family deductibles to aid in identification of the price elasticity of expenditure on medical care using data from the same firm (Kowalski (2009)). While the interaction is useful for identification in that context, it creates potentially severe measurement error in the budget set in this context. 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

²⁶A 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.

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 a large improvement in external validity because it allows me to consider members of 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.

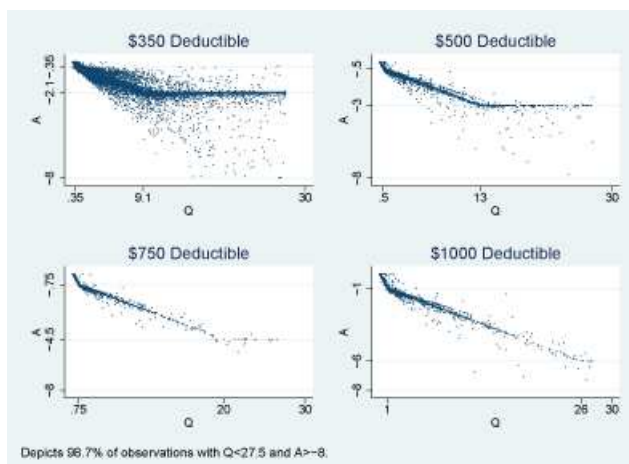
Another complication that arises when I apply the nonlinear budget set model to my empirical context is that the plans offered by this firm are preferred provider organization (PPO) plans that offer incentives for beneficiaries to go to providers that are part of a network. These plans are very common: 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.²⁷ 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.

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 $\zeta = 1.25$.²⁸ In Figure 6, I use this measure of the premium for each plan to shift each budget

²⁷Out-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.

²⁸This loading factor is motivated by Handel (2009) and Phelps (2010), page 350. 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.

Figure 5: Empirical Budget Set by Plan



set downward.²⁹ As shown, the \$350 deductible plan and the \$750 deductible plans are completely dominated by the other plans.³⁰ The advantage of modeling the premium is that I can also predict the premium for counterfactual plans. I also do 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.³¹ 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. Also, as is common in other claims studies, I do not observe anything about unenrolled family members or employees.

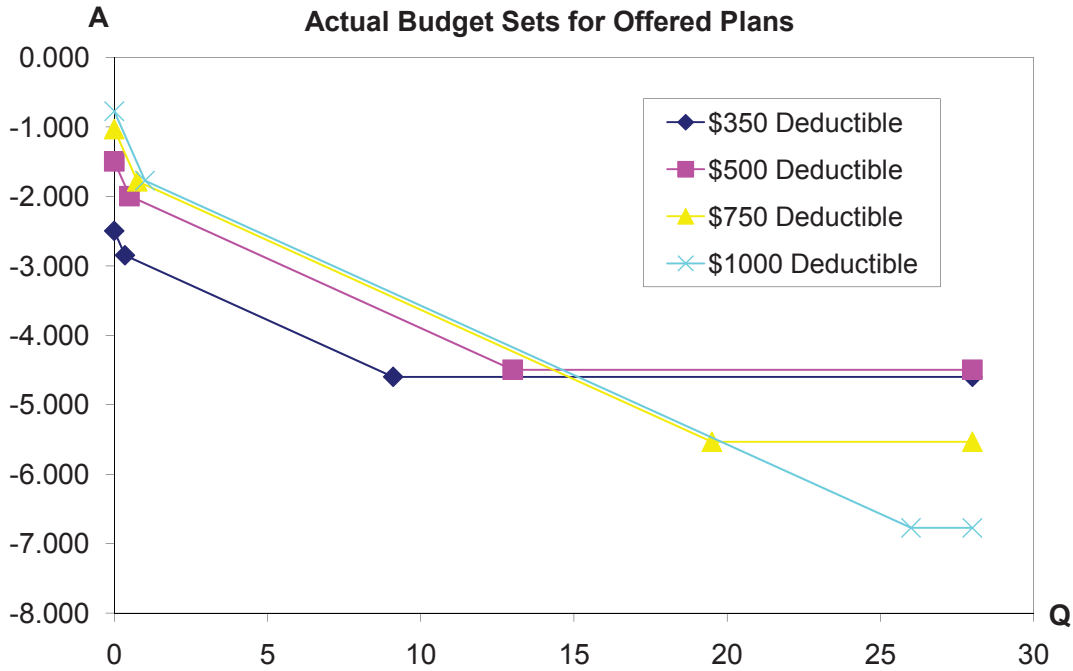
My selected sample consists of 101,343 employees enrolled in 2004. The selected sample reflects about one fifth of the employees that I ever observe in the 2004 data. I describe sample selection in detail in the Data Appendix, available upon request. In general, I lose approximately 30% of observations because I require everyone in the family to be contin-

²⁹This figure abstracts away from income variation because income shifts all plans up and down by the same amount.

³⁰In the region where the \$500 deductible budget set dominates the \$350 deductible budget set, the \$500 deductible budget set is \$100 higher. Even though the premia depicted are likely to be calculated with error, it is difficult to find a menu of premia in which two plans are not completely dominated.

³¹I 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.

Figure 6: Actual Budget Sets for Offered Plans



uously 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.³²

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. Average yearly spending by the beneficiary and insurer is \$2,335. Average spending decreases as plan generosity decreases, providing evidence of either moral hazard or adverse selection. I formalize the reduced form evidence on moral hazard and adverse selection in Appendix B. Average spending does not necessarily provide an accurate depiction of spending because spending is very skewed. We examine skewness in spending in more depth in Section 6.2. 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

³²The 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

Full Sample	All Plans	By Deductible			
		\$350	\$500	\$750	\$1,000
Total Spending/1,000	2.335	2.637	1.779	1.412	1.147
Consumer Spending/1,000	0.619	0.639	0.582	0.586	0.529
Insurer Spending/1,000	1.716	1.998	1.197	0.826	0.618
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
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
Salary	0.077	0.072	0.101	0.089	0.087
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
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
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
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
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 2003	0.053	0.004	0.011	0.017	0.475
In Family of 2	0.189	0.170	0.240	0.247	0.244
In Family of 3	0.085	0.070	0.119	0.130	0.131
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 2003	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.

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%). 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.

6 Results

6.1 Estimation Results

Table 3 reports the estimated coefficients from the full sample of 101,343 observations. 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 provide more evidence on the magnitude of the differences between demographic groups in the counterfactual simulations reported in Tables 7 and 8. 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 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, $Q_I - Q_{II}$, and compute a midpoint arc elasticity as follows:

$$arc = \frac{Q_I - Q_{II}}{Q_I + Q_{II}} \div \frac{p_I - p_{II}}{p_I + p_{II}} \quad (8)$$

With this calculation, I obtain an arc elasticity in my data of -0.0015, which is much

Table 3: Estimated Coefficients

Interpretation	Parameter	Simulated Minimum Distance		
		Estimate	95% confidence	
Mean of unobserved heterogeneity	mu	-1.0005 ***	-1.2174	-0.7836
Male	delta1	-0.5568 ***	-0.6088	-0.5048
Salary/1,000	delta2	-0.1129 ***	-0.1819	-0.0438
Census Division 2 - Middle Atlantic	delta3	-0.1290 **	-0.2575	-0.0004
Census Division 3 - East North Central	delta4	0.4612 ***	0.3576	0.5648
Census Division 4 - West North Central	delta5	0.2246 ***	0.1184	0.3308
Census Division 5 - South Atlantic	delta6	0.2912 ***	0.2019	0.3806
Census Division 6 - East South Central	delta7	0.2277 ***	0.1327	0.3227
Census Division 7 - West South Central	delta8	0.2511 ***	0.1616	0.3405
Census Division 8 - Mountain	delta9	0.0389 **	0.0068	0.0710
Census Division 9 - Pacific	delta10	-0.0456	-0.1018	0.0107
Age	delta11	0.1049 ***	0.0943	0.1155
Age Squared/100	delta12	-0.2102 ***	-0.2350	-0.1854
Age Cubed/1,000	delta13	0.2066 ***	0.1781	0.2351
Missing 2003	delta14	0.7034 ***	0.6422	0.7645
2003 Spending*Nonmissing 2003	delta15	0.3661 ***	0.3436	0.3886
2003 Spending*Nonmissing 2003 Squared/1,000	delta16	-3.0281 ***	-3.6727	-2.3835
2003 Spending*Nonmissing 2003 Cubed/1,000,000	delta17	1.4863	-1.1997	4.1722
In Family of 2	delta18	0.0873 **	0.0145	0.1602
In Family of 3	delta19	-0.0551 **	-0.0991	-0.0110
Standard deviation of unobserved heterogeneity	sigma	0.0371 **	0.0080	0.0662
Coefficient of Absolute Risk Aversion	gamma	0.0769 **	0.0157	0.1380
Price parameter	beta	0.3319 ***	0.1431	0.5207
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.

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. 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.

6.2 Model Fit

Table 4 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 the taking 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. (forthcoming), which examine “bunching” around kinks, except that my empirical context implies dispersion around the kinks. However, the distribution of predicted spending follows directly from

Table 4: Model Fit

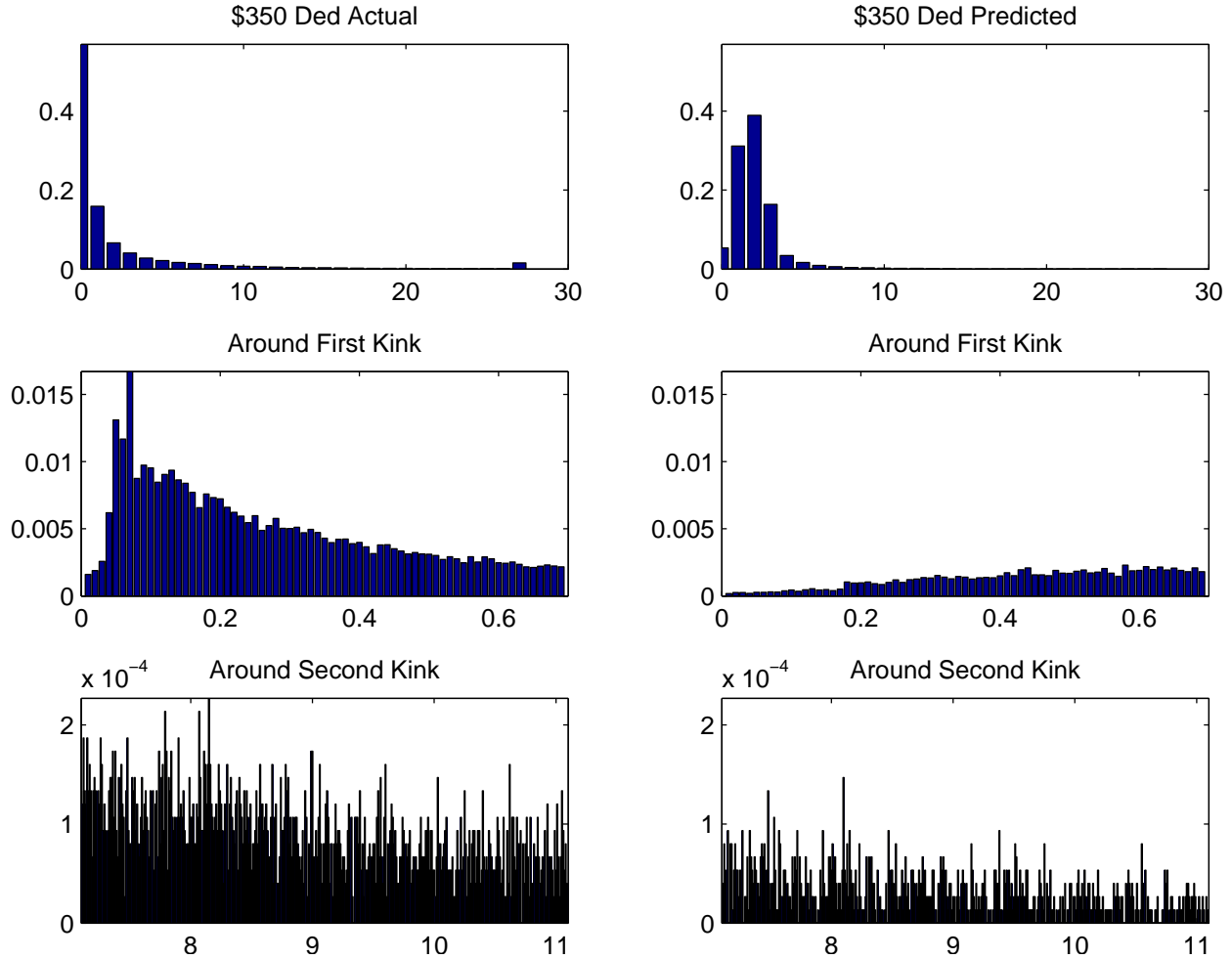
Regression of Actual Spending on Mean Predicted Spending Over All Draws

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

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

Figure 7: Actual and Predicted Spending in Thousands



my model, so I do not need to employ techniques to examine excess 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 4 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. It seems reasonable that the lack of dispersion in the actual distribution results in a small estimated price elasticity, which leads to a lack of predicted dispersion.

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 is much smaller than the largest actual value.³³ Furthermore, though the model predicts some expenditures on all segments, it overestimates the distribution in the middle and underestimates it at the extremes.

6.3 Counterfactual Simulations

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. As discussed above, I want to shut down adverse selection in my simulations, so in each simulation, I place all agents in a single plan. I then compare the results across plans.

In the bottom half of Table 1, I present the characteristics of all sixteen of the plans that I consider in turn. I consider the four offered plans as well as twelve hypothetical plans. 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 “0% Frac” in the tables. The full insurance and no insurance cases provide useful benchmarks, but I caution against taking their predictions literally because I do not observe anyone in the data with no insurance or full insurance. The other linear plans provide simple benchmarks that isolate the effect of each marginal price that I model in the nonlinear plans. Finally, the last four hypothetical plans have full cost sharing until a deductible of \$1,000, \$5,000, \$10,000, or \$20,000 and then have zero cost sharing. I label these plans as “Deductible/Stoploss” plans in the tables because the deductible and the stoploss are equal for these plans. These simple plans provide benchmarks that isolate the effect of varying the deductible.

Before using the counterfactual simulations from the model to predict expenditure and welfare in each counterfactual plan, to demonstrate the value of my model, I 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 5 presents total spending, spending by the insurer, and

³³The largest predicted value of spending in the data is approximately \$14,500, which is below the data censoring at \$27,500.

spending by the agent, in each plan.³⁴ In the more generous plans, the insurer spends more. In these results, the \$1,000 deductible plan appears to be more generous (insurer spending is higher) than the Feldstein plan given the distribution of total expenditure observed in the data. We will use these results as a baseline for comparison to the spending predictions from the model, which allow for moral hazard.

In the bottom half of Table 5, I present the results from the counterfactual simulations from the model.³⁵ Because these predictions allow for moral hazard, total agent plus insurer spending varies across plans. 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. This descriptive comparison of generosity does not allow us to make statements about consumer welfare across plans because consumers must pay for extra generosity, and this simple exercise does not tell us how much they value extra generosity.

In Table 6, I present results that move beyond analysis of spending to analysis of welfare. As calculated using the techniques discussed in Section 4.1, The first panel of Table 6 shows the distribution of DWL, the second panel shows the distribution of RPP, and the third panel shows the distribution of the tradeoff. Note that the distribution of the tradeoff at any quantile 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. In the last column, I divide all means by the money at stake measure (MAS) described above as total spending on under no insurance. Here, MAS=\$1,943, as shown in the bottom panel of Figure 5. 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. However, there is 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, and the bottom 1% of agents have a net loss from insurance that is ten times larger than the loss for the

³⁴For these simulations, I censor expenditure for each agent at \$27,500 for comparison to counterfactual simulations from the model.

³⁵For 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 4.

Table 5: Counterfactual Simulation Results: Spending

	Agent + Insurer Q_{ij} Mean	Insurer INS_{ij} Mean	Agent $INS_{ij}-Q_{ij}$ Mean
<i>Counterfactual Without Model*</i>			
<i>Offered</i>			
\$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
<i>Hypothetical</i>			
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
\$1,000 Deductible/Stoploss	1,963.20	1,536.89	426.31
\$5,000 Deductible/Stoploss	1,963.20	836.90	1,126.30
\$10,000 Deductible/Stoploss	1,963.20	451.37	1,511.83
\$20,000 Deductible/Stoploss	1,963.20	124.19	1,839.01
<i>Counterfactual Using Model</i>			
<i>Offered</i>			
\$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
<i>Hypothetical</i>			
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
\$1,000 Deductible/Stoploss	1,957.90	1,030.00	927.90
\$5,000 Deductible/Stoploss	1,946.00	84.20	1,861.80
\$10,000 Deductible/Stoploss	1,944.10	9.90	1,934.20
\$20,000 Deductible/Stoploss	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).

Table 6: Counterfactual Simulation Results: Welfare Across Distribution

DWL_{ij}	Quantiles										Mean as % of MAS
	Min	1	5	25	50	75	95	99	Max	Mean	
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
\$1,000 Deductible/Stoploss	0.00	0.00	0.00	0.58	3.78	8.88	25.33	59.61	600.82	7.39	0.380
\$5,000 Deductible/Stoploss	0.00	0.00	0.00	0.00	0.00	0.00	0.00	46.90	600.82	1.44	0.074
\$10,000 Deductible/Stoploss	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	600.82	0.49	0.025
\$20,000 Deductible/Stoploss	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.000
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
\$1,000 Deductible/Stoploss	0.00	0.00	0.00	0.01	0.03	0.05	0.09	0.13	0.29	0.03	0.002
\$5,000 Deductible/Stoploss	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.22	0.00	0.000
\$10,000 Deductible/Stoploss	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.28	0.00	0.000
\$20,000 Deductible/Stoploss	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
\$1,000 Deductible/Stoploss	-600.81	-59.59	-25.29	-8.85	-3.74	-0.53	0.00	0.03	0.24	-7.36	-0.379
\$5,000 Deductible/Stoploss	-600.81	-46.84	0.00	0.00	0.00	0.00	0.00	0.00	0.17	-1.44	-0.074
\$10,000 Deductible/Stoploss	-600.81	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.49	-0.025
\$20,000 Deductible/Stoploss	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.

RPP-DWL for each quantile is generally not equal to the quantile of RPP-DWL.

individuals at the mean.

Table 6 also shows welfare in the hypothetical plans in the lower rows of each panel. The average net welfare loss in the Feldstein plan is smaller than the welfare loss of the comparable \$1,000 deductible offered plan. We consider the net welfare loss in the remaining eleven hypothetical plans in Section 6.4. Here, we note that as in the offered plans, welfare in the hypothetical plans differs considerably across agents.

Tables 7 and 8 further investigate heterogeneity across agents by presenting average DWL and RPP for individuals in distinct demographic groups. The first column of Table 7 reproduces the mean results from the previous table. The second and third columns compare welfare for men and women. The results show that in all offered and hypothetical plans, though risk protection has a similar magnitude for women and men, the deadweight loss is generally twice as large for women. Overall, the net benefit of insurance is lower for women than it is for men because of the larger deadweight loss. Successive columns in tables 7 and 8 also show a wide amount of dispersion in welfare across income quartiles, ages, family sizes, and census regions. Dispersion in the tradeoff results mostly from dispersion in moral hazard (DWL) and less from dispersion in risk protection (RPP). This result follows from the model because only unobserved heterogeneity affects risk protection, but observed and unobserved heterogeneity affect moral hazard. From Tables 7 and 8, we can see that observed heterogeneity goes a long way in explaining the variation across the valuation quartiles shown in Table 6.

6.4 Implications for Optimal Insurance and Comparison to Trade-off Calculations in the Literature

In the literature, there is a strong sentiment that the optimal insurance policy should be somewhere between full and partial insurance.³⁶ However, I show that partial insurance need not be optimal. Indeed, my results suggest that zero insurance is optimal. If either zero or full insurance is optimal, the relevant welfare question is not what level of partial insurance yields the optimal balance; the relevant welfare question is how the magnitude of the net welfare gain or loss will change as generosity changes.

In Figure 8, 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,

³⁶For example, Feldstein (2006) states, “In principle, the optimal policy might involve a combination of deductibles and different coinsurance rates for different ranges of spending.”

Table 7: Counterfactual Simulation Results: Welfare By Covariates I

<i>DWL_{ij}</i>	Mean By												
	Mean	Gender		Mean By Type		Mean By Income Quartile				Mean By Age			
		Male	Fem.	Salary	Hourly	(Low)	1	2	3	(High)	4	Age< med	Age> med
<i>Offered</i>													
\$350 Deductible	5.52	3.58	6.68	4.08	5.64	12.43	5.70	2.96	0.80	3.32	7.84		
\$500 Deductible	5.36	3.42	6.52	3.96	5.48	12.07	5.54	2.88	0.78	3.23	7.61		
\$750 Deductible	5.23	3.21	6.44	3.80	5.36	11.78	5.41	2.81	0.76	3.07	7.52		
\$1,000 Deductible	5.04	2.90	6.32	3.53	5.17	11.34	5.22	2.70	0.74	2.75	7.45		
<i>Hypothetical</i>													
50% Frac to \$2,000 Deduct	4.35	2.40	5.51	2.81	4.48	9.77	4.51	2.34	0.63	2.03	6.80		
0% Frac (Full Insurance)	7.82	5.16	9.41	5.95	7.98	17.64	8.05	4.18	1.14	4.93	10.87		
20% Frac	5.44	3.55	6.57	4.08	5.56	12.26	5.62	2.92	0.79	3.35	7.65		
40% Frac	3.33	2.15	4.03	2.46	3.40	7.47	3.44	1.79	0.49	2.00	4.73		
50% Frac	2.41	1.55	2.92	1.76	2.46	5.40	2.49	1.30	0.35	1.43	3.44		
60% Frac	1.60	1.03	1.95	1.17	1.64	3.59	1.66	0.87	0.23	0.94	2.30		
80% Frac	0.43	0.28	0.53	0.31	0.44	0.97	0.45	0.23	0.06	0.25	0.63		
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		
\$1,000 Deductible/Stoploss	7.39	4.26	9.26	5.26	7.57	16.66	7.63	3.95	1.08	4.13	10.83		
\$5,000 Deductible/Stoploss	1.44	0.91	1.76	0.79	1.50	3.18	1.54	0.78	0.21	0.48	2.45		
\$10,000 Deductible/Stoploss	0.49	0.36	0.57	0.17	0.51	1.07	0.53	0.27	0.07	0.11	0.89		
\$20,000 Deductible/Stoploss	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
<i>RPP_{ij}</i>													
<i>Offered</i>													
\$350 Deductible	0.04	0.04	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	0.04	0.04
\$750 Deductible	0.04	0.03	0.04	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.04	0.04
\$1,000 Deductible	0.03	0.02	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.04
<i>Hypothetical</i>													
50% Frac to \$2,000 Deduct	0.03	0.03	0.04	0.03	0.04	0.03	0.03	0.04	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	0.03	0.04	0.04	0.04	0.04	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	0.01	0.02	0.02	0.02	0.02	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		
\$1,000 Deductible/Stoploss	0.03	0.02	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03		
\$5,000 Deductible/Stoploss	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
\$10,000 Deductible/Stoploss	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
\$20,000 Deductible/Stoploss	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
<i>Offered</i>													
\$350 Deductible	-5.48	-3.55	-6.63	-4.04	-5.60	-12.39	-5.66	-2.92	-0.76	-3.29	-7.80		
\$500 Deductible	-5.32	-3.38	-6.48	-3.93	-5.44	-12.03	-5.50	-2.84	-0.74	-3.19	-7.57		
\$750 Deductible	-5.20	-3.18	-6.40	-3.77	-5.32	-11.75	-5.37	-2.77	-0.73	-3.03	-7.48		
\$1,000 Deductible	-5.01	-2.87	-6.28	-3.51	-5.13	-11.31	-5.18	-2.67	-0.70	-2.73	-7.41		
<i>Hypothetical</i>													
50% Frac to \$2,000 Deduct	-4.32	-2.37	-5.48	-2.77	-4.45	-9.74	-4.48	-2.31	-0.60	-2.00	-6.76		
0% Frac (Full Insurance)	-7.78	-5.12	-9.36	-5.91	-7.94	-17.60	-8.01	-4.14	-1.09	-4.89	-10.82		
20% Frac	-5.40	-3.51	-6.53	-4.04	-5.52	-12.22	-5.58	-2.88	-0.75	-3.31	-7.61		
40% Frac	-3.29	-2.11	-3.99	-2.42	-3.36	-7.44	-3.41	-1.75	-0.45	-1.96	-4.69		
50% Frac	-2.37	-1.52	-2.89	-1.73	-2.43	-5.37	-2.46	-1.26	-0.32	-1.40	-3.41		
60% Frac	-1.58	-1.00	-1.92	-1.14	-1.61	-3.57	-1.64	-0.84	-0.21	-0.91	-2.27		
80% Frac	-0.42	-0.26	-0.51	-0.30	-0.43	-0.95	-0.44	-0.22	-0.05	-0.23	-0.61		
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		
\$1,000 Deductible/Stoploss	-7.36	-4.24	-9.22	-5.23	-7.54	-16.62	-7.60	-3.92	-1.04	-4.11	-10.79		
\$5,000 Deductible/Stoploss	-1.44	-0.91	-1.76	-0.78	-1.49	-3.18	-1.54	-0.78	-0.21	-0.48	-2.45		
\$10,000 Deductible/Stoploss	-0.49	-0.36	-0.57	-0.17	-0.51	-1.07	-0.53	-0.27	-0.07	-0.11	-0.89		
\$20,000 Deductible/Stoploss	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.

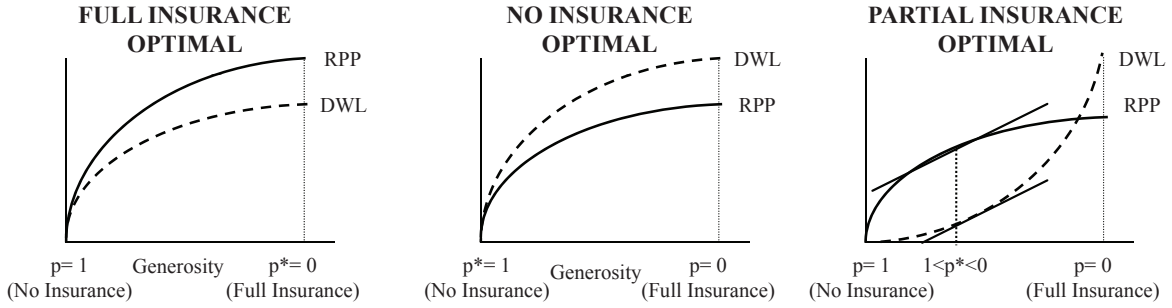
Table 8: Counterfactual Simulation Results: Welfare By Covariates II

DWL_{ij}	<u>Mean By Family Size</u>			<u>Mean By Census Region</u>								
	1	2	3	New Eng.	Mid Atlant.	East North Cent.	West North Cent.	South Atlant.	East South Cent.	West South Cent.	Mount.	Pacific
Offered												
\$350 Deductible	5.45	6.30	4.35	4.70	3.72	6.55	5.47	5.71	5.15	5.54	4.43	5.11
\$500 Deductible	5.29	6.15	4.25	4.56	3.56	6.38	5.31	5.55	5.05	5.37	4.26	4.88
\$750 Deductible	5.15	6.06	4.12	4.29	3.35	6.24	5.19	5.44	4.93	5.25	4.10	4.71
\$1,000 Deductible	4.94	5.94	3.89	4.04	3.11	6.10	4.96	5.26	4.73	5.04	3.90	4.53
Hypothetical												
50% Frac to \$2,000 Deduct	4.29	5.20	2.99	3.42	2.60	5.39	4.33	4.52	4.06	4.33	3.32	3.72
0% Frac (Full Insurance)	7.71	8.92	6.32	6.60	5.40	9.19	7.73	8.09	7.44	7.83	6.24	7.09
20% Frac	5.37	6.22	4.33	4.60	3.70	6.44	5.39	5.63	5.14	5.46	4.34	4.95
40% Frac	3.28	3.81	2.61	2.81	2.23	3.96	3.30	3.44	3.12	3.34	2.65	3.03
50% Frac	2.38	2.75	1.88	2.03	1.60	2.87	2.39	2.48	2.24	2.41	1.92	2.20
60% Frac	1.58	1.84	1.24	1.35	1.06	1.92	1.59	1.65	1.49	1.61	1.28	1.47
80% Frac	0.43	0.50	0.33	0.37	0.28	0.52	0.43	0.45	0.40	0.44	0.35	0.40
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	0.00
\$1,000 Deductible/Stoploss	7.24	8.71	5.79	5.90	4.61	8.90	7.26	7.72	6.99	7.38	5.70	6.61
\$5,000 Deductible/Stoploss	1.47	1.64	0.78	1.34	0.69	1.98	1.50	1.41	1.00	1.53	1.24	1.51
\$10,000 Deductible/Stoploss	0.52	0.48	0.23	0.61	0.34	0.62	0.56	0.46	0.24	0.52	0.58	0.69
\$20,000 Deductible/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
RPP_{ij}												
Offered												
\$350 Deductible	0.04	0.04	0.04	0.04	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04
\$500 Deductible	0.04	0.04	0.04	0.04	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04
\$750 Deductible	0.04	0.04	0.04	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.03	0.03
\$1,000 Deductible	0.03	0.04	0.03	0.03	0.02	0.04	0.03	0.03	0.03	0.03	0.03	0.03
Hypothetical												
50% Frac to \$2,000 Deduct	0.03	0.04	0.03	0.03	0.03	0.04	0.04	0.04	0.03	0.03	0.03	0.03
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	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	0.04
40% Frac	0.04	0.04	0.03	0.04	0.03	0.04	0.04	0.04	0.04	0.04	0.03	0.03
50% Frac	0.03	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	0.03
80% Frac	0.01	0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.02	0.01	0.01	0.01
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	0.00
\$1,000 Deductible/Stoploss	0.03	0.04	0.03	0.03	0.02	0.04	0.03	0.04	0.03	0.03	0.03	0.03
\$5,000 Deductible/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
\$10,000 Deductible/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
\$20,000 Deductible/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
Offered												
\$350 Deductible	-5.42	-6.26	-4.31	-4.66	-3.68	-6.51	-5.43	-5.67	-5.11	-5.50	-4.39	-5.07
\$500 Deductible	-5.25	-6.11	-4.21	-4.53	-3.52	-6.34	-5.27	-5.51	-5.01	-5.33	-4.23	-4.84
\$750 Deductible	-5.12	-6.02	-4.08	-4.25	-3.32	-6.20	-5.15	-5.41	-4.90	-5.22	-4.07	-4.68
\$1,000 Deductible	-4.91	-5.90	-3.86	-4.01	-3.09	-6.07	-4.93	-5.22	-4.70	-5.01	-3.87	-4.50
Hypothetical												
50% Frac to \$2,000 Deduct	-4.25	-5.16	-2.96	-3.39	-2.57	-5.35	-4.29	-4.49	-4.02	-4.29	-3.28	-3.68
0% Frac (Full Insurance)	-7.67	-8.88	-6.27	-6.55	-5.36	-9.14	-7.69	-8.05	-7.40	-7.79	-6.20	-7.05
20% Frac	-5.33	-6.18	-4.29	-4.56	-3.66	-6.40	-5.35	-5.59	-5.10	-5.42	-4.30	-4.91
40% Frac	-3.25	-3.77	-2.57	-2.77	-2.20	-3.92	-3.26	-3.40	-3.08	-3.30	-2.62	-3.00
50% Frac	-2.35	-2.72	-1.84	-2.00	-1.57	-2.84	-2.36	-2.45	-2.21	-2.38	-1.89	-2.17
60% Frac	-1.56	-1.81	-1.21	-1.33	-1.04	-1.89	-1.57	-1.63	-1.46	-1.58	-1.25	-1.44
80% Frac	-0.41	-0.48	-0.32	-0.35	-0.27	-0.51	-0.42	-0.43	-0.38	-0.42	-0.33	-0.38
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	0.00
\$1,000 Deductible/Stoploss	-7.20	-8.67	-5.76	-5.87	-4.59	-8.86	-7.23	-7.68	-6.96	-7.35	-5.67	-6.58
\$5,000 Deductible/Stoploss	-1.46	-1.64	-0.78	-1.34	-0.69	-1.98	-1.50	-1.41	-1.00	-1.53	-1.24	-1.51
\$10,000 Deductible/Stoploss	-0.52	-0.48	-0.23	-0.61	-0.34	-0.62	-0.56	-0.46	-0.24	-0.52	-0.58	-0.69
\$20,000 Deductible/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

Values in dollars.

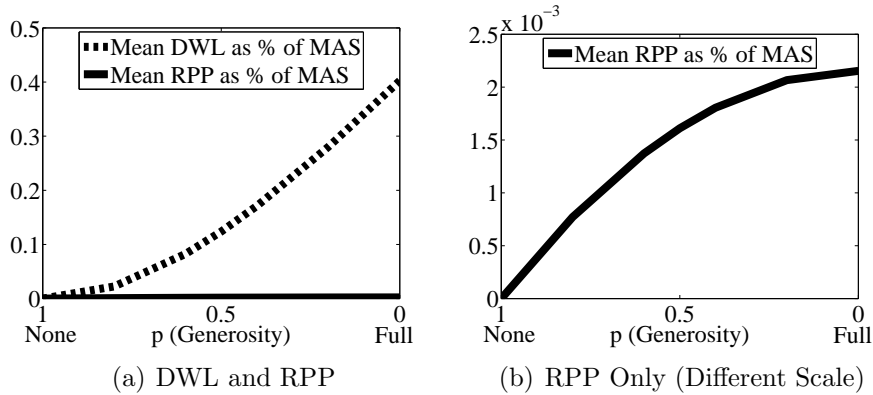
Median age is 43. Income first quartile: \$30,208; median: \$37,222; third quartile: \$49,113.

Figure 8: Optimal Insurance



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.

Figure 9: Estimates of Optimal Insurance with Varying Linear Price

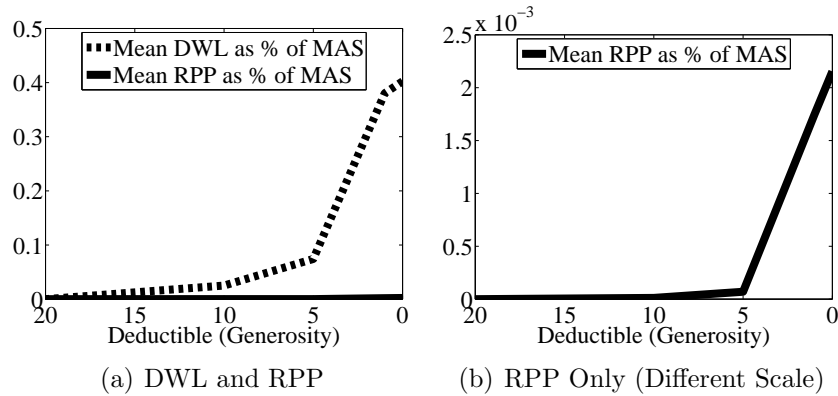


In Figure 9, I construct the same figure using the results from my counterfactual simulations, which vary the marginal price in plans with linear cost sharing from 0 to 1. The figure shows that in my empirical context, DWL always grows faster than RPP, implying that zero insurance is optimal.³⁷ This result stands in contrast to the result reported by Manning and Marquis (1996) from a similar exercise. They find that as generosity increases, DWL increases at an increasing rate, and RPP increases at a decreasing rate (as in the third subfigure in Figure 8), implying that partial 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.³⁸

³⁷In 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.

³⁸Manning and Marquis (1996) are limited to simple simulations in which plans only have two segments.

Figure 10: Estimates of Optimal Insurance with Varying Deductible



In Figure 10, I construct a related figure using simple simulations that vary the deductible from zero (given by the full insurance in the table) to \$20,000. I conduct this simulation because several theoretical papers, including Holmstrom (1980), suggest that plans with a deductible have better properties than linear plans. However, this simulation produces results that are similar to the previous simulation - DWL is much larger than RPP for more generous plans, implying that zero insurance (a very high deductible) is optimal.³⁹ Manning and Marquis (1996) find similar results when they vary the deductible.

Although the simple simulations in Figures 9 and 10 show smaller welfare in more generous plans, when generosity is measured in one dimension at a time (the coinsurance rate or the deductible), 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 we showed in Table 5, 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

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.

³⁹Although my deductible simulation results are similar to my coinsurance simulation results, it is interesting to note that in the deductible simulation, though the DWL increases monotonically with generosity, the path is not globally concave or convex.

to the Feldstein plan, the welfare calculations in Table 6 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 policy. Though the empirical results from the two sets of simple simulations presented here suggest that no insurance is optimal, 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. Thus, the findings provided here inform the optimal amount of insurance.

7 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. 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 losses from moral hazard outweigh the average welfare gains from risk protection. I find considerable variation

in the net loss across agents, suggesting that optimal insurance differs across individuals.

A 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.

$$S = \frac{\partial Q(y_s, p_s)}{\partial p_s} + \frac{\partial Q(y_s, p_s)}{\partial y_s} Q(y_s, p_s) \quad (9)$$

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.

B Appendix: Reduced form Evidence of Moral Hazard and Adverse Selection

Reduced form evidence on moral hazard is limited in this application because it does not quantify the magnitudes, and it does not address the associated welfare costs. I present this reduced form evidence to motivate the presence of moral hazard that I estimate in my model and to highlight the importance of my model in assessing magnitudes. The first reduced form evidence that I consider is the “bivariate probit” or “positive correlation” test proposed by Chiappori and Salanie (2000). This approach 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.

Table B1: Positive Correlation Test

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

Variable	Spending		
	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

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 B1, I find a statistically significant negative coefficient, which indicates that there is a positive correlation between generosity and claims. The magnitude of the coefficient is not important for this test. The coefficient indicates the presence of moral hazard and/or adverse selection.

The second reduced form evidence that I consider is 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”), that are correlated with coverage generosity as well as realized insurable spending, there is evidence of adverse selection. In my context, the premium does not vary with several characteristics that I observe, as is common in employer-sponsored health insurance plans. 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 B2. 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. Again, this test does not say anything about the magnitude of adverse selection.

Taken together, this reduced form evidence suggests that there is adverse selection and perhaps moral hazard in my empirical context. Given adverse selection, my decision to model plan choice in the estimation instead of taking it as exogenous seems merited. After taking adverse selection into account in the estimation, since I am interested in focusing on

Table B2: Unused Observables Test

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

Variable	Dependent Variable: Spending					
	Separate Regressions			Single Regression		
	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

Variable	Dependent Variable: Deductible					
	Separate Regressions			Single Regression		
	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

the tradeoff between moral hazard and risk protection, it will be important to shut down adverse selection in my counterfactual simulations.

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.
- 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*, forthcoming.

- Pierre Andre Chiappori and Bernard Salanie. Testing for asymmetric information in insurance markets. *The Journal of Political 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 Political Economy*, 80(4):623–48, July/August 1972.
- Matthew J. Eichner. Medical expenditures and major risk health insurance. *Massachusetts 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. *Working Paper*, September 2010c.
- 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.

- Bengt Holmstrom. Moral hazard and observability. *Bell journal of economics*, 10(1):74–91, October 1980.
- 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. Scheduling. *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 Paarsch. 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 elasticities 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.