A Semiparametric Analysis of Adverse Selection and Moral Hazard in Health Insurance Contracts *

Patrick Bajari, Department of Economics, University of Michigan and NBER
Han Hong, Department of Economics, Duke University
Ahmed Khwaja, Fuqua School of Business, Duke University

JEL Classification: C14, D82, I11
Keywords: Adverse Selection, moral hazard, health insurance, semiparametric estimation

Preliminary and Incomplete

Abstract

The last decade has been characterized by unprecedented inflation in health care expenditures. One commonly suggested source for this inflation is the distortion in the health insurance market due to asymmetric information about the latent health status of individuals that leads to adverse selection in health insurance contracts and run away medical expenditures due to moral hazard. In this paper, we investigate this claim by estimating a model of demand for health insurance and medical care in the presence of adverse selection and moral hazard under minimal parametric assumptions. We propose a computationally simple estimation strategy to recover the structural parameters of the model using a two step semiparametric approach. In particular we are able to non-parametrically recover the distribution of the latent health status. This allows us to formulate a non-parametric test for the presence of moral hazard and adverse selection, and we find evidence of both.

* We acknowledge excellent research assistance from Alvin Murphy and Ivan Shaliastovich. All remaining errors are our own.

I. Introduction.

Over the past decade, expenditures on health care have increased at an alarming rate, e.g., in 2003 the CPI rose by 1.9% while medical care inflation was 3.9%. Since health expenditures are now approximately 15% of the G.D.P. it is important for policy to analyze the underlying reasons for their rapid growth. Concurrently there has been a large increase in the number of individuals lacking insurance in the U.S. One commonly suggested source for this inflation is the distortion in the health insurance market due to asymmetric information that leads to adverse selection (see e.g., Akerlof 1970, Spence 1973, Rothschild and Stiglitz 1976, Wilson 1977, 1980) and moral hazard (see, e.g., Arrow 1963, Pauly 1968).[1] Most Americans do not directly pay for all of their health care costs. Instead, a large fraction of these costs are borne by employers or by public insurance programs such as Medicare or Medicaid. Insurance reimbursements, whether from private or public funds, make up more than 80% of medical expenditures in the U. S. (Cutler and Zeckhauser 2000). Since health care consumers do not typically bear all of these costs, their incentives to economize on costs is limited and an important reason for inflation in health care expenses. The situation is exacerbated by the large proportion of uninsured individuals who place an immense burden on the health care system, especially the resources for acute and emergency care.

In this paper, we investigate the distortionary effects of asymmetric information about the latent health status of individuals that leads to adverse selection in health insurance contracts and run away medical expenditures due to moral hazard (see e.g., Pauly 1974). We formulate a model of demand for health insurance and medical care in the presence of adverse selection and moral hazard. We propose a computationally

[1] See Cutler and Zeckhauser 2000 and Pauly 2000 provide an extensive review of the more recent literature.

simple estimation strategy to recover the structural parameters of the model using a two step semiparametric approach. In particular we are able to non-parametrically recover the distribution of the latent health status. Using our estimates, we non-parametrically test for the presence of moral hazard and adverse selection.

Our theoretical model builds on the theoretical work of Spence and Zeckhauser (1971) and Blomqvist (1997). We formulate a model of demand for medical care in the presence of both moral hazard and adverse selection. More precisely we make the following contributions in this paper. (1) We develop a model with *minimal* parametric assumptions that allows for the effects of both adverse selection and moral hazard in the simultaneous determination of the demand for health Insurance and medical utilization. (2) We develop a semiparametric procedure for the identification and estimation of the model that does not rely on ad hoc statistical assumptions about the selection process in to the different insurance categories. (3) We non-parametrically test for adverse selection in to various health plans. (4) We quantify the level of, and non-parametrically test for moral hazard between alternative health plans.

There has been a large empirical literature on the effects of health insurance on medical utilization (see Zweifel and Manning 2000 for an excellent review). The gold standard of this work is considered to be the RAND HIE of the late 1970's (Manning et al 1987, Newhouse 1993). Even though the RAND HIE was a randomized control trial, the data collected was limited in that it excluded the elderly, high income individuals and was limited to six sites in the U.S. More importantly, since the RAND HIE was conducted in the mid 1970s the U.S. health care system has undergone many changes so revisiting these questions with a more current data set is of great importance.

There is a growing body of empirical work on the structural estimation of models of medical utilization (e.g., Cameron et al. 1988, Gilleskie 1998, Blau and Gilleskie 2003). There is some research on the structural estimation of models of health insurance choices (e.g. Harris and Keane 1999). There is also a small literature that has estimated structural models of the joint decision to purchase health insurance and consume medical care. Cardon and Hendel (2001) do this in a static framework and while Khwaja (2002, 2005) does this in a dynamic life-cycle framework. This literature has relied exclusively on parametric methods to estimate these models e.g., maximum likelihood estimation. While this is an important strain of research, the empirical approach in these papers abstracts from the simultaneous examination of adverse selection and moral hazard. This has the advantage of making the estimators more tractable; however, it comes at the cost of abstracting away from potentially important jointly occurring market distortions. Moreover most of the empirical literature on this subject has relied on various parametric statistical assumptions in the analysis that may not hold in the data, e.g., Vera-Hernandez (2003). Hence a semiparametric analysis will shed light on the effects of various statistical assumptions.

There is very little empirical research on the effects of adverse selection in health insurance markets on medical utilization and expenditures. An exception is Cardon and Hendel (2001) who find no evidence of adverse selection in health insurance markets which is in contrast with our finding of evidence for adverse selection. Most of the existing, empirical literature however is descriptive in nature and has focused on testing for the presence of adverse selection and moral hazard. Very few empirical papers attempt to jointly estimate the demand for health insurance and medical utilization in the presence of adverse selection and/or moral hazard. Furthermore, most papers that attempt to estimate demand make the econometrically strong assumption that prices (and health premiums) are exogenous (e.g., Cardon and Hendel 2001), that is, these are uncorrelated with unobservables (such as private information about health status). This assumption is extremely heroic and implausible since the theoretical literature suggests that the choice of insurance and medical care is primarily a function of such unobservables. While this literature has made important contributions there are still limitations to the approaches used. First, because of the difficultly of the problems studied by these researchers, very strict parametric assumptions are used to estimate the models. While this simplifies the implementation of the estimator, it is well known that misspecification of parametric forms can lead to large biases in many contexts. Second, none of these papers is simultaneously able to deal with adverse selection, moral hazard and price endogeneity. Typically they attempt to make progress on at most one of these issues. The approach that we use is in the spirit of the empirical auctions literature that allows researchers to quantify the distortions from asymmetric information under minimal parametric assumptions.[2] We draw on this recent methodology developed in the auctions literature to develop a semiparametric two step estimator to recover the structural parameters of the model. We also modify the estimators to account for the endogeneity of the insurance contract and medical utilization decision of the individual as discussed below.

The rest of the paper is organized as follows. Section II discusses the data. The model is presented in section III with the identification and estimation strategy discussed in section IV. Results follow in section V, with examination of adverse selection and moral hazard in section VI. Section VII concludes.

[2] See e.g., Laffont and Vuong (1996), Guerre, Perrigne and Vuong (2002), Li, Perrigne and Vuong (2000), Campo, Perrigne and Vuong (2002), Bajari and Hortascu (2002, 2003), Athey and Haile (2003), and Haile, Hong and Shum (2003).

II. Data

The HRS is a nationally representative sample of men and women born between 1931 and 1941 and their spouses or partners, who could be of any age (Juster and Suzman 1995). As of wave 1 (in 1992), 12,652 people from 7,607 households were asked detailed questions about labor force participation, income and wealth, family structure, and health and health-relevant behaviors. The study also included detailed questions about respondents' health, health insurance characteristics, medical utilization, health care expenditures and employer characteristics. We also have data from HRS on out of pocket medical expenditures. The HRS oversampled blacks, Hispanics, and residents of Florida. We use data on individuals from wave 3 (1996) because this is the most recent wave for which we have out of pocket expenditure data. The publicly available data is supplemented by confidential data on location of residence to help create instrumental variables used in our analysis. The variables used in our analysis are described in appendix 1 with summary statistics in table 1. For additional detail, see Juster and Suzman (1995).

Our estimation sample was created in the following way. We started with 10,030 observations from wave 3 (1996) of the HRS. (1) Observations where either the household income (hhinc3), household insurance premium (hhincst3), or household out of pocket medical expenditure (hhoop3) were missing were dropped. This reduced sample size to 4645, a drop of approximately 54%. (2) Observations where household out of pocket medical expenditure (hhoop3) exceeded household total medical expenditure (hhmedex3) were dropped. This reduced sample size to 4540, a drop of approximately 2% of remaining observations. (3) Observations where household income (hhinc3) was less than the sum of household insurance premium (hhincst3), and household out of pocket medical expenditure (hhoop3) were dropped. This reduced sample size to 4412, a drop of approximately 3% of remaining observations. (4) Observations where the insurance category was either VA/Champus or Medicaid were dropped. This reduced the sample size to 4020, a drop of approximately 9% of remaining observations. (5) Outliers with excessive household total medical expenditure (hhmedex3) were dropped. Outliers were observations with values of household total medical expenditure (hhmedex3) below the second percentile or above the ninety eighth percentile. This reduced the sample size to 3849, a drop of approximately 4% of remaining observations. (6) Observations where the estimated derivative of out of pocket medical expenditure (hhoop3) with respect to household total medical expenditure (hhmedex3), i.e., z'(m), was negative were dropped as these violated an assumption of the model. This reduced the sample size to 3735, a drop of approximately 3% of remaining observations.

III. Model

III.A. Overview.

We specify a model of endogenous consumer demand for health insurance and medical utilization that allows for unobserved (to the econometrician) heterogeneity in the latent distribution of health status. It is not clear what the economic objective of insurers often is, i.e., whether insurers maximize profits (e.g., for profit insurance plans) or pooling of risk at the expense of profits (e.g., Medicare) or something in-between (e.g., employer provided insurance). Thus given its complex nature we do not model the objective function of the insurer. Consequently we are agnostic about the insurer's knowledge of the distribution of unobserved health status of the individuals. However in our empirical analysis we allow for the endogoeneity of the insurance contract with regard to the unobserved health status distribution. Thus our analysis is robust to adverse selection in to health insurance plans or alternatively the lack of adverse selection if insurers design insurance contracts contingent on their knowledge about the latent health distribution of the enrollees. The model will also incorporate moral hazard in medical utilization conditional on insurance status.

III.B. Structure

III.B.1. The Timing Structure.

Individuals in the model are assumed to make decisions about the purchase of health insurance and medical care in a staggered fashion. Following Khwaja (2001, 2005) there are two time periods in the model. Given a menu of insurance options, individuals in the first time period make a choice about an insurance plan. In the second time period, conditional on the first period insurance choice and a realization from the distribution of latent health status, the individuals make a choice about medical utilization.

III.B.2. The First Time Period: Insurance Choice.

The individuals face a menu of choices from a set of available insurance plans (*D)* that depends on whether they are younger or older than 65. For individuals younger than 65 the insurance choice set (*D)* includes the following choices: employer provided health insurance, self-employed with insurance through own business, privately purchased insurance, Medicare with or without private Medigap insurance to supplement Medicare and no insurance. For individuals older than 65 the insurance choice set (*D)* is the same as for those younger than 65 with two exceptions. It excludes the option to purchase private insurance and no one in this age group can be uninsured.

Each individual has an expectation about their distribution of latent health status. Given their age, each individual chooses the plan $d \in D$ that maximizes his or her utility from the expected choice about medical utilization conditional on the expected realization from the distribution of health status in the second time period. This may be represented as follows, $\max_{d \in D}[V_1, \ldots, V_D]$, where V_d is the expected utility associated with the optimal medical treatment choice in the second time period given the distribution of health status. While our model allows for adverse selection based on the unobserved distribution of health status, we do not explicitly model the functional form for V_d and therefore the process through which individuals self-select in to insurance plans.[3]

III.B.3. The Second Time Period: Medical Utilization Choice.

Conditional on the insurance choice in the first time period and having obtained a realization θ from the distribution of health status each individual makes a decision about medical utilization. The individuals choose the level of medical utilization that maximizes their utility from health status and the consumption of a composite good. The utility function and the budget constraint are described below.

III.B.4. The Utility Function.

Following Spence and Zeckhauser (1971) and Bloomqvist (1997) we specify the consumer's utility function, $U(c, m-\theta; \gamma)$, to depend on the level of composite good consumption c, the amount of medical utilization m, the consumer's unobserved health status θ, and the parameters γ that characterize the utility function of the consumer. As in Cardon and Hendel (2001), medical utilization is assumed to be a perfect substitute for the lack of good health status. This is a restrictive assumption but as found previously by Cardon and Hendel it captures the essential features of the data well. Also this allows for the preventive aspects of medical care, i.e., people incur medical expenditures even in

---

[3] It should be noted that the choice of health insurance is closely related to the employment decision of the individual. Our model does not include an employment decision in the interest of simplicity and due to the computational burden this would place on the estimator that we develop. An extensive review of the relationship between availability of health insurance coverage and the labor market decisions of individuals is provided by Currie and Madrian (1999) and Gruber (2000).

good health for preventive purposes. Both m and θ are assumed to be expressed in terms of monetary units. Therefore only the difference between m and θ directly enters the utility function. In particular we specify the utility function to be separable in c and m-θ and take the form,

$$U(c, m-\theta; \gamma) = \frac{1}{1-\gamma_1} c^{1-\gamma_1} + \frac{\gamma_2}{1-\gamma_3}(m-\theta)^{1-\gamma_3}\,. \tag{1}$$

The utility function allows for risk aversion in wealth (see e.g. Hubbard, Skinner and Zeldes 1995, Gertler and Gruber 2002) through the parameters on the composite consumption good as well as health status (see e.g., Khwaja 2001, 2005) through the parameters on the difference between medical utilization and latent health status. This helps to account for the decision to insure due to risk in two dimensions, i.e., financial and health capital. The budget constraint faced by the consumer is

$$c = y - p - z(m)\,,$$

where y is an exogenously given level of income, p denotes the premium or the fixed cost of participating in the health insurance policy, and $z(m)$ denotes the out of pocket expenditure of the individual. Alternatively, the reimbursement scheme used by the insurance plan is that \$(m – z) will be reimbursed to the consumer if the consumer incurs \$$m$ in medical expenses. We assume that the insurer specifies the reimbursement schedule (m - $z(m)$) prior to the realization of an individual's health shocks θ. The fact that the reimbursement schedule is only a function of the medical utilization m and does not directly depend on the health status realization θ creates a moral hazard problem. After the realization of the health status θ, the consumer chooses the level of health services m as a function of θ, $m(\theta)$, to maximize his or her utility. We further assume that *m* is a non-decreasing function of θ. Whenever $m(\theta)$ is strictly positive the following first order condition holds,

$$U_m(c, m-\theta;\gamma) - U_c(c, m-\theta;\gamma)\left[z'(m)\right] = 0\,. \qquad (2)$$

Equation (1) provides the condition for the optimal choice of medical utilization that maximizes an individual's utility conditional on the insurance status. This is the standard marginal rate of substitution (MRS) rule for allocating income between the composite commodity and medical care. It states that the ratio of marginal utilities from medical care and the composite commodity should be equal to their respective prices. The economic intuition underlying this is that conditional on the insurance contract an individual equates the marginal benefit of medical care measured in terms of improvement in utility to the marginal cost of medical utilization in terms of out of pocket expenditures using the composite commodity as the numeraire good. This optimality condition relates the unobserved health status θ to the observable medical utilization of the consumer *m*. Under the previous specification, this optimality condition becomes

$$\gamma_2(m-\theta)^{-\gamma_3} = c^{-\gamma_1}\left(z'(m)\right). \qquad (3)$$

We use this condition as a basis for developing our identification and estimation strategy.

IV. The Identification and Estimation Strategy.

We propose a semiparametric estimator for the parameters in the utility function, γ. The major advantage of this strategy is that we do not have to rely on parametric assumptions about the distributions in estimating the parameters of the model. The key insight of our identification strategy, is that both the coinsurance rate $z'(m)$ (and in turn the reimbursement schedule (m - $z(m)$), and the distribution of the health status θ can be non-parametrically identified using the optimality condition (equation 3) about the level of medical utilization. These non-parametric estimates can in turn be used to estimate the risk aversion parameters, γ_1 and γ_3. Therefore, our identification strategy depends only on the correct specification of the consumer utility function and on the validity of the economic hypothesis of utility maximization but not on the statistical hypotheses regarding the reimbursement schedule and the distribution of health status θ.

Our identification strategy makes use of instruments that provide exogenous variation in the characteristics of the health insurance plans. The specific instruments we use are (i) number of establishments in a county, (ii) the state level housing price index, and (iii) the county level malpractice insurance component of the Geographic Practice Cost Index (GPCI) developed by the Medicare Payment Advisory Commission to reimburse medical practices who treat Medicare beneficiaries (see appendix for more details. We assume that these instruments reflect county level variation in the costs of providing insurance and are independent of the distribution of unobserved health status of the consumers. Alternatively put, the identification assumption is that the patient's perceived latent health distribution is independent of characteristics of their residential location which we use as our instruments. We also tried other instruments (e.g., the number of employees in the firm that employs an individual, whether an individual is self-employed, local unemployment rate, average local income, HMO penetration) in preliminary work but without much success so these were not used in the final estimation.

The HRS data also offers another source of identification. In particular, identification is also achieved off the individuals who are eligible for Medicare coverage. Since (almost) everyone over the age of 65 is eligible for Medicare, there will not be adverse selection in Medicare. Hence the distribution of the unobservable health status of the Medicare beneficiaries (controlling for age and other observables) should be representative of the underlying population distribution of unobserved health status unconditional on health plan choices. Thus Medicare eligibility provides exogenous variation in insurance status that will aid in the estimation of the distribution of latent health status.

Our identification and estimation method proceeds in two steps. In the first step, we non-parametrically estimate the health insurance reimbursement schedules from medical utilization and insurance reimbursement data. This is estimated separately for each of the $d \in D$ plans in the data. The identification assumption is that insurance plans may design reimbursement policies given their expectation about the distribution of latent health status (in anticipation of adverse selection) but once individuals opt for a particular plan the reimbursement schedule cannot further discriminate against particular individuals. In other words the reimbursement schedule cannot be made individual specific though it may be group specific. We use the Nadaraya-Watson kernel estimator to recover the reimbursement schedule conditional on the insurance choice non-parametrically from the data.

Given data on $i = 1, \ldots, n$ individuals and their associated choices about insurance plans d_i, the amount of insurance reimbursement z_i and the level of medical utilization m_i., the Nadaraya-Watson kernel estimator for $z(m;d)$ for insurance plan d is given by

$$\hat{z}(m;d) = \left. \sum_{i=1}^{n} z_i k\left(\frac{m_i - m}{b_n}\right) 1(d_i = d) \middle/ \sum_{i=1}^{n} k\left(\frac{m_i - m}{b_n}\right) 1(d_i = d) \right., \qquad (4)$$

where b_n is a sequence of bandwidth parameters (see e.g., Hardle 1990 for details). This specification captures any non-linearity in the reimbursement schedule e.g., deductibles, maximum annual out of pocket limit, maximum life-time reimbursement limit (see e.g., Keeler, Newhouse and Phelps 1977 for the complications that arise when such non-linearities are ignored).

In the second step, we use the instrument variables (denoted by x) to identify and estimate both the utility function parameters γ and the distribution of the unobserved health status θ. The key identifying assumption is the independence between the instruments (x) and the health status shocks (θ). Even though the health status shocks (θ) are unobservable they can be uniquely recovered from the observable medical utilization by inverting the optimality condition given by equation 3. In particular, consider an insurance plan d. Given any initial value for the utility function parameters γ, for each individual $i = 1,\ldots,n$, the unobservable health status θ_i can be recovered from the observed medical utilization m_i using the consumer optimality condition (equation 3), where $z(m;d)$ is replaced by the estimated function $\hat{z}(m;d)$ from equation 4. Equation 3 is an implicit function of θ which we denote by $\varphi(\,)$. Using this implicit relationship and data for each individual $i = 1,\ldots,n$ we can consistently recover the realization of health status for each individual θ_i as follows,

$$\hat{\theta}_i = \varphi\big(m_i, p_i, y_i, \hat{z}(m_i;d_i), \gamma\big). \tag{5}$$

To illustrate, given our utility specification (equation 1) and under the additional assumption that $z'(m) \geq 0$, equation 5 is uniquely defined as,

$$\varphi\big(m_i, p_i, y_i, \hat{z}(m_i;d_i), \gamma\big) = m_i - \left[\frac{1}{\gamma_2}\big(y_i - p_i + \hat{z}(m_i;d_i) - m_i\big)^{-\gamma_1}\big(\hat{z}'(m_i;d_i)\big)\right]^{-\frac{1}{\gamma_3}}.$$

The economic intuition for the condition $z'(m) \geq 0$ is that as the medical utilization increases the coinsurance rate should not decrease.

Using the instrument variables x_i for each individual $i = 1, \ldots, n$ we formulate a method of moment estimator where the parameters of the utility function γ and the median of the unconditional distribution of health status (μ_θ) are jointly estimated by minimizing the following objective function,

$$(\hat{\gamma}, \hat{\mu}_\theta) = \arg\min \left\| \frac{1}{n} \sum_{i=1}^{n} g(x_i) \left\{ 1\left(\varphi\left(m_i, p_i, y_i, \hat{z}(m_i; d_i), \gamma \right) \le \mu_\theta \right) - 0.5 \right\} \right\| , \qquad (6)$$

where ||•|| is a typical quadratic norm i.e., $||x|| = x'Wx$ for a suitably chosen weighting matrix W that is used in generalized method of moments estimation, and $g(x)$ is a set of functions that generate different functional forms for a given vector of instrumental variables x. The estimation procedure iterates between using equations 5 and 6 to obtain consistent estimates of the utility parameters γ and the distribution of the latent health status θ.

In the HRS data we observe that a small but non-negligible proportion of individuals do not consume medical care. Hence we use a median based moment condition instead of the conventional mean based moment condition because the former are more robust to censoring at the upper and lower tails of the the conditional distribution of the observables (see e.g., Powell 1984, Hong and Tamer 2002 for details). The underlying assumption is that when medical utilization m is zero the first order condition does not hold but the corresponding draw from the latent health status distribution is less than the population median for that distribution.

Our estimation strategy is similar to the use of instrumental variables in conventional two stage least square methods for demand equation estimation. The unobserved health status θ is the analogue of the unobserved error term in a structural linear demand equation. In that case, once the parameters of the linear demand equation are estimated, the conditional or unconditional distributions of the error term can also be recovered non-parametrically by inverting the linear equation. This is similar to our procedure for recovering the distribution of θ by inverting the implicit function defined by equation 3. The difference between our method and the conventional two stage least square estimator, is that instead of relying on a reduced form specification of a linear functional form for the demand equation, we are using the optimality condition for a risk averse consumer in equation 3 to derive the functional form of the demand equation.[4]

V. Results

V. A. Estimates of the Demand for Out of Pocket Medical Expenditures and Supply of Health Insurance Contracts

We begin by discussing the estimates of a flexible parametric model of the demand for out of pocket medical expenditures and the supply of the health contracts using three stage least squares (Table 3). This controls for the endogeneity in the out of pocket medical expenditures and for the joint determination of the health insurance contract, as given by the level of reimbursement (which is the ratio of household out of pocket expenditures to total household medical expenditures). In this analysis we assume that the level of reimbursement (the supply of the health insurance contract) differs across insurance categories, due to unobserved heterogeneity in the individuals who pick different plans, but that there is no heterogeneity in the reimbursement schedule within plans. Table 2 presents the OLS results corresponding to the 3SLS specification. We report results for two specifications in table 3. In the first specification (cols. 1) we let the reimbursement rate only be a function of the insurance category. We find, as expected,

[4] Our semiparametric specification is similar in spirit to Campo, Guerre, Perrigne and Vuong (2003) who also assume a parametric form for the risk averse utility function but non-parametrically estimate the distribution of the latent valuations in a model of first price auctions.

that the level of reimbursement rate is the lowest for those without insurance, and highest for those with employer provided insurance. In the corresponding demand function for out of pocket medical expenditures (col. 2), we find that a 1% increase in the reimbursement would increase out of pocket medical expenditures by 0.192 percentage points, i.e., a more generous insurance plans leads to higher medical expenditures. The estimated elasticity of coinsurance of 0.192 is consistent with that found by the literature (see e.g., Zweifel and Manning 2000). It is also seen that household size, age and higher income lead to higher out of pocket medical expenditures, while being in better self reported health lowers expenditures.

In the second specification (cols. 3 and 4) we include additional instruments and let the reimbursement schedule (supply of the health, col. 3) also be a function of the local unemployment rate, number of employees per week in county, number of establishments in county, annual payroll for at county level, state housing price index, number of people covered by insurance in state and the three county level components of the Medicare GPCI.[5] We find that being uninsured leads to the highest out of pocket medical expenditures, where as the out of pocket expenditures are the least for those with employer provided insurance. Among the additional instruments we find that the log of state housing price index has a statistically significant and negative effect on the reimbursement rate, i.e. states that have higher housing prices also have lower reimbursement rates presumably because higher real estate costs raise the costs of providing insurance. Similarly states with higher values of the malpractice premium component from the Medicare GPCI have lower reimbursement rates, due to higher costs of providing insurance coverage in these states.

[5] The GPCI reflects the relative costs of practice expenses, malpractice insurance, and physician work in an area compared to the national average

In the analysis of the demand for out of pocket medical expenditures (col. 4) we find that a 1% increase in the reimbursement increases out of pocket medical expenditures by 0.426 percentage points, i.e., a more generous insurance plans leads to higher medical expenditures. This estimated elasticity of coinsurance is consistent but at the higher end of that found in the literature (see e.g., Zweifel and Manning 2000). The other determinants of the demand for out of pocket medical expenditures are very similar to those found in the previous specification (col. 2).

V. B. Estimates of the Structural Utility Parameters

We next discuss the results from the two step semiparametric estimation of the model presented in section III. Table 4, panel A presents the second step estimates of the utility parameters where in the first step the reimbursement schedule *z(m;d)* is estimated using OLS (Table 3, col. V). Table 4, panel B, presents results where the reimbursement schedule in the first stage is estimated using a local linear regression. As the difference between the estimates of the parameters in these two panels suggests, parametric assumptions in the first stage can have a considerable influence on the estimates of the structural parameters in the second stage. Table 5 reports the summary statistics for the local linear regression. We find that uninsured individuals on average incur an out of pocket cost of $1404,[6] which is higher than those with employer provided insurance have an average cost of $1165. Self employed individuals with insurance through their own business have a higher average cost of $1968, while those on Medicare have the lowest cost of $902. Those who live in a rural area incur an average lower cost of $238. This difference in out of pocket costs arises both due to differences in reimbursement as well as the difference in medical utilization, e.g., the uninsured have lower expenditures than

[6] This is the difference between the constant and the parameter on the dummy variable for insurance category 1.

those who are self-employed because even though they are not reimbursed they seek less medical care. Figures 4 reports the shape of the reimbursement schedule. We see that it has a concave shape that can easily be approximated by piecewise linear functions. Figure 4 provides a description of the gradient of the reimbursement schedule. We use the gradient which is a measure of the incentives in the insurance contract to infer the latent health status of an individual. The gradient is non-negative almost everywhere which is consistent with our assumption that the coinsurance rate does not decreases as total medical expenses (or medical utilization) increase, i.e., $z'(m) \geq 0$.[7] This gradient is in turn used in our objective function (eq. 6) in order to obtain the estimates of the parameters of the utility function which are discussed next. The coefficient of relative risk aversion for the aggregate consumption commodity is estimated to be 0.64, while that for the consumption of health care is 1.39. Thus individuals are more risk averse with respect to health care than the aggregate consumption commodity. These numbers are broadly within the range of the estimates found in the literature on consumption (see e.g., Zeldes 1989, Shea 1995, Hansen and Singleton 1982, Gourinchas and Parker 2002) The utility weight on the consumption of health care relative the aggregate consumption commodity is estimated to be 5.4. Thus individuals value consumption of the health care more than they do aggregate consumption commodity. Hence an individual would have to be compensated more than one dollar worth of the aggregate consumption commodity in order to give up one dollar worth of health care. The median monetary value of the latent health status (θ) is found to be \$3464 in 1996 dollars. This implies that a median person in perfect health would have an extra \$3464 to spend.

VI. Examination of Adverse Selection and Moral Hazard.

[7]As described in section in the estimation of the structural parameters we dropped the 114 observations for which the derivative of the reimbursement schedule as negative. This was 3% of the sample used in estimating the reimbursement schedule in the first stage.

VI.A. Examination of Adverse Selection.

The results of our estimation procedure allow us to test for the presence of adverse selection in health insurance plans. Given the estimated distribution of the unobserved health status θ, we recover conditional distributions of the latent health status for each health insurance plan. Table 7 reports the distribution of the unobserved health status conditional on insurance category. We find the mean of those on Medicare is the largest suggesting that these individuals on average are in worst health. These individuals also have the largest variance and hence they face the largest risk. The mean of those who are self employed is the lowest implying that these are the healthiest. They also have the lowest variance, i.e., the smallest risk. We then examine whether these conditional distributions differ significantly in a statistical sense. Figure 1 reports the estimates of the density of the latent health status conditional on insurance status. Figure 1 plots the densities and figure 7 the CDF of the elasticities by insurance category. Looking at fig. 7 we see that the CDF of individuals on Medicare strictly dominates the CDFs for all other insurance categories. This suggests that individuals on Medicare insurance are in the worst health. At the other extreme the CDF of individuals who are self-employed is dominated by the other CDFs suggesting that these individuals are in the best health.

In Table 8 we report the results of a formal Kolmogorov-Smirnov test among all the possible pairs of the insurance categories. We found that the health distributions between people who are not insured (insurance category 1) is statistically significantly different form people in all the other insurance categories (insurance categories 2, 3, 5, 6, which are, respectively, employer provided insurance, insurance of self employed, personal insurance and Medicare). Similarly, with only two exceptions, we also find evidence against the null hypothesis of no adverse selection in the pair-wise comparison among different insurance categories. The two exceptions are: first, there is no significant difference in the unobserved health distributions between individuals in insurance category 2 (employer provided insurance) and 3 (self employed); second, we find no difference between those who have personal insurance (category 5) and individuals in category 6 (Medicare).

VI.B. Examination of Moral Hazard.

We first develop and compute an alternative measure of moral hazard that is more general than that traditionally found in the empirical literature, e.g. Manning et al (1987).[8] Using the structural estimates of the utility parameters and the distribution of latent health status, we compute the elasticity of total medical expenditure with respect to a local change of the rate of reimbursement of medical expenditures:

$$\frac{\delta m}{\delta z'(m)} \times \frac{z'(m)}{m}.$$

This represents a counter factual policy experiment where for each individual, there is a marginal change in the reimbursement rate with the corresponding change in the demand for medical care. One advantage of this new measure of moral hazard is that it allows considerable nonlinearities in the change in behaviors with respect to changes in the reimbursement policy. This allows us to compute a distribution of elasticities for every individual in our sample, rather than computing a single statistic of the measure We compute this number for every individual $i = 1, \ldots, n$ in our data set using their observed level of total medical expenditures and their associated out of pocket costs, given the estimates of the model parameters. This elasticity is calculated by applying the implicit

[8] If the reimbursement rate was a constant c in which case the reimbursement schedule would be a linear function of the total expenditure, e.g., z = c.m, then our measure of moral hazard is identical to that used traditionally.

function theorem to the first order condition (2) and assuming $z(m_i) = m_i \cdot z'(m_i) + cons$, which yields:

$$\frac{\delta m_i}{\delta z'(m_i)} \times \frac{z'(m_i)}{m_i} = \frac{-\gamma_1 c_i^{-\gamma_1 - 1} m_i z'(m_i) - c_i^{-\gamma_1}}{\gamma_1 c_i^{-\gamma_1 - 1} z'(m_i)^2 + \gamma_2 \gamma_3 (m_i - \theta_i)^{-\gamma_3 - 1}} \times \frac{z'(m_i)}{m_i}.$$

Table 9 provides the summary statistics of this elasticity unconditional on the insurance category. We find that, on average, for a one percent change in the rate of reimbursement, total medical expenditures of individuals dropped by 0.56 percent. This number is slight larger than the demand elasticity of medical expenditure with respect to the log of reimbursement in column V of table 4 which was estimated to be 0.42. The fact that the mean of the estimated distribution of elasticities does not differ substantially from that in the existing literature (e.g. Manning et al. 1987) gives us confidence that our model, which is novel compared to the previous research, does not suffer from a large misspecification bias. On the other hand this allows us to examine the entire distribution of elasticities, e.g., we find that at the 25th percentile the elasticity is 0.64 and at the 75th percentile it is 0.12. Figure 2 plots the densities and figure 6 the CDF of the elasticities by insurance category. In fig. 6 the CDF of the self-employed is dominated by the CDFs of individuals in the other insurance categories. This suggests that these individuals are the most elastic, which seems quite intuitive. This is also consistent with the observation from fig. 7 that these individuals are in the best health.

Table 10 examines the difference in the distribution of elasticity conditional on insurance category. We find that there is a statistically significant difference in most pair wise comparisons of the elasticity of medical expenditures by insurance category. The exceptions are between categories 1 (uninsured) and 5 (personal insurance), 1 and 6 (Medicare), 2 (employer provided insurance) and 3 (insurance of self-employed), and 5 and 6. In the examination of adverse selection we found that there was no difference in pair wise comparisons of unobserved health status distribution by insurance category for insurance categories 2 and 3, and 5 and 6. Hence the finding that there is no difference in the elasticity of medical expenditures in a pair wise comparison of these insurance categories is not surprising. On the other hand we find that there is no difference in the elasticity of medical expenditures between individuals who are uninsured and those either on private insurance or on Medicare. This lack of difference arises in spite of a statistically significant difference in the latent health distribution of individuals in these plans. This implies that the reimbursement schedule of the private insurance plans and Medicare is designed to modify the behavior their insurees to be no different from individuals who are uninsured. This in turn suggests that these insurance plans are well designed to cope with adverse selection.

We describe the effects of various household characteristics on the elasticity of medical expenditure in tables 11 a (in levels) and b (in logs). It is seen that a higher health shock (i.e., worse health) makes individuals more inelastic in their medical expenditures. The same is the case as total and out of pocket medical expenditures increase. On the other hand higher incomes make individuals more elastic. In the interest of brevity we do not discuss the other coefficients. These results, and the correlations reported in table 12, as well the plots in figures 8 and 9 show that the relationship between the elasticity of medical expenditures and health shocks is monotonic but highly non-linear. In fact individuals with extremely large health shocks are almost inelastic. This also highlights the importance of allowing for non-linearities in estimating the reimbursement schedule.

VII. Conclusions, Limitations and Sensitivity Analysis.

We specify a model of demand for health insurance and medical utilization in the presence of unobserved heterogeneity in the latent health status of individuals. Using a semiparametric procedure we estimate the structural parameters of this model accounting adverse selection and moral hazard due to asymmetric information. Our estimation procedure does not rely on an explicit model of the choice of health insurance. We use the estimates of the model to examine the nature of adverse selection and moral hazard in health insurance contracts. We find evidence of adverse selection in health insurance plans. We also find evidence for moral hazard, though interestingly there is no evidence of moral hazard in private insurance plans and Medicare, suggesting that the design of these insurance contracts has controlled for the selection in these plans in such a manner that behavior does not differ between individuals in these plans and those who are uninsured.

Although our proposed semi-parametric method provides a more flexible and robust alternative for analyzing the empirical issues of adverse selection and moral hazard in health insurance, several limitations are acknowledged. The utility function specification we use is assumed to be separable in the consumption of the composite good and the difference between the medical utilization and the latent health status. While this specification captures the risk aversion features of consumer utilities in health status, it rules out more flexible interactions between the utility derived from composite good consumption and difference between the medical utilization and the health status. It is sometimes argued that the marginal utility for composite good consumption might decrease in the case of severe illness (e.g. Viscusi and Evans 1990). In future work we plan to extend our approach to other more general function forms for consumer utilities.

We also plan to consider extensions to incorporate stochasticity in to the production relationship between the medical utilization and the health status. However, since we allow for a flexible nonlinear relationship between medical utilization and health status, we do not expect that the incorporation of stochasticity in to the health production relationship will substantially change our empirical results. In conclusion, in spite of these limitations our research is novel in that it develops a computationally simple estimation procedure under minimal parametric assumptions to simultaneously examine adverse selection and moral hazard in health insurance contracts. Our research is also important as it provides a framework for similar analysis in other contexts where distortions exist due to asymmetric information.

Appendix: Computing standard errors for the semi-parametric estimator

In this appendix we discuss how to compute the correct standard errors for the two step semi-parametric estimator for the parameters γ and μ, where γ are the risk aversion parameters in the utility function and μ are the nuisance parameters that characterizes the moments of the distribution of the health status shock θ.

The estimator that we use in this paper is based on moment conditions of the form of

$$m_n(\hat{\gamma},\hat{\mu}) = \frac{1}{n}\sum_{i=1}^{n} h\left(x_i, \theta\left(w_i, \hat{z}(m_i;d_i), \hat{z}'(m_i;d_i), \hat{\gamma}\right), \hat{\mu}\right),$$

where x_i denote the instruments and w_i denote the data m_i, p_i, y_i and d_i. For example, equation (6) gives the particular form of the moment condition $h\left(x_i, \theta\left(w_i, \hat{z}(m_i;d_i), \hat{z}'(m_i;d_i), \hat{\gamma}\right), \hat{\mu}\right)$ we used in the estimation procedure. The GMM estimator $\hat{\alpha} = (\hat{\gamma},\hat{\mu})$ is calculated using the quadratic norm

$$\min_{\gamma,\mu} Q_n(\hat{\alpha}) \equiv m_n(\hat{\gamma},\hat{\mu})' W_n m_n(\hat{\gamma},\hat{\mu}).$$

The asymptotic distribution of $\hat{\alpha}$, as an usual two step GMM estimator, is given by

$$\sqrt{n}(\hat{\alpha}-\alpha) \xrightarrow{d} N\left(0, (G'WG)^{-1}(G'W\Omega WG)(G'WG)^{-1}\right),$$

where $G = \frac{\partial}{\partial\alpha} Em_n(\alpha)$, $W_n \to W$, and Ω is the limiting asymptotic variance of

$$\sqrt{n} m_n(\gamma,\mu).$$

To derive Ω, we will use the theory developed in Newey (1994), who showed that Ω does not depend on the particular nonparametric method that is being used to estimate $z(\cdot)$ and $z'(\cdot)$. Therefore we can derive the form of Ω regardless of whether we are using sieve or kernel methods to estimate the reimbursement function and its derivative.

To describe the asymptotic distribution using the framework of Newey (1994), define

$$\bar{h}\left(z\left(m_i;d_i\right), z'\left(m_i;d_i\right), m_i;d_i\right) = E\left[h\left(x_i, \theta\left(w_i, \hat{z}(m_i;d_i), \hat{z}'\left(m_i;d_i\right), \hat{\gamma}\right), \hat{\mu}\right) \mid m_i, d_i\right],$$

and

$$d_1\left(m_i;d_i\right) = \frac{\partial}{\partial z\left(m_i;d_i\right)} \bar{h}\left(z\left(m_i;d_i\right), z'\left(m_i;d_i\right), m_i;d_i\right),$$

and

$$d_2\left(m_i;d_i\right) = \frac{\partial}{\partial z'\left(m_i;d_i\right)} \bar{h}\left(z\left(m_i;d_i\right), z'\left(m_i;d_i\right), m_i;d_i\right).$$

Then using the calculations in Newey (1994), under suitable regularity conditions we can write, up to a term that converges to 0 in probability, $\sqrt{n}m_n(\gamma,\mu)$ as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\left(\delta_1\left(m_{i,}d_i\right)+\delta_2\left(m_{i,}d_i\right)\right)\left(z_i - z\left(m_i;d_i\right)\right) + h\left(x_i, \theta\left(w_i, \hat{z}(m_i;d_i), \hat{z}'\left(m_i;d_i\right), \hat{\gamma}\right), \hat{\mu}\right)\right),$$

where we define

$$d_1\left(m_i;d_i\right) = \delta_1\left(m_i;d_i\right)$$

and

$$\delta_2\left(m_i;d_i\right) = -\frac{1}{f\left(m_i \mid d_i\right)}\frac{\partial}{\partial m}\left[d_2\left(m_i,d_i\right)f\left(m_i \mid d_i\right)\right].$$

Therefore

$$\Omega = Var\left(\left(\delta_1\left(m_i,d_i\right)+\delta_2\left(m_i,d_i\right)\right)\left(z_i - z\left(m_i,d_i\right)\right) + h\left(x_i, \theta\left(w_i, \hat{z}(m_i;d_i), \hat{z}'\left(m_i;d_i\right), \hat{\gamma}\right), \hat{\mu}\right)\right).$$

Given that we know the form of the limiting variance, the next and final step is to obtain a consistent estimate of Ω, the asymptotic variance. For this we need to obtain consistent estimates of the elements of the asymptotic linear influence function, in particular the functions. Each of these components is related to the conditional expectation function,

$$\bar{h}(z(m_i;d_i),z'(m_i;d_i),m_i;d_i) = E[h(x_i,\theta(w_i,\hat{z}(m_i;d_i),\hat{z}'(m_i;d_i),\hat{\gamma}),\hat{\mu}) \mid m_i,d_i].$$

This can be estimated in a variety of ways, including using a kernel regression:

$$\hat{\bar{h}}(z(m_i;d_i),z'(m_i;d_i),m_i;d_i)$$
$$= \frac{\sum_{j=1}^{n} 1(d_j = d_i)\kappa\left(\frac{m_j - m_i}{h}\right)h(x_j,\theta(w_j,\hat{z}(m_j;d_j),\hat{z}'(m_j;d_j),\hat{\gamma}),\hat{\mu})}{\sum_{j=1}^{n} 1(d_j = d_i)\kappa\left(\frac{m_j - m_i}{h}\right)}.$$

Subsequently, we can estimate $d_1(m_i;d_i)$ by

$$\hat{d}_1(m_i;d_i) = \frac{\partial}{\partial z(m_i,d_i)}\hat{\bar{h}}(z(m_i;d_i),z'(m_i;d_i),m_i;d_i),$$

and estimate $d_2(m_i;d_i)$ by

$$\hat{d}_2(m_i;d_i) = \frac{\partial}{\partial z'(m_i,d_i)}\hat{\bar{h}}(z(m_i;d_i),z'(m_i;d_i),m_i;d_i).$$

We can also non-parametrically estimate $f(m_i \mid d_i)$ by

$$\hat{f}(m_i \mid d_i) = \frac{1/n \sum_{j=1} 1\left(d_j = d_i\right)\kappa\left(\frac{m_j - m_i}{h}\right)}{\frac{1}{n}\sum_{j=1}^{n} 1\left(d_j = d_i\right)}.$$

Then we can estimate $\delta_2(m_i;d_i)$ by

$$\hat{\delta}_2(m_i;d_i) = -\frac{1}{\hat{f}(m_i \mid d_i)}\frac{\partial}{\partial m}\left[\hat{d}_2(m_i,d_i)\hat{f}(m_i \mid d_i)\right],$$

where the derivative can be evaluated numerically. Finally, we can then estimate the asymptotic by the empirical sum of the outer product of the estimated influence function.

References:

Aguirregabiria, V. and Mira P. Sequential Simulated-Based Estimation of Dynamic Discrete Games, technical report, Boston University. 2002.

Akerlof, George. The Market for "Lemons": Quality Uncertainty and the Market Mechanism. The Quarterly Journal of Economics. 1970; 84(3): 488-500.

Arrow, K.J. Uncertainty and the Welfare Economics of Medical Care. American Economic Review. 1963; 53:941-973.

Athey, S. and P. Haile. Identification of Standard Auction Models. Econometrica. 2002; 70:2107-2140.

Bajari, Patrick and Lanier Benkard. Hedonic Price Indexes with Unobserved Product Characterstics. Working Paper, forthcoming at the Journal of Business and Economic Statistics. 2003.

Bajari, Patrick; Lanier Benkard, and Jon Levin. Estimating Dynamic Models of Imperfect Competition. Working Paper. 2003.

Bajari, Patrick and Ali Hortacsu. Are Structural Estimates of Auction Models Reasonable? Evidence from Experimental Data. Working paper under revision at The Journal of Political Economy. Duke University and University of Chicago 2002.

Bajari, Patrick and Matt Kahn. Estimating Housing Demand with an Application to Explaining Racial Segregation in Cities. Working paper accepted, subject to final revisions, at the Journal of Business and Economic Statistics. 2002.

---. The Winner's Curse, Reserve Prices and Endogenous Entry: Empirical Insights from eBay Auctions. The Rand Journal of Economics. 2003; 34(2):329-355.

Besley, T.J. Optimal Reimbursement Health Insurance and the Theory of Ramsey Taxation. Journal of Health Economics. 1988; 7:321-336.

Blau, D. and Donna B. Gilleskie. The Role of Retiree Health Insurance in the Employment Behavior of Older Males. Working Paper. UNC. 2003.

Blomqvist, A.G. Optimal Non-linear Health Insurance. Journal of Health Economics. 1997; 16(3):303-321.

Cameron. A. C.; P. K. Trivedi, Frank Milne, and J. Piggott. A Microeconometric Model of the Demand for Health Care and Health Insurance in Australia. Review of Economic Studies. 1988; 55(1):85-106.

Campo, Sandra M.; I. Perrigne, and Q. Vuong. Asymmetry and Joint Bidding in OCS Wildcat Auctions. Journal of Applied Econometrics. 2003; 18:179-207.

Campo, Sandra M.; E. Guerre, I. Perrigne, and Q. Vuong. Semiparametric Estimation of First-Price Auctions with Risk Averse Bidders. 2003 (last version), revised version submitted to Econometrica.

Cardon, James H. and Igal Hendel. Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey. RAND Journal of Economics. 2001; 32(3):408-27.

Chernozhukov, Victor and Chris Hansen. IV Model of Quantile Treatment Effects. MIT Working Paper. 2003.

Chiappori, Pierre-Andre and B. Salanié. Testing for Asymmetric Information on Insurance Markets. Journal of Political Economy. 2000; 108(1):56-78.

Currie, Janet and Brigitte. C. Madrian. Health, Health Insurance and the Labor Market. Ashenfelter, Orley and David Card (eds) Handbook of Labor Economics Vol. 3C. Amsterdam ; New York : Elsevier; 1999; pp. 3309-3416.

Cutler, David M. and Richard J. Zeckhauser. The Anatomy of Health Insurance. Culyer, Anthony J. and Joseph P. Newhouse (eds) Handbook of Health Economics Vol. 1A. Amsterdam ; New York : Elsevier; 2000; pp. 563-643.

Diamond, P. and J. Mirrlees. A Model of Social Insurance with Variable Retirement. Journal of Public Economics. 1978; 10: 295-336.

Ehrlich, I. and G.S. Becker. Market Insurance, Self-Insurance, and Self-Protection. Journal of Political Economy. 1972; 80:623-648.

Feldman, R. and D. Dowd. Must Adverse Selection Cause Premium Spirals? Journal of Health Economics. 1991; 10(3):350-357.

Feldstein, M.S. The Welfare Loss of Excessive Health Insurance. Journal of Political Economy. 1973; 81(1):251-280.

Ferrall, Christopher and Bruce Shearer. Incentives and Transactions Costs Within the Firm: Estimating an Agency Model Using Payroll Records. Review of Economic Studies. 1999; 66:309-338.

Gertler, Paul and Jonathan Gruber. Insuring Consumption Against Illness. American Economic Review. 2002; 92(1): 51-70.

Gilleskie, Donna B. A Dynamic Stochastic Model of Medical Care Use and Work Absence. Econometrica. 1998; 66(1):1-45.

Glied, Sherry. Managed Care. Culyer, Anthony J. and Joseph P. Newhouse (eds) Handbook of Health Economics Vol. 1A. Amsterdam ; New York : Elsevier; 2000; pp. 409-459.

Gourinchas, P. O. and J. Parker. Consumption Over the Life Cycle. Econometrica. 2002. 70(1): 47-89.

Grossman, M. On the Concept of Health Capital and the Demand for Health. Journal of Political Economy. 1972; 80(2):223-255.

Gruber, Jonathan. Health Insurance and the Labor Market. Culyer, Anthony J. and Joseph P. Newhouse (eds) Handbook of Health Economics Vol. 1A. Amsterdam ; New York : Elsevier; 2000; pp. 645-706.

Guerre, E.; I. Perrigne, and Q. Vuong. Optimal Nonparametric Estimation of First-Price Auctions. Econometrica. 2000; 68(3):525-574.

Haile, Phil; Han Hong, and Matthew Shum. Nonparametric Tests for Common Values in First Price Sealed-Bid Auctions. Working Paper. 2002.

Hansen, L. P. and K. J. Singleton. Generalized Instrumental Variables Estimation of Non Linear Rational Expectations Models. Econometrica. 1982; 50(5):1269-1285

Hardle, Wolfgang, Applied Nonparametric Regression, Cambridge University Press, 1990.

Harris, K. and M. Keane. A Model of Health Plan Choice: Inferring Preferences and Perceptions From a Combination of Revealed Preferences and Attitudinal Data. Journal of Econometrics. 1999; 89: 131-157.

Hendel, Igal and Alessandro Lizzeri. Adverse Selection in Durable Goods Markets. American Economic Review. 1999; 89(5):1097-1115.

---. The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance. forthcoming Quarterly Journal of Economics. 2003.

Hong, Han and Elie Tamer. Inference in Censored Models with Endogenous Regressors. Econometrica. 2003; 71(3):905-932.

Hotz, J. and Miller R.A. Conditional Choice Probabilities and the Estimation of Dynamic Models. Review of Economic Studies. 1993; 60:497-529.

Hubbard, R. Glenn, Jonathan Skinner, and Stephen P. Zeldes. Precautionary Savings and Social Insurance. Journal of Political Economy. 1995; 103(2): 360–99.

Juster, F.T. and R. Suzman. The Health and Retirement Study: An Overview. Journal of Human Resources. 1995; Supplement (JHR 30-S): S7-S56.

Keeler, E., J. Newhouse and C. Phelps. Deductibles and the Demand for Medical Care Services: The Theory of a Consumer Facing a Variable Price Schedule under Uncertainty. Econometrica 1977; 45: 641-655.

Khwaja, Ahmed. Health Insurance Habits and Health Outcomes: A Dynamic Stochastic Model of Investment in Health. Proceedings of 2002 North American Summer Meetings of the Econometric Society; 2002. edited by Case, Anne; Raquel Fernandez, John Rust, and Ken Wolpin

---. The Dynamic Trade-off Between Medical Expenditure Insurance, Moral Hazard and Mortality Risk: A Life Cycle Analysis of the Effects of Medicare on Individual Health Incentives and Health Outcomes. Working Paper. Duke University, Fuqua School of Business; 2005.

Laffont, J. J. and Q. Vuong. Structural Analysis of Auction Data. American Economic Review, Papers and Proceedings. 1996; 414-420.

Li, T.; I. Perrigne, and Q. Vuong. Conditionally Independent Private Values in OCS Wildcat Auctions. Journal of Econometrics. 2000; 98(1):129-161.

Lucas, Robert. Econometric Policy Evaluation: a Critique. In *The Phillips Curve and the Labor Market*, eds. K. Brunner and A. Meltzer, Vol. 1 of the Carnegie-Rochester Conferences in Public Policy, a supplemental series to the *Journal of Monetary Economy*. Amsterdam: North-Holland.

Manning, W.G.; J.P. Newhouse, N.Duan et al. Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment. American Economic Review. 1987; 77(3):251-277.

Mirrlees, M.A. An Exploration in the Theory of Optimum Income Taxation. Review of Economic Studies. 1971; 38:175-208.

---. Theory of Moral Hazard and Unobservable Behavior: Part I. Review of Economic Studies. 1999; 66(1):3-21

Newey, W. K. The Asymptotic Variance of Semiparametric Estimators. Econometrica. 1994; 62(6):1349-1382.

Newhouse, J.P. Medical Care Costs. How Much Welfare Loss? Journal of Economic Perspectives. 1992; 6(3):3-21.

Newhouse, J.P. and the Insurance Experiment Group. Free For All? Lessons from the Health Insurance Experiment. Cambridge: Harvard University Press. 1993.

Paarsch, Harry J. and Bruce Shearer. Piece Rates, Fixed Wages, and Incentive Effects: Statistical Evidence from Payroll Records. International Economic Review. 2000; 41(1):59-92.

Pakes, A. and P. McGuire. Stochastic Algorithms, Symmetric Markov Perfect Equilibrium, and the 'Curse' of Dimensionality. Econometrica, 2001; 69:1261-1282.

Pauly, M. The Economics of Moral Hazard: Comment. American Economic Review. 1968; 58:531-536.

---. Overinsurance and Public Provision of Insurance: the Roles of Moral Hazard and Adverse Selection. Quarterly Journal of Economics. 1974; 88(1):44-54.

---. Insurance Reimbursement. Culyer, Anthony J. and Joseph P. Newhouse (eds) Handbook of Health Economics Vol. 1A. Amsterdam ; New York : Elsevier; 2000; pp. 537-560.

Powell, James. Least Absolute Deviations Estimation for the Censored Regression Model. Journal of Econometrics. 1984; 25:303-325.

Pesendorfer, Martin and Philipp Schmidt-Dengler. Identification and Estimation of Dynamic Games. NBER working paper 9726. 2003.

Picard, Pierre. On the Design of Incentive Schemes under Moral Hazard and Adverse Selection. Journal of Public Economics. 1987; 33(3):305-31

Rothschild, M. and J.E. Stiglitz. Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. Quarterly Journal of Economics. 1976; 90(4):630-649.

Rust, J. and C. Phelan. How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets. Econometrica. 1997; 65(4): 781-831.

Shea, J. Union Contracts and the Life-Cycle/Permanent-Income Hypothesis. American Economic Review. 1995; 85(1): 186-200.

Spence, M. Job Market Signaling. The Quarterly Journal of Economics. 1973; 87(3): 355-374.

Spence, M. and R. Zeckhauser. Insurance, Information, and Individual Action. American Economic Review. 1971; 61(2):380-387.

Tamer, Elie. Incomplete Simultaneous Discrete Response Model with Multiple Equilibria. Review of Economic Studies. 2003; 70(1):147-165.

Theilen, B. Simultaneous Moral Hazard and Adverse Selection with Risk Averse Agents. Economics Letters. 2003; 79(2):283-89.

Viscusi, K. and W. Evans. Utility Functions that Depend on Health Status: Estimates and Economic Implications. American Economic Review. 1990; 80: 353-374.

Whinston, Michael D. Moral Hazard, Adverse Selection, and the Optimal Provision of Social Insurance. Journal of Public Economics. 1983; 22(1):49-71

Wilson, C. A Model of Insurance Markets with Incomplete Information. Journal of Economic Theory. 1977; 16:167-207.

Wilson, C. The Nature of Equilibrium in Markets With Adverse Selection. Bell Journal of Economics. 1980; 11(1):108-130.

Vera-Hernandez, Marcos. Estimation of a Principal-Agent Model: Moral Hazard in Medical Insurance. The Rand Journal of Economics. 2003. 34(4):670-693.

Zeldes, S. P. Consumption and Liquidity Constraints: An Empirical Investigation. Journal of Political Economy. 1989; 97(2):305-346.

Zeckhauser, R. Medical Insurance: a Case Study of the Tradeoff Between Risk Spreading and Appropriate Incentives. Journal of Economic Theory. 1970; 2(1):10-26.

Zweifel, Peter and Willard G. Manning. Moral Hazard and Consumer Incentives in Health Care. Culyer, Anthony J. and Joseph P. Newhouse (eds) Handbook of Health Economics Vol. 1A. Amsterdam ; New York : Elsevier; 2000; pp. 409-459.

Appendix: Variable Descriptions

Variable	Variable Description
HRS Data	
hosp_n3	hosp. nights
phys_v3	phys. Visits
hlth*	health
hsize3	HH size
hhinc3	HH income
ins3	insurance category
spoucvr3	insurance coverage from spouse
hhincst3	HH insurance cost (premium)
hhoop3	total HH out of pocket (OOP) cost for med. Care
hhmedex3	total imputed HH medical exp. (sum of OOP + insurer payment)
hhosoop3	HH hospital OOP costs
hdocoop3	HH physician OOP cost
prevemp*	employed at previous wave with same employer at wave *
nemploc*	number of employees at local plant with the firm at wave *
nempall*	number of employees at all locations with firm at wave *
hhinc	HH income
totalmed	total medical expenses. Analgous to hhmedex or "m"
totoop	a measure of OOP expenditures. (total medical expenses –insurance reimbursements)
oop	a measure of OOP expenditures. (similar to totoop but excludes drug expenditures and expenditure on special medical equipment etc. This measure only includes hospital and physician costs.)
UrbRural	A measure (0-9) of how rural an area is. 0. Central counties of metropolitan areas of 1 million population or more 1. Fringe counties of metropolitan areas of 1 million population or more 2. Counties in metropolitan areas of 250 thousand to 1 million population 3. Counties in metropolitan areas of than 250 thousand population 4. Urban population of 20,000 or more, adjacent to a metropolitan area 5. Urban population of 20,000 or more, not adjacent to a metropolitan area 6. Urban population of 2,500 to 19,999, adjacent to a metropolitan area 7. Urban population of 2,500 to 19,999, not adjacent to a metropolitan area 8. Completely rural or less than 2,500 urban population, adjacent to a metropolitan area 9. Completely rural or less than 2, urban population, not adjacent to a metropolitan area Note : In estimation only a dummy for UrbRural = 0 is used.
Non HRS Data	
AP	Annual Payroll-available at County level Source: US Census Bureau (County Business Patterns) http://www.census.gov/epcd/cbp/view/cbpview.html Total payroll includes all forms of compensation, such as salaries, wages, reported tips, commissions, bonuses, vacation allowances, sick-leave pay, employee contributions to qualified pension plans, and the value of taxable fringe benefits. For corporations, it includes amounts paid to officers and executives; for unincorporated businesses, it does not include profit or other compensation of proprietors or partners. Payroll is

	reported before deductions for Social Security, income tax, insurance, union dues, etc. First-quarter payroll consists of payroll during the January-to-March quarter.
Emp	Employees per week-available at County level Source : US Census Bureau (County Business Patterns) http://www.census.gov/epcd/cbp/view/cbpview.html Paid employment consists of full- and part-time employees, including salaried officers and executives of corporations, who are on the payroll in the pay period including March 12.Included are employees on paid sick leave, holidays, and vacations; not included are proprietors and partners of unincorporated businesses.
Est	Total Establishments –available at County level Source : US Census Bureau (County Business Patterns) http://www.census.gov/epcd/cbp/view/cbpview.html An establishment is a single physical location at which business is conducted or services or industrial operations are performed. It is not necessarily identical with a company or enterprise, which may consist of one or more establishments. When two or more activities are carried on at a single location under a single ownership, all activities generally are grouped together as a single establishment. The entire establishment is classified on the basis of its major activity and all data are included in that classification. Establishment-size designations are determined by paid employment in the mid-March pay period. The size group "1 to 4" includes establishments that did not report any paid employees in the mid-March pay period but paid wages to at least one employee at some time during the year. Establishment counts represent the number of locations with paid employees any time during the year. This series excludes governmental establishments except for wholesale liquor establishments (NAICS 4228), retail liquor stores (NAICS 44531), Federally-chartered savings institutions (NAICS 522120), Federally-chartered credit unions (NAICS 522130), and hospitals (NAICS 622).
HPI	Housing Price Index—Available at State level Source : Office of Federal Housing Enterprise Oversight http://www.ofheo.gov/HPI.asp HPI is a measure designed to capture changes in the value of single family homes. HPI is a weighted repeat sales index that measures average price changes in repeat sales or refinancings on the same properties. Data provided by Fannie Mae and Freddi
Inscov	People covered by insurance- Available at State level. Source : US Census Bureau CPS 1997 http://www.census.gov/hhes/www/hlthins/cover96/c96tabf.html Includes private and government sponsored plans
Insnotcov	People not covered by insurance- Source :US Census Bureau CPS 1997 http://www.census.gov/hhes/www/hlthins/cover96/c96tabf.html
GPCI Variables	work, malp, practice-Available at county level. Source : Medicare Payment Advisory Commission http://www.mgma.com/research/gpci.cfm A Geographic Practice Cost Index (GPCI) is used by Medicare to adjust for variance in operating costs of medical practices located in different parts of the country. Reimbursement of Physicians for services performed under Medicare is governed by a formula that considers the product of three factors: 1. A nationally uniform relative value unit (RVU) for the service; 2. A GPCI value which adjusts each RVU component (Work, Practice Expense, malpractice); 3. A nationally uniform conversion factor for the service.

	The Conversion Factor converts the relative values into payment amounts. For each physician fee schedule service, which is represented by an associated Health Care Common Procedure Coding System (HCPCS) code, there are three relative values: 1. An RVU for physician work; 2. An RVU for practice expense; 3. An RVU for malpractice expense. For each of these components, there is a GPCI which adjusts the RVU value based on a practices geographic location. The GPCIs reflect the relative costs of practice expenses, malpractice insurance, and physician work in an area compared to the national average for each component.

Table 1: Summary Statistics

Variable	Variable Description	Obs	Mean	Std. Dev.	Min	Max
HRS Data						
hosp_n3	Hospital nights	3735	1.247657	5.441511	0	176.00
phys_v3	Physician visits	3735	7.541365	8.541172	0	50.00
hlth3	Health status	3735	4.412048	1.110506	2.00	6.00
hsize3	Household size	3735	2.397323	1.191198	1	11
hhinc3	Household income	3735	51363.54	51112.21	0	515000
ins3	Insurance category	3735	2.598661	1.500883	1	6
spoucvr3	Ins coverage from spouse	3735	.1917001	.3984244	-1	1
hhincst3	Hh insurance premium	3735	1475.48	3117.755	0	121500
hhoop3	Total household OOP cost	3735	1970.326	3213.294	0	44854
hhmedex3	Total medical expenses	3735	16177.11	36864.79	0	600000
hhosoop3	Hospital OOP costs	3735	426.5023	1919.887	0	40000
hdocoop3	Physician OOP costs	3735	1445.469	2081.733	0	25000
prevemp3	Employed with same at w2	3735	.4064257	.4917765	-1	1
nemploc3	# employees at local plant	3735	229.2367	1034.756	0	20000
nempall3	# employees at all locations	3735	2738.709	29823.28	-1	999979
Mage3	Age	3735	58.64739	5.486718	25	85
UrbRural	Size of local area	3735	2.111647	2.424016	-1	9
Non HRS Data						
AP	Annual payroll state/county	3735	10494217	19966407	5460	109409701
Emp	# employees state/county	3735	339386.9	577239.6	242.00	3470100
Est	#establishments state/county	3735	20508.13	33989.78	42	215591
HPI	Housing price index	3735	195.4352	36.47388	125.81	305.82
Inscov	People covered by ins in state	3735	9165.549	6863.702	459	25853
Insnotcov	People not covered by ins in state	3735	1407.709	1771.727	19	6514
Work	Geog Practice Cost index	3735	.9898594	554.0259	0.936	1.068
Malp	Geog Practice Cost index	3735	.9458102	.4598232	0.356	3.051
Practice	Geog Practice Cost index	3735	.962483	.0973378	0.835	1.33
une	Unemployment Rate	3735	5.430335	2.760459	1.5	29.4

Table 2: OLS Results

	I	II	III	IV	V
	Supply	Supply	Demand	Demand	z(m)
	logreimburse	logreimburse	loghhoop3	loghhoop3	loghhoop3
No Insurance	0.533	0.55			0.315
	[4.95]**	[5.11]**			[3.24]**
Employer Provided	0.182	0.204			0.162
	[2.39]*	[2.67]**			[2.35]*
Self Employed	0.625	0.634			0.538
	[3.69]**	[3.73]**			[3.52]**
Privately Purchased	0.482	0.469			0.357
	[4.32]**	[4.21]**			[3.55]**
dumrural		0.002	-0.069	-0.042	0.068
		[0.03]	[1.87]	[1.15]	[1.24]
logemp		0.108			0.21
		[0.38]			[0.83]
logest		0.067			-0.046
		[0.50]			[0.38]
logap		-0.166			-0.188
		[0.83]			[1.04]
loghpi		-0.323			-0.398
		[2.03]*			[2.75]**
loginscov		-0.04			-0.061
		[1.20]			[1.60]
logwork		-3.138			-5.095
		[1.07]			[1.92]
logmalp		-0.205			-0.157
		[3.37]**			[2.79]**
logpractice		0.755			1.316
		[1.02]			[1.92]
logune		-0.032			-0.04
		[0.55]			[0.77]
logreimburse			0.612	0.629	
			[46.83]**	[48.94]**	
loghhmedex3					0.503
					[29.23]**
loginsnotcov					0.006
					[0.22]
Constant	-2.183	0.47	3.103	3.199	5.552
	[30.68]**	[0.48]	[3.64]**	[3.82]**	[6.10]**
loghsize3			0.188	0.153	
			[4.43]**	[3.68]**	
logmage3			0.665	0.562	
			[3.51]**	[3.03]**	
loghhinc3			0.191	0.295	
			[8.11]**	[12.07]**	
dumhlth3				-0.418	
				[4.60]**	
dumhlth4				-0.693	
				[8.07]**	
dumhlth5				-0.879	
				[10.15]**	
dumhlth6				-0.977	
				[10.56]**	
Observations	3597	3597	3597	3597	3597
R-squared	0.01	0.02	0.4	0.43	0.21

The above table presents the results from an OLS estimation of the model.
The dependent variable in columns I and II is log*reimburse*, which is the natural log of ratio of household out of pocket expenditure and total household medical expenses. For columns III and IV, and IV the dependent variable is the natural log of household out of pocket expenditure
The omitted insurance category is medicare.
The absolute value of the t-statistics are reported in brackets.

Table 3: 3SLS Results

	I Supply logreimburse	II Demand loghhoop	III Supply logreimburse	IV Demand loghhoop
No Insurance	0.931		0.914	
	(9.18)**		(8.83)**	
Employer Provided	0.163		0.213	
	(2.18)*		(2.75)**	
Self Employed	0.492		0.636	
	(2.97)**		(3.69)**	
Privately Purchased	0.548		0.646	
	(5.04)**		(5.74)**	
logreimburse		0.192		0.426
		(2.61)**		(7.05)**
loghsize3		0.204		0.203
		(4.67)**		(4.63)**
logmage3		0.537		0.642
		(2.63)**		(3.17)**
dumhlth3		-0.498		-0.516
		(5.14)**		(5.39)**
dumhlth4		-0.779		-0.802
		(7.90)**		(8.42)**
dumhlth5		-1.008		-1.038
		(9.44)**		(10.27)**
dumhlth6		-1.101		-1.129
		(9.59)**		(10.39)**
dumrural		-0.017	-0.091	-0.022
		(0.4)	(1.45)	(0.51)
loghhinc3		0.336		0.351
		(14.15)**		(14.87)**
logune			0.065	
			(1.08)	
Constant	-2.123	2.069	1.898	1.947
	(30.49)**	(2.32)*	(1.85)	(2.19)*
logemp			0.219	
			(0.75)	
logest			0.049	
			(0.35)	
logap			-0.261	
			(1.25)	
loghpi			-0.578	
			(3.49)**	
loginscov			-0.024	
			(0.71)	
logwork			-2.158	
			(0.71)	
logmalp			-0.201	
			(3.19)**	
logpractice			0.807	
			(1.04)	
Observations	4218	4218	4218	4218

The absolute value of the z-statistics are reported in parentheses.
** indicates a variable is statistically significant at the 1% level, * indicates significance at the 5% level.
The omitted insurance category is medicare.

The above table gives the two sets of results from a 3SLS estimation of a system of supply and demand equations.
For example in the first 3SLS estimation the equations estimated are

Supply : $\log(\text{reimburse}) = \beta_0 + \beta_1\text{dumins1} + \beta_2\text{dumins2} + \beta_3\text{dumins3} + \beta_4\text{dumins5} + \varepsilon$

Demand : $\log(\text{hoop}) = \alpha_0 + \alpha_1\log(\text{reimburse}) + \alpha_2\log(\text{householdsize}) + \alpha_3\log(\text{age}) + \alpha_4\text{health3} + \alpha_5\text{health4} + \alpha_6\text{health5} + \alpha_7\text{health6} + \alpha_8\text{dumrural} + \alpha_9\log(\text{householdincome})$

Table 4A: Grid Search Results – OLS in First Stage

γ_1	γ_2	γ_3	Median θ	# Grid Intervals
0.05	2	0.4	591.667	60

Grid Search results are using z'(m) estimated from log-log OLS
These are the grid search results obtained using a grid with intervals spread over:
0-3 for the gamma parameters and 500 - 1000 for median theta

Table 4B: Grid Search Results – Local Linear Regression in First Stage

γ_1	γ_2	γ_3	Median θ	Objective Function	# Grid Intervals
0.6429	5.4	1.3929	3464.3	3692.4	30

RHS variables in local linear regression are hhmedex3, insurance category dummies and a rural dummy
Bandwidth was h=1.
These are the grid search results obtained using a grid with intervals spread over:
0-3 for the gamma 1 and gamma3, 0-8 for gamma2 and 3000 - 4000 for median theta

Table 5: Summary Statistics of Local Linear Regression Results

Variable	Obs	Mean	Std. Dev.	Min	Max
Constant	3849	1706.24	114.74	915.59	2169.02
Medical Expenditure	3849	0.10	0.04	0.00	0.27
No insurance	3849	-372.35	511.99	-6119.66	3089.88
Employer Provided Insurance	3849	-541.19	322.81	-4239.85	2189.35
Self Employed Insurance	3849	261.57	1441.40	-7844.35	10415.22
Medicare	3849	-803.97	386.90	-4463.09	540.89
Rural / Urban	3849	-238.06	415.80	-3790.95	4993.11

The above table gives the results from a local linear regression estimation of the supply side of the model
The dependent variable is *OOP Medical Expenses.*
Each individual's constant can be interpreted as their expected *OOP Medical Expenses* conditional on their observable characteristics.
Each individual's coefficient on the covariates can be interpreted as the derivative of this conditional expectation with respect to the covariate.

Table 7: Summary Statistics for recovered Health Shocks by Insurance Category

Insurance Category	Mean	Std. Dev.	Min	Max	25th percentile	75 Percentile
All Insurance Categories	7698.67	14181.45	-11302.76	117886.72	-885.88	10265.90
No Insurance	7162.22	12596.34	-6102.15	60563.27	-299.39	9108.92
Employer Provided	7191.63	13478.25	-11302.76	117886.72	-1031.74	9754.85
Self Employed	5307.98	10973.44	-6615.71	45319.93	-2178.15	8550.30
Privately Purchased	8393.42	15224.46	-7353.95	98144.03	-507.29	10953.50
Medicare	11326.68	18378.06	-11099.07	93652.47	-459.40	13595.62

Table 8: Kolmogorov-Smirnov Tests for Equality of Distributions of Health Shocks

Insurance Category	2	3	5	6
1	0.1208 (0.000)	0.2257 (0.002)	0.1093 (0.053)	0.1389 (0.001)
2		0.1208 (0.165)	0.1077 (0.006)	0.0970 (0.002)
3			0.2062 (0.007)	0.1789 (0.018)
5				0.0822 (0.197)

The combined Kolmogorov-Smirnov Statistic is reported with corrected P-Values in parenthesis

Table 9: Summary Statistics for Elasticity of Total Medical Expenditure wrt the Effective Price of Medical Expenditure by Insurance Category

Insurance Category	Mean	Std. Dev.	Min	Max	25th percentile	75 Percentile
All Insurance Categories	-0.56	0.82	-11.18	-0.02	-0.64	-0.12
No Insurance	-0.50	0.79	-9.33	-0.02	-0.58	-0.07
Employer Provided	-0.58	0.81	-11.18	-0.02	-0.65	-0.13
Self Employed	-0.68	0.84	-3.79	-0.03	-0.83	-0.17
Privately Purchased	-0.52	0.96	-9.02	-0.02	-0.61	-0.09
Medicare	-0.50	0.81	-6.67	-0.02	-0.57	-0.08

Elasticity of medical expenditure wrt the effective price of medical expenditure is calculated as:

$$\frac{\delta m}{\delta z'(m)} \times \frac{z'(m)}{m}$$

Applying implicit function rule to First Order Condition and assuming $z(m) = m \cdot z'(m) + cons$ yields:

$$\frac{\delta m}{\delta z'(m)} \times \frac{z'(m)}{m} = \frac{-\gamma_1 c^{-\gamma_1 - 1} m z'(m) - c^{-\gamma_1}}{\gamma_1 c^{-\gamma_1 - 1} z'(m)^2 + \gamma_2 \gamma_3 (m - \theta)^{-\gamma_3 - 1}} \times \frac{z'(m)}{m}$$

Table 10: Kolmogorov-Smirnov Tests for Equality of Distributions of Price Elasticities

Insurance Category	2	3	5	6
1	0.1627 (0.000)	0.2038 (0.006)	0.0947 (0.130)	0.0655 (0.383)
2		0.1123 (0.231)	0.1341 (0.000)	0.1545 (0.000)
3			0.1911 (0.015)	0.1901 (0.010)
5				0.0715 (0.345)

The combined Kolmogorov-Smirnov Statistic is reported with corrected P-Values in parenthesis

Table 11a : OLS Results from Elasticity Regressions

	I	II	III	IV
	Elasticity	Elasticity	Elasticity	Elasticity
dumins1	-0.12815	-0.13421	-0.11404	-0.11218
	[2.25]*	[2.36]*	[2.07]*	[2.04]*
dumins2	-0.1338	-0.16671	-0.1198	-0.10358
	[3.20]**	[3.89]**	[2.95]**	[2.63]**
dumins3	-0.2234	-0.26042	-0.18456	-0.16427
	[2.39]*	[2.77]**	[2.02]*	[1.81]
dumins5	-0.13908	-0.17903	-0.14273	-0.1228
	[2.33]*	[2.96]**	[2.43]*	[2.14]*
dumrural	-0.02513	-0.03078	-0.00336	8.99E-05
	[0.95]	[1.16]	[0.14]	[0.00]
hsize3	-0.01405	-0.01224	-0.00732	-0.00866
	[1.32]	[1.15]	[0.74]	[0.87]
mage3	-0.0054	-0.00499	-0.00368	-0.00408
	[2.28]*	[2.11]*	[1.67]	[1.85]
une	-0.00042	0.000758	0.00092	0.000376
	[0.09]	[0.16]	[0.21]	[0.09]
hhinc3	2.6E-06	2.32E-06	2.68E-06	2.8E-06
	[9.54]**	[8.39]**	[5.17]**	[5.42]**
hpi	0.000234	0.000227	0.000144	0.000144
	[0.68]	[0.66]	[0.45]	[0.45]
malp	0.010683	0.012353	0.006958	0.004915
	[0.39]	[0.45]	[0.27]	[0.19]
theta	-2.3E-05	-2.2E-05	-2.4E-05	-2.5E-05
	[26.75]**	[25.19]**	[2.19]*	[2.37]*
dumhlth3		0.045176	-0.00341	
		[0.74]	[0.06]	
dumhlth4		0.091861	0.024417	
		[1.58]	[0.45]	
dumhlth5		0.138627	0.044813	
		[2.36]*	[0.81]	
dumhlth6		0.249955	0.140275	
		[3.99]**	[2.38]*	
hhmedex3			-3.4E-05	-3.4E-05
			[3.24]**	[3.19]**
hhmedex3sq			6E-10	6E-10
			[21.97]**	[22.30]**
hhoop3			-2.4E-05	-2.3E-05
			[5.01]**	[4.91]**
Constant	1.028283	0.907561	1.082249	1.139932
	[5.89]**	[5.06]**	[6.44]**	[6.96]**
Observations	3735	3735	3735	3735
R-squared	0.19	0.2	0.31	0.3

The above table presents the results from an OLS regression of the absolute value of the price elasticity on various explanatory variables

The absolute value of the t-statistics are reported in brackets.

** indicates a variable is statistically significant at the 1% level, * indicates significance at the 5% level.

Table 11b : OLS Results from Elasticity Regressions

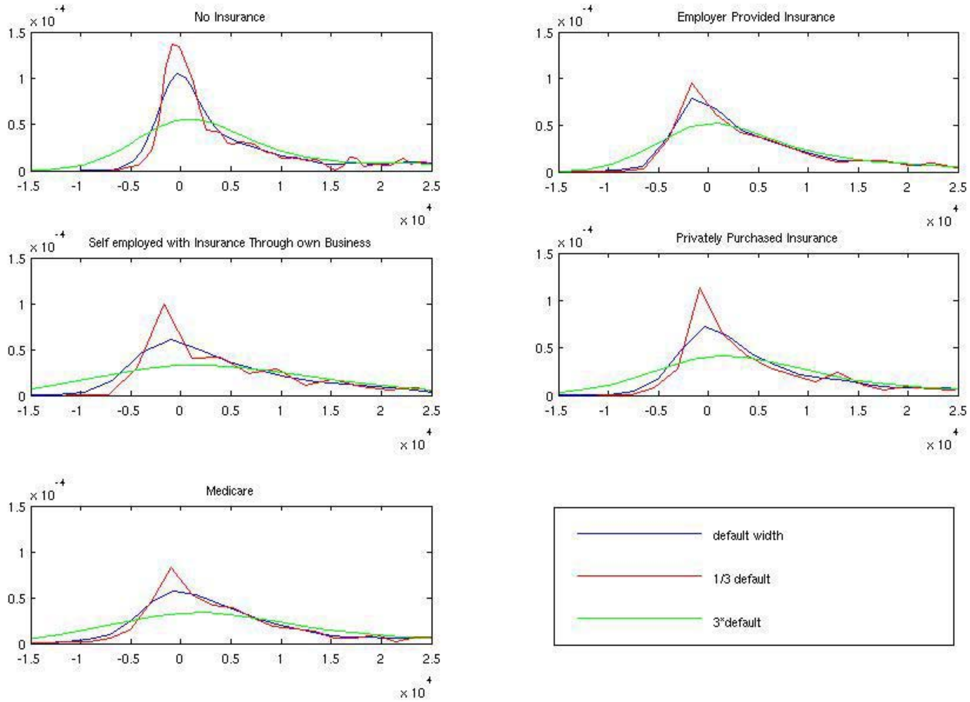
	I Log Elasticity	Log Elasticity	Log Elasticity	Log Elasticity
dumins1	-0.21284	-0.23185	-0.40458	-0.40275
	[3.93]**	[4.30]**	[39.45]**	[39.40]**
dumins2	-0.26147	-0.30393	-0.30705	-0.30426
	[6.38]**	[7.33]**	[39.37]**	[39.70]**
dumins3	-0.41682	-0.46974	-0.53236	-0.5299
	[4.67]**	[5.28]**	[32.77]**	[32.76]**
dumins5	-0.34417	-0.404	-0.49139	-0.4886
	[6.03]**	[7.01]**	[45.95]**	[46.37]**
dumrural	-0.05539	-0.0607	0.050656	0.050558
	[2.19]*	[2.41]*	[10.82]**	[10.81]**
loghsize3	-0.10427	-0.096	-0.00135	-0.00123
	[3.87]**	[3.58]**	[0.27]	[0.25]
logmage3	-0.32517	-0.32059	-0.0366	-0.03363
	[2.59]**	[2.56]*	[1.57]	[1.45]
logune	-0.02613	-0.01527	-0.00825	-0.00881
	[0.93]	[0.55]	[1.60]	[1.71]
loghhinc3	0.358099	0.32737	0.629229	0.62974
	[21.80]**	[19.38]**	[191.39]**	[198.98]**
loghpi	0.055096	0.048936	0.024124	0.024332
	[0.83]	[0.74]	[1.96]	[1.97]*
logmalp	0.024593	0.019203	0.015605	0.015944
	[0.89]	[0.70]	[3.09]**	[3.17]**
theta	-6E-05	-5.9E-05	1E-05	9.97E-06
	[73.02]**	[70.37]**	[37.44]**	[37.37]**
dumhlth3		0.154174	0.015736	
		[2.66]**	[1.44]	
dumhlth4		0.220448	0.022976	
		[3.95]**	[2.19]*	
dumhlth5		0.299081	0.018884	
		[5.28]**	[1.77]	
dumhlth6		0.370633	0.01465	
		[6.13]**	[1.29]	
loghhmedex3			-0.97423	-0.9744
			[297.24]**	[298.93]**
loghhoop3			-0.03743	-0.03746
			[21.61]**	[21.62]**
Constant	-3.15464	-3.06219	1.046792	1.046531
	[4.87]**	[4.74]**	[8.66]**	[8.68]**
Observations	3735	3735	3597	3597
R-squared	0.64	0.64	0.99	0.99

The above table presents the results from an OLS regression of the log of the absolute value of the price elasticity on various explanatory variables
The absolute value of the t-statistics are reported in brackets.
** indicates a variable is statistically significant at the 1% level, * indicates significance at the 5% level.

Table 12: Correlations Between Health Shocks and Elasticities by Insurance Category

Insurance Category	Correlation
All Insurance Categories	0.41
No Insurance	0.39
Employer Provided	0.43
Self Employed	0.51
Privately Purchased	0.34
Medicare	0.40

Figure 1: Density Estimates for Recovered Health Shock by Insurance Category



This figure plots estimated densities for the recovered health shock, theta, for each of the five insurance categories.
The estimated densities are calculated using a normal kernel. Each subplot estimates the density using three different bandwidths where the default bandwidth is the optimal bandwidth calculated in Matlab.

Figure 2: Density Estimates for Price Elasticities by Insurance Category

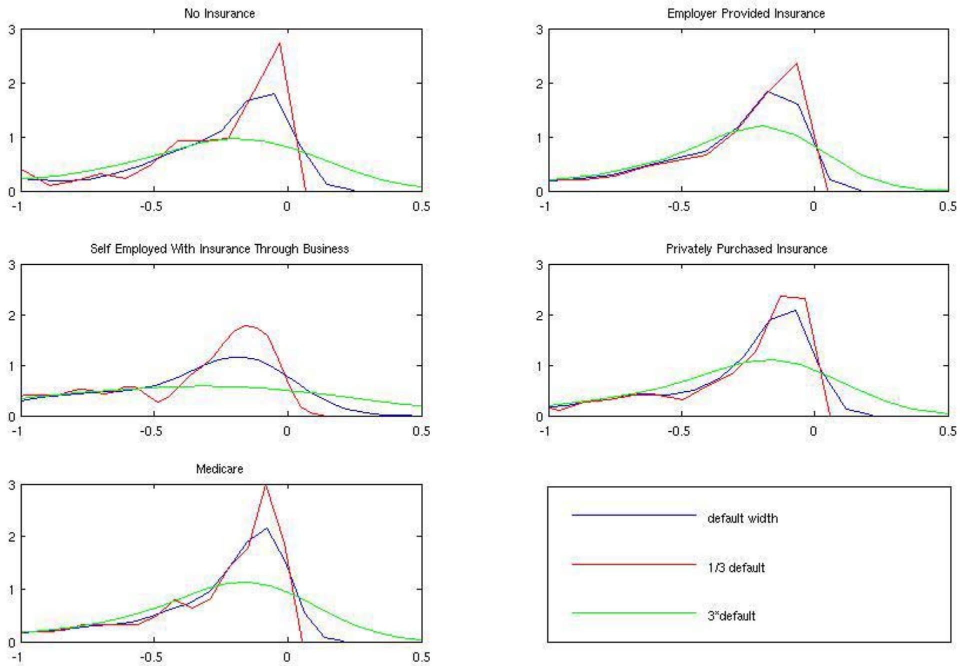
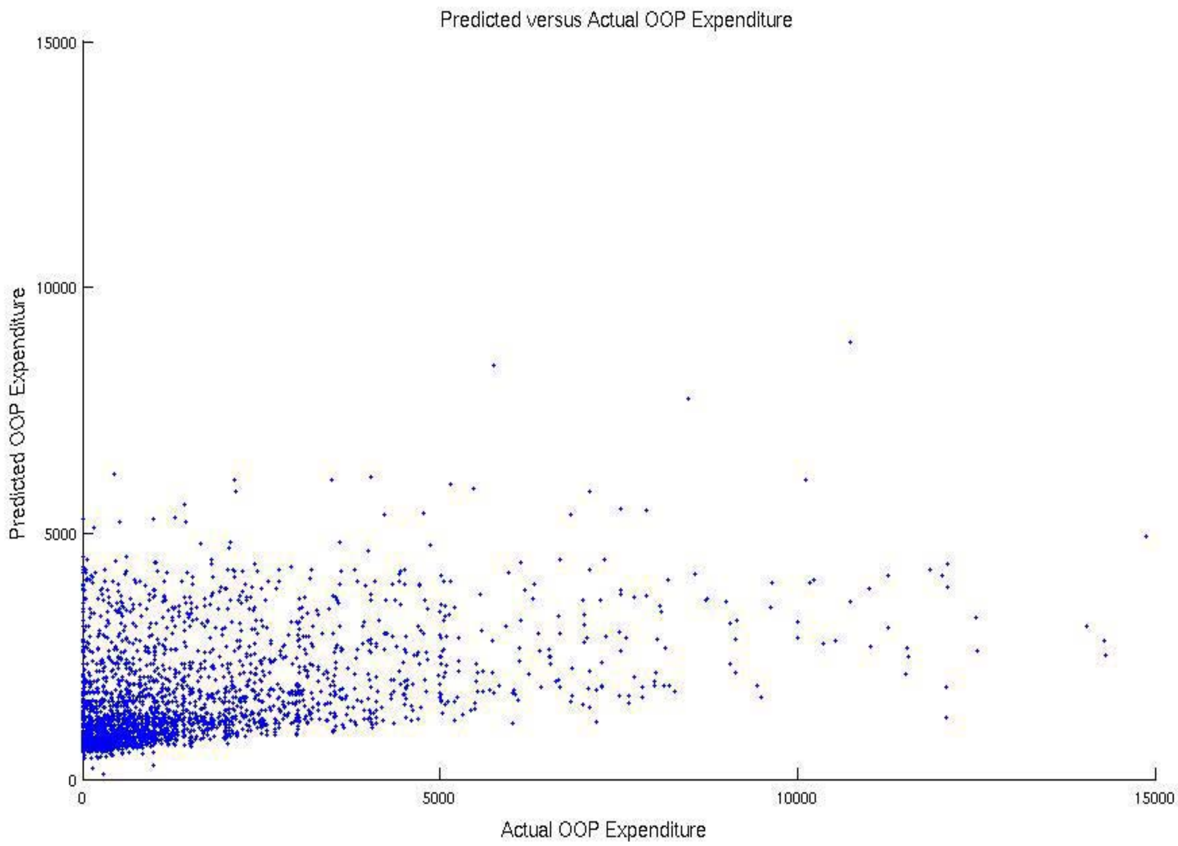
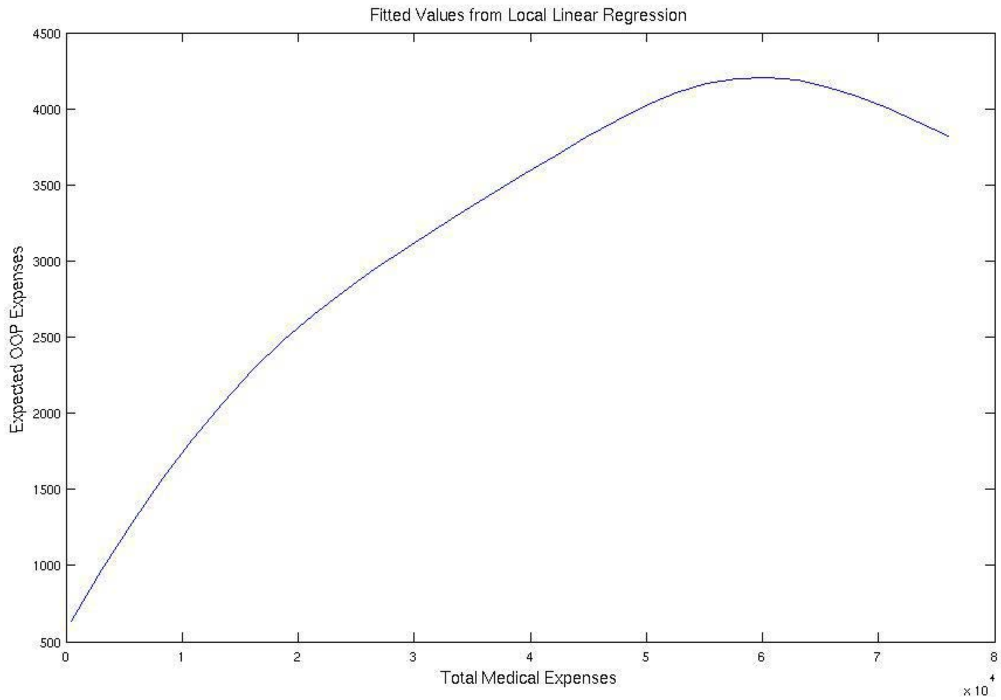


Figure 3:Scatter Plot of Predicted OOP Expenditure against Actual OOP Expenditure



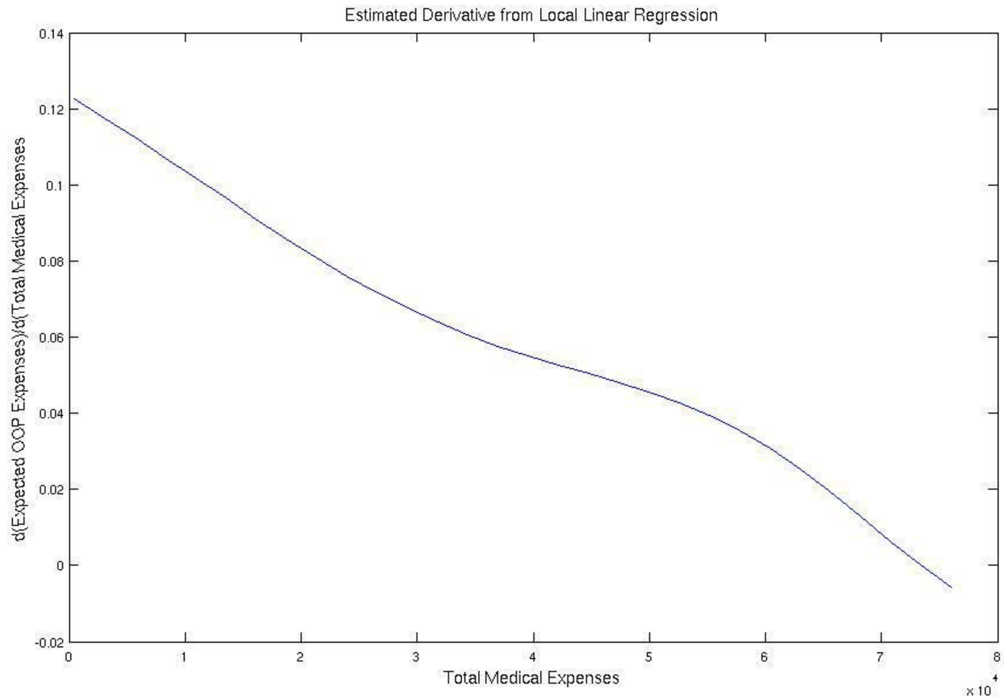
Predicted OOP Expenditure is calculated as the fitted values from a local linear regression of *OOP Expenditure* on *Total Medical Expenditure* and a series of dummy variables for insurance categories and rural location.

Figure 4: Expected OOP Medical Expenses conditional on Total Medical Expenses



Expected OOP Expenditure is calculated as the fitted values from a local linear regression of *OOP Expenditure* on *Total Medical Expenditure* and a series of dummy variables for insurance categories and rural location. Expected OOP Expenditure is calculated for 30 evenly spaced values of *Total Medical Expenditure* with the other covariates held at their means

Figure 5: Derivative of OOP Medical Expenses with respect to Total Medical Expenses



The derivative of Expected OOP Expenditure with respect to *Total Medical Expenditure* is calculated from a local linear regression of *OOP Expenditure* on *Total Medical Expenditure* and a series of dummy variables for insurance categories and rural location. The derivative of Expected OOP Expenditure with respect to *Total Medical Expenditure* is calculated for 30 evenly spaced values of *Total Medical Expenditure* with the other covariates held at their means

Figure 6: CDFs of Elasticity Distributions by Insurance Category

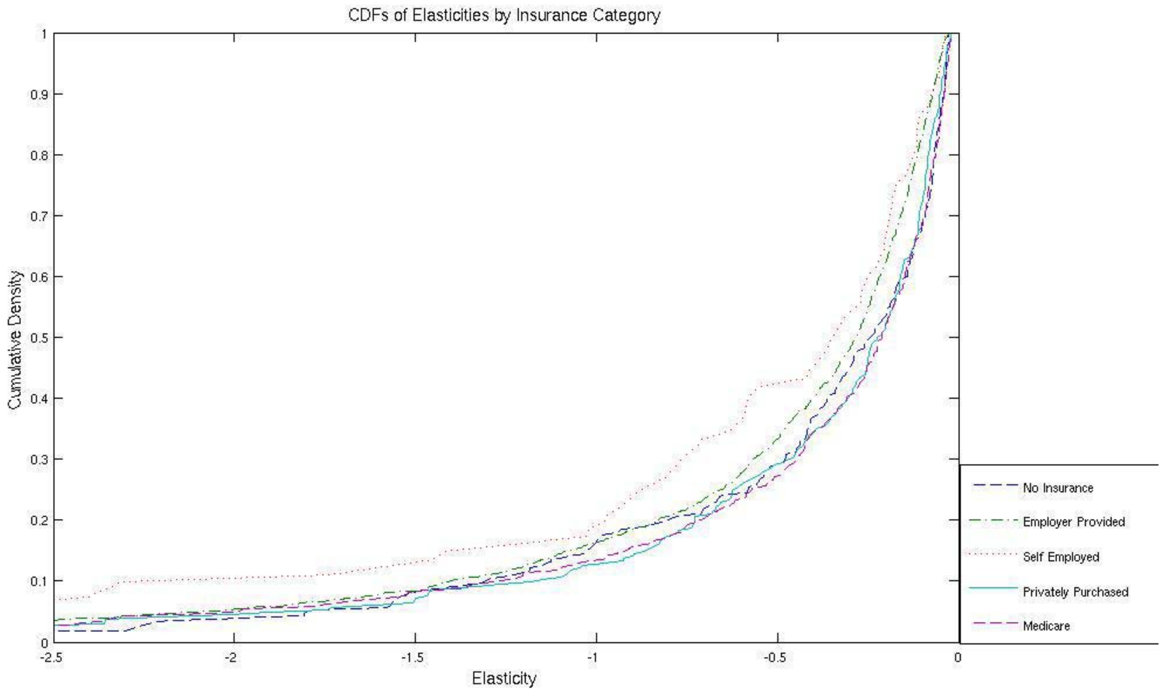


Figure 7: CDFs of Health Shock Distributions by Insurance Category

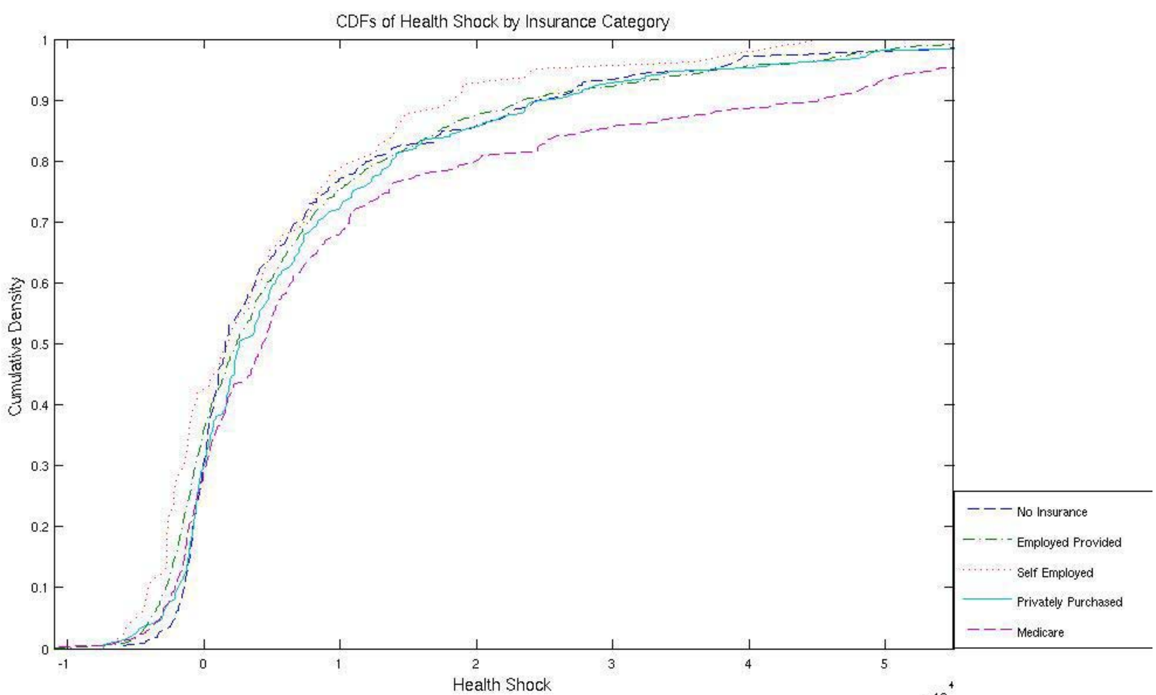


Figure 8: Scatterplots of Elasticities against Health Shocks by Insurance Category

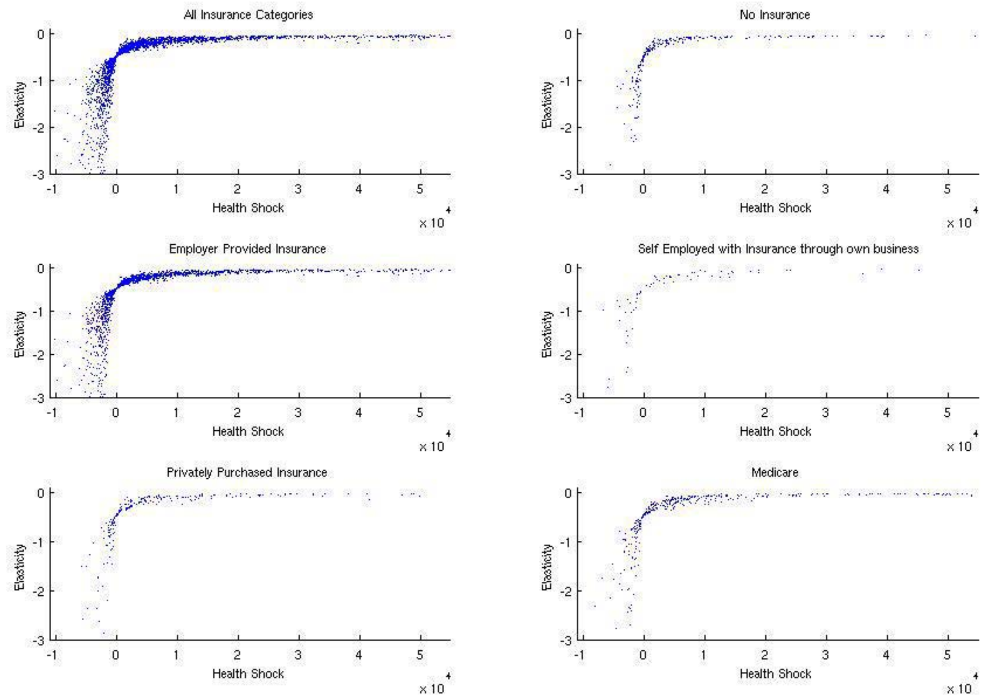


Figure 9: Kernel Regressions of Elasticity on Health Shock by Insurance Category