Adverse Selection, Moral Hazard and the Demand for Medigap Insurance

Michael Keane*
University of Oxford

Olena Stavrunova†
University of Technology, Sydney

November 2011

Abstract

The size of adverse selection and moral hazard effects in health insurance markets has important policy implications. For example, if adverse selection effects are small while moral hazard effects are large, conventional remedies for inefficiencies created by adverse selection (e.g., mandatory insurance enrolment) may lead to substantial increases in health care spending. Unfortunately, there is no consensus on the magnitudes of adverse selection vs. moral hazard. This paper sheds new light on this important topic by studying the US Medigap (supplemental) health insurance market. While both adverse selection and moral hazard effects of Medigap have been studied separately, this is the first paper to estimate both in an unified econometric framework.

We develop an econometric model of insurance demand and health care expenditure, where adverse selection is measured by sensitivity of insurance demand to expected expenditure. The model allows for correlation between unobserved determinants of expenditure and insurance demand, and for heterogeneity in the size of moral hazard effects. Inference relies on an MCMC algorithm with data augmentation. Our results suggest there is adverse selection into Medigap, but the effect is small. A one standard deviation increase in expenditure risk raises the probability of insurance purchase by 0.037. In contrast, our estimate of the moral hazard effect is much larger. On average, Medigap coverage increases health care expenditure by 32%.

Keywords: Health insurance, adverse selection, moral hazard, health care expenditure

JEL codes: I13, D82, C34, C35

*This research has been supported by Australian Research Council grant FF0561843 and by the ARC Centre of Excellence in Population Ageing Research (ARC grant CE110001029). But the views expressed are entirely our own.

†Corresponding author. School of Finance and Economics. University of Technology Sydney. PO Box 123, Broadway NSW 2007. Australia. Email: olena.stavrunova@uts.edu.au. T: +61295143597. F: +61295147711

1 Introduction

This paper studies adverse selection and moral hazard in the US Medigap health insurance market. Medigap is a collection of supplementary insurance plans sold by private companies to cover gaps in Medicare, the primary social insurance program providing health insurance coverage to senior citizens. While both the adverse selection and moral hazard effects of Medigap have been studied separately, this is the first paper to estimate both the selection and moral hazard effects of Medigap in an unified econometric framework.

One of the advantages of the Medigap market for studying adverse selection (a propensity of individuals with higher risk to purchase more coverage) is that it is relatively easy to identify what information about health expenditure risk is private to individuals. Because insurers can only price Medigap policies based on age, gender, state of residence and smoking status, expenditure risk due to other factors, including health status, can be considered private information of individuals for the purposes of the analysis.

The existence of private information is central to the analysis of insurance markets. Rothschild and Stiglitz (1976) show that if individuals have private information about their risk type, the competitive equilibrium (if it exists) is not efficient: adverse selection drives up premiums, so low-risk individuals remain underinsured. This suggests there may be scope for government intervention in insurance markets (e.g. mandatory social insurance financed by taxation). But the functioning of insurance markets can also be distorted by moral hazard, which is another type of informational asymmetry (Arrow (1963), Pauly (1968)). Moral hazard arises if ex-post risk of insured individuals is higher than the ex-ante risk. This occurs if insurance decreases incentives to avoid risky outcomes (or increases health care utilization conditional on health outcomes), by lowering health care costs to the insured.

Both adverse selection and moral hazard manifest themselves in a positive relationship between ex-post realization of risk and insurance coverage (Chiappori and Salanie (2000)).

But from a policy point of view the distinction between the two is very important. The same policies that can deal with adverse selection (e.g. mandatory enrolment) can lead to greatly increased aggregate health care costs if the moral hazard effect is strong.

Unfortunately, it is very challenging to isolate adverse selection and moral hazard empirically. While there are a large number of studies that examine these two effects in isolation, only a few attempt to address selection and moral hazard in the health insurance context in a unifying framework. Cutler and Zeckhauser (2000) review the literature that focuses on selection in health insurance markets and conclude that most studies find evidence in favour of adverse selection. These studies frequently utilize data from employers who offer different insurance plans to their employees, and examine risks across plans with different generosity. There is also empirical evidence that points to the importance of moral hazard. For example, Manning et al. (1987) use data from the RAND Health Insurance Experiment and find that individuals who were randomly given more generous plans had higher health care expenditure. Chiappori et al. (1998) document that an exogenous change in the generosity of health insurance coverage in France had an effect on some categories of health care expenditure. A large number of studies estimate substantial moral hazard effects of insurance on utilization of health care by employing parametric multiple equation models with exclusion restrictions (e.g., Munkin and Trivedi (2008, 2010), Deb et al. (2006)).

Only a couple of papers have estimated selection and moral hazard effects within a single structural model of health insurance choice and demand for health care. Cardon and Hendel (2001) was the first paper to adopt this approach. Using data from National Medical Expenditure survey, they find evidence of little adverse selection but of substantial moral hazard. But to estimate their model they rely on the restrictive assumption that the insurance choice set faced by an individual is exogenous. They also assume that the health shocks are lognormal. In contrast, recent papers by Bajari et al. (2011a,b) develop a semiparametric method for inference in a structural model of health insurance and health expenditure choice. They find evidence of substantial moral hazard and adverse selection in the HRS and in the insurance claims data from a large self-insured employer. However, while Bajari et al. (2011a,b) are flexible with respect to the distribution of expenditure risk, their framework is restrictive in that it does not allow for heterogeneity in risk preferences, or correlation of risk preferences with expenditure risk. Such features have been found to be important for explaining data regularities in several insurance markets (e.g., Fang et al. (2008), Finkelstein and McGarry (2006)). An extensive review of empirical studies of selection and moral hazard effects in other insurance markets is given in Cohen and Siegelman (2010).

We now consider prior work on the Medigap market in particular. The difficulty of disentangling selection and moral hazard effects empirically may be why existing studies of the Medigap market do not agree on their magnitudes. For example, Wolfe and Goddeeris (1991) find evidence of adverse selection and moral hazard in their 1977-1979 sample of Retirement History Survey respondents. In particular, they find that a one standard deviation health expenditure shock (i.e. the expenditure residual left after controlling for self-assessed health, disability, wealth and demographics) increases the probability of supplemental insurance by 3.3 percentage points in the first year, and by a further 7.8 percentage points in the following year. They also find that the moral hazard effect of supplemental insurance is a 37% increase in expenditure on hospital and physician services. Ettner (1997) also finds both adverse selection and moral hazard using the 1991 Medicare Current Beneficiary Survey (MCBS). In particular, she finds that total Medicare reimbursements of seniors who purchased Medigap plans independently were about 500 dollars higher than of those who received Medigap coverage through an employer. Assuming the former is a more selected group, this implies adverse selection. She also reported moral hazard effects of 10% and 28% of average total Medicare reimbursements for plans with lower and higher generosity of coverage, respectively. On the other hand, Hurd and McGarry (1997) find that the higher health care use by individuals with supplemental insurance in their 1993-1994 Asset and Health Dynamic Survey sample is mostly due to moral hazard, not adverse selection. Importantly, these studies only test for the presence of the adverse selection, rather than attempting to fully quantify it's effect.

Recently, Fang, Keane and Silverman (2008) (FKS) document *advantageous* selection into Medigap insurance. That is, seniors who purchase Medigap are (on average) in better health than those who have only Medicare. This finding contradicts the predictions of classic asymmetric information models of insurance markets (e.g. Rothschild and Stiglitz (1976)). These models predict that when individuals have private information about their risk type, the riskier types should be more likely to purchase insurance. But advantageous selection can arise if people are heterogeneous on dimensions other than risk type, and there exist unobservables that are positively correlated with both health and demand for insurance. Potential sources of advantageous selection (henceforth "SAS" variables for short) proposed by FKS include risk tolerance, income, education, the variance of health care expenditure, the interaction of risk tolerance and the variance of expenditure, financial planning horizon, longevity expectations and cognitive ability.

To test if these SAS variables explain advantageous selection, FKS first estimate an insurance demand equation that includes only pricing variables and expenditure risk. This yields the puzzling negative coefficient on expenditure. They then include the SAS variables, and test if the expenditure coefficient turns positive. To carry out such an analysis, one would ideally need a dataset which simultaneously contains information on health expenditure, insurance status and SAS variables for all respondents. However, as FKS point out, such a dataset does not exist. Instead, the following two datasets are available: the Medicare Current Beneficiary Survey (MCBS) which has information on health care expenditure and Medigap insurance status, but no information on risk tolerance or other SAS variables; and the Health and Retirement Study (HRS), which has information on a number of potential SAS variables as well as Medigap insurance status, but no information on health care expenditure. Both datasets have detailed demographic and health status characteristics. The empirical strategy of FKS is to first estimate the relationship between expenditure and demographic and health status characteristics using the MCBS. They then use the estimated relationship to predict expected health care expenditure in the Medicare only state for HRS respondents. This is their measure of health expenditure risk (in the absence of supplemental coverage). FKS then investigate how the relationship between Medigap insurance status and expenditure risk changes as potential sources of advantageous selection are added to the model.

FKS find that as more SAS variables are added to the insurance demand model, the relationship between Medigap status and expenditure risk turns from negative to positive. Thus, among individuals who are similar in terms of the SAS variables, it is indeed the less healthy who are more likely to buy Medigap insurance. This is just as classical asymmetric information models predict. Cognitive ability and income are found to be the most important SAS variables. Interestingly, risk tolerance was not very important - it affected demand but was not correlated with expenditure risk.[1]

The main limitation of FKS's analysis of adverse selection is they did not account for possibly non-random (conditional on observables) selection into insurance when estimating the prediction model for expenditure risk. To obtain the prediction equation for health expenditure, FKS estimate the following model by OLS using the MCBS:

$$E_i = \mathbf{H}_i\boldsymbol{\beta} + \gamma I_i + \varepsilon_i, \tag{1}$$

where E_i is expenditure, $\mathbf{H}_i$ is a vector of health measures and demographic characteristics,

[1]FKS propose three channels through which cognitive ability can affect demand for insurance: individuals with higher cognitive ability (i) may better understand the rules of Medicare and the costs and benefits of purchasing supplemental insurance; (ii) may have lower search costs; (iii) may be more aware of future health care expenditure risks. FKS also provide a brief discussion of informational policy interventions that might increase insurance coverage of high risk individuals in each of the three cases.

I_i is an indicator for Medigap coverage, and γ is the moral hazard effect of Medigap. Then, for HRS respondents, they predict total expected expenditure in the Medicare only state as follows:

$$\widehat{E}_i = \mathbf{H}_i \widehat{\boldsymbol{\beta}}.$$

They use $\widehat{E}_i$ as their measure of expenditure risk, and estimate the model for health insurance status in the HRS as:

$$I_i = \alpha_0 \widehat{E}_i + \mathbf{P}_i \boldsymbol{\alpha}_2 + \mathbf{SAS}_i \boldsymbol{\alpha}_3 + \eta_i. \tag{2}$$

Here $\mathbf{P}_i$ is a vector of variables that affect the price of Medigap insurance.[2] The degree of selection is captured by the sensitivity of insurance demand to expenditure risk, conditional on other variables (i.e. α_0 in equation 2).[3] However if ε_i is correlated with the insurance indicator I_i, then $\widehat{E}_i$ is an inconsistent estimate of expected total health expenditure in the Medicare only state, $\widehat{\gamma}$ is an inconsistent estimate of the moral hazard effect, and estimates of $\boldsymbol{\alpha}$ are inconsistent as well.

For example, if I_i and ε_i are negatively correlated (i.e. individuals with better unobserved health are more likely to buy insurance), the regression (1) will underestimate γ, and $\widehat{E}_i$ will overestimate the expected health care expenditure (in the Medicare only state) for individuals who actually have Medigap supplemental insurance. This will cause FKS to overstate the degree of advantageous selection (α_0 in model (2)), and to exaggerate the ability of the proposed SAS variables to explain the advantageous selection in the Medigap market.[4]

[2]Equation (2) can be interpreted as an insurance demand equation, in which $\widehat{E}_i$ is a measure of person's risk level. As Medicare only covers about 45% of costs, viewing expected total expenditure in the Medicare only state (of which one would have to cover 55% on average with supplementary insurance) as a measure of expenditure risk seems plausible. The implicit assumption here is that people can't predict if they are likely to need treatment that has a lower or higher coverage rate by Medicare.

[3]Note that in Medigap there may exist both selection on unobservables and selection on observables, because there are observables that insurance companies cannot legally price on (e.g., health status characteristics, race, etc.).

[4]Suppose there are individuals of low and high risk types, whose expenditure risk is equal to 1 and 5 thousand dollars, respectively. Also suppose that there is advantageous selection, i.e. each additional thousand dollars in risk decreases probability of supplemental insurance coverage by α. A random sample from this population is available, in which the proportions of uninsured and insured individuals are p_0 and

Unlike FKS, in this project we address the possibility of non-random selection into Medigap by explicitly modelling correlation between I_i and ε_i within a comprehensive model of demand for health insurance and health care expenditure. Our model for insurance demand and health care expenditure is a simultaneous equations model given by (1) and (2), where the parameters of interest (the selection and moral hazard effects) are identified via cross-equation exclusion restrictions. The key restrictions, apparent in (1) and (2), are (i) that the health status variables affect demand for insurance only through their effect on expenditure risk (not directly), and (ii) that selected demographic and behavioural characteristics (income, education, risk aversion, cognitive ability, financial planning horizon and longevity expectations) affect insurance demand but not expenditure risk (conditional on health status). That is, the SAS variables do not affect expenditure (conditional of health) except indirectly through their effect on insurance.

The first assumption appears plausible, as it is not clear why insurance demand would depend on health status measures per se, once one has conditioned on total expenditure risk. The second assumption, that **SAS** variables do not enter (1), also appears plausible given the very extensive set of health status controls we include in $\mathbf{H}_i$, but perhaps calls for further discussion. Although there is limited empirical evidence about the relationship between health care expenditure and the behavioral SAS variables (conditional on health status), what evidence there is seems consistent with our assumptions. FKS found no significant relationship between risk aversion and expenditure risk. Similarly, a recent paper by Fang et al. (2010) shows that in a large sample of HRS respondents the cross-sectional correlation between the total Medicare expenditure and cognitive ability largely vanishes when an extensive set of health status measures (similar to the ones utilized in this paper)

p_1 respectively. If expenditure risk is correctly measured then the relationship between risk and probability of supplemental insurance coverage can be estimated as $\frac{p_1 - p_0}{E_1 - E_0} = \frac{p_1 - p_0}{-4}$, which should be close to α if the sample size is large and expenditure risk is independent of other determinants of insurance status. However, if expenditure risk of the insured is incorrectly estimated to be equal to 2 thousand dollars (overstated), then the estimate of α will be equal to $\frac{p_1 - p_0}{-3}$, which will overstate the magnitude of advantageous selection.

are controlled for.[5]

As for income and education, our own analysis of the MCBS subsample suggests that these variables have little explanatory power for expenditure, conditional on other demographic and health controls. For example, when education and income are included in the expenditure equation which already contains $\mathbf{H}_i$ and I_i, the improvement in the R-squared, although statistically significant, is very modest (from 0.1649 to 0.1666). The effect of education is not statistically different from zero, and the effect of income is very small: e.g., an increase in income from the 10th to 90th percentile increases expenditure by \$281, which amounts to only a 3% increase from the sample mean level.[6] Hence, excluding income and education from the expenditure equation seems reasonable.[7]

In contrast to FKS, we combine information from the MCBS and HRS using multiple data imputation. To this end, we specify an auxiliary prediction model for SAS variables missing from the MCBS, conditional on exogenous variables common in the two datasets.[8] To deal with health expenditure data missing from the HRS, we use the expenditure distribution implicit in the joint model for insurance and expenditure. To capture the complex shape of

[5]A priori, it is tempting to think that higher cognitive ability people, who know more about medical conditions, will be more likely to seek treatment. But this is not at all clear. For example, if one understands that one can't really treat most viruses and that viruses usually just go away eventually, then one is less likely to waste time going to the doctor for virus-like symptoms. Similarly, the nature of the relationship between the expenditure and risk tolerance (conditional on health) is not at all clear ex ante. On the one hand, a more risk averse individual is probably more likely to seek treatment for a given health accident, but on the other, she may also know that treatments have risks, and may therefore want to avoid over-treatment. The results of FKS imply that these two effects roughly cancel.

[6]It is worth emphasizing, that the unconditional correlation between income and expenditure risk is large, but conditional on health it largely vanishes. That is, higher income people are healthier, and so tend to have lower expenditure. But they do not appear to demand more health care conditional on health.

[7]Also, as our model is cross-sectional, our specification implicitly assumes that the health status variables ($\mathbf{H}$) are exogenously given, and are not affected by health insurance status over time. That is, we assume that having insurance does not lead to a lower rate of investment in health, which causes health status to deteriorate over time. Under this dynamic scenario, we will underestimate the moral hazard effect (at least in the long run). However, Khwaja (2001) shows that in a dynamic model health insurance has two opposite effects. There is the ex-ante moral hazard effect, but there is also the "Mickey Mantle" effect: because insurance increases life expectancy, an individual has a greater incentive to invest in health. Khwaja finds that the two effects roughly cancel, so insurance has little effect on how health status evolves over time.

[8]We treat SAS variables as exogenous, so the model for insurance demand and expenditure is conditional on these variables. The auxiliary model for SAS variables is needed only for imputation of missing data.

the distribution of realized expenditure, which is positive and extremely skewed to the right, we employ a smooth mixture of Tobits (generalizing the smoothly mixing regressions (SMR) framework of Geweke and Keane (2007)). In the estimation we merge the two datasets, assume that the relevant variables are missing from the HRS and MCBS completely at random, and estimate the model using a MCMC algorithm with multiple imputations of the missing variables.[9]

Our approach to merging the two datasets can be described as "data fusion" - the combination of data from distinct datasets, which can have some variables in common as well as variables present in only one of the datasets. Rubin (1986) emphasized that the problem of data fusion can be cast as the problem of missing data, which, in turn, can be dealt with using Bayesian methods for multiple imputations from the posterior distribution of missing variables, conditional on common variables, as discussed in Gelman et al. (1995). This is the approach we adopt in this paper. Data fusion methods are often used in marketing to combine data from different surveys, such as product purchase and media exposure (e.g. Gilula et al. (2006)). Currently, there are few if any examples of data fusion in applied work in economics.

Our findings regarding selection confirm the main results of FKS - we find that income and cognitive ability are the most important factors explaining why higher-risk individuals are less likely to buy insurance. Both high income and high cognitive ability people tend to be (i) healthier and (ii) to demand more insurance conditional on health. But in addition to the SAS variables used in FKS, we also consider race and marital status as potential sources of adverse/advantageous selection. These variables can affect both tastes for insurance and health care expenditure, but cannot be legally used to price Medigap policies. We find that

[9]We will show below that the missing expenditure data (but not the missing SAS variables) can be integrated out analytically without complicating the MCMC algorithm for simulation from the posterior distribution of the parameters of the model. Therefore, we only have to perform multiple imputations of the SAS variables missing from the MCBS subset.

race is an important source of *adverse* selection: blacks and hispanics have both lower demand for Medigap insurance and lower health care expenditure. Overall, we find that, conditional on income, education, risk attitudes, cognitive ability, financial planning horizon, longevity expectations, race and marital status, there is adverse selection into Medigap insurance, but the effect is not very strong: a one standard deviation increase in expenditure risk in the Medicare only state increases the probability of buying insurance by only 3.7 percentage points (which is a 7.4% increase from the sample mean of Medigap coverage of 50%).

But we go beyond FKS in that our model allows estimation of the sample distribution of the effect of Medigap insurance on health care expenditure (i.e., the moral hazard effect). We find that, on average, an individual with Medigap insurance spends about $2,119 (32%) more on health care than his/her counterpart who does not have Medigap. The magnitude of this moral hazard effect is comparable to that found in the RAND Health Insurance Experiment. For example, Manning et al. (1987) find that decreasing the co-insurance rate from 25% to 0 increased total health care expenditure by 23%. The effect of adopting one of many typical Medigap insurance plans that cover co-pays is at least as big as this drop in the co-insurance rate,[10] and we see that it has a somewhat larger effect on expenditure. The moral hazard effect of Medigap varies with individual characteristics. In particular, it is lower for healthier individuals as well as for blacks and Hispanics, and it is largest in the New England region and smallest in the Pacific Coast region.

This paper is organized as follows. Section 2 describes the datasets used in the analysis; section 3 presents a model of the demand for Medigap insurance and health care expenditure and discusses an MCMC algorithm developed for Bayesian inference in this model; section 4 discusses the empirical results; section 5 concludes.

[10]For example, the average out of pocket expenses of individuals with Medigap coverage is about 1.8 thousand dollars (Kaiser Family Foundation 2005), which corresponds to about 23% of total health care expenditure in our data. This implies that on average adopting a Medigap policy decreases co-insurance by 32 percentage points, from 55% (co-insurance with Medicare only) to 23%.

2 Data: HRS and MCBS

While Medicare is the primary health insurance program for most seniors in the USA, on average it only covers about 45% of health care costs of beneficiaries. Medicare consists of two plans: plan A provides hospital insurance coverage, while plan B provides insurance for some physician services, outpatient services, home health services and durable medical equipment. Most beneficiaries are enrolled in both plans A and B. To cover the large gaps in Medicare, private companies offer Medigap insurance plans - private policies which cover some of the co-pays and deductibles associated with Medicare as well as expenses not covered by Medicare.[11] The Medigap market is heavily regulated - only 10 standardized Medigap plans are offered, and insurers can only price policies based on age, gender, smoking status and state of residence. They cannot use medical underwriting during six months after an individual is both at least 65 years old and is enrolled in Medicare plan B. Other institutional details of the Medigap market can be found in FKS. Medigap insurance status in our analysis is defined as equal to one if an individual purchases any additional private policy secondary to Medicare.

Our analysis uses data from the Medicare Current Beneficiary Survey (MCBS, years 2000 and 2001) and the Health and Retirement Study (HRS, year 2002). The MCBS contains comprehensive information about respondents' health care costs and usage, as well as detailed information about their health, demographic and socioeconomic characteristics. The HRS contains detailed information about health, demographic and socioeconomic characteristics as well as measures of risk attitudes, financial planning horizon, longevity expectations and

[11]For example, the basic Medigap plan A only covers Medicare parts A and B co-insurance costs, 365 additional hospital days during life time and blood products. In contrast, the most popular Medigap plan F, which is purchased by 37% of individuals with Medigap, additionally covers all Medicare plans A and B deductibles, part B balance billing, skilled nursing facility co-insurance and foreign travel emergency expenses. However, this plan does not cover the costs of preventative, home recovery or hospice care not covered by Medicare (Kaiser Family Foundation 2005). During the period of our study Medicare did not cover prescription drugs, and several Medigap plans offered partial prescription drugs coverage.

cognitive ability. The data used in the analysis includes only individuals covered by basic Medicare. Descriptive statistics for selected variables are presented in Table 2. We use the same MCBS sample as FKS, and the same HRS sub-sample used by FKS to obtain column (3) of Table 6 in their paper.[12] This is the sub-sample in which all individuals have non-missing information about all potential SAS variables, including risk aversion, financial planning horizon, cognitive ability and longevity expectations.

Our measure of risk attitude is the risk tolerance parameter estimated by Kimball et al. (2008) for all HRS respondents using their choices over several hypothetical income gambles. The variables which measure cognitive ability (one of the important SAS variables) in FKS include the Telephone Interview for Cognitive Status score, the word recall ability score, the numeracy score and the subtraction score. To decrease the number of auxiliary variables in our model we extract a common factor from these variable and use it as a scalar measure of cognitive ability in our analysis. We also use factor analysis do reduce the number of health status variables. Both datasets contain 76 health status measures which are detailed in the Data Appendix of FKS. These characteristics include self-reported health, smoking status, long-term health conditions (diabetes, arthritis, heart disease, etc.) and difficulties and help received for Instrumental Activities of Daily Living (IADLs). We use factor analysis to reduce these 76 variables to ten factors that best explain expenditure.[13]

The results of regressions of expenditure on different sets of health status characteristics

[12]FKS used three samples from the HRS in their analysis: (i) the full sample of 9973 observations, all of which have information on health, demographics and socioeconomic variables, but can have missing data on risk tolerance and other SAS variables; (ii) the subsample of 3467 observations which have information on risk-tolerance but not other SAS variables; (iii) the subsample of 1695 observations with information on all potential SAS variables. In our analysis we use the third HRS subsample.

[13]We first factor-analyze these 76 variables to extract 38 factors (using data in both the HRS (full sample) and MCBS samples) and then regress the health care expenditure in the MCBS on demographic characteristics and these 38 factors to select factors which are significant predictors of expenditure. We identify 16 such factors. We then select 10 factors out of these 16 such that the chosen 10 factors produce the highest possible adjusted regression R-squared (among all possible 10 factor subsets of the 16 factors). The factors that are selected are # 2, 3, 7, 8, 10, 11, 17, 20, 22 and 23 (not factors 1-10). Thus, the factors that explain the most covariance of the health indicators are not the same as the ones that explain most of the variance in expenditure.

Table 1: OLS results of total medical expenditure on Medigap coverage, demographic and health status characteristics in the MCBS

Variable	A. Without Health Controls	B. With Direct Health Controls	C. With Health Factor Controls
Medigap	979.4***	1951.2***	1948.2***
	(291.0)	(255.6)	(257.8)
Female	-933.6***	-834.7***	-734.3***
	(304.9)	(290.7)	(282.3)
Age-65	501.5***	408.0***	437.3***
	(125.8)	(115.1)	(116.5)
$(\text{Age-65})^2$	-23.3**	-28.8***	-31.0***
	(9.8)	(9.1)	(9.2)
$(\text{Age-65})^3$	0.43**	0.50**	0.51***
	(0.21)	(0.20)	(0.20)
Black	1212.9*	579.8	770.4
	(639.3)	(550.3)	(596.2)
Hispanic	-576.7	-843.8*	-622.2
	(511.7)	(431.6)	(467.4)
Married	-779.9***	-325.2	-213.5
	(299.0)	(268.7)	(275.3)
Health factor 2			4565.0***
			(252.4)
Health factor 3			-2544.6***
			(226.4)
Health factor 7			2049.0***
			(241.5)
Health factor 8			711.7***
			(213.1)
Health factor 10			-2047.0***
			(535.5)
Health factor 11			-961.6***
			(207.8)
Health factor 17			1176.3
			(931.4)
Health factor 20			-1339.2***
			(363.7)
Health factor 22			2144.6***
			(382.4)
Health factor 23			1254.7***
			(414.1)
Health status dummy	No	Yes	No
Region dummy	Yes	Yes	Yes
Year dummy	Yes	Yes	Yes
Observations	14128	14128	14128
Adjusted R^2	.017	.21	.18

Note: "Total medical expenditure" includes all expenditure, both covered and out-of-pocket. The regressions are weighted by cross-section sample weights. Robust standard errors clustered at the individual level are in parentheses. Statistical significance is indicated by * (10 percent), ** (5 percent) and *** (1 percent).

Table 2: Descriptive Statistics

Variable	MCBS All	MCBS Medigap	MCBS No Medigap	HRS All	HRS Medigap	HRS No Medigap
Medigap	0.50	1.00	0	0.43	1.00	0
Female	0.59	0.60	0.58	0.56	0.58	0.55
Age	76.57	77.02	76.11	68.70	68.67	68.72
	(7.50)	(7.29)	(7.69)	(3.10)	(2.98)	(3.20)
Black	0.10	0.04	0.17	0.14	0.06	0.20
Hispanic	0.08	0.03	0.12	0.07	0.02	0.11
Married	0.48	0.54	0.43	0.66	0.71	0.63
Education: Less than high school	0.36	0.27	0.45	0.28	0.22	0.33
Education: High School	0.27	0.31	0.24	0.38	0.41	0.35
Education: Some college	0.21	0.24	0.18	0.18	0.18	0.17
Education: College	0.08	0.10	0.06	0.08	0.08	0.08
Health factor 2 (Unhealthy)	0.04	-0.06	0.13	-0.32	-0.37	-0.28
	(1.01)	(0.89)	(1.10)	(0.51)	(0.43)	(0.56)
Health factor 3 (Healthy)	-0.12	-0.09	-0.15	0.17	0.23	0.13
	(-0.93)	(0.97)	(0.86)	(0.72)	(0.70)	(0.74)
Cognition				0.46	0.54	0.40
				(0.31)	(0.25)	(0.33)
Risk tolerance				0.234	0.228	0.236
(estimate from Kimball et al. (2008))				(0.142)	(0.138)	(0.146)
Financial planning horizon				4.46	4.83	4.18
(years)				(4.05)	(4.12)	(3.98)
Praliv75				67.32	69.57	65.59
(subjective probability to live to 75 or more)				(28.33)	(25.91)	(29.96)
Total medical expenditure*	8,085	8,559	7,605			
	(14,599)	(14,301)	(14,881)			
Number of observations	14128	7113	7015	1671	726	945

Note: "Total medical expenditure" includes all expenditure, both covered and out-of-pocket. Standard deviations are in parenthesis.

are presented in Table 1. Note that demographics explain only 1.7% of the variance of expenditure, but the inclusion of the 76 health measures increases this to 21%. When the 76 health status characteristics are replaced by our ten health factors, the adjusted regression R-squared drops from 0.21 to 0.18. This appears to be a reasonable price for reducing the number of covariates by 66. Health factors 2 and 3 turn out to be the most quantitatively important for predicting expenditure. Health factor 2 loads heavily on deterioration in health as well as difficulties and help with IADLs, and so is an unhealthy factor. It increases expenditures by about $4,500 per one standard deviation. Health factor 3 loads positively on good and improving self-reported health and negatively on difficulties with IADLs and thus is a healthy factor. It decreases expenditure by $2500 per one standard deviation.

Table 2 shows descriptive statistics for the HRS and the MCBS sub-samples. It can be seen that individuals in the HRS subsample are younger and healthier (have lower sample averages of unhealthy factor 2 and higher sample averages of healthy factor 3) than those in the MCBS subsample. The HRS data is used in our analysis as a source of information about behavioral SAS variables, such as risk tolerance, cognition, longevity expectations and financial planning horizon. From the HRS data we estimate the distribution of these SAS variables, conditional on exogenous characteristics common in the two datasets, and use it to impute the missing SAS variables in the MCBS sub-sample. The fact that the two subsamples have different characteristics does not create a problem for our analysis, provided the distribution of the SAS variables conditional on the exogenous characteristics used for imputation (including age and health) is the same in both subsamples.

Tables 1 and 2 suggest the presence of both advantageous selection and moral hazard. Table 2 shows that individuals with Medigap coverage are on average healthier than those without Medigap in both the HRS and the MCBS data (i.e. individuals with Medigap have lower values of unhealthy factor 2 and higher values of healthy factor 3 in both subsamples), while Table 1 shows that individuals with Medigap coverage spend more on health care than those without Medigap, both with and without conditioning on observed health status measures.[14] The Medigap coefficient *increases* when we add health status controls, stressing the *positive* correlation between health and Medigap coverage already evident in Table 2. We will investigate the magnitudes of the selection and moral hazard (or incentive) effects in the subsequent sections.

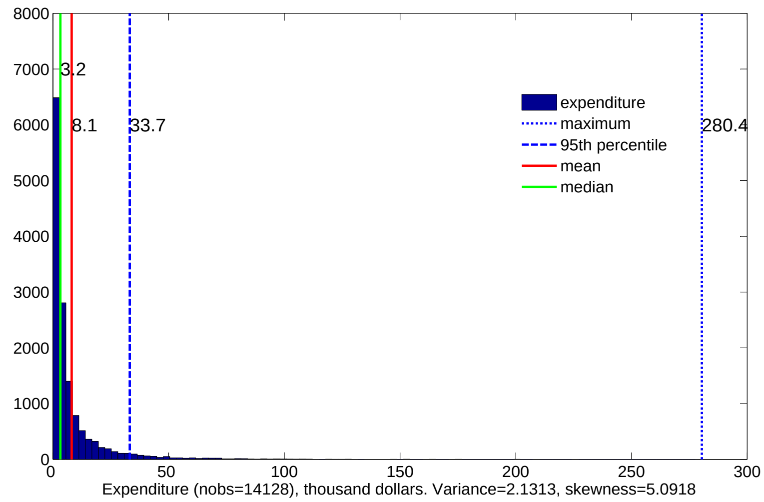
3 The Model

This section presents a model for the joint determination of insurance status and health care expenditure, in which we account for endogeneity of insurance by allowing the unobservable determinants of insurance status and expenditure to be correlated. But before developing the full model we first need to select a specification for the distribution of medical expenditure. It is well-known that econometric modelling of health care expenditures is challenging because of the properties of their empirical distribution. In particular, health care expenditures are non-negative, highly skewed to the right and have a point mass at zero. The histogram in Figure 1 shows that the empirical distribution of total health care expenditure of Medicare beneficiaries in our MCBS sample exhibits all these characteristics. The sample skewness is about 5.1 and the distribution has a long right tail. The proportion of observations with zero expenditure is about 0.025.

The literature on modelling health care expenditure has mainly focused on the problem of modelling it's conditional *expectation* in the presence of skewness and a mass of zero outcomes (e.g., Manning (1998); Mullahy (1998); Blough et al. (1999); Manning and Mullahy (2001);

[14] This is different from Table 2 of FKS, in which the Medigap coefficient changes from negative to positive as health controls are added to the insurance equation. The reason for the discrepancy is that FKS use different subsamples for regressions with and without health controls. In particular, the regression without health controls uses 15,945 observations, while the regression with health controls uses 14,129 observations (for which health status information is available) out of these 15,945. Table 1 in our paper uses the FKS sample of 14,129 observations to obtain the results with and without health controls. Hence, the 1,816 observations not used by FKS in the second regression have higher expenditure and lower Medigap coverage than the general Medicare population.

Figure 1: Histogram of total health care expenditure



Buntin and Zaslavsky (2004); Gilleskie and Mroz (2004); Manning et al. (2005)). The problem of modelling the entire conditional *distribution* of health care expenditure is less frequently addressed. When the context requires a probability model for expenditure, the preferred approach is a two-part model where the positive outcomes (the second part) are modelled using the log-normal distribution (e.g. Deb et al. (2006)). But because we are interested in the effect of the *level* of expenditure risk on Medigap insurance status, we prefer to model the level of expenditure rather than it's logarithm.

After trying several specifications of the distribution of expenditure, we decided to adopt a discrete mixture of Tobits in which the probability of a mixture component depends on an individual's observed characteristics. Because this model is a generalization of the Smoothly Mixing Regressions (SMR) framework of Geweke and Keane (2007) to the case of a Tobit-type limited dependent variable, we call it SMTobit (for Smooth Mixture of Tobits). With the appropriate number of mixture components, this specification can capture both skewness and non-negativity of the expenditure distribution, and provides a very good fit to various aspects of the conditional distribution of total health care expenditure in our MCBS sample, including conditional (on covariates) mean, variance, quantiles and probability of an extreme outcome. In section 4 we will discuss how the number of components was selected and examine the fit of the model to the distribution of expenditure.

In the next section we present the full specification of the model for insurance status and expenditure, where the insurance equation includes all potential sources of advantageous selection. We first present the model abstracting from the fact that not all variables of interest are available in both datasets and then discuss our approach to dealing with variables missing from the HRS or MCBS.

3.1 Complete data

We assume there are m types of individuals (types are indexed by j, $j = 1, \ldots, m$). A person's type is private information, i.e. individuals know their type, but from the point of view of the researcher these types are latent: given an individual's observable characteristics (i.e. demographics and health status) only her probability of belonging to type j can be inferred. Types differ in mean expenditure, in the effects of exogenous characteristics and insurance status on health care expenditure, as well as in the variance of expenditure.

Let I_i^* denote the utility that individual i derives from health insurance and let E_i^* denote her total expected health care expenditure if she remains without Medigap. As discussed in section 1, we assume that E_i^* is the expenditure risk relevant when individual i decides whether to purchase Medigap insurance, so henceforth we will refer to E_i^* as "expenditure risk". Both I_i^* and E_i^* are known to the individual but are unobserved by the econometrician, so they enter the model as latent variables.

Let I_i denote a binary variable which is equal to one if individual i has health insurance, and is equal to zero otherwise, and assume that $I_i = 0$ if $I_i^* < 0$ and $I_i = 1$ if $I_i^* >= 0$. Also, let $\widehat{E}_i$ denote *notional* health care expenditure of individual i (as in "notional demand", which can be negative). We assume that $\widehat{E}_i$ is determined as follows:

$$\widehat{E}_i|j = E_i^*|j + \gamma_j I_i + \eta_i|j \tag{3}$$

where γ_j denotes type-specific effect of health insurance on the notional health care expenditure (i.e. the price or moral hazard effect), and $\eta_i|j$ is the forecast error of individual i. Given the individual's type j, the forecast error $\eta_i|j$ is normally distributed with zero mean and variance σ_j^2:

$$\eta_i|j \sim N(0, \sigma_j^2)$$

The term σ_j^2 denotes the variance of actual expenditure around the expected expenditure risk (conditional on the insurance status) of an individual of type j. Thus σ_j^2 can be interpreted as the variance of the health care expenditure forecast error (i.e. η_i is a surprise health shock to individual i).

The *realized* expenditure $E_i|j$ is given by:

$$E_i|j = max\{0, \widehat{E}_i|j\}. \tag{4}$$

Hence, conditional on type j the model for the realized expenditure E_i is a Tobit. This specification ensures that the model does not predict negative expenditure for some individuals. Because in our data only 2.5% of observations have zero expenditure, the notional expenditure $\widehat{E}$ is equal to the realized expenditure E for 97.5% of the sample.

The model for the latent vector $[I_i^*, E_i^*]'$, conditional on type j, is specified as follows:

$$I_i^*|j = \alpha_0 E_i^*|j + \alpha_1 \sigma_j^2 + \alpha_2 \sigma_j^2 \cdot c_{1i} + \boldsymbol{\alpha}_3' \mathbf{xi}_i + \boldsymbol{\alpha}_4' \mathbf{c}_i + \varepsilon_{1i} \quad (5)$$

$$E_i^*|j = \boldsymbol{\beta}_j' \mathbf{xe}_i + \varepsilon_{2i}, \quad (6)$$

where the vector of disturbances $\boldsymbol{\varepsilon}_{12i} = [\varepsilon_{1i}, \varepsilon_{2i}]'$ is independent of η_i and follows a bivariate normal distribution:

$$\boldsymbol{\varepsilon}_{12i}|j \sim BVN\left(\mathbf{0}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}\right) \qquad \text{for all types } j = 1, \ldots, m.$$

The expenditure risk E_i^* consists of a part which depends on observable health status and demographics ($\boldsymbol{\beta}_j' \mathbf{xe}_i$) and a part which depends on unobservable characteristics (ε_{2i}). The disturbances ε_{1i} and ε_{2i} capture the heterogeneity in tastes for insurance and in health status, respectively, that are known to an individual, but not to the econometrician. We allow for ε_{1i} and ε_{2i} to be correlated with covariance given by σ_{12}.

In equations (5) and (6), $\mathbf{c}_i$ includes variables present in the HRS only (risk tolerance c_{1i}, financial planning horizon, cognition and longevity expectation), $\mathbf{xi}_i$ includes insurance pricing variables (age, gender, location of residence) as well as income, education, ethnicity and marital status, and $\mathbf{xe}_i$ includes demographic characteristics (age, gender, location of residence, marital status, race and ethnicity) and the ten health factors discussed in section 2. The variables $\mathbf{xi}_i$ and $\mathbf{xe}_i$ are present in both datasets.

There is heterogeneity in the effect of observable health status and demographic characteristics on expenditure risk because the $\boldsymbol{\beta}_j$ differ across different types of individuals. This is the smooth mixture of Tobits (SMTobit) described in the previous section. This specification allows for different marginal effects of covariates on expenditure for individuals with different health status (both observable and unobservable). As we will show in Section 4, a model with 5 latent types ($j = 1, ..., 5$) provides a very good fit to the data.

The parameter α_0 measures the effect of expenditure risk E_i^* on insurance demand when influences of other determinants of insurance status (including unobservables ε_{1i}) are held constant. A negative α_0 indicates advantageous selection, while a positive value indicates adverse selection. We also introduce the variance of the forecast error σ_j^2 and its interaction with risk tolerance into the insurance equation to capture that demand for insurance depends not only on expected expenditure but also on the variance of expenditure.

The model in equations (3)-(6) can be viewed as a simultaneous equations model where the parameters of interest (i.e., the selection and moral hazard effects) are identified via cross-equation exclusion restrictions. The restriction which allows us identify the selection effect α_0 is that the health status variables affect demand for insurance only through their effect on expenditure risk (not directly). If the health status variables were included in the insurance equation, we would not be able to isolate the effect of the expenditure risk α_0 from the independent effect of health status variables on insurance demand.

To identify the moral hazard effect we impose the restriction that selected demographic and behavioral characteristics (income, education, risk aversion, cognitive ability, financial planning horizon and longevity expectations) affect insurance demand but not expenditure risk (conditional on a rich set of health measures). Thus, these variables induce exogenous variation in the insurance choice conditional on expenditure risk E_i^*.[15] This permits us to

[15] Our approach to modelling health care expenditure and Medigap insurance status is related to that of Munkin and Trivedi (2010) (MT), who study the effect of supplemental drug insurance on drug expenditures. MT also used a discrete mixture model with covariate-dependent type probabilities to model drug expenditures, and they allow for correlation between unobservable determinants of drug expenditure and supplemental drug insurance status. However, our paper is quite different from theirs in a number of ways: (i) most obviously, we study a different market (i.e., Medigap supplemental insurance vs. drug coverage); (ii) MT only use the MCBS, while we merge the MCBS with the HRS in order to study effects of SAS variables, thus extending the application of MCMC methods to a rather novel selection/data fusion exercise; (iii) as MT note (see their conclusion), the expenditure distribution that they assume could be improved upon, and we do this by using the SMTobit specification, which turned out to be a very substantial improvement (see Keane and Stavrunova (2010)); (iv) we use a richer set of instruments for insurance status (not just price shifters but also the SAS variables); and (v) we use a much richer set of health status variables in the expenditure equation (this is made feasible by our factor analysis procedure). More importantly, MT only

account for endogeneity of insurance choice when estimating the parameters of the model. This, in turn, allows us to consistently estimate the moral hazard effect γ_j and the correlation between ε_{1i} and ε_{2i}.

To impute the missing $\mathbf{c}_i = [c_{1i}, ..., c_{4i}]'$ in the MCBS data (i.e. the 4 SAS variables) we specify an auxiliary model for $\mathbf{c}_i$ conditional on the exogenous variables common in the MCBS and HRS datasets. We assume the following relationship between $\mathbf{c}_{ki}$ and these exogenous variables:

$$c_{ki}|j = \mathbf{xc}_i'\boldsymbol{\lambda}_k + \varepsilon_{3ki}, \tag{7}$$

where $k = 1, \ldots, 4$. Here $\mathbf{xc}_i$ denotes the vector of exogenous variables common in the two datasets, such as demographics, income, health status and education.[16] The disturbances $[\varepsilon_{31i}, \ldots, \varepsilon_{34i}]' \equiv \boldsymbol{\varepsilon}_{3i}$ follow a multivariate normal distribution for all types $j = 1, \ldots, m$:

$$\boldsymbol{\varepsilon}_{3i}|j \sim N(0, V_c).$$

The disturbances $\boldsymbol{\varepsilon}_{3i}$ are independent of $\boldsymbol{\varepsilon}_{12i}$ and $\eta_i|j$. Hence,

$$\mathbf{c}_i|j = XC_i\Lambda + \boldsymbol{\varepsilon}_{3i}, \tag{8}$$

measure selection on unobservables, but what one needs to know for policy purposes also includes selection on *observables* that cannot (legally) be used for pricing insurance policies (i.e., health status), and which therefore should be treated as consumers' private information for this purpose. In contrast to MT, we estimate selection on both unobservables and observables that cannot be used for pricing Medigap policies. We find that selection on "observable private information" is much more important.

[16]The vector $\mathbf{xc}_i$ includes most of the variables in $\mathbf{xi}_i$ and $\mathbf{xe}_i$. The exception is that the second and third powers of age and interactions of age with gender and of marital status with gender as well as time trend are included in $\mathbf{xe}_i$ but not in $\mathbf{xc}_i$ to reduce the number of parameters to be estimated. See Table A-2.

where

$$XC_i = \begin{pmatrix} \mathbf{xc}_i' & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & \mathbf{xc}_i' & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & \mathbf{0}' & \mathbf{xc}_i' & \mathbf{0} \\ \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{xc}_i' \end{pmatrix},$$

and $\Lambda = [\boldsymbol{\lambda}_1', ..., \boldsymbol{\lambda}_4']'$. Thus, the disturbances of the structural system of equations (3)-(8), conditional on type j, follow a multivariate normal distribution with zero mean and variance-covariance matrix given by:

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & 0 & \mathbf{0} \\ \sigma_{12} & \sigma_{22} & 0 & \mathbf{0} \\ 0 & 0 & \sigma_j^2 & \mathbf{0} \\ \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & V_c \end{pmatrix}.$$

While type j is latent, we assume that the probability of being type j depends on an individual's exogenous characteristics by way of a multinomial probit model, as in Geweke and Keane (2007):

$$\begin{aligned} \widetilde{W}_{ij} &= \boldsymbol{\delta}_j' \mathbf{xw}_i + \zeta_{ij} \qquad j = 1, \ldots, m-1 \\ \widetilde{W}_{im} &= \zeta_{im}. \end{aligned} \tag{9}$$

The $\widetilde{W}_{ij}$ are latent propensities of being type j, and $\mathbf{xw}_i$ is a vector of individual characteristics including demographics and health status.[17] The ζ_{ij} are independent standard normal random variables. An individual i is of type j iff $\widetilde{W}_{ij} \geq \widetilde{W}_{il} \; \forall \; l = 1, ..., m$. The probability

[17]In our empirical specification $\mathbf{xw}_i$ is almost identical to $\mathbf{xe}_i$, with the exception that the second and third powers of age and interactions of age with gender and of marital status with gender are included in $\mathbf{xe}_i$ but not in $\mathbf{xw}_i$ to reduce the number of parameters to be estimated. See Table A-2.

of type j is given by:

$$P(\text{type}_i = j|\mathbf{xw}_i, \boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_m) = \int_{-\infty}^{\infty} \phi(y - \boldsymbol{\delta}_j'\mathbf{xw}_i) \prod_{l \neq j}^{m} \Phi(y - \boldsymbol{\delta}_l'\mathbf{xw}_i)dy, \tag{10}$$

where $\Phi(.)$ denotes standard normal cdf, $\phi(.)$ denotes standard normal pdf and $\boldsymbol{\delta}_m = \mathbf{0}$. This restriction resolves the well-known identification issue in multinomial choice models which stems from the fact that only differences in alternative-specific utilities affect the actual choice. That is, if no restrictions were placed on $\boldsymbol{\delta}_j$, the probability of being type j would not change if all $\boldsymbol{\delta}_j$ were replaced by $\boldsymbol{\delta}_j + \Delta$. To achieve identification, one of the alternative-specific vectors of coefficients is often normalized to zero, as we do here.

3.2 Combining data from the MCBS and the HRS

To estimate the model in section 3.1, a dataset containing information on $I_i, E_i, \mathbf{c}_i$ and the exogenous health status and demographic characteristics (which we denote by $\mathbf{x}_i : \mathbf{x}_i \supseteq \{\mathbf{xi}_i, \mathbf{xe}_i, \mathbf{xc}_i, \mathbf{xw}_i\}$) for all observations is required. Unfortunately, such a dataset is not available. But instead the following two datasets are available: the MCBS, which has information on I_i, E_i and $\mathbf{x}_i$ but does not have information on $\mathbf{c}_i$, and the HRS, which has information on $I_i, \mathbf{c}_i$ and $\mathbf{x}_i$ but does not have information on E_i. Our strategy is to combine information from the two datasets by assuming (i) that the joint distribution of $I_i^*, E_i^*, \widehat{E}_i, E_i, I_i, \mathbf{c}_i$ conditional on $\mathbf{x}_i$ and the parameters $\boldsymbol{\theta}$,

$$\boldsymbol{\theta} = [\alpha_0, \alpha_1, \alpha_2, \boldsymbol{\alpha}_3', \boldsymbol{\alpha}_4', \boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_m', \sigma_1^2, ...\sigma_m^2, \gamma_1, ..., \gamma_m, \boldsymbol{\delta}_1', ..., \boldsymbol{\delta}_m', \sigma_{12}, \sigma_{22}, V_c, \Lambda'],$$

is the same in the MCBS and HRS datasets, and is as specified in section 3.1, and (ii) that $\mathbf{c}_i$ and E_i are missing from the MCBS and the HRS respectively completely at random (using the definition of Gelman et al. (1995)).

Let $\mathbf{C}^o$ denote the collection of $\mathbf{c}_i$'s that are observed, and $\mathbf{C}^m$ denote the collection of $\mathbf{c}_i$'s that are missing. Similarly, let $\mathbf{E}^o$ denote the collection of E_i's that are observed, and $\mathbf{E}^m$ denote the collection of E_i's that are missing. Thus, $\mathbf{c}_i \in \mathbf{C}^m$ iff $i \in MCBS$, and $\mathbf{c}_i \in \mathbf{C}^o$ iff $i \in$ HRS. Similarly, $E_i \in \mathbf{E}^m$ iff $i \in HRS$, and $E_i \in \mathbf{E}^o$ iff $i \in$ MCBS. The assumption that the data are missing completely at random implies that the missing data mechanism is independent of $I_i, E_i, \mathbf{c}_i, \mathbf{x}_i$. Hence, there is no need to specify an auxiliary missing data process that is separate from the structural model in (3) - (10). Assuming that the HRS and the MCBS are non-overlapping random samples from the same population, the estimation can be carried out by stacking the variables from the two datasets and imputing missing variables using the assumed data generating process in (3) - (10).

Let S_i denote a survey indicator so that $S_i = 1$ if $i \in$ MCBS and $S_i = 0$ if $i \in$ HRS, and let N^M and N^H denote number of observations in the MCBS and HRS respectively. Let $N = N^M + N^H$ denote the number of observations in the combined dataset. The probability density function of the observables $\mathbf{I}$, $\mathbf{E}^o$ and $\mathbf{C}^o$ conditional on exogenous variables $\mathbf{X}$, survey indicators $\mathbf{S} \equiv [S_1, ..., S_N]$ and parameters $\boldsymbol{\theta}$ consists of two parts, corresponding to the MCBS and HRS subsets. To obtain the expression of this probability density we (i) substitute equations (6) and (8) into equation (5); and (ii) substitute (6) into (3). This gives us, conditional on type j, a system of equations for I_i^*, E_i^*, $\widehat{E}_i$, $\mathbf{c}_i$, in which the vector of disturbances has a multivariate normal distribution. At this point we can integrate out E_i^* (because it is a latent variable which is never observed by the econometrician), which leaves us with the multivariate normal distribution of I_i^*, $\widehat{E}_i$ and $\mathbf{c}_i$. We also have to integrate out $\mathbf{c}_i^m$ from the MCBS subsample because these variables are not available in the MCBS. So, in the MCBS subset we are left with the following reduced-form model, conditional on type j:

$$I_i^*|j = \alpha_0 \boldsymbol{\beta}_j' \mathbf{xe}_i + \alpha_1 \sigma_j^2 + \alpha_2 \sigma_j^2 \mathbf{xc}_i' \boldsymbol{\lambda}_1 + \boldsymbol{\alpha}_3' \mathbf{xi}_i + \boldsymbol{\alpha}_4' X C_i \boldsymbol{\Lambda} + \xi_{1i} \tag{11}$$

$$\widehat{E}_i|j = \boldsymbol{\beta}_j' \mathbf{xe}_i + \gamma_j I_i + \xi_{2i} \tag{12}$$

$$I_i|j = \iota(I_i^* > 0|j) \quad (13)$$

$$E_i|j = max\{0, \widehat{E}_i|j\}, \quad (14)$$

where $\iota(.)$ is an indicator function,

$$\xi_{1i} = \varepsilon_{1i} + \alpha_0\varepsilon_{2i} + \alpha_2\sigma_j^2\varepsilon_{31i} + \boldsymbol{\alpha}_4'\boldsymbol{\varepsilon}_{3i}$$

and

$$\xi_{2i} = \varepsilon_{2i} + \eta_i.$$

The reduced-form errors ξ_{1i} and ξ_{2i} have a bivariate normal distribution:

$$\begin{matrix}\xi_{1i} \\ \xi_{2i}\end{matrix} |j \sim N\left(\begin{matrix}0 \\ 0\end{matrix}, \begin{bmatrix}\sigma_{11} + 2\alpha_0\sigma_{12} + \alpha_0^2\sigma_{22} + \boldsymbol{\alpha}_4'V_c\boldsymbol{\alpha}_4 + \alpha_2^2\sigma_j^4 \cdot \mathrm{v}_c^{11} + 2\alpha_2\sigma_j^2\sum_{l=1}^4 \cdot\alpha_{4l}\cdot \mathrm{v}_c^{1l} & \sigma_{12} + \alpha_0\sigma_{22} \\ \sigma_{12} + \alpha_0\sigma_{22} & \sigma_{22} + \sigma_j^2\end{bmatrix}\right),$$

where v_c^{lk} denotes the lk^{th} element of V_c.

Let $\mu_1 \equiv \alpha_0\boldsymbol{\beta}_j'\mathbf{xe}_i + \alpha_1\sigma_j^2 + \alpha_2\sigma_j^2\mathbf{xc}_i'\boldsymbol{\lambda}_1 + \boldsymbol{\alpha}_3'\mathbf{xi}_i + \boldsymbol{\alpha}_4'XC_i\boldsymbol{\Lambda}$ and s_ξ denote the standard deviation of ξ_{1i}. The joint probability density of E_i and I_i, conditional on type j, in the MCBS subsample is that of a Tobit model (for E_i) with an endogenous binary explanatory variable (I_i). Its derivation is given in Wooldridge (ex.16.6):

$$\begin{aligned} & p(E_i, I_i|\mathbf{x}_i, j, \boldsymbol{\theta}, S_i = 1) = \\ & I_i \cdot \int_{-\mu_1}^{\infty} g(E_i|\boldsymbol{\beta}_j'\mathbf{xe}_i + \gamma_j + \frac{\sigma_{12} + \alpha_0\sigma_{22}}{s_\xi^2}\xi_{1i}, \sigma_{22} + \sigma_j^2 - \frac{(\sigma_{12} + \alpha_0\sigma_{22})^2}{s_\xi^2}) \cdot \frac{1}{s_\xi}\phi(\frac{\xi_{1i}}{s_\xi})d\xi_{1i} \\ + \; & (1 - I_i) \cdot \int_{-\infty}^{-\mu_1} g(E_i|\boldsymbol{\beta}_j'\mathbf{xe}_i + \frac{\sigma_{12} + \alpha_0\sigma_{22}}{s_\xi^2}\xi_{1i}, \sigma_{22} + \sigma_j^2 - \frac{(\sigma_{12} + \alpha_0\sigma_{22})^2}{s_\xi^2}) \cdot \frac{1}{s_\xi}\phi(\frac{\xi_{1i}}{s_\xi})d\xi_{1i}, \end{aligned}$$

where

$$g(E|\mu, \sigma^2) = \left(\frac{1}{\sigma}\phi\left(\frac{(E-\mu)}{\sigma}\right)\right)^{\iota(E>0)}\left(1 - \Phi\left(\frac{\mu}{\sigma}\right)\right)^{\iota(E=0)}. \quad (15)$$

In the HRS subset $\mathbf{c}_i$ is available, but E_i is not. Hence, we have to integrate out $\widehat{E}_i$ and E_i. After the integration we are left with the following reduced-form model for the HRS subset, conditional on type j:

$$I_i^*|j = \alpha_0 \boldsymbol{\beta}_j' \mathbf{xe}_i + \alpha_1 \sigma_j^2 + \alpha_2 \sigma_j^2 \cdot c_{1i} + \boldsymbol{\alpha}_3' \mathbf{xi}_i + \boldsymbol{\alpha}_4' \mathbf{c}_i + \xi_{2i} \quad (16)$$

$$\mathbf{c}_i|j = XC_i \Lambda + \boldsymbol{\varepsilon}_{3i} \quad (17)$$

$$I_i|j = \iota(I_i^* > 0|j) \quad (18)$$

where ξ_{2i} is normal with zero mean, variance $s_{\xi_2} \equiv \sigma_{11} + 2\alpha_0\sigma_{12} + \alpha_0^2\sigma_{22}$ and is independent of $\boldsymbol{\varepsilon}_{3i}$.

The joint probability density of I_i and $\mathbf{c}_i$ in the HRS subset, conditional on type j, is given by the product of the likelihood of a probit model for I_i and a multivariate normal probability density function for $\mathbf{c}_i$:

$$\begin{aligned} p(I_i, \mathbf{c}_i | \mathbf{x}_i, j, \boldsymbol{\theta}, S_i = 0) &= \Phi\left(\frac{\alpha_0 \boldsymbol{\beta}_j' \mathbf{xe}_i + \alpha_1 \sigma_j^2 + \alpha_2 \sigma_j^2 \cdot c_{1i} + \boldsymbol{\alpha}_3' \mathbf{xi}_i + \boldsymbol{\alpha}_4' \mathbf{c}_i}{\sqrt{\sigma_{11} + 2\alpha_0\sigma_{12} + \alpha_0^2\sigma_{22}}}\right)^{I_i} \\ & \left(1 - \Phi\left(\frac{\alpha_0 \boldsymbol{\beta}_j' \mathbf{xe}_i + \alpha_1 \sigma_j^2 + \alpha_2 \sigma_j^2 \cdot c_{1i} + \boldsymbol{\alpha}_3' \mathbf{xi}_i + \boldsymbol{\alpha}_4' \mathbf{c}_i}{\sqrt{\sigma_{11} + 2\alpha_0\sigma_{12} + \alpha_0^2\sigma_{22}}}\right)\right)^{1-I_i} \\ &\cdot \quad (2\pi)^{-\frac{K_c}{2}} |V_c|^{-\frac{1}{2}} \exp(-(\mathbf{c}_i - XC_i\boldsymbol{\Lambda})' V_c^{-1} (\mathbf{c}_i - XC_i\boldsymbol{\Lambda})/2). \end{aligned}$$

To obtain the probability density of the observables unconditional on type j we have to marginalize over the types by multiplying type-specific densities of observables by the type probabilities in (10) and summing the resulting products over the types. The probability density function of observables $\mathbf{E}^o$,$\mathbf{I}$, $\mathbf{C}^o$ conditional on exogenous variables $\mathbf{X}$, survey indicators $\mathbf{S} \equiv [S_1, ..., S_N]$ and parameters $\boldsymbol{\theta}$ is given by:

$$p(\mathbf{E}^o, \mathbf{I}, \mathbf{C}^o | \mathbf{S}, \mathbf{X}, \boldsymbol{\theta}) = \prod_{i=1}^{N} (\sum_{j=1}^{m} (\int_{-\infty}^{\infty} \phi(y - \boldsymbol{\delta}_j' \mathbf{xw}_i) \prod_{l \neq j}^{m} \Phi(y - \boldsymbol{\delta}_l' \mathbf{xw}_i) dy$$

$$\cdot \quad p(E_i, I_i | \mathbf{x}_i, j, \boldsymbol{\theta}, S_i = 1)^{S_i=1} \cdot p(I_i, \mathbf{c}_i | \mathbf{x}_i, j, \boldsymbol{\theta}, S_i = 0)^{S_i=0})), \tag{19}$$

where $\boldsymbol{\delta}_m = \mathbf{0}$. It is easy to see that σ_{11} is not identified separately from α_0, α_1, α_2, $\boldsymbol{\alpha}_3$, $\boldsymbol{\alpha}_4$ and σ_{12} in the sense that if we multiply $\sigma_{11}^{1/2}$ and all these parameters by a constant, the joint density will not change. Identification in such cases is usually achieved by the normalization $\sigma_{11} = 1$. But for the purposes of posterior simulation it is more convenient to normalize the variance of $\varepsilon_{1i}|\varepsilon_{2i}$, i.e. to set $\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}} = 1$, which implies the restriction $\sigma_{11} = 1 + \frac{\sigma_{12}^2}{\sigma_{22}}$.

3.3 Posterior Simulation Algorithm

Bayesian inference in this model can be simplified by data augmentation. In particular, both the MCBS and the HRS subsamples are augmented by the latent vectors $\mathbf{I}^* = [I_1^*, ..., I_N^*]'$; the MCBS data are also augmented by the missing values $\mathbf{c}_i^m$, $i = 1, ..., N_M$ and by notional expenditure $\widehat{\mathbf{E}} = [\widehat{E}_1, ..., \widehat{E}_{N^M}]'$. The notional expenditure $\widehat{E}_i$ differs from actual expenditure E_i only for observations with $E_i = 0$. Data augmentation which introduces artificial values of the dependent variable for observations with truncated outcomes is a standard approach to Bayesian inference in the Tobit model by way of the Gibbs sampler (due to Chib (1992)).

The fact that the $\mathbf{c}_i$ is missing from the MCBS subsample complicates simulation from the posterior distribution of parameters. In particular, after $\mathbf{c}_i^m$ is integrated out of the MCBS subsample, the usual normal and Wishart prior distributions for $\boldsymbol{\alpha}_4$ and V_c, respectively, are no longer conjugate to the probability density function of observables of the MCBS subset (as is clear from the expression for the variance of the reduced-form error ξ_{1i} in equation (11)). There are no other known distributions which would serve as conjugate priors for $\boldsymbol{\alpha}_4$ and V_c. Hence, the conditional (on other parameters) posterior distributions of the Gibbs sampler blocks involving $\boldsymbol{\alpha}_4$ and V_c would be of unknown form and would need to be sampled using a Metropolis-Hastings step. This involves a challenging task of choosing the proposal distribution for multidimensional vectors of parameters. For this reason in the estimation we perform multiple imputations of $\mathbf{c}_i^m$ rather than integrating it our analytically.

But we analytically integrate out E_i^m and $\widehat{E}_i$ in the HRS subsample, as well as E_i^* in both HRS and MCBS subsamples. Integrating out E_i^* (rather then augmenting data density with the latent E_i^*) destroys conjugacy of the data density to the normal and gamma priors for σ_{12}, σ^2 and σ_j^2. However, because these parameters are scalars, the proposal densities in the Metropolis-Hastings steps for these parameters are easier to choose. Both the HRS and MCBS are also augmented by latent type indicators $\mathbf{s} = [s_1, ..., s_N]'$, so that $s_i = j$ if i's type is j, and by latent type propensities $\mathbf{W} = [\widetilde{\mathbf{W}}'_1, ..., \widetilde{\mathbf{W}}'_N]$, where $\widetilde{\mathbf{W}}_i = [\widetilde{W}_{i1}, ..., \widetilde{W}_{im}]'$.

Let $\mathbf{I} = [I_1, ..., I_N]'$ and $\mathbf{E}^o = [E_1^o, ..., E_{N^M}^o]'$. Then the augmented data density conditional on $\mathbf{X}, \mathbf{S}$ and $\boldsymbol{\theta}$ can be written as follows:

$$
\begin{aligned}
& p(\mathbf{I}^*, \mathbf{I}, \widehat{\mathbf{E}}, \mathbf{E}^o, \mathbf{C}^o, \mathbf{C}^m, \mathbf{s}, \mathbf{W} | \mathbf{S}, \mathbf{X}, \boldsymbol{\theta}) \\
= & \prod_i^N \Big[p(I_i^* | \mathbf{c}_i^m, \mathbf{x}_i, s_i = j, \boldsymbol{\theta}) \cdot p(I_i | I_i^*, \mathbf{c}_i^m, \mathbf{x}_i, s_i = j, \boldsymbol{\theta}) \cdot p(\widehat{E}_i | I_i^*, I_i, \mathbf{c}_i^m, \mathbf{x}_i, s_i = j, \boldsymbol{\theta}) \\
\cdot \ & p(E_i^o | \widehat{E}_i, I_i^*, I_i, \mathbf{c}_i^m, \mathbf{x}_i, s_i = j, \boldsymbol{\theta}) \cdot p(\mathbf{c}_i^m | \mathbf{x}_i, s_i = j, \boldsymbol{\theta}) \Big]^{S_i} \\
\cdot \ & [p(I_i^* | \mathbf{c}_i^o, \mathbf{x}_i, s_i = j, \boldsymbol{\theta}) \cdot p(I_i | I_i^*, \mathbf{c}_i^o, \mathbf{x}_i, s_i = j, \boldsymbol{\theta}) \cdot p(\mathbf{c}_i^o | \mathbf{x}_i, s_i = j, \boldsymbol{\theta})]^{1 - S_i} \\
\cdot \ & p(s_i = j | \widetilde{\mathbf{W}}_i, \boldsymbol{\theta}) \cdot p(\widetilde{\mathbf{W}}_i | \mathbf{x}_i, \boldsymbol{\theta}). \qquad (20)
\end{aligned}
$$

The first two lines of (20) are for the MCBS observations, the third line is for the HRS observations, and the last line which involves type probabilities, is relevant for all observations.

After substitution of expressions for the probability distributions corresponding to the model specified in (3) - (10) the expression in (20) becomes:

$$
p(\mathbf{I}^*, \mathbf{I}, \widehat{\mathbf{E}}, \mathbf{E}^o, \mathbf{C}^o, \mathbf{C}^m, \mathbf{s}, \mathbf{W} | \mathbf{S}, \mathbf{X}, \boldsymbol{\theta}) = \prod_i [\frac{1}{\sqrt{2\pi(1 + \frac{\sigma_{12}^2}{\sigma_{22}} + 2\alpha_0\sigma_{12} + \alpha_0^2\sigma_{22} - \frac{(\sigma_{12} + \alpha_0\sigma_{22})^2}{\sigma_{22} + \sigma_{s_i}^2})}}
$$

$$
\cdot \ \exp(-\frac{(I_i^* - \alpha_0 \boldsymbol{\beta}'_{s_i}\mathbf{xe}_i - \alpha_1\sigma_{s_i}^2 - \alpha_2\sigma_{s_i}^2 c_{1i}^m - \boldsymbol{\alpha}'_3\mathbf{xi}_i - \boldsymbol{\alpha}'_4\mathbf{c}_i^m - \frac{\sigma_{12} + \alpha_0\sigma_{22}}{\sigma_{22} + \sigma_{s_i}^2}(\widehat{E}_i - \boldsymbol{\beta}'_{s_i}\mathbf{xe}_i - \gamma_{s_i} I_i))^2}{2(1 + \frac{\sigma_{12}^2}{\sigma_{22}} + 2\alpha_0\sigma_{12} + \alpha_0^2\sigma_{22} - \frac{(\sigma_{12} + \alpha_0\sigma_{22})^2}{\sigma_{22} + \sigma_{s_i}^2})})
$$

$$\cdot \quad (2\pi)^{-4/2}|V_c|^{-1/2}\exp(-(\mathbf{c}_i^m - XC_i\Lambda)'V_c^{-1}(\mathbf{c}_i^m - XC_i\Lambda)/2)$$
$$\cdot \quad \frac{1}{\sqrt{2\pi(\sigma_{22}+\sigma_{s_i}^2)}}\exp(-\frac{(\widehat{E}_i - \boldsymbol{\beta}'_{s_i}\mathbf{xe}_i - \gamma_{s_i}I_i)^2}{2(\sigma_{22}+\sigma_{s_i}^2)})\cdot(\iota(E_i^o = \widehat{E}_i)\cdot\iota(\widehat{E}_i \geq 0) + \iota(E_i^o = 0)\cdot\iota(\widehat{E}_i < 0))]^{S_i}$$
$$\cdot \quad [\frac{1}{\sqrt{2\pi(1+\frac{\sigma_{12}^2}{\sigma_{22}}+2\alpha_0\sigma_{12}+\alpha_0^2\sigma_{22})}}\exp(-\frac{(I_i^* - \alpha_0\boldsymbol{\beta}'_{s_i}\mathbf{xe}_i - \alpha_1\sigma_{s_i}^2 - \alpha_2\sigma_{s_i}^2 c_{1i}^o - \boldsymbol{\alpha}'_3\mathbf{xi}_i - \boldsymbol{\alpha}'_4\mathbf{c}_i^o)^2}{2(1+\frac{\sigma_{12}^2}{\sigma_{22}}+2\alpha_0\sigma_{12}+\alpha_0^2\sigma_{22})}) \quad (21)$$
$$\cdot \quad (2\pi)^{-4/2}|V_c|^{-1/2}\exp(-(\mathbf{c}_{1i}^o - XC_i\Lambda)'V_c^{-1}(\mathbf{c}_{1i}^o - XC_i\Lambda)/2)]^{1-S_i}$$
$$\cdot \quad (\iota(I_i^* \geq 0)\cdot\iota(I_i = 1) + \iota(I_i^* < 0)\cdot\iota(I_i = 0))\cdot\left(\sum_{j=1}^{m}\prod_{k=1}^{m}\iota(\widetilde{W}_{ik}\in(-\infty,\widetilde{W}_{ij}])\right)$$
$$\cdot \quad (\frac{1}{\sqrt{2\pi}})^m\cdot\exp(-\sum_{i=1}^{N}(\widetilde{W}_{im}^2/2))\cdot\exp(-\sum_{j=1}^{m-1}(\widetilde{\mathbf{w}}_j - \mathbf{XW}\boldsymbol{\delta}_j)'(\widetilde{\mathbf{w}}_j - \mathbf{XW}\boldsymbol{\delta}_j)/2),$$

where $\mathbf{XW} = [\mathbf{xw}_1, ..., \mathbf{xw}_N]'$ and $\widetilde{\mathbf{w}}_j = [\widetilde{W}_{1j}, ..., \widetilde{W}_{Nj}]'$ for $j = 1, ..., m$.

For the purposes of Bayesian inference it is convenient to split the parameter vector $\boldsymbol{\theta}$ into the following blocks:

1. α_0

2. $\boldsymbol{\alpha} \equiv [\alpha_1, \alpha_2, \boldsymbol{\alpha}'_3, \boldsymbol{\alpha}'_4]'$

3. $\boldsymbol{\beta}_j$ for $j = 1, ..., m$

4. γ_j for $j = 1, ..., m$

5. $h_j \equiv \frac{1}{\sigma_j^2}$, $j = 1, ..., m$.

6 $\boldsymbol{\Lambda} \equiv [\boldsymbol{\lambda}_1^{'}, ..., \boldsymbol{\lambda}_4^{'}]'$,

7. $H_c \equiv V_c^{-1}$;

8. $h_{22} \equiv \sigma_{22}^{-1}$;

9. σ_{12}

10. $\boldsymbol{\delta}_j$, $j = 1, ..., m-1$;

Where possible, we specify natural conjugate prior distributions for these parameters blocks, and specify that in the prior these blocks are independent, i.e

$$p(\boldsymbol{\theta}) = p(\alpha_0)p(\boldsymbol{\alpha}) \prod_{j=1}^{m} p(\boldsymbol{\beta}_j) \prod_{j=1}^{m} p(\gamma_j) \prod_{j=1}^{m} p(h_j) \prod_{k=1}^{4} p(\boldsymbol{\lambda}_k) p(\sigma_{12}) p(h_{22}) p(H_c) \prod_{j=1}^{m-1} p(\boldsymbol{\delta}_j). \quad (22)$$

We specify the hyperparameters of these prior distributions so as to allow a substantial prior uncertainty about the parameter values. These prior distributions are discussed in detail in Appendix A-1.

Let **data** denote the collection $\langle \mathbf{I}, \mathbf{E}^o, \mathbf{C}^o, \mathbf{X}, \mathbf{S} \rangle$. Then the joint posterior distribution of the parameters and the latent and missing data $p(\boldsymbol{\theta}, \mathbf{I}^*, \widehat{\mathbf{E}}, \mathbf{C}^m, \mathbf{W}, \mathbf{s} | \mathbf{data})$ is proportional to the product of (21) and (22). To simulate from this posterior distribution we construct a Gibbs sampling algorithm with Metropolis within Gibbs steps which cycles between the conditional posterior distributions of blocks of parameters and vectors of latent and missing variables $\mathbf{I}^*, \widehat{\mathbf{E}}, \mathbf{C}^m, \mathbf{W}, \mathbf{s}$. The details of the algorithm are given in Appendix A-2.

4 Results

The exact specification of the equations of the model in terms of the demographic and health status characteristics included in each equation is given in Appendix A-3. In particular, in Table A-2 we show our exclusion restrictions in tabular form. As for the expenditure distribution, we have specified $m = 5$. In a companion paper (Keane and Stavrunova (2011)) we discuss the SMTobit model in detail and show that a 6 component mixture provides the best fit to the expenditure distribution in the MCBS sample used in this paper, while a model with 5 components fares only slightly worse.[18] The number of components in this study is a compromise between model fit and the mixing properties of the posterior simulator. We use

[18]To compare models with different numbers of components we use the modified cross-validated log-scoring rule developed in Geweke and Keane (2007), which is less computationally demanding than the comparison based on marginal likelihoods, which is a standard approach to model selection in Bayesian statistics.

m=5 because the posterior simulator exhibited slow convergence when $m = 6$ was specified.

In order to investigate sources of advantageous selection we have estimated five alternative models by progressively adding more potential SAS variables to the insurance equation. This sequential procedure is similar to FKS. In the first specification the insurance equation contains only expenditure risk and insurance pricing variables. The second model adds income and education. The third model adds cognitive ability, financial planning horizon and longevity expectations. The fourth model adds risk tolerance, the variance of the expenditure forecast error and an interaction between risk tolerance and variance. Finally, the fifth model adds ethnicity, marital status and an interaction of gender with age. We also estimated the fifth specification with the covariance between the unobservable determinants of Medigap status and expenditure, σ_{12}, set to zero, to asses the bias in the selection and moral hazard effects from not accounting for correlation between the unobservables.

Due to the presence of latent variables and mixture components, the output of the posterior simulator exhibits a high degree of autocorrelation. To allow the simulator to explore the parameter space adequately, the algorithm was allowed to run for an extended period of time. We have obtained 1,200,000 draws from the posterior distribution, discarded the first 200,000 draws as a burn-in and used every 1000th of the remaining 1,000,000 draws for analysis. In the fifth model the autocorrelation in these 1,000 draws ranges from 0 (for parameters of the $\mathbf{c}^m$ distribution and the coefficients of the exogenous covariates in the insurance equation) to 0.11, 0.28 and 0.65 (for α_0, σ_2, σ_{12}) and to 0.81 (for parameters of the expenditure distribution for the type with the lowest probability). The serial correlation coefficients for the parameters σ_j^2 are between 0.04 and 0.16, while those for γ_j are between 0.26 and 0.02. Thus, the serial correlation is low for the parameters that are most important for our analysis (i.e. the insurance equation parameters and γ_j). The relative numerical efficiency ranges from 0.08 (for the parameters with the highest autocorrelation) to 1.6. All parameters pass the formal test of convergence suggested in Geweke (1992).

4.1 Model Fit

The fit of the model which includes all potential SAS variables to selected characteristics of the distribution of the Medigap insurance coverage rate and health care expenditure, conditional on observable characteristics $\mathbf{x}_i$, is examined in Figure 2. To produce Figure 2 we simulated artificial samples of E_i and I_i, conditional on $\mathbf{x}_i$, for 1000 draws from the posterior distribution of parameters. In particular, for each $\boldsymbol{\theta}^k$ drawn from the posterior distribution we simulate artificial data for each $i = 1, ..., N$ as follows:

1. Latent types $s_i^k \sim p(s_i|\mathbf{xw}_i, \boldsymbol{\delta}_1^k, ..., \boldsymbol{\delta}_m^k)$ using (9);
2. Missing SAS variables in the MCBS subsample $\mathbf{c}_i^{mk} \sim p(\mathbf{c}_i^m|XC_i, \Lambda^k, V_c^k)$ using (8);
3. Latent data $[I_i^{*k}, E_i^{*k}]' \sim p(I_i^*, E_i^*|\mathbf{xi}_i, S_i \cdot \mathbf{c}_i^{mk} + (1 - S_i) \cdot \mathbf{c}_i^o, \mathbf{xe}_i, s_i^k, \boldsymbol{\beta}_{s_i^k}^k, \sigma_{12}^k, \sigma_{22}^k, \sigma_{s_i^k}^{2k})$ using (5) and (6). Set $I_i^k = \iota(I_i^{*k} > 0)$.
4. Notional expenditure $\widehat{E}_i^k \sim p(\widehat{E}_i^k|E_i^{*k}, I_i^k, s_i^k, \gamma_{s_i^k}^k, \sigma_{s_i^k}^{2k})$ using (3). Set $E_i^k = \widehat{E}_i^k \cdot \iota(\widehat{E}_i^k > 0)$.

Panels (a) and (b) of Figure 2 plot the predicted expected expenditure and probability of insurance coverage against actual expenditure and insurance coverage. In particular, to plot panel (a), we split the MCBS subsample into ten groups by deciles of average predicted expenditure over draws k, that is $\overline{E}_i \equiv 10^{-3} \sum_{k=1}^{10^3} (E_i^k)$. Observations in the first group have average predicted expenditure ($\overline{E}_i$) that is less than the 10th percentile of the sample distribution of $\overline{E}_i$, observations in the second group have average predicted expenditure between 10th and 20th percentiles, and so on. We do this to identify individuals, whose exogenous characteristics $\mathbf{x}_i$ are likely to result in high or low expenditure.

Then for each decile subsample g, $g = 1, ..., 10$, and for each draw k, we compute average predicted expenditure $AE_g^k \equiv N_g^{-1} \sum_{i \in g} (E_i^k)$, where N_g is the number of observations in subsample g. In panel (a) we plot the average of AE_g^k over k against the averages of *actual* expenditure of individuals falling into decile subsample g (red dots). For each g, we also plot the 5th and 95th percentiles of the series AE_g^k to show the uncertainty about the predictions due to the posterior distribution of parameters (blue dots).

Panel (b) proceeds analogously to show the fit of the model to the probability of Medigap coverage. In particular, we group the whole sample by the deciles of $\overline{I}_i \equiv 10^{-3} \sum_{k=1}^{10^3} I_i^k$ and plot the subsample averages of actual I_i against the averages and 5th and 95th percentiles (over k) of $AI_g^k \equiv N_g^{-1} \sum_{i \in g}(I_i^k)$.

Panels (a) and (b) of Figure 2 indicate a good fit to both insurance coverage and expenditure because the relationships between the predicted and actual variables are close to the 45 degree line, and the actual values are almost always contained within the 5th and 95th percentiles of the predictions.

In panel (c) of Figure 2 we compare the actual and predicted relationships between expenditure and insurance coverage for individuals in the MCBS subsample. We split the MCBS data into 10 expenditure groups (the same groups used to construct panel (a)) and for each group compute and plot the average expenditure and the average probability of Medigap coverage (solid red line). We then use blue and black dots to plot AE_g^k and AI_g^k (computed for the same expenditure decile groups) for 20 random k. The figure suggests that the model fits the bivariate relationship between insurance and expenditure well, because the predictive distribution of this relationship is very close to the actual relationship. Note that the probability of Medigap coverage is rising with expenditure at low levels of expenditure, but falling with expenditure at higher levels of expenditure.

Finally, in panel (d) we show the fit of the model to the variance of expenditure. To construct panel (d) we use the same expenditure subsamples as for panel (a), and plot the variance of the actual expenditure against the variance of predicted expenditure within these subsamples. We are interested in the fit to the variance of expenditure because the fourth and fifth specifications of the model include the variance of expenditure as well as the interaction of the variance with risk tolerance among the SAS variables. Panel (d) suggests that the fit of the model to the variance of expenditure is also quite good.

Figure 3 plots kernel density estimates of actual expenditure (red) as well as of the distribution of predicted expenditure (E_i^k) for individuals in the MCBS subsample (black). Panel (a) of the figure shows the density plot for the entire support of the expenditure distribution. The expenditure distribution has an extremely long right tail, and the smooth mixture of Tobits does a good job in capturing this complex shape - i.e. in panel (a) the predicted and actual data density practically overlap. Panels (b) and (c) present density plots for the [0 20,000] and [20,000 100,000] intervals of the support of the expenditure distribution (these intervals together contain more than 99% of the sample distribution of expenditure). The fit is very good even for these more narrowly defined intervals.

We decided to order the different mixture components by levels of the health expenditure forecast error variance (i.e., with mixture component one having the lowest variance, while component 5 has the highest).[19] We order by variance because this parameter turned out to be able to separate the components of the mixture quite well. Munkin and Trivedi (2010) also

[19] A well-known feature of mixture models is that the parameters of mixture components are not identified with respect to permutations of component labels without further restrictions (e.g. a particular ordering of component-specific means and/or variances). For example, the value of the likelihood function of a mixture of two components will not change if component 1 is relabeled as component 2, and component 2 is relabeled as component 1. As a result, the likelihood function and the posterior distribution of parameters is multimodal with $m!$ modes corresponding to $m!$ permutations of the m component labels. This creates complications for posterior simulation via the Gibbs sampler, because the simulator can get stuck in one of the posterior modes and not fully explore the entire posterior distribution (Celeux et al. (2000)). One solution to this problem is random permutation of component labels after each iteration of the Gibbs sampler, as proposed by Fruhwirth-Schnatter (2001). Another solution, proposed by Geweke (2006), is to use the permutation-augmented simulator. In the case of permutation-invariant functions of interest, this simulator amounts to running the usual Gibbs sampler without the random permutation step and using the resulting output for inference. For permutation-sensitive functions of interest it amounts to reordering of the output from the usual Gibbs sampler according to inequality constraints, which identify the component labels, and using the reordered output for inference. In this paper we use the approach of Geweke (2006). We run the Gibbs sampling algorithm with the Metropolis-Hastings step as described in Appendix A-2 (i.e. without the random permutation step). We then use the resulting output directly for inference about permutation-invariant functions of interest, such as marginal effects of covariates on the expected expenditure, the moral hazard effect and the variance of the forecast error. For inference about permutation-sensitive functions of interest (e.g. the moral hazard effects for different health types j) we use the output reordered according to the inequality restrictions on σ_j^2.

Figure 2: Model fit: Conditional Moments of Medigap Probability and Expenditure

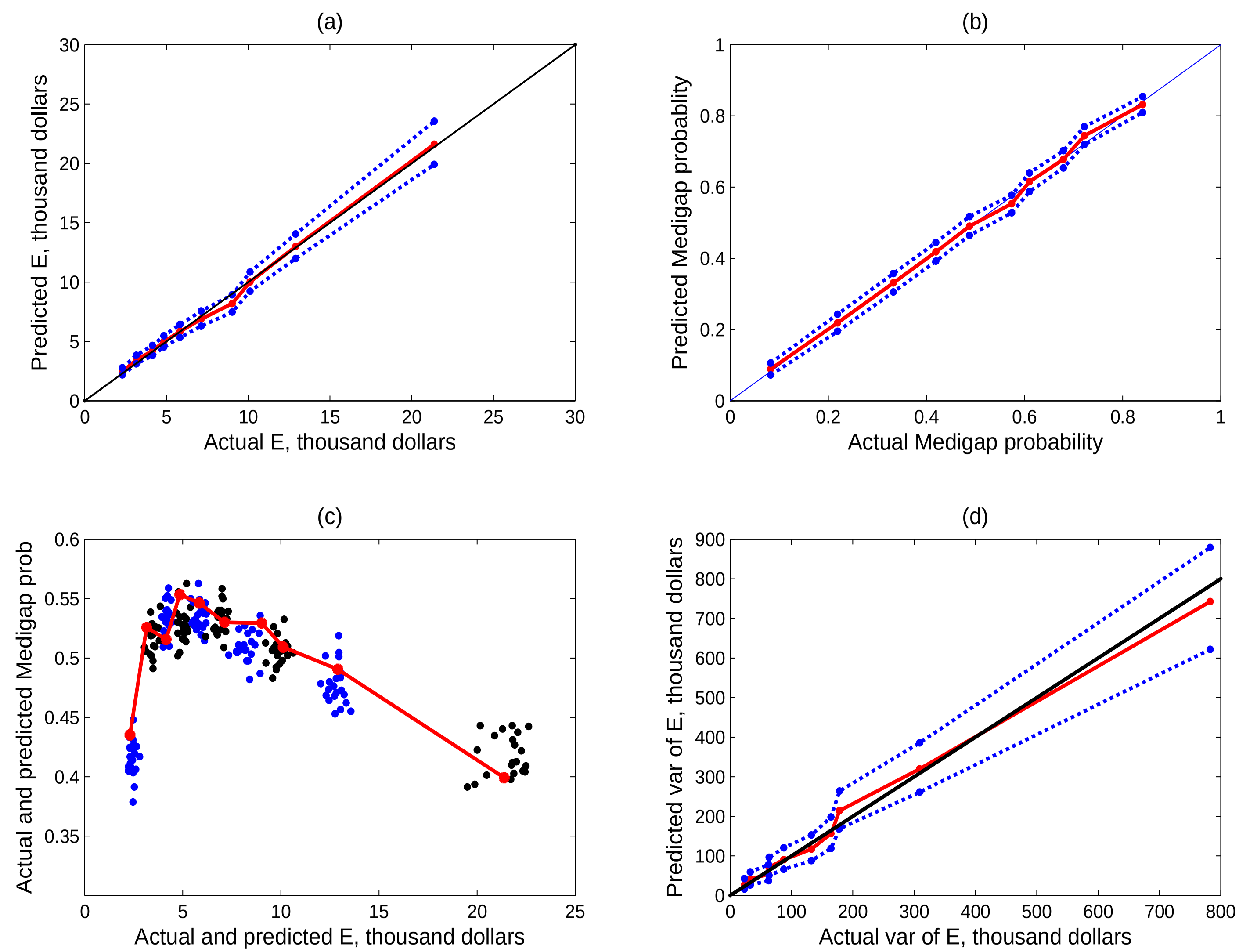
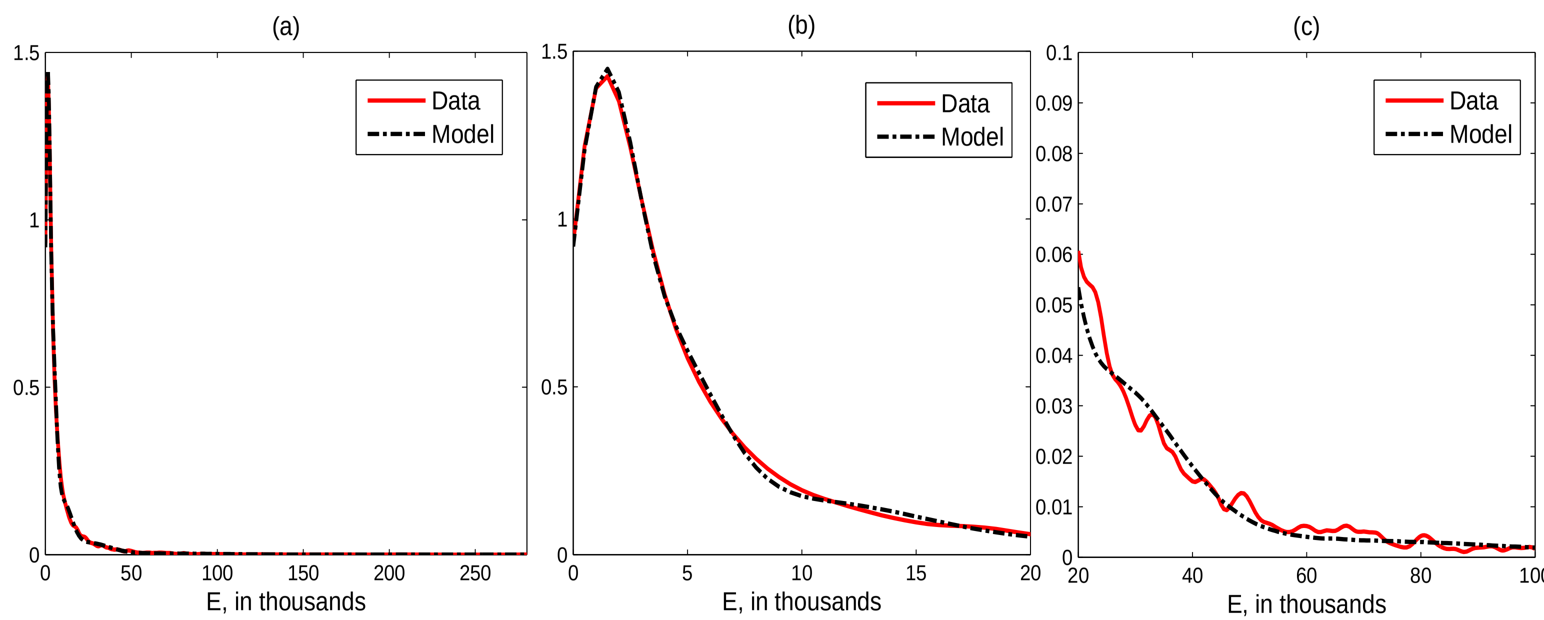


Figure 3: Model fit: Expenditure Distribution



found that mixture components were most easily separated by variance in their application of a discrete mixture model to drug expenditure of Medicare beneficiaries. Alternatively, we could have ordered the types by other parameters which differ substantially across the types, e.g. the type-specific coefficients of health factor 2 in the expected expenditure equation. Ordering by variance has the advantage that it renders our types easily interpretable, due to the fact that, for many types of health care expenditures, the conditional (on covariates) means and variances tend to be positively related (Deb et al., 2010).

Table 3 reports some type-specific parameters and functions of interest. Note that, as expected, the ranking of types by expenditure risk corresponds closely to that by variance. In fact, there is a perfect rank correlation. There is also a close relationship between health status and the size of the moral hazard effect. Types 1 and 2 are the healthiest and together make up about 71% of the sample.[20] These two types have lowest expenditure risks $E(E_i^*|\text{type}_i = j, \mathbf{data})$[21] and the smallest moral hazard effects of Medigap insurance γ_j. Type 5, which makes up about 3% of the sample, is the most unhealthy type, and also has the highest expenditure risk and the largest moral hazard effect.[22]

In the next two sub-sections we discuss, in turn, the adverse (or advantageous) selection and moral hazard effects implied by the model.

[20]The posterior mean of the type probability $p(\text{type}_i = j|\mathbf{data})$ was computed as the average of $\iota(s_i^k = j)$ over i and k, while the 5th and 95th percentiles are those of the series $\frac{1}{N}\sum_{i=1}^{N}\iota(s_i^k = j)$ for $k = 1, ...10^3$. These computations approximate the posterior mean and percentiles (over the posterior of parameters) of $\frac{1}{N}\sum_{i=1}^{N}P(\text{type}_i = j|\mathbf{xw}_i, \boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_m)$

[21]The posterior mean of the expenditure risk $E(E_i^*|\text{type}_i = j, \mathbf{data})$ is computed as the average of E_i^{*k} over i and k such that $\iota(s_i^k = j)$. The 5th and 95th percentiles are those of the series of E_i^{*k} averaged over i such that $\iota(s_i^k = j)$ for $k = 1, ...10^3$. These computations approximate the posterior mean and percentiles (over the posterior of parameters) of $\frac{1}{N}\sum_{i=1}^{N}\mathbf{xe}_i\boldsymbol{\beta}_j \cdot P(\text{type}_i = j|\mathbf{xw}_i, \boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_m)$

[22]Analogously, Munkin and Trivedi (2010) find that the size of the moral hazard effect is higher for the high-expenditure latent type than for the low-expenditure type in their study of supplemental drug insurance. Of course, since Medigap plans may cover other aspects of costs besides drugs (e.g., co-pays), it is not necessarily the case that these patterns would be the same in both markets.

Table 3: Type-specific characteristics: posterior means and 5th and 95th percentiles

Variable	Type 1	Type 2	Type 3	Type 4	Type 5
Std. deviation of η_i, $\sqrt{\sigma_j^2}$, thousand dollars	0.71 (0.68, 0.75)	1.8 (1.7, 1.9)	4.4 (4.1, 4.7)	10.9 (10.1, 11.8)	33.7 (31.5, 36.1)
$E(E^*\|\text{type}_i = j, \mathbf{data})$, thousand dollars	0.58 (0.51, 0.64)	2.86 (2.67, 3.005)	9.0 (8.3, 9.7)	22.9 (21.3, 24.7)	54.5 (49.2, 59.7)
Moral hazard effect γ_j, thousand dollars	1.25 (1.14, 1.37)	2.00 (1.8, 2.2)	2.15 (1.5, 2.8)	3.37 (1.5, 5.3)	10.2 (2.7, 17.8)
$P(\text{type}_i = j\|\mathbf{data})$	0.394 (0.38, 0.41)	0.315 (0.30, 0.33)	0.166 (0.15, 0.18)	0.094 (0.08, 0.10)	.031 (0.027, 0.036)

4.2 The Adverse (Advantageous) Selection Effect

One key focus of this paper is the relationship between expenditure risk and Medigap insurance status, conditional on pricing variables and potential SAS variables. Figure 4 shows how this relationship changes as we progressively add potential SAS variables to the insurance equation. This figure plots the distribution of the marginal effects of a one standard deviation increase in E_i^* (11.6 thousand dollars)[23] on the probability of having Medigap insurance, $\Phi(\frac{\alpha_0 E_i^* + \boldsymbol{\alpha}_1' \mathbf{P}_i + \boldsymbol{\alpha}_2' \mathbf{SAS}_i}{\sqrt{\sigma_{11}}})$, where $\mathbf{P}_i$ denotes pricing variables, and $\mathbf{SAS}_i$ denotes potential sources of advantageous selection. This probability can be derived from (5).[24]

Overall, the results in Figure 4 are consistent with the findings of FKS, both qualitatively and quantitatively. Panel (a) of the figure corresponds to the benchmark model with no SAS variables. In panel (a) the relationship is negative, suggesting that an increase in expenditure risk by 11.6 thousand dollars *decreases* the probability of Medigap coverage on average by 0.03. This negative relationship suggests advantageous selection. Adding income and education (panel (b)) weakens the relationship between risk and insurance to almost

[23]This is the standard deviation of $N \cdot 10^3$ simulated values E_i^{*k}.

[24]We evaluate the marginal effects for all individuals $i = 1, ..., N$ and for 1000 draws from the posterior distribution of parameters, replacing E_i^* and the unobserved components of $\mathbf{SAS}_i$ ($\sigma_{s_i}^2$ and $\mathbf{c}_i^m$) with E_i^{*k}, $\sigma_{s_i^k}^{2k}$ and $\mathbf{c}_i^{mk}$ simulated as discussed in the previous section. Figure 4 plots the histograms of the resulting $N \cdot 10^3$ marginal effects and indicates sample averages and standard deviations of these effects.

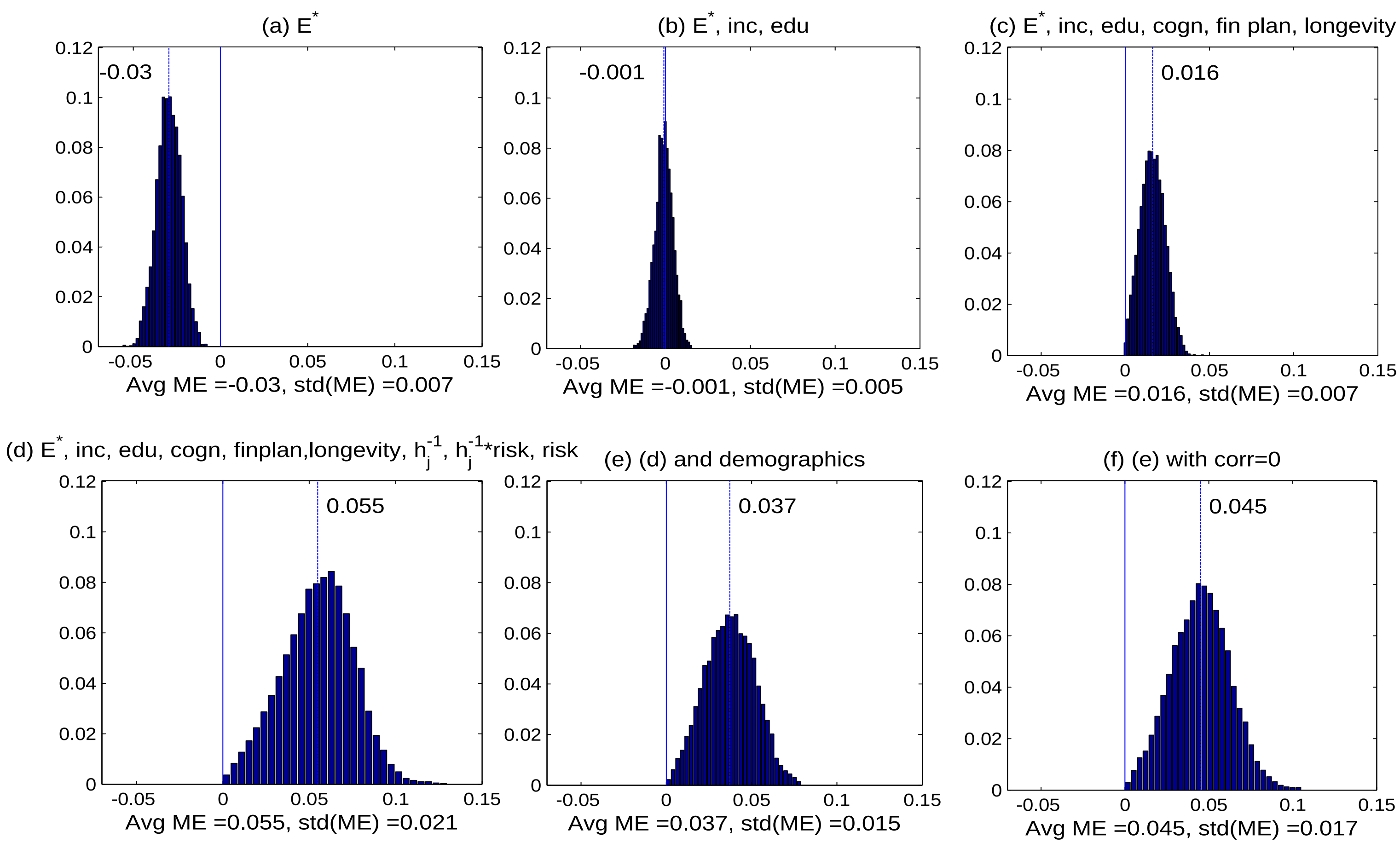
zero. Adding cognitive ability, financial planning horizon and longevity expectations (panel (c)) changes the sign of the relationship - it becomes positive, but the effect is small: a one standard deviation in E_i^* increases probability of Medigap coverage on average by 0.016. Adding risk tolerance, variance of the forecast error and the interaction of the risk tolerance with the variance (panel d) further increases the marginal effect to 0.055. Thus, our SAS variables can explain "advantageous selection" - once we condition on them, we find adverse selection as predicted by theory.

In addition to health measures, health insurance pricing variables and Medigap status I_i, our expenditure equation also includes some demographic variables (race, marital status and interactions of gender with age and marital status). These variables are included in the expenditure equation to improve model's predictive performance, but FKS excluded them from the baseline insurance equation because it is not legal to use them in pricing. However, just like the SAS variables (risk tolerance, cognition, income, education, etc.), these demographic variables may also affect tastes for insurance (and hence demand, conditional on price and expenditure risk). Hence, these demographics are also potential source of adverse/advantageous selection.

In our final exercise we include the demographics (race, marital status, etc) in the insurance equation. Doing so reduces the average marginal effect of E_i^* from 0.055 to 0.037 (panel (e)). Thus, these variables are a source of adverse selection. In particular, blacks and hispanics have a relatively low probability of purchasing Medigap (and they have relatively low expected expenditure (see Table 5). Overall, these results support the results of FKS - the set of SAS variables used in their study is enough to explain advantageous selection into Medigap insurance - but also imply that additional demographics like race may be an important source of selection.

Table 4 presents marginal effects of covariates on the probability of Medigap coverage $\Phi(\frac{\alpha_0 E_i^* + \boldsymbol{\alpha}_1' \mathbf{P}_i + \boldsymbol{\alpha}_2' \mathbf{SAS}_i}{\sqrt{\sigma_{11}}})$. The effects are evaluated for a median individual, i.e. an individual

Figure 4: Posterior distribution of the marginal effect of E_i^* on the probability of Medigap coverage



for whom: (i) exogenous characteristics are set to their sample medians, (ii) the E_i^* and $\sigma^2_{s_i}$ are set to their medians over E_i^{*k} and $\sigma^{2k}_{s_i^k}$, and (iii) $\mathbf{c}_i$ is set to it's median in the HRS subsample. We present the mean and the 5^{th} and 95^{th} percentile of the series of effects evaluated for 1000 draws from the posterior distribution. For continuous variables we report the change in Medigap probability brought about by a one standard deviation increase in the variable of interest from these median levels. For E_i^* and $\sigma^2_{s_i}$ the marginal effects correspond to a one standard deviation increase in E_i^{*k} and $\sigma^2_{s_i^k}$, while for $\mathbf{c}_i$ they correspond to a one standard deviation increase in the HRS subsample. Note that the marginal effects of E_i^* in Table 4 do not correspond to those in Figure 4, as the former are for a median individual, while the latter correspond to the whole sample distribution. Table 4 also summarizes the posterior distributions of the correlation coefficient between ε_1 and ε_2 (ρ) and the variance of ε_2, (σ_{22}).

The first column of results in Table 4 is for the basic model that contains only expected expenditure (E_i^*) and pricing variables (gender, age, region) in the insurance equation. The subsequent columns progressively add the demographic and behavioral variables that are potential sources of selection (SAS variables). Consistent with Figure 4, the effect of E_i^* goes from -0.03 in the basic model (advantageous selection) to +0.04 in the full model (adverse selection). The results in Table 4 column (2) suggest that, conditional on expenditure risk, the probability of Medigap coverage is higher for females, increases with age, education and income, and varies substantially by region. But in column (3) the inclusion of the **cogn**, **finplan** and **praliv75** variables eliminates the effect of education and greatly reduces the positive effect of income. Clearly cognitive ability has a much larger effect on the probability of Medigap coverage than the other behavioral variables. In the 4th column the model includes the risk tolerance and variance measures. This has little effect on the impacts of other variables, but it raises the effect of E_i^* to 0.07, which clearly implies adverse selection. Finally, the inclusion of race and marital status variables in column (5) causes the effects of cognition to drop from 0.17 to 0.08, and the effect of E_i^* to fall to 0.04. The indicators for black and hispanic are among the most important determinants of Medigap status - they both decrease Medigap probability by 0.24.

As we noted, cognition has a much bigger effect on Medigap coverage than other behavioral variables (e.g. risk tolerance, longevity expectation, etc.). In particular, a one standard deviation increase in the cognitive ability factor from the sample median level increases probability of Medigap coverage by 0.08 on average. This effects is estimated rather precisely - 90% of the support of it's posterior distribution is between 0.06 and 0.10.

Aside from cognition, the variance of the forecast error, σ_j^2, has a larger effect on insurance demand than any behavioral variable. It is one of the most important sources of advantageous selection. A one standard deviation increase in σ_j^2 (keeping E_i^* constant) decreases the probability of Medigap by 0.04 for individuals at the median of the risk tolerance distribution, and by 0.05 for individuals at the 90th percentile of risk tolerance distribution. Several potential explanations for the negative effect of the variance are given in FKS, including the crowding out of Medigap by Medicaid in the case of catastrophic health care expenses, as well as possible behavioral factors, such as underweighting of small probabilities of a large loss, and underestimation of the expenditure variance by individuals.

The probability of Medigap coverage also differs by region. In particular, residents of New England, the West South Central and Mountain census divisions are less likely to have Medigap than individuals living in other regions, while the East North Central and South Atlantic census divisions are the areas with the highest Medigap coverage. For example, the probability of having Medigap for individuals living in the East-North Central and South Atlantic census divisions is about 0.08 higher than that for residents of the Middle Atlantic census division, and it is about 0.36 higher than that for residents of the Mountain census division, conditional on other variables.

Interestingly, in all models the correlation between the *unobservable* determinants of insurance coverage and expenditure risk ε_1 and ε_2 is strongly negative, suggesting that selection with respect to *unobservable* determinants of the expenditure is advantageous (even when the SAS variables are included).[25] This may appear to contradict the finding of FKS that the observed SAS variables account for advantageous selection. This, in turn, raises a puzzle of why we obtained similar results to FKS in Figure 4.

The most plausible explanation for the similarity between our results and those of FKS is that the health status variables included in the prediction model of FKS capture most of the information relevant when individuals form an expectation about future health care costs and make a decision about Medigap insurance status. Indeed, our results indicate that the standard deviation of the unobservable component of expenditure risk, ε_{2i}, is very small compared to the standard deviation of expenditure risk E_i^* itself (i.e., 0.58[26] vs. 11.6 thousand dollars). This suggests that any systematic difference in expenditure risk between individuals with and without Medigap that is left unexplained by the observable health status characteristics is also small. Hence, results about the extent of adverse selection obtained from a model that does not account for the correlation between ε_{1i} and ε_{2i} should not be very different from the results reported in this paper.

In fact, we have also re-estimated our most general model with the covariance parameter σ_{12} set to zero. The fit of this restricted model to the data was very similar to that of the unrestricted model, and the posterior distribution of the marginal effect of E_i^* on the Medigap coverage probability was similar as well. This can be seen by comparing panels (e) and (f) of Figure 4. Note that the mean effect increases only slightly from 0.037 to 0.045.

[25] Note that $\sigma_{12} < 0$ means that, ceteris paribus, people with higher expected expenditure E_i^* tend to have lower demand for insurance.

[26] The posterior mean of σ_{22} is equal to 0.33, while the posterior mean of $\sqrt{\sigma_{22}}$ is equal to 0.576.

Table 4: Marginal effects of individual characteristics on the probability of Medigap coverage.

	(1) No SAS variables		(2) Add hgc and inc		(3) Add cogn, finpln, praliv75		(4) Add risktol and $\sigma^2_{s_i}$		(5) Add ethnicity and marst	
Variable	Post. mean	5th-95th prct.	Post. mean	5th-95th prct.	Post. mean	5th-95th prct.	Post. mean	5th-95th prct.	Post. mean	5th-95th prct.
E^*	-0.03	-0.04 , -0.02	0.00	-0.01 , 0.01	0.02	0.01 , 0.03	0.07	0.04 , 0.09	0.04	0.02 , 0.06
$\sigma^2_{s_i}$, risk 50th pr							-0.06	-0.09 , -0.04	-0.04	-0.07 , -0.02
$\sigma^2_{s_i}$, risk 90th pr							-0.08	-0.12 , -0.03	-0.05	-0.09 , -0.01
female	0.02	0.01 , 0.04	0.09	0.08 , 0.10	0.11	0.09 , 0.13	0.10	0.09 , 0.12	0.09	0.06 , 0.11
age	0.03	0.02 , 0.05	0.08	0.07 , 0.09	0.10	0.07 , 0.13	0.10	0.06 , 0.13	0.08	0.06 , 0.11
New Eng	-0.10	-0.13 , -0.06	-0.08	-0.12 , -0.04	-0.09	-0.14 , -0.05	-0.09	-0.14 , -0.04	-0.07	-0.11 , -0.03
Mid Atl	0.04	0.01 , 0.08	0.06	0.02 , 0.10	0.07	0.03 , 0.12	0.09	0.04 , 0.13	0.07	0.03 , 0.11
East North Cent	0.20	0.16 , 0.24	0.17	0.13 , 0.21	0.16	0.11 , 0.20	0.17	0.12 , 0.21	0.15	0.11 , 0.19
West North Cent	-0.03	-0.07 , 0.00	0.03	-0.01 , 0.06	0.04	0.00 , 0.08	0.05	0.01 , 0.10	0.06	0.02 , 0.1
South Atl	0.05	0.01 , 0.09	0.14	0.10 , 0.18	0.16	0.12 , 0.21	0.17	0.12 , 0.22	0.15	0.11 , 0.19
East South Cent	-0.02	-0.06 , 0.02	0.03	-0.01 , 0.07	0.05	0.01 , 0.10	0.07	0.02 , 0.11	0.06	0.03 , 0.1
West South Cent	-0.08	-0.12 , -0.04	-0.11	-0.15 , -0.07	-0.10	-0.15 , -0.05	-0.10	-0.16 , -0.05	-0.10	-0.14 , -0.05
Mountain	-0.18	-0.22 , -0.15	-0.23	-0.27 , -0.19	-0.24	-0.28 , -0.19	-0.23	-0.28 , -0.18	-0.21	-0.25 , -0.17
Pacific	-0.25	-0.31 , -0.19	-0.04	-0.11 , 0.03	0.08	0.02 , 0.15	0.10	0.03 , 0.16	0.15	0.10 , 0.21
hgc: ls8th			0.08	0.03 , 0.14	-0.05	-0.12 , 0.03	-0.05	-0.12 , 0.02	-0.04	-0.10 , 0.02
hgc: somehs			0.16	0.11 , 0.22	-0.04	-0.11 , 0.03	-0.04	-0.11 , 0.03	-0.02	-0.08 , 0.04
hgc: hs			0.25	0.19 , 0.3	-0.03	-0.10 , 0.04	-0.03	-0.10 , 0.04	0.01	-0.05 , 0.07
hgc: somecol			0.26	0.21 , 0.32	-0.05	-0.12 , 0.02	-0.05	-0.12 , 0.02	0.00	-0.06 , 0.06
hgc: college			0.28	0.23 , 0.34	-0.08	-0.15 , 0.00	-0.08	-0.15 , 0.00	0.00	-0.07 , 0.07
hgc: gradschl			0.31	0.26 , 0.38	-0.03	-0.11 , 0.05	-0.03	-0.11 , 0.05	0.04	-0.03 , 0.11
hgc: nr			0.12	0.02 , 0.21	0.15	0.08 , 0.22	0.15	0.07 , 0.23	0.11	0.03 , 0.18
inc 5k-10k			-0.14	-0.18 , -0.1	-0.19	-0.24 , -0.14	-0.19	-0.24 , -0.14	-0.16	-0.20 , -0.11
inc 10k-15k			0.09	0.06 , 0.13	-0.01	-0.06 , 0.04	-0.01	-0.06 , 0.05	0.02	-0.02 , 0.07
inc 15k-20k			0.16	0.12 , 0.19	0.02	-0.03 , 0.08	0.03	-0.03 , 0.08	0.06	0.02 , 0.11
inc 20k-25k			0.19	0.15 , 0.23	0.07	0.02 , 0.12	0.07	0.02 , 0.13	0.08	0.04 , 0.13
inc 25k-30k			0.24	0.20 , 0.28	0.11	0.06 , 0.16	0.11	0.06 , 0.17	0.12	0.08 , 0.17
inc 30k-35k			0.24	0.20 , 0.29	0.10	0.04 , 0.16	0.11	0.05 , 0.17	0.13	0.07 , 0.18
inc 35k-40k			0.25	0.21 , 0.30	0.09	0.03 , 0.15	0.09	0.03 , 0.15	0.12	0.07 , 0.17
inc 40k-45k			0.30	0.26 , 0.35	0.14	0.08 , 0.2	0.14	0.08 , 0.2	0.17	0.11 , 0.22
inc 45k-50k			0.28	0.23 , 0.32	0.16	0.10 , 0.22	0.16	0.09 , 0.21	0.16	0.11 , 0.22
inc 50plus			0.33	0.29 , 0.37	0.18	0.13 , 0.23	0.18	0.13 , 0.24	0.19	0.15 , 0.24
risktol							-0.02	-0.05 , 0.00	-0.01	-0.02 , 0.01
cogn					0.17	0.15 , 0.19	0.17	0.15 , 0.19	0.08	0.06 , 0.10
finpln					0.03	0.01 , 0.05	0.04	0.01 , 0.06	0.01	-0.01 , 0.03
praliv75					-0.02	-0.04 , 0	-0.01	-0.03 , 0.01	0.02	0.00 , 0.03
black									-0.24	-0.27 , -0.21
hispanic									-0.24	-0.27 , -0.21
married									0.06	0.03 , 0.09
age, female									0.06	0.04 , 0.08
married female									0.02	-0.01 , 0.06
ρ	-0.96	-0.98, -0.92	-0.90	-0.93, -0.87	-0.95	-0.96, -0.92	-0.94	-0.96, -0.92	-0.95	-0.93, -0.97
σ_{22}	0.20	0.15, 0.28	0.33	0.24, 0.39	0.30	0.24, 0.4	0.33	0.25, 0.4	0.33	0.25, 0.41

* Note: The omitted categories for the dummy variables are non-response for census divisions, zero schooling for education and less than 5 thousand dollars for income. The effects are computed for a median individual.

4.3 The Moral Hazard Effect

Row 2 of Table 3 presents the posterior means of type-specific moral hazard effects of Medigap insurance on health care expenditure (γ_j). The moral hazard effect increases as health status deteriorates (with the exception of type 3). Interestingly, however, the moral hazard effect makes up a smaller *proportion* of health care expenditure for unhealthy individuals compared to healthy individuals. For example, the individuals of type 1 who have Medigap insurance spend about 215% more than their counterparts with no Medigap, while individuals of type 5 who have Medigap spend only about 19% more.

The moral hazard effect of 215% for type 1 might seem very large, but note that this does not correspond to a large absolute expenditure increase (i.e. types 1 have average spending of $580 when uninsured and $1,830 when insured). Also note that for most individuals there is considerable posterior uncertainty about their type: e.g. almost no individual has a posterior probability of being type 1 equal to one. In the data, low expenditure individuals have high posterior probabilities of being types 1-2 and low posterior probabilities of being types 3-5, while the opposite is true for high expenditure individuals. When this individual-level uncertainty about the type is taken into account, estimates of the moral hazard effect are averages over type specific effects. For example, for individuals whose posterior type probability is highest for type 1, the average moral hazard effect is equal to 1,759 dollars, which makes up about 53% of their average expected expenditure in the Medicare only state (3,344 dollars). For individuals whose posterior modal type is 5, the moral hazard effect is equal to 5,334 dollars, which makes up about 17% of their average expected expenditure in the Medicare only state of 30,680 dollars.

This finding suggests that the price elasticity of health care demand decreases as health status deteriorates. This seems intuitive. For instance, much of the health expenditure for healthy low expenditure individuals may go towards treatment of minor ailments - treatment that one may fairly easily forgo due to cost. In contrast, expenditures for unhealthy individuals are presumably more often for essential treatment of serious illness.

The moral hazard effect of Medigap insurance for an individual with observable characteristics $\mathbf{x}_i$ can be computed as: $E(MH_i|\mathbf{x}_i, \boldsymbol{\theta}) = \sum_{j=1}^{m} \boldsymbol{\gamma}_j \cdot P(\text{type}_i = j|\mathbf{xw}_i, \boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_m)$.[27] The posterior mean (over the posterior of parameters) of $E(MH_i|\mathbf{x}_i, \boldsymbol{\theta})$ can be approximated as $10^{-3} \sum_{k=1}^{10^{-3}} \gamma^k_{s_i^k}$, where s_i^k are simulated as discussed in section 4.1. This posterior mean varies between \$1,333 and \$9,834 in our sample. The sample average of the moral hazard effect is equal to 2,119 dollars, which makes up about 32% of the average expenditure risk in the Medicare only state (6,476 dollars). The ratio of the average moral hazard effect to the average expenditure risk is at least comparable to the effect of insurance found in the RAND Health Insurance Experiment. For example, Manning et al. (1987) report that a decrease in the co-insurance rate from 25% to 0 increased total health care expenditure by 23%. Such a drop in co-pays is similar to the consequences of adopting many typical Medigap plans that cover co-pays.

It is interesting to see how different health types contribute to the aggregate increase in spending which would result from the moral hazard effect assuming we had universal Medigap coverage. As can be seen from Table 3, individuals of type 1 contribute about 23% of the increase in spending. This number is computed as the ratio of the type-specific moral hazard effect weighted by the type probability to the average moral hazard effect:

[27] Alternatively, we could define the moral hazard effect as the difference between the expected actual expenditure E_i of an individual with and without Medigap, i.e

$$E(MH_i|\mathbf{x}_i, \boldsymbol{\theta}) = \sum_{j=1}^{m} (E^1(E_i|\text{type}_i = j, \mathbf{xe}_i, \boldsymbol{\theta}) - E^0(E_i|\text{type}_i = j, \mathbf{xe}_i, \boldsymbol{\theta})) \cdot P(\text{type}_i = j|\mathbf{xw}_i, \boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_m)),$$

where

$$E^I(E_i|\text{type}_i = j, \mathbf{xe}_i, \boldsymbol{\theta}) = \Phi(\frac{\mathbf{xe}_i'\boldsymbol{\beta}_j + \gamma_j \cdot I}{(\sigma_j^2 + \sigma_{22})^{0.5}})[\mathbf{xe}_i'\boldsymbol{\beta}_j + \gamma_j \cdot I + (\sigma_j^2 + \sigma_{22})^{0.5} \cdot \frac{\phi(\frac{\mathbf{xe}_i'\boldsymbol{\beta}_j + \gamma_j \cdot I}{(\sigma_j^2 + \sigma_{22})^{0.5}})}{\Phi(\frac{\mathbf{xe}_i'\boldsymbol{\beta}_j + \gamma_j \cdot I}{(\sigma_j^2 + \sigma_{22})^{0.5}})}]$$

for $I = 0, 1$. This last expression is due to the Tobit specification for the distribution of actual expenditure. The two definitions of the moral hazard effect produce similar results.

$\frac{1250 \cdot 0.39}{2119} = 0.23$. Similarly, the contributions of individuals of types 2-5 to the aggregate increase in spending are 30%, 17%, 14% and 15%, respectively. Thus, the two healthiest types, who make up about 71% of the sample, give about 53% of the total spending increase. On the other hand, the three least healthy types, who make up 29% of population, account for 47% of the total increase in spending.

In contrast to the selection effect, restricting the covariance between unobservables ε_1 and ε_2, σ_{12}, to zero does have a noticeable effect on the estimate of the moral hazard effect. In such a specification the posterior mean of the moral hazard effect $E(MH_i|\mathbf{x}_i, \boldsymbol{\theta})$ is 1,315 dollars, compared to 2,119 in the full model. This drop in magnitude is not surprising given the large negative correlation between the unobervables shown in Table 4 (i.e. advantageous selection into Medigap). Once we control for this advantageous selection on unobservables, the moral hazard effect effect of insurance is revealed to be larger. This, even though ignoring σ_{12} does not have much impact on the estimated selection effect, it does significantly alter the estimate of the moral hazard effect.

It is also of interest to evaluate the potential effects on aggregate health expenditure of a policy which would expand Medigap coverage by making it more affordable. To this end we simulate a situation where the price of Medigap insurance drops sufficiently so that Medigap coverage increases by 10% (or 5 percentage points). According to the estimate of the price elasticity of health insurance demand in Buchmueller (2006), this would require approximately a 25$ drop in Medigap premiums.[28] The simulations suggest that the individuals who are attracted to Medigap insurance by this policy would on average spend 8.6 thousand dollars when Medigap-insured, compared to an average expenditure of 8 thousand dollars for

[28]Buchmueller (2006) estimates that a 5$ increase in an insurance premium would decrease a plan's enrollment by 2% in his sample of retirees over the age of 65. For estimation this study relies on changes in demand for different plans caused by an exogenous change in the retiree health insurance contributions policy of a single employer. Other studies which utilize natural experiment to estimate the elasticity of health insurance demand also suggest that this demand is relatively inelastic (e.g., Gruber and Washington (2005)). We use the estimate of Buchmueller (2006) because the demographic characteristics of individuals in his sample are similar to those in our data.

those who were already covered before the policy was implemented. The newly insured spend more in part because they have higher expenditure risk - their average expenditure risk in the Medicare only state (E_i^*) is 6.5 thousand dollars, compared to 5.9 thousand dollars for individuals who had Medigap coverage before the policy was implemented. But the newly insured also have a somewhat higher moral hazard effect (2.1 thousand dollars) than the previously insured (2.07 thousand dollars). This is because they come from a less healthy part of the population, and, as we have seen, moral hazard is inversely related to health. Thus, expanding Medigap coverage results in a somewhat higher cost per insured person due to both advantageous selection and moral hazard. But the increase in the average health expenditure of all insured individuals is still very small (from 8.02 to 8.06 thousand dollars). The policy increases per capita expenditure from 7.6 thousand dollars to 7.7 thousand dollars.[29]

In contrast, expanding Medigap coverage universally would have a large effect on expenditure, increasing per capita expenditure from 7.6 thousand dollars to 8.6 thousand dollars. The increase in expenditure is due to the moral hazard effect: the newly insured (who make up about 50% of the sample) increase their spending by 2.16 thousand dollars on average, which increases average expenditure by about 1 thousand dollars. Of course, the welfare consequences of expanding Medigap coverage cannot be evaluated using our model, but this is an interesting issue for future research.

[29]These calculations are based on artificial data samples of E_i^k, I_i^k and E_i^{*k} simulated as discussed in section 4.1 for two situations: (i) before the policy (the original posterior distribution of the parameters is used); (ii) after the policy (the intercept term in the insurance equation is increased by a constant to achieve the average Medigap coverage of 0.55, all random terms (e.g. $\varepsilon_{i1}, \varepsilon_{i2}, s_i, \eta_i$) are the same as in (i)). The average expenditure risk and moral hazard effects of the two groups, (i) with Medigap before the policy and (ii) with no Medigap before the policy but with Medigap after the policy, are computed as discussed above.

4.4 Marginal Effects of Covariates

Table 5 presents posterior means and the 5^{th} and 95^{th} percentiles of the posterior distributions of marginal effects of covariates on the following covariate-dependent functions of interest:

1. The expected expenditure risk (columns 1-3):[30]

$$E(E_i^*|\mathbf{xe}_i, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_m, \boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_m) = \sum_{j=1}^{m} E(E_i^*|\mathbf{xe}_i, \text{type}_i = j, \boldsymbol{\beta}_j) \cdot P(\text{type}_i = j|\mathbf{xw}_i, \boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_m), \tag{23}$$

   where $E(E_i^*|\mathbf{xe}_i, \text{type}_i = j, \boldsymbol{\beta}_j) = \mathbf{xe}_i'\boldsymbol{\beta}_j$ and $P(\text{type}_i = j|\mathbf{xw}_i, \boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_m)$ is given in equation (10);

2. The moral hazard effect of Medigap insurance on health care expenditure (columns 4-6):

$$E(MH_i|\mathbf{xw}_i, \gamma_1, ..., \gamma_m, \boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_m) = \sum_{j=1}^{m} \gamma_j \cdot P(\text{type}_i = j|\mathbf{xw}_i, \boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_m); \tag{24}$$

3. The unconditional standard deviation of the forecast error (columns 7-9):

$$SD(\eta_i|\mathbf{xw}_i, \sigma_1^2, ..., \sigma_m^2, \boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_m) = \left(\sum_{j=1}^{m} \sigma_j^2 \cdot P(\text{type}_i = j|\mathbf{xw}_i, \boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_m) \right)^{\frac{1}{2}}. \tag{25}$$

The marginal effects in Table 5 are computed for a median individual (i.e. an individual whose covariates $\mathbf{xe}_i$ are set to the sample median level) and are measured in thousands of dollars. For continuous covariates the effects are for a one standard deviation increase in the covariate from its sample median level, for categorical covariates the effect is from moving to the next category. The 5th and 95th percentiles reflect the uncertainty with respect to the posterior distribution of parameters, i.e. the effects of the covariates on the expressions

[30]Because in our analysis $\mathbf{xw}_i$ is a subset of $\mathbf{xe}_i$, as discussed in section (3.1), conditioning on $\mathbf{xe}_i$ is equivalent to conditioning on both $\mathbf{xw}_i$ and $\mathbf{xe}_i$

(23)-(25) were evaluated for 1000 draws from the posterior distribution of parameters, and the average and 5th and 95th percentiles of these 1000 values are reported in Table 5.

The expenditure risk in the Medicare only state $E(E_i^*|\mathbf{xe}_i, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_m, \boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_m)$ is lower for females, blacks and Hispanics. On average females are expected to spend 370 dollars less than males, while blacks and Hispanics are expected to spend 700 and 830 dollars less than other ethnic groups, respectively.[31] The expenditure risk is insensitive to age, conditional on detailed measures of health status, and is lower for married individuals. Unhealthy individuals are expected to spend more than their more healthy counterparts - a one standard deviation increase in the unhealthy factor 2 raises expenditure risk by 3.73 thousand dollars, while a one standard deviation increase in the healthy factor 3 decreases risk by 1.79 thousand dollars. Expenditure risk also varies by census division. In particular, residents of the New England census division have the highest expenditure risk, while residents of the East North Central, Pacific, South Atlantic and West South Central census divisions have the lowest expenditure risk, conditional on other variables.

The moral hazard effect of Medigap is higher for individuals in worse health, and is lower for Hispanics. Individuals living in Pacific census division have the lowest moral hazard effect, while individuals living in New England census division have the highest mean moral hazard effect. The standard deviation of the forecast error $SD(\eta_i|\mathbf{xw}_i, \sigma_1^2, ..., \sigma_m^2, \boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_m)$ is lower for females and is higher for less healthy individuals. The variance of the forecast error is the highest for individuals living in the New England census division, and is the lowest for individuals residing in the Pacific and West South Central census divisions.

[31] It is notable that race and marital status were not significant predictors of health expenditure in the OLS regression in Table 1, column C, but they are significant in the full model. This may be the result of a more flexible functional form for the conditional expectation of expenditure in the full model, compared to OLS. It may also be due to the bias in the Medigap coefficient due to failure of OLS to account for endogeneity of insurance.

Table 5: Marginal effects of individual characteristics on selected functions of interest

	$E(E_i^*\|\mathbf{xe}_i,\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_m,\boldsymbol{\delta}_1,...,\boldsymbol{\delta}_m)$		$E(MH_i\|\mathbf{xw}_i,\gamma_1,...,\gamma_m,\boldsymbol{\delta}_1,...,\boldsymbol{\delta}_m)$		$SD(\eta_i\|\mathbf{xw}_i,\sigma_1^2,...,\sigma_m^2,\boldsymbol{\delta}_1,...,\boldsymbol{\delta}_m)$	
Variable	Post. mean	$(5^{th},95^{th})$ prct.	Post. mean	$(5^{th},95^{th})$ prct.	Post. mean	$(5^{th},95^{th})$ prct.
	1	2	3	4	5	6
Female	-0.37	-0.86 , 0.08	-0.05	-0.13 , 0.02	-0.83	-1.32 , -0.37
Age	-0.10	-0.33 , 0.11	0.01	-0.01 , 0.04	-0.10	-0.32 , 0.13
New Eng	0.20	-0.53 , 0.96	0.14	0.02 , 0.30	1.01	-0.02 , 2.07
Mid Atl	-1.14	-1.87 , -0.48	-0.03	-0.15 , 0.07	-0.13	-1.16 , 0.81
East North Cent	-1.28	-2.03 , -0.59	-0.06	-0.18 , 0.07	-0.12	-1.25 , 1.03
West North Cent	-0.94	-1.67 , -0.28	0.00	-0.11 , 0.10	-0.12	-1.11 , 0.75
South Atl	-1.65	-2.45 , -0.89	-0.02	-0.15 , 0.10	-0.62	-1.71 , 0.33
East South Cent	-1.04	-1.76 , -0.36	-0.01	-0.13 , 0.10	0.09	-0.87 , 1.10
West South Cent	-1.37	-2.11 , -0.66	-0.11	-0.23 , 0.00	-0.97	-2.02 , -0.11
Mountain	-0.98	-1.68 , -0.31	-0.02	-0.14 , 0.08	-0.20	-1.19 , 0.68
Pacific	-3.25	-4.09 , -2.50	-0.26	-0.46 , -0.09	-2.49	-3.52 , -1.63
Health factor 2	3.73	2.98 , 4.64	0.32	0.13 , 0.53	2.06	1.33 , 2.92
Health factor 3	-1.79	-2.12 , -1.51	-0.15	-0.23 , -0.08	-1.03	-1.32 , -0.78
Health factor 7	0.85	0.61 , 1.15	0.07	0.02 , 0.13	0.44	0.10 , 0.92
Health factor 8	0.94	0.77 , 1.13	0.06	0.02 , 0.09	0.13	-0.12 , 0.41
Health factor 10	-0.36	-0.67 , -0.07	-0.02	-0.06 , 0.01	-0.36	-0.71 , 0.02
Health factor 11	-0.68	-0.86 , -0.53	-0.06	-0.10 , -0.03	-0.45	-0.66 , -0.23
Health factor 17	1.04	0.64 , 1.50	0.12	0.06 , 0.20	0.46	0.12 , 0.90
Health factor 20	0.05	-0.20 , 0.30	0.01	-0.02 , 0.04	-0.08	-0.37 , 0.26
Health factor 22	0.44	0.25 , 0.68	0.03	0.00 , 0.06	0.03	-0.22 , 0.35
Health factor 23	0.16	-0.04 , 0.40	0.04	0.00 , 0.08	0.28	-0.06 , 0.71
Black	-0.70	-1.21 , -0.14	-0.11	-0.21 , -0.01	-0.12	-0.77 , 0.81
Hispanic	-0.83	-1.32 , -0.33	-0.12	-0.20 , -0.05	-0.78	-1.32 , -0.24
Married	-0.37	-0.71 , -0.06	-0.01	-0.07 , 0.03	-0.42	-0.83 , -0.02
Year	-1.12	-1.41 , -0.84	-0.07	-0.12 , -0.02	-0.60	-0.94 , -0.28

* Note: The marginal effects are measured in thousand dollars and are evaluated for the median individual. For continuous covariates the effects are for a one standard deviation increase in the covariate from its sample median level. The omitted category for census division is the group with missing census division information.

5 Conclusion

This paper studies selection and moral hazard in the US Medigap health insurance market. Medigap is a collection of supplementary insurance plans sold by private companies to cover gaps in Medicare, a social insurance program providing health insurance coverage to senior citizens. We develop an econometric model for insurance demand and health care expenditure, in which the degree of selection is measured by the sensitivity of insurance demand to expected health care expenditure in the absence of Medigap insurance (our measure of expenditure risk). The model allows for correlation between unobservable determinants of expenditure risk and demand for insurance. This extends the analysis in Fang, Keane and Silverman (2008) who did not allow for correlated unobservables. To capture the complex shape of the expenditure distribution, we employ a smooth mixture of Tobit models generalizing the smoothly mixing regressions framework of Geweke and Keane (2007). To obtain the posterior distribution of parameters of the model we construct an MCMC algorithm with data augmentation.

We find that a specification which conditions Medigap insurance choice only on expenditure risk and insurance pricing variables suggests the existence of advantageous selection: that is, a one standard deviation increase in expenditure risk *decreases* probability of insurance coverage by 3 percentage points. However, when we condition on a range of potential sources of advantageous selection - including income, education, risk attitudes, cognitive ability, financial planning horizon, longevity expectation, race and marital status - we find that there is adverse selection into Medigap insurance. But this effect is modest: a one standard deviation increase in expenditure risk increases probability of insurance coverage by 3.7 percentage points. These findings are qualitatively and quantitatively similar to the results of Fang, Keane and Silverman (2008) (FKS). Hence, our first contribution is to show that the FKS results on the sources of advantageous selection are robust to correlation in unobservables.

Our second contribution, which goes beyond the findings of FKS, is that we provide estimates of the effects of "behavioral" SAS variables (i.e. risk attitudes, cognitive ability, financial planning horizon and longevity expectation) on the probability of Medigap coverage. We find that among these variables it is cognitive ability which has the largest effect: a one standard deviation increase in the cognitive ability factor increases the probability of having Medigap by 0.08 for a median individual. Variance in the health care expenditure forecast error also has a substantial effect on Medigap coverage - it decreases the Medigap probability by 4 percentage points per one sample standard deviation increase. Other "behavioral" SAS variables are less important: the effects of risk tolerance and financial planning horizon are close to zero, and longevity expectations have a modest effect. We also find that race is a source of adverse selection: blacks and hispanics are considerably less likely to purchase Medigap insurance (conditional on other determinants of Medigap status in the full model), and also have lower health care expenditure risk (conditional on health).

Notably this paper is the first to estimate selection and moral hazard effects jointly in the Medigap insurance market. We estimate the degree of moral hazard in the Medigap insurance market while accounting for the endogeneity of insurance choice. This third contribution goes well beyond the findings of FKS, who do not attempt to estimate moral hazard. We find that on average individuals with Medigap insurance coverage spend about \$2,119 more on health care than similar individuals without Medigap. This is a 32% increase, which is in the ballpark of moral hazard effects found in the RAND experiment.

We also estimate the sample distribution of the moral hazard (or price) effect of Medigap insurance on health care expenditure. We find that the moral hazard effect varies with individual characteristics - it is higher for individuals in worse health and is lower for Hispanics. Moreover, the moral hazard effect differs by risk level - individuals with higher expenditure risk tend to have a larger moral hazard effect in absolute terms, but a smaller moral hazard effect in *percentage* terms. This suggests that the price elasticity of health care demand decreases with health. This makes intuitive sense - healthy people are more likely to have ailments that can be relatively easy forgone due to price of health care, while people in poor health are more likely to have ailments which require treatments that are less easily forgone. We also find some interesting differences by region. Both the expenditure risk and the moral hazard effect of Medigap are greatest in New England, while the moral hazard effect is lowest in the Pacific and West South Central census division.

Finally, we simulate the effects of policies which would (i) reduce the price of the Medigap insurance to achieve a 10% increase in coverage, or (ii) implement universal coverage. Our results suggest that such policies will result in higher total health care expenditure, both per capita and per insured, because they attract individuals with a higher expenditure risk and a higher moral hazard effect to the pool of the insured. These effects are quantitatively very small for the policy that increases coverage by 10%. Expanding Medigap coverage to all individuals has a much larger effect: this policy would increase per capita health care expenditure by about $1,000 due to moral hazard effect. The welfare consequences of these policies is obviously an important issue that warrants further research.

Appendix

A-1. Prior Distributions

We specify the following prior distributions:

1. $\alpha_0 \sim N(\underline{\alpha}_0, \underline{h}_{\alpha_0}^{-1})$, where $\underline{\alpha}_0 = 0$ and the prior variance $\underline{h}_{\alpha_0}^{-1} = 0.4$. This specification implies that for an individual whose probability of Medigap coverage is equal to 0.5, the effect of a one sample standard deviation increase in expected expenditure (in Medicare only state) on this probability is centered at zero, while the 1st, 25th, 75th

and 99th percentiles of this effect are -0.31, -0.10, 0.10, 0.31, respectively.[32] That is, this prior reflects the belief that the effect of E_i^* is not very large, but still places a non-negligible probability on the event that a one standard deviation increase in E_i^* can change the Medigap coverage substantially (e.g., from 0.5 to 0.80 or 0.20).

2. $\boldsymbol{\alpha} \sim N\left(\underline{\boldsymbol{\alpha}}, \underline{\mathbf{H}}_{\alpha}^{-1}\right)$, where $\underline{\boldsymbol{\alpha}} = \mathbf{0}$ and the variance-covariance $\underline{\mathbf{H}}_{\alpha}^{-1}$ is a diagonal matrix which allows for reasonable prior uncertainty about the effects of the variables on the probability of Medigap coverage. In Table A-1 we present the diagonal elements of $\underline{\mathbf{H}}_{\alpha_1}^{-1}$ for the continuous variables, as well as the implied effects of a one sample standard deviation increase in these variables on the probability of Medigap coverage at the Medigap probability of 0.5.[33] The diagonal elements of $\underline{\mathbf{H}}_{\alpha_1}^{-1}$ for the intercept and for the indicator variables (i.e., census division, gender, education and income categories) are set to 1, so the prior distribution of the effect of increasing the indicator variables from 0 to 1 (evaluated at a Medigap probability of 0.5) are -0.49, -0.23, 0.23, 0.49 at 1st, 25th and 75th and 99th percentiles respectively.

3. $\boldsymbol{\beta}_j \sim N\left(\underline{\boldsymbol{\beta}}, \underline{\mathbf{H}}_{\beta}^{-1}\right) \quad$ for $j = 1, \ldots, m$. We specify that $\underline{\boldsymbol{\beta}} = [\overline{E}, \mathbf{0}_{K_E - 1}]'$, where $\overline{E}$ is the sample average of expenditure in the MCBS subsample and K_E is the size of $\mathbf{xe}_i$. The precision matrix $\underline{\mathbf{H}}_{\beta} = 0.1 \cdot \sum_{i \in MCBS} \mathbf{xe}_i \cdot \mathbf{xe}_i' / (N^M \cdot Var(E))$, where $Var(E)$ is the sample variance of expenditure in the MCBS subsample. This prior specification is

[32] Because the expected expenditure in the Medicare only state E_i^* is a latent variable, its distribution across individuals is unknown until the estimation is completed. To set the prior variance of α_0 we approximate E_i^* by the health expenditure risk computed using the FKS imputation methodology. In particular, we compute the expenditure risk as $\mathbf{xe}_i\mathbf{b}$, where $\mathbf{b}$ is a vector of least squares coefficients on health status characteristics $\mathbf{xe}_i$ from the regression of health care expenditure on $\mathbf{xe}_i$ and the Medigap insurance status in the MCBS subsample. The sample standard deviation of the imputed expenditure is equal to 0.59 for the expenditure measured in tens thousands of dollars.

[33] Similarly, the distribution of $\sigma_{s_i}^2$ across the individuals in the sample is not known until the estimation is completed. To set the prior variance of α_1 we approximate $\sigma_{s_i}^2$ by the imputed health expenditure variance computed using the FKS imputation methodology. In particular, we impute the expenditure variance as $\mathbf{xe}_i\mathbf{v}$, where $\mathbf{v}$ is a vector of least squares coefficients on heath status characteristics $\mathbf{xe}_i$ from the regression of $(E_i - \mathbf{xe}_i\mathbf{b} - I_i \cdot b_I)^2$ on $\mathbf{xe}_i$ and the Medigap insurance status in the MCBS subsample. The sample standard deviation of the imputed variance of expenditure is equal to 2.3 for the expenditure measures in tens thousands of dollars.

Table A-1: Prior distribution of $\boldsymbol{\alpha}_1$

Variable	Prior Variance	1st, 25th, 75th, 99th percentiles of marginal effect
$\sigma^2_{s_i}$	0.01	-0.20, -0.06, 0.06, 0.20
$\sigma^2_{s_i} \cdot$ risktol*	0.025	-0.09, -0.03, 0.03, 0.09
Age$^\sharp$	1.5	-0.36, -0.13, 0.13, 0.36
Age^2	3	
Age^3	3	
risktol	10	-0.35, -0.12, 0.12, 0.35
cogn	3	-0.39, -0.14, 0.14, 0.39
finpln	0.15	-0.39, -0.14, 0.14, 0.39
praliv75	0.35	-0.39, -0.14, 0.14, 0.39

* This marginal effect corresponds to the change in the effect of $\sigma^2_{s_i}$ brought about by two HRS sample standard deviations change in risk tolerance.

$^\sharp$ The marginal effect corresponds to the total effect when age changes by one sample standard deviation.

based on the prior for the normal linear regression model proposed in Geweke (2005), Chapter 5. It specifies considerable prior uncertainty about the effects of a one standard deviation change in a covariate on the expenditure. In particular, the implied 1st and 99th percentiles of the effect on expenditure of a one standard deviation increase in any of the covariates is $\pm$ 18.5 thousand dollars or more, which is enough to take expenditure from the 50th to the 90th percentile of it's sample distribution.

4. $\gamma_j \sim N(\underline{\gamma}, \underline{h}_\gamma^{-1})$, where $\underline{\gamma} = 0$ and $\underline{h}_\gamma = 0.01$ for $j = 1, ..., m$. This prior allows for substantial prior uncertainty about the effect of Medigap on health expenditure, e.g. the 1st and 99th percentiles of this effect are $\pm$23.3 thousand dollars.

5. $\underline{S} h_j \sim \chi^2(\underline{V})$, where $\underline{V} = 1$ and $\underline{S} = 0.59$ for $j = 1, ..., m$. This prior allows for substantial prior uncertainty about the type-specific variance of $\widehat{E}_i$. The interval constructed of the 10th and 90th percentiles of the prior distribution of $1/h_j$ is $[0.63, 108]$, which contains the variance of the observed expenditure (equal to 2.13 for expenditure measured in tens of thousands of dollars).

6. $\boldsymbol{\lambda}_k \sim N\left(\underline{\boldsymbol{\lambda}}_k, \underline{\mathbf{H}}_{\lambda_k}^{-1}\right)$ for $k = 1, ..., 4$, where $\underline{\boldsymbol{\lambda}}_k = [\bar{c}_k, \mathbf{0}_{K_c-1}]'$, $\bar{c}_k$ denotes sample average of c_{ki} in the HRS subsample, and K_c is the number of covariates in the vector $\mathbf{xc}_i$. The prior precision $\underline{\mathbf{H}}_{\lambda_k} = \sum_{i \in HRS} \mathbf{xc}_i \cdot \mathbf{xc}_i'/(N^H \cdot Var(c_k))$ for $k = 1, ..., K_c$, where $Var(c_k)$ is the sample variance of c_k in the HRS subsample. The prior distributions of $\boldsymbol{\lambda}_k$ are independent.

7. $V_c^{-1} \equiv H_c \sim Wishart\left(\underline{V}_c, \underline{S}_c^{-1}\right)$, so that the expectation of H_c is equal to $\underline{V}_c \cdot \underline{S}_c^{-1}$. We set $\underline{V}_c = 4$ and specify that S_c is a diagonal matrix with diagonal elements $h_{c,kk}$ equal to $\underline{V}_c \cdot 0.7 \cdot Var(c_k)$. This prior is based on that for the normal linear regression model proposed in Geweke (2005), Chapter 5, and specifies that for each c_k the population multiple correlation coefficient $1 - \frac{1}{Var(c_k)h_{c,kk}}$ is equal to 0.3 at the prior expectation of H_c. The prior probability that this coefficient is greater than 50% is 23%.

8. $s_{22}h_{22} \sim \chi^2(\underline{V}_\sigma)$, where $\underline{V}_\sigma = 1$ and $s_{22} = 0.039$. This prior sets the population multiple correlation coefficient $1 - \frac{1}{Var(E_i^*)h_{22}}$ to 0.90 at the prior expectation of h_{22}. This reflects a prior belief that the fraction of the variance in expected expenditure E_i^* due to unobserved determinants is much lower than that due to the wide array of observed health status characteristics that we use.[34] It seems plausible that expected expenditure is mostly due to observable health factors. However, our prior also allows for substantial prior variability in this coefficient. For example, the prior probability that it is less than 0.30 is 0.31. At the same time, the prior of h_{22} is flexible enough that unobserved factors can account for all variability in health care expenditure: the prior probability that $1/h_{22}$ is greater than the sample variance of E_i (equal to 2.13 for expenditure measured in tens thousands of dollars) is 0.11.

9. $\sigma_{12} \sim N\left(\underline{\sigma_{12}}, \underline{h}_{12}^{-1}\right)$, where $\underline{\sigma_{12}} = 0$ and $\underline{h}_{12}^{-1} = 50$. Together, the prior specifications

[34] To set this prior distribution we approximate the sample distribution of E_i^* by the imputed health care expenditure in Medicare only state, as described in the footnote to the discussion of the prior of α_0 (bullet point 1). Hence, we set $Var(E^*)$ to 0.59^2 for E^* measured in tens thousands of dollars.

of h_{22} and σ_{12} imply a significant uncertainty about the strength of the relationship between the unobservables, i.e. it specifies that the 1st, 25th, 75th and 99th percentiles of the prior distribution of the correlation coefficient between ε_1 and ε_2, $\frac{\sigma_{12}}{\sqrt{(1+h_{22}\cdot\sigma_{12}^2)h_{22}^{-1}}}$, are -0.91, -0.25, 0.25, 0.91, respectively.

10. $\boldsymbol{\delta}_j \sim N\left(\underline{\boldsymbol{\delta}}, \underline{\mathbf{H}}_\delta^{-1}\right)$ for $j = 1, \ldots, m-1$, where we specify $\underline{\boldsymbol{\delta}} = \mathbf{0}$ and $\underline{\mathbf{H}}_\delta = 0.1 \cdot \sum_{i \in MCBS} \mathbf{xw}_i \cdot \mathbf{xw}_i'/N^M$.

A-2. Posterior Simulation Algorithm

To obtain the posterior distribution of parameters of the model we construct a Gibbs sampling algorithm. We split the parameters vector into several blocks introduced in section 3.3 so that it is relatively easy to sample from the conditional posterior distributions of each block. Let $\boldsymbol{\theta}_{-\theta_k}$ denote the vector of parameters $\boldsymbol{\theta}$ with the block of parameters θ_k removed. The Gibbs sampler iteratively draws from the conditional posterior distributions of the following blocks of parameters and latent data:

1. The posterior conditional distribution of α_0, $p(\alpha_0|\boldsymbol{\theta}_{-\alpha_0}, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{s}, \mathbf{W}, \mathbf{data})$, is proportional to the product of it's prior density $p(\alpha_0)$ given in Section A-1 and the density of observable and latent data $p(\mathbf{I}^*, \mathbf{I}, \widehat{\mathbf{E}}, \mathbf{E}^o, \mathbf{C}^o, \mathbf{C}^m, \mathbf{s}, \mathbf{W}|\mathbf{S}, \mathbf{X}, \alpha_0, \boldsymbol{\theta}_{-\alpha_0})$ given in equation (21). This distribution is not of any known form and is sampled using the random walk Metropolis-Hastings algorithm. In particular, on iteration n the algorithm draws a proposal value $\widetilde{\alpha}_0$ from $N(\alpha_0^{n-1}, v_{\alpha_0})$, where the subscript $n-1$ indicates the value of α_0 from a previous iteration of the Gibbs sampler. The proposal $\widetilde{\alpha}_0$ is accepted as the new draw α_0^n with probability

$$\rho_{\alpha_0} = \min\{1, \frac{p(\mathbf{I}^*, \mathbf{I}, \widehat{\mathbf{E}}, \mathbf{E}^o, \mathbf{C}^o, \mathbf{C}^m, \mathbf{s}, \mathbf{W}|\mathbf{S}, \mathbf{X}, \widetilde{\alpha}_0, \boldsymbol{\theta}_{-\alpha_0})p(\widetilde{\alpha}_0)}{p(\mathbf{I}^*, \mathbf{I}, \widehat{\mathbf{E}}, \mathbf{E}^o, \mathbf{C}^o, \mathbf{C}^m, \mathbf{s}, \mathbf{W}|\mathbf{S}, \mathbf{X}, \alpha_0^{n-1}, \boldsymbol{\theta}_{-\alpha_0})p(\alpha_0^{n-1})}\}.$$

The variance of the proposal distribution v_{α_0} was set so that 45% of the new draws were accepted, as recommended in Roberts, Gelman, and Gilks (1997).

2. The posterior conditional distribution of $\boldsymbol{\alpha}$, $p(\boldsymbol{\alpha}|\boldsymbol{\theta}_{-\alpha}, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{s}, \mathbf{W}, \mathbf{data})$, is proportional to the product of it's prior density $p(\boldsymbol{\alpha})$ given in Section A-1 and the density of observable and latent data $p(\mathbf{I}^*, \mathbf{I}, \widehat{\mathbf{E}}, \mathbf{E}^o, \mathbf{C}^o, \mathbf{C}^m, \mathbf{s}, \mathbf{W}|\mathbf{S}, \mathbf{X}, \boldsymbol{\theta})$ given in equation (21). To derive the posterior conditional distribution of $\boldsymbol{\alpha}$ we first need to establish some notation.

   Let $V_{11} = 1 + \frac{\sigma_{12}^2}{\sigma_{22}} + 2\alpha_0\sigma_{12} + \alpha_0^2\sigma_{22}$ and $V_{12} = \sigma_{12} + \alpha_0\sigma_{22}$. Without loss of generality assume that the observations are arranged so that the first N^M observations belong to MCBS subset, and the last N^H belong to the HRS subset. Let $\widetilde{I}_i^* = I_i^* - \alpha_0\boldsymbol{\beta}_{s_i}'\mathbf{xe}_i$, and $\widetilde{\widetilde{I}}_i^* = \widetilde{I}_i^* - \frac{V_{12}}{\sigma_{s_i}^2+\sigma_{22}}(\widehat{E}_i - \boldsymbol{\beta}_{s_i}'\mathbf{xe}_i - \gamma_{s_i}I_i)$. Let $\widetilde{\mathbf{I}}_S^*$ denote the vector of elements $\widetilde{I}_i^*$ for $i = 1, ..., N^M$ (MCBS observations), while $\widetilde{\mathbf{I}}_{1-S}^*$ be the vector of elements $\widetilde{I}_i^*$ for $i = N^M + 1, ..., N$ (HRS observations). Let $\widetilde{\widetilde{\mathbf{I}}}_S^*$ be defined similarly for the elements $\widetilde{\widetilde{I}}_i^*$, $i = 1, ..., N^M$ (MCBS observations). Also, let $\mathbf{Z}_S$ be the matrix with the rows $[\sigma_{s_i}^2, \sigma_{s_i}^2 \cdot C_{1i}, \mathbf{xi}_i', C_{1i}, ..., C_{4i}]$ for $i = 1, ..., N_M$, while $\mathbf{Z}_{1-S}$ be the matrix with these rows for $i = N^M + 1, ..., N$, where $C_{ki} = c_{ki}^m \cdot S_i + c_{ki}^o \cdot (1 - S_i)$.

   Then, it can be shown that the posterior conditional distribution of $\boldsymbol{\alpha}$ is given by:

$$p(\boldsymbol{\alpha}|\boldsymbol{\theta}_{-\alpha}, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{s}, \mathbf{W}, \mathbf{data}) \sim N(\overline{\boldsymbol{\alpha}}, \overline{\mathbf{H}}_\alpha^{-1}),$$

   where

$$\overline{\mathbf{H}}_\alpha = \underline{\mathbf{H}}_\alpha + \frac{1}{V_{11}}\mathbf{Z}_{1-S}'\mathbf{Z}_{1-S} + \mathbf{Z}_S'\mathbf{Q}_\alpha\mathbf{Z}_S$$

   and

$$\overline{\boldsymbol{\alpha}} = \overline{\mathbf{H}}_\alpha^{-1}[\underline{\mathbf{H}}_\alpha\underline{\boldsymbol{\alpha}} + \frac{1}{V_{11}}\mathbf{Z}_{1-S}'\widetilde{\mathbf{I}}_{1-S}^* + \mathbf{Z}_S'\mathbf{Q}_\alpha\widetilde{\widetilde{\mathbf{I}}}_S^*],$$

   and where $\mathbf{Q}_\alpha$ is the $N^M \times N^M$ diagonal matrix with the ii^{th} element given by

$\frac{1}{V_{11}-\frac{V_{12}^2}{\sigma_{s_i}^2+\sigma_{22}}}$.

3. The posterior conditional distribution of $\boldsymbol{\beta}_j$, $p(\boldsymbol{\beta}_j|\boldsymbol{\theta}_{-\beta_j}, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{s}, \mathbf{W}, \mathbf{data})$, is proportional to the product of it's prior density $p(\boldsymbol{\beta}_j)$ given in Section A-1 and the density of observable and latent data given in equation (21). To derive this distribution for $j = 1, ..., m$ we need to establish the following notation:

   Let $\widehat{I}_i^* = I_i^* - [\sigma_{s_i}^2, \sigma_{s_i}^2 \cdot C_{1i}, \mathbf{xi}_i', C_{1i}, ..., C_{4i}]\boldsymbol{\alpha}$, $\widetilde{E}_i = \widehat{E}_i - I_i\gamma_{s_i}$, $\widetilde{\mathbf{xe}}_i = \alpha_0 \cdot \mathbf{xe}_i$. Define $\mathbf{XE}_S^j$ and $\widetilde{\mathbf{XE}}_S^j$ as the matrices with the rows $\mathbf{xe}_i$ and $\widetilde{\mathbf{xe}}_i$ respectively for observations i such that $s_i = j$ and $i \in MCBS$. Let $\mathbf{XE}_{1-S}^j$ and $\widetilde{\mathbf{XE}}_{1-S}^j$ be similarly constructed matrices for observations $i \in HRS$. Similarly, let $\widehat{\mathbf{I}}_S^{*j}$ and $\widehat{\mathbf{I}}_{1-S}^{*j}$ denote the vectors of $\widehat{I}_i^*$ for i with $s_i = j$ and $i \in MCBS$, or $s_i = j$ and $i \in HRS$, respectively. Let $\widetilde{\mathbf{E}}_S^j$ denote the vector of $\widetilde{E}_i$ for i such that $s_i = j$ and $i \in MCBS$. Also, let the matrix F^j with the elements f_{kl}^j be defined as

$$F^j = \begin{bmatrix} V_{11} & V_{12} \\ V_{12} & \sigma_{22} + \sigma_j^2 \end{bmatrix}^{-1}.$$

   Then for $j = 1, ..., m$ the posterior conditional distribution of $\boldsymbol{\beta}_j$ is independent of $\boldsymbol{\beta}_l$ for $l \neq j$ and is given by:

$$\boldsymbol{\beta}_j|(\boldsymbol{\theta}_{-\beta_j}, \widehat{\mathbf{E}}, \mathbf{E}^o, \mathbf{I}^*, \mathbf{C}^o, \mathbf{C}^m, \mathbf{s}, \mathbf{W}, \mathbf{data}) \sim N(\overline{\boldsymbol{\beta}}_j, \overline{\mathbf{H}}_{\beta_j}^{-1}),$$

   where

$$\overline{\mathbf{H}}_{\beta_j} = \underline{\mathbf{H}}_\beta + f_{11}^j \widetilde{\mathbf{XE}}_S^{j'} \widetilde{\mathbf{XE}}_S^j + 2f_{12}^j \widetilde{\mathbf{XE}}_S^{j'} \mathbf{XE}_S^j + f_{22}^j \mathbf{XE}_S^{j'} \mathbf{XE}_S^j + \frac{1}{V_{11}} \widetilde{\mathbf{XE}}_{1-S}^{j'} \widetilde{\mathbf{XE}}_{1-S}^j$$

and

$$\overline{\boldsymbol{\beta}}_j = \overline{\mathbf{H}}_{\beta_j}^{-1}[\underline{\mathbf{H}}_\beta\underline{\boldsymbol{\beta}} + f_{11}^j\widetilde{\mathbf{XE}}_S^{j'}\widehat{\mathbf{I}}_S^{*j} + f_{12}^j\widetilde{\mathbf{XE}}_S^{j'}\widetilde{\mathbf{E}}_S^j + f_{12}^j\mathbf{XE}_S^{j'}\widehat{\mathbf{I}}_S^{*j} + f_{22}^j\mathbf{XE}_S^{j'}\widetilde{\mathbf{E}}_S^j + \frac{1}{V_{11}}\widetilde{\mathbf{XE}}_{1-S}^{j'}\widehat{\mathbf{I}}_{1-S}^{*j}].$$

4. The posterior conditional distribution of γ_j, $p(\gamma_j|\boldsymbol{\theta}_{-\gamma_j}, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{s}, \mathbf{W}, \mathbf{data})$, is proportional to the product of it's prior density $p(\gamma_j)$ given in Section A-1 and the density of observable and latent data given in equation (21). To derive the posterior conditional distribution of γ_j we need to establish the following notation. Let $\breve{I}_i^* = I_i^* - \alpha_0\mathbf{xe}_i\boldsymbol{\beta}_{s_i} - [\sigma_{s_i}^2, \sigma_{s_i}^2 \cdot C_{1i}, \mathbf{xi}_i', C_{1i}, ..., C_{4i}]\boldsymbol{\alpha}$, and $\breve{\mathbf{I}}_S^{*j}$ denote a vector of $\breve{I}_i^*$ for i with $s_i = j$ and $i \in MCBS$. Also, let $\breve{E}_i = \widehat{E}_i - \mathbf{xe}_i\boldsymbol{\beta}_{s_i}$, and $\breve{\mathbf{E}}_S^j$ denote a vector of $\breve{E}_i$ for i with $s_i = j$ and $i \in MCBS$. Let $\mathbf{I}_S^j$ denote the vector of I_i for i with $s_i = j$ and $i \in MCBS$.

   Then, the posterior conditional distribution of γ_j for $j = 1, ...m$ is independent of $l \neq j$ and is given by:

$$\gamma_j|(\boldsymbol{\theta}_{-\gamma_j}, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{s}, \mathbf{W}, \mathbf{data}) \sim N(\overline{\gamma}_j, \overline{h}_{\gamma_j}^{-1}),$$

   where

$$\overline{h}_{\gamma_j} = \underline{h}_\gamma + f_{22}^j\mathbf{I}_S^{j'}\mathbf{I}_S^{j'}$$

   and

$$\overline{\gamma_j} = \overline{h}_{\gamma_j}^{-1}(\underline{h}_\gamma\underline{\gamma} + \mathbf{I}_S^{j'}(f_{22}^j\breve{\mathbf{E}}_S^j + f_{12}^j\breve{\mathbf{I}}_S^{*j})).$$

5. The posterior conditional distributions of h_j for $j = 1, ...m$ are proportional to the product of prior density of h_j, $p(h_j)$ given in Section A-1 and the density of observable and latent data as defined in (21). It is easy to see that the posterior conditional distributions of h_j for $j = 1, ...m$ are independent of those of h_l for $l \neq j$. For all j

the posterior conditional distribution of h_j is not of any known form and is sampled using the Metropolis-Hastings algorithm. In particular, on iteration n we draw the proposal value $\widetilde{h_j}$ from gamma distribution with the parameters $(\frac{v_j}{2}, \frac{2h_j^{(n-1)}}{v_j})$. Note, that the expected value of this distribution is equal to $h_j^{(n-1)}$. We set the parameters v_j for $j = 1, ..., m$ so that the new draws are accepted with a probability of 0.45. Denote the probability density of this proposal gamma distribution as $g(\widetilde{h_j}|h_j^{(n-1)})$. We accept $\widetilde{h_j}$ as the new draw h_j^n with probability

$$\rho_{h_j} = \min\{1, \frac{p(\mathbf{I}^*, \mathbf{I}, \widehat{\mathbf{E}}, \mathbf{E}^o, \mathbf{C}^o, \mathbf{C}^m, \mathbf{s}, \mathbf{W}|\mathbf{S}, \mathbf{X}, \widetilde{h_j}, \boldsymbol{\theta}_{-h_j})p(\widetilde{h_j})g(h_j^{(n-1)}|\widetilde{h_j})}{p(\mathbf{I}^*, \mathbf{I}, \widehat{\mathbf{E}}, \mathbf{E}^o, \mathbf{C}^o, \mathbf{C}^m, \mathbf{s}, \mathbf{W}|\mathbf{S}, \mathbf{X}, h_j^{(n-1)}, \boldsymbol{\theta}_{-h_j})p(h_j^{(n-1)})g(\widetilde{h_j}|h_j^{(n-1)})}\}$$

6. The posterior conditional distribution of Λ, $p(\Lambda|\boldsymbol{\theta}_{-\Lambda}, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{s}, \mathbf{W}, \mathbf{data})$, is proportional to the product of it's prior density $p(\Lambda)$ given in Section A-1 and the density of observable and latent data given in equation (21). To obtain the conditional posterior distribution of $\boldsymbol{\Lambda}$ we need to establish the following notation. Let $\mathbf{XC}$ denote the matrix of covariates $\mathbf{xc}_i$ for observations $i = 1, ..., N$, i.e. $\mathbf{XC} = [\mathbf{xc}_1, ..., \mathbf{xc}_N]'$. Let D_K denote the identity matrix of size K and let $\mathbf{Z}_\Lambda = D_4 \otimes \mathbf{XC}$, where $\otimes$ denotes the Kroneker product.

   Then, the posterior conditional distribution of $\boldsymbol{\Lambda}$ is given by:

$$\boldsymbol{\Lambda}|(\boldsymbol{\theta}_{-\Lambda}, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{s}, \mathbf{W}, \mathbf{data}) \sim N(\overline{\boldsymbol{\Lambda}}, \overline{\mathbf{H}}_\Lambda^{-1}),$$

   where

$$\overline{\mathbf{H}}_\Lambda = \underline{\mathbf{H}}_\Lambda + \mathbf{Z}_\Lambda'(V_c \otimes D_N)^{-1}\mathbf{Z}_\Lambda$$

   and

$$\overline{\boldsymbol{\Lambda}} = \overline{\mathbf{H}}_\Lambda^{-1}[\underline{\mathbf{H}}_\Lambda\underline{\boldsymbol{\Lambda}} + \mathbf{Z}_\Lambda'(V_c \otimes D_N)^{-1}\mathbf{C}].$$

In the above expressions $\underline{\mathbf{H}}_\Lambda$ is a block-diagonal matrix with the diagonal blocks $\underline{\mathbf{H}}_{\lambda_k}$, $k = 1, ..., 4$, $\underline{\mathbf{\Lambda}} = [\underline{\boldsymbol{\lambda}}_1', \underline{\boldsymbol{\lambda}}_2', \underline{\boldsymbol{\lambda}}_3', \underline{\boldsymbol{\lambda}}_4']'$, $\mathbf{C} = [\mathbf{C}_1', ..., \mathbf{C}_4']'$, and for $k = 1, ..., 4$ the vectors $\mathbf{C}_k$ consist of the elements $C_{ki} = c_{ki}^m \cdot S_i + c_{ki}^o \cdot (1 - S_i)$.

7. The posterior conditional distribution of the inverse of the variance-covariance matrix of the SAS variables missing from the MCBS data, $H_c \equiv V_c^{-1}$, $p(H_c|\boldsymbol{\theta}_{-H_c}, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{s}, \mathbf{data})$, is proportional to the product of its prior probability given in Section A-1 and the density of observable and latent data as defined in (21), and is given by:

$$H_c|(\boldsymbol{\theta}_{-H_c}, \widehat{\mathbf{E}}, \mathbf{E}^o, \mathbf{I}^*, \mathbf{C}^m, \mathbf{C}^o, \mathbf{W}, \mathbf{s}, \mathbf{data}) \sim W((\underline{S}_c + S_c)^{-1}, \underline{V}_c + N),$$

where

$$S_c = \begin{bmatrix} (\mathbf{C}_1 - \mathbf{XC}\boldsymbol{\lambda}_1)'(\mathbf{C}_1 - \mathbf{XC}\boldsymbol{\lambda}_1) & \cdots & (\mathbf{C}_1 - \mathbf{XC}\boldsymbol{\lambda}_1)'(\mathbf{C}_4 - \mathbf{XC}\boldsymbol{\lambda}_4) \\ \vdots & \ddots & \vdots \\ (\mathbf{C}_4 - \mathbf{XC}\boldsymbol{\lambda}_{4x})'(\mathbf{C}_1 - \mathbf{XC}\boldsymbol{\lambda}_1) & \cdots & (\mathbf{C}_4 - \mathbf{XC}\boldsymbol{\lambda}_4)'(\mathbf{C}_4 - \mathbf{XC}\boldsymbol{\lambda}_4) \end{bmatrix}.$$

8. The posterior conditional distribution of h_{22} is proportional to the product of it's prior density $p(h_{22})$ given in Section A-1 and the density of observable and latent data given in equation 21. This distribution is not of any known form and is sampled using the Metropolis-Hastings algorithm.

   In particular, on iteration n we draw the proposal value $\widetilde{h_{22}}$ from gamma distribution with the parameters $(\frac{v_{s_2}}{2}, \frac{2h_{22}^{n-1}}{v_{s_2}})$. Note, that the expected value of this distribution is equal to h_{22}^{n-1}. We set the parameter v_{s_2} so that the acceptance rate is about 45%, as recommended in Roberts, Gelman, and Gilks (1997). Denote the probability density of this proposal gamma distribution as $g(\widetilde{h_{22}}|h_{22}^{n-1})$. We accept $\widetilde{h_{22}}$ as the new draw

h_{22}^n with probability

$$\rho_{\sigma_{22}} = \min\{1, \frac{p(\mathbf{I}^*, \mathbf{I}, \widehat{\mathbf{E}}, \mathbf{E}^o, \mathbf{C}^o, \mathbf{C}^m, \mathbf{s}, \mathbf{W}|\mathbf{S}, \mathbf{X}, \widetilde{h_{22}}, \boldsymbol{\theta}_{-h_{22}})p(\widetilde{h_{22}})g(h_{22}^{n-1}|\widetilde{h_{22}})}{p(\mathbf{I}^*, \mathbf{I}, \widehat{\mathbf{E}}, \mathbf{E}^o, \mathbf{C}^o, \mathbf{C}^m, \mathbf{s}, \mathbf{W}|\mathbf{S}, \mathbf{X}, h_{22}^{n-1}, \boldsymbol{\theta}_{-h_{22}})p(h_{22}^{n-1})g(\widetilde{h_{22}}|h_{22}^{n-1})}\}$$

9. The posterior conditional distribution of σ_{12} is proportional to the product of it's prior density $p(\sigma_{12})$ given in Section A-1 and the density of observable and latent data given in equation (21). This distribution is not of any known form and is sampled using the random walk Metropolis-Hastings algorithm.

   In particular, on iteration n draw the proposal value $\widetilde{\sigma}_{12}$ from $N(\sigma_{12}^{n-1}, v_{\sigma_{12}})$. Accept $\widetilde{\sigma}_{12}$ as the new draw σ_{12}^n with probability

$$\rho_{\sigma_{12}} = \min\{1, \frac{p(\mathbf{I}^*, \mathbf{I}, \widehat{\mathbf{E}}, \mathbf{E}^o, \mathbf{C}^o, \mathbf{C}^m, \mathbf{s}, \mathbf{W}|\mathbf{S}, \mathbf{X}, \widetilde{\sigma}_{12}, \boldsymbol{\theta}_{-\sigma_{12}})p(\widetilde{\sigma_{12}})}{p(\mathbf{I}^*, \mathbf{I}, \widehat{\mathbf{E}}, \mathbf{E}^o, \mathbf{C}^o, \mathbf{C}^m, \mathbf{s}, \mathbf{W}|\mathbf{S}, \mathbf{X}, \sigma_{12}^{n-1}, \boldsymbol{\theta}_{-\sigma_{12}})p(\sigma_{12}^{n-1})}\}.$$

   The variance of the proposal distribution $v_{\sigma_{12}}$ was set so that 45% of the new draws are accepted, as recommended in Roberts, Gelman, and Gilks (1997).

10. The posterior conditional distribution of the vector of coefficients $\boldsymbol{\delta}_j$ which determine the latent type propensities $\widetilde{W}_{ij}$ is proportional to the product of the prior density of $\boldsymbol{\delta}_j$ given in Section A-1 and (21). It is easy to see that the posterior conditional distributions of $\boldsymbol{\delta}_j$ are independent across j and are given by:

$$\boldsymbol{\delta}_j|(\boldsymbol{\theta}_{-\delta_j}, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{s}, \mathbf{data}) \;\sim\; N(\overline{\boldsymbol{\delta}}_j, \overline{\mathbf{H}}_\delta^{-1})$$

   where

$$\begin{aligned}
\overline{\mathbf{H}}_\delta &= \underline{\mathbf{H}}_\delta + \mathbf{XW}'\mathbf{XW}, \\
\overline{\boldsymbol{\delta}}_j &= \overline{\mathbf{H}}_\delta^{-1}[\underline{\mathbf{H}}_\delta \underline{\boldsymbol{\delta}} + \mathbf{XW}'\widetilde{\mathbf{w}}_j] \text{ for } j = 1, ..., m-1.
\end{aligned}$$

11. Latent utility of health insurance $I_i^* \sim p(I_i^*|\boldsymbol{\theta}, \widehat{\mathbf{E}}, \mathbf{I}_{-i}^*, \mathbf{C}^m, \mathbf{s}, \mathbf{data})$. From (21) the kernel of this posterior distribution for $i \in MCBS$ is given by

$$\begin{aligned}
& \exp(-(I_i^* - \alpha_0 \mathbf{xe}_i \boldsymbol{\beta}_{s_i} - \alpha_1 \sigma_{s_i}^2 - \alpha_2 \sigma_{s_i}^2 C_{1i} - \boldsymbol{\alpha}_3' \mathbf{xi}_i - [C_{1i}, ..., C_{4i}]\boldsymbol{\alpha}_4 \\
- \quad & \frac{V_{12}}{\sigma_{s_i}^2 + \sigma_{22}} (\widehat{E}_i - \boldsymbol{\beta}_{s_i}' \mathbf{xe}_i - \gamma_{s_i} I_i))^2 / (2(V_{11} - \frac{V_{12}^2}{\sigma_{s_i}^2 + \sigma_{22}}))) \\
\cdot \quad & (\iota(I_i^* \geq 0) \cdot \iota(I_i = 1) + \iota(I_i^* < 0) \cdot \iota(I_i = 0)),
\end{aligned}$$

while for $i \in HRS$ it is given by

$$\begin{aligned}
& \exp(-(I_i^* - \alpha_0 \mathbf{xe}_i \boldsymbol{\beta}_{s_i} - \alpha_1 \sigma_{s_i}^2 - \alpha_2 \sigma_{s_i}^2 C_{1i} - \boldsymbol{\alpha}_3' \mathbf{xi}_i - [C_{1i}, ..., C_{4i}]\boldsymbol{\alpha}_4)^2 / (2V_{11})) \\
\cdot \quad & (\iota(I_i^* \geq 0) \cdot \iota(I_i = 1) + \iota(I_i^* < 0) \cdot \iota(I_i = 0)).
\end{aligned}$$

These can be recognized as kernels of truncated normal distributions. Thus,

$$I_i^* | (\boldsymbol{\theta}, \widehat{\mathbf{E}}, \mathbf{I}_{-i}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{s}, \mathbf{data}) \sim TN_{R(I_i)}(\overline{I}_i^*, V_{I^*}),$$

where $TN_{R(I)}(a, b)$ denotes normal distribution with mean a and variance b truncated to interval $R(I)$, $R(0) = (-\infty, 0]$, $R(1) = (0, \infty)$. For $i \in MCBS$ we have

$$\overline{I}_i^* = \alpha_0 \mathbf{xe}_i \boldsymbol{\beta}_{s_i} + \alpha_1 \cdot \sigma_{s_i}^2 + \alpha_2 \cdot \sigma_{s_i}^2 c_{1i}^m + \boldsymbol{\alpha}_3' \mathbf{xi}_i + \boldsymbol{\alpha}_4' \mathbf{c}_i^m + \frac{V_{12}}{\sigma_{s_i}^2 + \sigma_{22}} (E_i^* - \boldsymbol{\beta}_{s_i}' \mathbf{xe}_i - \gamma_{s_i} I_i),$$

$$V_{I^*} = V_{11} - \frac{V_{12}^2}{\sigma_{s_i}^2 + \sigma_{22}},$$

while for $i \in HRS$ we have

$$\overline{I}_i^* = \alpha_0 \mathbf{xe}_i \boldsymbol{\beta}_{s_i} + \alpha_1 \cdot \sigma_{s_i}^2 + \alpha_2 \cdot \sigma_{s_i}^2 c_{1i}^o + \boldsymbol{\alpha}_3' \mathbf{xi}_i + \boldsymbol{\alpha}_4' \mathbf{c}_i^o$$

and

$$V_{I^*} = V_{11}.$$

12. Notional expenditure $\widehat{E}_i \sim p(\widehat{E}_i|\boldsymbol{\theta}, \widehat{\mathbf{E}}_{-i}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{s}, \mathbf{data}, S_i = 1)$. The kernel of this posterior distribution is given by

$$\begin{aligned} &\exp(-\frac{(\widehat{E}_i - \mathbf{xe}_i\boldsymbol{\beta}_{s_i} - \gamma_{s_i} I_i - \frac{V_{12}}{V_{11}}(I_i^* - \alpha_0\mathbf{xe}_i\boldsymbol{\beta}_{s_i} - \alpha_1 \cdot \sigma_{s_i}^2 - \alpha_2 \cdot \sigma_{s_i}^2 c_{1i}^m - \boldsymbol{\alpha}_3'\mathbf{xi}_i - \boldsymbol{\alpha}_4'\mathbf{c}_i^m))^2}{2(\sigma_{s_i}^2 + \sigma_{22} - \frac{V_{12}^2}{V_{11}})}) \\ \cdot \quad &(\iota(E_i^o = \widehat{E}_i) \cdot \iota(\widehat{E}_i \geq 0) + \iota(E_i^o = 0) \cdot \iota(\widehat{E}_i < 0)). \end{aligned} \tag{26}$$

Thus, if $E_i = 0$ we draw notional expenditure from:

$$\widehat{E}_i|(\boldsymbol{\theta}, \widehat{\mathbf{E}}_{-i}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{s}, \mathbf{data}, S_i = 1) \;\sim\; TN_{(-\infty,0]}(\overline{E}_i, \sigma_{s_i}^2 + \sigma_{22} - \frac{V_{12}^2}{V_{11}})$$

where $\overline{E}_i = \mathbf{xe}_i\boldsymbol{\beta}_{s_i} + \gamma_{s_i} I_i + \frac{V_{12}}{V_{11}}(I_i^* - \alpha_0\mathbf{xe}_i\boldsymbol{\beta}_{s_i} - \alpha_1 \cdot \sigma_{s_i}^2 - \alpha_2 \cdot \sigma_{s_i}^2 c_{1i}^m - \boldsymbol{\alpha}_3'\mathbf{xi}_i - \boldsymbol{\alpha}_4'\mathbf{c}_i^m)$, while if $E_i > 0$ we simply set $\widehat{E}_i = E_i$.

13. SAS variables missing from the MCBS: $\mathbf{c}_i^m \sim p(\mathbf{c}_i^m|\boldsymbol{\theta}, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}_{-i}^m, \mathbf{W}, \mathbf{s}, \mathbf{data}, S_i = 1)$. The kernel of this posterior distribution is given by

$$\begin{aligned} &\exp(-(I_i^* - \alpha_0\mathbf{xe}_i\boldsymbol{\beta}_{s_i} - \alpha_1\sigma_{s_i}^2 - \alpha_2\sigma_{s_i}^2 c_{1i}^m - \boldsymbol{\alpha}_3'\mathbf{xi}_i - \boldsymbol{\alpha}_4'\mathbf{c}_i^m \\ - \quad &\frac{V_{12}}{\sigma_{s_i}^2 + \sigma_{22}}(\widehat{E}_i - \boldsymbol{\beta}_{s_i}'\mathbf{xe}_i - \gamma_{s_i} I_i)^2/(2(V_{11} - \frac{V_{12}^2}{\sigma_{s_i}^2 + \sigma_{22}}))) \\ \cdot \quad &\exp(-(\mathbf{c}_{1i}^m - XC_i\Lambda)'V_c^{-1}(\mathbf{c}_{1i}^m - XC_i\Lambda)/2) \end{aligned}$$

This kernel can be recognized as that of $p(\mathbf{c}_i^m|\boldsymbol{\theta}, \widehat{E}_i, \mathbf{xi}_i, \mathbf{xe}_i; I_i^*)$, where the joint conditional distribution $p(I_i^*, \mathbf{c}_i^m|\boldsymbol{\theta}, \widehat{E}_i, \mathbf{xi}_i, \mathbf{xe}_i)$ is multivariate normal with mean

$$\begin{bmatrix} \overline{I}_i^c \\ \overline{\mathbf{c}}_i \end{bmatrix} \equiv \begin{bmatrix} \alpha_0\mathbf{xe}_i\boldsymbol{\beta}_{s_i} + \alpha_1\sigma_{s_i}^2 + \alpha_2\sigma_{s_i}^2 \cdot \mathbf{xc}_i'\boldsymbol{\lambda}_1 + \boldsymbol{\alpha}_3'\mathbf{xi}_i + \boldsymbol{\alpha}_4'XC_i\Lambda + \frac{V_{12}}{\sigma_{s_i}^2 + \sigma_{22}}(\widehat{E}_i - \boldsymbol{\beta}_{s_i}'\mathbf{xe}_i - \gamma_{s_i} I_i) \\ XC_i\Lambda \end{bmatrix}$$

and variance matrix:

$$\mathbf{V}^c_{s_i} = \begin{pmatrix} v^c_{11s_i} & \boldsymbol{\alpha}'_4 V_c + \alpha_2 \sigma^2_{s_i} \mathrm{v}^{1\cdot}_c \\ V_c \boldsymbol{\alpha}_4 + \alpha_2 \sigma^2_{s_i} \mathrm{v}^{\cdot 1}_c & V_c \end{pmatrix} \equiv \begin{pmatrix} \mathbf{V}^c_{s_i 11} & \mathbf{V}^c_{s_i 12} \\ \mathbf{V}^c_{s_i 21} & V_c \end{pmatrix}$$

where

$$v^c_{11s_i} = V_{11} - \frac{V^2_{12}}{\sigma_{22} + \sigma^2_{s_i}} + \boldsymbol{\alpha}'_4 V_c \boldsymbol{\alpha}_4 + \alpha^2_2 \sigma^4_{s_i} \cdot \mathrm{v}^{11}_c + 2\alpha_2 \sigma^2_{s_i} \sum_{l=1}^{4} \cdot \alpha_{4l} \cdot \mathrm{v}^{1l}_c,$$

and where $\mathbf{v}^{kl}_c$ denotes kl^{th} element of V_c, while $\mathbf{v}^{k\cdot}_c$ and $\mathbf{v}^{\cdot k}_c$ denote k^{th} row and k^{th} column of V_c, respectively. Using the results for the multivariate normal distribution the posterior conditional distribution of $\mathbf{c}^m_i$ is given by

$$\mathbf{c}^m_i | \boldsymbol{\theta}, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}^m_{-i}, \mathbf{W}, \mathbf{s}, \mathbf{data}, S_i = 1 \sim N(\bar{\mathbf{c}}_i + \mathbf{V}^{c'}_{s_i 12} \mathbf{V}^{c^{-1}}_{s_i 11} (I^*_i - \bar{I}^c_i), V_c - \mathbf{V}^{c'}_{s_i 12} \mathbf{V}^{c^{-1}}_{s_i 11} \mathbf{V}^c_{s_i 12}).$$

14. The conditional posterior density kernel of latent type propensities $\widetilde{\mathbf{W}}_i$ is given by:

$$\exp(-\widetilde{W}^2_{im}/2 - \sum_{j=1}^{m-1} (\widetilde{W}_{ij} - \mathbf{x}\mathbf{w}'_i \boldsymbol{\delta}_j)^2/2) \tag{27}$$
$$\cdot \sum_{j=1}^{m} \left(\prod_{l=1}^{m} \iota(\widetilde{W}_{il} \in (-\infty, \widetilde{W}_{ij}]) \right) \tag{28}$$
$$\cdot \quad g_w(j),$$

where

$$\begin{aligned} g_w(j) &= \left\{ \exp(-\frac{(I^*_i - \alpha_0 \mathbf{x}\mathbf{e}_i \boldsymbol{\beta}_j - \alpha_1 \sigma^2_j - \alpha_2 \sigma^2_j C_{1i} - \boldsymbol{\alpha}'_3 \mathbf{x}\mathbf{i}_i - [C_{1i}, ..., C_{4i}] \boldsymbol{\alpha}_4)^2}{2V_{11}}) \right\}^{S_i = 0} \\ &\cdot \left\{ (V_{11} - \frac{V^2_{12}}{\sigma^2_j + \sigma_{22}})^{-\frac{1}{2}} \cdot \exp\Big(- \Big[I^*_i - \alpha_0 \mathbf{x}\mathbf{e}_i \boldsymbol{\beta}_j - \alpha_1 \sigma^2_j - \alpha_2 \sigma^2_j C_{1i} - \boldsymbol{\alpha}'_3 \mathbf{x}\mathbf{i}_i \right. \\ &- \left. [C_{1i}, ..., C_{4i}] \boldsymbol{\alpha}_4 - \frac{V_{12}}{\sigma^2_j + \sigma_{22}} (\widehat{E}_i - \alpha_0 \mathbf{x}\mathbf{e}_i \boldsymbol{\beta}_j - \gamma_j I_i) \Big]^2 / (2(V_{11} - \frac{V^2_{12}}{\sigma^2_j + \sigma_{22}})) \Big) \end{aligned}$$

$$\cdot \quad (\sigma_j^2 + \sigma_{22})^{-\frac{1}{2}} \cdot \exp(-\frac{(\widehat{E}_i - \alpha_0 \mathbf{xe}_i \boldsymbol{\beta}_j - \gamma_j I_i)^2}{2(\sigma_j^2 + \sigma_{22})}) \Big\}^{S_i=1}$$

Draws from this distribution are obtained by the Metropolis within Gibbs step suggested in Geweke and Keane (2007). The candidate draw $\widetilde{\mathbf{W}}_i^*$ is obtained from the normal density with the kernel given by (27). The function (28) then determines the candidate type $j^* : \widetilde{W}_{ij^*} \geq \widetilde{W}_{il}$ for all $l = 1, ..., m$. The candidate values are then accepted as new draws $\widetilde{\mathbf{W}}_i^n$ and s_i^n with probability

$$\min\left\{\frac{g_w(j^*)}{g_w(j^{(n-1)})}, 1\right\},$$

where $j^{(n-1)}$ denotes observation's i type from the previous iteration, i.e. $j^{(n-1)=s_i^{n-1}}$.

We checked that this algorithm was correctly implemented using the joint distribution tests of Geweke (2004).

A-3. Inclusion of Exogenous Variables in the Equations of the Model

Table A-2 shows specification of the equations of the model in terms of exogenous covariates included in each equation. As discussed in section 3.1, to identify selection and moral hazard effects we use cross-equation exclusion restrictions. In particular, we assume that (i) health status variables (i.e. health factors 2-23) and survey year indicator affect insurance status only indirectly (i.e. through their effect on expenditure risk E_i^*), and (ii) SAS variables, such as education, income, risk tolerance, cognitive ability, longevity expectations and financial planning horizon, enter the insurance equation but not the expenditure risk equation, once we condition on health status variables. That is, these SAS variables may affect one's health indirectly by shifting investment in health, but once we condition on health itself, they have no direct effect on one's health expenditure risk.

The demographic characteristics (i.e. marital status, ethnicity and interactions of gender with marital status and age) are included in both the expenditure and the insurance equations (in the full model). These variables are included in the expenditure equation to capture differences in health status and tastes for medical care between different demographic groups. Similarly, these variables are included in the final specification of the insurance equation to capture heterogeneity in tastes for insurance. We do not include these variables in the baseline model because insurers cannot legally price Medigap policies based on race or marital status.

The specification of the insurance equation (I_i^*) is the same as in FKS. In particular, in addition to expenditure risk E_i^*, the benchmark model includes only insurance pricing variables (polynomial in age, gender and location of residence). The potential SAS variables (education, income, risk tolerance, cognitive ability, longevity expectations, financial planning horizon, race and marital status) are progressively added to the insurance equation in extended specifications. Hence, variables indicated by "SAS" in column 3 of Table A-2 correspond to the vector $[\mathbf{xi}_i', \mathbf{c}_i]$ (see equation (5)) in the full specification of the insurance equation, and $\mathbf{c}_i$ consists of variables indicated in the last four rows of column 3 (risktol, cogn, finpln and praliv75).

The variables marked by "Yes" in column 4 of Table A-2 correspond to the vector $\mathbf{xe}_i$ of characteristics included in the specification of the expenditure risk E_i^* (see equation (6)).

The variables marked by "Yes" in column 5 of Table A-2 correspond to the vector $\mathbf{xw}_i$ of variables affecting type propensities $\widetilde{W}$ (see equation (9)). Note, that the equations for type propensities include most of the variables included in $\mathbf{xe}_i$, with the exception of the polynomial terms in age and the interactions of age with gender and gender with marital status. We omit these variables to reduce the number of parameters, as the specification for conditional mean of expenditure is already very flexible.

Finally, the model for missing SAS variable (SAS) includes most of the exogenous variables used in the analysis to maximize predictive power. The variables marked by "Yes" in column 6 of Table A-2 correspond to the vector $\mathbf{xc}_i$ of exogenous variables included in the prediction equation (7).

Table A-2: Exogenous variables included in equations for insurance status, expenditure risk, type probabilities and the prediction model for the SAS variables.

Variable	Description	I^*	E^*	$\widetilde{W}$	SAS
1	2	3	4	5	6
Female	Indicator for female	Yes	Yes	Yes	Yes
Age	Age, years	Yes	Yes	Yes	Yes
Age^2	Age squared	Yes	Yes		Yes
Age^3	Age cubed	Yes	Yes		Yes
Married	Indicator for being married	SAS	Yes	Yes	Yes
Age*Female	Interaction of age polynomial with Female	SAS	Yes		
Married*Female	Interaction of Married and Female	SAS	Yes		
Health factor 1	Health Status Factor		Yes	Yes	Yes
Health factor 3	Health Status Factor		Yes	Yes	Yes
Health factor 7	Health Status Factor		Yes	Yes	Yes
Health factor 8	Health Status Factor		Yes	Yes	Yes
Health factor 10	Health Status Factor		Yes	Yes	Yes
Health factor 11	Health Status Factor		Yes	Yes	Yes
Health factor 17	Health Status Factor		Yes	Yes	Yes
Health factor 20	Health Status Factor		Yes	Yes	Yes
Health factor 22	Health Status Factor		Yes	Yes	Yes
Health factor 23	Health Status Factor		Yes	Yes	Yes
Black	Indicator for race black	SAS	Yes	Yes	Yes
Hispanic	Indicator for Hispanic	SAS	Yes	Yes	Yes
Survey year	Year		Yes	Yes	
hgc: ls8th	Education: less than high school	SAS			Yes
hgc: somehs	Education: some high school	SAS			Yes
hgc: hs	Education: high school	SAS			Yes
hgc: somecol	Education: some college	SAS			Yes
hgc: college	Education: college	SAS			Yes
hgc: gradschl	Education: grad. school	SAS			Yes
hgc: nr	Education non-response	SAS			Yes
inc 5k-10k	Income: $5-10 thousand	SAS			Yes
inc 10k-15k	Income: $10-15 thousand	SAS			Yes
inc 15k-20k	Income: $15-20 thousand	SAS			Yes
inc 20k-25k	Income: $20-25 thousand	SAS			Yes
inc 25k-30k	Income: $25-30 thousand	SAS			Yes
inc 30k-35k	Income: $30-35 thousand	SAS			Yes
inc 35k-40k	Income: $35-40 thousand	SAS			Yes
inc 40k-45k	Income: $40-45 thousand	SAS			Yes
inc 45k-50k	Income: $45-50 thousand	SAS			Yes
inc 50plus	Income: $50+ thousand	SAS			Yes
risktol	Risk tolerance	SAS			
cogn	Congnition factor	SAS			
finpln	Financial planning horizon	SAS			
praliv75	Subjective probability to live to be 75 or more	SAS			

* Note: All equations include indicators for census divisions. The variables labelled "SAS" are not included in the baseline specification of the insurance equation (I_i^*). They are added later as potential sources of adverse/advantageous selection. The baseline insurance equation only includes pricing variables and expenditure risk.

References

[1] Arrow, Kenneth J. 1963. Uncertainty and the Welfare Economics of Medical Care. *American Economic Review* 53:941-973.

[2] Bajari, Patrick, Han Hong, Ahmed Khwaja. 2011a. A Semiparametric Analysis of Adverse Selection and Moral Hazard in Health Insurance Contracts. Working Paper.

[3] Bajari, Patrick, Han Hong and Ahmed Khwaja and Christina Marsh. 2011b. Moral Hazard, Adverse Selection and Health Expenditures: A Semiparametric Analysis. Working Paper.

[4] Blough, David K., Carolyn W. Madden and Mark C. Hornbrook. 1999. Modeling Risk Using Generalized Linear Models. *Journal of Health Economics* 18: 153171.

[5] Buchmueller, Thomas. 2006. Price and the Health Plan Choice of Retirees. *Journal of Health Economics* 25(1): 81-101.

[6] Beeuwkes Buntin, Melinda, and Alan M. Zaslavsky. 2004. Too Much Ado About Two-Part Models and Transformations? Comparing Methods of Modeling Medicare Expenditures. *Journal of Health Economics* 23: 525542.

[7] Cardon, James H., and Igal Handel. 2001. Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey. *The RAND Journal of Economics* 32 (Autumn): 408-427.

[8] Cawley, John, and Thomas Philipson. 1999. An Empirical Examination of Information Barriers to Trade in Insurance. *American Economic Review* 89 (September): 827-46.

[9] Celeux, Gilles., Merrilee Hurn and Christian. P. Robert. 2000. Computational and Inferential Difficulties with Mixture Posterior Distributions. *Journal of the American Statistical Association* 95: 957-970.

[10] Chib, Siddhartha. 1992. Bayes inference in the Tobit censored regression model. *Journal of Econometrics* 51: 79-99.

[11] Cohen, Alma. 2005. "Asymmetric Infromation and Learning: Evidence from the Automobile Insurance Market." *Review of Economics and Statistics* 87 (June): 197-207.

[12] Cohen, Alma and Peter Siegelman. 2010. "Testing for Adverse Selection in Insurance Markets." *The journal of Risk and Insurance* 77 (1): 39-84.

[13] Chiappori, Pierre-Andre, and Bernard Salanie. 2000. Testing for Asymmetric Information in Insurance Markets. *Journal of Political Economy* 108 (February): 56-78.

[14] Chiappori, Pierre-Andre, Franck Durand and Pierre Y. Geoffard. 1998. Moral Hazard and the Demand for Physician Services: First Lessons from a French Natural Experiment *European Economic Review* 42 (3-5): 499-511.

[15] Deb, Partha, Murat Munkin, and Pravin K. Trivedi. 2006. Bayesian Analysis of the Two-Part Model with Endogeneity: Application to Health Care Expenditure. *Journal of Applied Econometrics* 21: 1081-1099.

[16] Deb, Partha, Manning, Willard and Edward Norton. 2010. Preconference Course: Modeling Health Care Costs and Counts. *Lecture Slides Prepared for the ASHE-Cornell University Conference, 2010* .

[17] Ettner, Susan. 1997. Adverse Selection and the Purchase of Medigap Insurance by the Elderly. *Journal of Health Economics* 16(5): 543-562.

[18] Fang, Hanming, Michael P. Keane, and Daniel Silverman. 2008. "Sources of Advanatageous Selection: Evidence from the Medigap Insurance Market." *Journal of Political Economy* 116, no.2: 303-349.

[19] Fang, Hanming, Lauren Nicholas and Daniel Silverman. 2010. "Cognitive Ability and Retiree Health Care Expenditure." Michigan Retirement Research Center Working Paper 230.

[20] Finkelstein, Amy and Kathleen McGarry. 2006. Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market. *American Economic Review* 96 (September): 938-958.

[21] Fruhwirth-Schnatter, Sylvia. (2001). Markov Chain Monte Carlo estimation of classical and dynamic switching and mixture models. Journal of the American Statistical Association 96:194-209.

[22] Gelman, Andrew, John. B. Carlin, Hal. S. Stern, and Donald. B. Rubin. 1995. *Bayesian Data Analysis*. London: Chapman and Hall.

[23] Geweke, John. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Berger, J.O., Bernardo, J.M., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, vol. 4. Oxford University Press, Oxford, pp. 169194.

[24] Geweke, John. 2004. Getting it Right: Joint Distribution Tests of Posterior Simulators. *Journal of the American Statistical Association* 99: 799-804.

[25] Geweke, John. 2005. *Contemporary Bayesian Econometrics and Statistics* John Wiley & Sons, Inc., Hoboken, New Jersey.

[26] Geweke, John. 2006. Interpretation and Inference in Mixture Models: Simple MCMC Works. *Computational Statistics & Data Analysis* 51: 3529-3550.

[27] Geweke, John, and Michael Keane. 2007. Smoothly Mixing Regressions. *Journal of Econometrics* 138: 257-290.

[28] Gilleskie, Donna B., and Thomas A. Mroz. 2004. A Flexible Approach for Estimating the Effects of Covariates on Health Expenditures. *Journal of Health Economics* 23: 391418.

[29] Goldberger, Arthur S. 1991. *A Course in Econometrics.* Harvard University Press; Cambridge, Massachusetts; London, England.

[30] Gruber, Jonathan and Ebonya Washingtion. 2005. Subsidies to Employee Health Insurance Premiums and the Health Insurance Market. *Journal of Health Economics* 24(2): 253-176.

[31] Gulila Zvi, Robert. E. McCulloch, and Peter E. Rossi. 2006. "A Direct Approach to Data Fusion." *Journal of Marketing Research* 43: 73-83.

[32] Hurd, Michael D., and Kathleen McGarry. 1997. Medical Insurance and the Use of Health Care Services by the Elderly. *Journal of Health Economics* 16(2): 129154.

[33] Kaiser Family Foundation. 2005. Medicare Chart Book. Washington, DC: Kaiser Family Found.

[34] Khwaja, Ahmed. 2001. Health Insurance, Habits and Health Outcomes: A Dynamic Stochastic Model of Investment in Health. PhD dissertation, University of Minnesota.

[35] Kimball, Miles S., Sahm, Claudia R., and Matthew D. Shapiro. 2008. Imputing Risk Tolerance From Survey Responses. *Journal of American Statistical Association* 103(483): 1028-1038.

[36] Keane, Michael, and Olena Stavrunova. 2010. A Smooth Mixture of Tobits Model for Health Care Expenditure. 2011. *Health Economics* 20 (9): 1126-1153.

[37] Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, and Arleen Leibowitz. 1987. Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment. *The American Economic Review* 77: 251-277.

[38] Manning, Willard G. 1998. The Logged Dependent Variables, Heteroscedasticity, and the Retransformation Problem. *Journal of Health Economics* 17: 283-295.

[39] Manning, Willard G., and John Mullahy. 2001. Estimating log models: to transform or not to transform. *Journal of Health Economics* 20(4): 461-494.

[40] Manning, Willard G., Anirban Basu, and John Mullahy. 2005. Generalized modeling approaches to risk adjustment of skewed outcome data. *Journal of Health Economics* 24: 465-488.

[41] Mullahy, John. 1998. Much Ado About Two: Reconsidering Transformation and the Two-Part Model in Health Econometrics. *Journal of Health Economics* 17: 247281.

[42] Munkin, Murat, and Pravin K. Trivedi. 2008. Bayesian analysis of the ordered probit model with endogenous selection. *Journal of Econometrics* 143: 334-348.

[43] Munkin, Murat, and Pravin K. Trivedi. 2010. Disentangling Incentives Effects of Insurance Coverage from Adverse Selection in the Case of Drug Expenditure: a Finite Mixture Approach. *Health Economics* 19: 1093-1108.

[44] Pauly, Mark V. 1968. The Economics of Moral Hazard: Comment. *American Economic Review* 58: 531-537.

[45] Roberts, G.O., Gelman A., and W.R. Gilks. 1997. Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms. *The Annals of Applied Probability* 7: 110-120.

[46] Rothschild, Michael, and Joseph E. Stiglitz. 1976. Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *Quarterly Journal of Economics* 90 (November): 629-49.

[47] Rubin, Donald B. 1986. Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations. *Journal of Business and Economic Statistics* 4(1): 87-94.

[48] Wolfe, John R., and John H. Goddeeris. 1991. Adverse Selection, Moral Hazard, and Wealth Effects in the Medigap Insurance Market. *Journal of Health Economics* 10: 433-459.

[49] Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data.* The MIT Press. Cambridge, Massachusetts.