HEDG Health, Econometrics and Data Group
THE UNIVERSITY of York

HEDG Working Paper 10/14

Adverse Selection, Moral Hazard and the Demand for Medigap Insurance

Michael Keane
Olena Stavrunova

July 2010

york.ac.uk/res/herc/hedgwp

Adverse Selection, Moral Hazard and the Demand for Medigap Insurance

Michael Keane and Olena Stavrunova
University of Technology, Sydney

July 2010

Abstract

This paper studies selection and moral hazard in the US Medigap health insurance market. It develops 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 expenditure in uninsured state, conditional on other variables. The model allows for correlation between unobserved determinants of health care expenditure and the demand for insurance. To capture the complex shape of the distribution of the expenditure, a smooth mixture of Tobit models is employed (a generalization of the smoothly mixing regressions framework of Geweke and Keane (2007)). The model is estimated using a MCMC algorithm with data augmentation. The results suggest that conditionally on income, education, risk attitudes, cognitive ability, financial planning horizon and longevity expectations there is an adverse selection into Medigap insurance, but the effect is small: a one standard deviation increase in the expected expenditure in uninsured state increases the probability of buying insurance by 0.01. The model allows estimation of the sample distribution of the moral hazard effect of Medigap on health care expenditure. On average, an insured individual spends about $1,600 more on health care than her uninsured counterpart and the size of this effect is lower for healthier individuals as well as for blacks and hispanics. The smooth mixture of Tobits model developed in this paper predicts the conditional expectation of health care expenditure in our data better than the five standard models selected for comparison, and also correctly captures the complex shape of the expenditure distribution.

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, a social insurance program providing health insurance coverage to senior citizens in the US. One of the advantages of the Medigap market for studying adverse selection (a propensity of individuals with higher unobservable risk to purchase more coverage) is that in this market it is relatively easy to identify what part of information about relevant health care expenditure risk is private to individuals. Because insurers can only price Medigap policies based on age, gender, state of residence and smoking status only, 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: because adverse selection drives up the premiums, low-risk individuals remain underinsured. This result suggests that there can be a large scope for government intervention in insurance markets (e.g. mandatory social insurance financed by taxation). 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 in insurance markets 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, by lowering their 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)). It can be challenging to isolate these two effects empirically, which could be the reason why the existing studies of moral hazard and selection into the Medigap insurance do not always agree on the magnitude of these effects. For example, Wolfe and Goddeeris (1991) find evidence of adverse selection and moral hazard in their 1977-1979 sample of the Retirement History Survey respondents. Ettner (1997) also finds both adverse selection and moral hazard in her sample from the 1991 Medicare Current Beneficiary Survey (MCBS). In both studies the selection effects are found to be rather small. On the other hand, Hurd and McGarry (1997) find that the higher health care use by insured in their 1993-1994 Asset and Health Dynamic Survey sample is attributed to incentives (moral hazard) effect rather than to the selection effect. Recently Fang, Keane and Silverman (2008) (FKS) document advantageous selection into the Medigap insurance using the 2000 and 2001 waves of the MCBS and the 2002 wave of the Health and Retirement Survey (HRS). They also show that the advantageous selection could be explained by a number of individual characteristics, including risk preferences, cognition and financial planning horizon, which are correlated with both health expenditure risk and insurance demand. While providing a thorough analysis of selection, FKS did not attempt to estimate the incentives (moral hazard) effect of Medigap insurance on health care expenditure. In this paper we extend the analysis of FKS and attempt to evaluate the extent of both moral hazard and selection in the Medigap health insurance market.

FKS start by showing that seniors who purchase Medigap insurance are in better health than the uninsured, that is, there is advantageous selection in this market (individuals with higher expenditure risk are less likely to purchase insurance). Then they propose to investigate the relationship between Medigap status and ex-post health care expenditures conditionally on pricing variables only, and then with controls for potential sources of advantageous selection (SAS). These are variables which (i) can potentially have an independent effect on insurance demand and (2) are correlated with expenditure risk, but (iii) cannot be used by insurers for pricing policies. Potential sources of advantageous selection proposed by FKS include income, education, risk tolerance, variance of health care expenditure, interaction of risk tolerance and the variance of expenditure, financial planning horizon, longevity expectations and cognitive ability.

To carry out such analysis one would ideally need a dataset which simultaneously contains information on health expenditure, insurance status and SAS 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 as well as the 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, and then to use the estimated relationship to predict expected health care expenditure in uninsured state for HRS respondents as a measure of health expenditure risk. FKS then investigate how the relationship between the Medigap insurance status and the expected expenditure changes as various potential sources of advantageous selection are added to the model. They find that as more SAS variables are added to the insurance demand model, the relationship between Medigap status and expected expenditure turns from negative to positive, which suggests that among individuals who are similar in terms of SAS it is less healthy individuals 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 turned out to be not very important - it affected demand but was not correlated with expenditure risk.

The limitation of FKS's analysis is that they did not account for possibly non-random (conditional on observables) selection into insurance when estimating the prediction model for expected expenditure. To obtain the prediction equation for health expenditure in the uninsured state they estimate the following model by OLS in 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 I_i is an indicator for Medigap coverage. Then for HRS respondents they predict expected expenditure in the uninsured state as follows:

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

They use the prediction to estimate the model for health insurance status in HRS as:

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

where $\mathbf{P}_i$ is a vector of variables that affect the price of Medigap insurance. If ε_i is correlated with insurance indicator I_i, then $\widehat{E}_i$ is an inconsistent estimate of the expected health care expenditure risk in uninsured state. For example, if I_i and ε_i are negatively correlated (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 risk for individuals who actually have health insurance. This will cause FKS to overstate the degree of advantageous selection (α_0 in model (2)), thus being over-optimistic about the ability of the proposed variables to explain the advantageous selection in the Medigap market.

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 empirical strategy is the following. We assume that the risk that is relevant when an individual is considering buying a Medigap policy is her expected total health care expenditure in uninsured state, conditional on demographic and health status characteristics. The demand for insurance then depends on expected expenditure, insurance pricing variables and demographic and behavioral characteristics such as income, education, cognitive ability, risk aversion and other factors, which reflect costs of acquiring and tastes for insurance. The degree of selection is measured by the sensitivity of the insurance demand to the expected expenditure, conditional on other variables.

Unlike FKS, we allow for correlation between unobserved determinants of expected expenditure and the demand for insurance. We estimate the distribution of the expected expenditure taking into account that an insured individual can have higher realized expenditure than her uninsured counterpart because of the moral hazard (or price) effect of insurance. In particular, the realized expenditure is modelled as a sum of the expected expenditure, the moral hazard effect (if the individual is insured) and an additive random term which reflects individual's forecast error.

To capture the complex shape of the distribution of the realized expenditure, which is positive and extremely skewed to the right, we employ a smooth mixture of Tobit models (generalizing the smoothly mixing regressions (SMR) framework of Geweke and Keane (2007)). Hence, our model for insurance demand and health care expenditure is a simultaneous equations model where the parameters of interest (the selection and moral hazard effects) are identified via cross-equations restrictions. The key restrictions are (i) that the health status variables affect demand for insurance only through their effect on expected expenditure (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 expected expenditure.

To combine information from the MCBS and HRS we specify an auxiliary model for variables missing from the MSBC, conditional on the variables common in the two datasets. In the estimation we merge the two datasets, assume that the relevant variables are missing from HRS and MCBS completely at random, and estimate the model using a MCMC algorithm with multiple imputations of the missing variables.

Our findings 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. We find that conditionally on income, education, risk attitudes, cognitive ability, financial planning horizon and longevity expectations there is adverse selection into Medigap insurance but the size of the effect is small: a one standard deviation increase in expected expenditure in uninsured state increases probability of buying insurance by 0.01.

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 insured individual spends about $1,600 more on health care than her uninsured counterpart, and that this effect varies with individual characteristics. In particular, the moral hazard effect of Medigap is lower for healthier individuals as well as for blacks and hispanics.

Another contribution of this paper is a new econometric model for the conditional distribution of health expenditures, which for our expenditure data was found to be superior to the five standard models we used for comparison, including OLS, two-part lognormal model and several specifications of GLM. Our smooth mixture of Tobits model predicts the conditional expectation of health care expenditure better than these models, and also correctly captures the complex shape of the expenditure distribution as well as conditional variance and skewness - tasks which many existing models for expenditure are not suited to do.

This paper is organized as follows. Section (2) describes the datasets used in the analysis, section (3) discussed the choice of an econometric model for the health care expenditure data, section (4) presents a model of the demand for Medigap insurance and total health care expenditure and discusses MCMC algorithm developed for the Bayesian inference in this model, section (5) discusses the empirical results, and section (6) concludes.

2 Data: HRS and MCBS

While Medicare is a primary health insurance program for most seniors in the USA, on average it only covers about 45% of personal health care expenditure of beneficiaries. Medicare consists of two plans, plan A provides hospital insurance coverage, while plan B provides insurance coverage for some physician services, outpatient services, home health services and durable medical equipment. Most of the beneficiaries are enrolled in both plans A and B. To cover gaps in Medicare, private companies offer Medigap insurance plans - private policies to cover some of the co-pays and deductibles associated with Medicare, as well as uncovered expenses. There are 10 standardized Medigap plans, an open enrollment period within which insurers cannot deny coverage based on health status to qualified individuals, and pricing rules according to which insurers can price policies based on only age, gender, smoking status and state of residence. The institutional details of the Medigap market can be found in FKS. The Medigap insurance status in our analysis is defined as equal to one if an individual purchases additional private policy secondary to Medicare.

The analysis uses data from two datasets, the Medicare Current Beneficiary Survey (MCBS, 2000 and 2001 waves) and the Health and Retirement Study (HRS, 2002 wave). The MSBC contains comprehensive information about respondents' health care costs and use as well as detailed information about their health and demographic and socioeconomic characteristics. The HRS contains detailed information about health, demographics and socioeconomic characteristics as well as measures of risk attitudes, financial planning horizon, longevity expectations and cognitive ability - some of the potential sources of advantageous selection. The data used in the analysis includes only individuals covered by basic Medicare. Descriptive statistic of the selected variables are presented in Table 2. We use the same MCBS sample as FKS, and the HRS sub-sample used by FKS to obtain column (3) of Table 6 in their paper[1]. 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.

The variables which measure cognitive ability in FKS include Telephone Interview for Cognitive Status score, word recall ability score, numeracy score and subtraction score. To decrease the number of auxiliary variables to model we extract a common factor from these variable and use it as a measure of cognitive ability in our analysis. We also use factor analysis do reduce the size of the vector of health status characteristics. Both datasets contain 76 health status characteristics 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 (ISDLs). To reduce the number of parameters related to health status characteristics, we first factor-analyze these variables to extract 38 factors using data in both HRS (full sample) and MCBS samples. We than regress the health care expenditure in the MCBS on demographic characteristics and the 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 so that the chosen 10 factors produce the highest regression R-squared among all possible 10 factors subsets of the 16 factors.

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

[1]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 or 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.

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 Factors 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)
Fh2			4565.0***
			(252.4)
Fh3			-2544.6***
			(226.4)
Fh7			2049.0***
			(241.5)
Fh8			711.7***
			(213.1)
Fh10			-2047.0***
			(535.5)
Fh11			-961.6***
			(207.8)
Fh17			1176.3
			(931.4)
Fh20			-1339.2***
			(363.7)
Fh22			2144.6***
			(382.4)
Fh23			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

Table 2: Descriptive Statistics

	MCBS			HRS		
Variable	All	Medigap	No Medigap	All	Medigap	No Medigap
Medigap	0.50	1.00	0	0.43	1.00	0
	(0.50)	(0)	(0)	(0.50)	(0)	(0)
Female	0.59	0.60	0.58	0.56	0.58	0.55
	(0.49)	(0.49)	(0.49)	(0.50)	(0.49)	(0.50)
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
	(0.30)	(0.19)	(0.38)	(0.35)	(0.24)	(0.40)
Hispanic	0.08	0.03	0.12	0.07	0.02	0.11
	(0.27)	(0.17)	(0.33)	(0.26)	(0.13)	(0.32)
Married	0.48	0.54	0.43	0.66	0.71	0.63
	(0.50)	(0.50)	(0.50)	(0.47)	(0.46)	(0.48)
Education: Less than high school	0.36	0.27	0.45	0.28	0.22	0.33
	(0.48)	(0.44)	(0.50)	(0.45)	(0.42)	(0.47)
Education: High School	0.27	0.31	0.24	0.38	0.41	0.35
	(0.45)	(0.46)	(0.43)	(0.49)	(0.49)	(0.48)
Education: Some college	0.21	0.24	0.18	0.18	0.18	0.17
	(0.41)	(0.43)	(0.38)	(0.38)	(0.39)	(0.38)
Education: College	0.08	0.10	0.06	0.08	0.08	0.08
	(0.27)	(0.30)	(0.23)	(0.27)	(0.27)	(0.27)
Health factor Fh2	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 Fh3	-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				0.00	0.204	-0.157
				(0.78)	(0.61)	(0.84)
Risk tolerance				0.234	0.228	0.236
(estimate from Kimball et al. (2008))				(0.142)	(0.138)	(0.146)
Fin. 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

presented in Table 1. After 76 health status characteristics are replaced by ten health factors, the regression R-squared drops from 0.21 to 0.18, which is a reasonable price for reducing the number of covariates substantially. Factors 2 and 3 turn out to be the most quantitatively important for predicting expenditure. Factor 2 (Fh2) heavily loads on deterioration in health as well as difficulties and help with IADLs, and so is an unhealthy factor. It increases expenditures by about $4500 per one standard deviation. Factor 3 (Fh3) 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. Other factors are more difficult to interpret, but they help to improve the regression R-squared from 0.13 (in a regression of expenditure on Medigap coverage, demographic characteristics and Factors 2 and 3 alone) to 0.18.

Table 2 shows the descriptive statistics of the HRS and the MCBS sub-samples. It can be seen from the table that the individuals in the HRS subsample are healthier and younger than 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. Using the HRS data we estimate the distribution of these variables conditional on exogenous characteristics common in two datasets and use it to impute missing behavioral SAS variables in the MCBS sub-sample. Provided that this conditional distribution is the same in different age and health groups of the population, the fact that the two subsamples are different does not create a problem for our analysis.

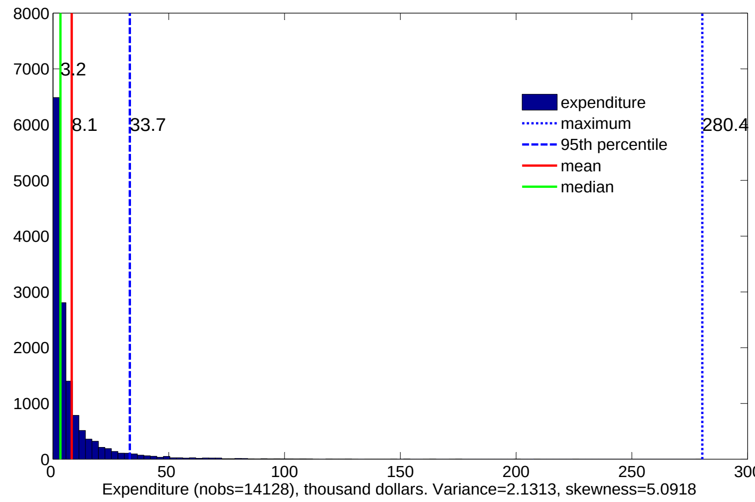
Tables 1 and 2 suggest the presence of both advantageous selection and moral hazard. Table 2 shows that insured individuals are on average healthier than the uninsured in both the HRS and the MCBS data, while Table 1 shows that the insured individuals spend more on health care than the uninsured. We will investigate the magnitudes of the selection and incentives effects in the subsequent sections.

3 Modelling Distribution of Health Care Expenditure

It is well-known in health economics that econometric modelling of health care expenditures can be challenging because of the properties of their empirical distribution. In particular, the health care expenditures are usually non-negative, highly skewed to the right and can have a point mass at zero. The empirical distribution of the total health care expenditure of Medicare beneficiaries in our MCBS sample exhibits all these properties, as can be seen on Figure 1. This figure plots the histogram of the expenditure and some properties of its distribution. 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 is to a large extent represented by studies which address the problem of modelling it's conditional *expectation* in the presence of skewness and a non-trivial fraction of zero outcomes (e.g. Manning (1998); Mullahy (1998); Jones (2000); Blough et al. (1999); Manning and Mullahy (2001); Buntin and Zaslavsky (2004); Manning et al. (2005)). The standard approach to modelling this conditional expectation is by way of a two-part model: the first part is a binary choice model for the indicator of whether a positive outcome is observed, and the second part is a regression model for positive outcomes. There is no single preferred specification for the second part. The specifications of the second part used in the literature include a linear regression model for the natural logarithm of positive outcomes, the generalized linear models (GLM), the exponential conditional mean models (ECM), etc. Usually

Figure 1: Histogram of total health care expenditure



it is recommended to try several specifications for the expenditure data at hand and select the one with the best fit (e.g. Buntin and Zaslavsky (2004)). The problem of modelling the conditional *distribution* of health care expenditure is less frequently addressed. When the context requires a model for the distribution of expenditure, the preferred approach is to specify a two-part model where the positive outcomes (the second part) are modelled using the lognormal distribution (e.g. Deb et al. (2006)). The log transformation is not appropriate for our purposes because we are interested in the effect of the *level* of expected expenditure on Medigap insurance status. Thus, we have to model the level of expenditure rather than it's natural logarithm. To select the appropriate model for expenditure we investigate several competing models in isolation and choose the one with the highest fit to the MCBS total medical expenditure data to be embedded into a comprehensive model of the joint determination of the demand for health insurance and health care expenditure. This section will discuss the choice of model for the expenditure distribution.

Because in our data the proportion of zero outcomes is small (0.025), we do not use a two-part structure. Instead, to capture the fact that the expenditure is nonnegative and there is some mass at zero, we adopt the Tobit specification. To capture skewness, we specify that the expenditure follows a discrete mixture of Tobits. We start with the specification where the probability of a mixture component is independent of exogenous characteristics and then proceed to the specification with covariate-dependent component probability. In particular, we specify the following model for expenditure E_i in the first case:

$$\begin{aligned} \widehat{E}_i|(\text{type}_i = j, \mathbf{xe}_i) &= \alpha_{0j} + \mathbf{xe}_i\boldsymbol{\alpha}_1 + \varepsilon_i|(\text{type}_i = j, \mathbf{xe}_i) \\ \varepsilon_i|(\text{type}_i = j, \mathbf{xe}_i) &\sim N(0, \exp(\gamma_{0j} + \mathbf{xv}_i\boldsymbol{\gamma}_1) \end{aligned}$$

$$
\begin{aligned}
Pr(\text{type=j}|(\mathbf{xe}_i) &= \pi_j, \ \sum_{j=1}^{m} \pi_j = 1 \\
E_i &= \max\{0, \widehat{E}_i\},
\end{aligned}
$$

and in the second case:

$$
\begin{aligned}
\widehat{E}_i|(\text{type}_i = j, \mathbf{xe}_i) &= \beta_{0j} + \mathbf{xe}_i\boldsymbol{\beta}_{1j} + \varepsilon_i|(\text{type}_i = j, \mathbf{xe}_i) &(3)\\
\varepsilon_i|(\text{type}_i = j, \mathbf{xe}_i) &\sim N(0, \sigma_j^2) &(4)\\
Pr(\text{type}_i = j|\mathbf{xe}_i) &= \int_{-\infty}^{\infty} \phi(d - \boldsymbol{\delta}_j'\mathbf{xw}_i) \prod_{l \neq j}^{m} \Phi(d - \boldsymbol{\delta}_l'\mathbf{xw}_i) dd &(5)\\
E_i &= \max\{0, \widehat{E}_i\}. &(6)
\end{aligned}
$$

The model in equations (3)-(6) is an extension of the smoothly mixing regressions framework of Geweke and Keane (2007) to the case of a Tobit-type limited dependent variable, so we'll call it SMTobit for Smooth Mixture of Tobits. The dependence of mixture component probabilities on covariates is achieved by assuming that the distribution of latent mixture component indicators is governed by a multinomial probit model, conditional on covariates. The expression (5) is the probability of a mixture component corresponding to this multinomial probit model. The vectors $\mathbf{xe}_i$, $\mathbf{xv}_i$ and $\mathbf{xw}_i$ contain socio-demographic characteristics (age, gender, marital status, location of residence) and ten health status factors. In the first model (we'll call it MixHetTobit) we explicitly allow the conditional variance of $\widehat{E}_i$ to depend on covariates which affect the mean of $\widehat{E}_i$, because in our data the variance of expenditure tends to increase as health status deteriorates. In the second model the conditional heteroscedasticity is allowed by way of a covariate-dependent type probability, so that unconditionally of type j the variance of ε_i is given by the following non-linear function of $\mathbf{xw}_i$:

$$
Var(\varepsilon_i|\mathbf{xw}_i) = \sum_{j=1}^{m} \sigma_j^2 \cdot \int_{-\infty}^{\infty} \phi(d - \boldsymbol{\delta}_j'\mathbf{xw}_i) \prod_{l \neq j}^{m} \Phi(d - \boldsymbol{\delta}_l'\mathbf{xw}_i) dd.
$$

We have estimated both models with up to 7 mixture components as well as the standard Tobit and the heteroscedastic Tobit by the Bayesian method using MCMC algorithms based on Chib (1992), Geweke (2005) chapter 6, and Geweke and Keane (2007). We compared the fit of the models using the modified cross-validated log scoring rule suggested in Geweke and Keane (2007). We found that the mixture models provided a better fit to the expenditure data than the standard and the heteroscedastic Tobit, and that the SMTobit with 6 components had the best fit. In general, the SMTobit outperformed the MixHetTobit with the same number of mixture components. While both models with 6 mixture components provided a good fit to the overall expenditure distribution, the SMTobit was much better than the MixHetTobit in fitting the conditional mean of the expenditure. The reason for the unsatisfactory performance of the MixHetTobit is that it attributes non-linear effects of covariates on expenditure to random variation due to belonging to different components of the mixture. Thus, unhealthy individuals with high expenditure were treated by the model as if they had a random draw from the high mean component of the mixture, and vice versa. As a results the model over-predicted expenditure for low expenditure individuals, and under-predicted it for high-expenditure individuals. Because the MixHetTobit is computationally simpler than the SMTobit and for this reason, ceteris paribus, would be preferred to SMTobit, we tried to improve the conditional mean predictions of MixHetTobit by introducing polynomials in the most important predictors of expenditure and by allowing the impact of covariates to be different for healthy and unhealthy individuals, without much success. In contrast, the rich functional form of the conditional mean in SMTobit allows the model to produce a good prediction of the conditional mean without the need to expand the vector of covariates. Hence, our preferred model for expenditure is SMTobit with six mixture components. The next section will present the joint model of the demand for insurance and health care expenditure, in which the expenditure is modelled as the SMTobit with six components.

4 The Model

This section will present the model for the joint determination of the Medigap insurance status and health care expenditure. We present the full specification of the model, in which the insurance equation includes all potential sources of advantageous selection, including risk tolerance, variance of health care expenditure and interaction of the variance and risk tolerance. In addition to estimating the full model, we also estimate several restricted versions of this model. We start with the benchmark specification in which the demand for insurance depends only on expenditure risk and pricing variables, and then progressively include other potential SAS variables into the insurance equation. Our goal is to estimate how the effect of the expected health care expenditure on the demand for insurance changes as we add the SAS variables. In this section we will first present the model abstracting from the fact that not all variables of interest are available in both dataset, and later discuss the imputation of variables missing from the HRS or the MCBS.

4.1 Complete data

We assume that there are m types of individuals (the types are indexed by j, $j = 1, \ldots, m$). The type is a private information of an individual, i.e. individuals know their own type, but from the point of view of a researcher these types are latent. The types differ by the effects of individual characteristics and insurance status on health care expenditure as well as by variance of the expenditure. Let I_i^* denote utility that individual i derives from health insurance and let E_i^* denote her expected health care expenditure if she remains uninsured. We treat E_i^* as the expenditure risk relevant when an individual decides to purchase a Medigap insurance plan. Both I_i^* and E_i^* are known to the individual but are unobserved by an econometrician, so they enter the model as latent variables. Let σ_j^2 denote the variance of actual expenditure around the expected expenditure, conditional on insurance status of individual of type j, so σ_j^2 can be interpreted as variance of health care expenditure forecast error. The model for the latent vector $[I_i^*, E_i^*]'$ conditional on type j is specified as follows:

$$I_i^*|j = \alpha_0 E_i^*|j + \alpha_1 \sigma_j^2 + \alpha_2 \sigma_j^2 \cdot c_{1i} + \boldsymbol{\alpha}_3' \mathbf{xi}_i + \boldsymbol{\alpha}_4' \mathbf{c}_i + \varepsilon_{1i} \quad (7)$$

$$E_i^*|j = \boldsymbol{\beta}_j' \mathbf{xe}_i + \varepsilon_{2i}, \quad (8)$$

where $\mathbf{c}_i$ includes variables present in 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 and education (all common in the two datasets), and $\mathbf{xe}_i$ includes demographic characteristics (age, gender, location of residence, marital status, race and ethnicity) and ten health factors, discussed in section (2). All these variables are also common in the two datasets.

The expected health care expenditure consists of the part which depends on the observable health status and demographics ($\boldsymbol{\beta}_j'\mathbf{xe}_i$) and the part which depends on unobservable characteristics (ε_{2i}). There is a heterogeneity in the effect of observable health status and demographic characteristics on the expected expenditure in the form of variation of $\boldsymbol{\beta}_j$ across the types of individuals. The parameter α_0 measures the effect of expenditure risk on insurance demand and indicates advantageous selection if negative, and adverse selection if positive. We also introduce the variance of forecast error σ_j^2 and its interaction with the risk tolerance in the insurance equation to make the full model consistent with FKS, who also included these variables among the potential SAS but found that the variance and the risk tolerance were not important for explaining advantageous selection. The vector of disturbances $\boldsymbol{\varepsilon}_{12i} = [\varepsilon_{1i}, \varepsilon_{2i}]'$ follows the bivariate normal distribution:

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

The parameter σ_{12} is the covariance between unobservable determinants of expected expenditure and insurance status[2].

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 the *potential* health care expenditure of individual i. We assume that $\widehat{E}_i$ is determined as follows:

$$\widehat{E}_i|j = E_i^* + \gamma_j I_i + \eta_i|j \tag{9}$$

where γ_j denotes type-specific effect of health insurance on the potential health care expenditure (price or moral hazard effect), and $\eta_i|j$ is the forecast error of individual i. Given individual's type j the forecast error $\eta_i|j$ is normally distributed with zero mean and variance σ_j^2 and is independent of $\boldsymbol{\varepsilon}_{12i}$:

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

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

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

Hence, conditional on type j the model for the realized expenditure E_i is Tobit. This model is adopted because it captures nonnegativeness and zero point mass of the empirical distribution of expenditure.

To impute missing $\mathbf{c}_i = [c_{1i}, ..., c_{4i}]'$ into the MCBS data we need to 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:

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

[2]The insurance equation (7) can be rewritten as follows:

$$I_i^*|j = \alpha_0\boldsymbol{\beta}_j'\mathbf{xe}_i + \alpha_1\sigma_j^2 + \alpha_2\sigma_j^2 \cdot c_{1i} + \boldsymbol{\alpha}_3'\mathbf{xi}_i + \boldsymbol{\alpha}_4'\mathbf{c}_i + \rho\varepsilon_2 + \xi_i,$$

where $\rho = \alpha_0 + \frac{\sigma_{12}}{\sigma_{22}}$ and ξ_i is independent of ε_2. Hence, the selection effect can be decomposed into the selection based on observable determinants of health expenditure (measured by α_0) and on unobservable determinants (measured by ρ).

where $k = 1, \ldots, 4$, $\mathbf{xc}_i$ denotes the vector of exogenous variables common in the two datasets, such as demographics, income, health status and education; the disturbances $[\varepsilon_{31i}, \ldots, \varepsilon_{34i}]' \equiv \boldsymbol{\varepsilon}_{3i} \sim N(0, V_c)$ for all $j = 1, \ldots, m$ and are independent of $\boldsymbol{\varepsilon}_{12i}$ and $\eta_i|j$. Hence,

$$\mathbf{c}_i|j = XC_i\Lambda + \boldsymbol{\varepsilon}_{3i}, \tag{10}$$

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 (7)-(10), conditionally on type j, follow 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 individual's exogenous characteristics by way of a multinomial probit model, as in Geweke and Keane (2007):

$$\begin{aligned} \widetilde{W}_{ij} &= \boldsymbol{\delta}_j'\mathbf{xw}_i + \zeta_{ij} \qquad j = 1, \ldots, m-1 \\ \widetilde{W}_{im} &= \zeta_{im} \end{aligned} \tag{11}$$

where $\widetilde{W}_{ij}$ are latent propensities of being type j, $\mathbf{xw}_i$ is a vector of individual characteristics including demographics and health status and ζ_{ij} are independent standard normal random variables. The individual i is of type j iff $\widetilde{W}_{ji} \geq \widetilde{W}_{li}\ \forall\ l = 1, ..., m$. The probability of type j is given by:

$$P(\text{type}_i = j) = \int_{-\infty}^{\infty} \phi(d - \boldsymbol{\delta}_j'\mathbf{xw}_i) \prod_{l \neq j}^{m} \Phi(d - \boldsymbol{\delta}_l'\mathbf{xw}_i) dd, \tag{12}$$

where $\Phi(.)$ denotes standard normal cdf, $\phi(.)$ denotes standard normal pdf and $\boldsymbol{\delta}_m = \mathbf{0}$. This restriction resolves the identification issue well-known in multinomial choice models which stems from the fact that only differences in alternative-specific utilities affect the actual choice. So, when no restrictions are 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 resolve this identification issue one of the alternative-specific vectors of coefficients is often normalized to zero, as we do here.

4.2 Combining data from the MCBS and the HRS

To estimate the model in section (4.1) a dataset containing information on $I_i, E_i, \mathbf{c}_i$ and exogenous health status and demographic characteristics (denoted by $\mathbf{x}_i$) for all observations is required. Such dataset is not available, and instead the following two datasets are available: the MCBS, which has information on I_i, E_i and 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 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 parameters $\boldsymbol{\theta}$ is the same in the MCBS and HRS datasets and is as specified in section (4.1), and 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. (1994)).

Let $\mathbf{C}^o$ denote the collection of $\mathbf{c}_i$'s subsequently observed, and $\mathbf{C}^m$ denote the collection of $\mathbf{c}_i$'s subsequently missing. Similarly, let $\mathbf{E}^o$ denote the collection of E_i's subsequently observed, and $\mathbf{E}^m$ denote the collection of E_i's subsequently 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 $I_i, E_i, \mathbf{c}_i$ and hence its modelling can be omitted. 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.

Let S_i denote a survey indicator so that $S_i = 1$ if $i \in$ MCBS and $S_i = 0$ if $i \in$ HRS. Let N^M and N^H denote number of observations in MCBS and HRS respectively, and $N = N^M + N^H$ denote the number of observations in the combined dataset. The probability density function of observables subsequently observed conditional on exogenous variables $\mathbf{X}$, survey indicators $\mathbf{S} \equiv [S_1, ..., S_N]$ and parameters $\boldsymbol{\theta}$ can be expressed as follows:

$$
\begin{aligned}
p(\mathbf{E}^o, \mathbf{I}, \mathbf{C}^o | \mathbf{S}, \mathbf{X}, \boldsymbol{\theta}) &= \prod_{i=1}^{N} \left[\int p(E_i^o, I_i, \mathbf{c}^m | \mathbf{X}, \boldsymbol{\theta}) d\mathbf{c}_i^m \right]^{S_i} \cdot \left[\int p(E_i^m, I_i, \mathbf{c}^o | \mathbf{X}, \boldsymbol{\theta}) dE_i^m \right]^{1-S_i} \\
= \quad & \prod_{i=1}^{N} \sum_{j=1}^{m} \int_{-\infty}^{\infty} \phi(d - \boldsymbol{\delta}_j \mathbf{x}\mathbf{w}_i) \prod_{l \neq j}^{m} \Phi(d - \boldsymbol{\delta}_l \mathbf{x}\mathbf{w}_i) dd \\
\cdot \quad & \left\{ \left\langle \Phi \left(\frac{\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} + \frac{\sigma_{12} + \alpha_0 \sigma_{22}}{\sigma_{22} + \sigma_j^2} (E_i - \boldsymbol{\beta}_j' \mathbf{xe}_i - \gamma_j I_i)}{\sqrt{\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}_x^{11} + 2\alpha_2 \sigma_j^2 \sum_{l=1}^4 \cdot \alpha_{4l} \cdot \mathrm{v}_x^{1l} - \frac{(\sigma_{12} + \alpha_0 \sigma_{22})^2}{\sigma_{22} + \sigma_j^2}}} \right)^{I_i} \right. \right. \\
\cdot \quad & \left. (1 - \Phi(\cdot))^{1-I_i} \cdot \frac{1}{\sqrt{2\pi(\sigma_{22} + \sigma_j^2)}} \exp \left(- \frac{(E_i - \boldsymbol{\beta}_j' \mathbf{xe}_i - \gamma_j I_i)^2}{2(\sigma_{22} + \sigma_j^2)} \right) \right\rangle^{E_i > 0} \\
\cdot \quad & \left. \left\langle \int_{-\infty}^{0} \Phi(\cdot)^{I_i} (1 - \Phi(\cdot))^{1-I_i} \cdot \frac{1}{\sqrt{2\pi(\sigma_{22} + \sigma_j^2)}} \exp \left(- \frac{(\widehat{E}_i - \boldsymbol{\beta}_j' \mathbf{xe}_i - \gamma_j I_i)^2}{2(\sigma_{22} + \sigma_j^2)} \right) d\widehat{E}_i \right\rangle^{E_i = 0} \right\}^{S_i} \\
\cdot \quad & \left\{ \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} (1 - \Phi(\cdot))^{1-I_i} \right. \\
\cdot \quad & \left. (2\pi)^{-\frac{K_c}{2}} |V_c|^{-\frac{1}{2}} \exp(-(\mathbf{c}_i - X C_i \boldsymbol{\Lambda})' V_c^{-1} \cdot (\mathbf{c}_i - X C_i \boldsymbol{\Lambda})/2) \right\}^{1-S_i},
\end{aligned}
\tag{13}
$$

where $\boldsymbol{\delta}_m = \mathbf{0}$ and the notation $\Phi(.)$ means that the parenthesis contain the expression which was inside the parenthesis at the previous occurrence of $\Phi()$.

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 likelihood will not change. The identification in such cases is usually achieved by the normalization $\sigma_{11} = 1$. For the purposes of posterior simulation it would be 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

$$\sigma_{11} = 1 + \frac{\sigma_{12}^2}{\sigma_{22}}.$$

4.3 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 $\mathbf{I}^* = [I_1^*, ..., I_N^*]'$ and $\mathbf{E}^* = [E_1^*, ..., E_N^*]'$; the MCBS data are augmented by the missing values $\mathbf{c}_i^m$, $i = 1, ..., N_M$ and by potential expenditure $\widehat{\mathbf{E}} = [\widehat{E}_1, ..., \widehat{E}_{NM}]'$. The latter is a standard approach to Bayesian inference in the Tobit model (due to Chib (1992)). It can be shown that the fact that E_i is missing from the HRS subsample does not create difficulties for the Bayesian inference about parameters of the model via Gibbs sampler - the probability density of observable and unobservable data retain their conjugacy to normal and gamma priors of parameters governing the distributions of I_i^*, E_i, E_i^* and $\mathbf{c}_i$ after E_i is integrated out of the HRS subsample analytically. On the other hand, the fact the $\mathbf{c}_i$ is missing from the MCBS subsample complicates the inference - the probability density of observable and unobservable data lose their conjugacy to normal and Wishart priors of the parameters governing the joint distribution of $\mathbf{c}_i$ when $\mathbf{c}_i$ is integrated out of the MCBS subsample. For this reason in the estimation we perform multiple imputations of $\mathbf{C}_i^m$, but not of E_i, which is integrated out analytically from the HRS subsample. Both HRS and MCBS are also augmented by the latent type indicators $\mathbf{s} = [s_1, ..., s_N]'$, so that $s_i = j$ if i's type is j, as well as by the latent type propensities $\mathbf{W} = [\widetilde{\mathbf{W}}_1', ..., \widetilde{\mathbf{W}}_N']$, where $\widetilde{\mathbf{W}}_i = [\widetilde{W}_{1i}, ..., \widetilde{W}_{mi}]'$.

Let $\mathbf{I} = [I_1, ..., I_N]'$ and $\mathbf{E} = [E_1, ..., E_N]'$. Then the augmented data density conditional on $\mathbf{X}, \mathbf{S}$ and $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \alpha_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_m, \sigma_1^2, ...\sigma_m^2, \gamma_1, ..., \gamma_m, \boldsymbol{\delta}_1, ..., \boldsymbol{\delta}_{m-1}, \sigma_{12}, \sigma_{22}, V_c, \Lambda)$ can be written as follows:

$$
\begin{aligned}
& p(\mathbf{I}^*, \mathbf{E}^*, \mathbf{I}, \widehat{\mathbf{E}}, \mathbf{E}, \mathbf{C}^o, \mathbf{C}^m, \mathbf{s}, \mathbf{W} | \mathbf{S}, \mathbf{X}, \boldsymbol{\theta}) \\
= & \prod_i^N [p(I_i^* | E_i^*, \mathbf{c}_i^m, \mathbf{x}_i, s_i = j, \boldsymbol{\theta}) \cdot p(E_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}) \\
\cdot & (I_i | I_i^*, E_i^*, \mathbf{c}_i^m, \mathbf{x}_i, s_i = j, \boldsymbol{\theta}) \cdot p(\widehat{E}_i | I_i^*, E_i^*, \mathbf{c}_i^m, I_i, \mathbf{x}_i, s_i = j, \boldsymbol{\theta}) \cdot p(E | \widehat{E}_i, I_i^*, E_i^*, \mathbf{c}_i^m, I_i, \mathbf{x}_i, s_i = j, \boldsymbol{\theta}) \Big]^{S_i} \\
\cdot & [p(I_i^* | E_i^*, \mathbf{c}_i^o, \mathbf{x}_i, s_i = j, \boldsymbol{\theta}) \cdot p(E_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}) \\
\cdot & p(I_i | I_i^*, E_i^*, \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}) \\
& = \prod_i [\frac{1}{\sqrt{2\pi}} \exp(-(I_i^* - \alpha_0 E_i^* - \alpha_1 \sigma_{s_i}^2 - \alpha_2 \sigma_{s_i}^2 c_{1i}^m - \boldsymbol{\alpha}_3' \mathbf{x} \mathbf{i}_i - \boldsymbol{\alpha}_4' \mathbf{c}_i^m - \frac{\sigma_{12}}{\sigma_{22}} (E_i^* - \boldsymbol{\beta}_{s_i}' \mathbf{x} \mathbf{e}_i))^2 / 2) \\
\cdot & \frac{1}{\sqrt{2\pi\sigma_{22}}} \exp(-\frac{(E_i^* - \boldsymbol{\beta}_{s_i}' \mathbf{x} \mathbf{e}_i)^2}{2\sigma_{22}}) \cdot (2\pi)^{-4/2} |V_c|^{-1/2} \exp(-(\mathbf{c}_i^m - X C_i \Lambda)' V_c^{-1} (\mathbf{c}_i^m - X C_i \Lambda)/2) \\
\cdot & \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp(-\frac{(\widehat{E}_i - E_i^* - \gamma_{s_i} I_i)^2}{2\sigma_j^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 & [\frac{1}{\sqrt{2\pi}} \exp(-(I_i^* - \alpha_0 E_i^* - \alpha_1 \sigma_{s_i}^2 - \alpha_2 \sigma_{s_i}^2 c_{1i}^o - \boldsymbol{\alpha}_3' \mathbf{x} \mathbf{i}_i - \boldsymbol{\alpha}_4' \mathbf{c}_i^o - \frac{\sigma_{12}}{\sigma_{22}} (E_i^* - \boldsymbol{\beta}_{s_i}' \mathbf{x} \mathbf{e}_i))^2 / 2) \qquad (14) \\
\cdot & \frac{1}{\sqrt{2\pi\sigma_{22}}} \exp(-\frac{(E_i^* - \boldsymbol{\beta}_{s_i}' \mathbf{x} \mathbf{e}_i)^2}{2\sigma_{22}}) \cdot (2\pi)^{-4/2} |V_c|^{-1/2} \exp(-(\mathbf{c}_{1i}^o - X C_i \Lambda)' V_c^{-1} (\mathbf{c}_{1i}^o - X C_i \Lambda)/2)]^{1-S_i}
\end{aligned}
$$

$\cdot \quad (\iota(I_i^* \geq 0) \cdot \iota(I_i = 1) + \iota(I_i^* < 0) \cdot \iota(I_i = 0))$

$$\cdot \quad \prod_{k=1}^{m} \iota_{(-\infty, \widetilde{W}_{ij}]}(\widetilde{W}_{ik}) \cdot (\frac{1}{\sqrt{2\pi}})^m \exp(-\sum_{i=1}^{N}(\widetilde{W}_{im}^2/2)) \cdot \exp(-\sum_{j=1}^{m-1}(\widetilde{\mathbf{w}}_j - \mathbf{XW}\boldsymbol{\delta}_j)'(\widetilde{\mathbf{w}}_j - \mathbf{XW}\boldsymbol{\delta}_j)/2),$$

where $\mathbf{XW} = \begin{pmatrix} \mathbf{xw}_1' \\ \vdots \\ \mathbf{xw}_N' \end{pmatrix}$ and $\widetilde{\mathbf{w}}_j = \begin{pmatrix} \widetilde{W}_{j1}' \\ \vdots \\ \widetilde{W}_{jN}' \end{pmatrix}$ for $j = 1, ..., m$.

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

1. $\boldsymbol{\alpha} \equiv [\alpha_0, \alpha_1, \alpha_2, \boldsymbol{\alpha}_3', \boldsymbol{\alpha}_4']'$;
2. $\boldsymbol{\beta}_j$, $j = 1, ..., m$;
3. γ_j, $j = 1, ..., m$;
4. $h_j \equiv \frac{1}{\sigma_j^2}$, $j = 1, ..., m$.
5. $\Lambda = [\boldsymbol{\lambda}_1', ..., \boldsymbol{\lambda}_4']'$;
6. $\nu \equiv \frac{\sigma_{12}}{\sigma_{22}}$;
7. $h_{22} \equiv \sigma_{22}^{-1}$;
8. $H_c \equiv V_c^{-1}$;
9. $\boldsymbol{\delta}_j$, $j = 1, ..., m-1$;

We specify that in the prior these vectors of parameters are independent, i.e

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\alpha}) \prod_{j=1}^{m} p(\boldsymbol{\beta}_j) \prod_{j=1}^{m} p(\gamma_j) \prod_{j=1}^{m} p(h_j) \prod_{k=1}^{4} p(\boldsymbol{\lambda}_k) p(\nu) p(h_{22}) p(H_c) \prod_{j=1}^{m-1} p(\boldsymbol{\delta}_j), \tag{15}$$

and have the following distributions:

1. $\boldsymbol{\alpha} \sim N\left(\underline{\boldsymbol{\alpha}}, \underline{\mathbf{H}}_{\alpha}^{-1}\right)$
2. $\boldsymbol{\beta}_j \sim N\left(\underline{\boldsymbol{\beta}}, \underline{\mathbf{H}}_{\beta}^{-1}\right) \qquad$ for $j = 1, \ldots, m$
3. $\boldsymbol{\gamma}_j \sim N\left(\underline{\boldsymbol{\gamma}}, \underline{h}_{\gamma}^{-1}\right)$ for $j = 1, ..., m$
4. $\boldsymbol{\lambda}_k \sim N\left(\underline{\boldsymbol{\lambda}}_k, \underline{H}_{\lambda_k}^{-1}\right)$ for k=1,...,4
5. $V_c^{-1} \equiv H_c \sim Wishart\left(\underline{V}_c, \underline{S}_c\right)$
6. $\nu \sim N\left(\underline{\nu}, \underline{h}_{\nu}^{-1}\right)$
7. $s_{22} h_{22} \sim \chi^2\left(\underline{V}_{\sigma}\right)$

8. $\boldsymbol{\delta}_j \sim N\left(\underline{\boldsymbol{\delta}}, \underline{H}_{\delta}^{-1}\right) \qquad \text{for } j = 1, \ldots, m-1$

9. $\underline{S}h_j \sim \chi^2\left(\underline{V}\right)$ for $j = 1, ..., m$

We specify the hyperparameters of this prior distribution so that to allow a substantial prior uncertainty about the parameter values[3]. This prior was checked by prior predictive analysis as discussed in Geweke (2005). Using prior distribution of the parameters, the model and the covariates $\mathbf{X}$ we have simulated prior distributions of a number of functions of interest, including averages of insurance coverage rate and health care expenditure, differences between insurance rates and health care expenditures of young and old, healthy and unhealthy individuals, variance and skewness of expenditure and differences between variances of expenditure of young and old and healthy and unhealthy individuals. These statistics computed for the actual data were all included into the intervals bounded by the 5^{th} and 95^{th} percentiles of the corresponding prior distributions. This can be interpreted as evidence that the prior distribution does not contradict the data, and that with high probability the prior and the model can generate a sample very similar to the actual observed data in terms of selected characteristics.

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 parameters and latent and missing data $p(\boldsymbol{\theta}, \mathbf{I}^*, \widehat{\mathbf{E}}, \mathbf{E}^*, \mathbf{C}^m | \mathbf{data})$, is proportional to the product of (14) and (15). To simulate from this posterior distribution we construct Gibbs sampling algorithm with Metropolis within Gibbs steps which cycles between the conditional posterior distributions of parameter blocks and the vectors of latent and missing variables $\mathbf{I}^*, \widehat{\mathbf{E}}, \mathbf{E}^*, \mathbf{C}^m$. The details of the algorithm are given in the Appendix.

5 Results

The specification of equations of the model in terms of the demographic and health status characteristics included in each equation are given in the Appendix. Because of the presence of a large number of latent variables the output of the posterior simulator exhibits high degree of autocorrelation. To allow the simulator to explore the parameter space well we ran the algorithm for an extended time period. We have obtained 5,000,000 draws from the posterior distribution and used every 1000th draw for the analysis. We have estimated four models, progressively adding potential SAS variables to the insurance equation. In the benchmark specification the insurance equation only contains expected health care expenditure and insurance pricing variables. The second model adds income and education to the first specification, the third model adds cognitive ability, financial planning horizon and longevity expectation to the second specification. Finally, the fourth model adds risk tolerance, variance of forecast error and an interaction between the risk tolerance and the variance to the third specification.

Among the primary interests of this paper is the relationship between the expected health care expenditure and the Medigap insurance status conditional on pricing variables and potential SAS variables. Figure 2 shows how this relationship changes as we progressively add potential SAS variables. This figure shows the relationship between the sample average of the probability of being insured $\Phi(\frac{\alpha_0 E_i^* + \boldsymbol{\alpha}_1' \mathbf{P}_i + \boldsymbol{\alpha}_2' \mathbf{SAS}_i}{\sqrt{\sigma_{11}}})$ (where $\mathbf{P}_i$ denotes insurance pricing variables) and the expected expenditure E^* for different sets of SAS variables included in the insurance equation. Overall, the results are consistent with the findings of FKS. Panel (a) of the figure corresponds to

[3] The details of the prior distribution are available from the authors upon request

the benchmark model with no SAS variables. In panel (a) the relationship is negative, suggesting that an increase in expected expenditure by \$10,000 (about one posterior standard deviation of E^*) decreases the probability of having Medigap insurance by 0.03. Adding income and education (panel (b)) weakens the relationship, but it remains negative. Adding cognitive ability, financial planning horizon and longevity expectation (panel (c)) changes the sign of the relationship - it becomes positive. Adding risk tolerance, variance of the forecast error and interaction of the risk tolerance with the variance (panel d) slightly increases the slope even further. Table 3 presents the posterior means, standard deviations and 5^{th} and 95^{th} percentiles of the posterior distributions of marginal effects of the variables which enter the insurance equation in the full model on the probability of Medigap coverage.

The results suggest that among the SAS variables income and cognitive ability factor are the most important sources of advantageous selection. To act as a source of advantageous selection a variable must be simultaneously correlated with E_i^* and the demand for insurance. Income and cognitive ability are the strongest sources of advantageous selection because (i) the effects on these variables on insurance demand is stronger than those of other SAS variables; and (ii) these variables are correlated with E_i^* independently of each other and the correlation of E_i^* with these variables is stronger than with other SAS variables. Adding cognitive factor to the insurance equation which already contains income and education changes the coefficient of E_i^* from negative to positive [4]; adding the remaining SAS variables increases the effect of E^* even further, but the magnitude of this change is smaller than of the change brought about by income or cognition. Hence, we have confirmed the results of FKS and identified the two most important sources of advantageous selection - income and cognitive ability. Results in Table 3 also suggest that the probability of Medigap coverage is higher for females, decreases with age and varies across locations of residence of individuals. Interestingly, in the full model the correlation between the unobservable determinants of the insurance demand and expected expenditure, $(\alpha_0 \cdot \varepsilon_2 + \varepsilon_1)$ and ε_2, is negative with the posterior mean equal to -0.25. So, the selection with respect to unobservable determinants of the expenditure is advantageous.

The results suggest that the health status variables included in the prediction model for expected expenditure include most of the information relevant to an individual when she forms expectation about the future health care costs. The posterior mean of the standard deviation of the unobserved component of the expected expenditure ε_{2i}, $\sqrt{\sigma_{22}}$, is very small compared to the standard deviation of the expected expenditure E_i^* (0.45 vs. 11.1 thousand dollars), so that any systematic difference in the expected expenditure between insured and uninsured individuals unexplained by observable health status characteristics is likely to be small. The posterior distribution of $\sqrt{\sigma_{22}}$ is shown in the middle of the right column of Figure 5.

We have estimated the model with number of types $m = 6$ because this number of mixture components provides the best fit to the expenditure data. We identify the types by the variance of the forecast error, with type 1 having the smallest variance, and type 6 having the largest variance. Table 4 shows some type-specific parameters and functions of interest. The table suggests that types 1 and 2 are the healthiest and together make up about 67% of the sample. These two types have the lowest expected expenditures and incentive effects of Medigap insurance. Type 6 makes up about 3% of the sample and is the most unhealthy type with the highest expected expenditure and the highest Medigap effect.

The fit of the model which includes all potential SAS variables is examined in Figure 3. The

[4]The results from this model are not presented but are available from the authors upon request

Figure 2: Relationship between expected expenditure and probability of Medigap coverage

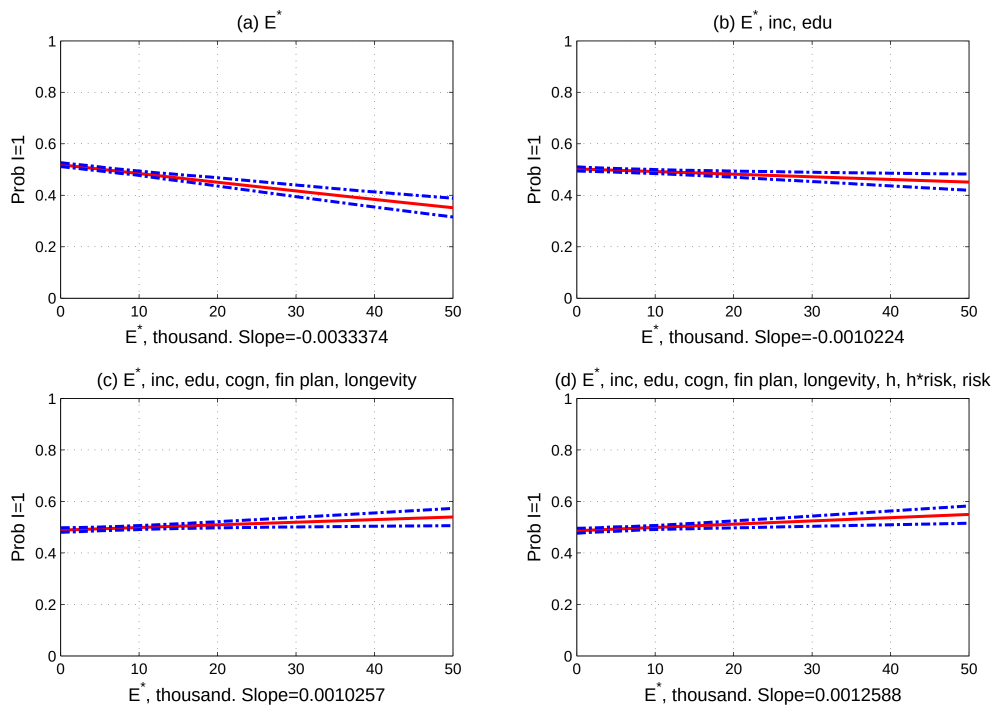


Table 3: Marginal effects of individual characteristics on the probability of Medigap insurance coverage computed for a median individual

Variable	Post. mean	Post.std	5th prct.	95th prct.
E^*	0.0177	0.0059	0.0083	0.0274
σ^2_η	-0.0012	0.0036	-0.0071	0.0048
Female	0.1070	0.0128	0.0859	0.1284
Age	-0.2805	0.1223	-0.4750	-0.0825
Cen: nweng	-0.0638	0.0341	-0.1234	-0.0110
Cen: midatl	0.0859	0.0290	0.0406	0.1357
Cen: encen	0.1500	0.0329	0.0990	0.2097
Cen: wncen	0.0579	0.0288	0.0120	0.1078
Cen: satl	0.1540	0.0354	0.0973	0.2149
Cen: escen	0.0687	0.0316	0.0177	0.1223
Cen: wscen	-0.0771	0.0375	-0.1415	-0.0155
Cen: mnt	-0.2079	0.0366	-0.2689	-0.1489
Cen: pac	0.1007	0.0380	0.0372	0.1641
hgc: ls8th	-0.0945	0.0736	-0.2238	0.0191
hgc: somehs	-0.0934	0.0740	-0.2269	0.0187
hgc: hs	-0.0919	0.0741	-0.2190	0.0230
hgc: somecol	-0.1169	0.0773	-0.2477	0.0027
hgc: college	-0.1517	0.0822	-0.2962	-0.0264
hgc: gradschl	-0.1027	0.0788	-0.2463	0.0185
hgc: nr	0.1230	0.0364	0.0665	0.1825
inc 5k-10k	-0.1082	0.0461	-0.1812	-0.0286
inc 10k-15k	0.0712	0.0418	0.0008	0.1396
inc 15k-20k	0.1026	0.0416	0.0329	0.1705
inc 20k-25k	0.1443	0.0407	0.0788	0.2075
inc 25k-30k	0.1797	0.0411	0.1125	0.2464
inc 30k-35k	0.1724	0.0430	0.1040	0.2424
inc 35k-40k	0.1537	0.0434	0.0826	0.2222
inc 40k-45k	0.1959	0.0445	0.1260	0.2699
inc 45k-50k	0.2118	0.0448	0.1393	0.2888
inc 50plus	0.2322	0.0413	0.1652	0.2990
risktol	-0.0192	0.0124	-0.0407	0.0009
cogn	0.1483	0.0238	0.1092	0.1887
finpln	0.0291	0.0123	0.0090	0.0492
praliv75	-0.0126	0.0115	-0.0320	0.0058

Table 4: Posterior means of selected type-specific characteristics

Variable	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
$\sqrt{\sigma^2_j}$, dollars	681	1,667	3,536	7,003	15,172	38,156
γ_j , dollars	578	1,384	2,225	3,346	2,260	7,035
Type sample proportion	0.3887	0.2843	0.1421	0.0856	0.0696	0.0297
Average of $E^*_i \mid \mathbf{xe}_i, j$, dollars	868	3,018	7,320	13,867	24,960	52,792

top two panels show the plot of the kernel density estimates of the actual and predicted health care expenditure for individuals in the estimation sample. The smooth mixture of Tobits does an amazing job in capturing the overall complex shape of the expenditure distribution. The two panels in the middle row and the left panel in the bottom row show how well the model captures various aspects of the overall and conditional distribution of expenditure. The intersection of two straight lines in these panels represent values corresponding to the actual data, and the red dots represent the posterior distributions of these statistics. The model can replicate well the 95th percentile of expenditure distribution, as well as skewness, overall variance and variance conditional on health status of individuals (defined by being in the top or bottom 50 percent of unhealthy factor F2). The model over-predicts the 99th percentile of the expenditure distribution, but still places a non-zero probability on the event that the 99th percentile of the predicted expenditure is less or equal to the 99th percentile of the actual expenditure. Finally, the right panel in the bottom row of Figure 3 shows the actual and predicted relationship between expenditure and insurance coverage of individuals in the estimation sample. The model can reproduce the actual relationship between the two variables well.

Figure 4 presents the fit of the model to the conditional mean of the health care expenditure and compares it to several standard models for the conditional mean of expenditure used in the literature. All panels of Figure 4 plot deciles of predicted expenditure against average expenditure corresponding to these deciles for different models. The panels are sorted by the mean squared prediction error (MSE) corresponding to the models. Our SMTobit model outperforms generalized linear models (GLM) with log link and Gaussian, Poisson and gamma distributional families, as well as OLS and a two-part log-normal model. All these models were included in the set of standard models for health expenditure compared in the study by Buntin and Zaslavsky (2004).

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

1. The expected health care expenditure in uninsured state (columns 1-3)[5]:

$$\begin{aligned} E(E_i^*|\mathbf{xe}_i) &= \sum_{j=1}^{m} E(E_i^*|\mathbf{xe}_i, \text{type}_i = j) \cdot Pr(\text{type}_i = j|\mathbf{xw}_i) \\ &= \sum_{j=1}^{m} (\mathbf{xe}_i'\boldsymbol{\beta}_j) \cdot \int_{-\infty}^{\infty} \phi(d - \boldsymbol{\delta}_j'\mathbf{xw}_i) \prod_{l \neq j}^{m} \Phi(d - \boldsymbol{\delta}_l'\mathbf{xw}_i) dd; \end{aligned} \tag{16}$$

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

$$E(MH_i|\mathbf{xe}_i) = \sum_{j=1}^{m} \gamma_j \cdot Pr(\text{type}_i = j|\mathbf{xw}_i) = \sum_{j=1}^{m} \gamma_j \cdot \int_{-\infty}^{\infty} \phi(d - \boldsymbol{\delta}_j'\mathbf{xw}_i) \prod_{l \neq j}^{m} \Phi(d - \boldsymbol{\delta}_j'\mathbf{xw}_i) dd; \tag{17}$$

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

$$Std.Dev(\eta_i|\mathbf{xe}_i) = \left(\sum_{j=1}^{m} \sigma_j^2 \cdot Pr(\text{type}_i = j|\mathbf{xw}_i) \right)^{\frac{1}{2}} = \left(\sum_{j=1}^{m} \sigma_j^2 \cdot \int_{-\infty}^{\infty} \phi(d - \boldsymbol{\delta}_j'\mathbf{xw}_i) \prod_{l \neq j}^{m} \Phi(d - \boldsymbol{\delta}_j'\mathbf{xw}_i) dd \right)^{\frac{1}{2}}. \tag{18}$$

[5] Because in our analysis $\mathbf{xw}_i$ is a subset of $\mathbf{xe}_i$, conditioning on $\mathbf{xe}_i$ is equivalent to conditioning on both $\mathbf{xw}_i$ and $\mathbf{xe}_i$

Figure 3: Model fit

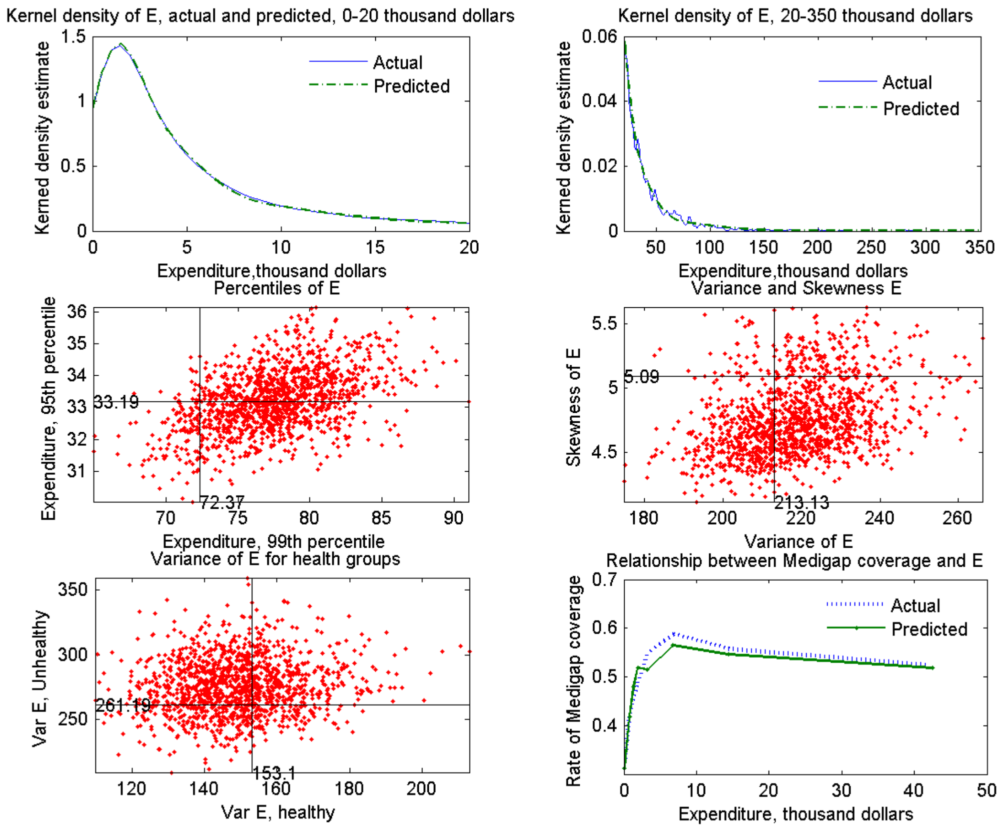


Figure 4: Conditional mean of expenditure: comparing SMTobit with other models for expenditure

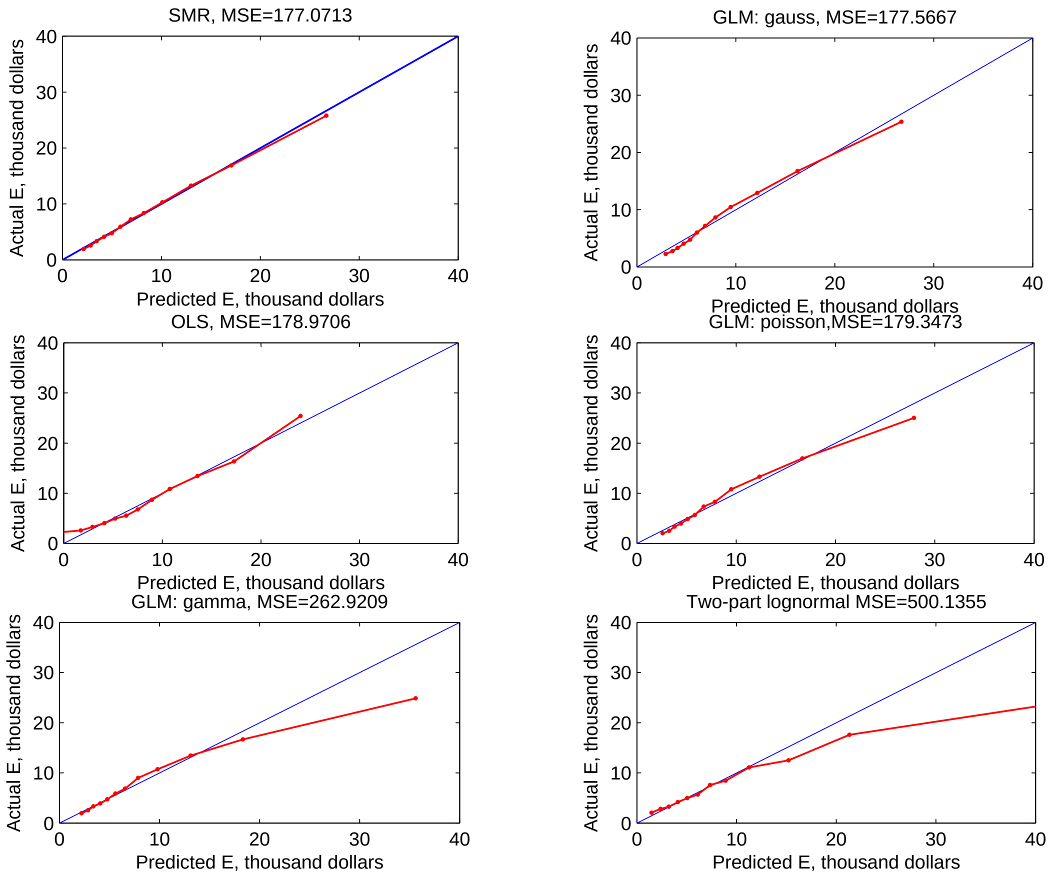


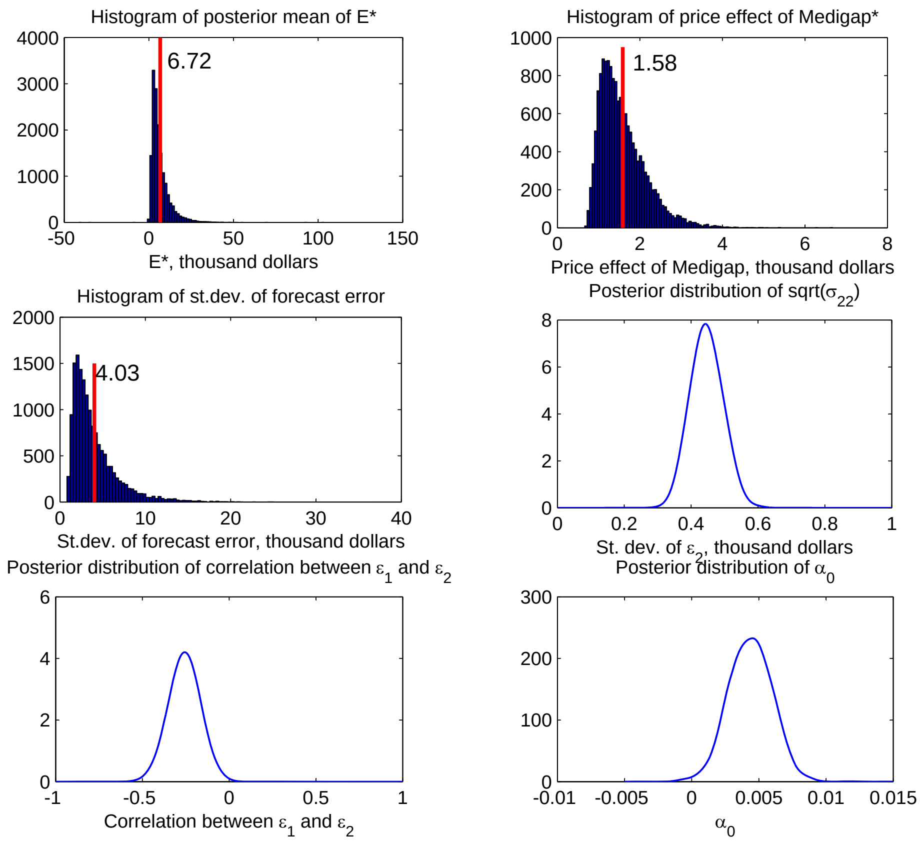
Table 5: Marginal effects of individual characteristics on the expected expenditure E_i^*, the effect of the Medigap insurance and the standard deviation of the forecast error η (in thousand dollars), computed for a median individual

	E^*			Medigap			St. dev. of η		
Variable	Post. mean 1	Post. std 2	$(5^{th}, 95^{th})$ prct. 3	Post. mean 4	Post. std 5	$(5^{th}, 95^{th})$ prct. 5	Post. mean 7	Post. std 8	$(5^{th}, 95^{th})$ prct. 9
Female	-0.65	0.26	(-1.12, -0.28)	0.01	0.05	(-0.07, 0.09)	-1.20	0.28	(-1.66, -0.77)
Age	0.29	0.14	(0.07, 0.51)	0.04	0.03	(-0.01, 0.08)	0.10	0.20	(-0.23, 0.45)
Cen: nweng	0.18	0.40	(-0.50, 0.82)	0.11	0.09	(-0.02, 0.26)	0.56	0.56	(-0.37, 1.52)
Cen: midatl	-0.83	0.35	(-1.42, -0.27)	-0.03	0.08	(-0.18, 0.11)	-1.37	0.48	(-2.17, -0.60)
Cen: encen	-1.68	0.52	(-2.60, -0.89)	-0.003	0.11	(-0.19, 0.17)	-1.91	0.53	(-2.85, -1.05)
Cen: wncen	-0.65	0.35	(-1.24, -0.10)	0.01	0.07	(-0.12, 0.12)	-1.29	0.47	(-2.09, -0.55)
Cen: satl	-1.44	0.49	(-2.29, -0.67)	0.02	0.10	(-0.16, 0.18)	-2.16	0.52	(-3.02, -1.35)
Cen: escen	-0.36	0.34	(-0.93, 0.17)	0.01	0.09	(-0.13, 0.16)	-1.1	0.50	(-1.95, -0.29)
Cen: wscen	-1.12	0.53	(-1.97, -0.27)	-0.06	0.11	(-0.25, 0.12)	-2.45	0.53	(-3.42, -1.63)
Cen: mnt	-0.76	0.40	(-1.46, -0.16)	0.01	0.10	(-0.16, 0.17)	-1.53	0.49	(-2.34, -0.75)
Cen: pac	-2.04	8.59	(-16.57, 12.76)	-0.26	0.17	(-0.50, -0.03)	-4.29	0.64	(-5.26, -3.43)
Fh2	2.81	0.42	(2.17, 3.56)	0.23	0.09	(0.10, 0.38)	2.64	0.40	(2.02, 3.33)
Fh3	-1.36	0.18	(-1.68, -1.09)	-0.16	0.04	(-0.23, -0.09)	-1.40	0.16	(-1.68, -1.14)
Fh7	0.33	0.12	(0.16, 0.55)	-0.005	0.03	(-0.05, 0.04)	0.59	0.18	(0.30, 0.89)
Fh8	0.75	0.12	(0.56, 0.96)	0.08	0.03	(0.04, 0.13)	0.72	0.18	(0.44, 1.03)
Fh10	-0.05	0.17	(-0.32, 0.23)	0.002	0.02	(-0.03, 0.04)	-0.06	0.20	(-0.39, 0.26)
Fh11	-0.42	0.08	(-0.56, -0.29)	-0.07	0.02	(-0.09, -0.04)	-0.26	0.15	(-0.50, -0.02)
Fh17	0.71	0.23	(0.35, 1.10)	0.11	0.05	(0.04, 0.19)	0.45	0.25	(0.01, 0.86)
Fh20	0.23	0.15	(-0.01, 0.51)	0.05	0.03	(0.01, 0.10)	0.07	0.19	(-0.24, 0.39)
Fh22	0.16	0.15	(-0.08, 0.40)	0.02	0.02	(-0.01, 0.06)	0.25	0.27	(-0.161, 0.72)
Fh23	0.01	0.11	(-0.15, 0.19)	-0.02	0.02	(-0.05, 0.005)	0.37	0.23	(0.01, 0.77)
Black	-0.72	0.32	(-1.25, -0.21)	-0.18	0.04	(-0.25, -0.115)	0.26	0.39	(-0.33, 0.97)
Hispanic	-0.91	0.33	(-1.44, -0.38)	-0.12	0.04	(-0.19, -0.055)	-0.63	0.43	(-1.35, 0.10)
Married	-0.13	0.23	(-0.53, 0.22)	0.02	0.04	(-0.05, 0.08)	-0.48	0.32	(-1.03, 0.02)
Year	-1.10	0.19	(-1.41, -0.82)	-0.13	0.04	(-0.19, -0.08)	-1.03	0.21	(-1.39, -0.69)

The effects are evaluated for a median individual, e.g. an individual with the covariates $\mathbf{xe}_i$ and $\mathbf{xw}_i$ set at their sample median. The uncertainty described by the standard deviations and the percentiles in Table 5 is with respect to the posterior distribution of parameters. For continuous variables we report the effects brought about by one standard deviation increase in the variable of interest from the sample median level. Figure 5 plots sample distributions of (16), (17) and (18).

The top leftmost panel of Figure 5 shows the distribution of the conditional expectation of expenditure in uninsured state $E(E_i^*|\mathbf{xe}_i, \mathbf{data}) = \int E(E_i^*|\mathbf{xe}_i, \boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{data})d\boldsymbol{\theta}$ among individuals in our sample. This expectation is conditional on observed health status and demographic characteristics, after integration with respect to the latent types and the posterior distribution of parameters has been carried out. The expected expenditure $E(E_i^*|\mathbf{xe}_i, \mathbf{data})$ varies between 0 and 50,000 dollars, with the average expenditure equal to 6,720 dollars. Columns 1-3 of Table 5 summarize the posterior distribution of the effects of demographic and health status characteristics on $E(E_i^*|\mathbf{xe}_i)$. The expected expenditure in uninsured state is lower for females, blacks and hispanics, increases

Figure 5: Posterior distributions of selected functions of interest



with age, is in general higher for respondents who did not report the Census Division of their residence (omitted group of the Census Division indicators), increases with the unhealthy factor Fh2 and decreases with the healthy factor Fh3.

Right panel of the top row of Figure 5 shows the histogram of the sample distribution of the moral hazard (or price) effect of Medigap coverage on potential health care expenditure $E(MH_i|\mathbf{xe}_i, \mathbf{data}) = \int E(MH_i|\mathbf{xe}_i, \boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{data})d\boldsymbol{\theta}$. The effect varies between \$700 and \$7,000, with the average effect equal to 1,580 dollars. Columns 4-6 of Table 5 summarize the posterior distribution of the impacts of individual characteristics on this effect. The Medigap effect is higher for unhealthier individuals and is lower for blacks and hispanics.

Left panel of the middle row of Figure 5 shows the histogram of the sample distribution of the standard deviation of forecast error η, $Std.Dev(\eta_i|\mathbf{xe}_i, \mathbf{data}) = \int Std.Dev(\eta_i|\mathbf{xe}_i, \boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{data})d\boldsymbol{\theta}$. The standard deviation of forecast error varies between 1 and 20 thousand dollars, with average standard deviation being equal to 4,030 dollars. The marginal effects in columns 7-9 of Table 5 suggest that this standard deviation is lower for females and is higher for unhealthier individuals and individuals with missing information about their location of residence.

6 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 in the US. 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 expenditure in uninsured state, conditional on other variables. Our model allows for correlation between unobserved determinants of health care expenditure and the demand for insurance. 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). The model is estimated using a MCMC algorithm with data augmentation.

We find that conditionally on income, education, risk attitudes, cognitive ability, financial planning horizon and longevity expectations there is a small adverse selection into Medigap insurance: a one standard deviation increase in expected health care expenditure in uninsured state increases probability of buying insurance by 0.01. We also estimate sample distribution of the moral hazard (or price) effect of Medigap insurance on health care expenditure. We find that on average an insured individual spends about \$1,600 more on health care than a similar individual who is uninsured, and that this effect varies with individual characteristics - it is higher for unhealthier individuals and lower for blacks and hispanics. Our econometric model for conditional distribution of health expenditures is found in our data to be superior to several standard models currently used to model expenditure. The smooth mixture of Tobits model which we develop predicts conditional expectation of health care expenditure better than these standard models and also correctly captures the shape of the expenditure distribution as well as conditional variance and skewness.

Appendix

6.1 Posterior Simulation Algorithm

The Gibbs sampling algorithm constructed for the model is section (4) iterates between the following conditional posterior distributions:

1. $\boldsymbol{\mu} \sim p(\boldsymbol{\mu}|\boldsymbol{\theta}_{-\mu}, \mathbf{E}^*, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{data})$ where $\boldsymbol{\mu} \equiv [\boldsymbol{\alpha}', \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}'_m, \boldsymbol{\lambda}_1, ..., \boldsymbol{\lambda}_4, \gamma_1, ..., \gamma_m]'$.

   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. To derive the conditional posterior distribution of $\boldsymbol{\mu}$ we need to establish the following notation. Let

   $$\widetilde{\mathbf{s}} = [s_{ij}], \text{ where } \widetilde{s}_{ij} = \iota(s_i = j),$$

   $$\mathbf{C}_k = [C_{ik}], \text{ where } C_{ik} = [(c^o_{k1} \cdot S_1 + c^m_{k1} \cdot (1 - S_1)), ..., (c^o_{kN} \cdot S_N + c^m_{kN} \cdot (1 - S_N))]', \ k = 1, ..., 4$$

   and

   $$\widetilde{\mathbf{E}} = [\widehat{E}^o_1 - E^*_1, ..., \widehat{E}^o_{N_M} - \widehat{E}^*_{N_M}].'$$

   Define vector $\mathbf{y}^*$ as

   $$\mathbf{y}^* = [\mathbf{I}^{*'}, \mathbf{E}^{*'}, \mathbf{C}'_1, ..., \mathbf{C}'_4, \widetilde{\mathbf{E}}']'.$$

   Then let

   $$\begin{aligned} \mathbf{Z}_1 &= [E^*_i, \sigma^2_{s_i}, \sigma^2_{s_i} \cdot c_{1i}, \mathbf{xi}'_i, C_{1i}, ..., C_{4i}], \\ \mathbf{Z}_2 &= (\boldsymbol{\iota}'_m \otimes \mathbf{XE}) \circ (\mathbf{S} \otimes \boldsymbol{\iota}'_{K_{XE}}) \\ \mathbf{Z}_3 &= \mathbf{D}_4 \otimes \mathbf{XC} \\ \mathbf{Z}_4 &= (\boldsymbol{\iota}'_m \otimes \mathbf{I}_{N_M}) \circ \widetilde{\mathbf{s}}_{N_M}, \end{aligned}$$

   where $\boldsymbol{\iota}_m$ is an identity vector of size m, $\mathbf{D}_K$ is an identity matrix of size $K \times K$, K_{XE} is a number of columns in the matrix $\mathbf{XE}$ and the notation $\mathbf{M}_K$ indicates the matrix obtained from matrix $\mathbf{M}$ by deleting the last $N - K$ rows. Then define matrix $\mathbf{Z}$ as follows:

   $$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & \mathbf{Z}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & \mathbf{0}' & \mathbf{Z}_3 & \mathbf{0} \\ \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{Z}_4 \end{bmatrix}.$$

   Finally, let

   $$\mathbf{V}_y = \begin{bmatrix} \widetilde{\mathbf{V}} \otimes I_N & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix},$$

   where $\widetilde{\mathbf{V}} = \begin{bmatrix} \Sigma_{12} & \mathbf{0} \\ \mathbf{0} & V_c \end{bmatrix}$, $\Sigma_{12} = \begin{bmatrix} 1 + \frac{\sigma^2_{12}}{\sigma_{22}} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$ and $\mathbf{Q}$ is $N_{MCBS} \times N_{MCBS}$ diagonal matrix with diagonal element $q_{ii} = \sigma^2_{s_i}$ for $i = 1, ..., N_{MCBS}$. Then the augmented data density in

(14) can be expressed:

$$
\begin{aligned}
& p(\mathbf{I}^*, \mathbf{E}^*, \widehat{\mathbf{E}}, \mathbf{C}^o, \mathbf{C}^m, \mathbf{s}, \mathbf{I}, \mathbf{E}^o, \mathbf{W}|\mathbf{S}, \mathbf{M}, \boldsymbol{\theta}) = \\
= \; & (2\pi)^{-\frac{1}{2}(2N+4+N_M)}\sigma_{22}^{-\frac{N}{2}}|V_c|^{-\frac{N}{2}}|\mathbf{Q}|^{-\frac{1}{2}}\exp(-(\mathbf{y}^* - \mathbf{Z}\boldsymbol{\mu})'\mathbf{V}_y^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\mu})/2) \\
\cdot \; & \prod_{i=1}^{N_M}(\iota(E_i^o = \widehat{E}_i)\cdot\iota(\widehat{E}_i \geq 0) + \iota(E_i^o = 0)\cdot\iota(\widehat{E}_i < 0)) \\
\cdot \; & \prod_{i=1}^{N}(\iota(I_i^* \geq 0)\cdot\iota(I_i = 1) + \iota(I_i^* < 0)\cdot\iota(I_i = 0))\cdot\prod_{k=1}^{m}\iota_{(-\infty,\widetilde{W}_{ij}]}(\widetilde{W}_{ik}) \\
\cdot \; & (\frac{1}{\sqrt{2\pi}})^m\exp(-\sum_{i=1}^{N}(\widetilde{W}_{im}^2/2))\cdot\exp(-\sum_{j=1}^{m-1}(\widetilde{\mathbf{W}}_j - \mathbf{XW}\boldsymbol{\delta}_j)'(\widetilde{\mathbf{W}}_j - \mathbf{XW}\boldsymbol{\delta}_j)/2). \quad (19)
\end{aligned}
$$

The conditional posterior distribution of $\boldsymbol{\mu}$ is proportional to the product of prior probabilities of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\gamma}$ and (19) and is given by:

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

where

$$
\begin{aligned}
\overline{\mathbf{H}}_\mu &= \underline{\mathbf{H}}_\mu + \mathbf{Z}'\mathbf{V}_y^{-1}\mathbf{Z}, \\
\overline{\boldsymbol{\mu}} &= \overline{\mathbf{H}}_\mu^{-1}[\underline{\mathbf{H}}_\mu\underline{\boldsymbol{\mu}} + \mathbf{Z}\mathbf{V}_y^{-1}\mathbf{y}^*].
\end{aligned}
$$

2. $\nu \sim p(\nu|\boldsymbol{\theta}_{-\nu}, \mathbf{E}^*, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{data})$. To express the posterior distribution of ν we will establish the following notation. Let

$$\widetilde{\mathbf{I}}^* = \mathbf{I}_i^* - \alpha_0\mathbf{E}_i^* - \alpha_1\boldsymbol{\sigma}^2 - \alpha_2\boldsymbol{\sigma}^2 \circ \mathbf{C}_1 - \mathbf{XI}\boldsymbol{\alpha}_3 - [\mathbf{C}_1, ..., \mathbf{C}_4]\boldsymbol{\alpha}_4,$$

where $\boldsymbol{\sigma}^2$ denotes a vector of individual type-specific variances: $\boldsymbol{\sigma}^2 = [\sigma_{s_i}^2]$. Also, let

$$\widetilde{\mathbf{E}}^* = \mathbf{E}^* - \mathbf{Z}_2\boldsymbol{\beta}.$$

The posterior distribution of ν is proportional to the product of prior probability of ν and (14) and is given by:

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

where

$$
\begin{aligned}
\overline{h}_\nu &= \underline{h}_\nu + \widetilde{\mathbf{E}}^{*'}\widetilde{\mathbf{E}}^*, \\
\overline{\nu} &= \overline{h}_\nu^{-1}(\underline{h}_\nu\underline{\nu} + \widetilde{\mathbf{E}}^{*'}\widetilde{\mathbf{I}}^*).
\end{aligned}
$$

3. $h_{22} \sim p(h_{22}|\boldsymbol{\theta}_{-h_{22}}, \mathbf{E}^*, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{s}, \mathbf{data})$. The posterior conditional distribution of h_{22} is proportional to the product of prior probability of σ_{22} and (14) and is given by:

$$\overline{s}^2h_{22}|(\boldsymbol{\theta}_{-h_{22}}, \mathbf{E}^*, \mathbf{I}^*, \mathbf{C}^m, \mathbf{data}) \;\sim\; \chi^2(\overline{v}_\sigma),$$

where

$$
\begin{aligned}
\overline{s}^2 &= \underline{s}^2 + \widetilde{\mathbf{E}}^{*'}\widetilde{\mathbf{E}}^*, \\
\overline{v}_\sigma &= \underline{v}_\sigma + N.
\end{aligned}
$$

4. $H_c \sim p(H_c|\boldsymbol{\theta}_{-H_c}, \mathbf{E}^*, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{s}, \mathbf{data})$. The posterior conditional distribution of H_c is the product of prior probability of H_c and (14) and is given by:

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

where

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

5. $\boldsymbol{\delta} \sim p(\boldsymbol{\delta}|\boldsymbol{\theta}_{-\delta}, \mathbf{E}^*, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{s}, \mathbf{data})$. This conditional posterior is proportional to the product of prior probability $\boldsymbol{\delta}$ and (14). It is easy to see that the posterior conditional distributions of $\boldsymbol{\delta}_j$ are independent:

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

where

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

6. $p(h_j|\boldsymbol{\theta}_{-h_j}, \mathbf{E}^*, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{s}, \mathbf{data})$. This conditional posterior distribution is proportional to the product of prior probability of h_j and (14). The kernel of this distribution can be expressed:

$$\begin{aligned} p(h_j|\boldsymbol{\theta}_{-h_j}, \mathbf{E}^*, \widehat{\mathbf{E}}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{s}, \mathbf{data}) &\propto \prod_{i:s_i=j}^{N} \exp(-(\widetilde{I}_i^* - \frac{\alpha_1}{h_j} - \frac{\alpha_2 c_{1i}}{h_j})^2/2) \\ &\cdot \prod_{i:s_i=j}^{N^M} h_j^{\frac{1}{2}} \exp(-h_j \widetilde{E}_i^2/2) \cdot h_j^{(\underline{V}-2)/2} \exp(-(\underline{S} h_j/2) \\ &= \exp(\sum_{i:s_i=j}^{N} \frac{-(\widetilde{I}_i^* - \frac{\alpha_1}{h_j} - \frac{\alpha_2 c_{1i}}{h_j})^2}{2}) \cdot (\frac{n_j^M + \underline{V} - 2}{2}) \exp(\frac{-h_j(\sum_{i:s_i=j}^{N^M} \widetilde{E}_i^2 + \underline{S})}{2}) \end{aligned} \quad (20)$$

where

$$\widetilde{\widetilde{I}}_i^* = I_i^* - \alpha_1 E_i^* - \boldsymbol{\alpha}_3' \mathbf{xi}_i - [C_{1i}, ..., C_{4i}]\boldsymbol{\alpha}_4 - \nu(E_i^* - \boldsymbol{\beta}_{s_i}' \mathbf{xe}_i)$$

and

$$\widetilde{E}_i = E_i^o - E_i^* - \gamma I_i.$$

The second term in this expression is the kernel of a gamma distribution with parameters $\frac{n_j^M + \underline{V}}{2}$ and $\frac{2}{\sum_{i:s_i=j}^{N^M} \widetilde{E}_i^2 + \underline{S}_j}$, while the first term cannot be recognized as a kernel of any known probability distribution. The draws from (20) will be obtained by Metropolis-Hastings algorithm. On n'th iteration of the sampler draw candidate $\widetilde{h}_j$ from the gamma distribution whose kernel is the second term of (20) and accept it with probability

$$\rho = \min\left\{\frac{\exp\left(-\sum_{i:s_i=j}^{N}(\widetilde{I}_i^* - \frac{\alpha_0}{\widetilde{h}_j} - \frac{\alpha_2 c_{1i}}{\widetilde{h}_j})^2/2\right)}{\exp\left(-\sum_{i:s_i=j}^{N}(\widetilde{I}_i^* - \frac{\alpha_1}{h_j^{n-1}} - \frac{\alpha_2 c_{1i}}{h_j^{n-1}})^2/2\right)}, 1\right\}.$$

7. $I_i^*|(\boldsymbol{\theta}, \mathbf{E}^*, \widehat{\mathbf{E}}, \mathbf{I}_{-i}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{s}, \mathbf{data})$. From (14) the kernel of this posterior distribution is given by

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

which can be recognized as a kernel of a truncated normal distribution. Thus,

$$I_i^*|(\boldsymbol{\theta}, \mathbf{E}^*, \widehat{\mathbf{E}}, \mathbf{I}_{-i}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{s}, \mathbf{data}) \sim TN_{R(I_i)}(\overline{I}_i^*, 1),$$

where $TN_R(a, b)$ denotes normal distribution with mean a and variance b truncated to interval R, $R(I_i = 0) = (-\infty, 0]$, $R(I_i = 1) = (0, \infty)$ and

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

8. $p(E_i^*|\boldsymbol{\theta}, \mathbf{E}_{-i}^*, \widehat{\mathbf{E}}_{-i}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{s}, \mathbf{data})$. This distribution needs to be derived for two cases: $S_i = 0$ and $S_i = 1$. From (14) the kernel of this distribution when $S_i = 0$ is given by

$$\exp(-(I_i^* - \alpha_0 E_i^* - \alpha_1 \sigma_{s_i}^2 - \alpha_2 \sigma_{s_i}^2 c_{1i}^o - \boldsymbol{\alpha}_3' \mathbf{xi}_i - \boldsymbol{\alpha}_4' \mathbf{c}_i^o - \frac{\sigma_{12}}{\sigma_{22}}(E_i^* - \boldsymbol{\beta}_{s_i}' \mathbf{xe}_i))^2/2)$$
$$\cdot \quad \frac{1}{\sqrt{2\pi\sigma_{22}}} \exp(-\frac{(E_i^* - \boldsymbol{\beta}_{s_i}' \mathbf{xe}_i)^2}{2\sigma_{22}}.$$

This is the kernel of the distribution $p(E_i^*|I_i^*; \boldsymbol{\theta}, \mathbf{c}_i^o, \mathbf{xi}_i, \mathbf{xe}_i)$ where the joint conditional distribution $p(I_i^*, E_i^*|\boldsymbol{\theta}, \mathbf{c}_i^o, \mathbf{xi}_i, \mathbf{xe}_i)$ is bivariate normal, i.e. $p([I_i^*, E_i^*]'|\boldsymbol{\theta}, \mathbf{c}_i^o, \mathbf{xi}_i, \mathbf{xe}_i)$

$$\sim N\left(\begin{array}{l} \alpha_0 \boldsymbol{\beta}_{s_i}' \mathbf{xe}_i + \alpha_1 \sigma_{s_i}^2 + \alpha_2 \sigma_{s_i}^2 c_{1i}^m + \boldsymbol{\alpha}_3' \mathbf{xi}_i + \boldsymbol{\alpha}_4' \mathbf{c}_i^m \\ \boldsymbol{\beta}_{s_i}' \mathbf{xe}_i \end{array}, \left[\begin{array}{cc} \sigma_{s_i}^2 + \sigma_{22} & \sigma_{12} + \alpha_0 \sigma_{22} \\ \sigma_{12} + \alpha_0 \sigma_{22} & 1 + \frac{\sigma_{12}^2}{\sigma_{22}} + \alpha_0^2 \sigma_{22} + 2\alpha_0 \sigma_{12} \end{array}\right]\right).$$

Using the standard result about the conditional distributions for a multivariate normal distribution (e.g. Goldberger (1991), chapter 18), the posterior distribution $p(E_i^*|\boldsymbol{\theta}, \mathbf{E}_{-i}^*, \widehat{\mathbf{E}}_{-i}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{s}, \mathbf{data}, S_i = 0)$ is $N(\overline{E}_i^*, Var_{E^*})$, where

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

and

$$Var_{E^*} = \sigma_{22} - \frac{(\sigma_{12} + \alpha_0 \sigma_{22})^2}{1 + \frac{\sigma_{12}^2}{\sigma_{22}} + \alpha_0^2 \sigma_{22} + 2\alpha_0^2 \sigma_{12}}.$$

The kernel of the conditional posterior distribution $p(E_i^*|\boldsymbol{\theta}, \mathbf{E}_{-i}^*, \widehat{\mathbf{E}}_{-i}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{s}, \mathbf{data}, S_i = 1)$ is given by

$$\exp(-(I_i^* - \alpha_0 E_i^* - \alpha_1 \sigma_{s_i}^2 - \alpha_2 \sigma_{s_i}^2 c_{1i}^o - \boldsymbol{\alpha}_3' \mathbf{xi}_i - \boldsymbol{\alpha}_4' \mathbf{c}_i^o - \frac{\sigma_{12}}{\sigma_{22}}(E_i^* - \boldsymbol{\beta}_{s_i}' \mathbf{xe}_i))^2/2)$$
$$\cdot \quad \frac{1}{\sqrt{2\pi\sigma_{22}}} \exp(-\frac{(E_i^* - \boldsymbol{\beta}_{s_i}'\mathbf{xe}_i)^2}{2\sigma_{22}} \cdot \exp(-\frac{(\widehat{E}_i - E_i^* - \gamma_{s_i} I_i)^2}{2\sigma_j^2}).$$

This is the kernel of the distribution $p(E_i^*|\widehat{E}_i, I_i^*, ; \boldsymbol{\theta}, \mathbf{c}_i^o, \mathbf{xi}_i, \mathbf{xe}_i, I_i)$ where the joint conditional distribution $p([\widehat{E}_i, I_i^*, E_i^*]'|\boldsymbol{\theta}, \mathbf{c}_i^o, \mathbf{xi}_i, \mathbf{xe}_i)$ is multivariate normal with the mean

$$\left(\begin{array}{l} \boldsymbol{\beta}_{s_i}'\mathbf{xe}_i + \gamma_{s_i} I_i \\ \alpha_0 \boldsymbol{\beta}_{s_i}'\mathbf{xe}_i + \alpha_1 \sigma_{s_i}^2 + \alpha_2 \sigma_{s_i}^2 c_{1i} + \boldsymbol{\alpha}_3'\mathbf{xi}_i + \boldsymbol{\alpha}_4'\mathbf{c}_i \ \ , \mathbf{V}_{s_i} \\ \boldsymbol{\beta}_{s_i}'\mathbf{xe}_i \end{array} \right),$$

and the variance

$$\mathbf{V}_{s_i} = \left[\begin{array}{ccc} \sigma_{s_i}^2 + \sigma 22 & \sigma_{12} + \alpha_0 \sigma_{22} & \sigma_{22} \\ \sigma_{12} + \alpha_0 \sigma_{22} & 1 + \frac{\sigma_{12}^2}{\sigma_{22}} + \alpha_0^2 \sigma_{22} + 2\alpha_0 \sigma_{12} & \sigma_{12} + \alpha_0 \sigma_{22} \\ \sigma_{22} & \sigma_{12} + \alpha_0 \sigma_{22} & \sigma_{22} \end{array} \right].$$

To derive the distribution of $p(E_i^*|\boldsymbol{\theta}, \mathbf{E}_{-i}^*, \widehat{\mathbf{E}}_{-i}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{s}, \mathbf{data}, S_i = 1)$ using the result for the multivariate normal distribution partition $\mathbf{V}_{s_i}$ as follows:

$$\mathbf{V}_{s_i} = \left[\begin{array}{cc} \mathbf{V}_{s_i 11} & \mathbf{V}_{s_i 12} \\ \mathbf{V}_{s_i 21} & \mathbf{V}_{s_i 22} \end{array} \right],$$

where $\mathbf{V}_{s_i 11}$ is 2×2 and $\mathbf{V}_{s_i 22} = \sigma_{22}$. With this notation the posterior conditional distribution of E_i^* when $S_i = 1$ is given by:

$$E_i^*|\boldsymbol{\theta}, \mathbf{E}_{-i}^*, \widehat{\mathbf{E}}_{-i}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{s}, \mathbf{data}, S_i = 1 \sim N(\boldsymbol{\beta}_{s_i}'\mathbf{xe}_i + \mathbf{V}_{s_i 12}' \mathbf{V}_{s_i 11}^{-1}(\mathbf{d}_i - \overline{\mathbf{d}}_i), \mathbf{V}_{s_i 22} - \mathbf{V}_{s_i 12}' \mathbf{V}_{s_i 11}^{-1} \mathbf{V}_{s_i 12}),$$

where

$$\mathbf{d}_i = [\widehat{E}_i, I_i^*]'$$

and

$$\overline{\mathbf{d}}_i = \left[\begin{array}{c} \boldsymbol{\beta}_{s_i}'\mathbf{xe}_i + \gamma_{s_i} I_i \\ \alpha_0 \boldsymbol{\beta}_{s_i}'\mathbf{xe}_i + \alpha_1 \sigma_{s_i}^2 + \alpha_2 \sigma_{s_i}^2 c_{1i} + \boldsymbol{\alpha}_3'\mathbf{xi}_i + \boldsymbol{\alpha}_4'\mathbf{c}_i \end{array} \right].$$

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

$$\exp(-\frac{(\widehat{E}_i - E_i^* - \gamma_{s_i} I_i)^2}{2\sigma_j^2}) \cdot (\iota(E_i^o = \widehat{E}_i) \cdot \iota(\widehat{E}_i \geq 0) + \iota(E_i^o = 0) \cdot \iota(\widehat{E}_i < 0)).$$

Thus,

$$\widehat{E}_i|(\boldsymbol{\theta}, \mathbf{E}^*, \widehat{\mathbf{E}}_{-i}, \mathbf{I}^*, \mathbf{C}^m, \mathbf{W}, \mathbf{s}, \mathbf{data}, S_i = 1) \ \sim \ TN_{(-\infty, 0]}(E_i^* + \gamma_{s_i} I_i, \sigma_{s_i}^2)$$

if $E_i = 0$, and $\widehat{E}_i = E_i$ if $E_i > 0$.

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

$$\exp(-(I_i^* - \alpha_0 E_i^* - \alpha_1\sigma_{s_i}^2 - \alpha_2\sigma_{s_i}^2 c_{1i}^m - \boldsymbol{\alpha}_3'\mathbf{xi}_i - \boldsymbol{\alpha}_4'\mathbf{c}_i^m - \frac{\sigma_{12}}{\sigma_{22}}(E_i^* - \boldsymbol{\beta}_{s_i}'\mathbf{xe}_i))^2/2)$$
$$\cdot \quad \exp(-(\mathbf{c}_{1i}^m - XC_i\Lambda)'V_c^{-1}(\mathbf{c}_{1i}^m - XC_i\Lambda)/2)$$

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

$$\begin{pmatrix} \alpha_0 E_i^* + \alpha_1\sigma_j^2 + \alpha_2\sigma_j^2 \cdot \mathbf{xc}_i'\boldsymbol{\lambda}_1 + \boldsymbol{\alpha}_3'\mathbf{xi}_i + \boldsymbol{\alpha}_4' XC_i\Lambda + \frac{\sigma_{12}+\alpha_0\sigma_{22}}{\sigma_{22}}(E_i^* - \boldsymbol{\beta}_{s_i}'\mathbf{xe}_i) \\ XC_i\Lambda \end{pmatrix}$$

and the variance

$$\mathbf{V}_{s_i}^c = \begin{pmatrix} v_{11j}^c & \boldsymbol{\alpha}_4' V_c + \alpha_2\sigma_j^2 \mathrm{v}_x^{1\cdot} \\ V_c\boldsymbol{\alpha}_4 + \alpha_2\sigma_j^2\mathrm{v}_x^{\cdot 1} & V_c \end{pmatrix} \equiv \begin{pmatrix} \mathbf{V}_{s_i11}^c & \mathbf{V}_{s_i12}^c \\ \mathbf{V}_{s_i21}^c & \mathbf{V}_c \end{pmatrix}$$

and

$$v_{11j}^c = 1 + \boldsymbol{\alpha}_4' V_c \boldsymbol{\alpha}_4 + \alpha_2^2\sigma_j^4 \cdot \mathrm{v}_c^{11} + 2\alpha_2\sigma_j^2 \sum_{l=1}^{4} \cdot \alpha_{4l} \cdot \mathrm{v}_c^{1l}.$$

Using the result for multivariate normal distribution the posterior conditional distribution of $\mathbf{c}_i^m$ is given by

$$\mathbf{c}_i^m|\boldsymbol{\theta}, \mathbf{E}^*, \widehat{\mathbf{E}}_{-i}, \mathbf{I}^*, \mathbf{C}_{-i}^m, \mathbf{W}, \mathbf{s}, \mathbf{data}, S_i = 1 \sim N(XC_i\Lambda + \mathbf{V}_{s_i12}^{c'}\mathbf{V}_{s_i11}^{c^{-1}}(I_i^* - \overline{I}_i^c), \mathbf{V}_c - \mathbf{V}_{s_i12}^{c'}\mathbf{V}_{s_i11}^{c^{-1}}\mathbf{V}_{s_i12}^c),$$

where

$$\overline{I}_i^c = \alpha_0 E_i^* + \alpha_1\sigma_j^2 + \alpha_2\sigma_j^2 \cdot \mathbf{xc}_i'\boldsymbol{\lambda}_1 + \boldsymbol{\alpha}_3'\mathbf{xi}_i + \boldsymbol{\alpha}_4' XC_i\Lambda + \frac{\sigma_{12}+\alpha_0\sigma_{22}}{\sigma_{22}}(E_i^* - \boldsymbol{\beta}_{s_i}'\mathbf{xe}_i).$$

11. The conditional posterior density kernel of $\widetilde{\mathbf{W}}_i$ is given by:

$$\exp(-\widetilde{W}_{im}^2/2 - \sum_{j=1}^{m-1}(\widetilde{W}_{ij} - \mathbf{xw}_i'\boldsymbol{\delta}_j)^2/2) \tag{21}$$
$$\cdot \quad \sum_{j=1}^{m}\left(\prod_{l=1}^{m}\iota_{(-\infty,\widetilde{W}_{ji}]}\widetilde{W}_{li}\right) \tag{22}$$
$$\cdot \quad \exp(-(I_i^* - \alpha_0 E_i^* - \alpha_1\sigma_j^2 - \alpha_2\sigma_j^2 C_{1i} - \boldsymbol{\alpha}_3'\mathbf{xi}_i - [C_{1i}, ..., C_{4i}]\boldsymbol{\alpha}_4 - \frac{\sigma_{12}}{\sigma_{22}}(E_i^* - \boldsymbol{\beta}_j'\mathbf{xe}_i))^2/2)$$
$$\cdot \quad \left(\exp(-\frac{(E_i^* - \boldsymbol{\beta}_j'\mathbf{xe}_i)^2}{2\sigma_{22}}) \cdot (\frac{1}{\sqrt{\sigma_j^2}}\exp(-\frac{(\widehat{E}_i - E_i^* - \gamma_j I_i)^2}{2\sigma_j^2}))\right)^{S_i=1} \tag{23}$$

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 (21). The function (22) then determines the candidate type $j^* : \widetilde{W}_{j^*i} \geq \widetilde{W}_{li}$

for all $l = 1, ..., m$. The candidate values are than accepted as new draws $\widetilde{\mathbf{W}}_i^n$ and s_i^n with probability

$$\min\left\{\frac{g_3(j^*)}{g_3(j)}, 1\right\},$$

where

$$\begin{aligned} g_3(k) &= \exp(-(I_i^* - \alpha_0 E_i^* - \alpha_1\sigma_k^2 - \alpha_2\sigma_k^2 C_{1i} - \boldsymbol{\alpha}_3'\mathbf{xi}_i - [C_{1i}, ..., C_{4i}]\boldsymbol{\alpha}_4 - \frac{\sigma_{12}}{\sigma_{22}}(E_i^* - \boldsymbol{\beta}_k'\mathbf{xe}_i))^2/2) \\ &\cdot \left(\exp(-\frac{(E_i^* - \boldsymbol{\beta}_k'\mathbf{xe}_i)^2}{2\sigma_{22}}) \cdot (\frac{1}{\sqrt{\sigma_k^2}}\exp(-\frac{(\widehat{E}_i - E_i^* - \gamma_k I_i)^2}{2\sigma_k^2}))\right)^{S_i=1} \end{aligned} \tag{24}$$

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

This algorithm was checked using the joint distribution tests of Geweke (2004).

6.2 Inclusion of Exogenous Variables in the Equations of the Model

Table A-1: Exogenous variables included in equations for insurance status, expected expenditure, probability of mixture component and in the prediction model for SAS variables.

Variable	Description	I^*	E^*	$\widetilde{W}$	SAS
Female	Indicator for female	Yes	Yes	Yes	Yes
Age	Age, years	Yes	Yes	Yes	Yes
Age^2	Age squared	Yes	Yes	No	Yes
Age^3	Age cubed	Yes	Yes	No	Yes
Married	Indicator for being married	No	Yes	Yes	Yes
Age*Female	Interaction of age polynomial with Female	No	Yes	No	No
Married*Female	Interaction of Married and Female	No	Yes	No	No
Cen: nweng	Indicator for New England Census Division	Yes	Yes	Yes	Yes
Cen: midatl	Indicator for Middle Atlantic Census Division	Yes	Yes	Yes	Yes
Cen: encen	Indicator for East North Central Census Division	Yes	Yes	Yes	Yes
Cen: wncen	Indicator for West North Central Census Division	Yes	Yes	Yes	Yes
Cen: satl	Indicator for South Atlantic Census Division	Yes	Yes	Yes	Yes
Cen: escen	Indicator for East South Central Census Division	Yes	Yes	Yes	Yes
Cen: wscen	Indicator for West South Central Census Division	Yes	Yes	Yes	Yes
Cen: mnt	Indicator for Mountain Census Division	Yes	Yes	Yes	Yes
Cen: pac	Indicator for Pacific Census Division	Yes	Yes	Yes	Yes
Fh2	Health Status Factor	No	Yes	Yes	Yes
Fh3	Health Status Factor	No	Yes	Yes	Yes
Fh7	Health Status Factor	No	Yes	Yes	Yes
Fh8	Health Status Factor	No	Yes	Yes	Yes
Fh10	Health Status Factor	No	Yes	Yes	Yes
Fh11	Health Status Factor	No	Yes	Yes	Yes
Fh17	Health Status Factor	No	Yes	Yes	Yes
Fh20	Health Status Factor	No	Yes	Yes	Yes
Fh22	Health Status Factor	No	Yes	Yes	Yes
Fh23	Health Status Factor	No	Yes	Yes	Yes
Black	Indicator for race black	Yes	Yes	Yes	Yes
Hispanic	Indicator for Hispanic	Yes	Yes	Yes	Yes
Year	Year	No	Yes	Yes	No
hgc: ls8th	Education: less than high school	Yes	No	No	Yes
hgc: somehs	Education: some high school	Yes	No	No	Yes
hgc: hs	Education: high school	Yes	No	No	Yes
hgc: somecol	Education: some college	Yes	No	No	Yes
hgc: college	Education: college	Yes	No	No	Yes
hgc: gradschl	Education: grad. school	Yes	No	No	Yes
hgc: nr	Education non-response	Yes	No	No	Yes
inc 5k-10k	Income: $5-10 thousand	Yes	No	No	Yes
inc 10k-15k	Income: $10-15 thousand	Yes	No	No	Yes
inc 15k-20k	Income: $15-20 thousand	Yes	No	No	Yes
inc 20k-25k	Income: $20-25 thousand	Yes	No	No	Yes
inc 25k-30k	Income: $25-30 thousand	Yes	No	No	Yes
inc 30k-35k	Income: $30-35 thousand	Yes	No	No	Yes
inc 35k-40k	Income: $35-40 thousand	Yes	No	No	Yes
inc 40k-45k	Income: $40-45 thousand	Yes	No	No	Yes
inc 45k-50k	Income: $45-50 thousand	Yes	No	No	Yes
inc 50plus	Income: $50+ thousand	Yes	No	No	Yes
risktol	Risk tolerance	Yes	No	No	No
cogn	Congnition factor	Yes	No	No	No
finpln	Financial planning horizon	Yes	No	No	No
praliv75	Subjective probability to live to be 75 or more	Yes	No	No	No

References

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

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

[3] 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.

[4] 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.

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

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

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

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

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

[10] Deb, Partha, Murat Munkin, and Pravin K. Trivedi. 2006b. Private Insurance, Selection, and Health Care Use: A Bayesian Analysis of a Roy-Type Model. *Journal of Business & Economic Statistics* 24 (October): 1081-1099.

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

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

[13] 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.

[14] Gelman, A, J.B. Carlin, H.S. Stern, and D.B. Rubin. 1995. *Bayesian Data Analysis.* London: Chapman and Hall.

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

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

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

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

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

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

[21] 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.

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

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

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

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

[26] Munkin, Murat, and Pravin K. Trivedi. 2009. Incentives and Selection Effects of Drug Coverage on Total Drug Expenditure: a Finite Mixture Approach *HEDG Working Paper 09/22.*

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

[28] 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.

[29] 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.