Incentives and Selection Effects of Drug Coverage on Total Drug Expenditure: a Finite Mixture Approach.

Murat Munkin

Pravin K. Trivedi

Incentives and Selection Effects of Drug Coverage on Total Drug Expenditure: a Finite Mixture Approach.

Murat K. Munkin
Department of Economics
4202 East Fowler Avenue, BSN 3426
University of South Florida
Tampa FL 33620-5500, U.S.A.
Email: mmunkin@coba.usf.edu

Pravin K. Trivedi
Department of Economics
Wylie Hall
Indiana University
Bloomington, IN 47405, U.S.A
Email: trivedi@indiana.edu

May 31, 2009

Abstract

This paper takes a finte mixture approach to model heterogeneity in incentives and selection effects of drug coverage on the total drug expenditure among the Medicare elderly US population. Evidence is found that the positive drug expenditures of the entire elderly popultion can be decomposed into two groups of relatively healthy with lower average expenditures and relatively unhealthy with higher average expenditures, accounting for approaximately 25% and 75% of the population, respectively. The incentive effects of drug insurance, i.e. ex post moral hazard, are much stronger in magnitude for the unhealthy group. There is also evidence of adverse selection into drug insurance, and this appears to be greater for the higher-expenditure component.

1. Introduction and background

In the United States, Medicare Part A, a public insurance program which covers hospitalization for specified duration, provides insurance coverage for most elderly after age 65. To cover physician services and most outpatient care, voluntary enrollment in Medicare Part B is necessary. The basic Medicare plan did not offer outpatient prescription drug coverage until an optional Part D plan for drug coverage was implemented in January 1, 2006. Prescription drug expenditure is an important component of healthcare spending for the elderly. Kaiser Family Foundation (2005) estimates that in 2005 the total prescription drug expenditure of the elderly was about $120.6 billion or an annual spending of $2,864 per person. The Medicare beneficiaries spent over 36 percent of total outpatient drug expenditure nationwide even though they constituted only 13 percent of the U.S. population (Cook, 1999). But, in 2002, nearly half of the Medicare beneficiaries lacked drug coverage for at least part of the year (Kaiser Family Foundation, 2005). Before 2006, those wanting prescription drug coverage did so using a variety of supplementary insurance plans that included employer-sponsored plans (ESI), Medigap plans and Medicare managed care plans (MMC). With the expansion of the Medicare benefits to include drug coverage, a change that is widely perceived to be a major policy initiative, it is of special interest to evaluate the contribution of drug coverage to the total prescription drug expenditure.

This paper attempts to study the effect of such insurance coverage on the total prescription drug expenditure using a model specification that allows for endogenous unobserved selection into health plans and discrete form of heterogeneity captured using a 2-component finite mixture model. Atherly (2001) indicates that a high proportion of the existing literature on supplemental insurance does not control for self-selection on unobservables. Whereas some studies of Medicare supplemental insurance attempt to control for unobserved selection (Sing et al., 2008; Curtis et al., 2004), others argue for limited or no role for such a selection effect (Lillard et al. (1999); Shea et al (2006)). Treating this as an open research issue, we propose and apply a Bayesian estimation procedure that attempts to decompose the selection effect from the pure incentive effect of insurance. When econometric models of drug expenditure treat the insurance status as exogenous, the

selection and incentive effects are confounded. An additional complication arises from possible heterogeneity in the expenditure behavior of the elderly. It is plausible that such heterogeneity cannot fully be controlled for by the standard observable covariates that are typically included in regression analysis. To allow for unobserved heterogeneity we propose to use a fairly flexible functional form, the latent class model that has been used in a number of previous studies of healthcare utilization, e.g., Deb and Trivedi (1997, 2002), Conway and Deb (2005), and Bago d'Uva (2005, 2006). However, these studies consider models in which the insurance status is exogenous. We extend the model specification to allow for endogenous insurance and then develop a Bayesian estimator for this new model.

The relationship between the Medicare supplemental insurance and the utilization of prescription drugs of the elderly has been investigated intensively in the US literature (e.g. Poisal and Murray, 2001; Piette et al., 2004; Goldman and Philipson, 2007) using both pre- and post 2006 data. A consistent finding is that the elderly with drug coverage spend more on prescription drugs. But the estimates of the incentive effect of insurance have a wide range. Li and Trivedi (2009), henceforth LT(2009), survey these findings and provide their own analysis using 2003 and 2004 data from Medicare Current Beneficiary Survey (MCBS). Using an extended version of the endogenous two-part model of Deb, Munkin and Trivedi (2006), LT (2009) provide a detailed empirical analysis of the effects of seven types of insurance status, controlling for endogenous sample selection. Such disaggregation allows one to study the diversity of responses among different types of insurance plans that vary in the drug benefits that they offer; that is, as plans vary in terms of coinsurance rates, coverage of brand-name versus generic drugs, use of expenditure caps and so forth, disaggregation by plan types is a way of capturing heterogeneity in responses. Overall, drug benefits offered by employer sponsored insurance plans (ESI) are the most generous and those from the Medigap plans the least generous. To reduce biases due to aggregation careful delineation of differences in plans is desirable. The estimated effects of prescription drug insurance may also vary with the approach used to control for self-selection. This consideration motivates LT (2009) to provide a comparative econometric analysis from Bayesian and classical viewpoints. However,

sometimes such disaggregation may not be possible due to lack of data, and at other times it may be avoided (as in this paper) for reasons of computational tractability.

Our approach is related to that of Shang and Goldman (2007) who analyzed the effect of drug coverage in Medigap market and implemented the discrete factor method in the two-part model. They assume that the population can be partitioned into three types with different sets of unobservable factors affecting plan choice and drug expenditure. The discrete factor model introduces the error terms for the plan choice equation, hurdle expenditure equation and conditional expenditure equation that jointly allow for endogeneity as well as unobserved individual heterogeneity. Our Bayesian analysis also allows for correlated unobserved individual heterogeneity, conditional on positive expenditure. However, we avoid the two-part structure and concentrate on conditional model of positive expenditures.

To develop the methodological framework for endogenous selection in a latent class model, we have made some simplifications. First, we will assume that a dichotomous indicator of drug coverage adequately captures the treatment effect of drug insurance in a meaningful way. This involves a strong assumption that combining the incentive effect for those who are only enrolled in the basic Medicare and those enrolled in a supplemental plan without drug coverage is a reasonable approximation that will not affect the identifiability of the key parameters we want to study; however, for a more cautious appraisal of this issue see (Atherly (2002), Sing et al. (2008); Khan et al. (2007)). We assume that our qualitative inferences regarding the aggregate incentive effects of drug insurance and the presence of adverse selection remain valid despite aggregation; see (Goldman et al, 2006). Previous work on expenditure modeling has extensively used the two-part ("hurdle") model which allows the data generating process for positive expenditures to differ from that for zero expenditures; see, for example, Deb et al (2006) and LT (2009). For the data analyzed here the share of positive expenditure is relatively high, around 93%. We will therefore concentrate on modeling the positive expenditures only. While extending our framework to the case of this hurdle model is a desirable objective, its success is likely to depend on having a better division into zeros and positives than in our sample.

4

The remainder of this paper is organized as follows. Section 2 develops the methodology for Bayesian analysis of 2-component latent class model with endogenous dummy variable, under assumption of bivariate normality. Section 3 provides the framework for calculating treatment effects. Section 4 presents our empirical application. Section 5 concludes.

2. The 2-component latent class model

2.1. Heteroskedasticity

Since most of individuals in our sample have positive drug expenditure we will conduct a conditional analysis considering only observations with positive expenditures. The expenditure variable usually displays skewness and it is standard to take its logarithmic retransformation and assume normality. Specify the logarithm of the positive values of Y to be linear in X, a set of exogenous explanatory variables, such that

$$E(\ln Y | Y > 0, X) = X\beta, \tag{2.1}$$

where β is a parameter vector. The main interest is usually centered on estimating $E(Y|X)$ and all associated with that partial effects and elasticities. Given the estimated conditional mean (2.1), one has to transform back from logarithm to variable Y to make inference about $E(Y|X)$.

The dependent variable is assumed to be linear in the set of explanatory variables X and normal error $e \sim N\left(0, \sigma^2\right)$ such as

$$\ln Y = X\beta + e. \tag{2.2}$$

After estimating the parameters of the model the main interest lies in calculating marginal effects of Y with respect to X for the actual expenditure variable. The assumption of lognormality allows us to retransform the dependent variable and consistently estimate $\varphi = E\left(\exp[e]\right)$ as

$$\widehat{\varphi} = \exp\left[0.5\sigma^2\right]. \tag{2.3}$$

Duan (1983) suggested that robustness of this estimator depends critically on whether the positive expenditure is indeed lognormally distributed. He proposed an alternative robust smearing

5

estimator defined as

$$\widehat{\varphi} = \frac{1}{N}\sum_{i=1}^{N} \exp \widehat{e}_i, \tag{2.4}$$

where $\widehat{e}_i$ are the fitted residuals. Mullahy (1998) emphasized that for consistent estimation of parameter β in the conditional mean (2.1) it is sufficient to assume that $E(e|Y > 0, X)$. However, this assumption does not imply statistical independence of e and X. Specifically, Mullahy considered several examples and showed analytically that $E(\exp[e])$ can depend on X and, therefore, the use of the (homoskedastic) smearing estimator will fail to address heteroskedasticity of the multiplying factor $\varphi = \varphi(X)$. This is likely to cause misleading estimates.

2.2. Finite Mixture of Normal Distributions

Assume that we have N independent observations $(i = 1, ..., N)$ and the dependent variable of interest, Y_i, the logarithm of drug expenditure, is continuous. Assume that the population consists of k homogenous subpopulations such that the dependent variable is generated by a k-component finite mixture of normal distributions. One of the modeling goals is to allow for a binary endogenous treatment variable in the conditional means of the components. Model the endogenous binary treatment variable d_i as generated by latent variable D_i that measures the difference in utility derived by individual i in the treated and untreated states respectively,

$$D_i = W_i\alpha + u_i \tag{2.5}$$

such that

$$d_i = I\{D_i \geqslant 0\},$$

where $I\{.\}$ is the indicator function, W_i is a set of exogenous variables and $u_i \overset{iid}{\sim} N[0,1]$. It is assume that W_i contains at least one variable which does not affect drug expenditure.

The main observed variable of interest Y_i, the logarithm of expenditure. Introduce k latent variables Y_{ij} $(j = 1, ..., k)$ distributed $N\left(\mu_{ji}, \sigma_j^2\right)$ and parameterize the means such that

$$Y_{ij} = X_i\beta_j + \rho_j d_i + \delta_j u_i + \varepsilon_{ij}, \ j = 1, ..., k \tag{2.6}$$

6

where X_i is a set of exogenous regressors, $\varepsilon_{ij} \sim N\left(0, \sigma_j^2\right)$. The set of exogenous variables W_i contains at least one variable not included in X_i. We utilize the approach of Geweke, Gowrisankaran, and Town (2003) and introduce random variable u_i directly in to the conditional mean. The way endogeneity of the treatment variable d_i can be modeled is by imposing a potential correlation between the errors in the treatment equation (2.5) and the mean (2.6). However, things are simplified substantially if u_i is introduced directly in the mean equation (2.6). The unobservable variables that affect the choice of the treatment variables also affect the expenditure variable, and the combined error term in the expression (2.6) conditional on d_i is not centered at zero

$$E\left(\delta_j u_i + \varepsilon_{ij} | d_i\right) \neq 0.$$

The variable of interest Y_i is distributed as a finite mixture of normals $Y_{ij} \sim N\left(\mu_{ij}, \sigma_j^2\right)$ with probability p_{ij}. We allow these probabilities to depend on regressors. Finite mixtures is a way of modeling heterogeneous individuals, however, there is a possibility that the probabilities of belonging to a component change with individuals as opposed to being a fixed number across the entire sample. The probabilities must satisfy the following constraint

$$\sum_{j=1}^{k} p_{ij} = 1$$

Introduce latent variable z_{ij} such that

$$z_{ij} = \begin{cases} 1 & \text{if observation } i \text{ belongs to component } j \\ 0 & \text{otherwise} \end{cases}.$$

Then the observability condition is

$$Y_i = Y_{ij} \text{ if and only if } z_{ij} = 1.$$

In order to allow the state probabilities to depend on covariates we follow Geweke and Keane (2007b) and specify the multinomial probit model to determine the state probabilities. Let $k-1$ latent variables be defined as

$$M_{ij} = V_i \gamma_j + \xi_{ij} \ (j = 2, ..., k), \ \xi_i \overset{iid}{\sim} N\left[0, I_{k-1}\right]. \tag{2.7}$$

and let M_{i1} be restricted to zero. Then the components are identified as

$$z_{ij} = 1 \text{ if and only if } M_{ij} \geqslant M_{il} \text{ (for } \forall l,\, l = 1, ..., k). \qquad (2.8)$$

The set of covariates V in general is different from X. The observability condition is

$$Y_i \sim N\left(\mu_{ij}, \sigma_j^2\right) \text{ if and only if } z_{ij} = 1.$$

Denote $z_i = (z_{i1}, z_{i2}, ...z_{ik})'$, $\gamma' = (\gamma_2', ..., \gamma_k')$, $\sigma = \left(\sigma_1^2, ..., \sigma_k^2\right)$, $R_i = (X_i, d_i, u_i)$, $\theta_j = (\beta_j, \rho_j, \delta_j)'$, $\theta = (\theta_1, ..., \theta_k, \gamma, \sigma)$ and $\Delta_i = (X_i, W_i, V_i, \theta)$. The joint density of observed and latent data for individual i is

$$
\begin{aligned}
f(D_i, d_i, Y_{ij}, M_i, z_i, Y_i | \Delta_i) \;=\; & p(D_i | \boldsymbol{\Delta}_i) \\
& \times p(d_i | D_i, \boldsymbol{\Delta}_i) \\
& \times p\left(Y_{ij} | d_i, D_i, \boldsymbol{\Delta}_i\right) \\
& \times p\left(M_i | d_i, D_i, Y_{ij}, \boldsymbol{\Delta}_i\right) \\
& \times p\left(z_i, Y_i | M_i, d_i, D_i, Y_{ij}, \boldsymbol{\Delta}_i\right)
\end{aligned}
$$

or

$$
\begin{aligned}
\;=\; & \frac{\exp\left(-0.5\left(D_i - W_i\alpha\right)^2\right)}{\sqrt{2\pi}} \left[d_i I\{D_i \geqslant 0\} + (1 - d_i)\, I\{D_i < 0\}\right] \\
& \times \prod_{j=1}^{k} \left[\frac{\exp\left(-0.5\sigma_j^{-2}\left(Y_{ij} - X_i\beta_j - \rho_j d_i - \delta_j\left(D_i - W_i\alpha\right)\right)^2\right)}{\sqrt{2\pi\sigma_j^2}}\right] \\
& \times \exp\left[-0.5\sum_{j=1}^{k}\left(M_{ij} - V_i\gamma_j\right)^2\right] \\
& \times \left[\sum_{j=1}^{k} I_{\{z_{ij}=1\}} I_{\{Y_i = Y_{ij}\}}\left(\prod_{l=1}^{k} I_{(-\infty, M_{ij}]}\left(M_{il}\right)\right)\right]
\end{aligned}
$$

The MCMC algorithm of the model requires a Metropolis-Hastings step as follows.

1. Draw M_{ij} $(j = 2, ..., k)$ independently from $N\left[V_i\gamma_j, 1\right]$.

2. Select state $j = j^*$ such that $M_{ij^*} \geqslant M_{ij} \ \forall j = 1, ..., k$.

3. We accept the draw with probability

$$
\min \left\{ \frac{\sigma_{j^*}^{-1} \exp\left(-0.5\sigma_{j^*}^{-2} \left(Y_i - X_i\beta_{j^*} - \rho_{j^*}d_i - \delta_{j^*}\left(D_i - W_i\alpha\right)\right)^2\right)}{\sigma_j^{-1} \exp\left(-0.5\sigma_j^{-2} \left(Y_i - X_i\beta_j - \rho_j d_i - \delta_j\left(D_i - W_i\alpha\right)\right)^2\right)}, 1 \right\},
$$

where j denotes the current state determining the component for observation i, corresponding to $z_{ij} = 1$.

4. The latent vectors D_i $(i = 1, ...N)$ are conditionally independent with normal distribution $Z_i \overset{iid}{\sim} N\left[\overline{D}_i, \overline{H}_i^{-1}\right]$ where

$$
\overline{H}_i = 1 + \sum_{j=1}^{k} \delta_j^2 \sigma_j^{-2},
$$

$$
\overline{D}_i = W_i\alpha + \overline{H}_i^{-1} \left[\sum_{j=1}^{k} \delta_j \sigma_j^{-2} \left(Y_{ij} - R_i\theta_j\right) \right]
$$

Each variable is truncated such that

$$
D_i \geqslant 0 \text{ if } d_i = 1 \text{ and } D_i < 0 \text{ if } d_i = 0.
$$

5. Draw Y_{ij} for each i and j when $z_{ij} = 0$, from the full conditional density $N\left(\mu_{ji}, \sigma_j^2\right)$. If $z_{ij} = 1$ then $Y_{ij} = Y_i$.

6. Let the prior distribution of α be $N\left[\underline{\alpha}, \underline{H}_\alpha^{-1}\right]$. Then the full conditional distribution of α is $\alpha \sim N\left[\overline{\alpha}, \overline{H}_\alpha^{-1}\right]$ where

$$
\overline{H}_\alpha = \underline{H}_\alpha + \sum_{i=1}^{N} W_i' \left(1 + \sum_{j=1}^{k} \delta_j^2 \sigma_j^{-2}\right) W_i
$$

$$
\overline{\alpha} = \overline{H}_\alpha^{-1} [\underline{H}_\alpha \underline{\alpha} + \sum_{i=1}^{N} \{W_i' \left(1 + \sum_{j=1}^{k} \delta_j^2 \sigma_j^{-2}\right) D_i
$$

$$
- W_i' \sum_{j=1}^{k} \delta_j \sigma_j^{-2} \left(Y_{ij} - R_i\theta_j\right)\}].
$$

7. Let the prior distributions be $\gamma_j \sim N\left[\underline{\gamma_j}, \underline{H}_{\gamma_j}^{-1}\right]$, $j = 1, ..., k$. Then the full conditional distribution of γ_j is $\gamma_j \sim N\left[\overline{\gamma_j}, \overline{H}_{\gamma_j}^{-1}\right]$ where

$$\overline{H}_{\gamma_j} = \underline{H}_{\gamma_j} + \sum_{i=1}^{N} R_i' \sigma_j^{-2} R_i$$

$$\overline{\gamma_j} = \overline{H}_{\gamma_j}^{-1}\left[\underline{H}_{\gamma_j}\underline{\gamma_j} + \sum_{i=1}^{N} R_i' \sigma_j^{-2} Y_{ij}\right]$$

8. Assign priors $\sigma_j^{-2} \sim G\left(n_j/2, (g_j/2)^{-1}\right)$, $j = 1, ..., k$. Then the full conditional of σ_j^{-2} is

$$G\left(\frac{n_j + N}{2}, \left[\frac{g_j}{2} + \sum_{i=1}^{N}\frac{(Y_{ij} - R_i\theta_j)^2}{2}\right]^{-1}\right).$$

3. Treatment Effects

Next we calculate the average treatment effect (ATE), the average treatment effect for the treated (ATET) and the local average treatment effect (LATE) for each component in the mixture. Definition of dependent variable Y_i establishes the link between the observed and counterfactual outcomes as

$$Y_i = d_i Y_i^1 + (1 - d_i)Y_i^0.$$

The average treatment effect is defined as the expected outcome gain from receipt of treatment for a randomly chosen individual. One can show that for a randomly chosen individual the ATE for component j is

$$E\left[Y_j^1 - Y_j^0|X\right] = \frac{1}{N}\sum_{i=1}^{N}\exp\left(X_i\beta_j + 0.5\left(\delta_j^2 + \sigma_j^2\right)\right)\left[\exp\left(\rho_j\right) - 1\right]. \tag{3.1}$$

We calculate the ATE parameters averaging with respect to the sample and posterior distributions of the parameters.

The treated group consists of individuals with drug coverage. Denote by $\Phi\left(\cdot\right)$ c.d.f. of the standard normal distribution. The expected outcome gain for those who actually receive the treatment

10

is calculated by the ATET,

$$E\left[Y_j^1 - Y_j^0 | X, W, d = 1\right]$$

as

$$= \frac{1}{N}\sum_{i=1}^{N} \frac{\Phi\left(W\alpha + \delta_j\right)}{\Phi\left(W\alpha\right)} \exp\left(X_i\beta_j + 0.5\left(\delta_j^2 + \sigma_j^2\right)\right)\left[\exp\left(\rho_j\right) - 1\right] \tag{3.2}$$

We will calculate the LATE effect with respect to the premium required to pay for the insurance plan. Denote it by W_k, the instrument component in the set of explanatory variable W in the insurance equation. The instrument should affect the decision to receive treatment when W_k changes from a "lower" value to a "higher" value. Denote W^0 and W^1 two realizations of W in which all the components are equal except for the instrument W_k for which it takes the "lower" and "higher" values respectively. Then the local average treatment effect is defined as the expected outcome gain for those induced to receive treatment at W^1 but who would not have received treatment at W^0,

$$E\left[Y_j^1 - Y_j^0 | X, W^0, W^1, d(W^0) = 0, d(W^1) = 1\right].$$

which is estimated as

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\Phi\left(W^1\alpha + \delta_j\right) - \Phi\left(W^0\alpha + \delta_j\right)}{\Phi\left(W^1\alpha\right) - \Phi\left(W^0\alpha\right)}$$
$$\times \exp\left(X_i\beta_j + 0.5\left(\delta_j^2 + \sigma_j^2\right)\right)\left[\exp\left(\rho_j\right) - 1\right]. \tag{3.3}$$

4. Application to drug expenditure of Medicare beneficiaries

4.1. Data

The data used in this paper are closely related to the 2003-04 MCBS those used in LT (2009), the main difference being that we also add the data for 2005. Further, our analysis only uses positive-valued expenditures. While we provide the essential information for making the current paper self-contained, the reader is referred to that earlier paper for a more complete description of methods and conventions.

Three data sources were linked to generate the data. The Medicare Current Beneficiary Survey (MCBS) is the main source of information regarding demographics, plan enrollment and prescription drug expenditure. The Area Resource File (ARF) and the State County File (SCF) provide extra sources for the instrumental variables.

4.1.1. MCBS data

MCBS, a continuous survey of a representative sample of the Medicare population, contains plan attribute information that can serve as instrumental variables for modeling plan choice. It also contains regional information that can be linked to other datasets such as ARF and SCF that are useful in the analyzing regional Medicare markets. The "Cost and Use" file provides information on Medicare utilization, expenditures, insurance coverage, health status and demographic characteristics. Our sample excludes enrollees with Medicaid or other public plans as well as those who hold more than one health plan or who switch plans during the year. Due to panel structure of the data we have multi-year observations on some individuals. As our analysis is cross sectional we only use one observation per year per individual, the convention being to use the first time the individual is sampled. The final sample contains 7,280 observations from three years, the breakdown being 3,503 from 2003, 1,825 from 2004 and 1,952 from 2005. The MCBS also provides information on supplemental insurance, source of the plan, the premium paid, and coverage of prescription drug expenses. This paper does not use all the information because of its level of aggregation.

MCBS provides information on demographic (age, gender, race, education level, family income, marital status, children), socioeconomic (e.g., income), and extensive list of health status variables (including chronic conditions, disability and activity limitations).

To handle endogeneity of plan choice we need individual-specific or plan-specific instrumental variables that affect plan choice, but are not directly correlated with drug expenditure. Premium is a natural instrument for insurance choice, as it is essentially a sunk cost after purchase of a plan and hence should not affect health care utilization. For additional detail, see LT (2009).

4.1.2. ARF and SCF data

Several county level variables can serve as individual specific instruments for plan choice. We use the county managed medicare (MMC) penetration rate in the preceding year as a measure of the market power of MMC when the elderly make the plan choice decision and this lagged variable should not affect the current year's drug expenditure. The information on the number of MMC enrollees and the number of eligible Medicare beneficiaries in each county in each year is provided by the SCF; this information is used to calculate the MMC penetration rate. For additional details see LT (2009).

4.1.3. Sample Summary Statistics

The summary statistics for enrollees by insurance status are provided in Table 1. The logarithm of total drug expenditure, LNAAMTTOT, is the dependent variable and drug coverage (COVRX) is the binary endogenous treatment variable. Figure 1 plots the dependent variable and it is clear that its distribution is non-central. The vector of covariates X in the both components of the expenditure variable is the same. It consists of the drug coverage dummy, self-perceived health status variables VEGOOD, GOOD, FAIR and POOR (excellent health status is the excluded category), indicators of present and past smoking habits, SMOKNOW and SMOKEVER, indicators for heart and other health conditions, HEARTCOND and OTHCOND, geographical location variables NOREAST, WEST, SOUTH and MSA, variables that proxy for socioeconomic status: AGE, WHITE, MALE, MARRY, WIDOW, LVALONE, DEGRCV, number of children living, CHNLNM, logarithm of personal income, LNINCOME, employment status, JOBSTAT, and year dummies YEAR2004, YEAR2005. Vector W of the insurance equation does not include COVRX, includes all X variables, plus the identifying instrumental variables PENET and MOAMT.

One needs a valid instrumental variable to identify the treatment effects and key endogeneity parameters. We use two such variables, PENET and MOAMT. Variable MOAMT is defined at the monthly premium amount paid for the insurance coverage. It is a plan attribute and should not affect the expenditure variable. For computational simplicity we have aggregated different types of

13

drug coverage plans. For example, some may cover both generics and brand-name drugs whereas others only cover generics. Of course, the monthly premiums on average would be higher for the more generous coverage. Using paid premium as one of the instruments helps to control for heterogeneity in the treatment variable. The decision to purchase drug insurance also depends upon access to plans. It is known that such access varies geographically. The ARF and SCF data base provide additional sources of information that potentially expands the set of instruments. Several county level variables can serve as individual specific instruments. For example, since the county MMC penetration rate in the preceding year can measure the market power of MMC that affects the plan choice but not the current year's drug expenditure. Combining information on the number of MMC enrollees and the number of eligible Medicare beneficiaries in each county, the MMC penetration rate can be generated; see LT (2009) for additional detail.

Finally, the set of covariates in the probability equation V includes indicators for heart and other chronic health conditions, HEARTCOND and OTHCOND, respectively, and logarithm of personal income, LNINCOME. If it is true that the two components are differentiated based on health conditions and ability to pay for the prescribed medicine, then these variables should have a strong effect on the probabilities.

4.2. Application

In general a finite mixture model with any number of components can be used to fit the data. However, in our application we concentrate on a two-component model. Our main specification is consistent with health economics literature which finds strong empirical support for two subpopulations of "higher" and "lower" level users. These correspond to relatively healthy and unhealthy individuals. Finite mixture models have some distinct features that make the estimation process non-trivial. For example, classical estimation requires carefully chosen starting values. The objective function is potentially multimodal which could lead the maximizer away from the global maximum. That makes the estimation somewhat unstable. Bayesian estimation, on the other hand, has its own problems with the Gibbs sampler because the objective function is invariant with re-

spect to all possible permutations of the components. In other words, it is not clear to allocate a draw from the Gibbs sampler with respect to the components and prevent label switching. This problem of estimation is especially serious when the number of components is misspecified (see Fruhwirth-Schnatter (2001), Geweke (2007a)). The Markov chains have bad mixing properties such that the entire support of the posterior distribution is not visited. For example, Celeux, Hurn and Robert (2000) report difficulties with the Gibbs sampler.

When the estimated function of interest is component invariant Fruhwirth-Schnatter (2001) shows that a random permutation sampler is able to solve the convergence problem and the produced Markov chains have very good properties. However, when the estimated function is component specific and the number of components is correctly identified Geweke (2007a) shows that finite mixture models display robust properties provided estimation is based on large samples and there exist justifiable constrains that allows one to separate the components at each draw of the Markov chains. In our application, there is a number of constraints that could be tested to produce practically identical results. These include $\rho_1 > \rho_2$, $\delta_1 > \delta_2$ and $\sigma_1^2 > \sigma_2^2$, and each serves as a good identifying restriction. The incentives and selection effects of COVRX are expected to be different between the "higher" and "lower" expenditure groups, which justifies the use of the constraints. We run our MCMC algorithm for 20,000 replications after discarding first 1000 replications of the burn-in phase. The chains show very good mixing properties with the autocorrelation function of the parameters dying after at most 10 lags. The results are presented in Table 2.

It is interesting that all three variables, HEARTCOND, OTHCOND and LNINCOME, have a negative and very strong impact on the probability to belong to a component. Our interpretation is that the higher mean components consists of relatively less healthy people with higher levels of income. The average probability of belonging to the higher expenditure subpopulation, taken over the entire sample, is estimated to be 0.75.

With age people in our sample are less likely to have drug coverage. On the other hand, whites, married, widowed and living in a metropolitan statistical area are more likely to have it. HEART-COND positively affect the coverage probability, but OTHCOND has no effect. Consistent with

15

expectation education and income increase drug coverage. Surprisingly smoking variables do not affect drug coverage, being employed decreases the probability of it.

We are able to identify two components. The first component has a smaller mean (5.98) than the second one (7.40), but the main difference between the components is in the variance parameter, effect of COVRX and the covariance parameter. We interpret these components as healthy and unhealthy. Males spend less on drug in both groups, while whites spend more. Marital status has no effect on the amounts spent, smokers spend more. Expenditure goes up with education and as health status indicators move from good health to poor health. This is consistent for both groups. HEARTCOND has a strong positive effect on expenditure but other chronic conditions increase expenditure only for the unhealthy group. Employment status has a negative effect on the amounts spent for both groups. There is no measurable impact of income on drugs for the healthy group and it is strong and positive for the unhealthy one. One qualification of the preceding description is that the mean of the lower-expenditure group is slightly overestimated as our analysis is for the population with positive expenditures. The sample estimate of the proportion with zero expenditure is around 7-8%.

We estimate completely different variance parameters for the healthy and unhealthy groups, 1.753 (0.103) and 0.497. This indicates that the healthy group is much more heterogeneous with a greater spread of expenditure values, relative to the unhealthy group. The covariance parameters δ_1, δ_2, 0.096 (0.090) and 0.089 (0.037) show that the selection effects are strong in the unhealthy group only, which is consistent with expectations of adverse selection into the drug coverage group. Finally the impact of drug coverage is strong and significant for both group, 0.341 (0.169) and 0.226 (0.066). However, in order to understand the magnitudes of the effects one has to calculate the treatment effect. These results are consistent with positive moral hazard or incentive effect of drug insurance. This result is consistent with many estimates available in the literature.

Our exclusion restrictions are strongly correlated with the drug coverage variable. The posterior means and standard deviations of PENET and MOAMT are 0.020 (0.002) and -0.001 (0.0002) respectively. These instruments drive the probability of treatment from 0.012 to 0.968.

16

4.3. Average Treatment Effects

The estimated treatment parameters are given in Table 3. As can be seen from the results the treatment effects are of very different magnitudes for the healthy and unhealthy components. The ATEs are \$397 and \$502, respectively, which is about 100 dollars higher for the second component. The ATET is interpreted as the average gain in the dependent variable from the treatment for the subpopulation which actually selected into the treatment. The ATET is higher than ATE by \$15 and \$24, respectively, for the healthy and unhealthy groups. Therefore, those who actually self-select into purchasing additional drug coverage derive a greater incentive effect to utilize the insurance. That is, evidence of adverse selection, consistent across both subpopulations, is suggested by the observed difference between ATE and ATET estimates.

An alternative estimate of adverse selection in the two components is generated by comparing estimates of ATE from the model that allows for endogenous selection and from that which does not. Denote by $\text{ATE}_j^{\text{End}}$, $j = 1, 2$, the estimate from the model with endogeneity, and by ATE_j^{Ex} the corresponding estimate from the model without endogeneity. Then the presence of adverse selection implies that

$$\text{ATE}_j^{\text{End}} < \text{ATE}_j^{\text{Ex}},$$

where as under favorable (or advantageous) selection the direction of the inequality is reversed. For consistency the second calculation is also based on Bayesian methodology implemented under the restriction that $\delta_1 = \delta_2 = 0$ which means that the insurance and expenditure decisions are independent. The model allowing for endogeneity of treatment may be interpreted as one which decomposes the raw estimate of the treatment effect into pure incentive and selection components. Under exogeneity this decomposition is not identified. The estimated average treatment effects under the exogeneity assumption are ATE_1^{Ex}=576 (4.6) and ATE_2^{Ex}=811 (3.6) for the healthy and unhealthy components respectively. This is additional evidence in favor of adverse selection.

Finally we calculate the LATE parameter using additional information not presented in the summary table. We know the information on the premiums paid by the individuals for the insurance plans that they have. The average premiums can be calculated for those with and without drug

17

coverage for employment sponsored insurance, Medigap and Medicare managed care plans. These averages are used as W_k^0 and W_k^1 in calculating the LATE parameters. Three different sets of LATE are calculated for the three different plans: ESI, Medigap and MMC. The local average treatment effect is calculated as the expected drug expenditure increase for those induced to receive drug coverage at W_k^1 value of the premium amount, which would not have been received at the W^0 level. The results indicate that such gains are about $390 for healthy component and a much higher $1880 for the unhealthy group.

5. Conclusion

Unobservable self-selection into drug coverage has been detected and controlled for in many empirical analyses. Based on simulated equilibrium premiums for drug coverage, Pauly and Zeng (2006) found the evidence of adverse selection into drug coverage. Khan et al. (2007) used fixed effects to control for self-selection and concluded that cross-sectional models for the effect of drug benefit on drug utilization are subject to substantial endogeneity bias. Our findings indicate that there is heterogeneity in incentives and selection effects among the elderly. In general the entire elderly population can be viewed as represented by two groups of relatively healthy and unhealthy individuals. The incentives effects are much stronger in magnitude for the unhealthy group which account for about 75 percent of the elderly population. We are able to identify a significant selection effect only for the second component which corresponds to the higher mean expenditure. The sign of the selection effect indicates adverse selection of relatively less healthy individuals into the higher expenditure group.

References

Atherly, A. 2001. Supplemental insurance: Medicare's accidental stepchild. *Medical Care Research and Review* 58(2):131-161.

Atherly, A., 2002. The effect of Medicare supplemental insurance on Medicare expenditures. *Journal International Journal of Health Care Finance and Economics*. 2(2): 1389-6563.

Bago d'Uva. T. 2005. Latent class models for use of primary care: evidence from a British panel. *Health Economics.* 14: 873–892

Bago d'Uva, T. 2006. Latent class models for use of health care. *Health Economics.* 15: 329–343.

Celeux, G., M. Hurn and C.P. Robert, 2000, Computational and Inferential Difficulties with Mixture Posterior Distributions. Journal of American Statistical Association 95, 957-970.

Conway, K. and P. Deb, (2005), Is Prenatal Care Really Ineffective? Or, is the 'Devil' in the Distribution?, *Journal of Health Economics*, 24, 489-513.

Cook, A.E. 1999. Strategies for containing drug costs: implications for a Medicare benefit. *Health Care Financing Review* 20(3):29-37.

Deb, P., Munkin M.K. and P.K. Trivedi. 2006. Bayesian Analysis of the Two-Part Model with Endogeneity: Application to Health Care Expenditure. *Journal of Applied Econometrics*, 21, 1081-1099.

Duan, N. 1983. Smearing estimate: a Nonparametric Retransformation Method, *Journal of the American Statistical Association,* 78, 605–610.

Duan, N., Manning, W.G., et al. 1983. A Comparison of Alternative Models for the Demand for Medical Care , *Journal of Business and Economics Statistics,* 1, 115–126.

Fruhwirth-Schnatter, S., 2001, Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models. Journal of American Statistical Association 96, 194-209.

Geweke, J., Gowrisankaran G. and R.J. Town 2003. Bayesian Inference for Hospital Quality in a Selection Model. *Econometrica*, 71(4), 1215-38.

Geweke, J., 2004. Getting it Right: Joint Distribution Tests of Posterior Simulators. Journal of the American Statistical Association 99, 799-804.

Geweke, J., 2007a, Interpretation and Inference in Mixture Models: Simple MCMC Works. Computational Statistics & Data Analysis 51, 3529-3550.

Geweke, J. and M. Keane, 2007b, Smoothly mixing regressions. Journal of Econometrics 138, 252-291.

Goldman, D.P., S.D. Hosek, L.S. Dixon, E.M. Sloss 1995. The Effects of Benefit Design and Managed Care on Health Care Costs", *Journal of Health Economics*, 14, 401-18.

Goldman, D.; Joyce, G.; Karaca-Mandic, P.; Sood, N. 2006. Adverse selection in retiree prescription drug plans. *Forum for Health Economics & Policy* 9(1):1062-1062.

Goldman, D.; Philipson, T. 2007. Integrated insurance design in the presence of multiple medical technologies. *American Economic Review* 97(2):427-432.

Jones, A., 2000. Health Econometrics. In: A. Culyer, J. Newhouse (Eds.), *Handbook of Health Economics*, Elsevier, Amsterdam.

Kaiser Family Foundation. 2005. Medicare chart book 2005.

Keane, M.P. 1992. A Note on Identification in the Multinomial Probit Model. *Journal of Business and Economic Statistics*, 10, 193-200

Khan, N.; Kaestner, R.; Lin, S.J. 2007. Prescription drug insurance and its effect on utilization and health of the elderly. NBER Working Paper No. 12848.

Li, Q.; Trivedi, P.K. 2009. Impact of Prescription Drug Coverage on Drug Expenditure of the Elderly - Evidence from a two-Part Model with Endogeneity. Working paper.

Lillard, L.A.; Rogowski, J.; Kington, R. 1999. Insurance coverage for prescription drugs: effects on use and expenditures in the Medicare population. *Medical Care* 37: 926-36.

Manning, W.G. and J. Mullahy 2001. Estimating Log Models: To Transform or Not to Transform? *Journal of Health Economics*, 20(4), 461-94.

Mullahy, J. 1998. Much Ado about Two: Reconsidering Retransformation and the Two-part Model in Health Econometrics. *Journal of Health Economics* 17, 247-281.

Poisal, J.A.; Murray, L.A. 2001. Growing differences between Medicare beneficiaries with and without drug coverage. *Health Affairs* 20(2):75-85.

Poisal, J.A.; Murray, L.A.; Chulis, G.S.; Cooper, B.S. 1999. Prescription drug coverage and spending for Medicare beneficiaries. *Health Care Financing Review* 20(3):15-25.

Shang, B; Goldman, D.P. 2007. Prescription drug coverage and elderly medicare spending. NBER Working Paper Series No. 13358.

Shea, D.G.; Terza, J.V.; , Stuart, B.C.; Briesacher, B. 2007. Estimating the effects of prescription drug coverage for Medicare beneficiaries. *Health Services Research* 42 (3):933-949.

Sing M; Miller E.; Banthin J.S. 2008. Estimated effects of increased coverage on prescription drug expenditures among seniors. Agency for Healthcare Research and Quality Working Paper No. 8005.

Table 1: Variable definition and summary statistics.

		COVRX=0	COVRX=1	
Number of observations		2821	4459	7280
Expenditure				mean (st. dev.)
AAMTTOT	total drug expenditure	1503 (1540)	2295 (2448)	1988 (2177)
LNAAMTTOT	logarithm of drug expend.	6.75 (1.28)	7.20 (1.23)	7.03 (1.27)
Insurance plan types				
COVRX	= 1 if cover prescribed drug	0	1	0.61 (0.49)
Demographic characteristics				
AGE	age	77.7 (7.5)	76.75 (7.1)	77.1 (7.3)
MALE	=1 if male	0.46 (0.49)	0.43 (0.49)	0.45 (0.49)
WHITE	=1 if white	0.89 (0.31)	0.91 (0.28)	0.91 (0.29)
HEARTCOND	=1 if one of heart cond.	0.79 (0.41)	0.79 (0.40)	0.79 (0.40)
OTHCOND	=1 if other chronic cond.	0.60 (0.49)	0.62 (0.48)	0.61 (0.49)
MARRY	if married	0.51 (0.50)	0.59 (0.49)	0.56 (0.49)
WIDOW	if widowed	0.39 (0.49)	0.34 (0.47)	0.36 (0.48)
LVALONE	if lives alone	0.37 (0.48)	0.30 (0.45)	0.33 (0.47)
CHNLNM	number of children living	3.05 (2.11)	3.02 (1.91)	3.03 (1.99)
MSA	=1 if metro area	0.61 (0.49)	0.77 (0.42)	0.71 (0.46)
NOREAST	=1 if northeast	0.12 (0.32)	0.16 (0.36)	0.15 (0.35)
WEST	= 1 if west	0.13 (0.33)	0.22 (0.42)	0.19 (0.39)
SOUTH	=1 if south	0.45 (0.49)	0.37 (0.48)	0.40 (0.49)
DEGRCV	highest grade completed	4.40 (1.96)	4.92 (2.01)	4.72 (2.01)
JOBSTAT	if working now	0.11 (0.31)	0.10 (0.30)	0.10 (0.31)
LNINCOME	logarithm of income	9.82 (1.03)	10.11 (0.95)	9.99 (0.99)
Health characteristics				
VEGOOD	= 1 if very good health	0.29 (0.45)	0.31 (0.46)	0.30 (0.46)
GOOD	= 1 if good health	0.36 (0.48)	0.34 (0.47)	0.35 (0.48)
FAIR	= 1 if fair health	0.16 (0.36)	0.15 (0.36)	0.16 (0.36)
POOR	= 1 if poor health	0.05 (0.22)	0.04 (0.21)	0.05 (0.21)
SMOKNOW	=1 if smoke now	0.10 (0.30)	0.09 (0.28)	0.09 (0.29)
SMOKEVER	=1 if ever smoked	0.59 (0.49)	0.59 (0.49)	0.59 (0.49)
Year dummies				
YEAR2004	= 1 if year 2004	0.29 (0.45)	0.23 (0.42)	0.25 (0.43)
YEAR2005	= 1 if year 2005	0.22 (0.41)	0.29 (0.45)	0.27 (0.44)
Instruments				
PENET	mmc penet. rate	7.43 (10.85)	13.93 (14.69)	11.4 (13.7)
MOAMT	monthly premium per person	88.7 (81.9)	77.9 (99.3)	82.1 (93.1)

Table 2. Posterior means and standard deviations.

Equation	COVRX	Comp 1	Comp 2		COVRX	Comp 1	Comp 2
CONST	-1.073	3.876	6.119	GOOD	0.000	0.544	0.324
	0.261	1.059	0.236		0.050	0.118	0.039
AGE	-0.013	0.004	-0.006	FAIR	0.071	1.189	0.522
	0.002	0.006	0.002		0.058	0.151	0.052
MALE	-0.203	-0.422	-0.083	POOR	0.010	1.271	0.742
	0.037	0.097	0.029		0.085	0.228	0.066
WHITE	0.135	0.285	0.159	SMOKNOW	-0.054	-0.354	-0.102
	0.055	0.139	0.041		0.057	0.141	0.041
HEARTCOND	0.095	0.744	0.234	SMOKEVER	0.042	0.173	0.024
	0.039	0.118	0.045		0.036	0.090	0.026
OTHCOND	0.056	0.194	0.257	DEGRCV	0.050	0.055	0.014
	0.033	0.113	0.028		0.008	0.023	0.006
MARRY	0.368	0.086	0.015	JOBSTAT	-0.138	-0.412	-0.165
	0.065	0.180	0.051		0.053	0.130	0.041
WIDOW	0.286	0.086	0.118	LNINCOME	0.136	-0.002	0.074
	0.060	0.158	0.050		0.017	0.098	0.019
LVALONE	-0.044	0.027	-0.053	YEAR2004	-0.091	0.018	0.086
	0.047	0.118	0.035		0.038	0.100	0.028
CHNLNM	-0.004	-0.006	0.006	YEAR2005	0.240	0.237	0.408
	0.008	0.021	0.006		0.038	0.101	0.027
MSA	0.205	-0.021	-0.063	PENET	0.020		
	0.039	0.094	0.027		0.002		
NOREAST	0.131	-0.228	0.043	MOAMT	-0.001		
	0.052	0.134	0.038		0.0002		
WEST	0.054	-0.474	-0.154	COVRX		0.341	0.226
	0.055	0.130	0.039			0.169	0.066
SOUTH	-0.025	0.083	0.033	δ_1, δ_2		0.096	0.089
	0.038	0.106	0.029			0.090	0.037
VEGOOD	0.012	0.373	0.066	σ_1, σ_2		1.753	0.497
	0.050	0.119	0.039			0.103	0.077

Table 3. Treatment effects.

$\left(W_k^1, W_k^0\right)$	ATE	ATET	LATE 1 $\overline{(91, 67)}$	LATE 2 $(176, 137)$	LATE 3 $(52, 48)$
Obs.	7280	4459	7280	7280	7280
Component 1					
	397	412	387	389	387
	3.1	4.2	3.1	3.1	3.1
Component 2					
	502	526	1882	1894	1878
	2.3	3.0	7.7	7.8	7.7

Figure 1

LNAAMTTOT