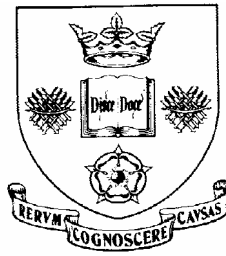# Sheffield Economic Research Paper Series
SERP Number: 2003005

Markus Jochmann* and Roberto Leon-Gonzalez**

Estimating the Demand for Health Care with Panel Data:

A Semiparametric Bayesian Approach

* Department of Economics, University of Konstanz

** Department of Economics, University of Sheffield

October, 2003

Estimating the Demand for Health Care with Panel Data:

A Semiparametric Bayesian Approach[*]

Markus Jochmann

Department of Economics

University of Konstanz

Roberto Leon-Gonzalez

Department of Economics

University of Sheffield

October 24, 2003

Abstract

This paper is concerned with the problem of estimating the demand for health care with panel data. A random effects model is specified in a semiparametric Bayesian fashion using a Dirichlet process prior. This results in a very flexible mixture distribution with an infinite number of components for the random effects. Therefore, the model can be seen as a natural extension of prevailing latent class models. A full Bayesian analysis using Markov chain Monte Carlo (MCMC) simulation methods is discussed. The methodology is illustrated with an application using data from Germany.

JEL classifications: C14, C23, I10

Keywords: random effects model, Dirichlet process prior, MCMC

1 Introduction

This paper is concerned with the problem of estimating the demand for health care. It advances on previous cross sectional studies by explicitly incorporating unobserved heterogeneity using a random effects panel data model (see Lopéz-Nicolás [1] and Riphahn et al. [2] for other studies using panel data to infer the demand for health care). This approach allows us to control for different behavioral attitudes or genetic diversity across individuals, which are both very likely to influence the demand for health care. One aspect of our analysis is to develop a semiparametric framework that avoids the arbitrary specification of a particular distribution for the random effects.

Another purpose of this paper is to expand the range of the recently advocated latent class models (e.g., Deb and Trivedi [3] or Jiménez-Martín et al. [4]) by allowing the population to be split into more than a small number of classes. An argument in favour of latent class models is that they allow for a heterogeneous population while avoiding the sharp distinction between "users" and "non-users" which is assumed in two-part hurdle models (see, for example, Pohlmeier and Ulrich [5] or Gurmu [6]). Deb and Trivedi [7] use data from the RAND Health Insurance Experiment (RHIE) and find that latent class models outperform two-part models in terms of in-sample and cross-validation model selection tests.

However, latent class models only allow for a small number of classes in practice. Moreover, the problem of selecting the number of classes is not straightforward, especially with small sample sizes. In the literature about health care demand, it is common to estimate models with just two classes representing the 'ill' and the 'healthy' (e.g., Deb and Trivedi [3], [7]). This assumption may be too restrictive in some circumstances.

Our model overcomes this fact by specifying a Dirichlet process prior (Ferguson [8] and Antoniak [9]) for the distribution of the random effects. The resulting mixture distribution of the random effects has a random number of components, and hence it is very flexible while remaining tractable.

We apply the proposed model to analyse equity in the delivery of health care using 5 waves from the *German Socio-Economic Panel Study* (GSOEP). In particular, we focus on the analysis of horizontal equity. The delivery of health care will be equitable in a horizontal sense if individuals with equal need, in terms of health status, are given the same treatment irrespective of their income and other socio-economic characteristics. For that purpose, we analyse the importance of income and socio-econonomic characteristics in explaining health care utilisation while controlling for health status.

The plan of the paper is as follows. Section 2 introduces a parametric random effects count data model which assumes a multivariate Normal distribution for the random effects. This model will serve as a benchmark throughout the paper. In Section 3 we present a semi-parametric extension of the model which allows for a wide range of distributions for the random effects. Section 4 describes the numerical procedures (Markov chain Monte Carlo techniques) that we use to obtain the model estimates. In Section 5, we describe the data and the results of the empirical analysis will be presented in Section 6. Section 7 draws some conclusions and provides an outlook on future research.

2 A Parametric Benchmark Model

In this section, we describe a parametric Bayesian model for panel count data, which sets the benchmark for the semiparametric extension discussed later (Chib and Winkelmann [10] analyse a similar model using Bayesian inference, Zeger and Karim [11] propose a Bayesian approach to generalized linear models). We assume that the observed count outcomes y_{it} for individual $i = 1, \ldots, N$ over time periods $t = 1, \ldots, T_i$ follow a Poisson distribution, that is,

$$y_{it}|\theta_{it} \sim \text{Poisson}(\exp(\theta_{it})). \tag{1}$$

The logarithm of the conditional mean θ_{it} is defined as

$$\theta_{it} = x_{it}'\beta + w_{it}'b_i + \varepsilon_{it}, \tag{2}$$

3

where x_{it} is a $k \times 1$ vector of covariates, β is the corresponding parameter vector, w_{it} is a $p \times 1$ vector of covariates for the corresponding vector of unobserved random effects b_i and ε_{it} is an error term. We assume that b_i and ε_{it} are independent and that each random effects vector b_i follows a p-dimensional multivariate Normal distribution with mean zero and variance-covariance matrix D:

$$b_i \sim \mathrm{N}_p(0, D). \tag{3}$$

The error term ε_{it} is drawn from a Normal distribution with mean 0 and precision parameter τ,

$$\varepsilon_{it} \sim \mathrm{N}(0, \tau^{-1}). \tag{4}$$

The model is completed by specifying the following priors for β, D and τ:

$$\beta \sim \mathrm{N}_k(\mu_0, \Sigma_0), \tag{5}$$

$$D^{-1} \sim \mathrm{Wishart}(\nu_0, S_0), \tag{6}$$

$$\tau \sim \mathrm{Gamma}\left(\frac{\alpha_0}{2}, \frac{\alpha_0}{2}\right). \tag{7}$$

3 A Semiparametric Extension

It has been shown both from the classical and the Bayesian perspectives that in many situations the assumption of a particular functional form for the random effects is too restrictive and may lead to wrong parameter estimates (see for example Heckman and Singer [12] who make this point for duration models or Verbeke and Lesaffre [13] in the context of linear mixed-effects models). For this reason, we now propose a Dirichlet process mixture (DPM) model in the spirit of Ibrahim and Kleinman [14] that generalises the parametric benchmark model of the previous section. In particular, we remove the parametric normal prior assigned to the random effects $\{b_i\}$ and replace it with a general distribution G:

$$b_i \sim G. \tag{8}$$

The prior distribution on G is then defined to be a Dirichlet process with concentration parameter M and base distribution G_0:

$$G \sim \mathcal{DP}(M \cdot G_0). \tag{9}$$

The base distribution G_0 is specified as a p-dimensional multivariate Normal distribution:

$$G_0 = \mathrm{N}_p(0, D). \tag{10}$$

We therefore add a further stage to the model that allows us to take into account possible deviations of the true distribution of the random effects G from the "baseline" multivariate normal distribution G_0. In other words, we approximate the true nonparametric shape of G by the base distribution G_0. The concentration parameter M reflects our prior belief about how similar G is to G_0. Large values of M lead to a G that is very likely to be close to G_0. Small values of M allow G to deviate more from G_0 and put most of its probability mass on just a few atoms.

Figure 1 illustrates this point by plotting several draws from the Dirichlet process with two different values for M (1.25 and 10). In order to sample from the prior of G we utilise a truncated version of the sum-representation of the Dirichlet process proposed by Sethuraman [15]. Note that each draw represents a probability density function. Figure 1 illustrates that each probability function is almost surely discrete (Sethuraman [15]). When $M = 1.25$ the number of mass points with non-negligible probability is smaller than when $M = 10$. As M increases, the probability mass will be more evenly distributed on a bigger set of mass points, and it would resemble more closely the continuous density G_0.

Looking at two key features of the Dirichlet process helps to clarify the implications of this setup. First, some of the b_i are identical with positive probability. Thus, each b_i takes one of $l < N$ distinct values which we denote by $\kappa = (\kappa_1, \ldots, \kappa_l)$. A so called cluster then contains all random effects which take the same value. In order to discuss the second fact, some additional no-

5

tation is necessary. Let b^{-i} denote the random effects excluding the random effect for individual i. Finally, let the set κ^{-i} consist of the distinct elements of b^{-i} with each value κ_j^{-i} appearing m_j^{-i} times. Now we can show that by integrating over G the prior distribution of b_i conditional on b^{-i} and G_0 can be expressed as:

$$b_i|b^{-i}, G_0 \sim \frac{M}{M+N-1}G_0 + \frac{1}{M+N-1}\sum_{j=1}^{l} m_j^{-i}\delta(\kappa_j^{-i}), \qquad (11)$$

where $\delta(\kappa)$ represents a degenerate distribution with point mass at κ. Therefore, a new value drawn from the base distribution is chosen for b_i with probability $M/(M+N-1)$, whereas b_i takes the value of an already existing cluster κ_j^{-i} with probability $m_j^{-i}/(M+N-1)$.

Combining this result with equations (2) and (4), we obtain the following expression for the conditional distribution of θ_{it} marginalized over b_i and G:

$$\theta_{it}|\beta, D, G_0, b^{-i} \sim \int f_N(\theta_{it}|x'_{it}\beta + w'_{it}b_i, \tau^{-1})d[b_i|b^{-i}, G_0]. \qquad (12)$$

Performing the integration we end up with:

$$\begin{aligned}
\theta_{it}|\beta, D, G_0, b^{-i} \sim & \frac{M}{M+N-1}f_N(\theta_{it}|x'_{it}\beta, w'_{it}Dw_{it} + \tau^{-1}) \\
& + \frac{1}{M+N-1}\sum_{j=1}^{l} m_j^{-i}f_N(\theta_{it}|x'_{it}\beta + w'_{it}\kappa_j^{-i}, \tau^{-1}),
\end{aligned} \qquad (13)$$

where f_N represents the normal density. We see that θ_{it} follows a mixture distribution with a random number of components, where the components differ both with respect to their means and variances. Equation (13) illustrates that the here proposed DPM model can be seen as a mixture model with an infinite number of classes (see Neal [16] for a more formal presentation of this point). Thus, it contributes to the existing literature on latent class models for estimating the demand for health care. Also note that by using the Dirichlet process as a prior on the distribution of the random effects, we are able to relax the restrictive parametric assumption inherent in the benchmark model

6

in a tractable manner.

4   Bayesian MCMC Sampling

Having specified the prior distribution and the likelihood function, we now turn to the analysis of the posterior distribution, which is proportional to the product of these two terms. In the Bayesian approach, the posterior distribution of a model contains all the relevant information and can be used to make probability statements about the parameters.

However, due to the complexity of the proposed models, we are not able to analyse their posterior distributions analytically. This problem can be overcome by applying Markov chain Monte Carlo (MCMC) techniques. This means that we draw large samples from the posterior distributions and then use these samples to summarise the posterior distributions. We do this by employing the Gibbs sampler where each element of the parameter vectors is updated conditional on the actual values of the other components. After discarding some number of initial draws, the resulting Markov chains have converged to the posterior distributions. We refer to Chen et al. [17] or Robert and Casella [18] for comprehensive surveys on MCMC methods.

In order to keep the Gibbs sampler computations simple, we apply the data augmentation technique put forward by Tanner and Wong [19]. This means that we include the random effects $\{b_i\}$ and the latent variables $\{\theta_{it}\}$ in the parameter space. Thus, we end up with full conditional distributions which take convenient functional forms.

The resulting Gibbs sampler for the parametric benchmark model can be summarised as follows (further details on the algorithm are given in the appendix of this paper):

0. Choose starting values for τ, $\{b_i\}$, D^{-1}, $\{\theta_{it}\}$.

1. Sample β from $[\beta|\{b_i\}, \tau, \{\theta_{it}\}]$, which is a Normal distribution.

2. Sample τ from $[\tau|\{b_i\}, \beta, \{\theta_{it}\}]$, which is a Gamma distribution.

3. Sample $\{\theta_{it}\}$ from $[\theta_{it}|b_i, \beta, \tau]$, using the Metropolis-Hastings algorithm, independently for $i = 1, \ldots, N$ and $t = 1, \ldots, T_i$.

4. Sample $\{b_i\}$ from $[b_i|\beta, \tau, D, \{\theta_{it}\}]$, which is a Normal distribution, independently for $i = 1, \ldots, N$.

5. Sample D^{-1} from $[D^{-1}|\{b_i\}]$, which is a Wishart distribution.

6. Repeat Steps 1-5 using the updated values of the conditioning variables.

Since G_0 is chosen to be a conjugate prior distribution (a conjugate prior distribution yields a posterior distribution that falls in the same class of distributions), we can easily set up a Gibbs sampler for the semiparametric model as well. Examples of MCMC methods applied to the semiparametric setting are Escobar and West [20] or MacEachern and Müller [21]. In particular, we have to modify steps 4 and 5 as follows (further details are also given in the appendix):

4'a. Sample $\{b_i\}$ from $[b_i|b^{-i}, G_0, D, \beta, \tau, \{\theta_{it}\}]$, independently for $i = 1, \ldots, N$.

5'. Sample D^{-1} from $[D^{-1}|\{\kappa_j\}]$, which is a Wishart distribution.

In order to improve the mixing behaviour of the modified algorithm, we follow a strategy proposed by Bush and MacEachern [22] and resample the cluster values $\{\kappa_j\}$ after determining how the b_is are grouped. This is achieved by including the following step:

4'b. Sample $\{\kappa_j\}$ from $[\kappa_j|\beta, \tau, D, \{\theta_{it}\}]$, which is a Normal distribution, independently for $j = 1, \ldots, l$.

We would like to point out that the Bayesian approach and its application via MCMC techniques offer several advantages. First, the Bayesian approach allows for full and exact small sample inference both in the parametric and the semiparametric version of the model and is not restricted to asymptotic approximations. Second, numerical integration methods are avoided in the

evaluation of the model. Finally, by using data augmentation we easily obtain estimates for the random effects. This becomes important when analysing extensions of the model in which the estimates of the random effects play a central role on their own (see Kleinmann and Ibrahim [14] and the cited literature therein). For example, one might think of a possible extension of the model in the direction of causal effect modelling. In this case, MCMC methods would allow us to calculate individual treatment effects (see Chib and Hamilton [23]).

5   The Data

In the following, the proposed methodology is used to estimate the demand for health care by the elderly in Germany. There are many existing studies analysing the demand for health care, but only few of them focus on the elderly population (Deb and Trivedi [3] is one exception). Nevertheless, this group is of particular interest, since elderly people typically have higher medical care needs and costs and their population share is steadily growing in many countries.

The data set used in this study stems from five waves (1997-2001) of the *German Socio-Economic Panel Study* (GSOEP). The GSOEP, conducted by the German Institute for Economic Research in Berlin, is a representative longitudinal survey of German households (for more information, see SOEP Group [24]). It contains detailed information about the health care utilisation of the respondents and insurance schemes under which they are covered.

We restrict our analysis to retired men who are older than 65 years. After eliminating all observations with missing values on any of the variables of interest, we obtain a final sample of 1854 individuals and 4761 person-year observations. Note that the observations are not equally distributed throughout the five years, since both in 1998 and 2000 the GSOEP was expanded with new sub-samples. The variable definitions and summary statistics are reported in Table 1.

The dependent variable in our study is the number of visits to a doctor in the last three months prior to the survey (VISITS). Note that visits to a dentist are subsumed under this definition as well. The explanatory variables consist of socioeconomic characteristics and variables that describe the health condition of the individual. In particular, we include a self-perceived health satisfaction index (SATISFAC), as well as variables measuring disability (HANDICAP and HDEGREE). In order to capture nonlinear and threshold effects of SATISFAC we include the dummy variables LOWS and HIGHS.

In the German health care system, only individuals above a certain earnings level (3,825 Euros gross monthly earnings in 2003), civil servants, or self-employed individuals can opt out the public insurance scheme (PUBLICIN) and choose a private insurance plan or remain uninsured. Individuals in the public insurance scheme can purchase add-on insurance (ADDON) that, for example, covers extra costs for dental prostheses or glasses.

Given this institutional setup, the decisions to choose a private insurance plan and to purchase add-on insurance may be endogenous. However, since we control for the health condition of the individual, the strength of this argument is reduced (see Deb and Trivedi [3], who argue in the same line). The possibility of endogeneity should nevertheless not be overlooked when interpreting the results.

6   Results

We analyse these data using both the parametric benchmark model and the semiparametric extension of it. Prior elicitation is done in the following way: we randomly choose 250 individuals from the data set and analyse this " training sample " using the parametric benchmark model with uninformative priors. In this way we mimic the usual Bayesian approach where the results of a previous study with different data are used to select prior distributions (Chib and Hamilton [23] and Ibrahim and Kleinman [14] also follow the 'training sample'

strategy).

To analyse the remaining data, we select a prior distribution on D^{-1} by setting $\nu_0 = 250$ and $S_0 = \frac{\hat{D}^{-1}}{\nu_0}$, where $\hat{D}$ is the training sample posterior mean. Cowles et al. [25] argue that a flatter prior on the variance matrix of the random effects can lead to a slow convergence of the algorithm (see also Ibrahim and Kleinman [14]). In addition, the prior means and variances of the slope parameters in β are the corresponding estimates obtained with the training sample. In order to facilitate the calculation of Bayes factors (Verdinelli and Wasserman [26]), the non diagonal elements in Σ_0 are set equal to zero. Finally, in order to represent prior ignorance, we set $\alpha_0 = 0.001$.

We then estimate the parametric benchmark model and the semiparametric model with M equal to 10. Recall that a Dirichlet process prior implies that we expect the density of the individual effects to be discrete (we showed several draws from the prior on the distribution of the random constant in Figure 1). Given our choice of M, S_0 and ν_0, the number of mass points with probability larger than 0.01 is between 2 and 9 with probability 0.95 (we calculate this "a priori" credible interval by Monte Carlo simulation).

We specify the models choosing VISITS as the dependent variable. All other variables (including the year dummies) plus a constant are included in the population mean vectors. The random effects include a constant and the effects of SATISFAC, HIGHS and LOWHS. The models are then estimated using the MCMC sampling algorithms described in Section 4. We ran each for 30,000 iterations keeping the last 25,000 iterations each time. To give an indication of the performance of the algorithm for the semiparametric model, Figure 2 reports the posterior histograms and autocorrelation functions of β_{AGE}, τ, and D_C, where D_C is the variance of the intercept in the base measure ($D_{SATISFAC}$, D_{HIGHS} and D_{LOWS} denote the variances of the SATISFAC, HIGHS and LOWS effects, respectively). It can be seen that the mixing behaviour of the sampler is satisfactory since autocorrelations decline steadily as the number of lags increases. The algorithm for the parametric model displays

11

an even better mixing behaviour.

Table 2 shows the posterior estimates for the parametric and semiparametric model. We observe that the point estimates of the semiparametric model tend to be associated with larger standard deviations and, for some parameters, they are substantially different to the parametric counterparts. This is illustrated in Figure 2, which compares the posterior density of $\beta_{SATISFAC}$ in both models. Note that the semiparametric point estimate receives very small density weight in the parametric model and that there is more uncertainty in the estimates when the parametric assumptions are relaxed.

The estimated coefficients on AGE and AGE2 imply that the number of doctor visits increases with age until the age of 85 and decreases thereafter. There is a large probability that the effect of NOPARTNER is negative, but positive values cannot be ruled out. Similarly, there is some uncertainty regarding the sign of the effect of education. The evidence on the effect of disability is twofold: the sign of the dummy variable (HANDICAP) is not clearly determined, whereas the degree of handicap (HDEGREE) has an unambiguously positive effect. An increase of 10 percentage points would lead to 0.2 visits more on average. The variable SATISFAC has as expected a negative effect, whereas the signs of the threshold effects (LOWS and HIGHS) are uncertain.

Note that the variance of the time variant error term ε_{it} is small when compared with the variance of the individual effects. Thus, individual heterogeneity accounts for a large proportion of the variability in the data, which illustrates the importance of modelling the distribution of the individual effects correctly.

There is substantial uncertainty regarding the signs of the coefficients of the variables FOREIGN, ADDON, PUBLICIN and PENSION. Riphahn et al. [2] argue that the result for ADDON is not surprising and can be explained by the benefit packages of the German add-on insurance plans. In order to determine whether the delivery of health care for the elderly is equitable, we test the hypothesis that the variables EDUCATION, FOREIGN, ADDON, PUBLICIN

and PENSION have all a zero effect. We calculate a Bayes factor for this hypothesis following the method proposed by Verdinelli and Wasserman [26]. We obtain that the hypothesis of equitable delivery of health care is much more likely than the alternative (the probability of this hypothesis versus the alternative is 0.9993). Note, however, that the model does not account for the possible endogeneous nature of the variable PUBLICIN. An extension in the direction of causal modelling using the potential outcomes approach is one direction for future research.

7 Conclusion

This paper developed a semiparametric Bayesian framework for estimating the demand for health care with panel data. This was done by specifying a Dirichlet process prior for the distribution of the random effects. Thus, the presented framework allowed explicitly for individual heterogeneity while it did not impose unreasonably strong constraints on distributional assumptions.

It was shown that the model can be seen as a natural extension of latent class models, which abound in the recent literature on health care demand. This results from the fact that the Dirichlet process prior leads to a mixing distribution with an infinite number of components.

The model was used to test for the existence of horizontal equity using German data. The estimation was carried out with MCMC methods. The results were largely in accordance with the previous literature.

The approach presented here can be extended in many directions, including the development of a potential outcomes model for inferring causal effects, or a model that allows for the endogenous nature of private insurance.

Appendix

The Algorithm for the Parametric Model

1. Sampling β from $[\beta|\{b_i\}, \tau, \{\theta_{it}\}]$:

$$p(\beta|\{b_i\}, \tau, \{\theta_{it}\}) \propto |\Sigma_0|^{-\frac{1}{2}} \exp\left(\frac{-1}{2}(\beta - \mu_0)'\Sigma_0^{-1}(\beta - \mu_0)\right)$$
$$\times \exp\left(\frac{-\tau}{2}\sum_{i=1}^{n}\sum_{t=1}^{T_i}(\theta_{it} - x_{it}'\beta - w_{it}'b_i)^2\right), \tag{14}$$

so that

$$[\beta|\{b_i\}, \tau, \{\theta_{it}\}] \sim N_k(\mu_\beta, \Sigma_\beta) \tag{15}$$

with

$$\Sigma_\beta = \left(\Sigma_0^{-1} + \tau\sum_{i=1}^{n}\sum_{t=1}^{T_i} x_{it}x_{it}'\right)^{-1} \tag{16}$$

and

$$\mu_\beta = \Sigma_\beta\left(\Sigma_0^{-1}\mu_0 + \tau\sum_{i=1}^{n}\sum_{t=1}^{T_i} x_{it}(\theta_{it} - w_{it}'b_i)\right). \tag{17}$$

2. Sampling τ from $[\tau|\{b_i\}, \beta, \{\theta_{it}\}]$:

$$p(\tau|\{b_i\}, \beta, \{\theta_{it}\}) \propto \tau^{\frac{\alpha_0}{2}-1} \exp\left(\frac{-\alpha_0\tau}{2}\right)\tau^{\frac{n}{2}}\exp\left(\frac{-\tau}{2}\sum_{i=1}^{n}\sum_{t=1}^{T_i}\varepsilon_{it}^2\right), \tag{18}$$

so that

$$[\tau|\{b_i\}, \beta, \{\theta_{it}\}] \sim \text{Gamma}\left(\frac{\alpha_0 + n}{2}, \frac{\alpha_0 + \sum_{i=1}^{n}\sum_{t=1}^{T_i}\varepsilon_{it}^2}{2}\right). \tag{19}$$

3. Sampling $\{\theta_{it}\}$ from $[\theta_{it}|\{b_i\}, \beta, \tau]$:

$$p(\theta_{it}|\{b_i\}, \beta, \tau) \propto \exp\Bigg(-\exp(\theta_{it}) + y_{it}\theta_{it}$$
$$-\frac{\tau(\theta_{it} - x_{it}'\beta - w_{it}'b_i)^2}{2}\Bigg). \tag{20}$$

4. Sampling $\{b_i\}$ from $[b_i|\beta, \tau, D, \{\theta_{it}\}]$:

$$p(b_i|\beta, \tau, D, \{\theta_{it}\}) \propto \exp\left(\frac{-1}{2}b_i'D^{-1}b_i\right)$$
$$\times \exp\left(\frac{-1}{2}\sum_{t=1}^{T_i}(\theta_{it} - x_{it}'\beta - w_{it}'b_i)^2\right), \tag{21}$$

so that

$$[b_i|\beta, \tau, D, \{\theta_{it}\}] \sim \mathrm{N}_p(\mu_b, \Sigma_b) \tag{22}$$

with

$$\Sigma_b = \left(\tau\sum_{t=1}^{T_i}w_{it}w_{it}' + D^{-1}\right)^{-1} \tag{23}$$

and

$$\mu_b = \Sigma_b\tau\sum_{t=1}^{T_i}w_{it}(\theta_{it} - x_{it}'\beta). \tag{24}$$

5. Sampling D^{-1} from $[D^{-1}|\{b_i\}]$:

$$p(D^{-1}|\{b_i\}) \propto |D^{-1}|^{\frac{\nu_0 - p - 1}{2}} \exp\left(\frac{-1}{2}\mathrm{tr}(S_0^{-1}D^{-1})\right)$$
$$\times |D^{-1}|^{\frac{n}{2}} \exp\left(\frac{-1}{2}\sum_{i=1}^{n}b_i'D^{-1}b_i\right), \tag{25}$$

so that

$$[D^{-1}|\{b_i\}] \sim \mathrm{Wishart}\left(\nu_0 + n, \left(S_0^{-1} + \sum_{i=1}^{n}b_ib_i'\right)^{-1}\right). \tag{26}$$

The Algorithm for the Semiparametric Model

4'a. Sampling $\{b_i\}$ from $[b_i|b^{-i}, G_0, D, \beta, \tau, \{\theta_{it}\}]$:

Sample $b_i|b^{-i}, G_0, D, \beta, \tau, \{\theta_{it}\}$ from the distribution

$$q_{i0}\pi_0(b_i|\beta, \tau, D, \{\theta_{it}\}) + \sum_j q_{ij}\delta(\kappa_j^{-i}), \tag{27}$$

where π_0 denotes the density of the p-variate Normal distribution:

$$\pi_0(b_i|\beta, \tau, D, \{\theta_{it}\}) = f_N(\mu_b, \Sigma_b), \tag{28}$$

where μ_b and Σ_b are defined above. The weights sum up to 1 and are given by

$$q_{i0} \propto M|\Sigma_b|^{\frac{1}{2}}|D|^{-\frac{1}{2}}\exp\left(\frac{\tau}{2}(\theta_i - X_i\beta)'U_i(\theta_i - X_i\beta)\right) \tag{29}$$

and

$$q_{ij} \propto m_j^{-i}\exp\left(-\frac{\tau}{2}(\theta_i - X_i\beta - W_i\kappa_j^{-i})'(\theta_i - X_i\beta - W_i\kappa_j^{-i})\right), \tag{30}$$

where $X_i \equiv (x_{i1}, x_{i2}, \ldots, x_{iT_i})'$, $W_i \equiv (w_{i1}, w_{i2}, \ldots, w_{iT_i})'$,
$\theta_i \equiv (\theta_{i1}, \theta_{i2}, \ldots, \theta_{iT_i})'$ and $U_i \equiv (\tau W_i\Sigma_b W_i' - I)$.

4'b. Sampling $\{\kappa_j\}$ from $[\kappa_j|\beta, \tau, D, \{\theta_{it}\}]$:

$$p(\kappa_j|\beta, \tau, D, \{\theta_{it}\}) \propto \exp\left(\frac{-1}{2}\kappa_j'D^{-1}\kappa_j\right)$$
$$\times \exp\left(\frac{-1}{2}\sum_{i\in j}\sum_{t=1}^{T_i}(\theta_{it} - x_{it}'\beta - w_{it}'\kappa_j)^2\right), \tag{31}$$

so that

$$[\kappa_j|\beta, \tau, D, \{\theta_{it}\}] \sim N_p(\mu_\kappa, \Sigma_\kappa) \tag{32}$$

with

$$\Sigma_\kappa = \left(\tau\sum_{i\in j}\sum_{t=1}^{T_i}w_{it}w_{it}' + D^{-1}\right)^{-1} \tag{33}$$

16

and

$$\mu_\kappa = \Sigma_\kappa \tau \sum_{i \in j} \sum_{t=1}^{T_i} w_{it}(\theta_{it} - x'_{it}\beta). \tag{34}$$

5'. Sampling D^{-1} from $[D^{-1}|\{\kappa_j\}]$:

$$\begin{aligned}
p(D^{-1}|\{\kappa_j\}) \propto & |D^{-1}|^{\frac{\nu_0 - p - 1}{2}} \exp\left(\frac{-1}{2}\text{tr}(S_0^{-1}D^{-1})\right) \\
& \times |D^{-1}|^{\frac{l}{2}} \exp\left(\frac{-1}{2}\sum_{j=1}^{l} \kappa'_j D^{-1} \kappa_j\right),
\end{aligned} \tag{35}$$

so that

$$[D^{-1}|\{\kappa_j\}] \sim \text{Wishart}\left(\nu_0 + l, \left(S_0^{-1} + \sum_{j=1}^{l} \kappa_j \kappa'_j\right)^{-1}\right). \tag{36}$$

17

References

[1] López-Nicolás A. Unobserved heterogeneity and censoring in the demand for private health care. *Health Economics* 1998; **7**:429–437.

[2] Riphahn RT, Wambach A, Million A. Incentive effects in the demand for health care: A bivariate panel count data estimation. *Journal of Applied Econometrics* 2003; **18**:387–405.

[3] Deb P, Trivedi P. Demand for medical care by the elderly: A finite mixture approach. *Journal of Applied Econometrics* 1997; **12**:313–336.

[4] Jimenez-Martin S, Labeaga J, Martinez-Granado M. Latent class versus two-part models in the demand for physician services across the European Union. *Health Economics* 2002; **11**:301–321.

[5] Pohlmeier W, Ulrich V. An econometric model of the two-part decision-making process in the demand for health care. *The Journal of Human Resources* 1995; **30**:339–361.

[6] Gurmu S. Semiparametric estimation of hurdle regression models with an application to Medicaid utilization. *Journal of Applied Econometrics* 1997; **12**:225–242.

[7] Deb P, Trivedi P. The structure of demand for health care: Latent class versus two-part model. *Journal of Health Economics* 2002; **21**:601–625.

[8] Ferguson TS. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* 1973; **1**:209–230.

[9] Antoniak CE. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics* 1974; **2**:1152–1174.

[10] Chib S, Winkelmann R. Markov chain Monte Carlo analysis of correlated count data. *Journal of Business and Economic Statistics* 2001; **19**:428–435.

[11] Zeger SL, Karim MR. Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association* 1991; **86**:79–86.

[12] Heckman JJ, Singer B. A method for minimizing the impact of distributional assumptions in econometric models of duration. *Econometrica* 1984; **52**:271–320.

[13] Verbeke G, Lesaffre E. A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association* 1996; **91**:217–221.

[14] Ibrahim JG, Kleinman KP. Semiparametric Bayesian infenrence for random effect models. In Dey D, Müller P, Sinha D, eds., *Practical Nonparametric and Semiparametric Bayesian Statistics*, Springer, New York. 1998; .

[15] Sethuraman J. A constructive definition of Dirichlet priors. *Statistica Sinica* 1994; **4**:639–650.

[16] Neal RM. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* 2000; **9**:249–265.

[17] Chen MH, Shao QM, Ibrahim JG. *Monte Carlo Methods in Bayesian Computation*. Springer, New York, 2000.

[18] Robert CP, Casella G. *Monte Carlo Statistical Methods*. Springer, New York, 1999.

[19] Tanner MA, Wong W. The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association* 1987; **82**:528–550.

[20] Escobar MD, West M. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* 1995; **90**:577–588.

[21] MacEachern SN, Müller P. Estimating mixtures of Dirichlet process models. *Journal of Computational and Graphical Statistics* 1998; **7**:223–238.

[22] Bush CA, MacEachern SN. A semi-parametric Bayesian model for randomized block designs. *Biometrika* 1996; **83**:275–285.

[23] Chib S, Hamilton BH. Semiparametric Bayes analysis of longitudinal data treatment models. *Journal of Econometrics* 2002; **110**:67–89.

[24] SOEP Group. The German Socio-Economic Panel (GSOEP) after more than 15 years - overview. In Holst E, Lillard DR, DiPrete TA, eds., *Proceedings of the 2000 Fourth International Conference of German Socio-Economic Panel Study Users (GSOEP2000)*. volume 70 of *Vierteljahrshefte zur Wirtschaftsforschung*, 2001; 7–14.

[25] Cowles MK, Carlin BP, Connett JE. Bayesian Tobit modeling of longitudinal ordinal clinical compliance data with nonignorable missingness. *Journal of the American Statistical Association* 1996; **91**:86–98.

[26] Verdinelli I, Wasserman L. Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *Journal of the American Statistical Association* 1995; **90**:614–618.

Variable	Definition	Mean	Std. Dev.
VISITS	Number of doctor visits in last 3 months	4.120	5.534
AGE	Age in years	72.371	6.041
AGE2	Age squared in years / 1000	5.274	0.913
EDUCATION	Years of education	11.300	2.306
SATISFAC	Self reported health satisfaction	5.667	2.323
	(0-low to 10-high)		
LOWS	1 if SATISFAC < 4	0.187	
HIGHS	1 if SATISFAC > 6	0.400	
HANDICAP	1 if individual is handicapped	0.337	
HDEGREE	Degree of handicap in percentage points	21.800	33.102
NOPARTNER	1 if individual has no partner	0.145	
PENSION	Monthly pension payments in DM / 1000	2.639	1.295
PUBLICIN	1 if individual is in public health insurance	0.920	
ADDON	1 if individual purchased add-on insurance	0.055	
FOREIGN	1 if individual is foreigner	0.056	
YEAR97	1 if year = 1997	0.118	
YEAR98	1 if year = 1998	0.138	
YEAR99	1 if year = 1999	0.153	
YEAR00	1 if year = 2000	0.298	
YEAR01	1 if year = 2001	0.293	
$N = 1854$			
$\sum T_i = 4761$			

Table 1: Variable definitions and summary statistics

Variable	Quantiles					
	2.5%	50%	97.5%	2.5%	50%	97.5%
		$M = \infty$			$M = 10$	
AGE	0.151	0.576	0.990	0.163	0.571	0.991
AGE2	-6.094	-3.343	-0.530	-6.140	-3.347	-0.648
EDUCATION	-0.006	0.066	0.140	-0.009	0.057	0.123
SATISFAC	-0.584	-0.460	-0.338	-0.708	-0.563	-0.423
LOWS	-0.354	0.108	0.571	-0.601	-0.073	0.440
HIGHS	-0.845	-0.424	0.002	-0.899	-0.379	0.126
HANDICAP	-0.690	-0.002	0.676	-0.709	-0.045	0.617
HDEGREE	0.010	0.019	0.029	0.011	0.020	0.030
NOPARTNER	-0.945	-0.493	-0.046	-0.783	-0.342	0.085
PENSION	-0.198	-0.057	0.086	-0.163	-0.034	0.092
PUBLICIN	-0.526	0.054	0.651	-0.383	0.209	0.767
ADDON	-0.442	0.124	0.699	-0.444	0.099	0.621
FOREIGN	-0.314	0.408	1.106	-0.434	0.240	0.890
τ	4.760	5.478	6.245	4.809	5.560	6.387
D_C	0.177	0.212	0.256	0.261	0.560	1.195
$D_{SATISFAC}$	0.016	0.018	0.021	0.018	0.027	0.045
D_{LOWS}	0.105	0.123	0.146	0.035	0.069	0.168
D_{HIGHS}	0.130	0.153	0.183	0.055	0.119	0.292

Note: We report marginal effects for the coefficient vector.

Table 2: Posterior estimates for the parametric benchmark model ($M = \infty$) and the MDP model with $M = 10$
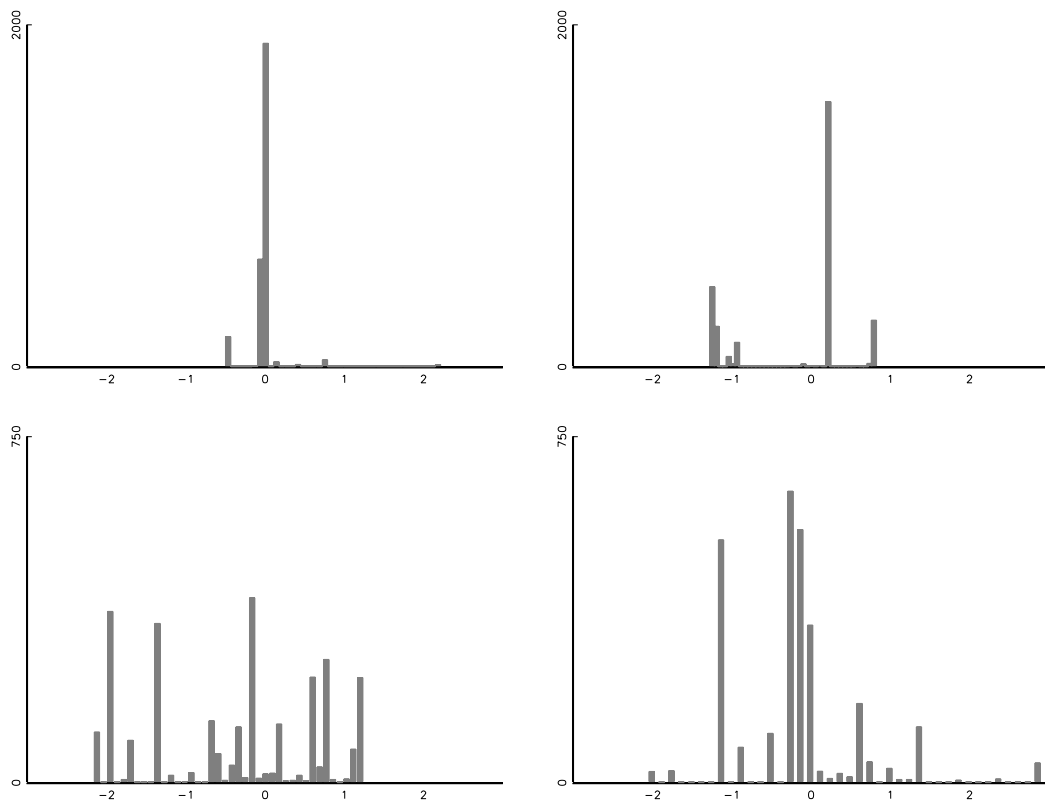
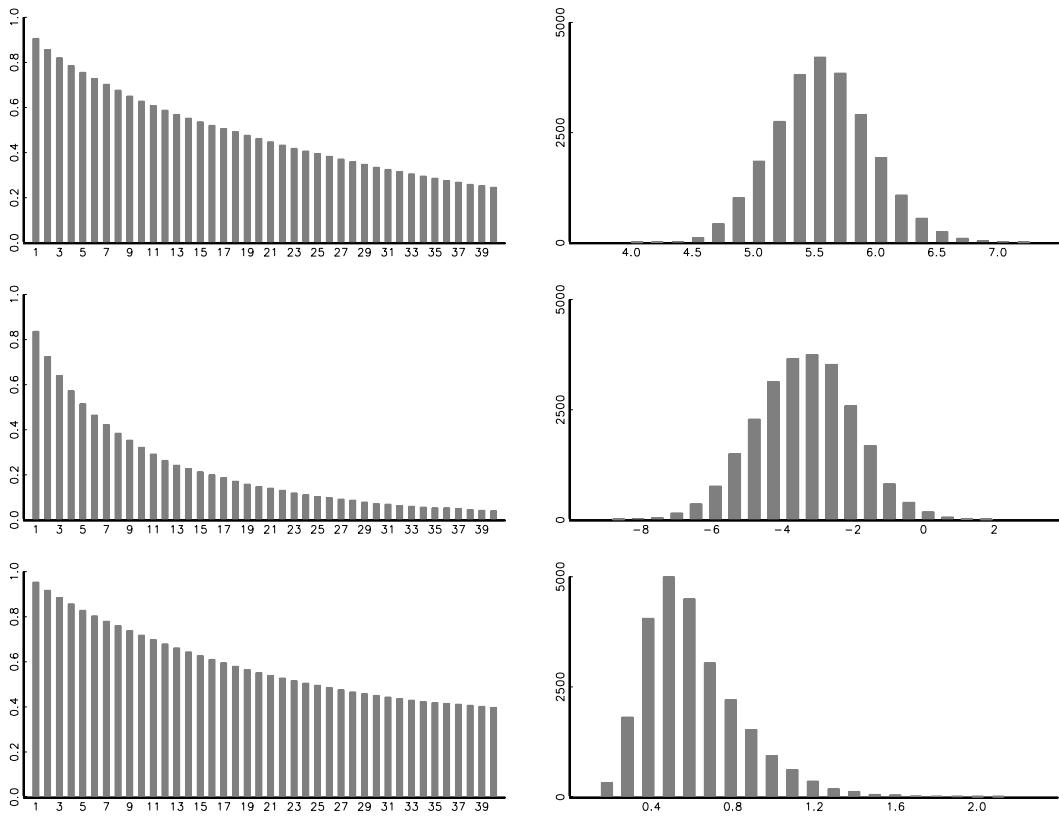Figure 1: Draws from the Prior with $M = 1.25$ (top row) and $M = 10$ (bottom row)

Figure 2: Autocorrelation functions and posterior histograms for τ (top row), the marginal effect of AGE2 (middle row) and D_C (bottom row)
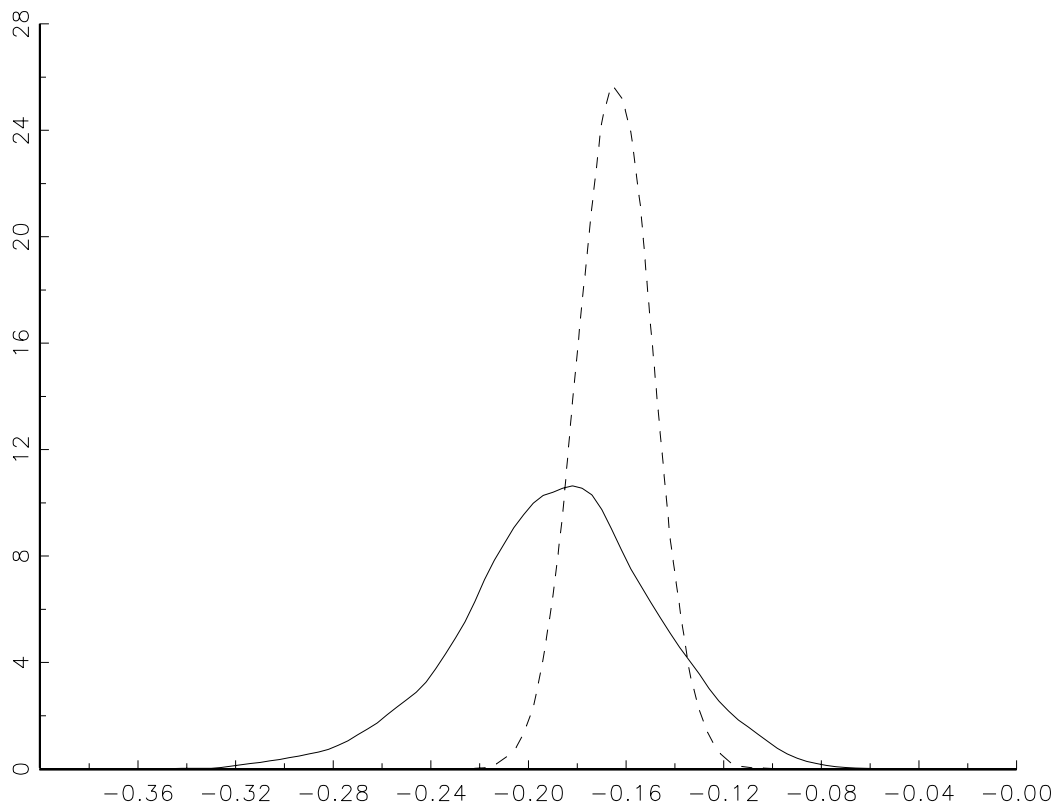
Figure 3: Posterior distributions of β_{SATISFAC}: Parametric benchmark model (dashed curve) and MDP model with $M = 10$ (solid curve)