Bayesian Procedures as a Numerical Tool for the Estimation of Dynamic Discrete Choice Models

Peter Haan
Daniel Kemptner
Arne Uhlendorff

DISCUSSION PAPER SERIES

I Z A

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Bayesian Procedures as a Numerical Tool for the Estimation of Dynamic Discrete Choice Models

Peter Haan
DIW Berlin

Daniel Kemptner
DIW Berlin

Arne Uhlendorff
University of Mannheim,
DIW Berlin and IZA

ABSTRACT

Bayesian Procedures as a Numerical Tool for the Estimation of Dynamic Discrete Choice Models[*]

Dynamic discrete choice models usually require a general specification of unobserved heterogeneity. In this paper, we apply Bayesian procedures as a numerical tool for the estimation of a female labor supply model based on a sample size which is typical for common household panels. We provide two important results for the practitioner: First, for a specification with a multivariate normal distribution for the unobserved heterogeneity, the Bayesian MCMC estimator yields almost identical results as a classical Maximum Simulated Likelihood (MSL) estimator. Second, we show that when imposing distributional assumptions which are consistent with economic theory, e.g. log-normally distributed consumption preferences, the Bayesian method performs well and provides reasonable estimates, while the MSL estimator does not converge. These results indicate that Bayesian procedures can be a beneficial tool for the estimation of dynamic discrete choice models.

JEL Classification: C11, C25, J22

Keywords: Bayesian estimation, dynamic discrete choice models, intertemporal labor supply behavior

Corresponding author:

Daniel Kemptner
DIW Berlin
Mohrenstraße 58
10108 Berlin
Germany
E-mail: dkemptner@diw.de

1 Introduction

Accounting for unobserved heterogeneity is important when estimating non-linear models. Numerous studies document that discrete choice models without unobserved heterogeneity require either very strong and often implausible assumptions or lead to biased estimates of central parameters. In particular, this is true for dynamic models which analyze the role of state dependence in the behavior of agents because in these models it is necessary to disentangle true state dependence from individual specific effects, see for example Heckman (1981a, 1981b).[1] For a recent study in labor economics see e.g. Prowse (2010), for an example from marketing see Dube, Hitsch, and Rossi (2010). The structure of unobserved heterogeneity in discrete choice models can be complex and therefore it is often necessary to allow for a general specification with potential correlations of the different processes. Estimation of such models often is computationally expensive and numerical problems with classical estimation procedures may arise (e.g. non-convergence or convergence at a local maximum of the optimization algorithm). These problems are particularly severe when unobserved heterogeneity is assumed to follow a less convenient distribution - such as a log-normal distribution - rather than a multivariate normal distribution which is the standard assumption. Train (2009) points out that in the classical approach locating the maximum of the likelihood is considerably more difficult with log-normal distributions. And, even if a maximum is found, it may happen that the Hessian is singular at this point.

However, economic theory clearly indicates that in many applications the assumption of a normal distribution of unobserved effects is not appropriate since this potentially leads to counterintuitive effects, e.g. negative preferences for income of consumption or positive price effects of normal goods. Bayesian procedures, such as Markov Chain Monte Carlo (MCMC) methods serve as an important alternative for the estimation of non-linear models with unobserved heterogeneity. Since the Bayesian MCMC estimator does not involve maximization of a likelihood function, the numerical problems of classical procedures such as maximum simulated likelihood (MSL) do not arise.[2]

---

[1]The same applies for hazard rate models which try to separate duration dependence from dynamic selection due to unobserved characteristics, see e.g. Lancaster (1990) and van den Berg (2001) for overviews.

[2]In the context of static mixed logit models Regier et al. (2009) demonstrate that MSL estimation

Consistency and efficiency can be achieved under more relaxed conditions.[3]

The use of MCMC methods in the analysis of dynamic discrete choice models has become more popular in recent years. Fitzenberger, Osikominu, and Paul (2010) and Troske and Voicu (2010) are two examples in the context of dynamic labor supply models, whereby Fitzenberger, Osikominu, and Paul (2010) focus on the impact of training incidence and duration on employment outcomes and Troske and Voicu (2010) analyze interdependencies between fertility choices and labor supply behavior of females. Imai, Jain, and Ching (2009) propose a method applying MCMC methods combined with dynamic programming for structural estimation of dynamic discrete choice models. However, the majority of applied micro-econometric studies estimating discrete choice panel data models with lagged states applies classical maximum likelihood approaches, see e.g. Akay (2009), Haan (2010) and Prowse (2010).

The contribution of our paper is to compare the numerical performance of Bayesian estimation procedures with MSL estimation of dynamic discrete choice models and to provide evidence about the advantages of Bayesian methods for practitioners. We consider an application of a dynamic discrete choice model of female labor supply with three distinct states and estimate dynamic mixed logit models. Our analysis is based on longitudinal data from the German Socioeconomic Panel (SOEP).[4]

The empirical comparison leads to two important conclusions which are highly relevant for practitioners estimating dynamic discrete choice models with unobserved heterogeneity. First, when considering a multivariate normal distribution for the unobserved heterogeneity both approaches, the MCMC estimator and the MSL estimation, yield almost identical results. This shows that for a finite sample of the size which is typical for common household panels, our findings are in line with the asymptotic re-

---

converge at a local maximum when unobserved effects follow a log-normal distribution. Therefore - similar to our strategy for dynamic models in this paper - they suggest to use Bayesian methods for the estimation of the static model.

[3]Zellner and Rossi (1984) were the first to apply Bayesian procedures for a logit model using importance sampling, and starting with Zeger and Karim (1991) various MCMC methods have been developed for the estimation of discrete choice models, see Albert and Chib (1993) and McCulloch and Rossi (1994) for probit models, and Allenby and Lenk (1994) and Allenby (1997) for logit models.

[4]From a practitioner's perspective, it seems more interesting to assess the numerical performance of the Bayesian and maximum likelihood procedures with real world data than with a simulated data set because the distributional assumptions of a simulated data set often appear a bit artificial.

2

sults of the Bernstein-von Mises Theorem. Hence, the Bayesian estimates can be given a classical interpretation. Second, we show that when imposing distributional assumptions about the unobserved heterogeneity which are consistent with economic theory, e.g. log-normally distributed consumption preferences, only the Bayesian method provides reliable estimates. The classical estimator does not converge because it fails to find an increase of the likelihood function after a few iterations. This is the case even for starting values that have been set close to the maximum of the likelihood function. With log-normally distributed consumption preferences the respective coefficient decreases significantly by about 20%. In our example, however, this translates not into significantly different labor supply elasticities for the population mean.

The paper is organized as follows. After a general comparison between Bayesian and classical estimation of discrete choice models with unobserved heterogeneity, we outline our MCMC procedure. Then, we describe the dynamic labor supply model, the detailed specification and the data for our specific application. We compare estimation results of the dynamic labor supply model with different specifications of unobserved heterogeneity using i) classical and ii) Bayesian estimation procedures. A final section summarizes our findings.

2 Classical versus Bayesian estimation of discrete choice models

In the following, we describe the classical and the Bayesian estimation procedure of discrete choice models. For convenience we use a similar notation as in Train (2009). We start with the classical perspective.

2.1 Maximum Simulated Likelihood Estimation

Let's assume that each individual i maximizes period t's utility U_{ijt} by making choice $j \in 1, ..., J$. The probability that individual i makes a series of choices $\mathbf{j} = (j_1, ..., j_T)$ in the T observed time periods is $P_{i\mathbf{j}} = Prob(U_{ijt} > U_{ikt}, \forall j \neq k, \forall j \in \mathbf{j}, \forall t = 1, ..., T)$. Allowing for random taste variation, we can set up a random utility model $U_{ijt} = \beta_i x_{ijt} + \epsilon_{ijt}$ where x_{ijt} is a vector of covariates, β_i is an individual-specific (random)

coefficient vector, and ϵ_{ijt} is an idiosyncratic error term. We denote the conditional probability of i making a series of choices $\mathbf{j}$ as $L_{i\mathbf{j}}(\beta_i)$. Then, we can write the unconditional probability $P_{i\mathbf{j}} = \int L_{i\mathbf{j}}(\beta_i)f(\beta_i|\theta)d\beta_i$ where $f(\beta_i|\theta)$ is the density function describing the distribution of β_i conditional on the parameters θ of the distribution. Thus, we integrate over all possible states of β_i. Usually, the integral must be approximated by simulation methods. Assuming that ϵ_{ijt} follows an extreme value distribution, the model becomes a mixed logit model. McFadden and Train (2000) show that any random utility model can be approximated to any degree of accuracy by a mixed logit model if the density function $f(\beta_i|\theta)$ is chosen appropriately.

The probability $P_{i\mathbf{j}}$ can be approximated by simulation procedures like the method of simulated likelihood.[5] The idea is that taking draws β_i^r from $f(\beta_i|\theta)$ allows calculating $L_{i\mathbf{j}}(\beta_i^r)$. The simulated probability is then $\hat{P}_{i\mathbf{j}} = \frac{1}{R}\sum_{r=1}^{R} L_{i\mathbf{j}}(\beta_i^r)$ where R is the number of draws. The simulated log likelihood function becomes

$$SLL = \sum_{i=1}^{N} \sum_{\mathbf{j}\in\mathbf{J}} d_{i\mathbf{j}} log\hat{P}_{i\mathbf{j}} \tag{1}$$

$d_{i\mathbf{j}}$ equals 1 if individual i makes the series of choices $\mathbf{j}$ and zero otherwise. $\mathbf{J}$ is the set of all possible series of choices. The maximum simulated likelihood estimator (MSLE) $\hat{\theta}^{MSL}$ of θ is the value that maximizes the SLL. The maximum of the SLL can be found using numerical methods. However, problems may arise if the optimization algorithm does not converge or converges at a local maximum of the likelihood function. The MSLE is consistent if R rises at any rate with $\sqrt{N}$. It is also efficient if R rises faster than $\sqrt{N}$. Note that for R fixed the MSLE is inconsistent. For the asymptotic distribution of $\hat{\theta}^{MSL}$, we have $\sqrt{N}(\hat{\theta}^{MSL} - \theta) \xrightarrow{d} N(0, -H^{-1})$. -H is the information matrix. Thus, the MSLE is consistent, efficient and asymptotically normal if R rises faster than $\sqrt{N}$.

2.2 Bayesian Estimation

When estimating a discrete choice model with unobserved heterogeneity, there is a Bayesian analog to the classical procedures such as the method of MSL. From a Bayesian perspective the prior beliefs of the researcher about the parameters θ are

---

[5]See Train (2009) for more details on this method as well as alternative approaches like the method of simulated moments and the method of simulated scores.

represented by a prior distribution $k(\theta)$. Given the observed series of choices Y of the N decision makers, there is a density $K(\theta|Y)$ that summarizes the information about θ that is provided by Y. This density is called posterior distribution. The relationship between $k(\theta)$ and $K(\theta|Y)$ is given by Bayes rule. If $P_{i\mathbf{j}}(\theta)$ is the probability that decision maker i makes the series of choices $\mathbf{j}$, the probability of observing the sample is given by $L(Y|\theta) = \prod\limits_{i=1}^{N} P_{i\mathbf{j}}(\theta)$. By Bayes rule $K(\theta|Y)L(Y) = L(Y|\theta)k(\theta)$ where $L(Y) = \int L(Y|\theta)k(\theta)d\theta$. It follows that $K(\theta|Y) = \frac{L(Y|\theta)k(\theta)}{L(Y)}$.

The relationship between the classical and the Bayesian perspective is stated by the Bernstein-von Mises Theorem. A formal outline of the theorem can be found in Train (2009). First, the posterior distribution of θ converges to a normal distribution with variance $\frac{-H^{-1}}{N}$ as the sample size increases. -H is the information matrix being well-known from classical statistics. Thus, the parameters' posterior distribution converges to the sampling distribution of the maximum likelihood estimator. Second, the mean of the posterior distribution $\bar{\theta} = \int \theta K(\theta|Y)d\theta$ converges to the maximum of the likelihood function.[6] As a consequence, $\bar{\theta}$ is asymptotically equivalent to the MSLE. Since the Bayesian MCMC-estimator does not involve maximization of a likelihood function, the numerical problems of classical procedures do not arise. Standard errors can easily be calculated recognizing that the variance of the posterior distribution of θ is the asymptotic sampling variance of the MSLE. Note that this is a classical perspective on Bayesian estimates. Simulation methods allow an approximation of $\bar{\theta}$ by taking R draws from the posterior distribution. Then, $\hat{\theta}^{MCMC} = \frac{1}{R} \sum\limits_{r=1}^{R} \theta^r$ is the simulated mean where θ^r is the rth draw from $K(\theta|Y)$. Since the standard errors of the estimates correspond to the standard deviation of $K(\theta|Y)$, they can be simulated by taking the standard deviation of the R draws. For a fixed number of draws, the simulated mean of the posterior distribution is consistent and asymptotically normal. We have $\sqrt{N}(\hat{\theta}^{MCMC} - \theta) \xrightarrow{d} N(0, -H^{-1})$ as for the MSLE. If the number of draws increases at any rate with sample size, the estimator is efficient. Thus, consistency and efficiency can be achieved under more relaxed conditions than for the MSLE. Draws from the posterior distribution $K(\theta|Y)$ can be taken by MCMC methods.

---

[6]This follows directly from the symmetry of the normal distribution. As has been pointed out, the posterior distribution of the parameters is asymptotically normal.

2.3 MCMC procedure

We follow an estimation procedure that has been developed by Allenby and Lenk (1994) and Allenby (1997), and generalized by Train (2001). The approach relies on diffuse priors and applies Gibbs sampling together with the more general, but computationally more expensive, Metropolis-Hastings algorithm (Metropolis et al. (1953) , Hastings (1970)). Gibbs sampling allows taking draws from the posterior distribution of subsets of the parameters conditional on the other parameters. We specify diffuse priors for the mean b and variance-covariance matrix W of the individual-level coefficients β_i: k(b) is $N(b_0, S_0)$ with S_0 being extremely large (i.e. flat prior) and k(W) is inverted Wishart IW(F,I) where F is the number of random coefficients and I is an F-dimensional identity matrix. In empirical applications, we often have no clear idea about the prior distribution. Therefore, it seems to be reasonable to assume an uninformative prior. If the utility that individual i obtains in period t from choosing alternative j is $U_{ijt} = \beta_i x_{ijt} + \epsilon_{ijt}$, drawing from the posterior distribution $K(b, W|Y)$ is fast and convenient when β_i is considered as parameters along with b and W (see Train (2009) for more details). Then, the posterior distribution is

$$K(b, W, \beta_i|Y) \propto L_{i\mathbf{j}}(\beta_i)\phi(\beta_i|b, W)k(b, W), \forall i \tag{2}$$

where ϕ is the normal density function with mean b and variance-covariance matrix W. Since some of the coefficients are kept fixed, this requires an additional layer of the Gibbs sampling. The fixed coefficients are denoted α and the utility becomes $U_{ijt} = \alpha z_{ijt} + \beta_i x_{ijt} + \epsilon_{ijt}$. Then, the individual-level coefficients β_i must be drawn conditional on α and vice versa. The prior distribution of α, k(α), is also assumed to be normal and flat. Then, we get the following four posterior distributions for the four sets of parameters β_i, b, W, and α:

1. $K(\beta_i|\alpha, b, W) \propto L_{i\mathbf{j}}(\alpha, \beta_i)\phi(\beta_i|b, W)$, $\forall i$. The Metropolis-Hastings algorithm is used to draw from this posterior distribution.

2. $K(b|W, \beta_i)$ is $N(\frac{1}{N}\sum_i \beta_i, \frac{W}{N})$, $\forall i$.

3. $K(W|b, \beta_i)$ is $IW(K + N, \frac{KI+N\bar{S}}{K+N})$, $\forall i$, where F is the number of random coefficients, I is an F-dimensional identity matrix and $\bar{S} = \frac{\sum_i(\beta_i-b)(\beta_i-b)'}{N}$.

4. $K(\alpha|\beta_i) \propto \prod_i L_{i\mathbf{j}}(\alpha, \beta_i)$, $\forall i$. The Metropolis-Hastings algorithm is used to draw from this posterior distribution.

Setting the initial values for all parameters to 0.1 and using Gibbs sampling to draw from the posterior distribution of the four sets of parameters, it takes a number of iterations until the draws converge to draws from the posterior distribution. The draws prior to convergence must be discarded. In our empirical application we discard the first 20,000 draws for burn-in and use the following 10,000 draws for the actual estimation. We always check on convergence by increasing the number of draws used for burn-in and, then, comparing the results. If results do not change even with a substantial increase of the draws used for burn-in, this suggests that convergence has been achieved. The acceptance rate in the Metropolis-Hastings algorithm is set to be 0.3.[7] Since the draws are serially correlated, we follow the suggestion of Train (2009) and save only every tenth draw reducing the correlation of the retained draws by an order of ten.

3 Economic model and specification

We specify a dynamic discrete choice model with random coefficients to model labor supply of married females. Our approach assumes that current labor supply is causally affected only by last period's labors supply, thus following a first-order Markov process, see, e.g. Keane and Wolpin (2001) or Haan (2010). We differentiate between three distinct states, $j \in 1, 2, 3$: full-time, part-time and non-employment. The median hours of work per week are 13.5 for part-time employees and 38 for full-time employees. We focus on married females in prime working age, defined as 25-59 years. The wife maximizes her utility conditional on her husband's behavior which is taken as given. Collective bargaining as suggested by e.g. Chiappori (1988) is not considered. Each period t, wife i receives a utility flow U_{ijt} depending on her choice of labor supply category j. U_{ijt} is a function of period t's leisure time and household consumption as well as of the previous period's labor market state. Since our model does not account for saving behavior, household consumption equals net household income. Net household

---

[7]Gelman, Carlin, Stern, and Rubin (1995) find that the optimal acceptance rate is about 0.44 if x_{ijt} contains only one variable and decreases with the number of variables in x_{ijt} towards 0.23.

income is simulated using the micro simulation model STSM (Steiner, Wrohlich, Geyer, and Haan, 2008), for more details, see below. We ensure concavity of the preferences by assuming a Cobb-Douglas functional form of the utility function:

$$U_{ijt} = \beta_{it}^L Log L_{jt} + \beta_i^I Log I_{ijt} + \epsilon_{ijt} \tag{3}$$

$$\beta_{it}^L = b_i^L + b^C C_{it} + b^E E_{it} + b_{10}^Z Z_{1i0} + b_{20}^Z Z_{2i0} + b_{1i}^Z Z_{1it-1} + b_{2i}^Z Z_{2it-1} \tag{4}$$

where L_{jt} is wife's leisure time and I_{ijt} is household consumption (i.e. net household income). Z_{1it-1} and Z_{2it-1} are dummy variables indicating the state which has been chosen in the previous period; Z_{1it-1}: non-employment, Z_{2it-1}: part-time. Analogously Z_{1i0} and Z_{2i0} are variables indicating the employment state of the first observed period (2001). Thus, we follow Wooldridge (2005) by modelling the distribution of unobserved heterogeneity conditional on the first observed state to account for the nonrandomness of the initial state.[8] The lagged dependent variables and the initial state are modeled as taste shifters of the leisure time in the current period, L_{jt}. Similarly, other individual specific characteristics are included as taste shifters for leisure. C_{it} is a dummy variable for children aged 0-3 years in the household and E_{it} is a dummy variable indicating residence in East Germany. ϵ_{ijt} is an i.i.d. error term that follows an extreme value distribution. Note that in the above specification β_i^I, b_i^L, b_{1i}^Z, and b_{2i}^Z are individual specific coefficients which introduce unobserved preference heterogeneity to the model. Distributional assumptions for these random coefficients must be imposed, and a standard approach in the literature is to assume a multivariate normal distribution for the unobserved components. This leads to a mixed logit model (McFadden and Train, 2000).

We estimate various specifications of the model. In a first specification, we neglect unobserved preference heterogeneity and assume that the variance-covariance matrix of the individual specific coefficients is zero. This results in a simple conditional logit model. The second specification only assumes two random coefficients, for b_i^L and β_i^I that are uncorrelated. Specification 3 extends specification 2 allowing the two random coefficients to be correlated. In specification 4, we estimate the full model

---

[8]See Akay (2011) for the performance of the method suggested by Wooldridge (2005).

allowing for four random coefficients (also on the state dependence) and for a free correlation structure. Specifications 2 to 4 assume a multivariate normal distribution of the random coefficients which is a standard approach in the literature. While this makes sense for the leisure preferences, it is problematic for consumption preferences, see e.g. van Soest (1995). Assuming a normal distribution of unobserved heterogeneity in consumption preferences predicts negative preferences for some individuals. In our example, about 15 % of the individuals are predicted to have negative consumption preferences when estimating specification 4. This is not plausible from an economic point of view. Therefore, we estimate a fifth specification that restricts consumption preferences to be positive by assuming them to be log-normally distributed. This can be easily achieved by setting β_i^I equal to $e^{b_i^I}$ and estimating b_i^I under the assumption that b_i^I is normally distributed. Thus, the utility function becomes:

$$U_{ijt} = \beta_{it}^L Log L_{jt} + e^{b_i^I} Log I_{ijt} + \epsilon_{ijt} \tag{5}$$

If b_i^I is normally distributed, $e^{b_i^I}$ follows a log-normal distribution. We can obtain draws from the posterior distribution of the log-normally distributed β_i^I by transforming the draws from the posterior distribution of b_i^I with $e^{(\cdot)}$. The variance of the log-normally distributed coefficient can be estimated transforming draws from the posterior distribution of the mean and the variance of b_i^I to draws from the posterior distribution of the variance of β_i^I. Note that $var(\beta_i^I) = e^{(mean(b_i^I) + var(b_i^I)/2)} * (e^{var(b_i^I)} - 1)$. Specification 5 also allows for a free correlation structure.

The first specification is estimated by the method of maximum likelihood using an analytic gradient in the optimization algorithm. Specifications 2 to 4 are estimated by the method of MSL using again an analytic gradient.[9] For comparison, we estimate specification 4 as well using the Bayesian MCMC estimator.[10] Specification 5 could not be estimated by ths MSL-estimator because the optimization algorithm does not converge because it fails to find an increase of the likelihood function after a few iterations. This is the case even for starting values that have been set close to the

---

[9]The simulation of the choice probabilities is based on 200 Halton draws. Estimating the model with other choices of R has shown that 200 Halton draws seems to be a lower bound to the number of draws required for the MSLE to have good statistical properties in our finite sample.

[10]We only compare the estimation of specification 4 with both estimation procedures, the comparison of the less flexible specifications 2 and 3 lead to the same results and are therefore omitted.

maximum of the likelihood function. Train (2009) points out that in the classical approach locating the maximum of the likelihood is considerably more difficult with log-normal distributions. And, even if a maximum is found, it may happen that the Hessian is singular at this point. However, specification 5 can be estimated easily by the Bayesian procedures. Since the Bayesian MCMC estimator does not involve maximization of a likelihood function, the numerical problems of the classical approach do not arise.

4 Data

The empirical analysis is based on data from the German Socio-Economic Panel (SOEP). The SOEP started in 1984 and annually collects information at the household and individual levels (Wagner, Frick, and Schupp, 2007). We construct a balanced panel of married females covering the years 2001 to 2008. The sample is restricted to married females in prime working age defined as 26-59 years whose labor supply may be considered as flexible, thus excluding pensioners, self-employed, and people in institutions. The final sample consists of 1,598 wives. The SOEP contains detailed information about employment behavior, as well as other individual and household characteristics. As mentioned above, we differentiate between non-employment, part-time and full-time work. The median hours of work per week are 13.5 for part-time employees and 38 for full-time employees. Leisure time is defined as 80 minus median hours of work per week of the employment category.

Table 1: Descriptive statistics

Variable	Mean	Std.	Min	Max
Age	44.6	7.1	26	59
Children from 0-3 years	0.054	0.227	0	1
German nationality	0.932	0.251	0	1
East Germany	0.232	0.422	0	1
Non-employment	0.21	0.407	0	1
Part-time	0.263	0.44	0	1
Full-time	0.527	0.5	0	1

Table 1 contains some descriptive statistics including the distribution of the working alternatives. Net household income is simulated for each employment category using the micro simulation model STSM (Steiner, Wrohlich, Geyer, and Haan, 2008). Net

household income is a complex nonlinear function of the gross wages of both partners.[11] The STSM allows taking into account the complexity of the German tax and transfer system. Simulation is based on detailed information on household characteristics as contained in the SOEP. Other explanatory variables entering directly our labor supply model are a regional dummy indicating whether a household is living in eastern or western Germany and a binary variable providing information on young children aged three or less in the household.

5 Results and discussion

5.1 Parameter estimates

Table 2 contains the parameter estimates for the five specifications that have been described above. Assuming no preference heterogeneity (see specification 1) seems to result in a substantial underestimation of the income coefficient, β_i^I. When allowing for correlation between random coefficients, the size of β_i^I increases significantly in comparison to the specification without correlated random coefficients (compare specifications 2 und 3). This finding supports previous studies, e.g. Prowse (2010) that specifications without unobserved heterogeneity lead to biased estimates of central structural parameters. Because of the multiple interactions of the leisure term the interpretation of coefficient b_i^L is not straight forward. However we find again a substantial difference in this estimated coefficient between models with and without unobserved heterogeneity. Neglecting heterogeneity in the coefficients of state dependence, b_{1i}^Z and b_{2i}^Z, results in significantly smaller estimates of these coefficients (compare specifications 3 and 4). This misspecification is in particular relevant for the interpretation of state dependence Heckman (1981b) and the prediction of short term labor supply elasticities. Looking at the information criteria AIC and BIC, specification 4 appears to have the better trade-off between flexibility and efficiency in comparison to specifications 1-3. Both a free correlation structure of the random coefficients and accounting for heterogeneity in the coefficients of state dependence improve the fit of the model. It appears that neglecting the correlation between random coefficients leads to an underestimation of the variances. The estimated covariances indicate a positive correlation between income

---

[11]Note, for women not employed in the month preceding the interview, gross hourly wages are estimated by applying a two-stage estimation procedure with a Heckman sample selection correction.

and leisure preferences. To summarize we find in line with the previous literature that it is important to account for unobserved heterogeneity in a flexible and general way.

However, a more detailed comparison of the results lead to important conclusions for practitioners related to the Bayesian estimations procedures that go beyond the standard findings. First, when considering a multivariate normal distribution for the unobserved heterogeneity (e.g. Specification 4) both approaches, the MCMC estimator and the MSL estimation, yield almost identical results. The estimated means of the posterior distribution of the parameters and of the variance-covariance matrix of the random coefficients differ only marginally from the parameter estimates of the MSL estimation. This shows that for a finite sample of the size which is typical for common household panels, the Bernstein-von Mises Theorem holds and the Bayesian estimates can be given a classical interpretation.

Second, we show that when imposing distributional assumptions about the unobserved heterogeneity which are more in line with economic theory only the Bayesian method provides reliable estimates. In particular, restricting consumption preferences to be positive by assuming them to be log-normally distributed yields an estimate of the respective coefficient, β_i^I, that is significantly smaller than in the case of normally distributed consumption preferences (compare specifications 4 and 5). The distribution of a normally distributed β_i^I is shifted to the right when fitting the model such that a smaller share of individuals is predicted to have negative consumption preferences. Restricting β_i^I to be positive is necessary to avoid an overestimation of the coefficient. The classical estimator does not converge because it fails to find an increase of the likelihood function after a few iterations. This finding even holds for starting values that have been set close to the maximum of the likelihood function. With log-normally distributed consumption preferences the respective coefficient decreases significantly by about 20 %.

5.2 Labor supply elasticities

For an economic interpretation of the female labor supply behavior it is necessary to compute labor supply elasticities. Therefore, we use the estimates for different specifications to derive elasticities for a permanent shock to gross wage. We only derive elasticities for the population mean. In particular we calculate the relative change in

Table 2: Parameter estimates

Parameters	ML (spec. 1)	MSL (spec. 2)		MSL (spec. 3)		MSL (spec. 4)		MCMC (spec. 4)		MCMC (spec. 5)	
	value	value	var.	value	var.	value	var.	mean	var.	mean	var.
b_{10}^Z	3.30	9.54		10.75		5.75		5.88		4.03	
	(0.19)	(0.58)		(0.59)		(0.66)		(0.55)		(0.32)	
b_{20}^Z	2.05	5.92		6.25		4.02		4.05		2.81	
	(0.15)	(0.42)		(0.43)		(0.45)		(0.38)		(0.20)	
b^C	2.75	4.73		5.24		5.06		4.90		4.17	
	(0.29)	(0.41)		(0.44)		(0.45)		(0.42)		(0.35)	
b^E	-1.36	-2.58		-2.53		-2.10		-2.20		-1.71	
	(0.17)	(0.35)		(0.39)		(0.31)		(0.31)		(0.26)	
b_{1i}^Z	9.34	5.64		5.60		11.07	51.34	11.02	55.87	12.27	67.72
	(0.23)	(0.32)		(0.32)		(0.72)	(7.80)[1]	(0.58)	(8.25)	(0.53)	(8.55)
b_{2i}^Z	4.54	2.46		2.32		4.18	5.87	4.07	5.14	4.87	6.91
	(0.15)	(0.21)		(0.21)		(0.37)	(1.78)[1]	(0.30)	(1.21)	(0.23)	(1.04)
b_i^L	-5.17	-4.31	8.51	-3.39	51.93	-3.30	18.21	-3.22	19.61	-4.21	7.63
	(0.15)	(0.29)	(1.40)[1]	(0.37)	(6.14)[1]	(0.33)	(3.74)[1]	(0.30)	(3.83)	(0.24)	(1.81)
β_i^I	1.89	7.54	26.61	10.38	167.95	9.39	92.46	9.41	89.49	6.97	81.38[2]
	(0.19)	(0.53)	(4.12)[1]	(0.69)	(22.94)[1]	(0.78)	(15.18)[1]	(0.64)	(14.96)	(0.54)	(24.52)
random coeff.	no	2		2		4		4		4	
free correlation	no	no		yes		yes		yes		yes	
log-normal β_i^I	no	no		no		no		no		yes	
ML / MSL	-6393.1	-6245.8		-6137.5		-6052.5		-		-	
AIC	12802.3	12511.5		12297.0		12140.9		-		-	
BIC	12845.3	12565.3		12356.1		12237.7		-		-	

Standard errors / standard deviations of the posterior distributions are given in parentheses.

[1] For numerical reasons, we estimate the elements of the Cholesky matrix of the variance-covariance matrix with the method of MSL. The standard errors of the derived variances, then, are approximated on the basis of 200 draws taken from the asymptotically normal sampling distribution of the elements of the Cholesky matrix.

[2] The variance of the log-normally distributed coefficient has been estimated transforming draws from the posterior distribution of the mean and the variance of b_i^I to draws from the posterior distribution of the variance of β_i^I. Note that $\beta_i^I = e^{b_i^I}$.

the labor supply behavior which result from a permanent relative increase in female gross wages. We predict for 10 consecutive periods of time for both a baseline scenario without a shock and a scenario where a permanent wage shock occurs in the first period, i.e. wages are 10% higher in the second scenario. Then, we use the differences between the predictions of the two scenarios to derive the labor supply elasticities. All predictions are based on the posterior distribution of the model parameters.[12] This allows investigating the dynamics of female labor supply. For a detailed description of the calculation see Haan (2010) and Haan and Uhlendorff (2012).

Table 3 contains the gross wage elasticities. The elasticities for period 1 can be inter-

---

[12]This is natural for predictions based on Bayesian estimation. For predictions based on MSL estimation, we have to compute the choice probabilities conditional on each individual's series of choices by a procedure that involves simulation within simulation.

Table 3: Gross wage elasticities

Period	ML (spec. 1)	MSL (spec. 2)	MSL (spec. 3)	MSL (spec. 4)	MCMC (spec. 4)	MCMC (spec. 5)
1	0.054	0.143	0.165	0.141	0.143	0.124
	[0.044,0.064]	[0.127,0.158]	[0.155,0.176]	[0.124,0.157]	[0.127,0.159]	[0.109,0.141]
2	0.080	0.184	0.207	0.190	0.193	0.163
	[0.065,0.096]	[0.165,0.203]	[0.195,0.219]	[0.170,0.211]	[0.174,0.214]	[0.144,0.186]
3	0.096	0.199	0.221	0.213	0.219	0.187
	[0.077,0.114]	[0.179,0.219]	[0.208,0.233]	[0.192,0.234]	[0.196,0.243]	[0.164,0.213]
4	0.105	0.205	0.225	0.225	0.234	0.202
	[0.085,0.126]	[0.184,0.225]	[0.212,0.239]	[0.204,0.247]	[0.210,0.261]	[0.177,0.230]
5	0.110	0.207	0.227	0.232	0.245	0.212
	[0.089,0.131]	[0.186,0.228]	[0.213,0.241]	[0.211,0.254]	[0.219,0.273]	[0.186,0.242]
6	0.113	0.208	0.228	0.236	0.252	0.220
	[0.092,0.135]	[0.187,0.229]	[0.214,0.241]	[0.215,0.258]	[0.225,0.281]	[0.193,0.251]
7	0.115	0.208	0.228	0.239	0.258	0.226
	[0.094,0.137]	[0.187,0.229]	[0.214,0.242]	[0.217,0.260]	[0.230,0.288]	[0.198,0.257]
8	0.117	0.208	0.228	0.240	0.262	0.231
	[0.095,0.138]	[0.187,0.229]	[0.214,0.242]	[0.219,0.262]	[0.234,0.293]	[0.201,0.263]
9	0.117	0.208	0.228	0.241	0.265	0.235
	[0.095,0.138]	[0.187,0.229]	[0.214,0.242]	[0.219,0.263]	[0.237,0.297]	[0.205,0.268]
10	0.118	0.208	0.228	0.242	0.268	0.239
	[0.096,0.139]	[0.187,0.229]	[0.214,0.242]	[0.220,0.263]	[0.239,0.300]	[0.207,0.272]

We calculate confidence intervals at the 5 percent significance level. Note that the interpretation of the Bayesian confidence intervals differs from the interpretation of the classical confidence intervals.

preted as short term elasticities, while the elasticities for period 10 can be interpreted as long term elasticities. Confidence intervals are estimated at the 0.95 significance level. The confidence intervals for the elasticities derived from the MCMC estimation have been computed directly using the draws from the posterior distribution of the model parameters. We also approximate confidence intervals for the elasticities derived from the ML and MSL estimation by taking 200 draws from the sampling distribution of the model parameters and deriving the elasticities for each of these draws. Unlike the Bayesian confidence intervals, this approach must assume that the asymptotically normal sampling distribution of the parameters holds for our finite sample. Note that the interpretation of Bayesian confidence intervals differs from the interpretation of classical confidence intervals. The Bayesian confidence interval represents the belief that the true elasticity lies within the interval with a probability of 0.95.

The estimated elasticities differ depending on whether random coefficients are allowed for. It results from the underestimation of the income coefficient in the conditional logit model that neglecting preference heterogeneity seems to induce a substantial downward bias to the estimated elasticities. For example, the conditional logit

model predicts a long term gross wage elasticity of about 0.12 as opposed to 0.24 for specification 4 (MSL). The difference is significant. The predicted elasticities do only differ marginally depending on whether the model has been estimated by the method of MSL or by the Bayesian MCMC estimator. This reflects the very similar parameter estimates of both estimation approaches. Finally, despite of the smaller income coefficient in specification 5 - related to log-normally distributed consumption preferences - labor supply elasticities for the population mean are not statistically different from specification 4.

6   Conclusion

This analysis provides evidence for the advantage of using Bayesian estimation procedures instead of classical maximum likelihood estimation for the estimation of dynamic discrete choice models. These models usually require a general specification of unobserved preference heterogeneity and therefore often relatively complex estimation routines need to be applied. We estimate a dynamic discrete choice model of female labor supply with three distinct states and different specifications of unobserved heterogeneity.

The empirical comparison leads to two important conclusions which are highly relevant for practitioners estimating dynamic discrete choice models with unobserved heterogeneity. First, when considering a multivariate normal distribution for the unobserved heterogeneity both approaches, the MCMC estimator and the MSL estimation, yield almost identical results. This shows that for a finite sample of the size which is typical for common household panels, our findings are in line with the asymptotic results of the Bernstein-von Mises Theorem. Hence, the Bayesian estimates can be given a classical interpretation. The second finding demonstrates the advantage of using Bayesian estimation procedures. We show that when imposing distributional assumptions which are consistent with economic theory, e.g. log-normally distributed consumption preferences, the Bayesian method performs well and provides reasonable estimates, while the MSL estimator does not converge. These results indicate that Bayesian procedures can be a beneficial tool for the estimation of dynamic discrete choice models.

References

Akay, A. (2009): "Dynamics of the Employment Assimilation of First-Generation Immigrant Men in Sweden: Comparing Dynamic and Static Assimilation Models with Longitudinal Data," *IZA Discussion Paper*, 4655.

——— (2011): "Finite-sample comparison of alternative methods for estimating dynamic panel data models," *Journal of Applied Econometrics*, forthcoming.

Albert, J. H., and S. Chib (1993): "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88(422), pp. 669–679.

Allenby, G. (1997): "An introduction to hierarchical Bayesian modelling," *Tutorial Notes*.

Allenby, G. M., and P. J. Lenk (1994): "Modeling Household Purchase Behavior with Logistic Normal Regression," *Journal of the American Statistical Association*, 89(428), 1218–1231.

Chiappori, P. (1988): "Rational Household Labor Supply," *Econometrica*, 56, 63–89.

Dube, J.-P., G. J. Hitsch, and P. E. Rossi (2010): "State dependence and alternative explanations for consumer inertia," *Rand Journal of Economics*, 41, 417–445.

Fitzenberger, B., A. Osikominu, and M. Paul (2010): "The Heterogeneous Effects of Training Incidence and Duration on Labor Market Transitions," *IZA Discussion Paper*, 5269.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995): *Bayesian Data Analysis*. Chapman and Hall.

Haan, P. (2010): "A Multi-state model of state dependence in labor supply: Intertemporal labor supply effects of a shift from joint to individual taxation," *Labour Economics*, 17(2), 323–335.

Haan, P., and A. Uhlendorff (2012): "Intertemporal Labor Supply and Involuntary Unemployment," *Empirical Economics*, forthcoming.

HASTINGS, W. K. (1970): "Monte Carlo sampling methods using Markov chains and their applications," 57(1), 97–109.

HECKMAN, J. (1981a): "Heterogeneity and State Dependence," in *Studies in Labor Markets*, ed. by S. Rosen, pp. 91–139. Chicago Press, Chicago, IL.

——— (1981b): "Statistical Models for Discrete Panel Data," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski, and D. McFadden, pp. 114–178. MIT Press, Cambridge, MA.

IMAI, S., N. JAIN, AND A. CHING (2009): "Bayesian Estimation of Dynamic Discrete Choice Models," *Econometrica*, 77, 1865–1899.

KEANE, M., AND K. WOLPIN (2001): "Estimating Welfare Effects Consistent with Forward-Looking Behavior: Part II: Empirical Results," *Journal of Human Resources*, 37, 600–622.

LANCASTER, T. (1990): *The Econometric Analysis of Transition Data*. Cambridge University Press, Cambridge.

MCCULLOCH, R., AND P. E. ROSSI (1994): "An exact likelihood analysis of the multinomial probit model," *Journal of Econometrics*, 64(1-2), 207–240.

MCFADDEN, D., AND K. TRAIN (2000): "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, 15(5), 447–470.

METROPOLIS, N., A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER, AND E. TELLER (1953): "Equation of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, 21(6), 1087–1092.

PROWSE, V. (2010): "Modeling Employment Dynamics with State Dependence and Unobserved Heterogeneity," *IZA Discussion-Paper*, 4489.

REGIER, D. A., M. RYAN, E. PHIMISTER, AND C. A. MARRA (2009): "Bayesian and classical estimation of mixed logit: An application to genetic testing," *Journal of Health Economics*, 28(598-610).

Steiner, V., K. Wrohlich, J. Geyer, and P. Haan (2008): "Documentation of the Tax-Benefit Microsimulation Model STSM: Version 2008," *Data Documentation 31*.

Train, K. (2001): "A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit," Discussion paper.

——— (2009): *Discrete Choice Models using Simulation.* Cambridge University Press, Cambridge, UK, second edn.

Troske, K., and A. Voicu (2010): "Joint Estimation of Sequential Labor Force Participation and Fertility Decisions Using Markov Chain Monte Carlo Techniques," *Labour Economics*, 17, 150–169.

van den Berg, G. J. (2001): "Duration Models: Specification, Identification, and Multiple Durations," in *Handbook of Econometrics, Vol.5*, ed. by J. J. Heckman, and E. Leamer, pp. 3381–3460. North-Holland, Amsterdam.

van Soest, A. (1995): "Structural Models of Family Labor Supply: A Discrete Choice Approach," *Journal of Human Resources*, 30, 63–88.

Wagner, G., J. Frick, and J. Schupp (2007): "The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements," *Schmollers Jahrbuch*, 127(1), 139–169.

Wooldridge, J. (2005): "Simple Solutions to the Initial Conditions Problem for Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity," *Journal of Applied Econometrics*, 20, 39–54.

Zeger, S. L., and M. R. Karim (1991): "Generalized Linear Models With Random Effects; A Gibbs Sampling Approach," *Journal of the American Statistical Association*, 86(413), pp. 79–86.

Zellner, A., and P. E. Rossi (1984): "Bayesian analysis of dichotomous quantal response models," *Journal of Econometrics*, 25(3), 365–393.