A Note on 'Bayesian analysis of the random coefficient model using aggregate data', an alternative approach

Zenetti, German

Goethe-University Frankfurt, School of Business and Economics, Department of Marketing

5 November 2010

A Note on 'Bayesian analysis of the random coefficient model using aggregate data', an alternative approach

German Zenetti *

Goethe-University Frankfurt, School of Business and Economics, Department of Marketing

November 5, 2010

Abstract

In this note on the paper from Jiang, Manchanda & Rossi (2009), I want to discuss a simple alternative estimation method of the multinomial logit model for aggregated data with random coefficients - the so-called BLP model, named for Berry, Levinsohn & Pakes (1995). The estimation is conducted through a Bayesian estimation similar to Jiang et al. (2009). However in contrast to Jiang et al. (2009) I omit the time-intensive contraction mapping for assessing the mean utility in every iteration step of the estimation procedure. The likelihood function is computed through a special case of the control function method (Park & Gupta (2009) and Petrin & Train (2002)). A full random walk MCMC approach is applied, that uses two random walk MCMC chains - one to draw the parameters of the model, and a second one to sampled an explicitly introduced uncorrelated error term. In total, the suggested simple procedure (i) permits the use of the full information from the data set, in contrast to Park & Gupta (2009), (ii) accelerates the Bayesian estimation by omitting the contraction mapping, in contrast to Jiang et al. (2009), and (iii) in contrast to both cited methods, allows the demand shock to be estimated without a distributional assumption, if desired.

Keywords: **Bayesian estimation, random coefficient logit, aggregate share** .

JEL Classification Numbers: C11, M3.

*Address: Grueneburgplatz 1, 60323 Frankfurt am Main, Germany, telephone: +49-69-798-34642, e-mail: *zenetti@wiwi.uni-frankfurt.de*.

1 Introduction

In this note[1] on the paper from Jiang et al. (2009), I want to discuss a simple alternative estimation method of the multinomial logit model for aggregated data with random coefficients - the so-called BLP model, named for Berry et al. (1995).

The estimation is conducted through a Bayesian estimation similar to Jiang et al. (2009). However in contrast to Jiang et al. (2009) I omit the time-intensive contraction mapping for assessing the mean utility in every iteration step of the estimation procedure. The likelihood function is computed through a special case of the control function method (Park & Gupta (2009) and Petrin & Train (2002)). A full random walk MCMC approach is applied, that uses two random walk MCMC chains - one to draw the parameters of the model, and a second one to sampled an explicitly introduced uncorrelated error term. In total, the suggested simple procedure (i) permits the use of the full information from the data set, in contrast to Park & Gupta (2009), (ii) accelerates the Bayesian estimation by omitting the contraction mapping, in contrast to Jiang et al. (2009), and (iii) in contrast to both cited methods, allows the demand shock to be estimated without a distributional assumption, if desired.

The rest of this paper is structured as follows. In the next section, the model setup is outlined, both for the usual BLP model and for the estimation through the Bayesian method, which includes the derivation of the likelihood function, the prior distribution and the Bayesian estimation through the posterior distribution. In the following section, a simulation study is employed to assess the performance of the introduced alternative estimation approach. Those results are compared with the outcome of the estimation according to Park & Gupta (2009) and Jiang et al. (2009). The paper ends with a conclusion.

[1] This study originates from a course paper for the lecture 'Bayesian Modeling for Marketing' held by Prof. Thomas Otter at the Goethe University Frankfurt for doctoral students in the winter semester of 2009/10.

2 Random coefficient logit model for aggregated data

2.1 General model

In the following, the choice model for aggregated data with random coefficients is considered, as developed by Berry et al. (1995), the so-called BLP model. Although Berry et al. (1995) study a model of markets in equilibrium, i.e. with explicit formulation and estimation of a demand side model and a supply side model, here only the demand side is taken into consideration.

This simulation concentrates on a normal distribution of unobserved individual preference[2], which is often used in applied work (e.g., recently Sovinsky Goeree (2008), Gowrisankaran & Rysman (2009), Albuquerque & Bronnenberg (2008)).

With a similar notation as in Nevo (2000) the utility of a product j=1,...,J for an individuum[3] i=1,...,I in a market t=1,...,T can be written as:

$$\begin{aligned} u_{ijt} &= x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt} \\ &= x_{jt}\beta + \xi_{jt} + x_{jt}(\Pi D_i + \Lambda\nu_i) + \varepsilon_{ijt} \\ &= \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}. \end{aligned} \tag{1}$$

x_{jt} is the vector of influence variables, including e.g. price or advertising, with the vector of random coefficients β_i, that is decomposed as $\beta_i = \beta + \Pi D_i + \Lambda\nu_i,\ \nu_i \sim N(0, I_K)$, where Λ is the lower-triangular Cholesky factor of the $K \times K$ covariance matrix, i.e. $\Sigma = \Lambda \cdot \Lambda'$.

ε_{ijt} is an independently and identically distributed (i.i.d.) extreme value distributed error term and ξ_{jt} is the surrogate of unknown product characteristics. Π are the influence parameters of the matrix of demographic variables D_i. Moreover, the mean value of the utility of product j in market t is $\delta_{jt} \equiv x_{jt}\beta + \xi_{jt}$ and thus $\mu_{ijt} \equiv x_{jt}(\Pi D_i + \Lambda\nu_i)$ is the individual specific deviation from δ_{jt}. Therefore, $\beta_i \sim N(\beta + \Pi D_i, \Sigma)$.

[2]However in the BLP model, any other distribution for individual heterogeneity can be employed.
[3]I.e. a costumer.

Define[4]

$$\Lambda = \begin{pmatrix} \gamma_{11} & 0 & \cdots & 0 \\ \gamma_{21} & \gamma_{22} & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ \gamma_{K1} & \cdots & & \gamma_{KK} \end{pmatrix}.$$

The expectation of the individual market shares s_{ijt},

$$\begin{aligned} s_{jt} = E_{D,\nu}(s_{ijt}) &= \int_D \int_\nu s_{ijt}(D,\nu) d\nu dP_D(D) \\ &= \int_D \int_\nu \frac{exp(x'_{jt}\beta + \xi_{jt} + x'_{jt}(\Pi D + \Lambda\nu))}{1 + \sum_{l=1}^{J} exp(x_{lt}\beta + \xi_{lt} + x_{lt}(\Pi D + \Lambda\nu))} \phi(\nu) d\nu dP_D(D) \end{aligned} \quad (2)$$

can be approximated by different simulation methods.

$D_i \sim P_D(D)$ is the distribution of the demographic variables and $\phi(\cdot)$ denotes the density of the standard normal distribution.

The most obvious kind of an approximation of the expectation is the mean value

$$\hat{s}_{jt} = \frac{1}{ns} \sum_{i=1}^{ns} w_i \frac{exp(x'_{jt}\beta + \xi_{jt} + x'_{jt}(\Pi D_i + \Lambda\nu_i))}{1 + \sum_{l=1}^{J} exp(x_{lt}\beta + \xi_{lt} + x_{lt}(\Pi D_i + \Lambda\nu_i))} \quad (3)$$

with ns [5] random draws or decisively chosen draws from the assumed distribution of ν and for given data observations from the empirical distribution of D with some appropriately chosen weights w_i for decisively chosen draws.

Later in the simulation study (section (3)) I use for the simulation of equation (3) the nested Gauss-Quadrature with nodes and weights on sparse grids as implemented and developed by Heiss & Winschel (2006)[6]. In the following, demographic variables are not considered, so D

[4] With respect to the exclusion restrictions for identification (Walker, Ben-Akiva & Bolduc 2007).

[5] ns for "number of simulations".

[6] Because the overall computation time depends mainly on the magnitude of ns (Dube, Fox & Su 2008), it is desirable to use the smallest necessary number of draws. Gauss-Quadrature with nodes and weights on sparse grids has the advantage over applying the product rule to the usual Gauss-Quadrature (e.g., Judd (1998)) in that the construction of sparse grids with far fewer nodes is needed for the integration of integrals

is dropped from equations. Because one or more of the known[7] product characteristics x_{jt} might be chosen on the basis of unknown product characteristics ξ_{jt}, the in the literature well-known problem of endogeneity arises.

2.2 Likelihood

In the likelihood function, endogeneity is explicitly incorporated through the set up used by Park & Gupta (2009), whose theory goes back to Hausman (1954) and Heckman (1978). Their approach uses a special case of the more general control function estimation method of Petrin & Train (2002). Park & Gupta (2009) proceed as follows:
Let z_{jt} be the vector of instrumental variables that are correlated with the known product characteristics x_{jt}, e.g., only price[8], but that are uncorrelated with the unknown product characteristics ξ_{jt}.

$$\begin{aligned} x_{jt} &= Z_{jt}\zeta_j + \tilde{\nu}_{jt}, \\ & Z_{jt} \equiv I_K \otimes z'_{jt}, \\ \tilde{\nu}_{jt} &\overset{i.i.d.}{\sim} N(0, \Sigma_{\tilde{\nu}_j}), \\ \xi_{jt} &\overset{i.i.d.}{\sim} N(0, \sigma^2_{\xi_j}), \\ cov(\tilde{\nu}_{jt}, \xi_{jt}) &= \lambda_j, \\ cov(z_{jt}, \xi_{jt}) &= 0, \forall t. \end{aligned} \tag{4}$$

of higher dimension. The numerical integration with Gauss-Quadrature on sparse grids leads to a relatively fast estimation compared with other simulation methods such as simple random draws from the normal distribution or randomized draws from Halton Sequences (e.g., Bhat (2000), Train (2000)).

[7] I.e. known respectively unknown to the data analyst.

[8] To simplify the notation, this notational difference is not made, but incorporating a set of exogenous variables and a set of endogenous variables is straightforward. In fact, one could treat all variables as endogenous and simply add the exogenous variables to the instrumental variables.

The distribution of the two error terms can be expressed as the product of a Choleski factor of their covariance and two independent error terms:

$$\begin{pmatrix} \tilde{\nu}_{jt} \\ \xi_{jt} \end{pmatrix} = \begin{pmatrix} b_{11,j} & 0 \\ b_{21,j} & b_{22,j} \end{pmatrix} \begin{pmatrix} \omega_{1,jt} \\ \omega_{2,jt} \end{pmatrix} \tag{5}$$

$$\begin{pmatrix} \omega_{1,jt} \\ \omega_{2,jt} \end{pmatrix} \overset{i.i.d.}{\sim} N(0, I_{K+1}).$$

$$\tilde{\Sigma}_j = \begin{pmatrix} b_{11,j} & 0 \\ b_{21,j} & b_{22,j} \end{pmatrix} \begin{pmatrix} b_{11,j} & 0 \\ b_{21,j} & b_{22,j} \end{pmatrix}' = \begin{pmatrix} \Sigma_{\tilde{\nu}_j} & \Sigma_{\xi_j,\tilde{\nu}_j} \\ \Sigma_{\tilde{\nu}_j,\xi_j} & \sigma^2_{\xi_j} \end{pmatrix}.$$

With (5) and with $b_{11,j} = \Sigma^{1/2}_{\tilde{\nu}_j}$, the equation (1) and (4) can be transformed to

$$u_{ijt} = x'_{jt}\beta_i + b_{21,j}\omega_{1,jt} + b_{22,j}\omega_{2,jt} + \varepsilon_{ijt}, \text{ and} \tag{6}$$

$$x_{jt} = Z_{jt}\zeta_j + b_{11,j}\omega_{1,jt}. \tag{7}$$

After plugging $\omega_{1,jt}$ from equation (7) in equation (6), the utility can be written as:

$$u_{ijt} = x'_{jt}\beta_i + \varrho'_j[x_{jt} - Z_{jt}\zeta_j] + b_{22,j}\omega_{2,jt} + \varepsilon_{ijt},$$

with $\varrho_j = b_{21,j}b^{-1}_{11,j}$.

Note that now along with the usual extreme value error term ε_{ijt}, there are two additional uncorrelated error terms in the utility: $\omega_{1,jt}$ and $\omega_{2,jt}$.

As $\omega_{2,t} = (\omega_{2,1t}, \ldots, \omega_{2,Jt})$ is uncorrelated with every other term in the equation Park & Gupta (2009) denote it as "exogenous unmeasured product characteristic" (EUPC).

Now, the logit probability for the individual i's choice of product j at time t, s_{ijt}, thus equals:

$$s_{ijt} = \frac{exp(x'_{jt}\beta + \varrho'_j[x_{jt} - Z_{jt}\zeta_j] + b_{22,j}\omega_{2,jt} + x'_{jt}\Lambda\nu_i)}{1 + \sum_{l=1}^{J} exp(x'_{lt}\beta + \varrho'_l[x_{lt} - (I_K \otimes z'_{lt})\zeta_l] + b_{22,l}\omega_{2,lt} + x'_{lt}\Lambda\nu_i)}.$$

Explicit inclusion of endogeneity thus leads to the ordinary form of the random coefficient logit model, apart from the bias correction term $[x_{jt} - Z_{jt}\zeta_j]$ and the EUPC shocks $\omega_{2,jt}$. Given the bias correction term $[x_{jt} - Z_{jt}\zeta_j]$ and conditional on $\omega_{2,jt}$, the likelihood can be written as:

$$L_{1,t}(\omega_{2,t}) = \left(\frac{q_t!}{q_{0t}! \dots q_{Jt}!}\right) \prod_{j=1}^{J+1} \left[\int_{\nu} s_{ijt}(\omega_{2,jt}, \nu_i)\phi(\nu)d\nu\right]^{q_{jt}}. \tag{8}$$

Let q_{jt} denote the observed chosen quantity of product j at time t. Because the error terms $\omega_{2,t}$ are not observed, Park & Gupta (2009) assume the error terms to be i.i.d. standard normally distributed and the authors integrate out those EUPC terms.

$$L_{1,t} = \int L_{1,t}(\omega_{2,t})\phi(\omega_2)d\omega_2. \tag{9}$$

Because of this integration, ordinarily only a subsample of the observed sold quantities can be used, e.g., 100 draws - not all of the information about the sold quantity in the data. Otherwise, the computation quickly reaches machine zero because the observed choice quantities enter the likelihood through the exponent (see equation (8)) and because the logarithmic transformation does not help to avoid as usual the computation of the exponentiated probabilities. Although Park & Gupta (2009) state that the estimation is not sensitive to that subsampling, it is generally desirable to use the full information of the data, if possible. This is especially important when one wants to recover the parameter values of a model with a complex setup. However, when integrating out ω_2, one has to additionally integrate over the dimension of J, the number of alternatives, which can be high. Moreover, because ν_i from equation (3) needs to be integrated out with a number of ns draws, the total number of evaluations to numerically integrate out both ν_i and ω_2 is $ns \cdot ns_2$, if ns_2 is the number of draws needed for integrating out ω_2. That increases the burden of computation.

In summary it is desirable to use the full available information from the data set and to

alleviate the burden of estimation, if possible.

The approach introduced here is to sample ω_2 through a random walk MCMC chain, which does not require a second numerical integration and which permits use of the full information of the data. Finally, the distribution of ω_2 does not need to be known, but it can be automatically estimated, if desired. In the following, the likelihood is written conditional on ω_2 and conditional on the bias correction term $[x_{jt} - Z_{jt}\zeta_j]$ which is then drawn in the Bayesian estimation. Let the conditional likelihood function be denoted as:

$$L(\theta|\omega_2) = \prod_{t=1}^{T} L_{1,t}(\omega_{2,t}).$$

With $\theta = (\beta, \gamma_{11}, ..., \gamma_{KK}, b_{11,1}, b_{21,1}, b_{22,1}, ..., b_{11,J}, b_{21,J}, b_{22,J})$. Therefore, except for a constant, the conditional log likelihood is:

$$l(\theta|\omega_2) = \sum_{t=1}^{T}\sum_{j=1}^{J+1} q_{jt} log\left(\int_{\nu} s_{ijt}(\omega_{2,jt}, \nu_i)\phi(\nu)d\nu\right).$$

The bias correction term $[x_{jt} - Z_{jt}\zeta_j]$ - that is the error term from the first-stage regression of the endogenous variable(s) on the instrumental variables - is not known. We can either include the bias correction term in the likelihood as the probability from the K dimensional normal distribution (Park & Gupta 2009) and estimate it simultaneously, or we can estimate the bias correction term in advance (Petrin & Train 2002). The latter has the advantage of simplifying the estimation procedure and reducing the number of parameters in the likelihood estimation respectively Bayesian estimation. That might be particularly important if the number of instrumental variables is very large as a result of interactions with dummy variables, as in some practical applications.

2.3 Priors

The standard choice for the priors is[9]:

$$\begin{aligned} \beta &\sim N(\beta_0, V_\beta), \\ (\zeta_j &\sim N(\zeta_0, V_\zeta).) \end{aligned}$$

As pointed out previously, one can optionally assume, that:

$$\omega_{2,jt} \overset{i.i.d.}{\sim} N(0,1).$$

To obtain a more equal distribution of the correlations of both variance-covariance matrices $\tilde{\Sigma}_j$ and Σ, the prior distributions are set as in Jiang et al. (2009):

$$\begin{aligned} \Sigma &= \Lambda\Lambda', \\ \Lambda &= \begin{pmatrix} e^{r_{11}} & 0 & \cdots & 0 \\ r_{21} & e^{r_{22}} & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ r_{K1} & \cdots & r_{K,K-1} & e^{r_{KK}} \end{pmatrix}. \end{aligned}$$

Where the priors of r_{mk} are specified as

$$\begin{aligned} r_{mm} &\sim N(0, \sigma^2_{mm}) \\ r_{mk} &\sim N(0, \sigma^2_{off}), m = 1, ..., K \text{ and } k = 1, ..., K, m > k. \end{aligned}$$

Analogously for $\tilde{\Sigma}_j$(same for all $j = 1, ..., J$):

$$\begin{aligned} \tilde{r}_{ll} &\sim N(0, \tilde{\sigma}^2_{ll}) \\ \tilde{r}_{lh} &\sim N(0, \tilde{\sigma}^2_{off}), l = 1, ..., J \text{ and } h = 1, ..., J, l > h. \end{aligned}$$

[9]If the estimation of the bias correction term is done separately before the full estimation procedure, as it is done here, the prior of ζ_j is not needed in the following.

The standard choice of hyper-parameters is:

$$\begin{aligned}
\beta_0 &= 0, V_\beta = 100 I_K, \\
\sigma^2_{mm} &= \frac{1}{4}\left(\frac{1+\sqrt{1-4(2(m-1)\sigma^4_{off}-c)}}{2}\right), \\
\tilde{\sigma}^2_{ll} &= \frac{1}{4}\left(\frac{1+\sqrt{1-4(2(l-1)\tilde{\sigma}^4_{off}-c)}}{2}\right), \\
\sigma^2_{off} &= 1, \tilde{\sigma}^2_{off} = 1, c = 50.
\end{aligned}$$

For the hyper-parameter of the variance-covariance matrices see Jiang et al. (2009) and otherwise, e.g., Rossi, Allenby & McCulloch (2005).

2.4 Bayesian estimation

With the inversion of the market shares that requires the contraction mapping according to Berry (1994), the parameters can be estimated with a hybrid MCMC algorithm as in Jiang et al. (2009).

They proceed as following. Given the draws for the Choleski factor of the variance-covariance matrix of the random coefficients Λ, the mean utility δ_{jt} can be computed by the contraction mapping.

After that, one can conduct a Bayesian linear instrumental variable regression of this mean utility on the influence variables and instrumental variables [10]. This permits an assessment of the structural error term ξ_{jt} and permits the establishment of the likelihood function with the assumed distribution of ξ_{jt}. The parameters of Λ are gained through a step of a random walk Metropolis chain.

The advantage of the procedure of Jiang et al. (2009) is, that one can easily obtain draws from all parameters other than those of Λ, while using information from the model set up

[10] This Bayesian linear instrumental variable regression can be conducted as e.g. outlined in Rossi et al. (2005). An adapted description for the case and the parameter notation used here can also be requested from the author.

and from the observed shares to obtain these draws.

A disadvantage of this approach is that for every iteration of the MCMC algorithm the contraction mapping[11] must be conducted, which leads to a three times larger estimation time than with the GMM approach, after Jiang et al. (2009). Additionally, the Jacobian of the shares according to ξ_{jt} must be computed.

In this paper, the likelihood is evaluated without conducting the contraction mapping. I use a method similar to that employed in Park & Gupta (2009) to calculate the likelihood function, but I apply a full random walk MCMC algorithm for all parameters θ in this model.

The approach is to sample ω_2 through a random walk MCMC chain, which (i) omits a second numerical integration, (ii) allows us to use the full information of the data and (iii) does not require us to assume of specific distribution of ω_2, if not desired. The posteriori probability for given ω_2 is:

$$
\begin{aligned}
\pi(\theta|s_t, X_{t_{t=1}}^{T}, \omega_2) &= \\
&\propto L(\theta|\omega_2)\pi(\theta) \\
&= L(\theta|\omega_2) \times \\
&\quad \times |V_\beta|^{-1/2} \, exp\left(-0.5(\beta-\beta_0)'V_\beta^{-1}(\beta-\beta_0)\right) \\
&\quad \times \prod_{m=1}^{K} exp\left(-\frac{r_{mm}^2}{2\sigma_{mm}^2}\right) \times \prod_{m=2}^{K}\prod_{k=1}^{j-1} exp\left(-\frac{r_{mk}^2}{2\sigma_{off}^2}\right) \\
&\quad \times \prod_{j=1}^{J+1}\left[\prod_{l=1}^{J} exp\left(-\frac{\tilde{r}_{ll}^2}{2\tilde{\sigma}_{ll}^2}\right) \times \prod_{l=2}^{J}\prod_{h=1}^{l-1} exp\left(-\frac{\tilde{r}_{lh}^2}{2\tilde{\sigma}_{off}^2}\right)\right].
\end{aligned}
$$

There are two random walk MCMC chains, that are iterated - one for updating θ and one for updating ω_2:

[11] That is the same contraction mapping procedure that is conveyed in the inner loop of the generalized method of moments (GMM) estimation of the BLP model.

(1) For given ω_2 the MCMC chain for θ is updated in this way:

$$\theta^{new} = \theta^{old} + \tau_\theta, \tau_\theta \sim N(0, \sigma^2 D_\theta).$$

Where σ^2 is a scaling constant and D_θ is the candidate variance-covariance matrix that is typically obtained from a calibration chain.

A new draw of θ is accepted with the probability α:

$$\alpha = min\left\{1, \frac{L(\theta^{new}|\omega_2)\pi(\theta^{new})}{L(\theta^{old}|\omega_2)\pi(\theta^{old})}\right\}.$$

(2) For given θ the random walk MCMC chain for ω_2 is updated as:

$$\omega_2^{new} = \omega_2^{old} + \tau_{\omega_2}, \tau_{\omega_2} \sim N(0, \sigma^2_{\omega_2} D_{\omega_2}).$$

Where $\sigma^2_{\omega_2}$ is the scaling constant and D_{ω_2} the candidate variance-covariance matrix for ω_2 which has the dimension of the number of alternatives J.

A new draw of ω_2 is accepted with the probability α_{ω_2}:

$$\alpha_{\omega_2} = min\left\{1, \frac{L(\omega_2^{new}|\theta)\pi(\omega_2^{new})}{L(\omega_2^{old}|\theta)\pi(\omega_2^{old})}\right\}. \quad (10)$$

3 Simulation study

3.1 Data generating process (DGP)

In this simulation study, I consider J=2 products (and an outside option) on T=50 markets. Thus there are $J \times T = 100$ observations. The data are generated as follows, in a manner similar to Park & Gupta (2009). Hence the way that endogeneity enters a certain variable, say price, is explicitly modeled:

$$\begin{aligned}
x_{jt}^{(1)} &= u_{jt}^{(1)} \cdot u_{jt}^{(2)} \text{ (Variable 1)} \\
u_{jt}^{(1)} &\overset{i.i.d.}{\sim} N(0,1) \\
u_{jt}^{(2)} &\overset{i.i.d.}{\sim} U(0,1) \\
x_{jt}^{(2)} &\overset{i.i.d.}{\sim} U(0,1) \text{ (Variable 2)} \\
x_{jt}^{(3)} &= Z_{jt}'\gamma_j + \tilde{\nu}_{jt} \text{ (an endogenous variable, say price)} \\
\omega_{1,jt}, \omega_{2,jt} &\overset{i.i.d.}{\sim} N(0,1).
\end{aligned}$$

Through the setting from equation (5) to account for endogeneity, there is a correlation between $x_{jt}^{(3)}$ and the structural error term ξ_{jt}, but ξ_{jt} also exhibits an additional variation component.

$b_{11} = \sqrt{.1/n_{iv}}$ and the $1+n_{iv}$ instrumental variables are set as the following, where $n_{iv} = 10$:

$$\begin{aligned}
Z_{jt} &= (2, z_{jt,1}, ..., z_{jt,n_{iv}}), \\
z_{jt,l} &\overset{i.i.d.}{\sim} N(0, .9/n_{iv}), l = 1, 2, ..., n_{iv} \\
b_{21} &= \sqrt{.5/n_{iv}}, b_{22} = \sqrt{.5/n_{iv}}.
\end{aligned}$$

Where γ_j is a column vector of values 1/2.

Thus, along with the two brand dummy variables d_j, there are five influence vari-

ables:

$(d_1, d_2, x_{jt}^{(1)}, x_{jt}^{(2)}, x_{jt}^{(3)})$ with mean parameter values determined as

$\beta = (-3.25, -3.5, .5, 1, -1)$.

The variance-covariance Σ of the coefficients of the five influence variables is specified as the following, to assure a model with different substitution patterns than those from the homogeneous logit (similar to Jiang et al. (2009)):

$$\Sigma = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1.5 & -1.5 \\ 0 & 0 & 1.5 & 2 & -1.5 \\ 0 & 0 & -1.5 & -1.5 & 2 \end{pmatrix}$$

To computed an observed aggregated market share in the data generating process, I take the average of the simulated decisions of 100,000 random sampled customers [12][13].

3.2 Estimation with an alternative approach

After the data were generated, I obtained first a candidate variance-covariance matrix for the MCMC algorithm by performing a calibration chain with some initial starting values[14]. The random walk sampling matrix[15] for θ is set as a diagonal matrix with the entries of

[12] With e.g. 20,000 random sampled customers the following results were quite similar.

[13] To check the robustness of the simulation study toward the parameter setup, different data generating parameters as shown below were used. The proposed estimation approach recovered the parameter values with an accuracy similar to that in the previously outlined parameter setup.

$$\begin{aligned} b_{11} &= 2\sqrt{.1/n_{iv}}, \quad b_{21} = 2\sqrt{.1/n_{iv}}, \quad b_{22} = b_{21} \\ \beta &= (-2, -3, 2, 2, -5) \\ \Sigma &= \begin{pmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 3 & .5 & .5 \\ 0 & 0 & .5 & 3 & -.5 \\ 0 & 0 & .5 & -.5 & 3 \end{pmatrix} \end{aligned}$$

[14] The starting values are deliberately taken from a prior ad-hoc unprecise estimation.

[15] i.e. the initial candidate variance-covariance.

1/100 on the diagonal, and then I took 10,000 draws, with 6,000 of them as burn-in. After that, every draw is kept and the mean of the draws is used as the starting value for the final MCMC chain. The variance-covariance matrix of the selected draws is employed as the candidate variance-covariance. Moreover, the scaling factor is set to the usually recommended value of $2.3/\sqrt{\text{number of parameters in } \theta}$ (Roberts & Rosenthal 2009), which is divided by 10. The same is done for the candidate variance-covariance matrix of ω_2. Finally, 20,000 draws are taken, with 20% of them as burn-in draws, i.e., 4,000.

To compute the simulated likelihood function, I used the nested Gauss-Quadrature with nodes and weights on sparse grids implemented and developed by Heiss & Winschel (2006), as mentioned earlier. The accuracy level is set to 6, which leads to ns=993 nodes for eight parameters for the integration of equation (2). This means that the approximation is exact up to a polynomial of degree 6+1.

3.3 Estimation with the reference model

The Bayesian estimation with the contraction mapping is conducted, as in Jiang et al. (2009), thus providing a comparison with the presented approach. The DGP, the overall procedure and the prior distributions as well as hyperparameters are the same as before, except the prior for the variance-covariance matrix $\tilde{\Sigma}_j$ [16], but now the likelihood function and posterior are obtained as in Jiang et al. (2009).

For every set of parameters Λ, the contraction mapping must be executed and Jacobian matrix must be computed, so this sampling procedure is considerably slower than the introduced alternative approach. The estimation time in general depends critically on the number of simulation draws for the simulation of the expected market share (Dube et al. 2008). In all three estimation methods discussed here, the same number of simulation draws for obtaining the market share is applied, but when employing the contraction

[16] To apply, as in Jiang et al. (2009) the usual Bayes linear instrumental variable regression as prior for $\tilde{\Sigma}_j$ a standard inverted Wishart distribution is employed.

mapping on average roughly about 70 to 100 iterations are conducted in this setting. Thus additionally 70 to 100 times more time is needed to simulate the expected market share. As a stopping criterion for the contraction mapping I use (similarly to Nevo (2000)) the condition that the mean respectively maximum of two consecutive values of δ is smaller than 1e-12 respectively 1e-15 to employ tight tolerance levels as recommended by Dube et al. (2008) [17].

Furthermore, for comparison, the parameters are also estimated with the maximum likelihood estimation from Park & Gupta (2009). In each replication of the data generation and estimation, a subsample of 100 choices is randomly drawn before the estimation to be able to numerically compute the likelihood[18] from equation (9), as recommended in Park & Gupta (2009). To integrate out the $J = 2$ dimensional EUPC term ω_2, a second grid of $ns_2 = 17$ values of the Gauss-Quadrature on sparse grids is used. The accuracy level is set to 4, which indicates that the approximation is exact up to a polynomial of degree 4+1.

3.4 Results

The DGP and the estimation were executed 50 times. The resulting biases and mean squared errors (MSE) are presented in table I [19].

[17] Dube et al. (2008) determine the effect of the stopping criterion in detail and recommend using tight tolerance levels, i.e. stopping the contraction mapping if the Euclidean norm of two consecutive values from this mapping is e.g. less than 1e-14. The applied tolerance level in the estimation thus is in the recommended range.

[18] The full likelihood is here the product of equation (9) over all observations $t = 1, ..., T$, i.e., $L = \prod_{t=1}^{T}(L_{1,t})$ and the first-stage regression of x_{jt} on Z_{jt} is conducted separately before the likelihood estimation. If the first-stage regression was estimated simultaneously, the obtained parameter estimates were less accurate.

[19] Because of parsimony the parameter estimates from the auxiliary instrumental variable regression are not presented.

TABLE I

Results from the maximum likelihood estimation, Bayesian estimation without contraction mapping and the alternative approach

			Likelihood estimation: with subsampling			Bayesian estimation: with contraction mapping			Bayesian estimation: alternative approach		
Description	Variable	True value	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
Product dummy 1	d_1	-3.25	-4.44	-1.19	4.68	-3.24	0.01	0.21	-3.25	0.00	0.15
Product dummy 2	d_2	-3.50	-4.98	-1.48	7.39	-3.41	0.09	0.20	-3.52	-0.02	0.17
Variable 1	$x_{jt}^{(1)}$	1.50	2.06	0.56	1.47	1.52	0.02	0.04	1.43	-0.07	0.04
Variable 2	$x_{jt}^{(2)}$	1.00	1.31	0.31	0.69	1.01	0.01	0.06	0.98	-0.02	0.04
Price	$x_{jt}^{(3)}$	-1.00	-1.61	-0.61	1.57	-1.24	-0.24	0.16	-1.07	-0.07	0.08
Covariance	r_{11}	0.35	0.38	0.04	2.84	0.23	-0.12	0.20	0.30	-0.05	0.10
Covariance	r_{22}	0.35	0.57	0.23	1.33	0.14	-0.21	0.21	0.33	-0.02	0.11
Covariance	r_{33}	0.35	0.65	0.30	0.28	0.32	-0.03	0.02	0.31	-0.03	0.02
Covariance	r_{43}	1.06	1.42	0.36	1.10	0.90	-0.16	0.14	0.97	-0.09	0.12
Covariance	r_{53}	-1.06	-1.40	-0.33	1.04	-0.99	0.07	0.04	-0.93	0.13	0.06
Covariance	r_{44}	-0.07	-1.49	-1.42	14.57	-0.62	-0.56	0.57	-0.74	-0.67	0.76
Covariance	r_{54}	-0.40	-0.98	-0.58	1.16	-0.66	-0.26	0.24	-0.58	-0.18	0.23
Covariance	r_{55}	-0.17	-0.71	-0.54	3.81	-0.16	0.00	0.13	-0.28	-0.11	0.16
Endogeneity	$\tilde{r}_{11,1}$	0.10	-1.97	0.33	0.11	0.22	0.12	0.01	0.10	0.00	0.00
Endogeneity	$\tilde{r}_{21,1}$	0.22	0.27	0.05	0.02	0.17	-0.06	0.00	0.21	-0.02	0.00
Endogeneity	$\tilde{r}_{22,1}$	0.22	0.22	1.72	3.00	0.34	0.12	0.02	0.26	0.04	0.00
Endogeneity	$\tilde{r}_{11,2}$	0.10	-2.37	-0.07	0.00	0.22	0.12	0.02	0.10	0.00	0.00
Endogeneity	$\tilde{r}_{21,2}$	0.22	0.29	0.06	0.05	0.17	-0.06	0.00	0.23	0.00	0.00
Endogeneity	$\tilde{r}_{22,2}$	0.22	0.19	1.69	2.91	0.35	0.12	0.02	0.28	0.06	0.01
Goodness of fit											
	mape		0.63			0.12			0.08		
	meape		0.36			0.12			0.04		
	maape		1.72			0.56			0.67		
	mse		0.70			0.03			0.03		

mape: mean absolute percentage error; meape: median absolute percentage error; maape: maximum absolute percentage error;mse: mean squared error; keep: every 100th draw was kept.

It can be seen, that with the alternative approach the parameters are captured with satisfying accuracy. The bias and MSE are in the range of the values of Jiang et al. (2009) and, in general, even slightly more precise (See goodness of fit of parameter estimates in Table (I)). It can be seen that the alternative approach is well able to recover the parameters from the DGP.

4 Conclusion

Conclusion

In summary, the proposed simple alternative estimation approach leads to accurate results, and the approach is shown to be capable of estimating the discussed model with a simpler estimation procedure and with less computational effort. In applied work, the reduced computation time is particularly helpful when many model specifications need to be estimated and analyzed. The introduced approach reduces the computation time by a factor roughly proportional to the iterations needed in the contraction mapping[20]. In general the number of draws used for the simulation of the market share integral predominantly determines the overall computation time[21]. For 1,000 iterations of the MCMC chain, the alternative approach needs 2,001 evaluations of that integral, whereas the procedure in Jiang et al. (2009) required about 95,000 evaluations for that chain in a test run. This means that the alternative approach is about 45 times faster in computation.

Moreover, in contrast to Park & Gupta (2009), the procedure enables us to use all of the information about sold quantities available in the data set.

Procedures as Jiang et al. (2009) and Park & Gupta (2009), in contrast to the GMM estimation of the BLP model Berry et al. (1995), rely on the additional assumption that the surrogate of unknown product characteristics ξ is normally distributed. More precisely, the assumption in Park & Gupta (2009) is that the residual $\tilde{\nu}$, for the mapping of price on its instrumental variable, and ξ are jointly normally distributed as in equation (5).

In fact, in the proposed approach, the distributional assumption of ω_1 and ω_2 is mainly irrelevant. In the first-stage regression of price on its instruments, the assumption of normality of ω_1 is not necessarily required for the validity of the OLS method. Otherwise, the assumption of normality of ω_2 is used only in the acceptance probability for ω_2 in equation (10), which can be easily be removed, if desired. Furthermore, the proposed

[20] The iterations needed in the contraction mapping are in a range of 70 to 100 iterations in this setting

[21] That is the case for a GMM estimation, a maximum likelihood estimation, a Bayesian estimation as in Jiang et al. (2009) and the approach discussed here.

estimation approach allows for a very flexible distribution of ω_2, as a result of the MCMC random walk estimation.

The assumption made here is that $\xi = b_{21}\omega_1 + b_{22}\omega_2$. Therefore, the specific distribution of ξ can be assessed freely through ω_2, and it does not need to be determined a priori.

Limitation and Outlook

The cost of the introduced approach is, that the dimension of the first random walk chain is larger than in Jiang et al. (2009), which may complicate the tuning of the random walk chain. Beyond that, a second random walk chain has to be tuned. This additional cost is comparable low, because the second candidate variance-covariance matrix has, by definition, only the dimension of the number of choice alternatives and because the candidate matrix can be scaled to obtain a reasonable acceptance rate.

Further investigation is needed to study how the supply side of the BLP model, as in Berry et al. (1995), can be introduced in a likelihood-based model and estimated.

Contribution

In summary, the introduced simple approach contributes in three ways:

(i) In contrast to the GMM estimation and the Bayesian estimation method in Jiang et al. (2009), the computation-intensive contraction mapping can be omitted. This results in a simpler estimation procedure and in an accelerated computation. This speed advantage becomes more important in practical applications, where many estimations must be conducted for model specification and testing, and in the handling of larger data sets containing many choice alternatives, a greater number of observations (markets) or random coefficients. In the simulation study discussed here, the proposed estimation was roughly about 45 to 50 times faster than the reference estimation as in Jiang et al. (2009).

(ii) In contrast to the maximum likelihood estimation in Park & Gupta (2009), a second numerical integration is left out and the full information of the data set about the market

share is used. Thus, subsampling resulting from the lack of computational precision is avoided.

(iii) Finally in contrast to estimation methods in Park & Gupta (2009) and Jiang et al. (2009), the introduced approach does not necessarily require a normally distributed demand shock ξ, but the approach does allow for it and does allow its distribution to be automatically assessed similar to the GMM estimation in Berry et al. (1995).

References

Albuquerque, P. & Bronnenberg, B. J. (2008), 'Market areas of car dealerships'.

Berry, S. T. (1994), 'Estimating discrete choice models of product differentiation', *RAND Journal of Economics* **25**(2), 242–262.

Berry, S. T., Levinsohn, J. & Pakes, A. (1995), 'Automobile prices in market equilibrium', *Econometrica* **63**(4), 841–890.

Bhat, C. R. (2000), 'Quasi-random maximum simulated likelihood estimation of the mixed logit model', *Transportation Research* **Part B**(35), 677–693.

Dube, J., Fox, J. & Su, C. (2008), 'Improving the numerical performance of blp static and dynamic discrete choice random coefficients demand estimation'.

Gowrisankaran, G. & Rysman, M. (2009), Dynamics of consumer demand for new durable goods. Working Paper.

Hausman, J. (1954), 'Specification tests in econometrics', *Econometrica* **46**(3), 1251–1271.

Heckman, J. J. (1978), 'Dummy endogenous variables in a simultaneous equation system', *Econometrica* **46**(6), 931–959.

Heiss, F. & Winschel, V. (2006), Estimation with numerical integration on sparse grids. Discussion paper 2006-15.

Jiang, R., Manchanda, P. & Rossi, Peter, E. (2009), 'Baysian analysis of random coefficients logit models using aggregate data', *Journal of Econometrics* **149**(2), 136–148.

Judd, K. L. (1998), *Numerical Methods in Economics*, MIT Press, Cambridge, Mass.

Nevo, A. (2000), 'A practitioner's guide to estimation of random coefficients logit models of demand', *Journal of Economics & Management Strategy* **9**(4), 513–538.

Park, S. & Gupta, S. (2009), 'Simulated maximumlikelihood estimator for the random coefficient logit model using aggregate data', *American Marketing Association* **46**, 531–542.

Petrin, A. & Train, K. (2002), Omitted product attributes in discrete choice models, Working paper, University of Chicago. December 16.

Roberts, G. O. & Rosenthal, J. S. (2009), 'Optimal scaling for various metropolis-hastings algorithms', *Statistical Science* **16**(4), 351–367.

Rossi, P. E., Allenby, G. M. & McCulloch, R. (2005), *Bayesian Statistics and Marketing*, Wiley Series in Probability and Statistics, West Sussex.

Sovinsky Goeree, M. (2008), 'Limited information and advertising in the u.s. personal computer industry', *Econometrica* **76**(5), 10171074.

Train, K. E. (2000), Halton sequences for mixed logit, Working paper, Department of Economics, University of California, Berkley.

Walker, L. J., Ben-Akiva, M. & Bolduc, D. (2007), 'Identification of parameters in normal error component logit-mixture (neclm) models', *Journal of Applied Econometrics* **22**(6), 1095–1125.