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, the so called BLP model, named after (Berry, Levinsohn & Pakes 1995).

The estimation is conducted through a bayesian estimation similar to (Jiang et al. 2009). But in difference to them here the time intensive contraction mapping for assessing the mean utility in every iteration step of the estimation procedure is not needed. This is because the likelihood function is computed via a special case of the control function method ((Petrin & Train 2002) and (Park & Gupta 2009)) and hence a full random walk MCMC algorithm is applied. In difference to (Park & Gupta 2009) the uncorrelated error, which is explicitly introduced through the control function procedure, is not integrated out, but sampled with a random walk MCMC. The introduced proceeding enables to use the whole information from the data set in the estimation and beyond that accelerates the computation.

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

JEL Classification Numbers: .

*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] I want to discuss a simple alternative estimation method of the multinomial logit model for aggregated data, as developed by (Berry et al. 1995), the so called BLP model.

The estimation is conducted with a Bayesian estimation similar to (Jiang et al. 2009). But in difference to them here the time intensive contraction mapping for assessing the mean utility in every iteration step of the estimation procedure is not needed. This is because the likelihood function is computed via a special case of the control function method ((Petrin & Train 2002) and (Park & Gupta 2009)) and hence a full random walk MCMC algorithm is applied. In difference to (Park & Gupta 2009) the uncorrelated error, which is explicitly introduced through the control function procedure, is not integrated out, but sampled with a random walk MCMC. The introduced proceeding enables to use the whole information from the data set in the estimation and beyond that accelerates the computation. In the remaining of this paper is structured as following. First in the next section the model setup is outlined, both for the usual BLP model and the estimation via the Bayesian method, which includes the derivation of the likelihood function, the prior and posterior distribution. After that in the following section a simulation study is employed to assess the performance of the introduced alternative estimation approach. These results are compared to the outcome the estimation according to (Jiang et al. 2009) as a reference. The text ends with referring to limitations and 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 Master and PhD respectively doctoral students in the winter semester 2009/10.

2 Random coefficient logit model for aggregated data

2.1 General model

In the following only the demand side is considered of a multinomial choice model for aggregated data, as developed by (Berry et al. 1995), the so called BLP model.

For marketing issues this framework can be applied for example to scanner data about consumer goods.

This simulation concentrates on the case where the unobserved individual preferences are normal distributed[2], which is the most relevant case for applications of the model (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 individal i=1,...,I in market t=1,...,T hence can be written as:

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

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

ε_{ijt} is an i.i.d. extreme value distributed error term and ξ_{jt} is the surrogate of unknown product characteristics[3]. $\boldsymbol{\Pi}$ are the influence parameters of the (matrix of) demographic variables $\boldsymbol{D_i}$. Moreover the mean value of utility of product j in market t is $\boldsymbol{\delta_{jt}} \equiv \boldsymbol{x_{jt}}\boldsymbol{\beta} + \xi_{jt}$ and thus $\boldsymbol{\mu_{ijt}} \equiv \boldsymbol{x_{jt}}(\boldsymbol{\Pi}\boldsymbol{D_i} + \boldsymbol{\Lambda}\boldsymbol{\nu_i})$ is the individual specific deviation from $\boldsymbol{\delta_{jt}}$.

[2] Although in the BLP model any other distribution for individual heterogeneity can be applied.

[3] These product characteristics are unknown by the data analyst, but known by the costumer and seller. ξ_{jt} is also called unobserved demand shock or structural error term.

$\mathbf{\Lambda}$ is the lower-triangular Cholesky factor of the $K \times K$ covariance matrix $\mathbf{\Sigma}$ so that $\mathbf{\Sigma} = \mathbf{\Lambda} \cdot \mathbf{\Lambda}'$ and $\boldsymbol{\beta_i} \sim N(\boldsymbol{\beta} + \mathbf{\Pi} \boldsymbol{D_i}, \mathbf{\Sigma})$.

Define[4]

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

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

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

can be approximated by different simulation methods.

Where $\boldsymbol{D_i} \sim P_{\boldsymbol{D}}(\boldsymbol{D})$ is the distribution of the demographic characteristics and $\phi(\cdot)$ denotes the density function of the standard normal distribution.

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

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

with ns[5] random or decisively chosen draws from the assumed distribution of $\boldsymbol{\nu}$ and for given data observations from the empirical distribution of $\boldsymbol{D}$ with some appropriately chosen weights w_i.

Later in the simulation study (section (3)) I use for the simulation the nested Gauss-Quadrature with nodes and weights on sparse grids as implemented and developed by (Heiss

[4]Define $\mathbf{\Lambda}$ with respect to possible exclusion restrictions for identification (Walker, Ben-Akiva & Bolduc 2007).

[5]ns stands for "number of simulations".

& Winschel 2006) and going back to (Smolyak 1963).

This method has the advantage over applying the product rule to the usual Gauss-Quadrature, e.g. in (Judd 1998), that due to the construction of sparse grids far less nodes are needed for the integration of integrals of higher dimension.

The numerical integration with Gauss-Quadrature on sparse grids leads to a relative fast estimation compared to other simulation methods such as simple random draws from the normal distribution or randomized draws from Halton Sequences (e.g. (Bhat 2000)) [6]; since the overall computation time depends mainly on the magnitude of ns (Dube, Fox & Su 2008). For all estimation methods of the BLP model, which require the simulation of the market share from equation (2) [7] and especially for both - the approach discussed here and a Bayesian estimation as in (Jiang et al. 2009) - the advantage of the nested Gauss-Quadrature cannot be taken, if also demographic variables enter the utility (in μ_{jt} from equation (1)). This is because otherwise the simulated individuals would have specific weights which are not connected with values of their demographic variables. Thus that would only make sense, if this assumption was reasonable.

In the following demographic variables are not considered and therefore $\boldsymbol{D}$ is dropped in equations.

2.2 Likelihood

In the likelihood function endogeneity is explicitly incorporated through a setup as (Park & Gupta 2009). Note that the derivation of the likelihood is strongly based on the theory of the cited paper, which goes back to (Hausman 1954) and (Heckman 1978). The approach uses a special case of the more general control function estimation method of (Petrin &

[6] (Train 2000) obtains similar results at the simulation with 1000 random draws and 100 draws from Halton Sequences.

[7] Which includes also the GMM estimation from (Berry et al. 1995).

Train 2002).

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

z_{jt} is the vector of instrumental variables which are correlated with the known product characteristics x_{jt} or e.g. only price[8], but uncorrelated with the unknown product characteristics ξ_{jt}.

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

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

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

$$
\Sigma_j = \begin{pmatrix} b_{11,j} & 0 \\ b_{21,j} & b_{22,j} \end{pmatrix} \begin{pmatrix} b_{11,j} & 0 \\ b_{21,j} & b_{22,j} \end{pmatrix}' = \begin{pmatrix} \Sigma_{\tilde{\nu}_j} & \Sigma_{\xi_j,\tilde{\nu}_j} \\ \Sigma_{\tilde{\nu}_j,\xi_j} & \sigma^2_{\xi_j} \end{pmatrix}. \qquad (6)
$$

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

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

$$u_{ijt} = x'_{jt}\beta_i + b_{21,j}\omega_{1,jt} + b_{22,j}\omega_{2,jt} + \varepsilon_{ijt}, \text{ and} \tag{7}$$

$$x_{jt} = Z_{jt}\zeta_j + b_{11,j}\omega_{1,jt}. \tag{8}$$

After plugging $\omega_{1,jt}$ from equation (8) in equation (9), the utility can be written as:

$$u_{ijt} = x'_{jt}\beta_i + \varrho'_j[x_{jt} - Z_{jt}\zeta_j] + b_{22,j}\omega_{2,jt} + \varepsilon_{ijt}, \tag{9}$$

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

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

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 hence 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 shocks $\omega_{2,jt}$ (EUPC).

Given the bias correction term $[x_{jt} - Z_{jt}\zeta_j]$ and conditional on $\omega_{2,jt}$ the likelihood can be written as:

$$L_{1,t}(\omega_{2,t}) = \left(\frac{q_t!}{q_{0t}! \ldots q_{Jt}!}\right) \prod_{j=1}^{J+1} \left[\int_{\nu} s_{ijt}(\omega_{2,jt}, \nu_i)\phi(\nu)d\nu\right]^{q_{jt}}. \tag{10}$$

The observe quantities of chosen products j at t time t are denoted by q_{jt}.

$\omega_{2,t}$ is standard normal distributed and unknown, therefore it can be integrated out as in

(Park & Gupta 2009):

$$L_{1,t} = \int L_{1,t}(\omega_{2,t})\phi(\omega_2)d\omega_2. \tag{11}$$

Since $\omega_{2,t}$ is not observed, one way is to integrate the error terms EUPC out. Because of this integration usually only a sample of the observed sold quantities, e.g. 100 draws, can be used and not the full information in the data. Since otherwise the computation reaches quickly machine zero, due the fact, that the observed choice quantities enter the likelihood through the exponent (see equation (10)) and the logarithmic transformation does not help to avoid the computation of the exponentiated probabilities as usual. Even though (Park & Gupta 2009) state, that the estimation is not sensitive to that sampling, it is on the one hand in general 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 somehow complex setup. On the other hand when integrating out ω_2 one has to integrate over the dimension of J, the number of alternatives which can be high per se. Additionally since 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 would is $ns \cdot ns_2$, if ns_2 is the number of draws needed for integrating out ω_2. Thus is procedure might limit the available information form the data set and aggravate the burden of the estimation.

The here used approach is to sample ω_2 through a random walk MCMC chain, which does not require a second numerical integration and furthermore allows to use the full information of the data. But here in the following the likelihood is written dependent on $\omega_{2,t}$ (and dependent on bias correction term $[x_{jt} - Z_{jt}\zeta_j]$), which is then drawn in the Bayesian estimation.

That is why the total likelihood function is:

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

With $\theta = (\boldsymbol{\beta}, \gamma_{11}, ..., \gamma_{K,K}, \boldsymbol{b_{11,1}}, \boldsymbol{b_{21,1}}, \boldsymbol{b_{22,1}}, \boldsymbol{...}, \boldsymbol{b_{11,J}}, \boldsymbol{b_{21,J}}, \boldsymbol{b_{22,J}})$.

Therefore log likelihood is except for a constant:

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

Because the bias correction term $[x_{jt} - Z_{jt}\zeta_j]$ - which is so to say the error term from the 1^{st} stage regression of the endogenous variable(s) on the instrumental variables - is not given we can either include it in the likelihood as the probability from the K dimensional normal distribution (Park & Gupta 2009) or estimate the bias correction term in advance (Petrin & Train 2002). It is an interesting issue, if the estimation method was more robust when the bias correction term would be simultaneous assessed. On the other hand estimating the bias correction term in advance simplifies the estimation procedure and especially reduces the number of parameters in the likelihood respectively Bayesian estimation. That might especially be important, if the number of instrumental variables is very large, due to interactions with dummy variables, as in some practical applications.

2.3 Priors

The priors are[9]:

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

As pointed out previously, it is assumed, that: (12)

$$\omega_{2,jt} \overset{i.i.d.}{\sim} N(0,1). \quad (13)$$

[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.

To obtain a more equal distribution of the correlations of both variance-covariance matrices $\tilde{\Sigma}_j$ and Λ, the prior distributions are set as in (Jiang et al. 2009) [10].

$$\Lambda = UU',$$

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

Where the priors of r_{mk} are specified as

$$r_{mm} \sim N(0, \sigma^2_{mm})$$

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

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

$$\tilde{r}_{ll} \sim N(0, \tilde{\sigma}^2_{ll})$$

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

[10] Frequencies of the correlations of elements of covariance-variance matrix from draws from the prior are shown in the appendix .

The standard choice of the hyper-parameters is (for the hyper-parameter of the variance-covariance matrices see (Jiang et al. 2009) and e.g. (Rossi, Allenby & McCulloch 2005)):

$$\begin{aligned}
\beta_0 &= 0, V_\beta = 100 I_K, \\
(\zeta_0 &= 0, V_\zeta = 100 I_K,) \\
\sigma^2_{mm} &= \frac{1}{4}\left(\frac{1+\sqrt{1-4(2(m-1)\sigma^4_{off}-c)}}{2}\right), \\
\sigma^2_{off} &= 1, \\
c &= 50, \\
\tilde{\sigma}^2_{ll} &= \frac{1}{4}\left(\frac{1+\sqrt{1-4(2(l-1)\tilde{\sigma}^4_{off}-c)}}{2}\right), \\
\tilde{\sigma}^2_{off} &= 1, \\
c &= 50.
\end{aligned}$$

2.4 Bayesian estimation

With the inversion of the market shares - the contraction mapping according to (Berry 1994) - the parameters can be estimated with a hybrid MCMC algorithm like 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 $\mathbf{\Lambda}$, the mean utility $\boldsymbol{\delta_{jt}}$ can be computed via the contraction mapping.

After that a Bayesian instrumental variable regression of this mean utility on the influence variables and instrumental variables can be conducted[11]. Through this they asses the structural error term ξ_{jt} and establish the likelihood function with the assumed distribution of ξ_{jt}. The parameters of $\mathbf{\Lambda}$ are gained through a step of a random walk Metropolis chain.

The advantage of the procedure of (Jiang et al. 2009) is, that they can easily obtain draws

[11] This Bayesian instrumental variable regression can be conducted as outlined in (Rossi et al. 2005), an adaption for the here used case and parameter notation can be requested from the author.

from all parameters, except those of $\mathbf{\Lambda}$, and use the information of the model setup and from the observed shares to obtain these draws.

A disadvantage of their approach is in my opinion, that for every iteration of the MCMC algorithm the contraction mapping - as the inner loop of the GMM estimation of the BLP model - has to be conducted, which leads after (Jiang et al. 2009) to a three times larger estimation time, than the GMM approach. Additionally the Jacobian of the shares according to ξ_{jt} has to be computed.

In approach of this paper the likelihood is evaluated without the conducting of the contraction mapping via a method similar to (Park & Gupta 2009) to compute the likelihood function and a full random walk MCMC algorithm for all parameters θ in the model, as discussed above.

The approach is to sample ω_2 through a random walk MCMC chain, which omits a second numerical integration and furthermore allows to use the full information of the data.

The posterior probability for given ω_2 is:

$$\begin{aligned}
\pi(\theta|s_t, X_{t_{t=1}}^T, \omega_2) &= \\
&\propto L(\theta|\omega_2)\pi(\theta) \\
&= L(\theta|\omega_2) \times \\
&\quad \times |V_\beta|^{-1/2} \, exp\left(-0.5(\beta-\beta_0)' V_\beta^{-1}(\beta-\beta_0)\right) \\
&\quad \times \prod_{m=1}^{K} exp\left(-\frac{r_{mm}^2}{2\sigma_{mm}^2}\right) \times \prod_{m=2}^{K}\prod_{k=1}^{j-1} exp\left(-\frac{r_{mk}^2}{2\sigma_{off}^2}\right) \\
&\quad \times \prod_{j=1}^{J+1}\left[\prod_{l=1}^{J} exp\left(-\frac{\tilde{r}_{ll}^2}{2\tilde{\sigma}_{ll}^2}\right) \times \prod_{l=2}^{J}\prod_{h=1}^{l-1} exp\left(-\frac{\tilde{r}_{lh}^2}{2\tilde{\sigma}_{off}^2}\right)\right]. \qquad (14)
\end{aligned}$$

There are two random walk MCMC chains, one for updating θ and the other one for updating ω_2.

(1) For given ω_2 the MCMC chain for θ is updated in the usual 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 the candidate variance-covariance matrix, 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 follows:

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

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 for the number of alternatives J.

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

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

3 Simulation study

3.1 Data generating process (DGP)

In this simulation study I consider J=2 products (and the outside option) on T=50 markets. The data is generated as follows, similar to (Park & Gupta 2009). This way how endogeneity enters a certain variable, say price, is explicitly modeled:

$$
\begin{aligned}
x_{jt}^{(1)} &= u_{jt}^{(11)} \cdot u_{jt}^{(12)} \text{ (Variable 1).} \\
u_{jt}^{(11)} &\overset{i.i.d.}{\sim} N(0,1) \\
u_{jt}^{(12)} &\overset{i.i.d.}{\sim} U(0,1) \\
x_{jt}^{(2)} &\overset{i.i.d.}{\sim} U(0,1) \text{ (Variable 2).} \\
x_{jt}^{(3)} &= \boldsymbol{Z}_{jt}'\boldsymbol{\gamma}_j + \nu_{jt} \text{ (an endogenous variable, say price).}
\end{aligned}
$$

To account for endogeneity the correlation of the error terms is:

$$
\begin{pmatrix} \nu_{jt} \\ \xi_{jt} \end{pmatrix} = \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} \omega_{1,jt} \\ \omega_{2,jt} \end{pmatrix}
$$

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

$$
j = 1, 2 \; ; \; t = 1, ..., T.
$$

Through this setting 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 $n_{iv} = 10$:

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

Where $\boldsymbol{\gamma}_j$ is a column vector of values $1/2$.

Thus there are with the two brand dummy variables d_j, five influence variables:
$(d_1, d_2, x_{jt}^{(1)}, x_{jt}^{(2)}, x_{jt}^{(3)})$ with mean parameter values determined as
$\boldsymbol{\beta} = (-3.25, -3.5, .5, 1, -1)$.
The covariance Σ of the coefficients of the five influence variables is specified as the following, to assure a model with different substitution patterns then 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 asses the aggregated market share (so to say the 'true simulated market share') I take in the data generating process the average of the simulated decisions of 100 000 random sampled costumers [12].

[12] With e.g. 20 000 random sampled costumers the following results are quite similar.

3.2 Estimation with an alternative approach

After the data is generated, first to obtain a candidate variance-covariance matrix for the MCMC algorithm a calibration chain is run with some initial starting values[13]. The random walk sampling matrix[14] for θ is set as a diagonal matrix with the entries of 1/100 on the diagonal and then 10 000 draws with 6 000 of them as burn-in are taken.
After that every draw is kept and the mean of the draws is used as the starting value for the final MCMC chain as well as 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 $2.3/\sqrt{\text{number of parameters in } \theta}$ 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 use the nested Gauss-Quadrature with nodes and weights on sparse grids implemented and developed by (Heiss & Winschel 2006) as discussed earlier. The accuracy level is set to 6, which leads to ns=993 nodes for eight parameters for the integration of equation (2) ; this involves that the approximation is exact up to a polynomial of degree 6+1.

3.3 Estimation with the reference model

As reference and comparison to the presented approach, the Bayesian estimation with contraction mapping is conducted as in (Jiang et al. 2009). The DGP, the overall procedure and the prior distributions as well as hyperparameters are the same as before, but now the likelihood function and posterior are obtained as in (Jiang et al. 2009).
Since here for every set of parameters Λ, the contraction mapping has to be executed and as well it is necessary to compute the Jacobian matrix, this sampling procedure is considerably

[13]Which are deliberately taken from a prior likelihood based estimation and are unprecise.
[14]i.e. the initial candidate variance-covariance.

slower than the earlier 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 both here discussed estimation methods the same number of simulation draws for obtaining the market share is applied, but when employing the contraction mapping on average roughly about 70-100 iterations are conducted in the setting here and thus additionally 70-100 times the time for simulating the expected market share is needed. As stopping criterion for the contraction mapping I use similar to (Nevo 2000) the condition that the mean respectively maximum of two consecutive values of δ are smaller than 1e-12 respectively 1e-15 to employ tight tolerance levels as recommended by (Dube et al. 2008) [15]

3.4 Results

The DGP and the estimation were executed 50 times. Bias and mean squared error (MSE) are presented in table (1). It can be seen, that the parameters[16] are captured with satisfying accuracy. The Bias and MSE are in the range of the values of e.g. (Jiang et al. 2009) and especially the parameters for the variance-covariance matrix are comparably well estimated. Even though the MSE of the parameter estimates with contraction mapping is slightly lower over the 50 replications, the means of the parameter estimates without contraction mapping on the other hand are slightly closer to the true values ('Goodness of fit of parameter estimates' in table (1)). In my personal experience the estimation accuracy depends mainly on the dimension and entries of Σ, which is the same issue for most estimation methods (i.e. GMM estimation, Bayesian estimation according to (Jiang et al. 2009) and the here discussed approach.).

[15] (Dube et al. 2008) determine the effect of the stopping criterion in detail and recommend to use tight tolerance levels, i.e. to stop 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.

[16] The parameter from the auxiliary instrumental variable regression are not shown here.

			With contraction mapping			Alternative approach		
Description	**Variable**	**True value**	**Mean**	**Bias**	**MSE**	**Mean**	**Bias**	**MSE**
Product dummy 1	d_1	**-3.25**	**-3.24**	0.01	0.21	**-3.25**	0.00	0.15
Product dummy 2	d_2	**-3.50**	**-3.41**	0.09	0.20	**-3.52**	-0.02	0.17
Variable 1	$x_{jt}^{(1)}$	**1.50**	**1.52**	0.02	0.04	**1.43**	-0.07	0.04
Variable 2	$x_{jt}^{(2)}$	**1.00**	**1.01**	0.01	0.06	**0.98**	-0.02	0.04
Price	$x_{jt}^{(3)}$	**-1.00**	**-1.24**	-0.24	0.16	**-1.07**	-0.07	0.08
Covariance	r_{11}	**0.35**	**0.23**	-0.12	0.20	**0.30**	-0.05	0.10
Covariance	r_{22}	**0.35**	**0.14**	-0.21	0.21	**0.33**	-0.02	0.11
Covariance	r_{33}	**0.35**	**0.32**	-0.03	0.02	**0.31**	-0.03	0.02
Covariance	r_{43}	**1.06**	**0.90**	-0.16	0.14	**0.97**	-0.09	0.12
Covariance	r_{53}	**-1.06**	**-0.99**	0.07	0.04	**-0.93**	0.13	0.06
Covariance	r_{44}	**-0.07**	**-0.62**	-0.56	0.57	**-0.74**	-0.67	0.76
Covariance	r_{54}	**-0.40**	**-0.66**	-0.26	0.24	**-0.58**	-0.18	0.23
Covariance	r_{55}	**-0.17**	**-0.16**	0.00	0.13	**-0.28**	-0.11	0.16
Endogeneity	$\tilde{r}_{11,1}$	**0.10**	**0.22**	0.12	0.01	**0.10**	0.00	0.00
Endogeneity	$\tilde{r}_{21,1}$	**0.22**	**0.17**	-0.06	0.00	**0.21**	-0.02	0.00
Endogeneity	$\tilde{r}_{22,1}$	**0.22**	**0.34**	0.12	0.02	**0.26**	0.04	0.00
Endogeneity	$\tilde{r}_{11,2}$	**0.10**	**0.22**	0.12	0.02	**0.10**	0.00	0.00
Endogeneity	$\tilde{r}_{21,2}$	**0.22**	**0.17**	-0.06	0.00	**0.23**	0.00	0.00
Endogeneity	$\tilde{r}_{22,2}$	**0.22**	**0.35**	0.12	0.02	**0.28**	0.06	0.01
Goodness	**mape**		0.125			0.084		
of	**meape**		0.118			0.039		
fit	**maape**		0.557			0.670		
	mse		0.031			0.028		

Table 1: Bias and mean squared error (MSE) for all parameters (except those from instrumental variable regression) from the Bayesian estimation without and with contraction mapping (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).

4 Limitations and conclusion

Limitations

Although the results were robust thus far to determinants of the setting of the simulation study, it would be interesting to examine these issues more extensively. Potential determinants are (i) the number of markets respectively number of periods (nmkt), (ii) the number of products (J).

Another interesting topic is the effect of 'misspecification' in the DGP on the estimation results, i.e. to analyze how this estimation approach does perform, if the estimated model is misspecified relating to the DGP.

The introduced approach is based on the procedure in (Jiang et al. 2009) and (Park & Gupta 2009) and its theory also relies - in opposite to the GMM estimation of the BLP model (Berry et al. 1995) - on the additional assumption, that the surrogate of unknown product characteristics ξ is normal distributed. More precisely the assumption made is that the residual ν, for the mapping of price on its instrumental variable, and ξ are distributed jointly normal. (Park & Gupta 2009) point out the economic implications of this assumption for pricing behavior and show that this assumption is not consistent with monopoly pricing and Nash pricing in differentiated products oligopoly, but consistent if prices are equal to marginal costs plus/times a fixed markups[17]. Otherwise the performance of an estimation method of this type depends on the robustness according to the violation of the assumption of the joint normality.

Conclusion

In summary the proposed simple alternative estimation approach leads to accurate results and thus shows to be capable of estimating the discussed model with a simpler estimation procedure and less computational effort. Moreover the procedure enables to use the whole

[17] I.e. markups do not depend on prices.

information about sold quantities available in the data set. In applied work the reduced time computation might be of special benefit when because of model selection lots of different model specifications have to be estimated and analyzed. The introduced approach roughly reduces the computation time by a factor proportional to the iterations needed in the contraction mapping which is in this setting in a range of 70-100 iterations.

Appendix

A Prior draws

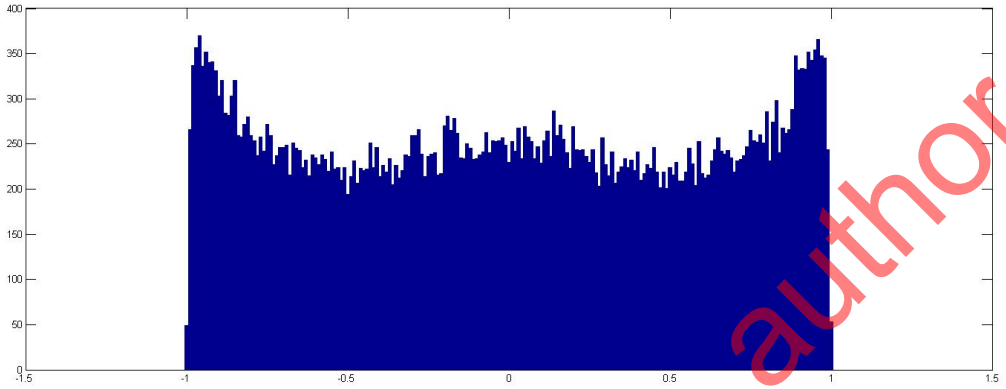

Figure 1: Frequencies of correlations of Element 4,3 from 50 000 draws from prior distribution of covariance-variance matrix.

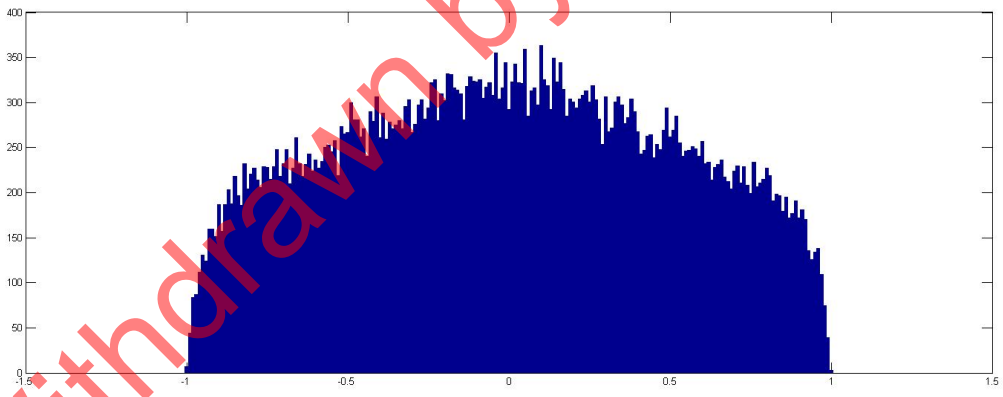

Figure 2: Frequencies of correlations of Element 5,3 from 50 000 draws from prior distribution of covariance-variance matrix.

B 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.

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

Smolyak, S. A. (1963), 'Quadrature and interpolation formulas for tensor products of certain classe of functions', *Soviet Mathematics Doklady* **4**, 240–243.

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.