Random Coefficient Demand Estimation by Optimal Instrument-Based Continuously Updated GMM

Zsolt Sándor*

15 February 2016

Abstract

We propose a continuously updated GMM estimator based on optimal instruments for estimating random coefficient demand, where the matrix of optimal instruments varies with the parameters. Under the condition that the optimal instrument matrix has full rank, this estimator avoids problems of inconsistency, from which GMM estimators that are based on instruments constructed from observed characteristics may suffer. In Monte Carlo simulations we find that when only product characteristics are available for instruments, the presence of some collinearity in the optimal instrument matrix affects the performance of the estimator in terms of bias and variance, but it still outperforms GMM based on characteristics instruments. In the case when also cost shifter instruments are available, our estimator performs very well.

Keywords: endogeneity, random coefficient logit, discrete choice.

*Email: zsosan@gmail.com. Department of Business, Sapientia University Miercurea Ciuc. Financial support from the grant PN-II-ID-PCE-2012-4-0066 of the Romanian Ministry of National Education, CNCS-UEFISCDI, is gratefully acknowledged.

1 Introduction

Empirical models of market demand have been used in a large number of applications after the influential work by Berry, Levinsohn and Pakes (1995, BLP hereafter). These authors constructed a market equilibrium model with differentiated products and price competition. An important feature of the demand specification is the presence of random coefficients that serve for modelling heterogeneity of consumers' tastes for the various characteristics. It has been shown (e.g., BLP, Nevo 2001) that these imply patterns of substitution between products that are more realistic. Another important feature is the presence of unobserved characteristics different from the previously used idiosyncratic errors, which facilitates the modelling of price endogeneity.

The estimation of the model commonly proceeds by GMM (i.e., generalized method of moments), which offers some robustness compared to other estimators; for example, GMM is potentially consistent without specifying the distributions of the underlying model errors. Allowing for price endogeneity in the model requires instruments in the GMM estimation. Instruments can be constructed either from cost shifters or observed product characteristics combined in a way to capture price competition. The latter are referred to as BLP instruments in the literature. In a market typically only a reduced number of cost shifters are available. However, in a typical empirical application besides the base coefficients also several random coefficients need to be instrumented. The literature has not yet fully clarified how the various types of instruments help identifying the various types of model parameters.

Armstrong (2015) provides clarification regarding how BLP instruments in large markets can help the identification of the price coefficient. He presents a situation in which the BLP instruments yield an inconsistent estimator of the price coefficient (see Armstrong 2015, p.3 for details). This happens in a large market, where the markups converge sufficiently fast to constants when the number of products goes to infinity, so the BLP instruments become weak instruments for the price. Armstrong finds that either cost shifter instruments or an increasing number of the markets at appropriate rate solves the problem of inconsistency. It should be noted, however, that in the cases when Armstrong establishes that BLP instruments identify the price coefficient, since there is a shortage of good instruments, it is still unclear how those instruments can identify the random coefficients.

Conlon (2013) proposes to select BLP-instruments by principle component analysis and uses GEL (i.e., generalized empirical likelihood) estimation, which is known to have lower higher order asymptotic bias (Newey and Smith 2004). Conlon studies the estimate of the price coefficient and shows that in the case of several markets his estimator outperforms GMM.

Both Conlon (2013) and Armstrong (2015) leave the issue of estimating the random coefficients open. The random coefficient on price potentially contains valuable information on price elasticity. The method proposed by Conlon uses information from BLP-instruments that is likely to improve the estimate of the mean price coefficient, but its effect on the random coefficients is unclear. Reynaert and Verboven (2014) study two-step GMM estimation based on Chamberlain's (1987) optimal instruments proposed by Berry, Levinsohn and Pakes (1999) by means of Monte Carlo simulations, and look at the estimates of the random coefficients as well. They assume perfect competition and the availability of several cost shifters. They find that in terms of finite sample properties their estimator outperforms the first step GMM estimator, but they admit that with several random coefficients this finding may not hold. This arises due to the fact that the second step GMM gets contaminated by the first step estimator, which, even if consistent, may have poor finite sample performance. In fact, the second step GMM estimator is not valid in the situations considered by Armstrong (2015) because the first step GMM estimator is not consistent.

In order to exploit the potential of optimal instruments we propose to use continuously updated GMM. This can potentially avoid the inconsistency issue of the first step GMM estimator. In the originally proposed version of this estimator (Hansen, Heaton and Yaron 1996) the weighting matrix varies with the parameters. In our version the matrix of optimal instruments as given by Chamberlain (1987) varies with the parameters. Optimal instruments are the most informative, at least asymptotically, and although Armstrong's (2015) findings suggest that the BLP instruments may not identify the parameters in large markets, these findings do not exclude that optimal instruments do identify all the model parameters (for an example and further clarification, see Domínguez and Lobato 2004). In a simple logit illustration we show that our estimator is consistent when the marginal cost specification is not linear in the observed characteristics (in the linear marginal cost case the matrix of optimal instruments may be collinear). That is, under the condition that the optimal instrument matrix has full rank, the estimator will be consistent, even if no cost shifters are available.

Our estimator is consistent in situations where Armstrong (2015) shows that GMM based on BLP instruments is not consistent. Our estimator is related to the one proposed by Conlon (2013) due to the fact that continuously updated GMM is in fact a special case of GEL. However, our estimator is different because Conlon's approach does not use optimal instruments, but applies GEL based on the selected BLP instruments.

In a Monte Carlo simulation we study the performance of our estimator. Similar to Armstrong (2015) we consider two cases; in the first case we assume that cost shifters are not observed, so only product characteristics are available for instruments, while in the second case cost shifters are observed. Our estimator performs very well in the second case, especially in comparison to the GMM estimator. In the first case we still get reasonable results, although, since the optimal instrument matrix shows some collinearity, the performance in terms of bias and variance is not as good as in the second case.

Our paper is related to a number of other recent papers. Petrin (2002) shows that the precision of the GMM estimator can be improved by combining the commonly used aggregate market data with consumer-level information related to product characteristics. Berry, Linton and Pakes (2004) provide conditions for the consistency and asymptotic normality of the GMM estimator in a market where the number of products goes to infinity. Freyberger (2015) develops many-small-market asymptotics with a special focus on the sampling error due to simulation estimation of market shares. Dubé, Fox and Su (2012) develop an estimation algorithm equivalent to GMM that improves the computation of the estimator by treating the market share equation as a constraint when optimizing the GMM objective function. Knittel and Metaxoglou (2014) report a number of problems related to the convergence of the GMM optimization algorithm. Berry and Haile (2014) provide conditions on nonparametric identification of demand and supply.

The next section presents the model. Section 3 presents the GMM estimation method and reviews the most common ways of constructing the instrument matrix. Section 4 presents the Monte Carlo study on the finite sample performance of the estimators. Finally, Section 5 concludes.

2 The Model

The model specifies demand based on consumers' latent utilities for the products. The latent utility of consumer i from purchasing product j in market t is assumed to be

$$u_{ijt} = \alpha_i p_{jt} + x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt},$$

where $\alpha_i = \alpha + \sigma_\alpha v_{\alpha i}, \quad \beta_i = \beta + V_{\beta i}\sigma$ with

$$V_{\beta i} = \begin{pmatrix} v_{1i} & & 0 \\ & \ddots & \\ 0 & & v_{Ki} \end{pmatrix}, \quad v = (v_{\alpha i}, v_{1i}, ..., v_{Ki})' \sim N\left(0, I_{K+1}\right),$$

p_{jt} is the price of product j in market t, x_{jt} is a $1 \times K$-vector of characteristics of product j in market t containing 1 for intercept, ξ_{jt} is a characteristic of product j in market t that is not observed by the econometrician, and ε_{ijt} is an iid type I extreme value distributed error term. In market t consumers can choose from J_t products and the outside alternative, which represents the option of not purchasing any of the J_t products, and is assumed to yield utility

$$u_{i0t} = \varepsilon_{i0t}.$$

We let $\mu_{jt} = \alpha p_{jt} + x_{jt}\beta + \xi_{jt}$ denote the part of utility that is common to all consumers.

These utility specifications imply the predicted market share expression

$$s_{jt}(\boldsymbol{\mu}_t, \mathbf{p}_t, \mathbf{x}_t, \delta) = \int q_{jt}(\boldsymbol{\mu}_t, \mathbf{p}_t, \mathbf{x}_t, \delta, v)\, \phi(v)\, dv \tag{1}$$

of product j in market t, where

$$q_{jt}(\boldsymbol{\mu}_t, \mathbf{p}_t, \mathbf{x}_t, \delta, v_i) = \frac{\exp\left(\mu_{jt} + \sigma_\alpha v_{\alpha i} p_{jt} + x_{jt} V_{\beta i}\sigma\right)}{1 + \sum_{r=1}^{J_t} \exp\left(\mu_{rt} + \sigma_\alpha v_{\alpha i} p_{rt} + x_{rt} V_{\beta i}\sigma\right)} \tag{2}$$

is the consumer-specific probability of purchasing product j. Here $\boldsymbol{\mu}_t = \alpha \mathbf{p}_t + \mathbf{x}_t\beta + \boldsymbol{\xi}_t$; $\boldsymbol{\xi}_t, \mathbf{p}_t$ are the $J_t \times 1$ vectors of unobserved characteristics and prices in market t, $\mathbf{x}_t$ is the $J_t \times K$ matrix of observed characteristics in market t, $\delta = (\alpha, \beta', \sigma')'$, where $\sigma = (\sigma_\alpha, \sigma_1, ..., \sigma_K)'$ and ϕ is the density function of the $K+1$-dimensional standard normal distribution. The predicted share of the outside alternative in market t is

$$s_{0t}(\boldsymbol{\mu}_t, \mathbf{p}_t, \mathbf{x}_t, \delta) = \int \frac{1}{1 + \sum_{r=1}^{J_t} \exp\left(\mu_{rt} + \sigma_\alpha v_\alpha p_{rt} + x_{rt}\Sigma^{1/2} v_\beta\right)} \phi(v)\, dv.$$

On the supply side of the market we assume Nash-Bertrand competition, where F firms determine their own prices optimally taking the prices of rival firms as given. Ignoring the market size and fixed production costs we assume that the profit of firm f in market t is

$$\Pi_{ft}(\mathbf{p}_{ft}, \mathbf{p}_{-ft}) = (\mathbf{p}_{ft} - \mathbf{c}_{ft})' \mathbf{s}_{ft},$$

where $\mathbf{p}_{ft}$, $\mathbf{c}_{ft}$, $\mathbf{s}_{ft}$ denote vectors of prices, marginal costs and market shares for firm f, and $\mathbf{p}_{-ft}$ denotes the vector of rival prices. Following the literature we assume that the price equilibrium arises as an inner solution of the first order conditions for profit maximization. These first order conditions can be rewritten as (see, e.g., BLP)

$$\mathbf{p}_t - \mathbf{c}_t = \Delta_t^{-1}\mathbf{s}_t, \tag{3}$$

where $\mathbf{c}_t$ and $\mathbf{s}_t$ are the column-vectors of all marginal costs and market shares in market t, provided that the matrix Δ_t, which is defined by $\Delta_{t,jh} = -\frac{\partial s_j}{\partial p_h}$, if j, h belong to the same firm and $\Delta_{t,jh} = 0$ otherwise, is invertible.

Although several empirical models specify marginal cost as log-linear (e.g., BLP in the car market, Sovinsky Goeree 2008 in the personal computer market), we adopt a linear specification used in several simulation experiments (e.g., Armstrong 2015, Reynaert and Verboven 2014):

$$c_{jt} = w_{jt}\gamma + \omega_{jt}, \tag{4}$$

where w_{jt} is a row vector of marginal cost covariates corresponding to product j and ω_{jt} is an unobserved cost characteristic.

3 Estimation by GMM

This section first presents the estimation procedure as proposed by BLP, then reviews the most common ways of constructing the instruments and finally presents our proposed continuously updated GMM estimator based on optimal instruments. It is important to note that, although –due to reasons that should become clear– in many cases we should use the wording "approximation of the optimal instruments," we just use "optimal instruments." The estimation of the model proceeds by first computing the demand and supply unobserved characteristics as a function of the model variables and parameters. On the demand side this consists of solving the nonlinear system of equations

$$s_{jt}\left(\boldsymbol{\mu}_t, \mathbf{p}_t, \mathbf{x}_t, \delta\right) = s_{jt}, \quad j = 1, ..., J_t \tag{5}$$

in μ_t in each market t, where s_{jt} is the observed share of product j in market t. According to Berry (1994) and BLP, the solution is unique and can be computed as the fixed point of the contraction

$$G_t : \mathbb{R}^{J_t} \rightarrow \mathbb{R}^{J_t}, \quad G_t\left(\boldsymbol{\mu}_t\right) = \boldsymbol{\mu}_t + \ln \mathbf{s}_t - \ln \mathbf{s}_t\left(\boldsymbol{\mu}_t, \mathbf{p}_t, \mathbf{x}_t, \delta\right),$$

where $\mathbf{s}_t\left(\boldsymbol{\mu}_t, \mathbf{p}_t, \mathbf{x}_t, \delta\right)$ is the column-vectors of the observed and predicted market shares in market t. We note that the solution $\mu_t\left(\mathbf{s}_t, \mathbf{p}_t, \mathbf{x}_t, \sigma\right)$ does not depend on α and β. Therefore, we can express the unobserved characteristics vectors in market t in terms of the model ingredients as

$$\begin{aligned} \boldsymbol{\xi}_t\left(\mathbf{s}_t, \mathbf{p}_t, \mathbf{x}_t, \delta\right) &= \boldsymbol{\mu}_t\left(\mathbf{s}_t, \mathbf{p}_t, \mathbf{x}_t, \sigma\right) - \left(\alpha \mathbf{p}_t + \mathbf{x}_t\beta\right), & (6) \\ \boldsymbol{\omega}_t\left(\mathbf{s}_t, \mathbf{p}_t, \mathbf{x}_t, \mathbf{w}_t, \alpha, \sigma, \gamma\right) &= \mathbf{p}_t - \Delta_t^{-1}\mathbf{s}_t - \mathbf{w}_t\gamma, & (7) \end{aligned}$$

where $\mathbf{w}_t$ is the column-vector of the marginal cost covariates in market t. This defines a simultaneous system of equations, where the first equation is linear in α and β and nonlinear in σ, and the second is linear in γ and nonlinear in α and σ.

With a few exceptions in the literature the model is estimated by GMM.[1] The GMM estimator is regarded appealing due to its robustness features. One such feature is that for consistency of the GMM estimator the distribution of the unobserved characteristics need not be specified. Another robustness feature is that the demand side can be estimated consistently without specifying the supply side exactly, a feature not necessarily shared by likelihood-based estimators.

Let $\rho_{jt} = \left(\xi_{jt}, \omega_{jt}\right)'$ and $\boldsymbol{\rho}_t = \left(\boldsymbol{\xi}_t', \boldsymbol{\omega}_t'\right)'$ be the vector of product-specific demand and supply unobserved characteristics and the vector of all demand and supply unobserved characteristics in market t, respectively. GMM estimation is based on the conditional moment restriction $E\left[\rho_{jt}|\mathbf{x}_t, \mathbf{w}_t\right] = 0$ that implies $E\left[Z_{jt}\rho_{jt}\right] = 0$ for an instrument matrix Z_{jt} of dimension $L \times 2$ that depends on $\mathbf{x}_t, \mathbf{w}_t$, for $t = 1, ..., T$ and $j = 1, ..., J_t$. In this context the GMM estimator commonly used is the value of $\theta = \left(\delta', \gamma'\right)'$ that minimizes the objective function

$$\left(\sum_{t=1}^{T} \boldsymbol{\rho}_t\left(\theta\right)' Z_t'\right) W \left(\sum_{t=1}^{T} Z_t \boldsymbol{\rho}_t\left(\theta\right)\right), \tag{8}$$

where Z_t is the $L \times 2J_t$ instrument matrix corresponding to the vectors $\boldsymbol{\xi}_t'$, $\boldsymbol{\omega}_t'$, and W is an $L \times L$ positive definite weighting matrix that may depend on $\mathbf{x}_t, \mathbf{w}_t$ but not on θ. For simplicity we omit the other arguments from $\boldsymbol{\rho}_t\left(\theta\right)$.

In many empirical applications of the model there is a shortage of instruments, on the one hand because the supply side characteristics w_{jt} are nonlinear (e.g., logarithmic) transformations of the demand observed characteristics x_{jt} while at most only a reduced number of cost shifters available, and on the other hand because the random coefficients of the model also need to be instrumented. Estimation of the random coefficients has been documented to be difficult even in models without endogeneity. Although there is an increasing number of papers in the literature on nonparametric identification of the distribution of random coefficients (e.g., Fox, Kim, Ryan and Bajari 2012), the literature has not yet fully clarified what drives the identification of random coefficients in practice (e.g., Chiou and Walker 2007).

Therefore, considering the use of optimal instruments is crucial. Chamberlain (1987) shows that the asymptotically optimal instrument matrix for the conditional moment restrictions $E\left[\boldsymbol{\rho}_t\left(\theta\right)|\mathbf{x}_t, \mathbf{w}_t\right] = 0$ is $B_t\left(\theta_0\right)$, where

$$B_t\left(\theta\right) = E\left[\left.\frac{\partial \boldsymbol{\rho}_t}{\partial \theta'}\left(\theta\right)\right| \mathbf{x}_t, \mathbf{w}_t\right]' \left(E\left[\boldsymbol{\rho}_t\left(\theta\right)\boldsymbol{\rho}_t\left(\theta\right)'|\mathbf{x}_t, \mathbf{w}_t\right]\right)^{-1}, \tag{9}$$

[1] For example, Jiang, Manchanda and Rossi (2009) develop a Bayesian estimator; Sándor (2013) develops a semiparametric maximum likelihood estimator.

and the optimal weighting matrix is $\left(\sum_{t=1}^{T} B_t(\theta_0) B_t(\theta_0)'\right)^{-1}$. In this case the number of instruments L is equal to the number of parameters. It is important to note that the expression $B_t(\theta_0)$ contains several unknowns like the true value of the parameter vector θ_0 and the two conditional expectations on the right hand side of $B_t(\theta_0)$ of which the second involves the conditional variance matrix of the vector of unobservables, which is a primitive unknown of the model.

3.1 BLP instruments

BLP take W from (8) equal to the identity matrix and approximate the optimal instruments following Newey (1990) and using the intuition suggested by Bresnahan (1987) that markups (3) respond differently to own and rival products. Newey (1990) proposes semiparametric estimation of optimal instruments in the form of conditional expectations employing polynomial series. BLP approximate the optimal instruments semiparametrically with linear polynomial series. Even in the case of linear series this is not easy due to the large number of basis functions that should be taken into account, because the linear basis functions of the polynomial space are the characteristics of all products. BLP reduce the dimension of the above mentioned polynomial space using the exchangeability of certain product characteristics as arguments of the unobserved characteristics (6), (7). The exchangeability of product characteristics tends to reduce the number of basis functions that are necessary in the semiparametric estimation of the optimal instruments, because, as Pakes (1994, Theorem 2) shows, the dimension of a polynomial space whose elements satisfy exchangeability conditions does not depend on the number of exchangeable arguments. Based on these arguments, for a given product BLP select as instruments the observed characteristics of the product, the sum of the characteristics of the other products produced by the same firm and the sum of the characteristics of the rival products.

These instruments are referred to as BLP instruments in the literature. They are useful since they can be applied in several situations without requiring assumptions on the distribution of the unobservables or the supply side. We note at the same time that these instruments are not expected to be precise approximations of the optimal instruments, for at least two reasons. First, the basis functions employed are linear and hence they are not able to cover the variation caused by the non-linearity of the optimal instruments. Second, the basis functions are not used eventually for estimating the optimal instruments in a semiparametric way but are taken directly as instruments.

3.2 Second step optimal instruments

BLP (1999) propose a two-step GMM estimator that relies on a first step consistent estimator $\tilde{\theta}$ of θ_0 in order to estimate $B_t(\theta_0)$. Then they replace all the unobservables by 0, recompute the prices and market shares and using these they compute the expressions involved in $B_t\left(\tilde{\theta}\right)$. Our procedure presented below follow similar steps, so we refer to that for details, or to the appendix of BLP (1999).

Some papers report cases where the second step GMM estimator has higher standard errors than the first step estimator based on BLP instruments (e.g., Sovinsky Goeree 2008). Relying on several cost shifter instruments, in a perfect competition supply setup Reynaert and Verboven (2014) find that the second step estimator outperforms the first step estimator in terms of both bias and standard deviation. Still, the most common situation is that the two-step optimal instruments approach is used after a first step based on BLP instruments. Therefore, under the conditions given by Armstrong (2015), the two-step approach fails due to the fact that there is no first step consistent estimator available.

3.3 Approximation of optimal instruments

Here we describe how we approximate the optimal instruments in order to obtain an easily computable continuously updated GMM estimator of the demand side parameters. We proceed by approximating the supply model in a way that is convenient for constructing valid and easily computable instruments. This motivation is in a way similar to BLP's goal to approximate the optimal instruments, but our instruments are a better approximation and are also continuously updated with respect to the parameters.

Next we provide details on the expressions involved in $B_t(\theta)$ from (9). First we replace both the conditional variance matrix $E\left[\boldsymbol{\rho}_t(\theta)\,\boldsymbol{\rho}_t(\theta)'\,|\mathbf{x}_t,\mathbf{w}_t\right]$ and the optimal weighting matrix $\sum_{t=1}^{T} B_t(\theta_0)\,B_t(\theta_0)'$ by the identity matrix; this only affects the efficiency of the estimator, not its consistency.[2] Second, we approximate the supply side model by omitting the markup term $\Delta_t^{-1}\mathbf{s}_t$ from $\boldsymbol{\omega}_t = \mathbf{p}_t - \Delta_t^{-1}\mathbf{s}_t - \mathbf{w}_t\gamma$.[3] The motivation for this is twofold. First, in the large market case discussed by Armstrong (2015), in some cases all components of this vector converge to the same constant, as $J_t \to \infty$, so they lose their power to explain the variation in prices. One simple case when this occurs is the simple logit when the number of firms goes to ∞ as $J_t \to \infty$ and the

[2] The disadvantage of this move is loss of efficiency. One gain is in terms of computation; a second gain is that the finite sample behaviour of the estimator improves; see a brief discussion on this below.

[3] Alternatively, one can include BLP instruments in the supply side equation as explanatory variables, as Reynaert and Verboven (2014) suggest. This could capture the effect of market power on price.

number of products of each firm stays bounded. Indeed, in the simple logit (i.e., when $\sigma = 0$)

$$\omega_{jt} = p_{jt} + \frac{1}{\alpha} \frac{1}{1 - \sum_f s_{jt}} - w_{jt}\gamma,$$

where $\sum_f s_{jt}$ is the sum of the market shares of the products that belong to firm f (see Konovalov and Sándor 2010). If the number of products of firm f stays bounded as $J_t \to \infty$, then $\sum_f s_{jt} \to 0$, so the supply side equation becomes $\omega_{jt} = p_{jt} + \frac{1}{\alpha} - w_{jt}\gamma$. The second motivation for omitting $\Delta_t^{-1}\mathbf{s}_t$ is that this way the supply equation becomes linear.

With these considerations we have

$$\frac{\partial \boldsymbol{\rho}_t}{\partial \theta'}(\theta) = \begin{pmatrix} \frac{\partial \boldsymbol{\xi}_t}{\partial \theta'}(\theta) \\ \\ \frac{\partial \boldsymbol{\omega}_t}{\partial \theta'}(\theta) \end{pmatrix} = \begin{pmatrix} \frac{\partial \boldsymbol{\xi}_t}{\partial \delta'}(\theta) & 0 \\ \\ 0 & \frac{\partial \boldsymbol{\omega}_t}{\partial \gamma'}(\theta) \end{pmatrix} = \begin{pmatrix} -\mathbf{p}_t & -\mathbf{x}_t & \frac{\partial \boldsymbol{\mu}_t}{\partial \sigma'}(\sigma) & 0 \\ \\ 0 & 0 & 0 & -\mathbf{w}_t \end{pmatrix}.$$

By the implicit function theorem

$$\frac{\partial \boldsymbol{\mu}_t(\sigma)}{\partial \sigma'} = \left(\frac{\partial \mathbf{s}_t}{\partial \boldsymbol{\mu}_t'}\right)^{-1} \frac{\partial \mathbf{s}_t}{\partial \sigma'},$$

where the derivatives involved can be computed as

$$\begin{aligned} \frac{\partial \mathbf{s}_t}{\partial \boldsymbol{\mu}_t'} &= \int (Q_t - q_t q_t') \phi(v)\, dv, \\ \frac{\partial \mathbf{s}_t}{\partial \sigma'} &= \int (Q_t - q_t q_t') \overline{X}_t \overline{V} \phi(v)\, dv, \end{aligned}$$

where $q_t \equiv q_t(\boldsymbol{\mu}_t(\mathbf{s}_t, \mathbf{p}_t, \mathbf{x}_t, \sigma), \mathbf{p}_t, \mathbf{x}_t, \delta, v)$, $\overline{X}_t = (\mathbf{p}_t, \mathbf{x}_t)$ and $\overline{V}$ is the diagonal matrix with main diagonal elements $(v_\alpha, v_1, ..., v_K)$. It will be useful to work with the variables $\boldsymbol{\xi}_t, \boldsymbol{\omega}_t$ instead of $\mathbf{s}_t, \mathbf{p}_t$. By doing so, we obtain

$$\begin{aligned} \boldsymbol{\mu}_t(\mathbf{s}_t, \mathbf{p}_t, \mathbf{x}_t, \sigma) &= \boldsymbol{\mu}_t(\mathbf{s}_t(\boldsymbol{\xi}_t, \boldsymbol{\omega}_t, \mathbf{x}_t, \mathbf{w}_t, \theta), \mathbf{p}_t(\boldsymbol{\xi}_t, \boldsymbol{\omega}_t, \mathbf{x}_t, \mathbf{w}_t, \theta), \mathbf{x}_t, \sigma) \\ &= \alpha(\mathbf{w}_t\gamma + \boldsymbol{\omega}_t) + \mathbf{x}_t\beta + \boldsymbol{\xi}_t, \end{aligned}$$

so

$$\begin{aligned} & q_{jt}(\boldsymbol{\mu}_t(\mathbf{s}_t, \mathbf{p}_t, \mathbf{x}_t, \sigma), \mathbf{p}_t, \mathbf{x}_t, \delta, v) \equiv \overline{q}_{jt}(\boldsymbol{\xi}_t, \boldsymbol{\omega}_t, \mathbf{x}_t, \mathbf{w}_t, \theta, v) \\ &= \frac{\exp\left((\alpha + \sigma_\alpha v_\alpha)(w_{jt}\gamma + \omega_{jt}) + x_{jt}(\beta + V\sigma_\beta) + \xi_{jt}\right)}{1 + \sum_{r=1}^{J_t} \exp\left((\alpha + \sigma_\alpha v_\alpha)(w_{rt}\gamma + \omega_{rt}) + x_{rt}(\beta + V\sigma_\beta) + \xi_{rt}\right)}. \end{aligned}$$

Therefore,

$$E\left[\frac{\partial \boldsymbol{\rho}_t}{\partial \theta'}(\theta)\middle| \mathbf{x}_t, \mathbf{w}_t\right] = -\begin{pmatrix} \mathbf{w}_t\gamma & \mathbf{x}_t & E\left[\frac{\partial \boldsymbol{\mu}_t}{\partial \sigma'}(\sigma)\middle| \mathbf{x}_t, \mathbf{w}_t\right] & 0 \\ \\ 0 & 0 & 0 & \mathbf{w}_t \end{pmatrix},$$

where

$$E\left[\frac{\partial \boldsymbol{\mu}_t}{\partial \sigma'}(\sigma)\bigg|\mathbf{x}_t,\mathbf{w}_t\right]=E\left[\left(\left(\int\left(\overline{Q}_t-\overline{q}_t\overline{q}_t'\right)\phi(v)\,dv\right)^{-1}\int\left(\overline{Q}_t-\overline{q}_t\overline{q}_t'\right)\overline{X}_t\overline{V}\phi(v)\,dv\right)\bigg|\mathbf{x}_t,\mathbf{w}_t\right].$$

Let

$$G(\boldsymbol{\xi}_t,\boldsymbol{\omega}_t,\mathbf{x}_t,\mathbf{w}_t,\sigma)=\left(\int\left(\overline{Q}_t-\overline{q}_t\overline{q}_t'\right)\phi(v)\,dv\right)^{-1}\int\left(\overline{Q}_t-\overline{q}_t\overline{q}_t'\right)\overline{X}_t\overline{V}\phi(v)\,dv. \tag{10}$$

Then

$$E\left[\frac{\partial \boldsymbol{\mu}_t}{\partial \sigma'}(\sigma)\bigg|\mathbf{x}_t,\mathbf{w}_t\right]=\int G(\boldsymbol{\xi}_t,\boldsymbol{\omega}_t,\mathbf{x}_t,\mathbf{w}_t,\sigma)\,f(\boldsymbol{\xi}_t,\boldsymbol{\omega}_t|\mathbf{x}_t,\mathbf{w}_t)\,d\boldsymbol{\xi}_t d\boldsymbol{\omega}_t,$$

which can be estimated by Monte Carlo integration as

$$\frac{1}{R}\sum_{i=1}^{R}G\left(\boldsymbol{\xi}_t^i,\boldsymbol{\omega}_t^i,\mathbf{x}_t,\mathbf{w}_t,\sigma\right), \tag{11}$$

where $\left(\boldsymbol{\xi}_t^i,\boldsymbol{\omega}_t^i\right)$, $i=1,...,R$ are draws from the distribution with density $f(\boldsymbol{\xi}_t,\boldsymbol{\omega}_t|\mathbf{x}_t,\mathbf{w}_t)$.

However, this Monte Carlo estimator cannot be computed because the conditional distribution of $(\boldsymbol{\xi}_t,\boldsymbol{\omega}_t|\mathbf{x}_t,\mathbf{w}_t)$ is not known. We approximate the Monte Carlo estimator (11) by $G(0,0,\mathbf{x}_t,\mathbf{w}_t,\sigma)$, that is, by taking $R=1$, $\boldsymbol{\xi}_t^i=0$, $\boldsymbol{\omega}_t^i=0$, as suggested by BLP (1999), and we take the corresponding Monte Carlo estimators of the integrals involved in G in equation (10). So we propose the estimator that minimizes the objective function

$$\left(\sum_{t=1}^{T}\boldsymbol{\rho}_t(\theta)'Z_t^*(\sigma,\gamma)'\right)\left(\sum_{t=1}^{T}Z_t^*(\sigma,\gamma)\boldsymbol{\rho}_t(\theta)\right), \tag{12}$$

where

$$Z_t^*(\sigma,\gamma)=-\begin{pmatrix}\mathbf{w}_t\gamma & \mathbf{x}_t & G(0,0,\mathbf{x}_t,\mathbf{w}_t,\sigma) & 0\\ 0 & 0 & 0 & \mathbf{w}_t\end{pmatrix}'.$$

Note that the estimator defined in (12) is a continuously updated GMM estimator because the instrument matrix $Z_t^*(\sigma,\gamma)$ is allowed to vary with the unknown parameters σ, γ. However, this estimator is different from the original continuously updated GMM estimator proposed by Hansen, Heaton and Yaron (1996), because that estimator uses fixed instruments and the optimal weighting matrix depends on the parameters.

The motivation for introducing Hansen's et al. (1996) estimator was to obtain an estimator that is less biased in finite samples than the GMM estimator. Newey and Smith (2004) show that the continuously updated GMM estimator has smaller bias than the two-step GMM estimator in terms of higher order asymptotics. This feature has been noted for the continuously updated GMM estimator in some Monte Carlo simulations along with the feature that –especially when instruments are weak– its distribution may have fat tails (e.g., Hansen et al. 1996, Guggenberger 2008).

These observations are partially relevant for our estimator. We experimented with the optimal instruments as given in (9) and indeed in a small proportion of the replications we obtained estimates that are far from the true values. After replacing the conditional variance matrix $E\left[\boldsymbol{\rho}_t(\theta)\,\boldsymbol{\rho}_t(\theta)'\,|\mathbf{x}_t,\mathbf{w}_t\right]$ and the optimal weighting matrix $\sum_{t=1}^{T} B_t(\theta_0)\,B_t(\theta_0)'$ by the identity matrix, this phenomenon disappeared completely. We believe that in these experiments the fat tail feature occurs due to some collinearity in the optimal instrument matrix accompanied by a somewhat low correlation between price and its optimal instrument and not as much due to weak instruments.

3.3.1 Illustration: Simple logit

The simple logit residuals can be written as

$$\begin{aligned}\boldsymbol{\xi}_t(\mathbf{s}_t,\mathbf{p}_t,\mathbf{x}_t,\delta) &= \mathbf{y}_t-(\alpha\mathbf{p}_t+\mathbf{x}_t\beta)\,,\\ \boldsymbol{\omega}_t(\mathbf{p}_t,\mathbf{w}_t,\gamma) &= \mathbf{p}_t-\mathbf{w}_t\gamma,\end{aligned}$$

where $\mathbf{y}_t$ is the column-vector of $y_{jt}=\ln\left(\frac{s_{jt}}{s_{0t}}\right)$. For the optimal instrument matrix we obtain

$$Z_t^*=-\begin{pmatrix}\mathbf{w}_t\gamma & \mathbf{x}_t & 0\\ 0 & 0 & \mathbf{w}_t\end{pmatrix}',$$

which implies the objective function

$$\begin{pmatrix}\gamma'\mathbf{w}_t'(\mathbf{y}_t-\alpha\mathbf{p}_t+\mathbf{x}_t\beta)\\ \mathbf{x}_t'(\mathbf{y}_t-\alpha\mathbf{p}_t+\mathbf{x}_t\beta)\\ \mathbf{w}_t'(\mathbf{p}_t-\mathbf{w}_t\gamma)\end{pmatrix}'\begin{pmatrix}\gamma'\mathbf{w}_t'(\mathbf{y}_t-\alpha\mathbf{p}_t+\mathbf{x}_t\beta)\\ \mathbf{x}_t'(\mathbf{y}_t-\alpha\mathbf{p}_t+\mathbf{x}_t\beta)\\ \mathbf{w}_t'(\mathbf{p}_t-\mathbf{w}_t\gamma)\end{pmatrix}.$$

For $\gamma=\gamma_0$ the consistency of the estimators of α and β depends on the rank of the instrument matrix $\begin{pmatrix}\mathbf{w}_t\gamma & \mathbf{x}_t\end{pmatrix}$. If this matrix has full rank, and the supply side equation is well approximated in order for $\mathbf{w}_t\gamma$ to correlate with $\mathbf{p}_t$, then the estimators of α and β will be consistent. In the case when the marginal cost covariates are the same as the demand observed characteristics, $\mathbf{w}_t=\mathbf{x}_t$, then the instrument matrix $\begin{pmatrix}\mathbf{w}_t\gamma & \mathbf{x}_t\end{pmatrix}$ has deficient rank, and the estimator will be inconsistent.

3.3.2 Log-linear marginal cost

Here we discuss the log-linear marginal cost specification, which is also relevant for the random coefficient logit demand model. Often (e.g., BLP, Sovinsky 2008) the following marginal cost specification is used:

$$\ln c_{jt}=w_{jt}\gamma+\omega_{jt},$$

where $w_{jt}=\ln x_{jt}$ for continuous variables and $w_{jt}=x_{jt}$ for dummy variables. Taking component j of (3) $p_{jt}=c_{jt}+\overline{\Delta}_{jt}$, where $\overline{\Delta}_{jt}=\left(\Delta_t^{-1}\mathbf{s}_t\right)_j$, we have the relationship

$$p_{jt} = \exp(w_{jt}\gamma)\exp(\omega_{jt}) + \overline{\Delta}_{jt}.$$

Write $\exp(\omega_{jt}) = a + b\zeta_{jt}$ with $E\left[\zeta_{jt}\right] = 0,\ var\left(\zeta_{jt}\right) = 1$; then

$$p_{jt} = a\exp(w_{jt}\gamma) + b\exp(w_{jt}\gamma)\zeta_{jt} + \overline{\Delta}_{jt},$$

that is, w_{jt} also effects the variance of the error term in the model of p_{jt}. This suggests the following nonlinear model of prices:

$$p_{jt} = \lambda_1\exp(w_{jt}\gamma) + z_{jt}\lambda_2 + \varpi_{jt},$$

where z_{jt} is a vector of covariates (like BLP instruments) that aims to proxy for the omitted $\overline{\Delta}_{jt}$ term. The unknown parameters γ and $\lambda = (\lambda_1, \lambda_2')'$ can be estimated consistently by nonlinear least squares. Even if there are no cost shifters available, this marginal cost specification yields consistent estimates of the demand parameters because the instrument matrix $\left(\begin{array}{cc}\lambda_1\exp(\mathbf{w}_t\gamma) & \mathbf{x}_t\end{array}\right)$ has full rank.

4 Monte Carlo Study

Here we study the finite sample properties of the estimator specified in (12), which we denote below by CUE-OI, in two relevant cases. In the first case we assume that the cost shifters are not observed; in the second case we assume that the cost shifters are observed by the researcher.[4] In these cases we compare the performance of this estimator to the GMM estimator used in the literature, which we refer to below as GMM-IV. These comparisons are for data generating processes that are similar to those used by Armstrong (2015), where there is only one random coefficient included in the model. In addition to this, we look at the performance of the CUE-OI in the case when all three demand coefficients are random and the cost shifters are observed. The results obtained in this case are meant to show the performance of the CUE-OI in a situation that is more relevant from a practical point of view.

In the Monte Carlo study we focus on the large markets case. The large markets case is more interesting because in practically relevant situations, that is, when the researcher

[4]For several markets data on cost shifters are scarce, but in the case of some markets some cost shifters are available. One example is the car market, which is one of the mostly studied markets in the literature. For cars one can find several pairs of variables of which one variable is likely to affect utility while the other is likely to affect marginal cost. For example, fuel consumption in terms of quantity of fuel "miles/gallon" affects marginal cost while fuel consumption in money terms "miles/dollar " affects utility; the dummy variable "4-wheel drive" affects marginal cost while the dummy variable "SUV" = "sport utility vehicle" affects utility. These pairs of variables are highly correlated but not completely the same. For more cost shifters in the car market see Reynaert and Verboven (2014).

has data on a reduced number of large markets, there may be no consistent estimator available, as shown by Armstrong (2015). In the Monte Carlo study we allow for one market only. While pooling data from several markets is a necessity in the case of small markets in order to obtain sufficiently high degrees of freedom, there are at least two arguments why estimation in a single large market is relevant. First, there is no reason to assume a priori that consumer preferences conditional on demographics in distinct markets are the same. Second, in practice there is little variation in the characteristics of products in different markets that are close to each other either in time or space, so empirical data from a reduced number of markets are not likely to show more variation than simulated data from a single market.

4.1 Simulation design

For the data generating processes whose results are presented in Tables 1 and 2 we follow Armstrong (2015) with minor differences. We consider a single market (so we drop the market subscript t) with 100 products that belong to 10 firms, each having 10 products. There is one observed demand characteristic x_{j2} and one cost shifter denoted z_j both generated as uniform on $(0, 1)$. The unobserved characteristics are generated as $\xi_j = U_1 + U_2 - 1$, $\omega_j = U_1 + U_3 - 1$, where U_1, U_2, U_3 are generated as uniform on $(0, 1)$. The supply side observed characteristic is $w_j = x_j^2$ in order to destroy perfect collinearity of the optimal instruments. The true values of the parameters are: $\alpha = -1,\ \beta = (-3, 6)',\ \sigma = 3,\ \gamma = (2, 1, 1)'$.

In the case when we assume that the cost shifters are not observed the CUE-OI only uses w_j in the supply side specification while the GMM-IV estimator employes the instruments: $x_j,\ \sum_{h\neq j;f} x_{j2}$ (BLP instruments), $\left(\sum_{h\neq j;f} x_{j2}\right)^2$. In the case when we assume that the cost shifters are observed by the researcher the CUE-OI uses w_j and z_j in the supply side specification while the GMM-IV estimator employes the instruments: $x_j,\ z_j,\ z_j^2$. For the latter estimator we use the same instruments as Armstrong (2015) in order to have a reference for the results.

For the data generating processes whose results are presented in Table 3 we consider a single market first with 50 products that belong to 5 firms, each having 10 products and then with 100 products that belong to 10 firms, each having 10 products. We generate both observed and unobserved product characteristics in the same way as above. In this case all coefficients are random, so the true values of the parameters are: $\alpha = -3,\ \beta = (-3, 6)',\ \sigma = (1, 2, 3),\ \gamma = (2, 1, 1)'$. Here we assume that the cost shifters are observed.

4.2 Results

The results are based on 1000 replications. In approximately 1% of the replications the GMM-IV estimator did not converge, while the CUE-OI converged in all cases. The tables present the results as the means and standard deviations of the estimates. Table 1 presents the results in the case when cost shifters are not observed. The CUE-OI shows relatively small bias, while apart from the price coefficient ($= -1$) the GMM-IV estimates are rather biased. The standard deviations of the CUE-OI are reasonably low, while those of the GMM-IV estimates are rather high.

Table 1. Cost shifters are not observed

	# products = 100		# products = 100	
True	CUE-OI		GMM-IV	
value	Mean	St. dev.	Mean	St. dev.
-1	-0.804	0.492	-0.934	3.098
-3	-3.523	1.277	-1.948	8.252
6	5.703	0.653	10.785	10.637
3	2.740	0.575	11.215	18.377

Table 2 presents the results in the case when cost shifters are observed. The CUE-OI shows no bias and low standard deviation for all the parameters. The GMM-IV estimator of the price coefficient ($= -1$) has low mean bias and relatively low standard deviation, which is still about three times as high as that of the CUE-OI. Both the bias and standard deviation of the GMM-IV estimates of the other three parameters are rather high. The standard deviations of the CUE-OI are reasonably low, while those of the GMM-IV estimates are rather high. For the sake of comparison we note that the median bias for the GMM-IV of the price coefficient is -0.066 (not reported in the table), while the median bias obtained by Armstrong (2015) is -0.020, which is quite close.

Table 2. Cost shifters are observed

	# products = 100		# products = 100	
True	CUE-OI		GMM-IV	
value	Mean	St. dev.	Mean	St. dev.
-1	-1.011	0.156	-1.202	0.461
-3	-2.969	0.435	-1.517	3.389
6	6.033	0.359	10.217	9.837
3	3.029	0.498	9.644	17.034

Table 3 presents the results in the case when all the coefficients are random and cost shifters are observed. In general the biases are rather low; we can see the lowest biases for the mean and standard deviation parameters of the observed characteristic (6 and 3); we can see the highest biases for the mean and standard deviation parameters of the constant (-3 and 2). In the case of 50 products the standard deviations of the estimates for the latter two parameters are rather high. However, as one expects, the standard deviations get lower in the 100 products case.

Table 3. The model with three random coefficients when cost shifters are observed

	# products = 50		# products = 100	
True	CUE-OI		CUE-OI	
value	Mean	St. dev.	Mean	St. dev.
-3	-3.182	1.155	-2.916	0.913
1	1.077	0.876	0.857	0.756
-3	-3.354	2.605	-3.467	2.005
6	5.927	0.861	5.944	0.659
2	2.467	2.637	2.277	1.758
3	2.998	1.333	3.089	0.739

The results in Tables 1 and 2 suggest that the CUE-OI clearly outperforms the GMM-IV estimator irrespective of whether the researcher observes cost shifters or not. The most striking difference in performance between the two estimators appears to occur for the standard deviation parameter (= 3). The GMM-IV estimator does quite well regarding the bias of the price coefficient ($= -1$), but it is relatively poor regarding its standard deviation.

5 Conclusions

This paper proposes a continuously updated GMM estimator based on optimal instruments for estimating demand models based on the random coefficient logit. This estimator can be employed in several versions of this model, but this estimator is the most useful in the case of large markets when markups tend to have a decreasing correlation with instruments based on observed product characteristics. In such cases the traditionally employed GMM estimator has been shown to be inconsistent (Armstrong 2015). We argue that our proposed estimator is consistent unless the marginal cost is specified to be linear in the demand characteristics.

In a Monte Carlo study we compare this estimator to the traditional GMM estimator regarding finite sample performance both when cost shifters are observed by the researcher and when they are not. We find that in both cases the proposed estimator clearly outperforms the traditional GMM estimator in terms of both bias and standard deviation. In a separate Monte Carlo study we look at the performance of the proposed estimator in a version of the model where all coefficients are random. The results are not as good as for the simpler model, but the improvement in performance when the number of products is doubled suggests that the estimator is promising in this case as well.

References

[1] Armstrong, Timothy B. (2015). Large Market Asymptotics for Differentiated Product Demand Estimators with Economic Models of Supply, Mimeo.

[2] Berry, S.T. (1994). Estimating discrete-choice models of product differentiation, *RAND Journal of Economics* 25, 42-162.

[3] Berry, S.T. and P.A. Haile (2014). Identification in differentiated products markets using market level data, *Econometrica* 82: 1749–1797.

[4] Berry, Steven, James Levinsohn, and Ariel Pakes (1995). Automobile Prices in Market Equilibrium, *Econometrica* 63, 841–890.

[5] Berry, S., J. Levinsohn and A. Pakes (1999). Voluntary export restraints on automobiles: Evaluating a trade policy, *American Economic Review* 89: 400–430.

[6] Berry, S., O.B. Linton and A. Pakes (2004). Limit theorems for estimating the parameters of differentiated product demand systems, *Review of Economic Studies* 71: 613–654.

[7] Bresnahan, T. (1987). Competition and Collusion in the American Automobile Market: The 1955 Price War, *Journal of Industrial Economics* 35, 457-482.

[8] Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions, *Journal of Econometrics* 34: 305–34.

[9] Chiou, L. and J.L. Walker (2007). Masking identification of discrete choice models under simulation methods, *Journal of Econometrics* 141: 683–703.

[10] Conlon, C. (2013). The Empirical Likelihood MPEC Approach to Demand Estimation, Mimeo.

[11] Domínguez, M.A. and I.N. Lobato (2004). Consistent estimation of models defined by conditional moment restrictions, *Econometrica* 72: 1601-1615.

[12] Dubé, J-P., J.T. Fox and C-L. Su (2012). Improving the numerical performance of BLP static and dynamic discrete choice random coefficients demand estimation, *Econometrica* 80: 2231–2267.

[13] Freyberger, J. (2015). Asymptotic theory for differentiated products demand models with many markets, *Journal of Econometrics* 185: 162-181.

[14] Fox, J.T. Kyoo il Kim, S.P. Ryan and P. Bajari (2012). The random coefficients logit model is identified, *Journal of Econometrics* 166, 204–212.

[15] Guggenberger, P. (2008). Finite Sample Evidence Suggesting a Heavy Tail Problem of the Generalized Empirical Likelihood Estimator, *Econometric Reviews* 26, 526–541.

[16] Hansen, L.P., J. Heaton and A. Yaron (1996). Finite-sample properties of some alternative GMM estimators, *Journal of Business & Economic Statistics* 14: 262-280.

[17] Jiang, R., P. Manchanda and P.E. Rossi (2009). Bayesian analysis of random coefficient logit models using aggregate data, *Journal of Econometrics* 166, 204–212.

[18] Knittel, C.R., K. Metaxoglou (2014). Estimation of Random-Coefficient Demand Models: Two Empiricists' Perspective, *Review of Economics and Statistics* 96: 34-59.

[19] Konovalov, A. and Z. Sándor (2010). On price equilibrium with multi-product firms, *Economic Theory* 44: 271–292.

[20] Newey, W.K., (1990). Efficient instrumental variables estimation of nonlinear models, *Econometrica* 58, 809-837.

[21] Newey, W.K. and R.J. Smith (2004). Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators, *Econometrica* 72: 219–255.

[22] Pakes, A. (1994). Dynamic Structural Models, Problems and Prospects: Mixed Continuous Discrete Controls and Market Interactions, in *Advances in Econometrics: The Sixth World Congress of the Econometric Society*, Vol. II, ed. by J.-J. Laffont and C. Sims. New York: Cambridge University Press, 171-260.

[23] Petrin, A. (2002). Quantifying the Benefits of New Products: The Case of the Minivan, *Journal of Political Economy* 110: 705-729.

[24] Reynaert, M. and F. Verboven (2014). Improving the performance of random coefficients demand models: The role of optimal instruments, *Journal of Econometrics* 179: 83–98.

[25] Sándor, Z. (2013). Semiparametric Maximum Likelihood Estimation of Differentiated Product Demand Systems, Mimeo.

[26] Sovinsky Goeree, M. (2008). Limited information and advertising in the U.S. personal computer industry, *Econometrica*, 76, 1017–1074.