The Stata Journal (2015)
15, Number 3, pp. 854–880

The Berry–Levinsohn–Pakes estimator of the random-coefficients logit demand model

David W. Vincent
Deloitte Economic Consulting
London, UK
davivincent@deloitte.co.uk

Abstract. In this article, I describe the algorithm proposed by Berry, Levinsohn, and Pakes (1995, *Econometrica* 63: 841–890) to fit the random-parameters logit demand model from product market shares. I present a new command, `blp`, for this estimator.

Keywords: st0408, blp, logit model, elasticities, contraction mapping, GMM, random coefficients, optimal instruments

1 Introduction

The estimation of consumer demand in differentiated product industries plays a central role in applied economic analysis. The traditional approach is to specify a system of demand functions consistent with economic theory and estimate the parameters using aggregate market data. Popular examples include the Rotterdam model formulated by Theil (1965) and the almost-ideal demand system by Deaton and Muellbauer (1980b).

Despite their flexibility, a major hurdle that practitioners often face concerns the large number of parameters that need to be estimated, even after the restrictions of adding-up homogeneity and symmetry have been imposed. This dimensionality problem can be solved if preferences are assumed to be separable and consumers adopt a two-stage budgeting process. The first assumption allows products to be allocated into broad groups, such that preferences within one group are independent of the consumption decisions in another. The second assumption then permits expenditure to be allocated in two discrete stages: first to the broad groups, given total expenditure and group price indices, and then to the products within each group, given group outlay and within-group prices. Although empirically appealing, the assumption of separability places strong restrictions on the degree of substitutability between goods in different groups. Furthermore, the requirement of a valid group price index, such that two-stage budgeting corresponds to the allocation if made in one step, places severe restrictions on the permissible forms of utility function (see Deaton and Muellbauer [1980a, chap. 5]).

 st0408

The logit demand model, first proposed by McFadden (1974), is another way to address the dimensionality problem by projecting consumer preferences onto a finite set of product characteristics. This model uses observed market shares and can be easily fit following a transformation of the dependent variable. However, despite this computational simplicity, the model imposes strong restrictions on the nature of consumer heterogeneity that again lead to patterns of substitution that are generally unrealistic.

In this article, I discuss the new command `blp`, which estimates the parameters of the random-coefficients logit demand model from product market shares. This command uses the generalized method of moments (GMM) estimator proposed by Berry, Levinsohn, and Pakes (1995) (henceforth, BLP) and allows for endogenous prices and consumer heterogeneity in the valuation of product characteristics. This creates flexible patterns of substitution and leads to more realistic estimates of own- and cross-price elasticities. Finally, to reduce bias and improve both the efficiency and stability of the estimator, `blp` follows Reynaert and Verboven (2014) and provides an option to fit the model using Chamberlain (1987) optimal instruments.

The remainder of this article is organized as follows. In section 2, I describe the demand model and follow the exposition by Nevo (2000b), who popularized BLP. In section 3, I discuss details of the BLP algorithm that constitute the GMM estimator. In section 4, I describe `blp`. In section 5, I provide examples, and in section 6, I use Monte Carlo simulation to investigate the small-sample properties of the estimator.

2 The model

Following the exposition by Nevo (2000b), it is assumed that there are $t = 1, \dots, T$ observable markets consisting of $i = 1, \dots, I_t$ consumers, facing $j = 1, \dots, J$ alternative products.[1] For each market, aggregate data are available on product demand, prices, and product characteristics. Markets are assumed to be independent and can be cross-sectional (for example, geographic), time series, or longitudinal.

2.1 Demand

Let u_{ijt} denote the indirect utility that individual i receives from the consumption of product j in market t. This is assumed to be a linear function of a $K \times 1$ vector of product characteristics $\mathbf{x}_{jt}$, price p_{jt}, an unobserved (to the econometrician) component ξ_{jt}, and an idiosyncratic error ϵ_{ijt}. Hence,

$$u_{ijt} = \alpha_i(y_i - p_{jt}) + \mathbf{x}'_{jt}\boldsymbol{\beta}_i + \xi_{jt} + \epsilon_{ijt} \tag{1}$$

where y_i is individual income, $\boldsymbol{\beta}_i$ is a $K \times 1$ vector of coefficients, and α_i is the marginal utility of income. The term ξ_{jt} can be regarded as deviations from observed product quality that are common to all individuals. Consumer i can also choose to buy the outside product $j = 0$, with normalized utility $u_{i0t} = \alpha_i y_i + \epsilon_{i0t}$.

1. `blp` accommodates the case where markets are unbalanced, that is, J_t.

Both $\boldsymbol{\beta}_i$ and α_i are assumed to be linear functions of a $d \times 1$ vector of demographic factors, $\mathbf{D}_i$, and a $(K+1) \times 1$ vector of unobservable components, $\mathbf{v}_i$. In particular,

$$\begin{pmatrix} \boldsymbol{\beta}_i \\ \alpha_i \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta} \\ \alpha \end{pmatrix} + \boldsymbol{\Pi}\mathbf{D}_i + \mathbf{L}\mathbf{v}_i \tag{2}$$

where $\boldsymbol{\Pi}$ and $\mathbf{L}$ are $(K+1) \times d$ and $(K+1) \times (K+1)$ matrices, respectively. Although both $\mathbf{D}_i$ and $\mathbf{v}_i$ are typically unobserved, the distribution of the demographics, $P(\mathbf{D}_i)$, is assumed to be known. For $\mathbf{v}_i$, it is assumed that $\mathbf{L}\mathbf{v}_i \sim \text{i.i.d.}N(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^{'}$ is the covariance matrix of the coefficients $\boldsymbol{\beta}_i, \alpha_i$ conditional upon $\mathbf{D}_i$.[2]

Define the set $A_{ijt} = (\epsilon_{it} : u_{ijt} \geq u_{imt}, \forall m \neq j)$, where $\epsilon_{it} = (\epsilon_{ij0}, \ldots, \epsilon_{iJt})$, then the probability that individual i selects product j, in market t, given $\mathbf{D}_i$ and $\mathbf{v}_i$ is

$$\text{Pr}_{ijt} = \int_{A_{ijt}} dF(\epsilon_{it} \mid \mathbf{D}_i, \mathbf{v}_i) \tag{3}$$

Integrating out the unobservables $\mathbf{D}_i$ and $\mathbf{v}_i$ in (2) yields

$$\text{Pr}_{jt} = \int_{\mathbf{D}_i} \int_{\mathbf{v}_i} \text{Pr}_{ijt} dF(\mathbf{D}_i \mid \mathbf{v}_i) dF(\mathbf{v}_i) \tag{4}$$

The probability Pr_{jt} is the same for all i and can be estimated by the product market shares $s_{jt} = q_{jt}/I_t$, where q_{jt} denotes the sales. The error in this approximation is $O_p(I_t^{-1/2})$ and will be negligible for large I_t, which is often the case.

To evaluate the integrals in (4), it is assumed that the errors ϵ_{ijt} are independent and identically distributed (i.i.d.) and have a type-I extreme-value distribution. Then, from (2),

$$\text{Pr}_{ijt} = \frac{\exp\left(\mathbf{x}_{jt}^{'}\boldsymbol{\beta}_i - \alpha_i p_{jt} + \xi_{jt}\right)}{1 + \sum_{m=1}^{J} \exp\left(\mathbf{x}_{mt}^{'}\boldsymbol{\beta}_i - \alpha_i p_{mt} + \xi_{mt}\right)} \tag{5}$$

Because income y_i appears in the (indirect) utility function for all alternatives, including the outside option, $\alpha_i y_i$ cancels in the expression for Pr_{ijt}.

The integrals in (4) cannot be evaluated analytically, but they can be approximated by Monte Carlo integration with R random draws of $(\mathbf{D}_i, \mathbf{v}_i)$ from the distributions $P(\mathbf{D}_i)$ and $N(0, \mathbf{I}_{K+1})$. Letting $\delta_{jt} = \mathbf{x}_{jt}^{'}\boldsymbol{\beta} - \alpha p_{jt} + \xi_{jt}$ denote the mean utility,

$$s_{jt} = \frac{1}{R}\sum_{i=1}^{R} \text{Pr}_{ijt} = \frac{1}{R}\sum_{i=1}^{R} \frac{\exp\left\{\delta_{jt} + \left(\mathbf{x}_{jt}^{'}, -p_{jt}\right)(\boldsymbol{\Pi}\mathbf{D}_i + \mathbf{L}\mathbf{v}_i)\right\}}{1 + \sum_{m=1}^{J} \exp\left\{\delta_{mt} + \left(\mathbf{x}_{mt}^{'}, -p_{mt}\right)(\boldsymbol{\Pi}\mathbf{D}_i + \mathbf{L}\mathbf{v}_i)\right\}} \tag{6}$$

The simulation error can be reduced by increasing the number of draws, R.[3] This has a convergence rate of $O(R^{-1/2})$ and requires R to be increased by a factor of 100 for every additional digit of accuracy.

2. In practice, it is not necessary for all coefficients to be random or determined by the same set of demographic variables in $\mathbf{D}_i$. `blp` allows for each case.
3. The model in (4) is the aggregate counterpart of the random-coefficients logit model for individual-level choice data. The command `mixlogit` by Hole (2007) fits such models by maximum simulated likelihood using Halton draws.

An alternative method of variance reduction is to use quasi-random numbers. These are draws that represent nonrandom points within the domain of integration and provide a more-uniform coverage of the sampling distribution. A leading example are draws derived from Halton sequences. This method was introduced to the econometrics literature by Train (2000) and has a convergence rate of $O\{R^{-1}\log(R)^a\}$, where a is the dimension of the integral. This suggests that for modest R and small a, Monte Carlo integration using Halton draws will provide more-accurate approximations to the integrals in (4) compared with approximations based on random variates. Drukker and Gates (2006) examine this issue and conclude that Halton draws should be used for $a \leq 10$.

Polynomial-based integration is an additional method that offers further improvements over simulation. This approach is adopted by Skrainka and Judd (2011), who approximate the market-share integrals in (4) using three multidimensional polynomial-based rules.

Finally, for models that contain demographic variables, Halton draws or polynomial-based integration over $\mathbf{v}$ should offer no improvement when the integrals over $\mathbf{D}$ are approximated using random draws.[4]

Elasticities

To illustrate the advantages of the BLP model, consider the case of the logit specification, where preferences are homogeneous across consumers, ($\boldsymbol{\beta}_i = \boldsymbol{\beta}$, $\alpha_i = \alpha$). Then, from (3), the own- and cross-price elasticities are given by

$$e_{jkt} = \begin{cases} -\alpha p_{jt}(1 - s_{jt}) & \text{if } j = k \\ \alpha p_{kt} s_{kt} & \text{if } j \neq k \end{cases}$$

Because shares are often small, own-price elasticities will be proportional to price. This suggests that cheaper products are less elastic and have higher markups over marginal costs. This assumption is clearly implausible for many industries. A further limitation is implied by the cross-price elasticities, which impose unrealistic restrictions on the patterns of substitution between products. For example, if a red wine and a white wine have similar market shares, the logit model restricts the increase in their sales for a rise in the price of another brand of red wine to be the same. In practice, we would expect more consumers to substitute toward products that are similar and less toward products whose characteristics are different.

As Nevo (2000b) explains, the restrictive nature of the cross-price elasticities is due to the i.i.d. structure of the error terms, ϵ_{ijt}. Although the ranking of the products will differ across consumers, the probability of selecting a particular alternative will be the same as the population average, which is simply the market share.

To avoid this problem, the utilities in (1) must be correlated across brands. This is introduced by including $\mathbf{D}_i$ and $\mathbf{v}_i$ in (2), which generates correlation between products

4. This results from the property $O(a) + O(b) = O(\max\{a, b\})$. Hence, if using Halton draws for $\mathbf{v}$, the error will be $O[\max\{R^{-1/2}, R^{-1}\log(R)^{k+1}\}]$, which is at least as large as $O(R^{-1/2})$.

with similar characteristics. Furthermore, consumers with the same demographics will have similar preferences and, hence, similar patterns of substitution. From (4), the own- and cross-price elasticities for the BLP model become the following:

$$e_{jkt} = \begin{cases} -\frac{p_{jt}}{s_{jt}} \int \alpha_i \mathrm{Pr}_{ijt}(1 - \mathrm{Pr}_{ijt}) dF(\mathbf{D}_i, \mathbf{v}_i) & \text{if } j = k \\ \frac{p_{kt}}{s_{jt}} \int \alpha_i \mathrm{Pr}_{ijt} \mathrm{Pr}_{ikt} dF(\mathbf{D}_i, \mathbf{v}_i) & \text{if } j \neq k \end{cases} \tag{7}$$

The price sensitivity is now a probability-weighted average and can differ over products. As such, the model allows for flexible patterns of substitution that are more likely to be observed in the data. The integrals in (5) are again approximated by simulation.

2.2 Supply

Following Reynaert and Verboven (2014), let the marginal cost of product j in market t be independent of output and described by

$$c_{jt} = \mathbf{x}_{jt}^{'} \boldsymbol{\gamma}_1 + \mathbf{w}_{jt}^{'} \boldsymbol{\gamma}_2 + \zeta_{jt} \tag{8}$$

where $\mathbf{w}_{jt}$ is a vector of variables that affects marginal costs and ζ_{jt} is an unobserved (to the econometrician) component. Under perfect competition, $p_{jt} = c_{jt}$, and in vector notation, the supply side is described by

$$\mathbf{p}_t = \mathbf{X}_t \boldsymbol{\gamma}_1 + \mathbf{W}_t \boldsymbol{\gamma}_2 + \boldsymbol{\zeta}_t \tag{9}$$

For markets characterized by differentiated yet substitutable products, it may be reasonable to assume that firms engage in multiproduct Bertrand price competition. In this case, each company selects prices to maximize profits, given own-product attributes and the prices and attributes of other products in the market. The supply side under this model of imperfect competition is described by

$$\mathbf{p}_t = \mathbf{c}_t + \boldsymbol{\Delta}_t^{-1} (\boldsymbol{\delta}_t, \mathbf{L}, \boldsymbol{\Pi}) \, \mathbf{s}_t (\boldsymbol{\delta}_t, \mathbf{L}, \boldsymbol{\Pi}) \tag{10}$$

where $\boldsymbol{\Delta}_t$ is a $J \times J$ matrix with i, j-element $\Delta_{tij} = -\partial s_{jt} / \partial p_{it}$ if products i and j are manufactured by the same company and 0 otherwise. Prices now include a markup $\boldsymbol{\Delta}_t^{-1} \mathbf{s}_t$, which depends on $\boldsymbol{\delta}_t(\mathbf{p}_t, \mathbf{X}_t, \boldsymbol{\xi}_t)$. Thus, under this form of competition, the Nash-equilibrium prices will be a function of $\mathbf{X}_t$, $\mathbf{W}_t$, and the unobservables $\boldsymbol{\zeta}_t$ and $\boldsymbol{\xi}_t$.

`blp` follows the recent literature and estimates the parameters of the demand system by GMM. Unlike BLP, who estimate the parameters of both the demand and supply side, the relationships in (7) and (8) are used only to generate additional instruments to accommodate the endogeneity of prices. This will occur in perfect competition in (6) if ξ_{jt} is correlated with ζ_{jt}, and it is likely to be exacerbated under imperfect competition in (8), because the markups $\boldsymbol{\Delta}_t^{-1} \mathbf{s}_t$ contain the vectors of unobservables $\boldsymbol{\zeta}_t$ and $\boldsymbol{\xi}_t$.

3 GMM estimation

To identify the demand parameters, the product characteristics, $\mathbf{X}_t$, and the cost shifters, $\mathbf{W}_t$, are assumed to be mean independent of the unobserved component, ξ_{jt}.

$$E\left(\xi_{jt} \mid \mathbf{X}_t, \mathbf{W}_t\right) = 0, \;\; j = 1, \ldots, J \tag{11}$$

The conditional-moment restriction in (9) implies an infinite number of unconditional-moment restrictions,

$$E\left(\mathbf{z}_{jt}\xi_{jt}\right) = \mathbf{0} \tag{12}$$

where $\mathbf{z}_{jt} = g_{jt}\left(\mathbf{X}_t, \mathbf{W}_t\right)$ is a vector of instruments that are functions of the product characteristics and cost drivers of all products (see section 3.1).

Following the exposition by Nevo (2000a), the GMM estimator is carried out in three steps, including an initial stage, as follows:

0. For each market, draw R individuals $\mathbf{v}_1, \ldots, \mathbf{v}_R$ and $\mathbf{D}_1, \ldots, \mathbf{D}_R$. These are used to approximate the integrals in (4) in step 1.[5]
1. For given values of $\mathbf{\Pi}$ and $\mathbf{L}$, solve each market for the vector $\boldsymbol{\delta}_t = (\delta_{1t}, \ldots, \delta_{Jt})^{'}$, such that shares from (4) equal observed market shares.
2. Compute the sample-moment conditions $T^{-1}\sum_{t=1}^{T}\mathbf{Z}_t^{'}\boldsymbol{\xi}_t$, where $\mathbf{Z}_t$ is a $J \times l$ set of instruments and $\boldsymbol{\xi}_t = (\xi_{1t}, \ldots, \xi_{Jt})^{'}$, and form the GMM objective function.
3. Search for the values $\boldsymbol{\beta}$, α, $\mathbf{\Pi}$, and $\mathbf{L}$ that minimize the GMM objective function.

The `blp` command restricts the off-diagonal elements of the covariance matrix $\mathbf{\Sigma}$ to be zero. Thus, $\mathbf{L} = \text{diag}(\sigma_1, \ldots, \sigma_{K+1})$, where σ_k denotes the standard deviation of the kth random coefficient for individual i, conditional upon $\mathbf{D}_i$. As such, any correlation between the taste parameters operates through the demographics only.

To simplify the exposition, let $\mathbf{x}_{jt}^{\dagger} = (\mathbf{x}_{jt}^{'}, p_{jt})^{'}$ be a $K+1$ vector of observed product characteristics that now includes price, and let $\mathbf{X}_t^{\dagger} = (\mathbf{x}_{1t}^{\dagger}, \ldots, \mathbf{x}_{Jt}^{\dagger})^{'}$ be the $J \times (K+1)$ matrix of observations for market t and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^{'}, \boldsymbol{\theta}_2^{'})^{'}$, where $\boldsymbol{\theta}_1^{'} = (\boldsymbol{\beta}^{'}, \alpha,)$ and $\boldsymbol{\theta}_2^{'} = \{\sigma_1, \ldots, \sigma_{K+1}, \text{vec}(\mathbf{\Pi}^{'})^{'}\}$. Each stage of the estimation is described in detail below.

Step 0: Drawing v and D

`blp` uses R random draws of the demographic variables $\mathbf{D}$ and Halton or pseudorandom draws (as an option) for the $K+1$ vector $\mathbf{v}$. These differ across markets, but do not vary during estimation (steps 1 through 3). To generate the Halton draws for market t, a matrix containing a Halton sequence of length R and dimension $K+1$ is created from the first $K+1$ primes. The set is started at point $1 + B + R(t-1)$, where B is

5. As Nevo (2000b) explains, identification of the demographic parameters $\mathbf{\Pi}$ requires data on several markets and with variation in the distribution of the demographics across markets.

discarded to reduce correlation between the sequences.[6] `blp` sets $R = 200$ by default, although it is recommended that practitioners investigate the sensitivity of parameter estimates to increasing numbers of draws. Reynaert and Verboven (2014, tab. 8) report that for $R = 50$, estimates based on pseudo-Monte Carlo integration are less precise and considerably more biased than estimates that result when $R = 200$.

Step 1: Contraction mapping

For each market $t = 1, \dots, T$, compute the $J \times 1$ vector of mean utilities $\boldsymbol{\delta}_t$ such that

$$\mathbf{s}(\boldsymbol{\delta}_t, \boldsymbol{\theta}_2) = \mathbf{s}_t \tag{13}$$

where $\mathbf{s}(\boldsymbol{\delta}_t, \boldsymbol{\theta}_2)$ are the predicted shares from (4) and $\mathbf{s}_t = \{s_{1t}, \dots, s_{Jt}\}'$ are the observed counterparts. This system of J equations is then solved using the contraction mapping suggested by BLP. For a given vector $\boldsymbol{\delta}_t^n$, this involves computing

$$\boldsymbol{\delta}_t^{n+1} = \boldsymbol{\delta}_t^n + \log \mathbf{s}_t - \log \{\mathbf{s}(\boldsymbol{\delta}_t^n, \boldsymbol{\theta}_2)\} \tag{14}$$

where n denotes the nth iteration of the process. Updating then continues using (3) and (11) until $\|\boldsymbol{\delta}_t^n - \boldsymbol{\delta}_t^{n-1}\|$ is below a specified tolerance level.

As Dubé, Fox, and Su (2012) demonstrate, a loose tolerance for the contraction mapping propagates into the GMM objective function and its derivatives and may lead to a nonconvergence of the optimization routine.[7] To avoid this situation, `blp` sets a default inner-loop tolerance to 10E–15. Iteration is also over $\exp(\boldsymbol{\delta}_t)$, and $\boldsymbol{\delta}_t$ is recovered at convergence. This removes the need to compute logs and exponentials within each loop and, therefore, saves considerable computational time.

Step 2: GMM objective function

Let $\mathbf{Z}_t = (\mathbf{z}_{1t}, \dots, \mathbf{z}_{Jt})'$ be a $J \times l$ matrix of instruments that satisfy (10) such that $E\{\mathbf{h}_t(\boldsymbol{\theta}_0)\} = E\{\mathbf{Z}_t' \boldsymbol{\xi}_t(\boldsymbol{\theta}_0)\}$ where $\boldsymbol{\xi}_t(\boldsymbol{\theta}_0) = \boldsymbol{\delta}_t(\boldsymbol{\theta}_{2,0}) - \mathbf{X}_t^{\dagger} \boldsymbol{\theta}_{1,0}$ and $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{1,0}', \boldsymbol{\theta}_{2,0}')'$ are the true population parameters. The GMM method replaces the population-moment conditions with the sample counterparts $\overline{\mathbf{h}}(\boldsymbol{\theta}) = T^{-1} \sum_{t=1}^{T} \mathbf{h}_t(\boldsymbol{\theta})$ and selects $\boldsymbol{\theta}$ as the values that minimize the objective function

$$Q = \overline{\mathbf{h}}(\boldsymbol{\theta})' \mathbf{A}_T \overline{\mathbf{h}}(\boldsymbol{\theta})$$

where $\mathbf{A}_T$ is a positive-definite weighting matrix that is independent of $\boldsymbol{\theta}$. The subscript T indicates reliance on the data, and identification of the parameters requires that $l \geq (K+1)(2+d)$.

6. Hole (2007) sets $B = 15$ by default in `mixlogit`.
7. The contraction mapping is a nested fixed-point algorithm. Dubé, Fox, and Su (2012) eliminate this step using mathematical programming with equilibrium constraints. This minimizes the GMM objective function subject to the market-share equations as constraints.

Step 3: Parameter search

For the case where the model is just identified as $\{l = (K+1)(2+d)\}$, the estimator $\widehat{\boldsymbol{\theta}}$ is the solution to $\overline{\mathbf{h}}(\widehat{\boldsymbol{\theta}}) = 0$, which does not involve $\mathbf{A}_T$. For overidentified models, selection of the weighting matrix yields different estimators that are consistent under regulatory assumptions, but with varying degrees of efficiency.

Letting $\mathbf{G}_T(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^{T}(\partial \boldsymbol{h}_t/\partial \boldsymbol{\theta}^{'})$, the GMM estimator for the overidentified model solves the following first-order conditions:

$$2\mathbf{G}_T(\widehat{\boldsymbol{\theta}})^{'}\mathbf{A}_T\overline{\mathbf{h}}(\widehat{\boldsymbol{\theta}}) = \mathbf{0} \tag{15}$$

The elements of $\mathbf{G}_T$ are given by

$$\mathbf{G}_T(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^{T}\mathbf{Z}_t^{'}\left\{\mathbf{X}_t^{\dagger}, \boldsymbol{D}_{\boldsymbol{\theta}_2}\boldsymbol{\delta}_t(\boldsymbol{\theta}_2)\right\} \tag{16}$$

where $\boldsymbol{D}_{\boldsymbol{\theta}_2}\boldsymbol{\delta}_t$ denotes the $J \times (K+1)(1+d)$ matrix of derivatives of $\boldsymbol{\delta}_t$ with respect to $\boldsymbol{\theta}_2^{'}$. To reduce the search time, $\boldsymbol{\theta}_1$ can be written as an explicit function of $\boldsymbol{\theta}_2$,

$$\widehat{\boldsymbol{\theta}}_1 = (\mathbf{X}^{\dagger'}\mathbf{Z}\mathbf{A}_T\mathbf{Z}^{'}\mathbf{X}^{\dagger})^{-1}\mathbf{X}^{\dagger'}\mathbf{Z}\mathbf{A}_T\mathbf{Z}^{'}\boldsymbol{\delta}(\boldsymbol{\theta}_2) \tag{17}$$

where $\mathbf{X}^{\dagger} = (\mathbf{X}_1^{\dagger'}, \ldots, \mathbf{X}_T^{\dagger'})^{'}$, $\mathbf{Z} = (\mathbf{Z}_1^{'}, \ldots, \mathbf{Z}_T^{'})^{'}$, and $\boldsymbol{\delta} = (\boldsymbol{\delta}_1^{'}, \ldots, \boldsymbol{\delta}_T^{'})^{'}$. Ignoring proportionality constants, the search is now limited to $\boldsymbol{\theta}_2$, where the estimator $\widehat{\boldsymbol{\theta}}_2$ solves

$$\left\{\sum_{t=1}^{T}\mathbf{Z}_t^{'}\boldsymbol{D}_{\boldsymbol{\theta}_2}\boldsymbol{\delta}_t\left(\widehat{\boldsymbol{\theta}}_2\right)\right\}^{'}\mathbf{A}_T\left\{\sum_{t=1}^{T}\mathbf{Z}_t^{'}\boldsymbol{\xi}_t\left(\widehat{\boldsymbol{\theta}}\right)\right\} = \mathbf{0} \tag{18}$$

To use a Newton method, the analytical derivatives $D_{\theta_2}\boldsymbol{\delta}_t$ in the first-order conditions in (15) are required. These can be computed by an application of the implicit function theorem to $\mathbf{s}\{\boldsymbol{\delta}_t(\boldsymbol{\theta}_2), \boldsymbol{\theta}_2\} = \mathbf{s}_t$, which yields the following:

$$\boldsymbol{D}_{\boldsymbol{\theta}_2}\boldsymbol{\delta}_t = -(\boldsymbol{D}_{\boldsymbol{\delta}_t}\mathbf{s}_t)^{-1}\boldsymbol{D}_{\boldsymbol{\theta}_2}\mathbf{s}_t \tag{19}$$

The elements inside the matrices of (16) are

$$\boldsymbol{D}_{\boldsymbol{\theta}_2}\boldsymbol{\delta}_t = -\begin{pmatrix} \frac{\partial s_{1t}}{\partial \delta_{1t}} & .. & \frac{\partial s_{1t}}{\partial \delta_{Jt}} \\ \vdots & \vdots & \vdots \\ \frac{\partial s_{Jt}}{\partial \delta_{1t}} & .. & \frac{\partial s_{Jt}}{\partial \delta_{Jt}} \end{pmatrix}^{-1}\begin{pmatrix} \frac{\partial s_{1t}}{\partial \sigma_1} & .. & \frac{\partial s_{1t}}{\partial \sigma_{K+1}}, & \frac{\partial s_{1t}}{\partial \pi_{11}} & .. & \frac{\partial s_{1t}}{\partial \pi_{(K+1)d}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial s_{Jt}}{\partial \sigma_1} & .. & \frac{\partial s_{Jt}}{\partial \sigma_{K+1}}, & \frac{\partial s_{Jt}}{\partial \pi_{11}} & .. & \frac{\partial s_{Jt}}{\partial \pi_{(K+1)d}} \end{pmatrix}$$

where, for example, $\partial s_{jt}/\partial \pi_{kd}$ is the derivative of s_{jt} with respect to π_{kd}, where π_{kd} measures the impact of demographic variable d on the kth stochastic coefficient.

From the simulator in (4), the derivatives in the matrices of (16) are given by

$$\begin{aligned}
\frac{\partial s_{jt}}{\partial \delta_{jt}} &= R^{-1}\sum_{i=1}^{R}\mathrm{Pr}_{ijt}(1-\mathrm{Pr}_{ijt})\\
\frac{\partial s_{jt}}{\partial \delta_{mt}} &= R^{-1}\sum_{i=1}^{R}\mathrm{Pr}_{ijt}\mathrm{Pr}_{imt}\\
\frac{\partial s_{jt}}{\partial \sigma_{k}} &= R^{-1}\sum_{i=1}^{R}\mathrm{Pr}_{ijt}v_{ik}\left(x_{jtk}-\sum_{m=1}^{J}x_{mtk}\mathrm{Pr}_{imt}\right)\\
\frac{\partial s_{jt}}{\partial \pi_{kd}} &= R^{-1}\sum_{i=1}^{R}\mathrm{Pr}_{ijt}D_{id}\left(x_{jtk}-\sum_{m=1}^{J}x_{mtk}\mathrm{Pr}_{imt}\right)
\end{aligned}$$

where Pr_{ijt} is defined in (3), and x_{mtk} denotes the observation for alternative m in market t on product characteristic k.[8]

To implement the estimation algorithm, initial values of $\boldsymbol{\theta_2}$ are needed. `blp` sets these to 0.5 but provides the user with the option to specify alternative starting values if required. This is important, because recent literature suggests that the objective function may not be globally concave, leading to convergence at local extrema. Knittel and Metaxoglou (2014) explore this issue using pseudo-Monte Carlo integration with $R \in (25, 50)$ and for different starting values and alternative optimization routines. They report considerable variation in objective-function values and multiple instances of convergence failure. These issues are associated with the nonlinearity of the model, which leads to an objective function that is not globally concave. They suggest that in the absence of prior information on $\boldsymbol{\theta_2}$, a large number of starting values should be generated using random draws. Estimation should then proceed using several different optimization algorithms and stopping rules based on tight outer tolerances.[9] Starting values and optimization algorithms that yield the lowest objective-function values should then be selected for the final parameter search.

Skrainka and Judd (2011) report similar issues using pseudo-Monte Carlo integration and attribute these to ripples on the surface of the GMM objective function that lead to false local minimums. Experiments are carried out using different starting values for the parameters, $R \in (1000, 10000)$ in the pseudo-Monte Carlo integration, and approximations to (4) using polynomial-based rules. The results show that while larger R increases stability, variation persists in both point estimates and objective-function values. This is not the case when (4) is approximated using polynomial-based integration, where results are virtually identical across all sets of starting values.

8. To speed up the algorithm, `blp` uses vectorization throughout.
9. `blp` sets the outer tolerance to 10e–12.

3.1 Instruments

For the demand-side model where $E(\boldsymbol{\xi}_t\boldsymbol{\xi}_t^{'} \mid \mathbf{X}_t, \mathbf{W}_t) = \mathbf{I}_J\sigma_\xi^2$, the Chamberlain (1987) optimal instruments are

$$\mathbf{z}_{jt}^* = E\left(\frac{\partial \xi_{jt}}{\partial \boldsymbol{\theta}} \mid \mathbf{X}_t, \mathbf{W}_t\right)$$

This is the conditional expectation of the derivative of the conditional-moment restriction with respect to vector $\boldsymbol{\theta}$. As demonstrated by Chamberlain (1987), $\mathbf{z}_{jt}^*$ will minimize the asymptotic covariance matrix of the estimator $\widehat{\boldsymbol{\theta}}$.

Reynaert and Verboven (2014) evaluate the small-sample properties of the BLP–GMM estimator for different sets of instruments $\mathbf{z}_{jt} = \mathbf{g}_{jt}(\mathbf{X}_t, \mathbf{W}_t)$. In particular, their analysis uses simulation to compare the bias, efficiency, and stability of the estimator using optimal instruments with two sets based on series approximations that are commonly applied in the literature. They conclude that cost-side drivers should be used to identify the price parameters[10] and that optimal instruments should be applied to identify the variances of the random coefficients. Computation of these instrument sets are described in the following.

Standard instruments

Most applications of the BLP model use instruments that represent series approximations of $\mathbf{z}_{jt}^*$. In the Monte Carlo experiments by Reynaert and Verboven (2014), two sets of suboptimal instruments are constructed. The first, denoted $\mathbf{z}_{jt}^1$, is a second-order polynomial of $(\mathbf{x}_{jt}, \mathbf{w}_{jt})$ and resembles the instruments used by Dubé, Fox, and Su (2012). These comprise $\mathbf{x}_{jt}$, $\mathbf{w}_{jt}$, and their squares and interactions. Hence, for one demand characteristic and one cost component, $\mathbf{z}_{jt}^1 = (x_{jt}, w_{jt}, x_{jt}^2, w_{jt}^2, x_{jt}w_{jt})^{'}$.

Their second set, denoted $\mathbf{z}_{jt}^2$, extends $\mathbf{z}_{jt}^1$ by adding the sums of the characteristics of other products $\sum_{m=1,m\neq j}^{J} \mathbf{x}_{mt}$. These additional instruments resemble those used by BLP[11] and are assumed to assist in the identification of the heterogeneity parameters, $\boldsymbol{\theta}_2$.[12] The results of their simulations show that for a model with one random coefficient, the estimator using suboptimal instruments is biased, inefficient, and unstable.[13] Furthermore, while including BLP-type instruments improves performance, the benefits are small. These results generalize to models with multiple random coefficients.

10. This finding was initially reported by Armstrong (2014).
11. More precisely, BLP use the following instruments: $\mathbf{x}_{jt}$, $\mathbf{w}_{jt}$, $\sum_{m\in\mathcal{F}_j,m\neq j}(\mathbf{x}_{mt}, \mathbf{w}_{mt})$, and $\sum_{m\notin\mathcal{F}_j,m\neq j}(\mathbf{x}_{mt}, \mathbf{w}_{mt})$, where $\mathcal{F}_j$ contains the set of all products that are manufactured by the same company that produces product j.
12. `blp` requires the user to create and specify all suboptimal instruments.
13. This refers to the case where the distribution of the standard deviation of the taste parameter exhibits a large number of spikes around zero.

Optimal instruments

In vector notation, the unconditional-moment restrictions for each market satisfy $E(\mathbf{Z}_t^{*'}\boldsymbol{\xi}_t) = 0$, where the matrix of optimal instruments $\mathbf{Z}_t^*$ is described as

$$\mathbf{Z}_t^* = E\left(\boldsymbol{D}_{\boldsymbol{\beta}}\boldsymbol{\xi}_t, \boldsymbol{D}_{\alpha}\boldsymbol{\xi}_t, \boldsymbol{D}_{\boldsymbol{\theta}_2}\boldsymbol{\xi}_t \mid \mathbf{X}_t, \mathbf{W}_t\right) \tag{20}$$

The expectation of the first component in (5) is simply

$$E\left(\boldsymbol{D}_{\boldsymbol{\beta}}\boldsymbol{\xi}_t \mid \mathbf{X}_t, \mathbf{W}_t\right) = -E\left(\mathbf{X}_t \mid \mathbf{X}_t, \mathbf{W}_t\right) = -\mathbf{X}_t \tag{21}$$

For the second component in (5), a supply-side assumption is required. Under perfect competition in (7), this expectation becomes the following:

$$E\left(\boldsymbol{D}_{\alpha}\boldsymbol{\xi}_t \mid \mathbf{X}_t, \mathbf{W}_t\right) = -E\left(\mathbf{p}_t \mid \mathbf{X}_t, \mathbf{W}_t\right) = -(\mathbf{X}_t\boldsymbol{\gamma}_1 + \mathbf{W}_t\boldsymbol{\gamma}_2) \tag{22}$$

A consistent estimator of the expectation in (22) is then the predicted price from an ordinary least-squares (OLS) regression of p_{jt} on $\mathbf{x}_{jt}$ and $\mathbf{w}_{jt}$. As noted by Reynaert and Verboven (2014), the instruments for the linear parameters $\boldsymbol{\theta}_1$ in (21) and (22) are the same as those from the first stage in a two-stage least-squares estimator.

To compute $E(\mathbf{p}_t \mid \mathbf{X}_t, \mathbf{W}_t)$ under imperfect competition in (8), it is necessary to first solve for the vector of equilibrium prices, $\mathbf{p}_t = \mathbf{f}(\mathbf{W}_t, \mathbf{X}_t, \boldsymbol{\xi}_t, \boldsymbol{\zeta}_t)$. The expectation is then with respect to the density of the demand and the cost-side unobservables. Reynaert and Verboven (2014) follow BLP and approximate the expectation by setting $\boldsymbol{\xi}_t = \boldsymbol{\zeta}_t = 0$. However, they also report more accurate predictions from an OLS regression of prices p_{jt} on a polynomial of the demand characteristics, cost drivers, and $\sum_{m\neq j}\mathbf{x}_{mt}$.

The final component in (5) is the optimal set of instruments for the heterogeneity parameters, $\boldsymbol{\theta}_2$. This is given by

$$E\left(\boldsymbol{D}_{\boldsymbol{\theta}_2}\boldsymbol{\xi}_t \mid \mathbf{X}_t, \mathbf{W}_t\right) = E\left(\boldsymbol{D}_{\boldsymbol{\theta}_2}\boldsymbol{\delta}_t \mid \mathbf{X}_t, \mathbf{W}_t\right) \tag{23}$$

Reynaert and Verboven (2014) approximate the expectation using simulation, but report similar gains in efficiency using the method proposed by Berry, Levinsohn, and Pakes (1999). This evaluates the derivatives in (23) at the expected value of the unobservables $\boldsymbol{\xi}_t = \boldsymbol{\zeta}_t = 0$ and is computationally faster. This procedure is described as follows:

1. Estimate the parameters of the model with suboptimal instruments to obtain $\widehat{\boldsymbol{\theta}}$.

2. Predict prices $\widehat{\mathbf{p}}_t = \mathbf{X}_t\widehat{\boldsymbol{\gamma}}_1 + \mathbf{W}_t\widehat{\boldsymbol{\gamma}}_2$, where $\widehat{\boldsymbol{\gamma}}_1$ and $\widehat{\boldsymbol{\gamma}}_2$ are the OLS estimates. Functions of $\mathbf{W}_t$ and $\mathbf{X}_t$ can be included if necessary. For expectations that are nonlinear in the parameters, (22) will be an approximation.

3. Compute the predicted mean utility $\widehat{\boldsymbol{\delta}}_t = \mathbf{X}_t\widehat{\boldsymbol{\beta}} - \widehat{\alpha}\widehat{\mathbf{p}}_t$, and substitute for $\boldsymbol{\delta}_t$ in the components of (16). This provides $E(\mathbf{D}_{\boldsymbol{\theta}_2}\boldsymbol{\xi}_t \mid \mathbf{X}_t, \mathbf{W}_t) \approx \mathbf{D}_{\boldsymbol{\theta}_2}\widehat{\boldsymbol{\delta}}_t$.

Reynaert and Verboven (2014) report considerable improvements in the performance of the estimator when using optimal instruments. In particular, they find that optimal instruments reduce the small-sample bias and improve the efficiency and stability of the estimator when compared with estimation using the series approximations $\mathbf{z}_{jt}^1$ and $\mathbf{z}_{jt}^2$.

`blp` includes an option to compute and use optimal instruments. The program follows the approximation in steps 1 through 3 and permits one or more variables to be endogenous (not just price). This requires the user to specify subsets or functions of the instruments that appear in step 2 ($\mathbf{x}_{jt}$ is included by default). The iterative GMM estimator can also be specified, which updates the values in step 3 from the previous round of parameter estimates. This process continues until convergence is achieved or until the maximum number of iterations has been reached.[14]

3.2 Distribution of the GMM estimator

For fixed J and $T \to \infty$, the GMM estimator

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_{\text{GMM}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} N\left\{0, \left(\mathbf{G}_0^{'}\mathbf{A}_0\mathbf{G}_0\right)^{-1}\mathbf{G}_0^{'}\mathbf{A}_0\mathbf{S}_0\mathbf{A}_0\mathbf{G}_0\left(\mathbf{G}_0^{'}\mathbf{A}_0\mathbf{G}_0\right)^{-1}\right\} \tag{24}$$

where $\mathbf{G}_0 = \text{plim}\mathbf{G}_T(\boldsymbol{\theta}_0)$, $\mathbf{A}_0 = \text{plim}\mathbf{A}_T$, and $\mathbf{S}_0 = \text{plim}T^{-1}\sum_{t=1}^{T}\mathbf{h}_t(\boldsymbol{\theta}_0)\mathbf{h}_t^{'}(\boldsymbol{\theta}_0)$, is the asymptotic variance matrix of the sample-moment conditions. The optimal weighting matrix sets $\mathbf{A}_0 \propto \mathbf{S}_0^{-1}$, which yields

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_{\text{GMM}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} N\left\{0, \left(\mathbf{G}_0^{'}\mathbf{S}_0^{-1}\mathbf{G}_0\right)^{-1}\right\} \tag{25}$$

Inference can proceed by replacing $\mathbf{G}_0$ with $\mathbf{G}_T(\widehat{\boldsymbol{\theta}})$ from (13) and $\mathbf{S}_0$ with

$$\mathbf{S}_T\left(\widehat{\boldsymbol{\theta}}\right) = T^{-1}\widehat{\sigma}_\xi^2\mathbf{Z}^{'}\mathbf{Z} \tag{26}$$

if the errors ξ_{jt} are i.i.d., where $\widehat{\sigma}_\xi^2 = (JT)^{-1}\sum_{t=1}^{T}\widehat{\boldsymbol{\xi}}_t^{'}\widehat{\boldsymbol{\xi}}_t$ and $\widehat{\boldsymbol{\xi}}_t = \boldsymbol{\delta}_t - \mathbf{X}_t\widehat{\boldsymbol{\theta}}_1$, or with

$$\mathbf{S}_T(\widehat{\boldsymbol{\theta}}) = T^{-1}\sum_{t=1}^{T}\left(\mathbf{Z}_t^{'}\widehat{\boldsymbol{\xi}}_t\widehat{\boldsymbol{\xi}}_t^{'}\mathbf{Z}_t\right) \tag{27}$$

if ξ_{jt} is considered to be heteroskedastic across markets and correlated over products within each market.

14. `blp` uses the original starting values for the parameter search following step 3. This avoids the optimizer converging at local minimums for poor first-round estimates of $\boldsymbol{\theta}_2$. Subsequent iterations use estimates from the previous round. Hall (2005, sec. 2.4 and 3.6) mentions that there may be gains to finite-sample efficiency with the iterative GMM estimator.

When (6.3) is a consistent estimator of $\mathbf{S}_0$, the weighting matrix in (14) and (15) can be replaced by $\mathbf{A}_T = (T^{-1}\mathbf{Z}'\mathbf{Z})^{-1}$. Then, from (25), the GMM estimator is asymptotically normally distributed, with mean $\boldsymbol{\theta}_0$ and estimated asymptotic variance.

$$\widehat{\mathbf{V}}\left(\widehat{\boldsymbol{\theta}}\right) = T^{-1}\left\{\mathbf{G}_T\left(\widehat{\boldsymbol{\theta}}\right)\mathbf{S}_T\left(\widehat{\boldsymbol{\theta}}\right)^{-1}\mathbf{G}_T\left(\widehat{\boldsymbol{\theta}}\right)\right\}^{-1} \tag{28}$$

If $\mathbf{S}_0$ is consistently estimated by (27), but $\mathbf{A}_T = (T^{-1}\mathbf{Z}'\mathbf{Z})^{-1}$ has been imposed, a panel–robust estimate of $\mathbf{V}(\widehat{\boldsymbol{\theta}})$ can be computed from (24). This is given by

$$\begin{aligned}\mathbf{V}\left(\widehat{\boldsymbol{\theta}}\right) = T^{-1}&\left\{\mathbf{G}_T'\left(\widehat{\boldsymbol{\theta}}\right)\mathbf{A}_T\mathbf{G}_T\left(\widehat{\boldsymbol{\theta}}\right)\right\}^{-1}\mathbf{G}_T\left(\widehat{\boldsymbol{\theta}}\right)'\mathbf{A}_T\mathbf{S}_T\left(\widehat{\boldsymbol{\theta}}\right)\mathbf{A}_T\mathbf{G}_T\left(\widehat{\boldsymbol{\theta}}\right)\\ &\left\{\mathbf{G}_T'\left(\widehat{\boldsymbol{\theta}}\right)\mathbf{A}_T\mathbf{G}_T\left(\widehat{\boldsymbol{\theta}}\right)\right\}^{-1}\end{aligned}$$

where $\mathbf{S}_T(\widehat{\boldsymbol{\theta}})$ is computed postestimation.

For the more-efficient GMM estimator (using suboptimal instruments), the weighting matrix $\mathbf{A}_T = \mathbf{S}_T(\widehat{\boldsymbol{\theta}})^{-1}$ from (27). This method requires the model to be fit twice. In the first round, $\mathbf{A}_T = (T^{-1}\mathbf{Z}'\mathbf{Z})^{-1}$ to obtain a consistent estimate $\widehat{\boldsymbol{\theta}}$. The second round then uses $\widehat{\boldsymbol{\theta}}$ to compute $\mathbf{S}_T(\widehat{\theta})$ and the optimal weighting matrix $\mathbf{A}_T$. Equation (28) provides the estimated asymptotic variance of this optimal two-step GMM estimator.

Finally, when the model is exactly identified as $\{l = (K+1)(2+d)\}$, the first-order conditions in (12) can be premultiplied by $(\mathbf{G}_T'\mathbf{A}_T)^{-1}$, and the estimator will be the solution to $\overline{\mathbf{h}}(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$. As such, the same estimator is obtained for any full-rank weighting matrix $\mathbf{A}_T$. This will be the case if optimal instruments are used, because $\dim(\mathbf{Z}_t^*) = (K+1)(2+d)$ after the first round of estimation.[15] The estimated asymptotic variance matrix is then provided by (28) and will be robust if $\mathbf{S}_0$ is estimated by (27).

4 The blp command

`blp` fits the random-parameters logit demand model from product market shares using the algorithm discussed in section 3. The program requires the data to be in long form, where each market contains observations on product shares and characteristics. Products are not required to be the same across markets; hence, the panel can be unbalanced. The user can specify which coefficients are stochastic and include different sets of demographic variables in the equations for each random parameter.

The market-share integrals are approximated by quasi-Monte Carlo integration using Halton draws, with an option to switch to pseudorandom numbers. To investigate the effects of simulation error, it is recommended that the user tests the stability of the parameter estimates for progressively larger draws. If demographic variables are used, it is necessary to include a random sample of individuals from each market (in long

15. The solution $\widehat{\boldsymbol{\theta}}$ should, therefore, set the objective function $Q(\widehat{\boldsymbol{\theta}}) = 0$.

form) in a separate dataset. The dimensions of this sample must be the same for each market.

The parameters are estimated by GMM, and for overidentified models, `blp` permits the user to specify a weighting matrix that is optimal when the errors are correlated between products and heteroskedastic across markets. The default assumes that the errors are i.i.d., and to guard against misspecification, robust standard errors (SEs) can be reported.

Instruments must be included to identify the heterogeneity parameters and any coefficients associated with endogenous regressors. Thus, for a model that contains one endogenous regressor, two random coefficients, and no demographic variables, a minimum of three additional instruments must be specified. `blp` also provides an option to fit the model using Chamberlain (1987) optimal instruments. This requires the user to include standard instruments and subsets and functions of these instruments that appear in the linear (in parameters) conditional expectations in (22) when the model contains endogenous regressors.

Finally, the program provides the user with an option to compute the matrix of demand elasticities for any market and any variable with a stochastic coefficient. These are computed by Monte Carlo simulation, using the same set of draws for the market-share integrals. The user can also include a string variable that identifies the products and is used to label the rows and columns of the elasticity matrix.

4.1 Syntax

The syntax for `blp` is as follows:

`blp` *depvar* [*varlist*] [*if*] [*in*] , `endog(`[*varlist_endog*]`=`*varlist_inst1*`)`
`stochastic(`*varname_s1*`=`*varlist_s1*`,` *varname_s2*`=`*varlist_s2*`, ...)`
`markets(`*varname_m*`)` [`optinst`[`(`*varlist_inst2*`)`] `tolin(`*#*`)` `tolout(`*#*`)`
`draws(`*#*`)` `burn(`*#*`)` `iter`[`(`*#*`)`] `demofile(`*filename_d*`)` `initsd(`*initvals*`)`
`initdemo(`*initvals_d*`)` `elast(`*varname_e*`,` *#*[`,` *varname_p*]`)` `robustweight`
`robust` `random` `noisily` `nocons`]

4.2 Options

`endog(`[*varlist_endog*]`=`*varlist_inst1*`)` identifies *varlist_endog* as any endogenous variables and *varlist_inst1* as the instruments for *varlist_endog* and the parameters in `stochastic()`. `endog()` is required.

`stochastic(`*varname_s1*`=`*varlist_s1*`,` *varname_s2*`=`*varlist_s2*`, ...)` identifies *varname_s1*, *varname_s2*, ... as product characteristics with random coefficients and *varlist_s1*, *varlist_s2*, ... as demographic variables that appear in these equations.

Random coefficients can be associated with only the variables that appear in *varlist* or *varlist_endog*. For a random constant, it is necessary to generate and include a variable named `cons`. `stochastic()` is required.

`markets(`*varname_m*`)` identifies *varname_m* as the market variable in the data and in the demographic file *filename_d*, if used. `markets()` is required.

`optinst` [`(`*varlist_inst2*`)`] fits the model using Chamberlain (1987) optimal instruments. For models that include endogenous variables, *varlist_inst2* contains subsets and functions of *varlist_inst1* that appear in the linear (in parameters) conditional expectation of *varlist_endog*.

`tolin(`#`)` specifies the tolerance level used to define convergence of the contraction-mapping algorithm. The default is `tolin(10E-15)`.

`tolout(`#`)` specifies the tolerance level used to define convergence of the GMM estimator. The default is `tolout(10E-12)`.

`draws(`#`)` specifies the number of Halton draws used to approximate the market-share integrals. The default is `draws(200)`, and the Halton sequence is created from the first K primes, where K denotes the number of stochastic coefficients.

`burn(`#`)` specifies the number of initial elements to drop when creating Halton sequences. The default is `burn(15)`. This helps to reduce correlation between the sequences.

`iter` [`(`#`)`] specifies the iterative instead of the two-step GMM estimator. It is available for `optinst()` or `robustweight`, and estimation will continue until the relative difference between estimates in successive iterations is below `tolout()`. Alternatively, # specifies the number of iterations.

`demofile(`*filename_d*`)` identifies *filename_d* as the path to the file that contains the random draws of the demographic variables for each market.

`initsd(`*initvals*`)` identifies *initvals* as the starting values for the standard deviations of the random coefficients. These must be separated by a comma. The default is `initsd(0.5,0.5)`.

`initdemo(`*initvals_d*`)` identifies *initvals_d* as the starting values for the coefficients on the demographic variables. The order will correspond to *varlist_s1*, *varlist_s2*, ..., and values must be separated by a comma. The default is `initdemo(0.5,0.5)`.

`elast(`*varname_e*`,` # [`,` *varname_p*]`)` provides the matrix of demand elasticities for a 1% increase in variable *varname_e* in market number #. These are available for *varname_s1*, *varname_s2*, etc., only. An optional string variable *varname_p* can be specified to identify the products.

`robustweight` specifies a weighting matrix in the GMM estimator that is optimal when the errors are correlated between products and heteroskedastic across markets. This option cannot be used when the number of instruments equals the number of parameters or when `optinst()` is specified (because the model is exactly identified).

`robust` computes an estimate of the SEs that are robust when the errors are correlated between products and heteroskedastic across markets. The default assumes that the errors are i.i.d.

`random` specifies that pseudorandom draws be used to approximate the market-share integrals instead of those based on Halton sequences. Following Drukker and Gates (2006), it is suggested that this option be selected if the number of stochastic coefficients exceeds 10.

`noisily` displays the iteration log during estimation. This indicates convergence of the contraction mapping by market and displays the current values of the heterogeneity parameters and associated analytical gradients.

`nocons` fits the model without the constant term in the mean utility.

4.3 Remarks

`blp` requires the data to be in long form, where each market contains observations on product shares and characteristics. Products are not required to be the same across markets. The number of Halton draws is set to 200, but it is recommended that the user test the stability of the estimates using sequentially larger values.

If demographic variables are included, *filename_d* must contain an equal number of draws across markets in long form and have the same numeric market identifier, *varname_m*. The number of draws contained in this file will also override those specified by the user. Identification of the demographic coefficients requires data on multiple markets and variation in the distribution across markets. There is no requirement to demean the demographic draws to identify the parameters.

Finally, for models that include a random coefficient on the constant, Monte Carlo experiments indicate that it is very difficult to identify the standard deviation of this characteristic in addition to the associated coefficients on the demographic variables.

4.4 Stored results

`blp` stores the following in `e()`:

Scalars	
`e(N)`	number of observations
`e(T)`	number of markets
`e(conv)`	`1` if converged, `0` otherwise
`e(draws)`	number of simulation draws
`e(burn)`	dropped initial Halton sequences
`e(obj)`	objective-function value
Macros	
`e(cmd)`	`blp`
Matrices	
`e(b)`	coefficient vector
`e(V)`	variance–covariance matrix of the estimators
`e(initsd)`	initial values for standard deviations of random coefficients
`e(initd)`	initial values for coefficients on demo variables
`e(elast)`	elasticity matrix
Functions	
`e(sample)`	marks estimation sample

5 Examples

5.1 No demographic variables

In this example, consumers can select from $J = 10$ alternatives, excluding the outside good. Data are simulated for $T = 25$ markets, and utility is determined by a constant, one product characteristic `x1`, and price `p` (which is endogenous). The supply side is characterized by perfect competition, where marginal costs are a linear function of the product characteristics and three exogenous cost drivers `w1`, `w2`, and `w3`. Heterogeneity is restricted to the coefficient on `x1`, which has a true mean valuation of 2 and a standard deviation of 1. The constant is set to 2, and the coefficient on price is -2.

Standard instruments

The model is initially fit by generating the instrument set $\mathbf{z}_{jt}^2$ (section 3.1). This contains the exogenous variables, their squares and interactions, and the sums of the characteristics of other products. Construction of these instruments is given below.

```
. use blp_nodemo
. generate w12=w1^2
. generate w22=w2^2
. generate w32=w3^2
. generate x12=x1^2
. generate x1w1=x1*w1
. generate x1w2=x1*w2
. generate x1w3=x1*w3
. bysort mkt: egen x1s=sum(x1)
```

```
. replace x1s=x1s-x1
(250 real changes made)
. blp s x1, stochastic(x1) endog(p=w1 w2 w3 w12 w22 w32 x12 x1w1 x1w2 x1w3 x1s)
> markets(mkt)
Iteration 0:   f(p) =  13.131515
  (output omitted)
Iteration 4:   f(p) =  12.941619
GMM estimator of BLP-model
GMM weight matrix: unadjusted             Number of obs           =  250
                                          Number of markets       =  25
                                          Number of Halton draws  =  200
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Mean utility						
cons	1.86419	.7104963	2.62	0.009	.4716428	3.256737
x1	2.411279	.6008964	4.01	0.000	1.233544	3.589014
p	-2.040571	.0488951	-41.73	0.000	-2.136403	-1.944738
x1						
SD	.7360524	.4879043	1.51	0.131	-.2202225	1.692327

Optimal instruments

Prices are a linear function of `w1`, `w2`, `w3`, and `x1` and a constant. Hence, to fit the model using optimal instruments, `optinst(w1 w2 w3)` is specified, where `x1` and the constant is included by default.

```
. blp s x1, stochastic(x1) endog(p=w1 w2 w3 w12 w22 w32 x12 x1w1 x1w2 x1w3 x1s)
> markets(mkt) optinst(w1 w2 w3)
Iteration 0:   f(p) =  13.131515
  (output omitted)
Iteration 4:   f(p) =  12.941619
Estimation iteration with optimal instruments: 1
Iteration 0:   f(p) =   .6931999
  (output omitted)
GMM estimator of BLP-model
Instruments: Chamberlain optimal          Number of obs           =  250
                                          Number of markets       =  25
                                          Number of Halton draws  =  200
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Mean utility						
cons	2.133752	.4518272	4.72	0.000	1.248187	3.019317
x1	2.210469	.3299087	6.70	0.000	1.56386	2.857078
p	-2.05896	.0450369	-45.72	0.000	-2.147231	-1.970689
x1						
SD	.9145327	.1890028	4.84	0.000	.544094	1.284971

The parameter estimates are now closer to the true values and have smaller SEs compared with those estimated using the suboptimal set.

5.2 Demographic data

This example extends the previous model to include a random coefficient on price `p`, with a standard deviation of 1 and two demographic variables `d1` and `d2` in the coefficient equation for `x1`. The marginal effects of `d1` and `d2` are both 1, and samples are drawn from independent normal distributions. To permit parameter identification, the mean and variance are allowed to differ across markets. `demodata.dta` contains 500 draws (per market) selected at random from the simulated population of individuals used to construct the product shares.

The model is fit with optimal instruments, and price elasticities are reported for market 1 by specifying `elast(p,1,product)`, where `product` is a string variable that contains product names to label the elasticity matrix.

```
. use blp_demo, clear
. generate w12=w1^2
. generate w22=w2^2
. generate w32=w3^2
. generate x12=x1^2
. generate x1w1=x1*w1
. generate x1w2=x1*w2
. generate x1w3=x1*w3
. bysort mkt: egen x1s=sum(x1)
. replace x1s=x1s-x1
(500 real changes made)
. blp s x1, stochastic(x1=d1 d2,p) endog(p=w1 w2 w3 w12 w22 w32 x12 x1w1 x1w2
> x1w3 x1s) markets(mkt) optinst(w1 w2 w3) demofile(demodata) initdemo(1,1)
> initsd(1,1) elast(p,1,product)
  (output omitted)
Iteration 0:   f(p) =  18.345114  (not concave)
Iteration 1:   f(p) =   11.51863
Iteration 2:   f(p) =  10.731222
  (output omitted)
Estimation iteration with optimal instruments: 1
Iteration 0:   f(p) =  330.01219  (not concave)
Iteration 1:   f(p) =  26.894109
  (output omitted)
```

```
GMM estimator of BLP-model
Instruments: Chamberlain optimal           Number of obs          =  500
                                           Number of markets      =  50
                                           Number of Halton draws =  500

                    Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]

Mean utility
        cons     1.562526   .4870916    3.21   0.001     .6078444    2.517208
          x1     2.144829   .1669901   12.84   0.000     1.817535    2.472124
           p    -1.938477   .1165748  -16.63   0.000     -2.16696   -1.709995

x1
          d1     .9794679   .0728124   13.45   0.000     .8367581    1.122178
          d2     .9406059   .0740145   12.71   0.000     .7955401    1.085672
          SD     .9448503    .204845    4.61   0.000     .5433616    1.346339

p
          SD     .9726345   .0751903   12.94   0.000     .8252641    1.120005
```

The estimated marginal effects of `d1` and `d2` are approximately 0.98 and 0.94.

Elasticities

To view the price elasticity matrix, type

```
. matrix list e(elast)
e(elast)[10,10]
                       1% rise i~p:
                          product1      product2      product3      product4
% change in:product1    -3.1977628     .65508213     .07305337     .00991889
            product2     1.1187173    -4.2095821      .0890736     .00896675
            product3      1.029381     .73495422    -4.7211944     .00551456
  (output omitted)
```

The (i, j) element in the above matrix represents the percentage change in the demand for product i caused by a 1% rise in the price of product j.

For a given j, the logit model restricts the cross-price elasticities to be the same for all $i \neq j$ (section 2.1). As illustrated above, this is not the case for the BLP model, which permits patterns of substitution that are more likely to be observed in the data.

5.3 Estimation time

For each evaluation of the GMM objective function, `blp` first inverts the system of demand equations by calling the inner-loop contraction mapping from (3) and (11). Because this process is numerically intensive, the convergence time of `blp` will increase significantly with every parameter in $\boldsymbol{\theta}_2$. The number of draws R may also have an effect on the computational time of the contraction mapping, despite the use of vectorization in (4).

To investigate these issues, the estimation times of the examples are recorded for $R \in (100, 500)$. For the first example with one heterogeneity parameter (section 5.1), the computational time using standard and optimal instruments is less than 1 second for $R = 100$. This increases to under 6 seconds when $R = 500$.

For the second model containing four heterogeneity parameters (section 5.2), the time increases to 203 seconds using optimal instruments, but it remains almost unchanged for $R = 500$. To investigate the impact of a larger inner-loop tolerance, the process is repeated for this model, setting `tolin(10E-10)`. The computation time falls to 138 seconds and again shows no significant increase when $R = 500$.

5.4 Applications with real data

Simulated data help illustrate `blp` but avoid issues that arise in empirical applications of the estimator. These issues include decisions on model specification, sourcing the market-level data, selection of appropriate instruments, and data preparation.

Of particular importance is a need to construct the product market shares. While prices and characteristics can be observed in the data, market shares must be computed from quantities sold and a prior estimate of the number of potential consumers. This will depend on the definition of the market (that is, location, time period) and the nature of the products being investigated.

BLP, for example, define a national annual market for automobile sales and measure the total size as the number of U.S. households. Nevo (2001), on the other hand, defines a quarterly city market for cereal brands and measures the potential size as one serving per capita per day.[16] The quantities sold are then converted into the same unit of measurement (for example, servings), and product market shares are computed as the ratio of sales to market size. In all cases, however, the size of the potential market must be set to ensure that the outside good has a nonzero share.

Finally, and as suggested by Nevo (2000b), it is recommended that practitioners test the sensitivity of their results to the market definition. If these are found to be sensitive, then alternative definitions should be considered.

6 Monte Carlo experiments

In this section, we use Monte Carlo simulation to investigate the bias, efficiency, and stability of `blp` using the standard ($\mathbf{z}^1_{jt}$, $\mathbf{z}^2_{jt}$) and optimal ($\mathbf{z}^*_{jt}$) instrument sets as defined in section 3.1. For each experiment, 1,000 datasets are generated from the process specified by Reynaert and Verboven (2014) and extended to include one demographic variable in the stochastic coefficient equations. We also consider the role of starting values in attaining global convergence (Knittel and Metaxoglou 2014) and the importance

16. This implies that consumers face daily purchase decisions between servings of alternative brands. In practice, they face choices between brands and pack sizes; however, it would be difficult to measure the market size when quantity is defined in this way, because packs contain multiple servings.

of accurate approximations to the share integrals in (4) in avoiding false local minimums (Skrainka and Judd 2011).

6.1 The data-generating process

Data are generated for $J = 10$ products and $T = 25$ markets. The variables consist of product characteristics $\mathbf{x}_{jt} = (1, x_{1jt})$, independent cost drivers $\mathbf{w}_{jt} = (w_{jt1}, w_{jt2}, w_{jt3})'$, an endogenous price p_{jt} described by (7), and the product market share s_{jt}. The characteristic x_{1jt} is drawn from the $U(1,2)$ distribution, the cost drivers $\mathbf{w}_{jt}$ from the $U(0,1)$ distributions, and the unobservables (ξ_{jt}, ζ_{jt}) from the bivariate normal distribution with unit variance and covariance 0.7.

For all experiments, the mean valuations are equal to $\boldsymbol{\beta} = (2, 2, -2)'$, and the cost parameters are set at $\boldsymbol{\gamma}_1 = (0.7, 0.7)'$ and $\boldsymbol{\gamma}_2 = (3, 3, 3)'$. One design also includes a demographic variable $\mathbf{D}_{it}$,[17] which is drawn from the $N(\mu_{Dt}, \sigma^2_{Dt})$ distribution, where μ_{Dt} and σ^2_{Dt} are sampled from the $N(0,1)$ and $\mid N(0,1) \mid$ distributions, respectively.[18]

Finally, the market shares are approximated by Monte Carlo integration using $R = 300000$ random draws of $\mathbf{D}_{it}$ and $\mathbf{v}_{it}$ for each market.

6.2 A single stochastic coefficient

No demographics

The first design sets $\mathbf{L} = \text{diag}(\sigma_0, \sigma_{x_1}, \sigma_p) = \text{diag}(0, 1, 0)$ and $\boldsymbol{\Pi} = (0, 0, 0)'$, and it draws starting values from the $U(0.1, 2)$ distribution. As such, all heterogeneity is associated with the random coefficient on x_{1jt} with a standard deviation of 1. For each instrument set, table 1 provides the estimated bias, the root mean squared error (RMSE), and the mean of the asymptotic SEs for $R \in (50, 100, 200)$ Halton draws.

The results reflect the findings of Reynaert and Verboven (2014, tab. 1), and they report moderate bias using standard instruments $\mathbf{z}^1_{jt}$, a small reduction from $\mathbf{z}^2_{jt}$ (which includes BLP-type instruments), and a large improvement using Chamberlain (1987) optimal instruments $\mathbf{z}^*_{jt}$. Note that as the approximation of the share integrals in (4) improves, the bias of the GMM estimator using optimal instruments reduces. For example, the estimator of σ_{x_1} (denoted `sdx1`) reports a bias of 7% for $R = 50$, which falls to 4% when R is increased to 200. This is not the case when using standard instruments, where the bias tends to persist throughout.

The simulations also suggest that suboptimal instruments lead to biased estimates of the SEs for $\text{SE}(\widehat{\sigma}_{x_1})$, which increases as the approximation of the share integrals in (4) improves.[19] This is not the case when using optimal instruments, although the estimated SEs exhibit modest downward bias when compared with the RMSE.

17. The subscript t in $\mathbf{D}_{it}$ indicates that the demographic draws are market specific.
18. As aforementioned, variation in the distribution of demographics is required to identify $\boldsymbol{\Pi}$.
19. This finding was first noted by Skrainka and Judd (2011) and then confirmed in the simulation study by Reynaert and Verboven (2014, tab. 8).

Using optimal instruments also offers considerable improvements in efficiency. This is especially the case for the estimator of the heterogeneity parameter σ_{x_1}, where for $R = 200$, the RMSE is 46% lower than the value using $\mathbf{z}^1_{jt}$. Figure 1 visualizes these efficiency gains by comparing histograms of the parameter estimates of σ_{x_1} for each set of instruments. It also displays a large spike in the distribution of the estimator around 0 using suboptimal instruments $\mathbf{z}^1_{jt}$ and $\mathbf{z}^2_{jt}$, which is not the case using optimal instruments $\mathbf{z}^*_{jt}$. This confirms the findings of Reynaert and Verboven (2014, fig. 1) that optimal instruments improve both the efficiency and stability of the estimator $\widehat{\sigma}_{x_1}$.

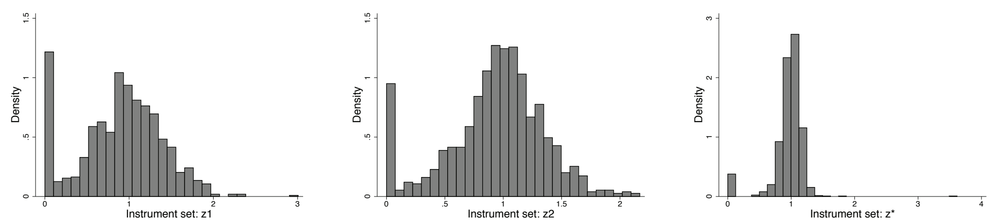

Figure 1. Histograms of $\widehat{\sigma}_1$

Finally, to investigate the properties of the objective function, the experiment was repeated with different starting values for the heterogeneity parameter. The results using the same datasets were virtually identical to those in table 1, suggesting no local extrema.

Table 1. Bias and efficiency—random coefficient on `x1`

draws and parameters	z1 bias	z1 st.err	z1 rmse	z2 bias	z2 st.err	z2 rmse	opt bias	opt st.err	opt rmse
50									
con	−0.122	0.765	0.754	−0.102	0.663	0.678	−0.085	0.444	0.511
x1	0.048	0.587	0.585	0.030	0.489	0.515	0.066	0.275	0.356
p	0.012	0.055	0.053	0.012	0.050	0.050	0.005	0.043	0.044
sdx1	−0.106	0.690	0.505	−0.068	0.457	0.423	−0.070	0.129	0.289
100									
con	−0.113	0.778	0.757	−0.090	0.673	0.679	−0.049	0.445	0.502
x1	0.043	0.597	0.587	0.021	0.499	0.516	0.036	0.276	0.345
p	0.012	0.055	0.054	0.012	0.051	0.050	0.004	0.043	0.044
sdx1	−0.106	0.993	0.501	−0.064	0.596	0.416	−0.045	0.129	0.260
200									
con	−0.110	0.791	0.758	−0.084	0.678	0.682	−0.040	0.446	0.509
x1	0.039	0.609	0.587	0.016	0.503	0.518	0.029	0.276	0.359
p	0.011	0.056	0.054	0.011	0.051	0.050	0.003	0.044	0.044
sdx1	−0.105	1.487	0.498	−0.064	0.808	0.416	−0.042	0.136	0.267

Including demographics

The second design sets $\mathbf{L} = \mathrm{diag}(0, 1, 0)$ and $\mathbf{\Pi} = (\pi_{0D_1}, \pi_{x_1 D_1}, \pi_{pD_1})' = (0, 1, 0)'$ to permit the stochastic coefficient on x_{1jt} to be a linear function of $\mathbf{D}_{it}$, with a marginal effect of 1. Table 2 displays the bias, RMSE, and the mean of the asymptotic SEs for $R \in (50, 100, 200)$ Halton draws.

Table 2. Bias and efficiency—random coefficient on `x1` with demographics

draws and parameters	z1 bias	z1 st.err	z1 rmse	z2 bias	z2 st.err	z2 rmse	opt bias	opt st.err	opt rmse
50									
con	-0.186	0.851	0.760	-0.183	0.743	0.724	-0.132	0.455	0.530
x1	0.153	0.726	0.665	0.142	0.599	0.608	0.124	0.279	0.353
p	0.014	0.064	0.058	0.014	0.057	0.055	0.006	0.045	0.048
d1x1	-0.077	0.369	0.347	-0.070	0.282	0.293	-0.048	0.070	0.105
sdx1	-0.185	1.277	0.562	-0.145	0.834	0.507	-0.102	0.156	0.346
100									
con	-0.138	0.854	0.750	-0.128	0.740	0.706	-0.083	0.450	0.504
x1	0.097	0.731	0.656	0.082	0.600	0.585	0.073	0.278	0.324
p	0.012	0.063	0.057	0.012	0.056	0.054	0.004	0.044	0.047
d1x1	-0.051	0.369	0.341	-0.042	0.282	0.277	-0.026	0.071	0.095
sdx1	-0.167	1.469	0.556	-0.129	0.938	0.493	-0.087	0.162	0.315
200									
con	-0.124	0.844	0.761	-0.107	0.732	0.711	-0.047	0.447	0.496
x1	0.081	0.727	0.673	0.058	0.594	0.592	0.042	0.277	0.336
p	0.011	0.063	0.057	0.011	0.056	0.054	0.002	0.044	0.046
d1x1	-0.042	0.368	0.351	-0.028	0.280	0.280	-0.016	0.071	0.099
sdx1	-0.175	1.868	0.559	-0.130	1.169	0.487	-0.061	0.153	0.291

The results reflect the findings in table 1, and they report reductions in bias and gains in efficiency when the GMM estimator uses optimal instruments, $\mathbf{z}^*_{jt}$, instead of standard sets $\mathbf{z}^1_{jt}$ and $\mathbf{z}^2_{jt}$. For the estimator of $\boldsymbol{\pi}_{x_1 D_1}$ (denoted `d1x1`) with $R = 200$, the bias falls from 4% using $\mathbf{z}^1_{jt}$ to around 1.5% using $\mathbf{z}^*_{jt}$ and has a reduction of over 70% in the RMSE.

6.3 Multiple stochastic coefficients

The final experiment considers the performance of the GMM estimator when the coefficients on x_1 and p are allowed to be random. The standard deviations of these parameters are set to $\mathbf{L} = (0, 1, 1)$, the demographic variables are excluded from the model, and the number of markets is increased to $T = 50$. Table 3 provides the bias, RMSE, and the mean of the asymptotic SEs for $R \in (100, 200)$ Halton draws.

Table 3. Bias and efficiency—random coefficients on `x1` and `p`

draws and parameters	z1 bias	z1 st.err	z1 rmse	z2 bias	z2 st.err	z2 rmse	opt bias	opt st.err	opt rmse
100									
con	-0.437	1.537	1.375	-0.439	1.217	1.141	-0.199	0.608	0.666
x1	-0.079	0.286	0.309	-0.095	0.263	0.308	-0.014	0.174	0.189
p	0.151	0.496	0.458	0.159	0.395	0.390	0.066	0.183	0.209
sdx1	-0.185	2.566	0.779	-0.124	1.666	0.696	-0.187	0.284	0.482
sdp	-0.093	0.317	0.292	-0.097	0.258	0.250	-0.039	0.114	0.133
200									
con	-0.402	1.527	1.346	-0.385	1.208	1.106	-0.024	0.620	0.643
x1	-0.071	0.279	0.293	-0.086	0.258	0.302	0.002	0.176	0.186
p	0.138	0.492	0.445	0.140	0.392	0.375	0.007	0.190	0.201
sdx1	-0.187	4.034	0.767	-0.121	2.659	0.693	-0.152	0.343	0.484
sdp	-0.085	0.313	0.283	-0.086	0.252	0.239	-0.002	0.118	0.127

The bias is much higher relative to the results in table 1, and the estimator is less efficient. Optimal instruments, $\mathbf{z}^*_{jt}$, continue to provide considerable gains in efficiency over suboptimal sets, although the reduction in bias is minimal for the heterogeneity parameter $\widehat{\sigma}_{x_1}$. This is not the case for the estimator of σ_p (denoted `sdp`), where the estimated bias falls from 8.5% using $\mathbf{z}^1_{jt}$ to 0.2% using $\mathbf{z}^*_{jt}$, with a 55% reduction in the RMSE for $R = 200$.

As before, an increase in the number of draws leads to improvements in the performance of the estimator using optimal instruments, $\mathbf{z}^*_{jt}$, which is not the case for the estimator using both suboptimal sets, $\mathbf{z}^1_{jt}$ and $\mathbf{z}^2_{jt}$.

7 Conclusion

The aggregate random-coefficients logit demand model by Berry, Levinsohn, and Pakes (1995) has become increasingly popular since the publication of Nevo (2000b). In this article, I describe a user-written command, `blp`, that estimates the parameters of this model, using standard or Chamberlain (1987) optimal instruments. Monte Carlo experiments confirm the findings by Reynaert and Verboven (2014) and report reductions in bias and improvements in the stability and efficiency of the estimator using optimal instruments relative to the series approximations typically found in the literature.

8 Acknowledgment

I thank an anonymous referee for bringing my attention to the recent work of Reynaert and Verboven (2014). This referee also made several other valuable suggestions to expand the code and to highlight potential limitations of the estimator.

9 References

Armstrong, T. B. 2014. Large market asymptotics for differentiated product demand estimators with economic models of supply. Working paper, Yale University. http://armstrong.commons.yale.edu/files/largemarket_r2.pdf.

Berry, S., J. Levinsohn, and A. Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* 63: 841–890.

———. 1999. Voluntary export restraints on automobiles: Evaluating a trade policy. *American Economic Review* 89: 400–430.

Chamberlain, G. 1987. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics* 34: 305–334.

Deaton, A. S., and J. Muellbauer. 1980a. *Economics and Consumer Behavior*. Cambridge: Cambridge University Press.

———. 1980b. An almost ideal demand system. *American Economic Review* 70: 312–326.

Drukker, D. M., and R. Gates. 2006. Generating Halton sequences using Mata. *Stata Journal* 6: 214–228.

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

Hall, A. R. 2005. *Generalized Method of Moments*. Oxford: Oxford University Press.

Hole, A. R. 2007. Fitting mixed logit models by using maximum simulated likelihood. *Stata Journal* 7: 388–401.

Knittel, C. R., and K. Metaxoglou. 2014. Estimation of random-coefficient demand models: Two empiricists' perspective. *Review of Economics and Statistics* 96: 34–59.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, ed. P. Zerembka, 105–142. New York: Academic Press.

Nevo, A. 2000a. Appendix to "A practitioner's guide to estimation of random-coefficients logit models of demand". http://faculty.wcas.northwestern.edu/~ane686/supplements/Ras_guide_appendix.pdf.

———. 2000b. A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics and Management Strategy* 9: 513–548.

———. 2001. Measuring market power in the ready-to-eat cereal industry. *Econometrica* 69: 307–342.

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.

Skrainka, B. S., and K. L. Judd. 2011. High performance quadrature rules: How numerical integration affects a popular model of product differentiation. Working Papers CWP03/11, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Theil, H. 1965. The information approach to demand analysis. *Econometrica* 33: 67–87.

Train, K. 2000. Halton sequences for mixed logit. Economics Department, University of California, Berkeley, Working Paper E00-278.
http://repositories.cdlib.org/iber/econ/E00-278/.

About the author

David Vincent is an applied econometrician for Deloitte Economic Consulting in London. He has used Stata for more than 10 years and enjoys developing new commands for estimation and specification testing. His current interests include consumer demand models and Bayesian econometrics.