Notes on BLP

Adam N. Smith
July 31, 2021

This note reviews the canonical random coefficients logit or "BLP" model à la Berry et al. (1995). We outline details of the model, the contraction mapping, and both classical and Bayesian approaches to estimation.

1 Model

Let u_{ijt} denote the indirect utility of consumer i for product $j = 0, 1, \ldots, J$ at time $t = 1, \ldots, T$, which is specified as:

$$u_{ijt} = x'_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt} \tag{1}$$

where $u_{i0t} = \varepsilon_{i0t}$ is a normalization for the outside good $j = 0$. Here β_i denotes the consumer's preference for characteristics $x_{jt} \in \mathbb{R}^K$ (including price), ξ_{jt} is a market-level error term representing unobserved shocks to the demand for good j, and ε_{ijt} is an error term with distribution $F_\varepsilon(\cdot)$ representing all other unobservables associated with consumer i.

Assume that individual taste parameters belong to a location-scale family:

$$\beta_i = \bar{\beta} + \eta_i, \qquad \eta_i \sim F_\eta(\cdot|\Sigma) \tag{2}$$

where $\bar{\beta}$ represents average preferences in the population (the "linear" parameters) and Σ parameterizes the variances of the distribution of tastes (the "nonlinear" parameters). We can then rewrite utility as:

$$\begin{aligned} u_{ijt} &= x'_{jt}(\bar{\beta} + \eta_i) + \xi_{jt} + \varepsilon_{ijt} \\ &= x'_{jt}\bar{\beta} + \xi_{jt} + x'_{jt}\eta_i + \varepsilon_{ijt} \\ &= \delta_{jt}(\bar{\beta}, \xi_{jt}) + x'_{jt}\eta_i + \varepsilon_{ijt} \end{aligned}$$

where $\delta_{jt} \equiv \delta_{jt}(\bar{\beta}, \xi_{jt})$ captures the "mean utility" for j at time t in the population. Under the assumption of utility maximization and iid TIEV errors, the market share for good j at time t is given by:

$$\sigma_{jt} = \int 1\left(u_{ijt} \geq u_{ikt}\ \forall k \neq j\right) dF_\eta(\eta_i|\Sigma) dF_\varepsilon(\varepsilon_{it}) \tag{3}$$
$$= \int \frac{\exp\left(\delta_{jt}(\bar{\beta}, \xi_{jt}) + x'_{jt}\eta_i\right)}{1 + \sum_{k=1}^{J} \exp\left(\delta_{kt}(\bar{\beta}, \xi_{kt}) + x'_{kt}\eta_i\right)} dF_\eta(\eta_i|\Sigma).$$

2 Share Inversion and Contraction Mapping

Many estimation approaches (both classical and Bayesian) require an inversion of the system in (3). That is, we want to exchange an expression of the shares as a function of mean utility parameters with an expression of the mean utility parameters as a function of the shares:

$$s_{jt} = \sigma_{jt}(\delta_{1t}, \ldots, \delta_{Jt}, \Sigma) \implies \delta_{jt} = \sigma_{jt}^{-1}(s_{1t}, \ldots, s_{Jt}, \Sigma). \tag{4}$$

The inversion in (4) does not have an analytic solution because of the integral over nonlinear parameters.

BLP solve for δ_{jt} via a contraction mapping. Suppose we know the value of nonlinear parameter vector Σ and have an initial guess of the mean utilities $\{\delta_{jt}^1\}$. Then for $h = 1, 2, \ldots$ do the following:

$$\delta_{jt}^{h+1} = \delta_{jt}^{h} + \log(s_{jt}) - \log\left(\int \frac{\exp\left(\delta_{jt}^{h} + x'_{jt}\eta_i\right)}{1 + \sum_{k=1}^{J} \exp\left(\delta_{kt}^{h} + x'_{kt}\eta_i\right)} dF_\eta(\eta_i|\Sigma)\right) \tag{5}$$

and stop only when $||\delta_{jt}^{h+1} - \delta_{jt}^{h}|| < \epsilon$. Note that because we are conditioning on Σ, the integral on the right-hand-side of (5) can be computed using Monte Carlo integration. That is, we can draw a set of $\eta_i \sim F_\eta(\cdot|\Sigma)$ prior to the contraction mapping and then, for each iteration, plug them into the share expression on the right-hand-side of (5) and take an average. Also note that some care must be taken when choosing the tolerance ϵ for the inside loop. If the tolerance is too small, then the contraction mapping algorithm will not converge. If the tolerance is too large, then numerical errors in the inside loop can propagate to the outer loop and affect estimation and inference. For guidance, see the discussion in Section 3 of Conlon and Gortmaker (2020).

3 Classical Estimation

3.1 GMM Obejective Function

To address the potential endogeneity of product characteristics, assume we have instruments z_{jt} that are excluded from the demand equation. BLP propose a GMM estimator based on the moment condition $\mathbb{E}[\xi_{jt}|z_{jt}, x_{jt}] = 0$. The empirical analog can be written as:

$$\mathbb{E}[\mathcal{G}(\Sigma)] = 0 \text{ where } \mathcal{G}(\Sigma) = \frac{1}{TJ}\sum_{t=1}^{T}\sum_{j=1}^{J}\xi_{jt}(\Sigma)z_{jt}. \tag{6}$$

Then for a given weighting matrix W, the GMM estimator is defined as the solution to the problem:

$$\min_{\Sigma} Q(\Sigma) = \mathcal{G}(\Sigma)' W \mathcal{G}(\Sigma). \tag{7}$$

However, given a candidate value of Σ, the evaluation of sample moments $\mathcal{G}(\Sigma)$ requires first finding the implied values of the demand shocks ξ_{jt}. This motivates BLP's nested fixed point algorithm.

3.2 Nested Fixed Point Approach

- The **outer loop** searches over Σ (via derivative-based methods like quasi-Newton or nonderivative, simplex-based methods like Nelder-Mead).[1]
- The **inner loop** constructs the GMM objective function $Q(\Sigma)$ to be minimized given a candidate Σ.

  (a) Use the BLP contraction mapping to solve for mean utilities $\hat{\delta}_{jt}$.

  (b) Use 2SLS to estimate linear coefficients $\bar{\beta}$.

  $$\hat{\delta}_{jt} = x_{jt}'\bar{\beta} + \xi_{jt}$$

  (c) Obtain residuals $\hat{\xi}_{jt} = \hat{\delta}_{jt} - \hat{\bar{\beta}}$ and construct the sample moments:

  $$\hat{\mathcal{G}}(\Sigma) = \frac{1}{TJ}\sum_{t=1}^{T}\sum_{j=1}^{J}\hat{\xi}_{jt}(\Sigma)z_{jt}.$$

[1] Derivative-based methods have been found to generally outperform simplex-based methods (Dubé et al., 2012; Knittel and Metaxoglou, 2014; Conlon and Gortmaker, 2020).

(d) Evaluate the GMM objective function:

$$Q(\Sigma) = \hat{\mathcal{G}}(\Sigma)' W \hat{\mathcal{G}}(\Sigma).$$

3.3 MPEC Approach

Dubé et al. (2012) study the numerical performance of the nested fixed point algorithm outlined above and show how error inside the BLP contraction mapping propagates into the GMM optimization problem in the outer loop, which can lead to biased parameter estimates. The authors propose an alternative based on mathematical programming with equilibrium constraints (MPEC).

$$\begin{aligned} &\min_{\Sigma} Q(\Sigma) = \mathcal{G}(\Sigma)' W \mathcal{G}(\Sigma) \\ &\text{s.t. } \sigma(\delta, \Sigma) = s \end{aligned}$$

An MPEC approach can be faster and avoids the numerical issues associated with the nested inner loop.

4 Bayesian Estimation

Jiang et al. (2009) propose a Bayesian approach to estimation and inference for the random coefficients logit model. We outline details of the likelihood, prior, and posterior sampling strategies below.

Model Likelihood The model definition and underlying assumptions mirror those made in BLP. Note that at any point in time t, market shares $(s_{1t}, \ldots, s_{Jt})$ are stochastic only because of the demand shocks $(\xi_{1t}, \ldots, \xi_{Jt})$. Therefore, one additional parametric assumption is made about demand shocks in order to define a model likelihood:

$$\xi_{jt} \sim N(0, \tau^2). \tag{8}$$

Then using the definition of the mean utilities $\delta_{jt} = x_{jt}'\bar{\beta} + \xi_{jt}$, we have:

$$\delta_{jt} \sim N(x_{jt}'\bar{\beta}, \tau^2). \tag{9}$$

Now the density of observed shares is determined by the mean utility parameters through the familiar share inversion:

$$s_{jt} = \sigma_{jt}(\delta_{1t}, \ldots, \delta_{Jt}, \Sigma) \implies \delta_{jt} = \sigma_{jt}^{-1}(s_{1t}, \ldots, s_{Jt}, \Sigma). \tag{10}$$

We can then define the likelihood function using the change-of-variables theorem:

$$\begin{aligned} p(s|\bar{\beta}, \Sigma, \tau^2) &= \prod_{t=1}^{T} \prod_{j=1}^{J} p(s_{jt}|\bar{\beta}, \Sigma, \tau^2) \\ &= \prod_{t=1}^{T} \prod_{j=1}^{J} \phi\left(\sigma_{jt}^{-1}(s_{1t}, \ldots, s_{Jt}, \Sigma)/\tau\right) |\mathbf{J}_{\delta_t \to s_t}| \\ &= \prod_{t=1}^{T} |\mathbf{J}_{s_t \to \delta_t}|^{-1} \prod_{j=1}^{J} \phi\left(\delta_{jt}/\tau\right) \end{aligned} \tag{11}$$

where $\phi(\cdot)$ is the standard normal pdf and $\mathbf{J}_{s_t \to \delta_t}$ is the Jacobian from the mapping of shares to mean utilities. Specifically, the Jacobian is defined as:

$$\mathbf{J}_{s_t \to \delta_t} = \begin{bmatrix} \frac{\partial \sigma_{1t}(\delta_t, \Sigma)}{\partial \delta_{1t}} & \frac{\partial \sigma_{1t}(\delta_t, \Sigma)}{\partial \delta_{2t}} & \cdots & \frac{\partial \sigma_{1t}(\delta_t, \Sigma)}{\partial \delta_{Kt}} \\ \frac{\partial \sigma_{2t}(\delta_t, \Sigma)}{\partial \delta_{1t}} & \frac{\partial \sigma_{2t}(\delta_t, \Sigma)}{\partial \delta_{2t}} & \cdots & \frac{\partial \sigma_{2t}(\delta_t, \Sigma)}{\partial \delta_{Kt}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \sigma_{Kt}(\delta_t, \Sigma)}{\partial \delta_{1t}} & \frac{\partial \sigma_{Kt}(\delta_t, \Sigma)}{\partial \delta_{2t}} & \cdots & \frac{\partial \sigma_{Kt}(\delta_t, \Sigma)}{\partial \delta_{Kt}} \end{bmatrix} \tag{12}$$

where

$$\frac{\partial \sigma_{jt}(\delta_t, \Sigma)}{\partial \delta_{kt}} = \begin{cases} \int \sigma_{ijt}(\delta_t, \Sigma)\sigma_{ikt}(\delta_t, \Sigma) dF_\eta(\eta_i|\Sigma) & \text{if } k \neq j \\ \int \sigma_{ijt}(\delta_t, \Sigma)\Big[1 - \sigma_{ijt}(\delta_t, \Sigma)\Big] dF_\eta(\eta_i|\Sigma) & \text{if } k = j. \end{cases} \tag{13}$$

Note that the Jacobian depends on both linear and nonlinear parameters. However, given Σ and observed shares, the mean utilities can be uniquely determined through the share inversion (i.e., contraction mapping). Therefore, the process of computing the Jacobian follows from the inner loop of BLP's nested fixed point algorithm. Specifically, we can do the following.

i. Given a candidate Σ^*, first generate Monte Carlo draws of η_i^* for $i = 1, \ldots, H$.

ii. Solve for δ_t using BLP contraction mapping (where shares are evaluating using

Monte Carlo draws of η_i^*). Also save out individual-level choice probabilities σ_{ijt} for $i = 1, \ldots, H$ evaluated at "final" δ_t^*.

iii. Compute elements of Jacobian as

$$\frac{\partial \sigma_{jt}(\delta_t, \Sigma)}{\partial \delta_{kt}} = \begin{cases} \frac{1}{H} \sum_{i=1}^{H} \sigma_{ijt}(\delta_t^*, \Sigma^*) \sigma_{ikt}(\delta_t^*, \Sigma^*) & \text{if } k \neq j \\ \frac{1}{H} \sum_{i=1}^{H} \sigma_{ijt}(\delta_t^*, \Sigma^*) \left[1 - \sigma_{ijt}(\delta_t^*, \Sigma^*)\right] & \text{if } k = j. \end{cases} \tag{14}$$

Prior In addition to a likelihood function, a Bayesian analysis also requires the specification of priors for all model parameters: $\bar{\beta}$, Σ, and τ^2. Jiang et al. (2009) use conditionally conjugate priors for $\bar{\beta}$ and τ^2:

$$\bar{\beta} \sim N(\bar{\beta}_0, V_{\bar{\beta}}) \tag{15}$$

$$\tau^2 \sim \frac{\nu_0 s_0^2}{\chi^2_{\nu_0}}. \tag{16}$$

To specify a prior for Σ, they first parameterize Σ in terms of the $K(K+1)/2$ elements of its Cholesky root:

$$\Sigma = U'U \text{ where } U = \begin{bmatrix} \exp(r_{11}) & r_{12} & \cdots & r_{1K} \\ 0 & \exp(r_{22}) & \cdots & r_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \exp(r_{KK}) \end{bmatrix}. \tag{17}$$

Note that the diagonals are enforced to be positive to ensure Σ is positive definite. Then we can place normal priors on the elements in the upper diagonal of U:

$$r_{kk} \sim N(0, s^2_{\text{diag}}) \tag{18}$$

$$r_{k\ell} \sim N(0, s^2_{\text{off}}) \tag{19}$$

where separate variances are used for the diagonal and off-diagonal terms to account for the fact that the diagonals are on a log scale.

Posterior Given the model likelihood in (11) and priors in (15), (16), and (18), the complete posterior is defined as:

$$
\begin{aligned}
p(\bar\beta,\Sigma,\tau^2|s) &\propto p(s|\bar\beta,\Sigma,\tau^2)p(\bar\beta,\Sigma,\tau^2)\\
&= \left[\prod_t \prod_j p(s_{jt}|\bar\beta,\Sigma,\tau^2)\right] p(\bar\beta)p(\Sigma)p(\tau^2)\\
&\propto \left[\prod_t |\mathbf{J}_{s_t\to\delta_t}|^{-1} \prod_j \phi\left(\delta_{jt}/\tau\right)\right]\\
&\qquad \times |V_{\bar\beta}|^{-1/2} \exp\left(-\frac{1}{2}(\bar\beta-\bar\beta_0)' V_{\bar\beta}^{-1}(\bar\beta-\bar\beta_0)\right)\\
&\qquad \times \prod_{k=1}^{K} \exp\left(-\frac{r_{kk}^2}{2s_{\text{diag}}^2}\right) \prod_{k=1}^{K-1} \prod_{\ell=k+1}^{K} \exp\left(-\frac{r_{k\ell}^2}{2s_{\text{off}}^2}\right)\\
&\qquad \times (\tau^2)^{-(\nu_0/2+1)} \exp\left(-\frac{\nu_0 s_0^2}{2\tau^2}\right).
\end{aligned}
$$

Computaiton Sampling from the posterior above can be achieved with a Metropolis-within-Gibbs sampler which iteratively draws from the following full conditionals.

$$
\Sigma|\bar\beta,\tau^2,\text{data} \tag{20}
$$

$$
\bar\beta,\tau^2|\Sigma,\text{data} \tag{21}
$$

Specifically, draws from each full conditional can be made as follows.

1. (Nonlinear Parameters) Conjugate priors for nonlinear parameters Σ do not exist and so we must draw from the first full conditional using a Metropolis-Hastings step.

   (a) Propose $r_{jk}^* \sim N(r, s_{\text{step}}^2)$ and set $\Sigma^* = U^{*\prime}U^*$ where the upper diagonals are filled with r_{jk}^*.

   (b) Use the BLP contraction mapping to solve for δ_{jt}^*. Then, as discussed above, use the stored draws of choice probabilities $\sigma_{ijt}(\delta_t^*,\Sigma^*)$ to construct the Jacobian which is is used to evaluate the likelihood.

   (c) Accept (Σ^*,δ_t^*) with probability:

$$
\alpha = \min\left\{1, \frac{p(s|\bar\beta,\Sigma^*,\tau^2)p(\Sigma^*)}{p(s|\bar\beta,\Sigma,\tau^2)p(\Sigma)}\right\}.
$$

2. (Linear Parameters) Draw the pair $(\bar{\beta}, \tau^2)$ from the posterior of a univariate Bayesian regression model (or Bayesian IV regression model):

$$\delta_{jt} = x'_{jt}\bar{\beta} + \xi_{jt}, \qquad \xi_{jt} \sim N(0, \tau^2) \tag{22}$$

with conjugate priors given in (15) and (16).

References

Berry, S. T., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Conlon, C. and Gortmaker, J. (2020). Best practices for differentiated products demand estimation with PyBLP. *The RAND Journal of Economics*, 51(4):1108–1161.

Dubé, J. P., Fox, J. T., and Su, C. (2012). Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica*, 80(5):2231–2267.

Jiang, R., Manchanda, P., and Rossi, P. E. (2009). Bayesian analysis of random coefficient logit models using aggregate data. *Journal of Econometrics*, 149(2):136–148.

Knittel, C. R. and Metaxoglou, K. (2014). Estimation of random-coefficient demand models: Two empiricists' perspective. *The Review of Economics and Statistics*, 96(1):34–59.