1 Setup

This document serves as a companion text to BLP (1995) by focusing only on implementation matters and making those details more explicit. Our notation follows that of BLP (1995) until section 2.3, at which point we depart from their notation for clarity.

Throughout the creation of this library, we made a number of choices over implementation details when the correct option was ambiguous. In these cases we consider both the details provided in BLP (1995) and the methods found in BLP (1999) code.[1] See this Confluence page for details about these choices. This document will make note of instances where we purposefully depart from details stated in BLP (1995).

1.1 Raw Inputs

To begin, suppose we have n cars and n_m car models, each with K demand-side attributes x and some supply-side attributes w. In the BLP (1995) data set, $n = 2217$, $n_m = 999$,[2] and $K = 5$ (constant, horsepower/weight, air conditioning standard, miles/dollar, and space). The supply-side attributes are a constant, log of horsepower/weight, air conditioning, log of miles/gallon, log of space, and a time trend. There are a total of 20 independent markets in the data, each of which corresponds to a year between 1971 and 1990. We index markets by t, cars within each market by j (so that a market-car will have index tj), car models by m, and individuals (unobservables) by i.

Firm/Model data: Using the BLP (1999) replication code and data, we are able to extract the following variables for each of the 20 markets from 1971 to 1990: car id, model id, firm id, year, market share, quantity, price, indicators for production location, horsepower per weight, whether air conditioning comes standard, miles per dollar, size of car, miles per gallon, and the logit dependent variable $(\ln(s_j) - \ln(s_0))$.[3] All other firm/model variables used are transformations of these base variables.

Unobservables: For each individual in the model, we draw a vector of $v_i = (v_{iy}, v_{i1} \ldots, v_{iK})$ unobservables from a standard multivariate normal distribution with identity variance-covariance matrix, where we set $y_i = e^{m_t + \hat{\sigma}_y v_{iy}}$. Note that the unobservables are fixed over the time period of the panel (see BLP (1995) p. 868). Define $n_s = 200$ to be the number of unobservables used in simulation.[4] Importance sampling is done directly in BLP (1999) code, which gives a corresponding importance sampling weight w_i^u for each unobservable i.[5] We use the weight when taking integrals across unobservables. The BLP (1999) code also provides the mean of income for each year m_t and standard deviation of income σ_y across all years in the sample.

[1] Accessed on July 16, 2014 via Archive.org's cached version: https://web.archive.org/web/20050430123542/http://www-personal.umich.edu/~jamesl/verstuff/instructions.html

[2] BLP (1995) mention that they use 997 models (see p. 869), but BLP (1999) code returns 999 models instead.

[3] We redefine a car to be from Japan if it is not produced in Europe and not produced domestically in the US. BLP (1999) makes additional restrictions, but our definition matches BLP (1995)'s summary statistics.

[4] Note that this number matches the discussion in BLP's NBER Working Paper No. 4264, footnote 17.

[5] Note that importance sampling requires computation of market shares and thus relies on a set of parameter inputs. By default, BLP (1999) code uses a set of parameter guesses, which it also uses as the starting values for estimation. Since we ultimately decided to use BLP (1995) published parameters as our starting values, we have altered BLP (1999) code so that it uses the published parameters for importance sampling instead. Thus, our unobservable draws differ slightly from the default ones from BLP (1999) code. For our bootstrap procedure, we draw 10 sets of 10 unobservables for each year following the same procedure. Then, for each block bootstrap at the market level, we use the first set of unobservables for the first time a market year is selected, the second set of unobservables for the second time a market year is selected, etc. This gives a sample of 200 unobservables that are fixed across the bootstrap panel. We repeat this for each bootstrap iteration.

1.2 Utility

Let utility be defined by (see BLP (1995) equation 6.14a)

$$u_{itj} = \alpha \ln \left(e^{m_t + \hat{\sigma}_y v_{iy}} - p_{tj} \right) + x_{tj}\overline{\beta} + \xi_{tj} + \sum_k \sigma_k x_{tjk} v_{ik} + \varepsilon_{itj}.$$

Note that income is assumed to be log-normal distributed, with parameters m_i (mean) and $\hat{\sigma}_y$ (standard deviation) estimated from the March Current Population Survey (CPS) for each year of the panel. We use income parameters extracted from BLP (1999) code. We can separate the utility into a product-specific component that does not vary with consumer characteristics

$$\delta_{tj} = \delta \left(x_{tj}, p_{tj}, \xi_{tj}; \theta \right) = x_{tj}\overline{\beta} + \xi_{tj}$$

and the component that contains interactions between product specific attributes and consumer characteristics

$$\mu_{itj} = \mu \left(x_{tj}, p_{tj}, v_i; \theta \right) = \alpha \ln \left(e^{m_t + \hat{\sigma}_y v_{iy}} - p_{tj} \right) + \sum_k \sigma_k x_{tjk} v_{ik}.$$

Thus, we can rewrite the utility as

$$u_{itj} = \delta \left(x_{tj}, p_{tj}, \xi_{tj}; \theta \right) + \mu \left(x_{tj}, p_{tj}, v_i; \theta \right) + \varepsilon_{itj}.$$

In practice, taking the natural log in the income term can result in imaginary numbers. Following BLP (1999) code, we can approximate it with

$$\begin{aligned} \alpha \ln \left(e^{m_t + \hat{\sigma}_y v_{iy}} - p_{tj} \right) &= \alpha \ln \left(e^{m_t + \hat{\sigma}_y v_{iy}} \right) + \alpha \ln \left(1 - \frac{p_{jt}}{e^{m_t + \hat{\sigma}_y v_{iy}}} \right) \\ &\approx \alpha \ln \left(e^{m_t + \hat{\sigma}_y v_{iy}} \right) - \frac{\alpha p_{jt}}{e^{m_t + \hat{\sigma}_y v_{iy}}}. \end{aligned}$$

Since utility is with respect to the outside good, we can remove the constant $\alpha \ln \left(e^{m_t + \hat{\sigma}_y v_{iy}} \right)$ from the expression,[6] leaving us with just the $-\frac{\alpha p_{jt}}{e^{m_t + \hat{\sigma}_y v_{iy}}}$ term.[7] Thus the utility computed is

$$u_{itj} = -\frac{\alpha p_{tj}}{e^{m_t + \hat{\sigma}_y v_{iy}}} + x_{tj}\overline{\beta} + \xi_{tj} + \sum_k \sigma_k x_{tjk} v_{ik} + \varepsilon_{itj}.$$

1.3 Instruments

Suppose we have some attribute x_j. Then we define instruments to be the attribute itself, the sum of the attribute for all other cars from the same firm and year ("own" instruments), and the sum of the attribute for all cars from rival firms in the same year ("all" instruments). In mathematical notation, they are (see also

[6] Note that by (BLP (1995) equation 6.14b), utility of the outside good is

$$u_{i0t} = \alpha \ln \left(e^{m_t + \hat{\sigma}_y v_{iy}} \right) + \xi_{0t} + \sigma_0 v_{i0} + \varepsilon_{i0t}.$$

Removing the $\alpha \ln \left(e^{m_t + \hat{\sigma}_y v_{iy}} \right)$ term from every utility term has the effect of normalizing u_{i0t} to 0.

[7] See BLP (1999) code, STRUC-INST-STEVE2.ARC, procedure DEFDATA for confirmation.

BLP (1995), section 5.1):

$$x_j, \quad \sum_{r \neq j, r \in \{\text{firm, year}\}} x_r, \quad \sum_{r \notin \{\text{firm}\}, r \in \{\text{year}\}} x_r.$$

The BLPDATA object takes care of these calculations. We do this for each of the K demand attributes to generate the demand-side instruments, and each of the supply attributes to generate the supply-side instruments.[8] For the supply side, we also add as an instrument the excluded demand-side variable, miles / dollar (but not the sums of this variable across other cars).

This set of instruments is likely to be highly collinear. To solve this problem, we remove collinearity as follows (see REMOVECOLLINEARITY.M for the implementation):

1. Choose a R^2 tolerance.
2. Start with the first column in instruments, let this be x.
3. Choose the next column in instruments, let this be y.
4. Regress y onto x. If the R^2 is lower than the tolerance, accept y as a new instrument. If the R^2 is higher than the tolerance, drop the column.
5. Redefine y to be the next column in instruments, and redefine x to be the columns of the accepted instruments.
6. Redo steps 4-6 until all instruments have either been accepted or dropped.

Following BLP (1999) code, we choose a R^2 tolerance of 0.99, and do this collinearity check separately for the demand- and supply-side instruments.[9]

In TRANS-244, we found that demeaning the instruments has the effect of reducing mechanical correlation among the moments and leads to standardized sensitivities smaller in magnitude. Thus, after removing collinear instruments, we demean all remaining, non-constant instruments prior to estimation to reduce mechanical correlation among the moments.

2 Evaluation of distance vector $G_j(\theta, s^n, P_{n_s})$

2.1 Solve for mean utility (δ)

Start with some estimate of δ, δ^0. We use the logit estimate (estimate with no random coefficients)[10]

$$\delta^0_{tj} = \ln\left(s^n_{tj}\right) - \ln\left(s^n_0\right),$$

[8] Based on BLP (1999) code, it appears as if BLP (1995) had a mistake in the implementation of their instruments. When creating ‘own’ firm instruments, instead of summing across a firm’s other models, they multiplied the instrument by the number of models the firm was producing.

[9] Consider $z_s^{own} = \{$constant, horsepower per weight, air conditioning, miles per dollar, and size$\}$ and $z_d^{own} = \{$constant, log of horsepower per weight, air conditioning, log of mpg, log of size, and year trends$\}$. For the supply-side instruments, the order is: each element of z_s^{own} in order, own firm instruments in same order, rival firm instruments in same order, and own miles per dollar. For the demand-side instruments, the order is: each element of z_d^{own} in order, own firm instruments in same order, and rival firm instruments in same order.

[10] Note that we use the logit estimate as the starting point for *every* iteration.

where s^n is the observed vector of sampled market shares, and s_0^n is the outside share. Then use the contraction mapping

$$\delta_{tj,\text{new}} = \delta_{tj,\text{prev}} + \ln\left(s_{tj}^n\right) - \ln\left[s_{tj}\left(p_{tj}, x_{tj}, \delta_{tj,\text{prev}}, P_{n_s}; \theta\right)\right]$$

to solve for δ. In each iteration of the contraction mapping, we estimate the market shares as follows. First, compute the individual market shares as (see BLP (1995) equation 6.6)

$$f\left(v_i, \delta_{tj}, p_{tj}, x_{tj}; \theta\right) = \frac{e^{\delta_{tj} + \mu\left(x_{tj}, p_{tj}, v_i; \theta\right)}}{1 + \sum_{j=1}^{J} e^{\delta_{tj} + \mu\left(x_{tj}, p_{tj}, v_i; \theta\right)}}.$$

Note that each market t is treated separately, i.e. the sum in the denominator is taken within each market. BLP (1995) is ambiguous on this treatment across markets,[11] but procedures in BLP (1999) code address this ambiguity. Second, integrate out over the distribution of v to obtain the market shares conditional only on product attributes. The simulator for this market share is

$$s\left(p_{tj}, x_{tj}, \xi_{tj}, P_{n_s}; \theta\right) = \frac{1}{n_s} \sum_{i=1}^{n_s} f\left(v_i, \delta_{tj}, p_{tj}, x_{tj}; \theta\right) w_i^u,$$

where w_i^u is the weight for individual i from importance sampling.

The contraction mapping continues until a stopping condition is met. Since the market shares are independent across markets, we can conduct a separate contraction mapping for each market.[12] In each iteration, we obtain vectors $\delta_{tj,\text{prev}}$ and $\delta_{tj,\text{new}}$. We continue onto the next iteration unless the following stopping condition is met:

$$\max_j \left(\left| \frac{\delta_{tj,\text{prev}}}{\delta_{tj,\text{new}}} - 1 \right| \right) \leq \delta_{\text{tol}},$$

where δ_{tol} is the tolerance, set to 10^{-14}.

2.2 Compute marginal cost

First, note that the price is additively separable in marginal cost and markup as follows

$$p = mc + \Delta\left(p, x, \xi; \theta\right)^{-1} s\left(p, x, \xi; \theta\right).$$

Let $b\left(p, x, \xi; \theta\right) = \Delta\left(p, x, \xi; \theta\right)^{-1} s\left(p, x, \xi; \theta\right)$ be defined as markup. Given markup, we then calculate marginal cost as $mc = p - b\left(p, x, \xi; \theta\right)$. Note that $\Delta\left(p, x, \xi; \theta\right)$ is a J by J matrix where whose (j, r) element is (see BLP (1995) equations 6.9a and 6.9b)

$$\Delta_{jr} = -\begin{cases} \int f\left(v_i, \delta_j, x_j, p_j; \theta\right)\left(1 - f\left(v_i, \delta_j, x_j, p_j; \theta\right)\right)\left[\partial \mu_{ij} / \partial p_j\right] P_0\left(dv\right), & j = r \\ \int -f\left(v_i, \delta_j, x_j, p_j; \theta\right) f\left(v_i, \delta_r, x_r, p_r; \theta\right)\left[\partial \mu_{ir} / \partial p_r\right] P_0\left(dv\right), & \text{if } \{j, r\} \in \text{same firm/market}, \\ 0, & \text{otherwise} \end{cases}$$

[11] At the beginning of section 5, they make the simplifying assumption that there is only a single cross section of autos in the data. Later details are written with this simplification in mind.

[12] Indeed, this is what is done in BLP (1999) code. Another advantage of this separation is that the markets can then be done in parallel, which in practice tends to decrease estimation run time by about 5-6x.

where

$$\partial\mu_{ij}/\partial p_j = \frac{-\alpha}{e^{m_t+\hat{\sigma}_y v_{iy}}}$$

(note that $\partial\mu_{ij}/\partial p_j$ is derived from the approximation of utility from prices as seen in section 1.2). As described in BLP (1995), the non-zero terms in Δ represent $-\partial s_r/\partial p_j$. BLP (1995) only explicitly states that the derivative is zero when cars j and r are made by different *firms*. The extension of this statement to *markets* comes from the fact that (as mentioned in section 2.1) market shares are independent across markets (i.e., changing the price of a car in one market has no effect on market shares in a different market).

We simulate the above integrals in a manner similar to our simulations for market shares. For example, the first term is

$$\frac{1}{n_s}\sum_{i=1}^{n_s} f\left(v_i,\delta_j,x_j,p_j;\theta\right)\left(1-f\left(v_i,\delta_j,x_j,p_j;\theta\right)\right)\left[\partial\mu_{ij}/\partial p_j\right]w_i^u,$$

where $f\left(v_i,\delta_j,x_j,p_j;\theta\right)$ is the *unweighted* individual market share as described in section 2.1, and w_i^u is the weight for individual i from importance sampling.

2.3 Obtain error terms from IV regression

In BLP (1995), the discussion splits the demand- and supply-side errors as follows. Given $\delta\left(x_{tj},p_{tj},\xi_{tj};\theta\right)$, the demand-side unobservable is

$$\xi_{tj} = \delta\left(x_{tj},p_{tj},\xi_{tj};\theta\right) - x_{tj}\beta,$$

where β is the estimate of an IV regression of δ_{tj} onto x_{tj}.

Then the supply-side unobservable is

$$\omega_{tj} = \ln\left(mc_{tj}\right) - w_{tj}\gamma,$$

where γ is the estimate of an IV regression of $\ln\left(mc_{tj}\right)$ onto w_{tj}. This means that β and γ are linear for any given (α,σ).

We combine the supply-side and demand-side IV regressions together into a single regression, which corresponds to the actual procedure done in the BLP (1999) code.[13] Let x (z_1) denote the demand-side variables (instruments) and w (z_2) denote the supply-side variables (instruments). Denote

$$X = \begin{pmatrix} x & 0 \\ 0 & w \end{pmatrix}$$

and

$$Z = \begin{pmatrix} z_1 & 0 \\ 0 & z_2 \end{pmatrix}.$$

Perform 2SLS on

$$\begin{bmatrix} \delta \\ \log(mc) \end{bmatrix} = X\begin{bmatrix} \beta \\ \gamma \end{bmatrix} + \begin{bmatrix} \xi \\ \omega \end{bmatrix}.$$

[13] See BLP (1999) code, STRUC-INST-STEVE2.ARC, procedure OBJ for confirmation. See BASE.PRG, line 739 for instrument construction.

The 2SLS projection matrix used is ZWZ', where W is the appropriate GMM weight matrix from below (see section 3), unless explicitely stated otherwise. Using the 2SLS coefficient estimates $(\hat{\beta},\hat{\gamma})'$, one can back-out estimates of the unobservables $(\hat{\xi},\hat{\omega})'$.

2.4 Interact instruments with error terms from IV regression

For a given car model, the moment condition associated with instrument ℓ for model m is

$$g_{\ell,m}(\theta) = \sum_{tj} z_{tj,\ell}\varepsilon_{tj},$$

where the sum is taken over all tj that belongs to the model.[14] Note that due to the stacking of the demand- and supply-side matrices, we actually need to consider two sets of tj's. For example, for the first model, we need to consider both the first car and the corresponding supply-side row n rows below. The portion of the GMM distance vector g associated with instrument ℓ is then defined as the mean across models:

$$g_{\ell}(\theta) = \frac{1}{n_m}\sum_{m} g_{\ell,m}(\theta).$$

Note that $g_{\ell}(\theta)$ is a scalar. g is then defined as the vector of these scalars across all instruments ℓ.

3 Estimation

As in BLP (1999) code, we use two-stage GMM to estimate the model. ESTIMATE.M contains all the necessary code to perform both stages, calling WTD_DIST and COMPUTEDISTANCEVECTOR.M as the objective function. We use Artelys KNITRO's Interior/Direct algorithm as our solver.

In addition to the data, the solver requires two inputs to begin — starting parameters and an initial GMM weight matrix. Given starting parameters, we first compute the distance vector for *each model*, using the standard 2SLS weight matrix for the IV regression portion of the distance vector computation (see section 2.3). Then, the initial weight matrix is calculated as the inverse of the variance-covariance matrix for the matrix of distance vectors, taking each model as one observation.

Next, the starting parameters and initial weight matrix are passed to the solver for the first-stage GMM estimation (minimizing $g'Wg$ where g is as defined in section 2.4, and W is the weight matrix).[15] Once we have the first-stage parameter estimates, we compute the distance vector again at these estimates (now using the GMM weight matrix for the 2SLS weights instead), and use this information to obtain a new weight matrix in the same manner as before. The first-stage parameter estimates and new weight matrix is then passed to the solver for the second-stage GMM estimation, which produces the final parameter estimates.

[14] According to BLP (1995) p. 863, the GMM distance vector g is defined as the average across models m of

$$g_m(\theta) = \sum_t \left[f_{mt}(z)^T \otimes I_2\right]\begin{bmatrix}\xi_{mt}(\theta)\\ \omega_{mt}(\theta)\end{bmatrix}$$

where $f_{mt}(z)$ is the basis function for the instruments. Note that $f_{mt}(z)^T \otimes I_2$ is simply the block-diagonal matrix with $f_{mt}(z)$ duplicated on the top left and the bottom right blocks. The only difference between this discussion and the one we implement (following BLP (1999) code) is the construction of the instruments matrix Z.

[15] Note that the weight matrix is also used in the 2SLS estimation, as described in section 2.3.

Note that the final parameter estimates contain a random coefficient (α, σ) portion (which is the entire parameter space from the perspective of the solver) and a mean coefficient (β, γ) portion (which is can be computed directly from (α, σ)).

In summary, the steps are:

1. Determine a set of starting random coefficient parameters, $(\alpha, \sigma)_{\text{start}}$. We use BLP (1995) published values as our starting parameters.

2. Using the standard 2SLS projection matrix for IV regression $Z(Z'Z)^{-1}Z'$, compute the distance vector $\hat{g}_m$ for *each* model m at $(\alpha, \sigma)_{\text{start}}$.

3. Define the initial GMM weight matrix W_{GMM1} as the inverse of the variance-covariance matrix for the distance vectors $\hat{g}_m$.

4. Using $(\alpha, \sigma)_{\text{start}}$ and W_{GMM1}, estimate the first-stage parameter estimates $(\alpha, \sigma)_{\text{GMM1}}$ via GMM.

5. Compute the new weight matrix W_{GMM2} in the same manner as before, but at $(\alpha, \sigma)_{\text{GMM1}}$ and using W_{GMM1} for the 2SLS weight matrix instead.

6. Using $(\alpha, \sigma)_{\text{GMM1}}$ and W_{GMM2}, estimate the final parameter estimates $(\alpha, \sigma)_{\text{GMM2}}$ via GMM. Extract β and γ using $(\alpha, \sigma)_{\text{GMM2}}$.

We constrain all elements of σ to be non-negative.

4 Post-estimation objects

Table V from BLP (1995) details demand elasticities for a sample of cars at the estimated parameters. Though we are unable to reproduce these numbers, BLP (1999) code shows precisely how this should be calculated (see BASE.PRG, near lines 960 and 1008 for price and attribute elasticities, respectively). We implement the BLP (1999) code procedures described in the sections below.

4.1 Price elasticities

Recall that in section 2.2, we already computed the derivative of market shares with respect to price. In this case, we only need own-price elasticities, which is calculated the same way as before:

$$\frac{\partial s_j}{\partial p_j} = \int f(v_i, \delta_j, x_j, p_j; \theta)(1 - f(v_i, \delta_j, x_j, p_j; \theta))[\partial \mu_{ij} / \partial p_j] P_0(dv),$$

where

$$\partial \mu_{ij} / \partial p_j = \frac{-\alpha}{e^{m_t + \hat{\sigma}_y v_{iy}}}.$$

We then simulate the above integrals the same way as before. Note that we multiply the end result by -1 to give positive price elasticities.

4.2 Attribute elasticities

Consider an increase of 0.10 in product attribute k. From section 2.1, the individual market share is entirely determined by mean utilities δ and random utilities μ in the following way:

$$f(v_i, \delta_{tj}, p_{tj}, x_{tj}; \theta) = \frac{e^{\delta_{tj} + \mu\left(x_{tj}, p_{tj}, v_i; \theta\right)}}{1 + \sum_{j=1}^{J} e^{\delta_{tj} + \mu\left(x_{tj}, p_{tj}, v_i; \theta\right)}},$$

where δ_{tj} is a function of the mean parameter β:

$$\delta_{tj} = x_{tj}\beta + \xi_{tj}.$$

Thus, we only need the new values of δ and μ to obtain the associated new market shares. To that end, we increase the mean utility δ by $0.10\beta_k$ and compute a new μ after increasing x_k by 0.10 for every car. We then compute the new market shares using the updated δ and μ. Then, the partial derivative of market shares with respect to characteristic k is numerically estimated as

$$\frac{\partial s}{\partial x_k} = \frac{\text{new shares} - \text{old shares}}{0.10}.$$

The elasticity of demand with respect to characteristic k is then

$$\frac{x_k}{s} \frac{\partial s}{\partial x_k}.$$

5 Sensitivity

In this section, note that since in section 2.4 the distance vectors are means across *models* as opposed to individual *cars*, we essentially have n_m as opposed to n observations. As such, all of our computations and scalings use n_m instead of n. Note that with the exception of the final $1/n_m$ scaling, using $1/n$ would make no difference to the final standardized sensitivities as all the scalings cancel out anyway.

The Jacobian of the distance vectors with respect to the parameters, $\hat{G}$, is computed in the COMPUTE_JACOBIAN function in ESTIMATE.M numerically using small perturbations to the parameters. When computing the Jacobian for the random coefficient (α, σ) parameters, the IV regression is re-estimated to obtain new error terms. When computing the Jacobian for the mean (β, γ) parameters, the IV regression is skipped and the error terms are instead computed with the perturbed mean parameters. Note that since the GMM distance vector g is defined as the mean across models, naturally the Jacobian is in these units as well.

For the weight matrix, we use $\hat{W}_g = W_{\text{GMM2}}$ (see section 3). The moment variance-covariance matrix, $\hat{\Omega}_{gg}$, is just the sample variance-covariance matrix of $\hat{g}(\theta)$ taken across models. Note that each observation corresponds to a model rather an an individual car. This is computed in COMPUTE_OMEGA function in ESTIMATE.M.

Our sample sensitivity and standarized sample sensitivity computations use the matrix $\hat{A}$ in addition to these objects. When computing the pth column of $\hat{A}$ as $\hat{G}_p' \hat{W}_g \hat{g}_m$, we determine each matrix $\hat{G}_p$ numerically by calculating the change in $\hat{G}$, the Jacobian as computed using the method described above, associated with small perturbations to the estimated parameters. We then compute the parameter variance-covariance matrix

$\Sigma_{\theta\theta}$ using theorem (13.1) in Greene (2012), where θ contains both the random coefficient paramters (α, σ) and the mean parameters (β, γ). (Sample) sensitivity and standardized sensitivity are then computed using these objects.