From Market Shares to Consumer Types: Duality in Differentiated Product Demand Estimation

Myrto Kalouptsidi*

Department of Economics
Yale University

May 2008

Abstract

A widely applied method for differentiated product demand estimation, introduced by Berry, Levinshon and Pakes (1995), is founded on matching observed and theoretical market shares of products. In this paper, we derive an equivalent formulation to BLP by assuming that consumer tastes are discrete, rather than continuous. Our approach is based on a matching occurring in the consumer type space, rather than the product space. The equivalence between the two formulations expresses a duality between market shares and consumer types. The dual formulation, introduced in this paper is computationally more efficient than the primary, while the adoption of discrete tastes allows for correlation of tastes across different product characteristics. Simulation exercises show that the dual method is significantly faster.

*I would like to thank my advisors, Steve Berry and Phil Haile for numerous comments and suggestions.
1 Introduction

Firms, regulating authorities, economists and several operating agents are interested in knowing demand in particular markets. Indeed, understanding how consumers operate and how the resulting market shares of firms are formed is an important question in economic thought and practice. Moreover, estimating demand poses a difficult problem for economists, as individual consumer choice data are not frequently available; often the only available data include the market shares of firms operating in a market.

One of the most well-known methods for differentiated product demand estimation was developed by Berry, Levinshon and Pakes (1995). This method corrects for the major disadvantages of the logit model (namely, independence of irrelevant alternatives) and provides an intuitive, unified framework for analyzing demand. Indeed, it has found numerous applications in a wide variety of markets, such as automobiles (BLP (1995) and (2004), Petrin (2002)), ready-to-eat cereal (Nevo (2001)), satellite television (Goolsbee and Petrin (2004)), personal computers (Goeree (2005)) and others. The BLP model rests upon a random coefficients utility maximization paradigm and postulates the presence of an unobserved product characteristic. It then proceeds by picking the value of these unobservable product characteristics to match the observed market shares to those predicted by the theoretical model. Finally, it estimates the demand parameters via a moment condition expressing the independence of these product unobservables from available instruments. In typical markets with a large number of products, the complexity of the BLP method becomes a critical issue: one needs to use Monte Carlo integration techniques to derive the theoretical market shares, as well as solve a system of high dimensional non-linear equations, via a fixed point algorithm, repeatedly in the GMM minimization procedure. Indeed, the dimension of this system is equal to the number of products which is very large. Moreover, we show that the limit of the speed of convergence of this fixed point algorithm is greater than one, as the number of products goes to infinity.

The BLP formulation places emphasis on the market shares and the corresponding space of products. In this paper, we introduce an equivalent formulation based upon consumer "types" and their corresponding space. We refer to the former as the "prime" BLP formulation and the latter as the "dual" BLP formulation. The equivalence between the two expresses a duality between consumer types and product market shares. We assume that consumer tastes for products are drawn from a discrete distribution, rather than a continuous one, as in BLP. This allows us to transform the \((J \times J)\) system of market share equations to an equivalent \((T \times T)\) system of consumer types equations which we can solve for some new, "consumer unobservables". In many markets of interest, \(T\) is significantly smaller than
rendering computations in the dual domain more attractive. Indeed, correlations between tastes of different product characteristics are allowed. On the other hand, the number of parameters to be estimated increases with the number of types. Consumer types have been extensively used in the literature, especially in marketing applications (e.g. Chintagunta et al (1991), Kamakura and Russel (1989)).

Bajari, Fox, Kim and Ryan (2007) in a recent working paper explore other ways to reduce the computational costs of differentiated product demand estimation via the usage of discrete consumer tastes: they propose using linear regression methods to gauge these the consumer type weights, allowing for non-parametric assumptions about the taste parameters.

Finally, we perform a series of simulation exercises aiming to compare the primary and the dual method. First, we retrieve the product unobservables in several combinations of numbers of types and products using the true parameter values. This experiment shows that the dual exhibits higher speed of convergence than BLP under certain conditions. Indeed, the dual formulation requires considerably fewer iterations (up to an order of thousands) than the BLP approach. Our main simulation exercise consists of the estimation of the model via GMM. We present two examples (that differ on the magnitude of the share of the outside good, which is the driving force of the BLP speed of convergence) which confirm that the dual is considerably faster than BLP. Indeed, the dual method’s implementation takes some minutes, while BLP requires several hours.

In section 2, the BLP approach to demand estimation is described. In section 3 the dual formulation is derived. Finally, section 4, presents the results from the simulation exercises performed. The dependence of the speed of convergence on the number of products for the BLP algorithm is established in the appendix.

2 The BLP Approach to Demand Estimation

Demand estimation is one of the most important issues that applied IO is concerned with. Knowledge of market demand is crucial to both firms and regulating authorities. A widely used estimation method was first introduced by Berry, Levinshon and Pakes (1995), henceforth BLP. This model allows estimation even when the available data contain only the market shares of the operating firms and the product characteristics. A main contribution of the BLP model is the treatment of endogenous prices in a random coefficients utility framework, which in turn allows for plausible substitution patterns. We briefly present this method below and refer the interested reader to Berry, Levinshon and Pakes (1995) for a more detailed analysis.

In the BLP framework, the utility individual \(i\) derives from product \(j\) takes the following
form:

\[ u_{ij} = x_j \beta_i + \xi_j + \epsilon_{ij} \]  \hspace{1cm} (1)

where \( x_j \) is a vector of observed (by the researcher) product characteristics (including price), \( \beta_i \) is a vector of "tastes" for consumer \( i \), capturing how person \( i \) values each of \( j \)'s observed characteristics, \( \xi_j \) represents an unobserved (by the researcher) product characteristic, such as style or quality and \( \epsilon_{ij} \) is the idiosyncratic taste that individual \( i \) holds for product \( j \). \( \epsilon_{ij} \) is assumed to be iid across agents and products and to follow the double exponential distribution ("logit error term"). The taste parameters \( \beta_i \) may include observed consumer characteristics (such as income, gender, etc.), as well as unobserved consumer characteristics.

In the standard BLP, the unobserved components of \( \beta \) are assumed to be continuous random variables, usually taken to be normally distributed. Each consumer \( i \) (who can observe all the elements in her utility function) purchases the product that yields the highest utility out of all alternatives (including an "outside good", representing the option of not buying any product).

The first step of the estimation procedure consists of inferring the product unobservables by matching observed and predicted-by the model-market shares. Initially, the predicted market shares of each product must be computed. To this end, we calculate the purchase probabilities for consumers conditional on \( \beta_i \), by "integrating out" the logit error terms and then we form the aggregate market shares by integrating out the \( \beta_i \). Some of the taste parameters have known distributions (income, gender, etc.), while distributional assumptions are made for some others (usually normal). Typically, to calculate the integrals present in the market shares, we resort to Monte Carlo integration techniques. These market share equations depend on the product unobservables \( \xi \):

\[
\sigma_j(\xi) = \int \frac{e^{x_j b + \xi_j}}{\sum_k e^{x_k b + \xi_k}} f(b) \, db, \quad j = 0, 1, \ldots, J
\]  \hspace{1cm} (2)

where \( f(\cdot) \) is the cdf of the consumer tastes \( \beta \). We, therefore, end up with \( J \) nonlinear equations which can be solved for the \( J \) unknowns \( \xi \). Indeed, BLP show that the operator

\[
T_j(\xi) = \xi_j + \log(s_j) - \log \left( \int \frac{e^{x_j b + \xi_j}}{\sum_k e^{x_k b + \xi_k}} f(b) \, db \right), \quad j = 1, \ldots, J
\]  \hspace{1cm} (3)

has a unique fixed point \( \xi \) (for every value of \( \beta \)).

The next step in the estimation procedure is to gauge the parameters of interest \( \beta \) via GMM. We assume that the product unobservables are independent of some variables \( w \) at
the true parameter value \( \beta_0 \):

\[
E \left[ \xi_j (\beta_0) \right] = 0
\]

The vector \( w \) is then used as instruments and usually includes observed characteristics of product \( j \) as well as the characteristics of the competing products (excluding prices). It is reasonable to expect that \( \xi_j \) is correlated with prices.

It is customary to supplement the demand analysis with a price-setting equilibrium equation derived from the firm side of the market. Joint estimation of a demand and a "supply" side yields more precise estimates. Although it is straightforward to add a pricing equation to our estimation procedure, we will not do so here.

In this paper, we are concerned with the BLP formulation delineated above, but with random consumer tastes represented by discrete random variables. Adopting discrete tastes (which we call discrete consumer "types") has several advantages:

- they offer computational savings (more on that below)
- they allow for correlation of tastes across different product characteristics
- they can approximate continuous random variables
- they often match our intuition better.

Berry, Carnall and Spiller (2006)- henceforth BCS- assume that \( \beta_i \) are drawn from a bimodal distribution, so that there are two discrete (unobserved) consumer types. They analyze the airlines market and assume that there are two types of consumers: "business" and "tourists". Business customers are individuals who don’t care much about the price of the trip, but are mostly interested in the time of the flight, whether it is direct, whether the airport is a hub, etc. In contrast, tourists are price sensitive, but don’t care much about the frequency of flights or the number of connections.

3 The Dual Formulation: From Market Shares to Consumer Types

3.1 Derivation

The BLP formulation described above places emphasis on the market shares and the corresponding space of products \((\mathbb{R}^J)\). In this section we introduce an equivalent formulation based upon consumer types and their corresponding space \((\mathbb{R}^T)\). We refer to the former as
the "prime" BLP formulation and the latter as the "dual" BLP formulation. The equivalence between the two expresses a duality between consumer types and product shares.

Next, we reformulate the market share equations ("primary") to this new set of "dual" equations. The dual equations have the same form as the market share equations, but hold in a much lower dimensional space, as we expect that $T << J$. We adopt the discrete consumer tastes setting and employ a finite number of consumer types to derive our results.

We assume that the utility individual $i$ derives from product $j$ is given by equation (1), but that $\beta_i$ can take on one of $T$ possible values ($T$ consumer types). For instance, in BCS, $T = 2$ and travellers can be either tourists or business. In addition, let $\gamma_t$ denote the probability with which type $\beta_t$ appears.

Consider a finite random variable taking values in the finite set of products $\{0, ..., J\}$ whose density is given by $\{\sigma_0, ..., \sigma_J\}$. Likewise, consider the finite random variable taking values in the finite set of consumer tastes $\{\beta_1, ..., \beta_T\}$, each $\beta_t$ being a $k$-dimensional vector ($k$ is the number of product characteristics), with density $\{\gamma_1, ..., \gamma_T\}$. The conditional probability $p(j|\beta_t)$ represents the odds of product $j$ being chosen by a consumer of type $\beta_t$.

The law of total probability gives:

$$\sigma_j = \sum_{t=1}^{T} \gamma_t p(j|\beta_t), \quad j = 0, 1, ..., J \quad (4)$$

In the case of logit error terms, the market shares become:

$$\sigma_j = \sum_{t=1}^{T} \gamma_t \frac{e^{x_j \beta_t + \xi_j}}{\sum_{j=0}^{J} e^{x_k \beta_t + \xi_k}} \quad (5)$$

The BLP estimation procedure is founded on these equations: $\sigma_j$ is replaced by its observed value and the product unobservables are retrieved.

Instead of working with market shares, we can calculate the consumer type probabilities by applying the law of total probability:

$$\gamma_t = \sum_{j=1}^{J} \sigma_j p(\beta_t|j), \quad t = 1, ..., T \quad (6)$$

We denote by $p(\beta_t|j)$ the probability that tastes are of type $\beta_t$, conditional on purchase of product $j$. By Bayes’ rule,

$$p(\beta_t|j) = \frac{\gamma_t \Pr(j|\beta_t)}{\sum_{m=1}^{T} \gamma_m p(j|\beta_m)} \quad (7)$$
Equations (6), then, match consumer type frequencies to their theoretical counterpart. The significant advantage of the new set of equations is their lower dimensionality, which we discuss at the end of this section.

In the widely used case in which \( \varepsilon_{ij} \) is distributed \( iid \) double exponential, (6) take a very special form: they become identical to the corresponding market share equations. Indeed, in this case,

\[
p(\beta_t | j) = \frac{e^{x_j \beta_t + \xi_j}}{\sum_{k=0}^J e^{x_k \beta_t + \xi_k}} \tag{8}
\]

Moreover, since only differences in utilities of different products are necessary to determine conditional purchase probabilities, we normalize the observed and unobserved characteristics of the outside good to zero, as is customary, so that \( x_0 = \xi_0 = 0 \) and thus \( u_{t0} = \varepsilon_{t0} \).

Notice, then, that the purchase probability of the outside good (or rather, the probability of no purchase) by consumer type \( \beta_t \) is:

\[
p(0 | \beta_t) = \frac{1}{\sum_{k=0}^J e^{x_k \beta_t + \xi_k}} \tag{9}
\]

Therefore, by (7):

\[
p(\beta_t | j) = \frac{\gamma_t p(0 | \beta_t) e^{x_j \beta_t + \xi_j}}{\sum_{m=1}^T \gamma_m p(0 | \beta_m) e^{x_j \beta_m + \xi_j}}
\]

or

\[
p(\beta_t | j) = \frac{\gamma_t p(0 | \beta_t) e^{x_j \beta_t}}{\sum_{m=1}^T \gamma_m p(0 | \beta_m) e^{x_j \beta_m}}
\]

Now let

\[
q_t = p(0 | \beta_t) \gamma_t, \ t = 1, \ldots, T \tag{10}
\]

Replacing above, the consumer type equations become

\[
\gamma_t = \sum_{k=0}^J s_k \frac{e^{x_k \beta_t + \log q_t}}{\sum_{m=1}^T e^{x_k \beta_m + \log q_m}}, \ t = 1, \ldots, T \tag{11}
\]

The system of equations (11) constitutes the system of "dual" equations. We call these equations "dual", because they have exactly the same form as the original market share equations (5), but lie in the type space, instead of the product space. Indeed, notice that the type probabilities \( \gamma_t \) in the primary equations, have been replaced in the dual by the observed market shares, \( s_k \). Similarly, the product characteristics \( x_j \) in the primary equations, have
given their place to the consumer tastes $\beta_i$ in the dual equations. Finally, the unobserved characteristic $\xi_j$ has been substituted by a new element, $\log q_t$, which is a function of the vector $\xi$ and can be thought of as a "consumer unobservable".

In the primary problem, the researcher faces a highly non-linear system of $J$ equations in $J$ unknowns. These equations match the observed market shares to the market shares that the model predicts. We are therefore looking for a solution $\xi$ that sets these two quantities as close as possible. The $T$ dual equations contain the $T$ unknowns, $q$. These equations match the consumer type probabilities to their theoretical counterpart. In doing so we have significantly reduced the dimensionality of the problem. Indeed, the number of products $J$ is likely to be much higher than the number of consumer types, $T$. In contrast, the number of products in a market can be immense. For instance, BCS face 14,000 markets out of which 140 include more than 100 products (with a max of 874), BLP face 20 markets with a number of products between 72 and 150. Reducing the dimensionality of the problem, facilitates and accelerates estimation- remember this problem needs to be solved repeatedly a large number of times by the maximization algorithm used in the GMM optimization. Indeed, the simulation exercises (see Section 4) confirm that the dual methods requires a few minutes when BLP requires several hours.

As already mentioned, in the original problem we have set $\xi_0 = 0$. Similarly, a constraint is necessary in the dual problem, which is implied by the condition $\xi_0 = 0$. Indeed, writing up equation (4) for $j = 0$ we get that

\[ s_0 = \sum_{t=1}^{T} q_t \tag{12} \]

It is straightforward to show that this relation is equivalent to assuming $\xi_0 = 0$ in the original problem.

### 3.1.1 Solving the Dual Problem

As already mentioned, BLP show that the operator (3) is a contraction mapping and thus has a unique fixed point $\xi^*$. In particular, the proof is applicable for discrete densities of consumer tastes, where the integral becomes a sum (the calculations are straightforward).

Since the dual equations have the same form as the original ones, one might deduce that the same arguments apply and that the corresponding operator has a unique fixed point $(\log q)^*$. The constraint (12), however, requires caution.

Consider the vector-valued function with components:
\[ F_t(q) = q_t \frac{\gamma_t}{\sum_{k=0}^{J} s_k e^{x_k \beta_t q_t} / \sum_{m=1}^{T} e^{x_k \beta_m q_m}}, \quad t = 1, ..., T - 1 \]  

(13)

and

\[ F_T(q) = s_0 - \sum_{t \neq T} q_t \]

Restriction (12) implies that \( q \)’s generated by \( F_t(q) \) must be in the set:

\[ C = \left\{ (q_1, ..., q_{T-1}) : \sum_{t=1}^{T-1} q_t \leq s_0 \right\} \]

If we could guarantee that \( q \)’s remain in \( C \), we could use the following iterative algorithm:

\[ q_t^{n+1} = q_t^n \frac{\gamma_t}{\sum_{k=0}^{J} s_k e^{x_k \beta_t q_t} / \sum_{m=1}^{T} e^{x_k \beta_m q_m}}, \quad t = 1, ..., T - 1 \text{ and } n = 1, 2, ... \]

Unfortunately, however, it is not guaranteed that \( q \)’s remain in \( C \) in every step of the algorithm. Indeed, in simulation exercises we observed that in the second iteration, \( q \)’s fell outside of \( C \), resulting in \( q_T < 0 \) which rendered all \( q \)’s negative in the next step.

The "logit transformation" will solve this problem, though. Indeed, let \( \{a_1, ..., a_T\} \) be such that:

\[ q_t = s_0 \sum_i e^{a_i} \]

Replacing this expression in the dual equations (11), we get:

\[ \gamma_t = \sum_{k=0}^{J} s_k \frac{e^{x_k \beta_t + a_t}}{\sum_{m=1}^{T} e^{x_k \beta_m + a_m}}, \quad t = 1, ..., T \]

Now we can use the following algorithm to solve for \( \{a_1, ..., a_T\} \):

\[ F_t(a) = a_t \frac{\gamma_t}{\sum_{k=0}^{J} s_k e^{x_k \beta_t + a_t} / \sum_{m=1}^{T} e^{x_k \beta_m + a_m}}, \quad t = 1, ..., T - 1 \]  

(14)

and \( a_T = 0 \). The corresponding \( q \)’s are in \( C \), while convergence of the algorithm is known from BLP.

Nevertheless, simulations show that using (13) and constraining the \( q \)’s in \( C \) often convergences (for instance if \( T = 2 \) it almost always works) and is always faster than (14). Moreover, its speed of convergence does not seem to depend on \( T \), in contrast to (14). We
therefore propose and implement the following algorithm: begin using (13) and check in each iteration whether the generated \(q\)'s are in \(C\). At the first iteration where they fall outside of \(C\), switch to (14). This way, one can benefit from the speed of (13) as long as possible (and many times throughout the GMM procedure) and switch to (14) which is guaranteed to converge if necessary and which is still faster than the original BLP.

The computational burden of an algorithm is defined by two factors: the number of calculations performed in each step and the number of steps necessary for convergence (speed of convergence). It is immediate that the dual formulation corresponds to an algorithm that performs fewer computations in each step (iteration) than the primary. Indeed, the dual performs \(T\) instead of \(J\) calculations in each iteration, so that if each calculation requires \(m\) seconds, the dual takes \(Tm\) seconds, while BLP takes \(Jm\) seconds. What is less obvious is that the dual formulation exhibits a higher speed of convergence, as well. Indeed, we show that the speed of convergence for the BLP algorithm depends on the dimension of the problem, \(J\): as \(J \to \infty\), the speed of convergence becomes higher than unity. Simulations exercises confirm this dependence: dual formulation requires much fewer iterations to converge. For instance, in a market with 100 products and 2 consumer types, the BLP formulation may require about 10,000 iterations while the dual about 30 (see Table 1 below).

### 3.2 Estimation Procedure: Return to product unobservables \(\xi\)

The proposed estimation procedure works first on the dual space of consumer types to determine \(q_t\) from (11) and then determines \(\xi_j\) via the following formula:

\[
s_j e^{-\xi_j} = \sum_{t=1}^{T} e^{x_j \beta_t} q_t, \quad j = 1, \ldots, J
\]

or

\[
\xi_j = \log \left( \frac{s_j}{\sum_{t=1}^{T} e^{x_j \beta_t} q_t} \right), \quad j = 1, \ldots, J
\]  

(15)

and \(\xi_0 = 0\). Having determined \(\xi_j\) GMM is performed as in BLP. The moment condition is of the form

\[
E \left[ \xi_j (\beta_0, \gamma_0) f_j(w) \right] = 0
\]

for some function \(f(\cdot)\) of the instruments, \(w\). In practice, we seek for the values of \(\beta\) and \(\gamma\) that set the sample moments as close to zero as possible. We, therefore, solve the following
minimization problem:

$$\min_{\beta} \| \sum_{j=0}^{J} \xi_j (\beta, \gamma) f_j(w) \|^2$$  \hspace{1cm} (16)$$

There are numerous references concerning the instruments. Instruments usually include own and competing products’ observable characteristics, as well as factors present in the marginal cost equations when the firm side is added to the estimation procedure. Moreover, if multiproduct firms operate in the market examined, instruments may consist of own (product and cost) characteristics as well as the characteristics of the products produced by the same firm and by rival firms.

One issue that arises when using the dual formulation is the increase of the number of parameters to be estimated. For instance, with two consumer types one has to estimate \((2k + 1)\) parameters instead of \(2k\) parameters if normality is assumed. In the general case, we need to estimate \((Tk + T - 1)\) (when \(T = 2\) there is only one extra parameter). A way out of this, is attaching random coefficients only to a subset of the product characteristics. Another solution is assuming the type probabilities are known. Moreover, note that the small number of parameters to be estimated in the case of normal consumer tastes is a consequence of the independence of tastes across different product characteristics. This assumption is somewhat unrealistic and adding such correlations increases the parameters to be estimated.

4 Simulations

In this section we present results from simulation exercises performed to compare the performance of the dual and the BLP formulation.

Consider a fictional market with \(J\) products (the outside good is subsumed in \(J\)). There are two product characteristics, including price. We use \(x \sim U [0, 2]\) and \(\xi_{true} \sim U [-\frac{1}{2}, \frac{1}{2}]\). The price of each good \(j\) is a function of \(\beta x_j\) and "instruments" \(w\) (vectors in the null space of \(\xi_{true}\)). Finally, we used closed-form solutions for the market shares (as if no sampling error).

Table 1 shows the number of iterations necessary for convergence. In particular, we consider one market and calculate the product unobservables at the true parameter values with both methods. In a sense, this is the "best case scenario" as the number of iterations is expected to increase at other (wrong) parameter values. The BLP and dual fixed point algorithms were performed 100 times (parameter values remain fixed, while \(x\) and \(\xi_{true}\) change). The number of iterations varies with the range of observed and unobserved product characteristics (hence the high variance in some cases), as well as the values of the parameters \((\beta, \gamma)\), but the proportion between BLP and dual remains approximately constant.
In Table 1, the logit transformation of the $q$'s is used to guarantee that the algorithm always converges. Adopting the "mixed" algorithm proposed above, one can gain significant speed benefits. Indeed, in the GMM exercise presented below, the "mixed" algorithm is used. In addition, we should mention that the magnitude of the outside is the major determinant of the speed of convergence (as is clear from the appendix). Therefore, should we repeat this experiment with a large share for the outside alternative, the two methods become almost equivalent in terms of the speed of convergence. Nevertheless, the dual still performs fewer calculations in each step of the algorithm and therefore, as the number of products increases, the dual remains faster (see below). Finally, we should mention that the computer used is an Aspire 5612 WLMi Inter Core Duo processor T2300, 2GB RAM.

<table>
<thead>
<tr>
<th>$J$</th>
<th>$T = 2$</th>
<th>$T = 5$</th>
<th>$T = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BLP</td>
<td>DUAL</td>
<td>BLP</td>
</tr>
<tr>
<td>50</td>
<td>5311.12</td>
<td>30.3</td>
<td>1462.94</td>
</tr>
<tr>
<td></td>
<td>(1063.68)</td>
<td>(0.95)</td>
<td>(122.8)</td>
</tr>
<tr>
<td>Avg $s_0$</td>
<td>0.0038</td>
<td>0.014</td>
<td>0.006</td>
</tr>
<tr>
<td>100</td>
<td>9769</td>
<td>30.29</td>
<td>2806.19</td>
</tr>
<tr>
<td></td>
<td>(1609.5)</td>
<td>(0.64)</td>
<td>(211.2)</td>
</tr>
<tr>
<td>Avg $s_0$</td>
<td>0.002</td>
<td>0.007</td>
<td>0.003</td>
</tr>
<tr>
<td>300</td>
<td>27869.37</td>
<td>30.1</td>
<td>7723.38</td>
</tr>
<tr>
<td></td>
<td>(3123.58)</td>
<td>(0.33)</td>
<td>(536.42)</td>
</tr>
<tr>
<td>Avg $s_0$</td>
<td>0.0006</td>
<td>0.0023</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Averages and st. deviations based on 100 inversions at the true parameter values. $x$'s and $\xi$'s drawn uniform random. By ’ we denote minutes and by ” seconds.

Subsequently, we perform GMM by minimizing the objective function (16). We report results from two different GMM experiments. In the first experiment, we generate data from ten markets with $J \in \{20, 30, 40, 50\}$, and $T = 2$ consumer types while $x$, $\xi_{true}$, and prices are generated as above. We employ the Nelder-Mead algorithm. $[\beta_{11}, \beta_{12}]$ correspond to consumer type $\beta_1$, while $[\beta_{21}, \beta_{22}]$ correspond to consumer type $\beta_2$. We show results from 4 different initial conditions ($bg01$, $bg02$, $bg03$, $bg04$) used to initiate the optimization.
algorithm\(^1\). The last two columns present the results from GMM (they are identical for all initial conditions). The last two rows present the time needed for each of the two methods to complete the exercise. We want to estimate \( T \times k + (T - 1) = 5 \) parameters. In this example, the share of the outside alternative is small (as in Table 1 above) and thus the dual is faster from the BLP both due to its higher speed of convergence and due to the fact that in each iteration it performs fewer calculations. Table 2 indicates that even in this example with only a small number of products, the dual offers major computational advantages\(^2\):

<table>
<thead>
<tr>
<th>( \beta_{11} )</th>
<th>bg01</th>
<th>bg02</th>
<th>bg03</th>
<th>bg04</th>
<th>BLP</th>
<th>Dual</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_{12} )</td>
<td>-1</td>
<td>-1</td>
<td>-0.9</td>
<td>-0.8</td>
<td>-1.3</td>
<td>-1</td>
</tr>
<tr>
<td>( \beta_{21} )</td>
<td>3</td>
<td>3</td>
<td>3.1</td>
<td>3.2</td>
<td>2.7</td>
<td>3</td>
</tr>
<tr>
<td>( \beta_{22} )</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.4</td>
<td>-0.3</td>
<td>-0.8</td>
<td>-0.5</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>0.65</td>
<td>0.65</td>
<td>0.75</td>
<td>0.85</td>
<td>0.35</td>
<td>0.65</td>
</tr>
</tbody>
</table>

| BLP: | 6 hours | 11 hours | 14 hours | 41 min | 18 hours |
| Dual: | 1 min 36 sec | 4 min 20 sec | 5 min 28 sec | 9 min 8 sec |

As the share of the outside good is often large in actual applications, we generate a different "dataset" for the second experiment, where this share is large: to this goal, we choose \( x \sim U [-7, 0] \) and \( \xi_{true} \sim U [-2.5, 0] \). In this case, the two methods have a comparable speed of convergence. Therefore, we expect the dual to be faster only due to the fact that it involves fewer calculations at each step of the algorithm \((T < J)\). We use again 10 markets and \( J \in \{100, 200, 300, 400\}\)\(^3\). Table 3 shows that even though BLP becomes faster, the dual method still offers significant computational benefits:

---

1. The Nelder-Mead algorithm (as most optimization algorithms) requires a set of initial conditions for the variables that is solved for. The closer to the true values that the initial condition is, the faster the algorithm converges.
2. The parameters are retrieved exactly due to the fact that \( \xi_{true} \) is chosen significantly smaller than \( x \).
3. In fact, 5 markets have 100 products, 3 have 300 products and the last two have 200 and 400 products.
Table 3: Runs GMM, Several Initial Conditions, Time. Big Outside Share

<table>
<thead>
<tr>
<th></th>
<th>TRUE</th>
<th>bg01</th>
<th>bg02</th>
<th>bg03</th>
<th>bg04</th>
<th>BLP</th>
<th>Dual</th>
</tr>
</thead>
<tbody>
<tr>
<td>β11</td>
<td>2</td>
<td>2</td>
<td>2.1</td>
<td>2.2</td>
<td>1.7</td>
<td>2</td>
<td>-2</td>
</tr>
<tr>
<td>β12</td>
<td>-1</td>
<td>-1</td>
<td>-0.9</td>
<td>-0.8</td>
<td>-1.3</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>β21</td>
<td>3</td>
<td>3</td>
<td>3.1</td>
<td>3.2</td>
<td>2.7</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>β22</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.4</td>
<td>-0.3</td>
<td>-0.8</td>
<td>-0.5</td>
<td>-0.5</td>
</tr>
<tr>
<td>γ1</td>
<td>0.65</td>
<td>0.65</td>
<td>0.75</td>
<td>0.85</td>
<td>0.35</td>
<td>0.65</td>
<td>0.65</td>
</tr>
</tbody>
</table>

BLP: 2 hours 4 hours 9 min 4 hours 21 min 6 hours 12 min
Dual: 17 min 35 min 37 min 54 min

5 Conclusion

In this paper, we introduce an equivalent formulation to the BLP model which lies upon consumers types, rather than market shares. The dual formulation is computationally less demanding when the number of types is smaller than the number of products and exhibits higher speed of convergence under certain conditions. Our simulations results show that computation time can be reduced to some minutes when using the dual formulation, compared to several hours when using the BLP formulation.

Appendix

A Speed of Convergence

The rate of convergence for the BLP contraction mapping is governed by the parameter ζ, given by:

\[ \zeta = \sum_{m=1}^{J} \frac{\partial T_j(\xi)}{\xi_m} \]

for suitable \( j \) and \( \xi \). Then,

\[ \zeta = 1 - \sum_{m=1}^{J} \frac{1}{s_j \xi_m} \]

where \( s_j \) is given by (5). Note that

\[ \frac{\partial s_j}{\partial \xi_m} = \begin{cases} - \sum_t \gamma_t p(j|\beta_t) p(m|\beta_t) & \text{for } m \neq j \\ s_j - \sum_t \gamma_t p(j|\beta_t)^2 & \text{for } m = j \end{cases} \]

See the Appendix in BLP (1995).
Substituting this in (18) we get that:

\[
\zeta = \frac{1}{s_j} \sum_{m=1}^{J} \sum_{t} \gamma_t p(j|\beta_t) p(m|\beta_t) = \\
= \frac{1}{s_j} \sum_{t} \gamma_t p(j|\beta_t) (1 - p(0|\beta_t)) = \\
= 1 - \frac{1}{s_j} \sum_{t} \gamma_t p(j|\beta_t) p(0|\beta_t)
\]

Now notice that

\[
\zeta \geq 1 - \frac{1}{s_j} \max_{\beta_t} p(0|\beta_t) \sum_{t} \gamma_t p(j|\beta_t) = \\
= 1 - \max_{\beta_t} p(0|\beta_t)
\]  \hspace{1cm} (19)

\[
= 1 - \max_{\beta_t} p(0|\beta_t)
\]  \hspace{1cm} (20)

In the preceding analysis, the number of products \( J \) remains fixed. Next, we consider the behavior of \( \zeta \) as \( J \) becomes large. For this purpose, we explicitly indicate the dependence of \( \zeta \) and \( p(0|\beta_t) \) on \( J \) by writing \( \zeta_J \) and \( p_J(0|\beta_t) \). We show that:

\[
\lim_{J \to \infty} p_J(0|\beta_t) = 0
\]

provided that \( x_J \) and \( \xi_J \) remain bounded\(^5\).

Indeed, the sequence \( p_J(0|\beta_t) \) is updated as follows:

\[
p_{J+1}(0|\beta_t) = \frac{1}{1 + \sum_{j=1}^{J+1} e^{x_j \beta_t + \xi_j}} = \\
= \frac{p_J(0|\beta_t)}{1 + p_J(0|\beta_t) e^{x_{J+1} \beta_t + \xi_{J+1}}}
\]

It follows that \( p_J(0|\beta_t) \) is decreasing and bounded for all \( \beta_t \) and, thus, convergent to limit \( M \). Taking limits in both sides of:

\[
p_{J+1}(0|\beta_t) \left( 1 + p_J(0|\beta_t) e^{x_{J+1} \beta_t + \xi_{J+1}} \right) = p_J(0|\beta_t)
\]

we get that:

\[
M \left( 1 + M \lim_{J \to \infty} e^{x_{J+1} \beta_t + \xi_{J+1}} \right) = M
\]

If \( M \neq 0 \), we must have \( x_{J+1} \beta_t + \xi_{J+1} \to -\infty \), which can’t hold if \( x_J \) and \( \xi_J \) remain

\(^5\)Analogous considerations are given in Bajari and Benkard (2003).
bounded. Therefore, \( M = 0 \) and by (19),

\[
\lim_{J \to \infty} \zeta(J) \geq 1
\]

References


