FROM MARKET SHARES TO CONSUMER TYPES: DUALITY IN DIFFERENTIATED PRODUCT DEMAND ESTIMATION

MYRTO KALOUPTSIDI*

Department of Economics, Yale University, New Haven, CT, USA

SUMMARY
A widely applied method for differentiated product demand estimation, introduced by Berry, Levinsohn and Pakes in 1995, is founded on matching observed and theoretical market shares of products. In this paper, we allow for discrete consumer tastes and derive an equivalent matching occurring in the consumer type space. The equivalence between the two formulations expresses a duality between market shares and consumer types. In applications where a large number of products and a small number of consumer types is natural, the dual formulation introduced in this paper is computationally more efficient than the primal. Indeed, simulation exercises show that the dual method can be significantly faster. Copyright © 2010 John Wiley & Sons, Ltd.

Received 15 January 2009; Revised 18 January 2010

1. INTRODUCTION

One of the most well-known methods for differentiated product demand estimation was developed by Berry et al. (1995; henceforth, BLP). In markets with a large number of products, the complexity of the BLP method becomes a critical issue: for each tried parameter, one needs to use Monte Carlo integration techniques to derive the theoretical market shares of products, as well as solve a high-dimensional system of nonlinear equations via a fixed-point algorithm. The dimension of this system is equal to the number of products, which may be very large.

The formulation introduced in BLP focuses on the market shares of products. In this paper, we derive an equivalent formulation based upon consumer types. We refer to the former as the primal 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. Both continuous (e.g. BLP; Nevo, 2001) and discrete (e.g. Berry et al., 2006; Hastings, 2008) consumer types have been used in the literature.¹ In this paper, we assume that consumer tastes for products are drawn from a discrete distribution. This allows us to transform the system of J market share equations to an equivalent system of T consumer type equations which we solve for some new consumer unobservables. In many markets of interest, a large number of products and a small number of consumer types may be natural, rendering computations in the dual domain more attractive. On the other hand, in markets where many consumer types are needed, the dual method loses its computational benefits. Adopting discrete consumer types has the additional advantages of avoiding the Monte Carlo integration techniques in computing market shares, potentially approximating continuous

* Correspondence to: Myrto Kalouptsidi, Department of Economics, Yale University, 37 Hillhouse Ave, New Haven, CT 06511, USA. E-mail: myrto.kalouptsidi@yale.edu

¹ The marketing literature uses discrete consumer types extensively (e.g. Dubé et al., 2008).

Copyright © 2010 John Wiley & Sons, Ltd.
consumer tastes and allowing for correlation of tastes across different product characteristics.\(^2\) Finally, Heckman and Singer (1984) stress the sensitivity of estimates to parametric assumptions of unobserved heterogeneity, as well as the flexibility of discrete heterogeneity distributions, while noting that in practice it is often difficult to find more than a few different types.

The numerical performance of the BLP estimation procedure has recently attracted increased interest (e.g. Dubé et al., 2009; Knittel and Metaxoglou, 2008). Dubé et al. (2009) suggest recasting estimation as a mathematical program with equilibrium constraints (MPEC), in order to eliminate the numerical error of the nested fixed-point algorithm, as well as increase the speed of estimation. The dual method we propose can be used both with the fixed-point algorithm as well as the MPEC formulation.\(^3\)

We propose a fixed-point algorithm for the dual formulation and show in simulation exercises that it can be up to 20 times faster than the BLP algorithm. It benefits both from the fewer calculations made in each step, when \(T < J\), as well as a higher rate of convergence, especially when the share of the outside good is small. Compared to the MPEC method of Dubé et al. (2009), the dual formulation benefits from having fewer unknown parameters: \(T\) instead of \(J\).

Section 2 derives the dual formulation and Section 3 describes its solution for the consumer unobservables. Section 4 presents the results from the simulation exercises performed, while Section 5 concludes. In the Appendix we examine the relation between the number of products and the contraction modulus of the BLP fixed-point algorithm.

2. THE DUAL FORMULATION

BLP provide a model and method for demand estimation that has become central to the empirical I.O. literature.\(^4\) In the BLP framework, the utility that individual \(i\) derives from product \(j\) takes the following form:

\[
\begin{align*}
  u_{ij} &= x_j \beta_i + \xi_j + \epsilon_{ij} \\
  &\text{where } x_j \in \mathbb{R}^k \text{ is a vector of observed product characteristics (including the price), } \\
  &\xi_j \text{ is a scalar unobserved (by the researcher) product characteristic, } \\
  &\beta_i \text{ is a vector of individual attributes of consumer } i \text{ and } \\
  &\epsilon_{ij} \text{ is the idiosyncratic taste that individual } i \text{ holds for product } j. \\
\end{align*}
\]

A critical assumption for our derivation is that \(\epsilon_{ij}\) is independent and identically distributed across agents and products and follows the double exponential distribution.\(^5\) Each consumer chooses among all products, including an outside good that represents the option of not buying any of the products, the alternative that yields the highest utility. We assume that consumer tastes \(\beta_i\) can take on one of \(T\) possible values in \(\mathbb{R}^k\), corresponding to the \(T\) consumer types. Let \(\gamma_i\) denote the probability with which type \(\beta_i\) appears. Finally, we adopt the standard normalization \(x_0 = \xi_0 = 0\).

Consider a finite random variable taking values in the finite set of products \(\{0, \ldots, J\}\) with probabilities \(\sigma_0, \ldots, \sigma_J\). \(\sigma_j\) represents the market share of product \(j\). Likewise, consider the finite

\(^2\)Even though continuous consumer tastes offer a more general specification, it is most often assumed that they follow independent normal distributions. This assumption, adopted in order to decrease the number of unknown parameters, may be problematic. For example, Dubé et al. (2008) stress the necessity for correlations and non-normality of tastes to fit their data.

\(^3\)Bajari et al. (2009) reduce the BLP algorithm to a linear regression, using (a large number of) discrete consumer types.

\(^4\)For more details on the framework and estimation procedure, see BLP.

\(^5\)Note that the adoption of random coefficients \(\beta_i\) allows for a flexible specification and avoids the well-known limitations of the simple logit model.
random variable taking values in the finite set of consumer tastes \( \{ \beta_1, \ldots, \beta_T \} \). Each \( \beta_t \in \mathbb{R}^k \) occurs with probability \( \gamma_t \). The conditional probability \( p(j|\beta_t) \) represents the odds of product \( j \) being chosen by a consumer of type \( \beta_t \). Since \( \varepsilon_{ij} \) follows the double exponential distribution we have

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

As \( x_0 = \xi_0 = 0 \), the purchase probability of the outside good by consumer type \( \beta_t \) is

\[
p(0|\beta_t) = \frac{1}{\sum_{k=0}^{J} e^{x_k \beta_t + \xi_k}} \tag{2}
\]

The law of total probability gives

\[
\sigma_j = \sum_{t=1}^{T} \gamma_t p(j|\beta_t), \quad j = 0, 1, \ldots, J \tag{3}
\]

or

\[
\sigma_j = \sum_{t=1}^{T} \gamma_t \frac{e^{x_j \beta_t + \xi_j}}{\sum_{k=0}^{J} e^{x_k \beta_t + \xi_k}}, \quad j = 0, 1, \ldots, J \tag{4}
\]

The BLP estimation procedure is founded on the market share equations (4). In particular, observed market shares, \( s_j \), replace \( \sigma_j \), \( 1 \leq j \leq J \), leading to solutions for \( \xi_j \), \( 1 \leq j \leq J \), as a function of the parameters of interest \( (\beta_t, \gamma_t) \), \( 1 \leq t \leq T \), or in vector form \( (\beta, \gamma) \).

Instead of working with market shares, we similarly calculate the consumer type probabilities:

\[
\gamma_t = \sum_{j=0}^{J} s_j p(\beta_t|j), \quad t = 1, \ldots, T \tag{5}
\]

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

\[
p(\beta_t|j) = \frac{\gamma_t p(j|\beta_t)}{\sum_{m=1}^{T} \gamma_m p(j|\beta_m)} \tag{6}
\]

Since \( \varepsilon_{ij} \) is a distributed i.i.d. double exponential, (5) takes a very special form: it becomes identical to the corresponding market share equation. Indeed, combining (1) and (2) we obtain

\[
p(j|\beta_t) = p(0|\beta_t) e^{x_j \beta_t + \xi_j}
\]
Substituting this into (6) leads to

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

Now let

\[
q_t = p(0|\beta_t) \gamma_t, \quad t = 1, \ldots, T
\]

Replacing in (5), the consumer type equations become

\[
\gamma_t = \sum_{j=0}^{J} s_j \frac{e^{x_j,\beta_t+\log q_t}}{\sum_{m=1}^{T} e^{x_j,\beta_m+\log q_m}}, \quad t = 1, \ldots, T
\]

As already mentioned, in the original formulation we set \( \xi_0 = 0 \). This translates in the dual domain in the following condition:\(^6\)

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

The system of equations (8) constitutes the system of dual equations. We call these equations dual, because they have the same form as the original market share equations (4), but lie in the consumer type space instead of the product space. Indeed, note that the type probabilities, \( \gamma_t \), in the primal equations, have been replaced in the dual by the observed market shares, \( s_j \). Similarly, the product characteristics \( x_j \) in the primal equations have given their place to the consumer tastes \( \beta_t \) 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 \( \xi \in \mathbb{R}^J \) and can be thought of as a consumer unobservable.

The proposed estimation procedure works first on the dual space of consumer types to determine \( q_t, 1 \leq t \leq T \) from (8) and (9), as described in Section 3. It then determines \( \xi_j, 1 \leq j \leq J \) from

\[
s_j = \sum_{t=1}^{T} \gamma_t \frac{e^{x_j,\beta_t+\xi_j}}{\sum_{k=0}^{J} e^{x_k,\beta_t+\xi_k}} = \sum_{t=1}^{T} \gamma_t p(0|\beta_t) e^{x_j,\beta_t+\xi_j} = \sum_{t=1}^{T} q_t e^{x_j,\beta_t+\xi_j}
\]

or

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

\(^6\) Equation (9) obtains when we substitute the definition of \( q_t \), (definition of 7), in (3) written for \( j = 0 \).

Copyright © 2010 John Wiley & Sons, Ltd.

*J. Appl. Econ.* (2010)

DOI: 10.1002/jae
and $\xi_0 = 0$. Having determined $\xi_j, 1 \leq j \leq J$, GMM is performed as in BLP, i.e. using the mean independence assumption $E[\xi_j(\beta_0, \gamma_0)h(z_j)] = 0$, where $h(\cdot)$ is a function of appropriate instruments, $z_j, 1 \leq j \leq J$. In practice, we seek the values of $(\beta, \gamma)$ that set the sample moments, $g(\beta, \gamma) = \frac{1}{J} \sum_{j=1}^{J} \xi_j(\beta, \gamma)h(z_j)$, as close to zero as possible. We therefore solve the minimization problem:

$$\min_{\beta, \gamma} g(\beta, \gamma)' W g(\beta, \gamma)$$

(11)

where $W$ is the GMM weighting matrix.

Adopting discrete consumer tastes requires estimation of $(Tk + T - 1)$ parameters. The formulation in BLP that uses continuous normally distributed consumer tastes requires $(k + k (k + 1)/2)$ parameters, corresponding to the means and the covariance matrix. Often it is assumed, somewhat unrealistically, that tastes for product characteristics are uncorrelated, resulting in $2k$ parameters. The number of parameters in that case becomes roughly the same as the discrete case for 2 consumer types. Discrete types may offer a good compromise by allowing for correlations among the random coefficients, while maintaining a reasonable number of parameters.

The following proposition summarizes the above derivation:

**Proposition 1** The BLP market share equations (4) and the dual consumer type equations (8), along with (10), are equivalent, in that numerically identical $\xi_j (\beta, \gamma), 1 \leq j \leq J$ are obtained from the two formulations.

In the primal problem, the researcher faces a nonlinear system of $J$ equations and $J$ unknowns. These equations match the observed to the theoretical market shares. The $T$ dual equations contain the $T$ unknowns, $q_t, 1 \leq t \leq T,$ significantly reducing the dimensionality of the problem, in cases where $T << J$. Indeed, in some applications it is natural to expect that the number of products $J$ is much higher than the number of consumer types, $T$. For instance, in the airline application found in Berry et al. (2006), there are 14,000 markets that represent origin–destination cities air travel. Many markets include more than 100 products (i.e. combination of airline, fare and itinerary), with a maximum of 874 products, while it is reasonable to assume there are two consumer types who have systematically different tastes for observed characteristics: ‘business’ (not price sensitive, interested in the time of the flight, whether it is direct, etc.) and ‘tourists’ (price sensitive, care less about the frequency of flights or the number of connections, etc.). BLP face 20 markets consisting of 72–150 products each. Reducing the dimensionality of the problem facilitates and accelerates estimation, as it needs to be solved repeatedly by the GMM optimization algorithm and for multiple starting values.

The above analysis assumes that the number of types is known. If the number of types is unknown, robustness of results should be examined, as in Berry et al. (2006). In simulation exercises conducted, increasing the number of types beyond its true value led to very small probabilities, $\gamma_t$, for the excess types.

---

7 We abuse notation slightly here as we really have $T - 1$ unknowns: as in the primal, one of the unknowns is given by the normalization, (9).
3. SOLUTION OF THE DUAL EQUATIONS

In this section we propose a fixed-point algorithm for the dual equations, which differs from BLP, due to the normalization (9).

BLP show that the market share equations can be solved for $\xi^2 / Ropen J$, via a fixed-point algorithm of the form $\xi^n = H(\xi^{n-1})$, where $H: \mathbb{R} \to \mathbb{R}^J$. They show that $H(\cdot)$ is a contraction mapping and, therefore, has a unique fixed point $\xi^* \in \mathbb{R}^J$.

Consider the vector-valued function $F: \mathbb{R}^{T-1} \to \mathbb{R}^{T-1}$:

$$F_t(r) = r_t + \log(\gamma_t) - \log \left( \frac{\sum_{j=0}^J s_j e^{x_j \beta_t + r_t}}{\sum_{m=1}^M e^{x_m \beta_m + r_m}} \right), \quad t = 1, \ldots, T - 1 \tag{12}$$

and let $r = \log(q), q \in \mathbb{R}^{T-1}$, while $r_T$ is defined by $r_T = \log \left( s_0 - \sum_{t \neq T} e^r_t \right)$.

$r_T$ is well defined only if $s_0 - \sum_{t \neq T} e^r_t \geq 0$. In order to use (12) as an iterative algorithm, it must be that $r^n$ calculated in the $n$th iterative step remains in the following set:

$$C = \left\{ (r_1, \ldots, r_{T-1}) : \sum_{t=1}^{T-1} e^{r_t} \leq s_0 \right\}$$

In general, the set $C$ is not invariant under $F$. A simple transformation, though, guarantees convergence of the dual algorithm: let $r_t$ be such that $r_T = 0$ and $\{r_1, \ldots, r_{T-1}\}$ satisfy

$$r_t = \frac{e^{r_t}}{\sum_i e^{r_i}}, \quad t = 1, \ldots, T - 1 \tag{13}$$

$F(\cdot)$ solves for the unobservables $r$ and thus $q$ as well, through the iterative algorithm $r^{n+1} = F(r^n)$ and $r^n_T = 0$. Convergence of this algorithm to a unique fixed point follows from the analysis of BLP. The fixed point found by $F(\cdot)$ leads to $q_t, 1 \leq t \leq T - 1$, via (13).

Simulations show, however, that when using $r = \log(q)$, instead of the logit transformation above, $F(\cdot)$ converges frequently and is faster. We therefore implement the following ‘mixed’ algorithm: use $r = \log(q)$, but check in each iteration whether $F(r)$ is in $C$. If/when $r$ falls outside of $C$, switch to the logit transformation specified by (13). The mixed algorithm benefits from the speed of $F(r), r = \log(q)$, for as long as possible and multiple times throughout the GMM procedure, while it switches to the logit transformation specified in (13) if necessary, which is still faster than the original BLP algorithm (see Section 4).

The computational burden of the iterative algorithm is characterized by two factors: the number of calculations performed in each step and the number of steps necessary for convergence: the rate of convergence. It is immediate that the dual formulation is accompanied by an algorithm

---

It is easy to see that $r_t, 1 \leq t \leq T - 1$ are uniquely determined by (13).

The BLP proof, carried out for continuous consumer types, can be adapted to discrete densities of consumer tastes; calculations are straightforward.
that performs fewer computations in each iteration than the primal, as long as \( T < J \). Moreover, the rate of convergence for the BLP algorithm depends on the dimension of the problem, \( J \): as shown in the Appendix, as \( J \to \infty \), the contraction modulus approaches 1 and hence the speed of convergence becomes low. Indeed, the simulation exercises in Section 4 show that the dual algorithm requires up to 30 times fewer iterations to converge. The mixed algorithm in (12) exhibits a high speed of convergence due to restriction (9). Indeed, the smaller the market share of the outside good, i.e. the stricter the requirement imposed by (9), the faster the mixed algorithm becomes. In contrast, when the share of the outside good is large, the difference in the rate of convergence of the dual and the BLP algorithm decreases considerably. Nevertheless, the dual still performs fewer calculations in each iteration.

Finally, instead of the fixed-point algorithm, one can use the MPEC approach. Dubé et al. (2009) suggest maximizing the GMM objective function over \((\xi, \gamma)\), but also \( x_i, 1 \leq j \leq J \) and \( (\xi, 1 \leq j \leq J) \) for multiple combinations of \( T \) and \( J \). Table I shows the number of iterations required for convergence. As the magnitude of the outside good is a key determinant of the rate of convergence (see Appendix), we keep \( s_0 \) constant as \( J \) increases, by decreasing the mean value of \( x_j \).\(^\text{10}\) We perform the exercise for a small outside share of about 5%, as well as a large outside share of about 90%. Each product has \( k = 2 \) observed characteristics, including price, which is constructed as \( p_j = |0.5\xi_j + \epsilon_j + 1.1x_j| \), where \( \xi_j \sim N(0,0.4) \) and \( \epsilon_j \sim N(0,1) \). The taste for price is drawn from a uniform distribution \( U(-2,0) \), while the taste for the other observed characteristic is drawn from \( U(0,7) \). \( \gamma_0 \) is drawn from \( U(0,1) \) so that their sum equals 1. When the outside share is small, the BLP algorithm requires 20–30 times the number of iterations that the dual requires. When the outside share is large, this ratio falls to about 2.5.

In the second simulation exercise, we perform GMM by minimizing the objective function (11). We again report results for both small and large outside share. We construct 50 independent markets with \( J = 100 \) products each. We let \( T = 2 \) consumer types, while \( x_j, \xi_j \) and prices are generated as above. Six instruments are generated as \( z_jd \sim U(0,1) + 0.25(\epsilon_j + 1.1x_j) \), \( d = 1, \ldots, 6 \) and we include \( \left\{ z^2_{jd}, z^3_{jd}, x^2_j, x^3_j, \prod_{d=1}^6 z_{jd}, x_j, z_{jd}x_j \right\} \) as moments, similar to Dubé et al., (2009). Results are reported in Table II. \([\beta_{11}, \beta_{12}]\) correspond to consumer type \( \beta_1 \), while \([\beta_{21}, \beta_{22}]\) correspond to

\(^{10}\)See the footnotes in Table I for the construction of \( x_j \) at each \((J, T)\) combination.
Table I. Number of iterations for one market at the true parameter values

<table>
<thead>
<tr>
<th></th>
<th>$T = 2$</th>
<th></th>
<th>$T = 3$</th>
<th></th>
<th>$T = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BLP</td>
<td>Dual</td>
<td>BLP</td>
<td>Dual</td>
<td>BLP</td>
</tr>
<tr>
<td>$J = 50$</td>
<td>BLP</td>
<td>Dual</td>
<td>BLP</td>
<td>Dual</td>
<td>BLP</td>
</tr>
<tr>
<td></td>
<td>560</td>
<td>18</td>
<td>20</td>
<td>7</td>
<td>831</td>
</tr>
<tr>
<td></td>
<td>(316)</td>
<td>(5)</td>
<td>(15)</td>
<td>(2)</td>
<td>(585)</td>
</tr>
<tr>
<td>$\bar{e}$</td>
<td>5%$^a$</td>
<td>81%$^b$</td>
<td>4%</td>
<td>85%</td>
<td>5%</td>
</tr>
<tr>
<td>$J = 100$</td>
<td>BLP</td>
<td>Dual</td>
<td>BLP</td>
<td>Dual</td>
<td>BLP</td>
</tr>
<tr>
<td></td>
<td>462</td>
<td>15</td>
<td>15</td>
<td>6</td>
<td>597</td>
</tr>
<tr>
<td></td>
<td>(215)</td>
<td>(4)</td>
<td>(8)</td>
<td>(1.3)</td>
<td>(351)</td>
</tr>
<tr>
<td>$\bar{e}$</td>
<td>6%$^c$</td>
<td>87%$^d$</td>
<td>5%</td>
<td>90%</td>
<td>6%</td>
</tr>
<tr>
<td>$J = 300$</td>
<td>BLP</td>
<td>Dual</td>
<td>BLP</td>
<td>Dual</td>
<td>BLP</td>
</tr>
<tr>
<td></td>
<td>507</td>
<td>14</td>
<td>14</td>
<td>6</td>
<td>604</td>
</tr>
<tr>
<td></td>
<td>(205)</td>
<td>(4)</td>
<td>(5)</td>
<td>(1)</td>
<td>(333)</td>
</tr>
<tr>
<td>$\bar{e}$</td>
<td>5%$^e$</td>
<td>90%$^f$</td>
<td>5%</td>
<td>92%</td>
<td>6%</td>
</tr>
</tbody>
</table>

$a \sim N(-1,1)$.
$b \sim N(-3,1)$.
$c \sim N(-1.5,1)$.
$d \sim N(-3.5,1)$.
$e \sim N(-2,1)$.
$f \sim N(-4,1)$.

Table II. GMM simulation results

<table>
<thead>
<tr>
<th></th>
<th>$\beta_{11}$</th>
<th>$\beta_{12}$</th>
<th>$\beta_{21}$</th>
<th>$\beta_{22}$</th>
<th>$\gamma_1$</th>
<th>time BLP</th>
<th>time DUAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small $s_0$</td>
<td>True</td>
<td>2</td>
<td>-1</td>
<td>3</td>
<td>-0.5</td>
<td>0.65</td>
<td>24.4</td>
</tr>
<tr>
<td></td>
<td>Est</td>
<td>1.93</td>
<td>-1.04</td>
<td>3</td>
<td>-0.55</td>
<td>0.53</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>s.e.</td>
<td>(0.12)</td>
<td>(0.06)</td>
<td>(0.34)</td>
<td>(0.17)</td>
<td>(0.2)</td>
<td>(2)</td>
</tr>
<tr>
<td>Big $s_0$</td>
<td>Est</td>
<td>1.98</td>
<td>-0.99</td>
<td>2.92</td>
<td>-0.55</td>
<td>0.57</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>s.e.</td>
<td>(0.05)</td>
<td>(0.02)</td>
<td>(0.28)</td>
<td>(0.14)</td>
<td>(0.12)</td>
<td>(2)</td>
</tr>
</tbody>
</table>

5. CONCLUSION

The dual formulation provides a transformation to the BLP estimation procedure for differentiated products demand estimation. It is equivalent to the BLP formulation, but computationally more

---

11 Standard errors are calculated based on the estimate of the asymptotic variance–covariance matrix of the parameters in BLP.
12 The actual times required are heavily dependent on the software, the tolerance levels and the starting values. The ratio of times, however, is robust. The KNITRO optimization package is dramatically faster than the Nelder–Mead algorithm (MATLAB’s fminsearch command). Tight tolerance levels were set, following Dubé et al. (2009).
13 We also perform a simulation exercise that examines the case where the number of consumer types is unknown. Data are generated by three consumer types, while four types are used in the estimation. The fourth type has negligible probability, $\gamma$, while three types are recovered correctly.

tractable when $T < J$. Indeed, simulation exercises, aimed at comparing run times for the two algorithms, suggest that the dual can be significantly faster. In applications where a specification with a large number of products and a small number of consumer types is natural, the dual formulation provides important practical advantages.

ACKNOWLEDGEMENTS

I would like to thank my advisors, Steve Berry and Phil Haile, for their continuous encouragement, as well as their valuable comments and suggestions. I would also like to thank three anonymous referees, as well as the participants at the Yale IO seminar for their helpful comments.

REFERENCES

APPENDIX

We show that the limit of the contraction modulus of the BLP fixed-point algorithm approaches one as the number of products grows to infinity. The rate of convergence for the BLP contraction mapping is governed by the contraction modulus $\zeta$:

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

(14)

**Proposition 2** If the $(J \times k)$ matrix $x$ and the $(J \times 1)$ vector $\xi$ remain bounded as $J \to \infty$, $\lim_{J \to \infty} \zeta \geq 1$.

**Proof** Note that

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

Substituting in (14) we get

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

\[
= 1 - \frac{1}{\sigma_j} \sum_t \gamma_t j p(j|\beta_t) p(0|\beta_t)
\]

Now note that

\[
\zeta \geq 1 - \frac{1}{\sigma_j} \max_{\beta_t} p(0|\beta_t) \sum_t \gamma_t j p(j|\beta_t) = 1 - \max_{\beta_t} p(0|\beta_t)
\]

(15)

Consider the behavior of $\zeta$ as $J$ becomes large. We show that $\lim_{J \to \infty} p_J(0|\beta_t) = 0$, provided that $x_J$ and $\xi_J$ remain bounded.$^{14}$ Indeed, the sequence $p_J(0|\beta_t)$ is updated as follows:

\[
p_{J+1}(0|\beta_t) = \frac{1}{1 + \sum_{j=1}^{J+1} e^{x_t j + \xi_J}}
\]

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

It follows that $p_J(0|\beta_t)$ is decreasing and bounded for all $\beta_t$ and thus converges to a limit $M$. Taking limits in both sides of the following expression:

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

we get

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

If $M \neq 0$, we must have $x_{J+1} j + \xi_{J+1} \to -\infty$, which cannot hold if $x_J$ and $\xi_J$ remain bounded. Therefore, $M = 0$ and by (15), $\lim_{J \to \infty} \xi_J \geq 1$.

---

$^{14}$ Analogous considerations are given in Bajari and Benkard (2003).