The flexible coefficient multinomial logit (FC-MNL) model of demand for differentiated products

Peter Davis*

and

Pasquale Schiraldi**

We show FC-MNL is flexible in the sense of Diewert (1974), thus its parameters can be chosen to match a well-defined class of possible own- and cross-price elasticities of demand. In contrast to models such as Probit and Random Coefficient-MNL models, FC-MNL does not require estimation via simulation; it is fully analytic. Under well-defined and testable parameter restrictions, FC-MNL is shown to be an unexplored member of McFadden’s class of Multivariate Extreme Value discrete-choice models. Therefore, FC-MNL is fully consistent with an underlying structural model of heterogeneous, utility-maximizing consumers. We provide a Monte-Carlo study to establish its properties and we illustrate its use by estimating the demand for new automobiles in Italy.

1. Introduction

In this paper, we describe a fully parametric, flexible, discrete choice model of demand, together with the methods required to estimate it using aggregate data. The contribution of the paper is to provide a discrete choice demand model with an analytic market share function that, in contrast to existent analytic models, is flexible – in the sense that it can match own- and cross-price elasticities of demand for a well-defined set of data generating processes (DGP). Under well-defined, testable, parameter restrictions, the model is a previously unexplored member of the class of structural discrete choice demand models developed in the series of seminal papers by McFadden (1978, 1981).

* Compass Lexecon and University College London Faculty of Laws; pdavis@compasslexecon.com.
** London School of Economics and Political Science & CEPR; p.schiraldi@lse.ac.uk.

Thanks are due to The British Academy, grant number LRG-39888, for their generous funding, and to Ricardo Ribeiro for his excellent research assistance. Naturally, this paper is a research paper, representing solely the view of the authors. This paper has evolved and been expanded but also draws upon a paper first circulated in 2001 under the title of “Demand Models for Market Level Data,” mimeo MIT (http://web.mit.edu/pjdavis/www/papers/mainbody_paper.pdf) and subsequently expanded and recirculated in 2006 as CEPR working paper #5880 under the title of “The Discrete Choice Analytically Flexible (DCAF) Model of Demand for Differentiated Products” (http://www.cepr.org/pubs/new-dps/dplist.asp?dpno=5880.asp). Last, but certainly not least, we should add our wholehearted thanks to the editor and three referees whose constructive comments and suggestions have materially improved the paper.
The Flexible Coefficient-MNL (FC-MNL) model developed here has some potentially significant advantages relative to existent parametric discrete choice models currently used in practice. For example, popular models such as the Multinomial Logit (MNL) and the Nested Multinomial Logit models (NMNL) models are well known to impose severe restrictions on the substitution patterns (estimated own- and cross-price elasticities) between goods that can possibly be estimated. Ideally, in any demand study, we would like the data to drive flexible models to capture the truth about the nature of substitutability between any pair of goods. In practice, popular demand models implicitly impose a great deal of implausible structure on the nature of estimated substitution patterns. See, for example, Berry and Pakes (2007) for a critique of the NMNL model.

A number of recent authors have attempted to address these concerns using three approaches. Firstly, authors have proposed using more flexible parametric models. Secondly, authors have introduced unobserved consumer heterogeneity in an attempt to “free-up” the models’ elasticities. And thirdly, authors have proposed using semiparametric methods. This paper follows authors using the first of these three approaches, which we now discuss in turn.

Recent papers proposing “more” flexible parametric forms for discrete choice models include Ben-Akiva and Bielaire (1999), Bresnahan, Stern, and Trajtenberg (1997), Chu (1989), Verboven and Brenkers (2002), Wen and Koppelman (2001), and Koppelman and Sethi (2000) among others. The papers in this tradition are closest to our paper. In particular, each of these authors considers a less restrictive member of the class of MEV models than MNL and moreover, this branch of the literature has specifically developed with the aim of building parametric models with more flexible substitution patterns than MNL. However, perhaps surprisingly, since such claims would form the bedrock for justifying particular parametric models in the continuous choice demand literature (see, e.g., Christensen Jorgenson & Lau, 1975; Deaton & Muelbauer, 1980; Banks Blundell & Lewbel, 1996; and/or Pollak & Wales, 1995), to our knowledge none of the authors writing in what we might call the parametric discrete choice literature has made Diewert’s (1974) flexibility claims for their proposed MEV specifications. In contrast, we demonstrate that the FC-MNL model provides a flexible functional form in the sense of Diewert (1974), ensuring that it can match a well-defined class of possible own- and cross-price elasticities of demand.

An additional advantage of the FC-MNL model is that it nests the familiar Multinomial Logit (MNL) model, and hence Wald or Lagrange Multiplier tests for the validity of MNL can be constructed. Imposing fewer restrictions on a subset of the model’s parameters yields previously unexplored members of the class of Multivariate Extreme Value (MEV) discrete choice demand models developed in the series of seminal papers by McFadden (1978, 1981). As such, the FC-MNL is a fully structural discrete choice model under well-defined parameter restrictions. Naturally, the parameter constraints required to make FC-MNL consistent with an underlying MEV model may or may not be imposed in estimation. This allows the researcher to test whether the data is consistent with the parametric restrictions which embody restrictions on underlying consumer behaviour like Slutsky symmetry and possibly use the unrestricted model. Such an approach is motivated solely by pragmatism, and it would certainly be preferable to have an underlying utility model from which the resulting demand model is derived and welfare implications could be analyzed.

---

1. Despite their rather severe disadvantages, MNL and NMNL are popular in a wide variety of contexts where the estimated substitution patterns are crucial for guiding policy and where billions of pounds are at stake. For example, MNL and NMNL have been used to evaluate market definition questions and also to simulate the impact on prices of horizontal mergers between firms in concentrated industries. See, for example, Werden and Froeb (1994), Epstein and Rubinfeld (2001), Ivaldi and Verboven (2005), and the related literature described in Davis and Garces (2009) and Davis (2000).

2. The recent industrial organization literature has followed McFadden (1978) in describing the class of models we consider here as the class of Generalized Extreme Value (GEV) models. We shall use the term MEV models because that is used in the more recent statistics literature (see, e.g., Joe, 1996, and in particular Kotz, Balakrishnan, & Johnson, 2000, chapter 47, and citations therein). The term GEV is used in the statistics literature to mean a unifying framework for the three types of univariate extreme value distributions.
The second approach constitutes a very significant attempt to provide models with potentially flexible substitution patterns by introducing observed and unobserved consumer heterogeneity across individuals to model the correlation of utilities across options and individuals. Well-known examples include the Probit model proposed by Hausman and Wise (1978) and the Mixed-MNL model which evolved from McFadden (1978, 1981) and which is also known as the Random-Coefficient Multinomial Logit (RC-MNL) model or the Error-Component Logit (ECL) model (Ben-Akiva & Bolduc, 1996; Train, 2003; Green et al., 2006). The advantage of mixed-MNL models is that their empirical implementation can use random coefficients sparingly by incorporating a base level of consumer heterogeneity in the closed form MNL (or other base) model. Doing so reduces computational complexity. See in particular Boyd and Mellman (1980) and Cardell and Dunbar (1980) for the first applications of the aggregate demand RC-MNL model and the more recent seminal contribution in Berry, Levinsohn, and Pakes (1995) (henceforth BLP).

The FC-MNL model has both advantages and disadvantages relative to such models. A potential disadvantage is that, at least in principle, a general enough version of the RC-MNL class of models may be more flexible. Specifically, while we will show that the FC-MNL model is Diewert (1974) flexible, McFadden and Train (2000) show that in principle, RC-MNL models are flexible in a far more general sense. That said, we show that a large subclass of the RC-MNL models — including the specifications typically estimated — for example that are used in BLP — do not always make the resulting demand model particularly flexible. More favourably, an advantage of the model discussed here is that it entirely avoids using simulation techniques, which (i) introduce an additional layer of computational intensity, meaning they typically take far longer to estimate than models which do not require simulation and (ii) generate instability and numerical errors which have been widely discussed in the literature in the context of RC-MNL estimation (see, e.g., Freyberger, 2012; Knittel & Metaxoglou, 2011, 2012; Judd & Skrainka, 2011). The fact that the FC-MNL model avoids simulation entirely also makes it potentially particularly useful in a variety of practical policy settings, for example, those where computational constraints currently mean simulation based models are too costly to apply. One such important arena involves the estimation of dynamic choice models. Recent papers in this area (see Gowrisankaran & Rysmann, 2012; Schiraldi, 2011, among others) have combined RC-MNL within an optimal stopping framework in the spirit of Rust (1987) to estimate demand systems in a durable goods context. Nesting the BLP-style algorithm within a Rust-Nested Fixed-point (NFXP) algorithm turns out to be challenging from a computational point of view. Clearly, allowing for a flexible

---

3 Diewert (1974) flexibility involves the ability to match market shares and price elasticities at a point in price and income space while McFadden and Train (2000) show that the RC-MNL demand model may make arbitrary continuous functions of prices and indeed other product characteristics.

4 The reason is that in most actually estimated models, the conditional indirect utilities are assumed linear and separable in product characteristics. This parametric structure places strong restrictions on the potential substitution patterns even in the presence of a considerable number of random coefficients.

5 We note that some authors (following Bhat, 1997) have suggested a “middle ground” wherein base model parameters that control substitution patterns are allowed to be described as parametric functions of observed consumer heterogeneity. More specifically, the idea in Bhat (1997) is a simple one — to allow the Nested Logit’s nesting parameter \( \sigma \) to vary with a consumer’s characteristics \( z_i \), so that we model \( \sigma_i = F(\alpha + \gamma z_i) \) where now \((\alpha, \gamma)\) are parameters to be estimated and \(F()\) is a transformation function which must be chosen to ensure \( 0 < \sigma_i < 1 \) as required for the Nested Logit model. Subsequent papers in this tradition include the Heterogenous Generalized Nested Logit model (Koppelman & Sethi, 2005) who build upon Swait and Adamowicz (2001). Such models, designed for use with individual-level data sets, have the advantage of using a base-line GEV model allowing for closed form probabilities of purchasing a given option, while allowing for observed heterogeneity across individuals to affect purchase probabilities and in particular, the covariance between utilities of different options. Hess at al. (2010) examine a general and flexible form of covariance in the mixed-covariate generalized extreme value (MCGEV) by accommodating random covariance heterogeneity in addition to deterministic covariance heterogeneity, although introducing random coefficients means losing the closed form for the choice probabilities. We note that the FC-MNL model proposed here could similarly be adopted as a base model in the tradition of this literature.
discrete choice model which does not require simulation will greatly simplify the implementation of such models.

The third significant branch of the discrete choice literature has attempted to estimate semiparametric and nonparametric discrete choice models, most notably in the seminal contributions from Matzkin (1992, 1993), discussed, for example, by Horowitz and Savin (2001). The approach we take in this paper in some ways contrasts with the authors in this tradition and in other ways may complement their work. The case for a nonparametric analysis is easy to make in principle. However, as is well understood, the kinds of data sets we typically have in industrial organization or competition policy would not usually be sufficient to estimate fully nonparametric models. Briesch, Chintaguntam and Matzkin (2002) provide an application of such techniques allowing some components of the model to be either semi- or nonparametric, and compare the results with MNL. Unfortunately, the estimation precision of nonparametric models decreases rapidly as the number of explanatory variables increases, the so-called curse of dimensionality. The FC-MNL model is fully parametric, but it could also be used in the spirit of this branch of the literature since it could (in principle) provide the basis for a semiparametric model wherein only the distribution of consumer tastes in the population is assumed parametric and provided below. This paper’s aim of improving model flexibility is very much in the spirit of this branch of the discrete choice literature – in the sense that we are aiming to reduce the constraints being imposed on the data by the assumptions embedded in the model.

The paper proceeds as follows. First, we introduce notation and review McFadden’s results on the MEV class of models, emphasizing that this class of models provide analytic expressions for market shares and thereby avoids simulation. In Section 3, we formally describe the FC-MNL model and demonstrate (i) that it is a member of the MEV class of models under specific parameter restrictions and (ii) that it provides a flexible functional form for a well-defined class of DGP’s. Section 4 discusses identification and estimation. Section 5 provides a Monte-Carlo study. Section 6 illustrates the use of the model in estimating the demand for automobiles in Italy. Finally, we conclude. Proofs are relegated to Appendix A.

2. The MEV class of models

Consider a discrete choice demand model where consumers, indexed by $i$, each choose between an outside option (option 0) and $J$ inside options indexed from $j \in \{0, 1, \ldots, J\} = J$. Each consumer is assumed to choose the option which provides the greatest conditional indirect utility, $\max_{j \in J} v_{ij}(w_j, y - p_j; \theta_i)$ where $w_j = (x_j, \xi_j)$ denotes the observed and unobserved characteristics of product $j$ (respectively denoted $x_j$ and $\xi_j$ following BLP), $p_j$ denotes good $j$’s price and $p_0$ is the price of the outside good, which we normalize to 1, so that the outside good is assumed to be in monetary units and $\theta_i$ parameterizes the conditional indirect utility functions. We further suppose that $v_{ij}(w_j, y - p_j; \theta_i) = \delta_j(w_j, y - p_j; \theta_i) + \epsilon_{ij}$ with $\epsilon_{ij}$ can be an option- and an individual-specific idiosyncratic component of conditional indirect utility.

In the familiar MNL model, $\epsilon_{ij}$ is assumed as i.i.d. across individuals and products with a type I extreme value distribution so that its cumulative distribution function has the form $F(\epsilon_{ij}) = \exp\{-\exp\{-\epsilon_{ij}\}\}$. In that case, the aggregate market shares of each of the $J$ choices are well known to have an analytic form: $s_j = \frac{e^{\delta_j(w_j, y - p_j; \theta_i)}}{\sum_{k=0}^J e^{\delta_k(w_k, y - p_k; \theta_i)}}$. It will sometimes be useful to define the functions $r_j = e^{\delta_j(w_j, y - p_j; \theta_i)}$ so that we can write $s_j = r_j / \sum_{k=0}^J r_k$. Next we state the result due to McFadden (1981), which generalizes this result substantively to establish the MEV class of models. We opted for the version of the theorem with the generalization to homogeneous degree $\tau$ functions due to Ben-Akiva and François (1983) for reasons that will become clear below.

---

6 More formally, we could follow McFadden (1981) and argue that the conditional indirect utility specification here can in turn be motivated by an underlying utility function maximized subject to a budget constraint. See, for example, the derivation provided in Davis and Garces (2009), pages 462–467.
Theorem 1. (McFadden, 1978; Ben-Akiva and François, 1983) Suppose \( H(r_0, r_1, \ldots, r_J; \theta_2, \xi) \) is a non-negative, homogeneous of degree \( \tau > 0 \), function of \((r_0, r_1, \ldots, r_J) \geq 0\), where \( \theta_2 \) is a vector of parameters and \( \xi = \{0, 1, 2, \ldots, J\} \). Suppose (i) \( \lim_{r_j \to +\infty} H(r_0, r_1, \ldots, r_J; \theta_2, \xi) = +\infty \) for \( j \in \xi \) (ii) that for any distinct \((i_1, \ldots, i_k)\) from \( \xi \), \( \frac{\partial^k H(r_0, r_1, \ldots, r_J; \xi)}{\partial r_{i_1} \ldots \partial r_{i_k}} \) is non-negative if \( k \) is odd and nonpositive if \( k \) is even. Then if the joint distribution of consumer heterogeneity has the form \( F(\varepsilon_0, \ldots, \varepsilon_J; \theta_2, \xi) = \exp[-H(e^{-\varepsilon_0}, \ldots, e^{-\varepsilon_J}; \theta_2, \xi)] \), and consumers solve \( \max_{j=0,1,\ldots,J} \delta_j(w_j, y - p_j; \theta_1) + \varepsilon_{ij} \) then market shares have the analytic form:

\[
s_j = \frac{r_j^{\delta_j(w_j, y - p_j; \theta_1)} \varepsilon_{ij}}{\sum_{l=0}^{J} r_l^{\delta_l(w_l, y - p_l; \theta_1)} \varepsilon_{il}}, \quad j \in \xi,
\]

evaluated at \( r_j = e^{\delta_j(w_j, y - p_j; \theta_1)} \) for all \( j \in \xi \), and where \( r = (r_0, r_1, \ldots, r_J) \).

This important theorem describes a simple way of generating structural discrete choice models with analytic market share functions. The structural model consists of two components. First, each consumer is assumed to solve a conditional indirect utility problem \( \max_{j=0,1,\ldots,J} \delta_j(w_j, y - p_j; \theta_1) + \varepsilon_{ij} \) – which can in turn be motivated by an underlying utility function subject to a budget constraint – and second consumer heterogeneity is assumed to follow a particular distribution across the population of consumers, \( F(\varepsilon_0, \ldots, \varepsilon_J; \theta_2, \xi) = \exp[-H(e^{-\varepsilon_0}, \ldots, e^{-\varepsilon_J}; \theta_2, \xi)] \).

The MEV class of models relaxes the independence from irrelevant alternatives (IIA) property of the MNL model by relaxing the independence assumption between the error terms of alternatives while maintaining a closed-form expression for the market shares. Several specific MEV structures have been formulated in an attempt to increase flexibility of the models and applied within the MEV class. The first class of models include the classic Nested Multinomial Logit (NMNL) model (Williams, 1977; McFadden, 1978; Daly & Zachary, 1978) and subsequent modifications (see, e.g., the nest-specific parameters used by Verboven & Brenkers, 2002). Subsequent authors noted that the NMNL’s partitioning of the lists of products into separate non-overlapping nests of products seriously restricts substitution patterns. As a result, authors have developed models where the product nests can overlap; see the Cross-Nested Logit (CNL) model (Vovsha, 1997; Ben-Akiva & Bierlaire, 1999; Papola, 2004; Bielaire, 2006) and the Generalized Nested Logit (GNL) model (Wen & Koppelman, 2001) which Bielaire (2006) shows are equivalent. The empirical industrial organization literature has also contributed to this branch of literature in the form of the Product Differentiation Logit (PDL) model (Bresnahan et al., 1997). With the notable exception of the PDL, authors writing in this literature are primarily writing in the transportation demand literature and therefore have also adopted terminology around “Networks” to motivate and describe classes of MEV type models. In particular, the Paired Combinatorial Logit (PCL) model (Chu, 1989; Koppelman & Wen, 2000) and the Network GEV (NetGEV) model (Bierlaire, 2002; Daly & Bierlaire, 2006). In case it is helpful for the reader, Table 1 shows the MEV generating functions for a variety of models including those proposed in both transportation and industrial organization literatures.

This paper follows the tradition of these papers and indeed the model we propose is somewhat of a “hybrid” variety, with elements of both the PCL and the Cross-Nested Logit classes of models but has been developed to have some nice properties. Of course, understanding the right base models to take to data is important for applied researchers and which one to use depends on the properties a researcher wishes her model to have. We motivate the FC-MNL specification with very deliberate aims in mind, in particular, we would like (i) to generate a model which is known to be Diewert (1974) flexible while introducing a minimal number of parameters to estimate and (ii) to have a model specification which is easy to work with computationally. In the latter regard

---

7 Bielaire (2006) also shows that the Ordered GEV model (Small, 1987) is a special case of the CNL. See also the Multinomial Logit-Ordered GEV model (Bhat, 1998), the ordered GEV-nested logit model (Whelan et al., 2002), and also more recently the application to the car market of the Ordered Nested GEV model (Grigolon, 2012).
<table>
<thead>
<tr>
<th>Model Type</th>
<th>Author(s)</th>
<th>Description of (H) function</th>
<th>Parameters in the distribution of tastes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Multinomial Logit (MNL)</strong></td>
<td>McFadden (1978)</td>
<td>(H(r_0, r_1, \ldots, r_J; \mathcal{I}) = \sum_{j \in \mathcal{I}} r_j)</td>
<td>None</td>
</tr>
<tr>
<td><strong>One Level Nested Multinomial Logit (NMNL)</strong></td>
<td>McFadden (1978)</td>
<td>(H(r_0, r_1, \ldots, r_J; \theta, \mathcal{I}) = \sum_{g=1}^{G} \left( \sum_{j \in \mathcal{I}_g} r_j \right)^{1-\sigma_g})</td>
<td>(\sigma_g) for (g = 1, \ldots, G) (G parameters)</td>
</tr>
<tr>
<td><strong>Two Level Nested Multinomial Logit with nest specific parameters</strong></td>
<td>Verboven and Brenkers (2002)</td>
<td>(H(r_0, r_1, \ldots, r_J; \theta, \mathcal{I}) = \sum_{g=1}^{G} \left( \sum_{h=1}^{H} \left( \sum_{j \in \mathcal{I}<em>{gh}} r_j \right)^{1-\sigma</em>{gh}} \right)^{1-\sigma_g})</td>
<td>(\sigma_{gh}, \sigma_g) for (g = 1, \ldots, G) and (h = 1, \ldots, H). (G+G*H) parameters. (\mathcal{I}_{gh}) partition (\mathcal{I}).</td>
</tr>
<tr>
<td><strong>Cross Nested Logit Model</strong></td>
<td>Vovsha (1997)</td>
<td>(H(r_0, r_1, \ldots, r_J; \theta, \mathcal{I}) = \sum_{g=1}^{G} \left( \sum_{j \in \mathcal{I}<em>g} r_j \right)^{a_g}) where (a_g \geq 0, \sum</em>{g=1}^{G} a_g = 1)</td>
<td>(0 \leq a_{jg} \leq 1) for all (j, g); (0 \leq \sum_{g} a_{jg} \leq 1) and (\sum_{g} a_{jg} = 1). (\mathcal{I}_g) do not partition (\mathcal{I}).</td>
</tr>
<tr>
<td><strong>Product-Differentiation model (PD)</strong></td>
<td>Bresnahan, Stern, and Trajtenberg (1997)</td>
<td>(H(r_0, r_1, \ldots, r_J; \theta, \mathcal{I}) = \sum_{g=1}^{G} \left( \sum_{j \in \mathcal{I}<em>{gh}} r_j \right)^{a_g}) where (\sum</em>{g} a_g = 1)</td>
<td>(a_g, \sigma_g) for (g = 1, \ldots, G) (2G parameters). (\mathcal{I}_{gh}) do not partition (\mathcal{I}).</td>
</tr>
<tr>
<td><strong>Paired Combinatorial Logit model (PCL)</strong></td>
<td>Chu (1989)</td>
<td>(H(r_0, r_1, \ldots, r_J; \theta, \mathcal{I}) = \sum_{i=0}^{J-1} \sum_{j=1}^{J} \left( r_i^{1/\sigma_{ij}} + r_j^{1/\sigma_{ij}} \right)^{1-\sigma_{ij}}) where (\sum_{ij} a_{ij} = 1)</td>
<td>(0 \leq \sigma_{ij} &lt; 1) for all (ij) pairs</td>
</tr>
<tr>
<td><strong>Generalized Nested Logit (GNL) or Cross Nested Logit model (CNL)</strong></td>
<td>Wen &amp; Koppelman (2001) or Ben-Akiva and Bierlaire (1999); see Bierlaire (2006)</td>
<td>(H(r_0, r_1, \ldots, r_J; \theta, \mathcal{I}) = \sum_{g=1}^{G} \left( \sum_{j \in \mathcal{I}_g} r_j \right)^{a_g})</td>
<td>(a_{jg} \geq 0, \sum_{g} a_{jg} &gt; 0, \sum_{g} a_{jg} = e^{\gamma}) where (\gamma = -0.5772) is Euler’s constant and (\sigma_{jg} \geq \sigma &gt; 0). (\mathcal{I}_g) do not partition (\mathcal{I}).</td>
</tr>
</tbody>
</table>

Notes: In the one-level nested logit model, \(J_g\) denotes the set of products in the \(g^{th}\) nest. Similarly, in the two level nested logit model, \(\mathcal{I}_{gh}\) denotes the set of products that are in top level nest \(g\) and in second level nest \(h\).
we will, for example, show in Appendix B that our model specification can be easily programmed using matrix algebra.

It is perhaps surprising that only a relatively small number of possible members of the class of MEV models have previously been actively taken to data. In this paper, we develop some properties of this general class of models and also propose using a “new” \( H(r) \) function, one that provides researchers with demand systems with some desirable properties. Along the way, we discuss some inherent limitations of the MEV class of models, and present a model which will be a member of the MEV class of discrete choice demand models under well-defined parametric restrictions which may or may not be imposed in estimation.

In many policy applications, including merger simulation, the key object of interest is the matrix of own- and cross-price elasticities. Given the analytic market share function for the wide class of MEV models, it is possible to compute analytic expressions for the own- and cross-price elasticities of demand. Doing so provides an insight into at least one useful property of the function \( H \). We begin by providing an expression for the matrix of own- and cross-price elasticities that are generated by the MEV class of models. Define \( H_j(r; \theta_2, \Theta) = \frac{\partial H(r; \phi_2, \Theta)}{\partial \theta_j} \) and \( H_{jk}(r; \theta_2, \Theta) = \frac{\partial H_j(r; \phi_2, \Theta)}{\partial \theta_k} \).

**Corollary 1.** (i) Own- and cross-price elasticities of demand in the class of MEV models have the following analytic form:

\[
\frac{\partial \ln s_j}{\partial \ln p_k} = \left( I(k = j) + \frac{r_k H_{kk}(r; \theta_2, \Theta)}{H_j(r; \theta_2, \Theta)} - \tau s_k(r) \right) \frac{\partial \delta_k}{\partial \ln p_k}
\]

for \( j \in \mathcal{I} = \{0, 1, 2, \ldots, J\}, k \in \mathcal{I} / 0 = \{1, 2, \ldots, J\} \) (i.e., \( A/B \) denotes the set ‘A not B’) and with each element evaluated at \( r_j = e^{(\omega_j + y_j - p_j \phi_2) + \epsilon_j} \) for \( j \in \mathcal{I} \).

(ii) Income elasticities of demand in the class of MEV models have the form:

\[
\frac{\partial \ln s_j}{\partial \ln y} = \frac{\partial \delta_j}{\partial \ln y} + \frac{1}{H_j(r; \theta_2, \Theta)} \sum_{l=0}^{J} r_l H_{jl}(r; \theta_2, \Theta) \frac{\partial \delta_l}{\partial \ln y} - \tau \sum_{l=0}^{J} s_l(r) \frac{\partial \delta_l}{\partial \ln y}
\]

for \( j \in \mathcal{I} \).

Using the expressions provided in Corollary 1, it is possible to see why the MNL model, wherein \( H(r_1, \ldots, r_J; \Theta) = \sum_{j=1}^{J} r_j \) suffers from severely restricted predicted cross-price elasticities. In particular, note that the \( H \) function for the MNL model has all of its second cross-derivatives as zero. As a result, the MNL model provides a matrix of own- and cross-price elasticities that for \( k > 0 \), collapses to the well-known and often criticized expression

\[
\frac{\partial \ln s_j}{\partial \ln p_k} = (I(j = k) - s_k(r)) \frac{\partial \delta_k}{\partial \ln p_k}
\]

This, for example, is entirely independent of \( j \) for any \( j \neq k \), so that the MNL model predicts that an increase in the price of good \( k \) will result in an identical increase in demand for every other product. Clearly, the model’s predicted cross-price elasticities generically will not reflect the true level of substitutability between the products in the market. Rather, the model imposes a great deal of structure on the estimated demand system.

An implication of Corollary 1 is that, of the MEV models considered in the literature, very few (e.g., Bresnahan, Stern, & Trajtenberg, 1997; Verboven & Brenkers, 2002; Wen & Koppelman, 2001; Ben-Akiva & Bielaire, 1999) can potentially avoid the restrictiveness of the MNL model in the substitution patterns between goods. The reason is that these models allow the second cross derivatives between any pair of goods is potentially nonzero for \( j, k \).

On the other hand, provided we make the last term parametric \( \frac{\partial \delta_k}{\partial y_k} = \alpha_k \), the MNL can in principle match arbitrary own-price elasticities of demand. To state this first very
limited – result, we use the notation that \( s^*_k \) is the true (observed) vector of market shares \( \left( \frac{\partial \ln s^*_k}{\partial \ln p_j} \right) \) is the true matrix of own- and cross-price elasticities. We also use the assumption, which will be proved below in Lemma 3, that there exists a vector \( r \) such that the model can exactly match the vector of true market shares, \( s_k(r) = s^*_k \) for all \( k = 0, 1, \ldots, J \).

**Lemma 1.** The MNL model with option specific price parameters can match any vector of own-price elasticities provided \( 0 < s^*_j < 1 \) and provided there exists an \( r \) (or following BLP an underlying vector of unobserved product characteristics \( \xi = (\xi_1, \ldots, \xi_J) \)) such that \( s_k(r) = s^*_k \), with \( r_j = \exp((s_j - s_k)^*) \).

Lemma 1 provides a simple demonstration that a suitably parameterized version of the MNL can match arbitrary own-price elasticities. Unfortunately, such a parameterization does not help to simultaneously match the cross-price elasticities unless the Data Generating Process (DGP) satisfies very stringent conditions. To see that, notice that such an MNL model imposes the requirement that \( \frac{\partial \ln s^*_j}{\partial \ln p_k} = -s^*_j s^*_k \alpha_k \) for all \( j \neq k \), while setting \( \alpha_k = \frac{1}{1-s^*_j} \left( \frac{\partial \ln s^*_j}{\partial \ln q_k} \right)^* \). These relationships mean we can only fit both if the DGP satisfies the restriction that \( \left( \frac{\partial \ln s^*_j}{\partial \ln q_k} \right)^* = -\frac{s^*_j}{(1-s^*_j)} \left( \frac{\partial \ln s^*_j}{\partial \ln q_k} \right)^* \) for all \( j \neq k \), a set of restrictions for which there is no obvious theoretical motivation. For data sets which involve small market shares \( s^*_j \) and \( s^*_k \), these restrictions, implicitly imposed by the model, ensure that the estimated cross-price elasticities will be small relative to own-price elasticities, a highly undesirable state of affairs for most applications. On the other hand, such a model would be very useful to a firm who wanted to have its proposed merger approved; the restriction would ensure that the firms appeared to have low cross-price elasticities and hence were producing products in distinct markets. The standard MNL model will impose this restriction and more. Once we attempt to match cross-price elasticities as well, we find an alternative approach (to using option specific coefficients) that has some advantages. It is to that case which we now turn.

### 3. A flexible discrete choice demand model

In this section, we provide a sequence of results concerning a member of the MEV class of models, one that has some considerable ability to match own- and cross-price elasticities of demand. Specifically, we propose developing a discrete choice demand model based around the member of the class of MEV models with \( H \) function:

\[
H(r; \theta_2, \Xi) = \sum_{j \in \Xi} \sum_{k \neq j} \sum_{b_{jk}} \left( \frac{r^j_j + r^k_k}{2} \right)^{\tau \sigma} + \sum_{j=0}^{b_{jk}} r^j_j
\]

Intuitively, following the discussion above, the central element motivating this choice of \( H \) function is that the matrix of cross derivatives of \( H_{jk}(r) \) can be varied greatly by varying the parameters \( b_{jk} \). In this member of the MEV class of models, the parameters \( \theta_2 = (B, \sigma, \tau) \) are taste parameters controlling the distribution of tastes in the population, where \( B \) is the \((J+1) \times (J+1) \) matrix with \( jk \)th element \( b_{jk} \). The parameter \( b_{jk}^\tau \) is indexed by the choice set \( \Xi \) for reasons that will become apparent.

As discussed below, a nice feature of the FC-MNL model is that the parameter controlling the substitution pattern between goods \( j \) and \( k \) will be \( b_{jk} \) while \( b_{jk}^\tau \) will control the own-price elasticity for product \( j \). This model nests the standard MNL model. To see that, simply set \( b_{jk} = 0 \) for all \( j \neq k, b_{jk} = 1 \) and \( \tau = 1 \). The FC-MNL can also be motivated as a specific instance of a very general version of the GNL/CNL model written down by Wen and Koppelman (2001) and Ben-Akiva and Bierlaire (1999). To see the mapping, recall that the CNL has a general MEV generating function

\[
H(r_0, r_1, \ldots, r_J; \theta_2, \Xi) = \sum_{g \in \Xi} \left( \sum_{j \in \Xi} a_{jk} r^j_k \right)^{\tau \sigma} \] which, if we (i) put every pair of products into a nest so that nest \( g \) is made up of products \( \Xi_g = \{j, k\} \) for all combinations of products \( j \neq k \) and
in addition, (ii) put every product in its own nest so that we also have $\mathcal{J}^g_j = \{j\}$ for all products $j = 0, \ldots, J$; and (iii) set $\mu = \tau$, $\mu_g = \frac{1}{\tau}$ for all $g$. Doing so allows us to write the expression $(a_{jk} r_j^{\mu_1} + a_{kl} r_l^{\mu_2})^{\tau} / \tau$ for each $j \neq k$ nest since $(a_{jk} r_j^{\mu_1} + a_{kl} r_l^{\mu_2})^{\tau} = (2 a_{jk})^{\tau} / \tau$. That is, we can set $b_{jk} = (2 a_{jk})^{\tau} / \tau$ for $j \neq k$ and $b_{jj} = (a_{jj})^{\tau} / \tau$ for the parameters on the nests consisting on each individual product. To our knowledge, no one has ever previously estimated such a general version of the CNL model.

We now state a lemma which establishes the parameter restrictions under which the FC-MNL model can generate members of the class of MEV models.

Lemma 2. The function $H(r; \theta, \mathcal{J})$ as defined above can be used to generate members of the class of MEV models if (i) $b_{jk} \geq 0$, for all $j, k = 0, \ldots, J$, (ii) for each $j$ there is a good indexed $m_j \in \mathcal{J}$ such that $b_{jm_j} > 0$, (iii) $\tau > 0$, $\sigma > 0$ and $\tau \sigma \leq 1$. Any member of this class of models with an asymmetric $B$ matrix is observationally equivalent to a member of the MEV class of models with a symmetric $B^*$ matrix where $b^*_{jk} = b^*_{kj} = (b_{jk} + b_{kj})/2$.

The first part of the lemma is proven in Appendix A but also follows implicitly from Bierlaire (2006) who proves the result for a general CNL model. The second part of the lemma is both important and easy to see since $\frac{\partial H(r; \theta, \mathcal{J})}{\partial g}$ only depends on the averages of the off-diagonal parameters, $(b_{jk} + b_{kj})/2$. As a result, it will sometimes be interesting to estimate a model which is more general than the MEV based model described thus far since doing so will, for example, allow us to test the restrictions required to support the structural MEV interpretation of our parameter estimates. Note that when symmetry is not imposed, $b_{jk} \neq b_{kj}$, the model with market share function, $s_j(r) = r_j N_j(r)/\sum_{l=0}^{J} r_l N_l(r)$ and $N_j(r; \theta) = \sum_{k \neq j} b_{jk} (r_j/r_k)^{\tau-1} r_j^{\tau-1} + \tau b_{jj} r_j^{\tau-1}$ is more general than could be derived from an MEV model because it does not require that each of the functions $N_j(r)$ are derived from a single common function according to the relationship $N_j(r) = \frac{\partial H(r; \theta)}{\partial g}$. To see that, notice that since $H(r; \theta)$ is homogeneous of degree $\tau$, $\tau H(r; \theta) = \sum_{k=0}^{J} r_k \frac{\partial H(r; \theta)}{\partial r_k}$ and so the share function becomes one generated by an MEV model wherein $s_j(r) = r_j H(r)/\tau$. This market share model will therefore be a member of the MEV class of models only when the parametric symmetry restrictions $b_{jk} = b_{kj}$ all $j \neq k$ are imposed. We will discuss the implications of these symmetry restrictions more extensively below. For now, we note in particular that Lemma 2 does not imply that the more general model we outline below is observationally equivalent to a symmetric member of the MEV class of models.

In the next subsection we establish sufficient conditions on the parameters such that there is a value of $r$ which will equate predicted market and actual market shares for all products, $s_j(r^*) = s_j^*$ for $j = 0, \ldots, J$. We then progress to demonstrate that the model can, in addition to market shares, also match any matrix of own- and cross-price elasticities of demand for a well-defined class of DGP. In particular, we show that the MEV restriction that $b_{jk} = b_{kj}$ restricts the flexibility result only to apply to DGP’s which satisfy the familiar “symmetry” restriction, $(\frac{\partial s_j}{\partial p_k})^* = (\frac{\partial s_k}{\partial p_j})^*$ for all j,k. We also show that the current generation of models such as the member of the class of RC-MNL models popularized in the seminal contributions from Berry (1994) and BLP (1995) implicitly also frequently impose symmetry restrictions. In contrast, we will see that the symmetry restrictions need not necessarily be imposed on the general market share function considered here.

Flexibility. We consider the flexibility of the proposed demand system in two steps. First, we show that using this class of models we can always solve for the levels of mean utilities⁶

⁶ Notice that the term “mean utility” (of product $j$) to denote $\delta_j$ is misleading in our context because it is not difficult to show that the above discussed MEV distribution implies $E[e_{ij}] = (\ln(\frac{1}{2})^{\tau-1} \sum_{k \neq j} b_{jk} + b_{kj}) + \gamma) / \tau$ where $\gamma$ is the
associated with each option that make predicted and actual market shares equal for all the goods. More precisely, we can show that we can always solve for the monotonic transformation of the mean utilities \( r_j = e^{s_j(x_j, y, p_j; \theta_i)} \) \( j \in \Im \). This result also provides the first step in establishing flexibility results about the demand system since it ensures the model can always match the vector of observed market shares, one requirement for a model to be a Dievert (1974) flexible functional form. We then proceed to provide a result establishing the models’ ability to also arbitrarily match any own- and cross-price elasticities.

Notice that given any value for \( r_j^* \), and any fixed values of \((x_j, y, p_j; \theta_i)\), the equation \( r_j = e^{s_j(x_j, y, p_j; \theta_i)} \) implicitly defines a solution \( s_j^* \) which exists and is recoverable provided \( \frac{\partial s_j(x_j, y, p_j; \theta_i)}{\partial x_j} > 0 \). Thus we may entirely equivalently think of picking the \( r_j^* \) or \( s_j^* \) amount to finding the level of utility that must be associated with product \( j \) in order to explain its observed market share, where \( s_j^* = (e_{j_0}^*, e_{j_1}^*, \ldots, e_{j_\ell}^*) \). An implication of Lemma 3 below therefore will be that we can use an estimation strategy based on the suggestion from Nakanishi and Cooper (1974), Berry (1994), and BLP (1995); see section 4. We first state a general result and then apply it to the particular demand system under consideration wherein

\[
s_j(r) = \frac{r_j N_j(r)}{\sum_{l=0}^\ell r_l N_l(r)} \quad \text{with} \quad N_j(r; \theta_2) = \tau \sum_{k \neq j} b_{jk} \left( \frac{r_j^1 + r_k^1}{2} \right)^{\tau \sigma - 1} r_j^{\frac{1}{\sigma} - 1} + \tau b_{j}^{1} r_j^{\frac{1}{\sigma} - 1}.
\]

Lemma 3. Let \( \Im = \{0, 1, \ldots, J\} \) be the set of products and \( \Im^+ = \{j|s_{j\text{obs}} > 0, j \in \Im\} \) be the set of products with strictly positive observed market shares. Let \( f(r; \Im) \) be continuous, differentiable, and homogenous of degree \( \tau > 0 \) in \( r \in \Re^J_+ \) and for any \( \Im^+ \subseteq \Im \) and any \( r_{\Im^+} \in \Re^J_{+}\), define \( f(r_{\Im^+}; \Im^+) \equiv f(r_{\Im^+}, 0, \ldots, 0; \Im) \). Further suppose \( f(r; \Im) \) has the following properties: (i) if \( r_j = 0 \) then \( f_j(r; \Im) = 0 \), (ii) \( \lim_{j \rightarrow \infty} f_j(r; \Im) \geq 1 \) for all \( j \in \Im \), (iii) for each \( j \in \Im^+ \), and any \( r_j > 0 \), \( f_j(r; \Im) > 0 \) and (iv) for each \( k \in \Im^+ / f_j \), \( \frac{\partial f_j(r_{\Im^+}; \Im^+)}{\partial r_k} \leq 0 \). Then there exists a finite vector of \( r \) ’s that solve the \( J+1 \) vector of equations \( s_{j\text{obs}} = f_j(r; \Im) \) for \( j \in \Im \). If \( s_{j\text{obs}} = 0 \) then a solution sets \( r_j^* = 0 \) for \( j \in \Im \). Moreover, the solution to the subset of equations \( s_{j\text{obs}} = f_j(r_{\Im^+}; \Im^+) \) with \( j \in \Im^+ \) is unique.

We now provide a result which allows this lemma to be applied to the particular model under consideration in this paper.

Corollary 2. Let \( s_j(r; \theta_2, \Im) = \frac{r_j N_j(r; \theta_2, \Im)}{\sum_{l=0}^\ell r_l N_l(r; \theta_2, \Im)} \) and let \( s_{j\text{obs}} < 1 \) for all \( j \). If (i) \( b_{jk} \geq 0 \), for all \( j,k = 0, \ldots, J \), (ii) \( b_{j} > 0 \) for each \( j \in \Im^+ \) (iii) \( \sigma > 0, \tau > 0 \) and \( \tau \sigma \leq 1 \). Then, there exists a finite vector of \( r \) ’s that solve the equations \( s_{j\text{obs}} = s_j(r; \theta_2, \Im) \) for \( j \in \Im \). If \( j \in \Im / \Im^+ \), so \( s_{j\text{obs}} = 0 \), then a solution sets \( r_j^* = 0 \) while if \( j \in \Im^+ \), so \( s_{j\text{obs}} > 0 \), the solution sets \( r_j^* > 0 \). Moreover, the solution \( r_{\Im^+}^* \) to the subset of equations \( s_{j\text{obs}} = s_j(r_{\Im^+}; \theta_2, \Im^+) \) (i.e., with \( s_{j\text{obs}} > 0 \)) is unique.

Notice that starting from a general framework where consumer preferences are represented by a random utility model, Berry and Haile (forthcoming) and Berry, Gandhi, and Haile (2013) prove the invertibility of a demand system under mild conditions: (i) the index restriction and (ii) the connected substitute assumptions.10 Lemma 3 and Corollary 2 prove the uniqueness results in

---

10 The index restriction requires that \( x_{ji}^{(i)} \) (where \( x_{ji}^{(i)} \) is at least an observed exogenous characteristic of product \( j \)) and \( \xi_{ji} \) affect the distribution of utilities only through an (potentially nonlinear) index, for example, \( \delta = x_{ji}^{(i)} \beta + \xi_{ji} \) with

© RAND 2014.
the present context also when the demand system is not necessarily generated by random utility models, i.e., $b_{jk} \neq b_{kj}$.\footnote{Proposition 2 in appendix B provides an alternative approach to establish the invertibility of the demand system.}

An important aspect of the result provided in Lemma 3 and Corollary 2 is that we need not impose the symmetry conditions, $b_{jk} = b_{kj}$, in order to compute the model and, as a result, we need not impose the symmetry restriction when estimating the market share equation. That means that we need not impose such assumptions \emph{a priori} on the data but rather can test whether the data are consistent with it by testing the restriction that the B parameters satisfy the symmetry restriction using conventional tests such as LM or Wald tests.\footnote{Typically, if symmetry is rejected when estimating say an almost ideal demand system (AIDS) or Translog model we do not impose the symmetry restriction required for the underlying theory but rather prefer to use the estimated demand system – for say merger analysis – even in the absence of an underlying model of utility and consumer heterogeneity.} A referee raised the interesting question of whether, if a researcher finds that their data rejects the symmetry restriction, she would effectively be in the same position as if she had just estimated a computationally simpler model, say a simple linear regression of quantities on prices. Clearly, as in the continuous choice demand literature, the asymmetric model has no underlying welfare foundation, and so we agree in that respect with the point being made. However, we are of the view that the main approach taken by the literature involves estimating discrete choice models which imposes symmetry \emph{a priori}, rather than estimating computationally simple models without an underlying utility motivation. In that respect, it is a distinct advantage to be able to test whether the symmetry restrictions implicit in many models are, in truth, rejected by the data. Authors will not know at the start whether symmetry is rejected, but the fact that the model we propose remains estimable means that the effort undertaken to get to that point will not immediately be wasted since not all applications (e.g., merger simulation) require a utility theory. Moreover, the process of testing may provide indications of the pattern in the data that is leading to a rejection of symmetry and therefore possibly suggest ways to accommodate such asymmetries, perhaps by introducing some elements of consumer heterogeneity.\footnote{Consumer heterogeneity will be helpful because aggregate demand curves need not satisfy symmetry. We do not pursue the option in detail here, but we hypothesize that an aggregate demand model based on two distinct types of FC-MNL consumers, each with distinct but symmetric matrices of parameters B and B\textsuperscript{*} may provide a flexible but relatively easy-to-estimate option for generating asymmetric models.}

Next we state the proposition which establishes that the model we propose in this paper is a flexible functional form within a well-defined class of DGP's and therefore in the sense of Diewert (1974).

\textbf{Proposition 1. (Flexibility of the discrete choice model)} Consider the MEV model $H(r; \theta_j) = \sum_{j=0}^{J} \sum_{k \neq j} b_{jk} \left( \frac{r_j}{2} \right)^{\alpha} + \sum_{j=0}^{J} b_{jj} r_j$ where $r_j = e^{\delta_j(s_j, \xi_j, y, p_j, \theta_j)}$ with $\frac{\partial \delta_j(s_j, \xi_j, y, p_j, \theta_j)}{\partial \xi_j} > 0$ for $j \in \mathbb{N}$ with parameters $\sigma, \tau$ taking some fixed values with $0 < \sigma < 1$, $\tau > 1$ and $\sigma \tau < 1$ and where: (i) $\frac{\partial \delta_j}{\partial p_j} = \alpha < 0$, (ii) $b_{jk} \geq 0$, for all $j, k \in \mathbb{N}$ with $b_{jj} > 0$ for all $j \in \mathbb{N}^+$, and (iii) $b_{jk} = b_{kj}$. There exists a $(\xi, B, \alpha)$ which satisfy properties (i)-(iii) above such that this MEV model can match any vector of market shares and any matrix of own- and cross-price elasticities, provided that in the DGP market shares are positive and goods are (a) demand substitutes for one another so that $\frac{\partial s_j}{\partial p_k} > 0$ for all $k \neq j$ and own-price sensitive so $\left( \frac{\partial s_j}{\partial p_j} \right)^* < 0$, and (b) satisfy the Slutsky Symmetry restriction, $\left( \frac{\partial s_j}{\partial p_k} \right)^* = \left( \frac{\partial s_j}{\partial p_j} \right)^*$. When the essential requirement that the index that is strictly monotonic in $\xi_j$, which in turn implies $\partial \delta_j / \partial \xi_j > 0$. The connected substitutes structure requires two conditions. First, goods must be weak substitutes in the characteristics and in the index $\delta$. Second, there must be sufficient strict substitution among the goods to require treating them all in one demand system. Under the symmetry restriction of the B matrix, i.e., $b_{jk} = b_{kj}$, theorem 1 in Berry and Haile (forthcoming) proves the uniqueness of the vector $\delta$ that rationalized the market shares in present context.
This proposition establishes that for any fixed $\tau > 1$ and $0 < \sigma < 1$ with $\tau \sigma < 1$ there are values of $(\xi, B, \alpha)$, so that the model can match an arbitrary vector of positive market shares and an arbitrary matrix of symmetric own- and cross-price elasticities. However, this result holds with a caveat. While the theoretical result in proposition 1 is valid in case we study a single market or time period, the model will not be able to match different arbitrary matrices of own- and cross-price elasticities if multiple markets or time periods are considered. Naturally, the same caveat applies to the analogous flexibility results provided by the continuous choice demand literature. We further study this limitation with a Monte-Carlo study in section 5.2.

Finally, we show that all members of this MEV class of models will impose symmetry-like restrictions which should not be expected to hold in general (Diewert, 1977, 1980). Davis (2006b) similarly shows that a wide variety of RC-MNL models, including all of the famous examples in the recent industrial organization literature, also impose symmetry. Specifically, the specifications estimated in BLP (1995), Nevo (2001), and Petrin (2002) each therefore implicitly impose symmetry while Goolsbee and Petrin (2004) explicitly impose it. Unfortunately, symmetry restrictions should not in general be expected to hold in aggregate demand models and should not be imposed a priori on aggregate demand systems, but rather the validity of such restrictions should be tested in any given context. In contrast, the condition $(\frac{\partial G_j}{\partial p_{kj}})^\tau > 0$, while strict, does allow cross price effects to be arbitrarily small and so is not a substantive constraint on the DGP.

Lemma 4.

(i) All members of the RC-MNL model with $v(w_j, p_j, y_j, \alpha_j, \beta_j, e_{ij}) = \bar{v}_j(w_j, p_j, y_j, \alpha_j, \beta_j) + e_{ij}$ have cross-price derivatives equal to the average across individual of an identical multiplicative constant (the product of market shares) times the derivative of product $k$’s utility with respect to its price. If $\bar{v}_j(w_j, p_j, y_j, \alpha_j, \beta_j) = f_j(w_j, y_j, \alpha_j, \beta_j) - g(y_j, \alpha_j, \beta_j)p_j$, i.e. the models satisfy an additive separability condition in $(w_j, p_j)$ and linearity in $p_j$ then a symmetry restriction, $\frac{\partial g_j}{\partial p_j} = \frac{\partial g_j}{\partial p_k}$, is imposed and the cross-price derivatives are only a function of the average product of market shares.

(ii) All members of the MEV class of models will impose a ‘proportional’ symmetry restriction, namely that $(\frac{\partial a_{jk}}{\partial p_{kj}})^{-1} \frac{\partial a_{ij}}{\partial p_{ij}} = (\frac{\partial a_{jk}}{\partial p_{kj}})^{-1} \frac{\partial a_{ij}}{\partial p_{ij}}$. If $\frac{\partial a_{jk}}{\partial p_{kj}} = \alpha$ for all $j > 0$, then the MEV class of models will impose the symmetry restriction, $\frac{\partial a_{ij}}{\partial p_{ij}} = \frac{\partial a_{ij}}{\partial p_{ij}}$.

Lemma 4(i) suggests that cross-derivatives in the RC-MNL models are not a function of how close a product’s characteristics are to other products’ characteristics (to the extent such similarities are not captured in shares). Moreover, most of the RC-MNL applications impose the further symmetry restriction on the cross-price derivatives. An implication of Lemma 4(ii) is the invariant proportion of substitution (IPS) discussed in Steenburgh (2008). The IPS property represents one of the researcher’s implicit assumptions about how an individual consumer will substitute away from competing alternatives, if improvements are made to one of the available goods. It holds if the proportion of demand generated by substituting away from a given competing alternative is the same no matter which own good attribute is improved. Formally, IPS implies $\frac{\partial a_{jk}}{\partial x_j} / \frac{\partial a_{ij}}{\partial y_j} = \frac{\partial a_{jk}}{\partial x_j} / \frac{\partial a_{ij}}{\partial y_j}$ for any two attributes $x_j$ and $y_j$ with $x_j \neq y_j$ which is implied by the proportional symmetry restriction.

To summarize, in this section we have shown that flexible (in the sense of Diewert, 1974) substitution patterns can be obtained by using members of the MEV class of functions that do not restrict the matrix of second derivatives of the $H$ function. In particular, if we use the homogeneous degree $\tau$ function where $H(r; \theta) = \sum_{j=0}^{J} \sum_{k \neq j} b_{jk} \left(\frac{\theta_j}{\theta_k}\right)^\tau + \sum_{j=0}^{J} b_{j} x_j^\tau$ which, with fixed $(\sigma, \tau)$ has at least $(J+1)(J+2)/2$ parameters in $B$ in the distribution of consumer tastes.

---

14 This result and the proof could be extended to asymmetric cross-price elasticities in the demand system by not imposing the restriction $b_{jk} = b_{kj}$, however several technical complications would arise.
as well as the parameters \( \theta_1 \) in the transformed function \( r_j = e^{\beta_j(x_j, \xi_j; p_{-j}, y, \theta_1)} \) then we will have a discrete choice demand system that can match a general matrix of own- and cross-price elasticities provided the DGP satisfies symmetry.

Under symmetry, the expected maximum utility function equals to \( \frac{1}{2} \ln H(r; \theta, \xi) + \gamma \) and the model can be used to compute the distribution of welfare changes in response to a change in the environment (e.g. a change in price or in the number of products). Furthermore in the spirit of the recent literature of demand estimation in differentiated product since Lancaster’s (1996) seminal work, we can map the substitution matrix \( B \) down to be parametric functions of “distance” of goods \( j \) and \( k \) on a characteristics space\(^15\) by following the approach recently suggested by Pinkse et al. (2002).\(^16\) Specifically in the Monte-Carlo and in the estimation in Section 6, we find a convenient specification for each element of the \( B \) matrix is as follows:

\[
b_{jkl} = \begin{cases} 
1/d_{jk}(x_j, x_k; \alpha_j) & \text{if } j \neq k \\
\exp(x_j/\alpha_j) & j = k 
\end{cases}
\]

where \( d_{jk}(x_j, x_k; \alpha_j) = (\sum_{l=1}^{L} \alpha_{jl}(x_{jl} - x_{kl})^2)^2 \) measure the “distance” between products \( j \) and \( k \) in characteristics space and \( (\alpha_{jl}, \alpha_{kj}) \) are the parameters to estimate.\(^17\) Such procedure will reduce the set of parameters to estimate and it will preserve the ability of the present framework to predict the change in demand and welfare following changes in characteristics, the entry of new products, the exit of old ones etc.\(^18,19\)

### 4. Identification and estimation

Following the approach proposed in Berry (1994) and BLP (1995), estimation is based on a two-step procedure nested within a minimization over the nonlinear parameters,\(^20\) those in the distribution of tastes, \( \theta = (B, \tau, \sigma) \). Specifically, for any given value of \( \theta_2 \), the vector of \( r_j \)’s are chosen so that predicted market shares are equal to actual market shares in each period (or market) \( t \). Since \( r_j = e^{\beta_j(w_j, \xi_j; \theta_2)} \) defining \( \delta_{jt} = \ln r_j \), we can compute \( \delta_{jt} \) for each product, run a Berry (1994) style regression based on the linear relationship \( \delta_{jt} = x_j \beta + \alpha p_j + \xi_j \), and then employ a generalized method of moments (GMM) estimator via a forming of conditional moment restrictions, \( E[\xi_j, \theta^J(x_j)] = 0 \), where \( \theta = (\theta_1, \theta_2), \xi_j, \theta^j = \delta_{jt}(\theta) = x_j \beta + \alpha p_j, \) and \( z_j \) denotes the \((1 \times q)\) vector of available instruments. Specifically, we define \( G_n(\theta) = \frac{1}{n} \sum_{j=1}^{J} \xi_j, (\theta) z_j \)

\(^{15}\)See Section 5 for an explicit application.

\(^{16}\)Such an approach is also the standard method of estimating the Probit discrete choice model where the vector of utilities \( u \) associated with the set of products in the market is assumed to follow the distribution \( u = (u_0, u_1, \ldots, u_J)^T \rightarrow N(\mu, \Sigma) \) and where the distributional parameters \( (\mu, \Sigma) \) are subsequently typically mapped to be assumed functions of product characteristics. By their nature, as a variance-covariance matrix, the distributional parameters in \( \Sigma \) will always be symmetric.

\(^{17}\)Notice that the parameters \( \alpha_j \) are identified up to their sign.

\(^{18}\)Since the \( B \) parameters describe only the distribution of tastes across the population, there are no restrictions from random utility theory on the way in which these parameters may vary with product characteristics with one exception: the \( B \) parameters must satisfy the restrictions required to ensure that the distribution of consumer tastes is a proper cumulative distribution function (cdf). In the MEV class of models, recall that Theorem 1 described that the cdf is \( F(x_0, \ldots, x_j; \theta_2, 3) = \exp\{-H(e^{-\mu}, \ldots, e^{-\mu}; \theta_2, 3)\} \). Thus, in the case of our \( H \) function, the \( B \) parameters must satisfy the symmetry and sign restrictions required for the model to be in the class of MEV models – conditional on the observed product characteristics.

\(^{19}\)Notice that in the RC model a change in characteristic \( j \) will change the joint distribution of utilities for all products but it will not change the distribution of the subvector of utilities for products other than \( j \). Instead in the FC-MNL, when \( B \) is assumed to be a parametric function of product characteristics with \( b_{jk} = b(X_j, X_k) \), a change in characteristics of product \( j \) will change the joint distribution of utilities as well as the distribution of any subvector of utilities (see also the earlier discussion in footnote 10). As a result a referee raised the interesting question of whether the version of the model which maps the \( B \) parameters to product characteristics – was truly “structural” with respect to changes in product characteristics. We believe that the model is clearly well motivated by an underlying utility theory, but we agree that the question of whether the last step – mapping the \( B \) parameters to product characteristics – can be motivated theoretically more explicitly provides an interesting avenue for further research.

\(^{20}\)Alternatively, a one-step estimator following Dube, Fox, and Che-Lin (2011) could be used.
formulate the GMM criterion function to minimize where \( \tilde{\xi}_{jt} = \xi_{jt} x_{jt}, \tilde{z}_{jt} = (z_{1jt} x_{jt}, \ldots, z_{qjt} x_{jt}) \), 
\( n = \sum_{j=1}^{J} \sum_{t=1}^{T} x_{jt} \) and \( x_{jt} = 1 \) if product \( j \) is sold in period (or market) \( t \) and zero otherwise so that \( x_{jt} \) provides a missing value indicator.\(^{21}\)

In our parameterization of the model, the market share function is clearly invariant to scalar multiples of the entire matrix of parameters \( B \). Thus, for identification it is necessary to fix one parameter of this matrix. Relatedly, the parameter \( b_{00} \) controls the nature of the own-price elasticity of demand for the outside good. Clearly, such a parameter will not be identified in data sets we typically have where the price of the outside good is normalized to 1 in every period or market. As a result of these two factors, a natural normalization for estimation is to choose \( b_{00} = 1 \) in estimation and, in so doing, fix the scale of the matrix of parameters \( B \).

The fact that we will typically observe no variation in the price of the outside good also suggests that it will be very difficult to identify in practice parameters \( b_{jj} \) for all \( j \). These parameters control the way in which substitution occurs from product \( j \) to product 1 when the price of the outside good increases. Since the price of the outside good is typically assumed fixed in our data sets, we do not observe such variation. A natural restriction appears to be to require that substitution is to be symmetric in these parameters since, in contrast, we can observe the way in which the demand for the outside good varies with the price of the inside goods – i.e., we can expect to be able to learn about the parameters \( b_{jj} \) for all \( j \) inside goods. In estimating the model, we therefore impose the symmetry conditions that \( b_{jj} = b_{jj} \) for all \( j \).

Furthermore, notice that since the model provides a flexible functional form for any fixed \( \tau > 1, 0 < \sigma < 1 \) and \( \tau \sigma < 1 \), in practice the values of the taste \( \tau, \sigma \) parameters can be fixed ex-ante. In the flexibility proofs, these taste parameters only serve to ensure that the constraint that the matrix of parameters in \( B \) can be all positive does not inappropriately constrain the flexibility properties of the model. Hence, we fix them ex-ante to limit the number of nonlinear parameters to estimate which clearly reduces the computational time and reduces the need of additional exclusion restrictions without affecting the flexibility of the model. The flexibility proof suggests that it may be advantageous to set \( \tau \) above but close to 1, so we set \( \tau = 1.1 \) while \( \sigma = 0.5 \) appears to be a natural choice.\(^{22}\)

The parameters are then identified providing the presence of sufficient instruments, i.e. \( q \geq \dim(\theta) \), but this is only necessarily whereas sufficient condition for local identification (and a necessary condition for global identification) is that the matrix \( \frac{\partial E(Z' \xi(\theta))}{\partial \theta} \bigg|_{\theta = \theta^*} \) has rank equal to its number of columns (see Rothenberg, 1971).\(^{23}\) In our context, we have

\[
\frac{\partial}{\partial \beta'} E(Z' \xi(\theta)) \bigg|_{\theta = \theta^*} = -E(Z' X) \tag{1}
\]

\[
\frac{\partial}{\partial \alpha} E(Z' \xi(\theta)) \bigg|_{\theta = \theta^*} = E(Z' p) \tag{2}
\]

\[
\frac{\partial}{\partial \theta_2} E(Z' \xi(\theta)) \bigg|_{\theta = \theta^*} = -E \left( Z' \left( \frac{\partial s(\delta, \theta_2, \theta^*)}{\partial \delta'} \right)^{-1} \left( \frac{\partial s(\delta, \theta_2, \theta^*)}{\partial \theta_2^*} \right) \right) \bigg|_{\theta = \theta^*}
\]

\(^{21}\) To simplify the exposition, we assume that the number of products is constant across markets (or periods) and equal to \( J \), so that \( n = J T \).

\(^{22}\) While it may feel slightly unnatural to some readers to “fix” parameter values exogenously, it is important to note that we could initially have written down our \( H() \) function with \( \sigma = 0.5 \) and \( \tau = 1.1 \) already substituted in and then have proven the flexibility of the resulting model. Doing so would clearly be equivalent. The model with sigma and tau estimated is clearly more flexible and so, for those who prefer, it may sometimes be possible to estimate these parameters. In this paper we have chosen not to do so since the model is already Diewert flexible with them fixed.

\(^{23}\) See also the survey paper by Newey and McFadden (1994). Nonparametric identification of demand in a general set-up is shown in Berry and Haile (forthcoming).

© RAND 2014.
\[
\begin{align*}
N & = 0 \\
\theta & = \theta^1 \\
\end{align*}
\]

where \( \xi(\theta) = (\xi_1(\theta), \ldots, \xi_T(\theta)) \) and in equation (3) each term \( \frac{\partial \xi_i(\theta)}{\partial \theta_j} \) is established by using the Implicit Function Theorem. This expression is immediately useful because it provides intuition regarding the data variation required for identification. Specifically, equation (3) indicates that the key requirement for identification, beyond the presence of sufficient instruments, is that predicted market shares will change if we change the parameters in \( \theta_2 \). That is, predicted market shares will vary as we move each of the up to \((J + 1)^2 - J - 1\) free parameters in the asymmetric matrix \( B \), away from their true values. Recalling that the \( B \) matrix control substitution patterns in the model, a change in the parameter \( b_{jk} \) will imply a different degree of substitution between products \( j \) and \( k \) following a price movement and this will affect the levels of market shares predicted by the model. In the case of a large number of products, it will typically be necessary to map the parameters in the matrix \( B \) down to be functions of a smaller set of parameters, i.e., \( (\alpha, \beta) \), which requires less data availability for identification. Notice also that if the characteristics for the outside option are set to zero, then this formulation also ensures that \( b_{00} = 1 \). Please see appendix A for further derivation and discussion.

Berry and Haile’s (forthcoming) paper clarifies the types of exclusion restrictions needed for identification. Specifically, identification requires not only an excluded instrument for each \( p_{jt} \), but also a sufficient number of instruments (BLP or more generally Chamberlin instruments) to identify all the remaining parameters, those in the distribution of tastes, \( B \). The need for additional instruments arises from the presence of the unknown taste parameters interacting with the endogenous variables: prices and quantities (or market shares), i.e., \((s_t, p_t)\), in the inverse demand map from market shares to mean utilities.

In contrast to a model based on simulation such as BLP, the estimator for this model will not need to be corrected for simulation error and hence the standard formulae for a GMM estimator apply. Following Hansen (1982), if we choose \( \theta \) to minimize \( G_s(\theta) AG_s(\theta) \) possibly subject to the linear restrictions \( R\theta = r \) then, under standard regularity conditions, \( ^{24} \)

\[
\sqrt{n}(\hat{\theta} - \theta_0) \sim N(0, D\Gamma A V_\theta \Gamma^\prime D)
\]

where \( D = C^{-1} - C^{-1} R (RC^{-1} R')^{-1} C^{-1}, C^{-1} = (\Gamma' A \Gamma')^{-1}, A \) is a non-negative definite weighting matrix with rank at least equal to the dimension of \( \theta \), \( \Gamma = (\Gamma_{1\theta}, \Gamma_{2\theta}) = (\frac{\partial G_s(\theta)}{\partial \theta_1}, \frac{\partial G_s(\theta)}{\partial \theta_2}) \) and

\( ^{24} \) See, for example, McFadden (1999).
$V_{G} = \text{Var}[G_{x}(\theta)]$ is the variance of the moment conditions. It is particularly useful to estimate allowing for linear restrictions since we will often want to estimate models with $b_{jk} = b_{kj}$. We may follow the traditional literature on the “continuous” demand for differentiated products, wherein asymptotic arguments are assumed to work in the number of independent time periods or markets or alternatively follow the BLP assumption that asymptotic arguments work in the number of products.\textsuperscript{25}

Finally, to solve the $J$-dimensional vector of share equations to compute $\delta_j$, we use a slightly amended BLP contraction mapping:

$$\delta_{j+1}^{k} = \delta_{j}^{k} + \rho(\ln s_{obs}^j - \ln(s_j(\delta_{j}^{k})))$$

for $j = 0, 1, \ldots, J$ with $0 < \rho < \tau^{-1}$, see appendix B for discussion and derivation. A more efficient contraction mapping based on convex programming methods procedure is proved to converge faster. Appendix B discusses these methods in details and shows how these algorithms are more efficient than the one based on BLP’s contraction mapping.

5. Monte-Carlo study

To illustrate the use of the model, we provide two sets of Monte-Carlo studies. In the first Monte-Carlo, we study the performance of the estimation algorithm highlighted in section 4 in retrieving the true underlying parameters; in the second Monte-Carlo, we assess the flexibility of the model by matching the elasticity matrix when the model is miss-specified.

□ Estimation algorithm. In this section, we provide a Monte-Carlo study which considers two variants of the model – one designed for data sets involving a small number of products from a larger number of markets and a second specification of the model appropriate for data sets where there are a large number of products. In each case, we set $\tau^{DGP} = 1.1, \sigma^{DGP} = 0.5$, we consider the (L-1) nonprice observed product characteristics $x_{jlt}^{DGP}$ and a price vector $p_{jt}^{DGP}$. The mean taste parameters are potentially option specific, but for the Monte-Carlo study we follow the majority of the literature and equalize them across inside options so that $\beta_{1}^{DGP} = \ldots = \beta_{L-1}^{DGP} = (1, 1, \ldots, 1, -1)$ and $\beta_{\text{price}}^{DGP} = -1$. Identification of the parameters requires a set of available instruments, we then use BLP instruments and, following the reasoning in the web-appendix, weighted sums of distances between product $j$ and its rivals $\frac{1}{2} \sum_{k \neq j} (x_{j} - x_{k})^2 (\sum_{l=1}^{L} (x_{jl} - x_{kl})^2)^{-1}$, and their interactions. We also use cost shifters to control for price endogeneity in specification B. The tolerance for the inner loop is set at 1-e10. Please see the web-appendix for further computational details.

Specification A: In this specification, we consider a small number of products and a large number of markets. Moreover, we assume price is exogenous and treat it as any other product characteristics. This exercise emphasizes the importance of the instruments to identify the demand system independently from the availability of any cost shifters as discussed in section 4. The unobserved product characteristic $\xi_j$, is drawn from $N(0, \sigma_{\xi})$ where we set $\sigma_{\xi} = 0.5$ whereas $x_{jlt}^{DGP}$ are assumed $N(-1, 1)$ for $l = 1, \ldots, L-1$ and $p_{jt}^{DGP}$ is assumed positive, $|N(-1, 1)|$. In this specification, we consider two products and the outside option for which all characteristics are set to zero. Hence, we set $J = 2$ and $L = 3$. Finally, since typical samples do not show all products selling in all markets and time periods, we assumed that a given product was missing in a given time period with probability 0.2.\textsuperscript{26} Products were assumed to be missing at random, following the certainly heroic assumption familiar from both the differentiated product and unbalanced panel data literatures.\textsuperscript{27} This specification sets the true data generating taste parameters $b_{jk}^{DGP}$ to

\textsuperscript{25} See the web-appendix for a detailed discussion of the two cases.

\textsuperscript{26} Specifically, a product was assumed missing in period t if the realization of a uniform random variable on [0,1] was less than 0.2, subject to the constraint that every market was assumed to have at least one inside good.

\textsuperscript{27} See Coublucq (2010) for a first attempt to address this issue.
be 1. For the purposes of this Monte-Carlo experiment, we do not force the matrix B to be fully symmetric so the only restriction imposed on the matrix B are to set \( b_{00} = 1. \)\(^{28}\)

Evidently, the results suggest first that the estimators are consistent; all the biases are reassuringly small at large sample sizes and, in fact, appear to be relatively modest even at relatively small sample sizes. Two features of the results stand out. First, that there can be significant small sample biases in the estimates of \( b_{21} \) – but these reduce dramatically as sample sizes grow. Second, the “linear” parameters in \( \beta \) appear much easier to identify – with bias, variance and MSE smaller at any given sample size.\(^{29}\)

Specification B: Next we turn to a discussion of the “large-J” case when the number of products is too large to estimate the elements of the matrix B, but instead those parameters are mapped down to be functions of parameters as discussed at the end of section 3. In this specification, we allow for price endogeneity. Therefore, along with three exogenous characteristics (\( L = 4 \)) drawn form \( N(1, 1) \), we allow for two additional characteristics that only shift costs, \( w_{j\tau}^{DGP} + 1.1 \sum_l |x_{jlt}^{DGP}| \), where each element of \( w_{j\tau}^{DGP} \) is independently drawn from a uniform distribution \( U(0, 2) \). The unobserved demand and cost characteristics are drawn from a bivariate normal distribution

\[
\begin{bmatrix}
\xi_{j\tau} \\
\omega_{j\tau}
\end{bmatrix} \sim N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.5 & 0.7 \\ 0.7 & 0.5 \end{bmatrix} \right)
\]

To generate the endogenous price variable, we use the competitive price specification \( p_{j\tau} = (w_{j\tau} + 1.1 \sum_l |x_{jlt}^{DGP}|)\lambda + \omega_{j\tau} \) with \( \lambda = (1, 1) \). Hence, we instrument price with the predicted price, \( \hat{p}_{j\tau} \), from a first-stage OLS regression on the linear competitive supply equation. The taste parameters, \( b_{j\tau} \), are a function of the exogenous characteristics as described in the main text.

Table 3 reports the results of the Monte-Carlo in two scenarios: as \( J \) increases for a given number of markets \( T \), and as \( T \) increases for a given large number of products \( J \). Table 3 makes clear that the estimates for the “large \( J \)” example do converge properly as the sample size gets large. Specifically, we can see that bias and mean square error falls for each of the parameters as the number of products or markets rises. It is however also clear that to achieve a given MSE, estimating the parameters which control own and cross-price effects \( \alpha_1 \) and \( \alpha_2 \) requires systematically larger samples than is required to estimate the “linear” parameters, \( \beta_1, \beta_2, \beta_3, \beta_{\text{price}} \).\(^{30}\)

□ Model flexibility. We consider whether the model is successfully exhibiting flexibility by fitting the model to data generated by a RC-MNL. We focus our attention to the large-\( J \) case which is more common in reality. The three exogenous characteristics, the unobserved demand and cost characteristics and the endogenous prices are generated as in Specification B. In the

\footnotesize{28} We also included the constraints that each \( b_{j\tau} \) be in the range [0,60] for the estimation routine to ensure the parameter space is compact. Such constraints did not bind in practice but did provide reassurance that the optimisation algorithm would do its job appropriately.

\footnotesize{29} The small sample (\( T = 100 \)) biases in \( b_{10} \) and \( b_{20} \) are unsurprising given we have not restricted the substitution towards the outside good parameters to be symmetric, i.e., we have not imposed \( b_{10} = b_{11} \) or \( b_{20} = b_{21} \) in the table. More surprising is the result that these parameters do appear identified in larger samples even without these restrictions. Since there is no observed variation in the price of the outside good (it is normalized to 1 throughout) the surprise is that these parameters do appear to be identified in large samples whereas in principle this author had expected to need to impose symmetry restrictions in order to allow actual “observed” variation in \( \partial s_{jt}/\partial p_{jt} \) to identify the (unobserved) variation in \( \partial s_{jt}/\partial p_{jt} \). Unreported results show that when the restrictions \( b_{10} = b_{11} \) and \( b_{20} = b_{21} \) are imposed, the small sample (\( T = 100 \)) bias in \( b_{10} \) and \( b_{20} \) is much reduced (from 0.596 and 0.439 as reported in the table to 0.073 and 0.119, respectively) and in addition, the small sample bias reported in \( b_{11} \) reduces from the –0.399 reported in the table to just –0.031 while the bias in \( b_{13} \) and \( b_{23} \) fall to 0.113 and 0.006, respectively. The variance and MSEs of the estimated B parameters also fall. The small sample bias in \( \beta \) also falls considerably (to –0.00732, –0.02196 and 0.02808 if \( T = 100 \)).

\footnotesize{30} Examination of the underlying Monte-Carlo runs makes clear that the immediate cause of the reported small sample bias in \( \alpha_1 \) and \( \alpha_2 \) is a tendency on occasion for the optimization routine to push an individual parameter estimate in \( \alpha_1 \) or \( \alpha_2 \) to zero. This does not occur in any Monte-Carlo experiments once the sample size is sufficiently large. Small sample bias in the estimates of \( \alpha_1 \) and \( \alpha_2 \) does not appear to contaminate the estimates of \( \beta \).
TABLE 2  Monte-Carlo simulation results for Specification A with J = 2 and asymptotic assumed to be working in the number of time periods or markets, T. Each figure in the table is calculated using 50 Monte-Carlo simulations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Bias</th>
<th>MSE</th>
<th>Bias</th>
<th>MSE</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>-0.06454</td>
<td>0.01019</td>
<td>-0.00683</td>
<td>0.0005</td>
<td>-0.00409</td>
<td>0.00022</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.07707</td>
<td>0.01198</td>
<td>-0.00252</td>
<td>0.0007</td>
<td>-0.00157</td>
<td>0.00028</td>
</tr>
<tr>
<td>$\beta_{price}$</td>
<td>0.08093</td>
<td>0.01768</td>
<td>0.00528</td>
<td>0.0013</td>
<td>-0.00111</td>
<td>0.00073</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Bias</th>
<th>MSE</th>
<th>Bias</th>
<th>MSE</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_00$</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>$b_{10}$</td>
<td>0.596</td>
<td>0.835</td>
<td>0.038</td>
<td>0.070</td>
<td>0.034</td>
<td>0.031</td>
</tr>
<tr>
<td>$b_{21}$</td>
<td>0.439</td>
<td>0.831</td>
<td>0.067</td>
<td>0.060</td>
<td>0.007</td>
<td>0.024</td>
</tr>
<tr>
<td>$b_{01}$</td>
<td>-0.083</td>
<td>0.454</td>
<td>0.009</td>
<td>0.111</td>
<td>-0.006</td>
<td>0.002</td>
</tr>
<tr>
<td>$b_{11}$</td>
<td>-0.399</td>
<td>0.372</td>
<td>-0.049</td>
<td>0.023</td>
<td>-0.004</td>
<td>0.012</td>
</tr>
<tr>
<td>$b_{21}$</td>
<td>0.101</td>
<td>0.861</td>
<td>0.023</td>
<td>0.028</td>
<td>-0.004</td>
<td>0.023</td>
</tr>
<tr>
<td>$b_{02}$</td>
<td>0.066</td>
<td>0.314</td>
<td>-0.021</td>
<td>0.010</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td>$b_{12}$</td>
<td>-0.215</td>
<td>0.388</td>
<td>-0.001</td>
<td>0.026</td>
<td>0.001</td>
<td>0.012</td>
</tr>
<tr>
<td>$b_{22}$</td>
<td>-0.231</td>
<td>0.466</td>
<td>-0.019</td>
<td>0.040</td>
<td>0.005</td>
<td>0.011</td>
</tr>
</tbody>
</table>

TABLE 3  Monte-Carlo simulation results using 50 replications for each experiment.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Bias</th>
<th>MSE</th>
<th>Bias</th>
<th>MSE</th>
<th>Bias</th>
<th>MSE</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>0.0463</td>
<td>0.0360</td>
<td>0.0056</td>
<td>0.0097</td>
<td>0.0076</td>
<td>0.0029</td>
<td>-0.0196</td>
<td>0.0122</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.0344</td>
<td>0.0323</td>
<td>-0.0047</td>
<td>0.0101</td>
<td>-0.0014</td>
<td>0.0031</td>
<td>0.0145</td>
<td>0.0082</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.0869</td>
<td>0.0267</td>
<td>0.0102</td>
<td>0.0068</td>
<td>-0.0037</td>
<td>0.0034</td>
<td>0.0134</td>
<td>0.0121</td>
</tr>
<tr>
<td>$\beta_{price}$</td>
<td>-0.0782</td>
<td>0.0070</td>
<td>0.0028</td>
<td>0.0003</td>
<td>0.0012</td>
<td>0.0001</td>
<td>-0.0009</td>
<td>0.0002</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>0.0463</td>
<td>0.0786</td>
<td>0.0065</td>
<td>0.0256</td>
<td>0.0029</td>
<td>0.0114</td>
<td>0.0373</td>
<td>0.0247</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>0.0344</td>
<td>0.0682</td>
<td>0.0192</td>
<td>0.0298</td>
<td>0.0237</td>
<td>0.0143</td>
<td>-0.0175</td>
<td>0.0188</td>
</tr>
<tr>
<td>$\alpha_{13}$</td>
<td>0.0229</td>
<td>0.0690</td>
<td>-0.0045</td>
<td>0.0225</td>
<td>0.0146</td>
<td>0.0172</td>
<td>-0.0139</td>
<td>0.0247</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>-0.0869</td>
<td>0.1176</td>
<td>-0.0537</td>
<td>0.0518</td>
<td>-0.0076</td>
<td>0.0404</td>
<td>0.0768</td>
<td>0.0556</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>-0.0782</td>
<td>0.1284</td>
<td>-0.0426</td>
<td>0.0512</td>
<td>-0.0224</td>
<td>0.0366</td>
<td>-0.0391</td>
<td>0.0386</td>
</tr>
<tr>
<td>$\alpha_{23}$</td>
<td>-0.0307</td>
<td>0.118</td>
<td>-0.0151</td>
<td>0.0587</td>
<td>-0.0431</td>
<td>0.0413</td>
<td>-0.0123</td>
<td>0.04</td>
</tr>
</tbody>
</table>

TABLE 4  Monte-Carlo simulation results using 50 replications for each experiment.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Bias</th>
<th>MSE</th>
<th>Truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>T = 3, J = 35</td>
<td>-0.0613</td>
<td>0.3453</td>
<td>-3.8869</td>
</tr>
<tr>
<td>T = 10, J = 35</td>
<td>-0.0630</td>
<td>0.2838</td>
<td>-3.8951</td>
</tr>
<tr>
<td>T = 10, J = 25</td>
<td>-0.1570</td>
<td>0.3998</td>
<td>-3.8807</td>
</tr>
</tbody>
</table>

simulated data we assume as before, that there are three dimensions of consumer preferences, $(\beta_1, \beta_2, \beta_3)$, each distributed independently normal with means and variances $E[\beta_i] = \{1, 1, -1\}$ and $Var[\beta_i] = \{0.5, 0.5, 0.5\}$. Finally, we assume that the price parameter $\beta_{price,jt} = \frac{\beta_{price}}{\gamma_{jt}}$ where $\beta_{price} = -1$ and $\gamma_{jt}$ is drawn from a log-$N(\mu_t, \sigma_t)$ where $(\mu_t, \sigma_t)$ are market specific and drawn from U(0,1). We simulate the integral in the market share equation with $n_s = 1000$ independent standard normal draws in each market. We then use the simulated data (market shares, exogenous characteristics and prices) to estimate a FC-MNL where $\tau_{DGP} = 1.1, \sigma_{DGP} = 0.5$ and the taste parameters, $b_{jkt}$, are function of the exogenous characteristics and a set of parameters $\alpha_{lt}$ as before.31

31 The estimation is performed as in specification B including the same IVs plus the mean income which divided the predicted prices in each market. Please refer to the web-appendix for the further computational details.
Table 5: Descriptive Statistics

<table>
<thead>
<tr>
<th>Years</th>
<th>No. of Models</th>
<th>Quantity '000</th>
<th>Price in €</th>
<th>Size</th>
<th>Engine Size (cc)</th>
<th>Weight</th>
<th>Nominal GDP per Capita €</th>
<th>Inflation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1991</td>
<td>77</td>
<td>2,233</td>
<td>7,810</td>
<td>632.56</td>
<td>7.83</td>
<td>1184.17</td>
<td>866.46</td>
<td>12,992</td>
</tr>
<tr>
<td>1992</td>
<td>76</td>
<td>2,273</td>
<td>8,740</td>
<td>634.76</td>
<td>8.03</td>
<td>1271.61</td>
<td>887.05</td>
<td>13,647</td>
</tr>
<tr>
<td>1993</td>
<td>78</td>
<td>1,775</td>
<td>9,426</td>
<td>634.93</td>
<td>8.07</td>
<td>1261.55</td>
<td>898.30</td>
<td>14,034</td>
</tr>
<tr>
<td>1994</td>
<td>77</td>
<td>1,613</td>
<td>9,981</td>
<td>640.90</td>
<td>8.17</td>
<td>1259.05</td>
<td>923.83</td>
<td>14,795</td>
</tr>
<tr>
<td>1995</td>
<td>80</td>
<td>1,556</td>
<td>11,097</td>
<td>645.01</td>
<td>8.28</td>
<td>1272.85</td>
<td>940.14</td>
<td>15,974</td>
</tr>
<tr>
<td>1996</td>
<td>84</td>
<td>1,518</td>
<td>12,249</td>
<td>659.29</td>
<td>8.39</td>
<td>1296.22</td>
<td>960.56</td>
<td>16,855</td>
</tr>
<tr>
<td>1997</td>
<td>85</td>
<td>2,166</td>
<td>11,500</td>
<td>644.91</td>
<td>8.24</td>
<td>1253.70</td>
<td>934.46</td>
<td>17,515</td>
</tr>
<tr>
<td>1998</td>
<td>87</td>
<td>2,065</td>
<td>12,137</td>
<td>655.54</td>
<td>9.09</td>
<td>1269.25</td>
<td>958.02</td>
<td>18,268</td>
</tr>
<tr>
<td>1999</td>
<td>91</td>
<td>1,874</td>
<td>12,408</td>
<td>655.81</td>
<td>9.14</td>
<td>1253.60</td>
<td>989.47</td>
<td>19,110</td>
</tr>
</tbody>
</table>

Table 6: Estimated parameters of the demand. The tolerance for the inner loop is set at 1-e10

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Variables</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear parameters in δ</td>
<td>Constant</td>
<td>-13.64</td>
<td>3.28</td>
</tr>
<tr>
<td></td>
<td>Fuel Consumption (L/100 Km)</td>
<td>-2.88</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>Log(size)</td>
<td>3.95</td>
<td>2.15</td>
</tr>
<tr>
<td></td>
<td>Log(cc)</td>
<td>4.95</td>
<td>1.22</td>
</tr>
<tr>
<td></td>
<td>Price</td>
<td>-2.52</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>Manufacturer Location dummies</td>
<td>Yes</td>
<td>–</td>
</tr>
<tr>
<td>Parameters in B diagonal α₂</td>
<td>Fuel</td>
<td>3.55</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>Log(size)</td>
<td>-6.10</td>
<td>2.62</td>
</tr>
<tr>
<td></td>
<td>Log(cc)</td>
<td>-4.63</td>
<td>1.20</td>
</tr>
<tr>
<td>Parameters in B off-diagonal α₁</td>
<td>Fuel</td>
<td>4.27</td>
<td>7.73</td>
</tr>
<tr>
<td></td>
<td>Log(size)</td>
<td>4.82</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Log(cc)</td>
<td>2.23</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Table 7: Average (weighted by sales) own- and cross-price elasticity across segments

<table>
<thead>
<tr>
<th>Segment</th>
<th>Own-Price Elasticity</th>
<th>Within-Market Segment Cross-Price Elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subcompact</td>
<td>-1.59</td>
<td>0.0066</td>
</tr>
<tr>
<td>Domestic</td>
<td>-1.67</td>
<td></td>
</tr>
<tr>
<td>Foreign</td>
<td>-1.48</td>
<td></td>
</tr>
<tr>
<td>Compact</td>
<td>-2.54</td>
<td>0.0050</td>
</tr>
<tr>
<td>Domestic</td>
<td>-2.62</td>
<td></td>
</tr>
<tr>
<td>Foreign</td>
<td>-2.36</td>
<td></td>
</tr>
<tr>
<td>Mid-Size</td>
<td>-3.32</td>
<td>0.0143</td>
</tr>
<tr>
<td>Domestic</td>
<td>-3.66</td>
<td></td>
</tr>
<tr>
<td>Foreign</td>
<td>-2.87</td>
<td></td>
</tr>
<tr>
<td>Sedan</td>
<td>-4.39</td>
<td>0.0015</td>
</tr>
<tr>
<td>Domestic</td>
<td>-4.58</td>
<td></td>
</tr>
<tr>
<td>Foreign</td>
<td>-3.99</td>
<td></td>
</tr>
</tbody>
</table>

Table 4 conforms the ability of the model to match the matrix of own and cross-price elasticities arising from the RC-MNL models. In fact, the model does so rather well. Thus for example, in the Monte-Carlo with T = 10 and J = 25, the average value of the true own-price elasticity was –3.8807 while the model estimated this elasticity with a bias of just –0.157 which
<table>
<thead>
<tr>
<th></th>
<th>BMW 3</th>
<th>Citroen XM</th>
<th>Fiat Uno</th>
<th>Honda Civic</th>
<th>Mitsubishi Lancer</th>
<th>Opel Vectra</th>
<th>Rover Montego</th>
<th>Suzuki Swift</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMW 3</td>
<td>−3.179536</td>
<td>0.000320</td>
<td>0.007156</td>
<td>0.000069</td>
<td>0.000016</td>
<td>0.000815</td>
<td>0.000020</td>
<td>0.000052</td>
</tr>
<tr>
<td>Citroen XM</td>
<td>0.001182</td>
<td>−4.843053</td>
<td>0.007156</td>
<td>0.000069</td>
<td>0.000117</td>
<td>0.000815</td>
<td>0.000020</td>
<td>0.000052</td>
</tr>
<tr>
<td>Fiat Uno</td>
<td>0.001082</td>
<td>0.000320</td>
<td>−1.199981</td>
<td>0.000069</td>
<td>0.000016</td>
<td>0.000815</td>
<td>0.000020</td>
<td>0.000058</td>
</tr>
<tr>
<td>Honda Civic</td>
<td>0.001082</td>
<td>0.000320</td>
<td>0.007174</td>
<td>−2.442549</td>
<td>0.000017</td>
<td>0.000817</td>
<td>0.000020</td>
<td>0.001378</td>
</tr>
<tr>
<td>Mitsubishi Lancer</td>
<td>0.001182</td>
<td>0.000320</td>
<td>0.007156</td>
<td>0.000070</td>
<td>−2.931113</td>
<td>0.000822</td>
<td>0.000086</td>
<td>0.000052</td>
</tr>
<tr>
<td>Opel Vectra</td>
<td>0.002182</td>
<td>0.000320</td>
<td>0.007156</td>
<td>0.000069</td>
<td>0.000017</td>
<td>−2.060492</td>
<td>0.000047</td>
<td>0.000052</td>
</tr>
<tr>
<td>Rover Montego</td>
<td>0.001183</td>
<td>0.000320</td>
<td>0.007156</td>
<td>0.000069</td>
<td>0.000071</td>
<td>0.001930</td>
<td>−2.537783</td>
<td>0.000052</td>
</tr>
<tr>
<td>Suzuki Swift</td>
<td>0.001182</td>
<td>0.000321</td>
<td>0.007917</td>
<td>0.001814</td>
<td>0.000016</td>
<td>0.000815</td>
<td>0.000020</td>
<td>−2.099886</td>
</tr>
</tbody>
</table>
is very encouraging. The results for the cross-price elasticities are similarly encouraging – we get a relatively small downward bias of −0.0099 on the average true value of 0.0918.


We use the model studied in section 3 to estimate the demand for new automobiles in Italy between 1991 and 1999. The variables in the data set include quantity, price, dummies for where the firm that produced the car is headquartered, the engine size (cc), the consumption in terms of number of litres necessary to drive 100 km, size (measured as length times width), and weight. The data set includes this information on the majority of models marketed during the 9-year period (models with extremely small market shares, such as the Ferrari and the Rolls Royce, are not included in the data). Since models both appear and exit over this period, this gives us an unbalanced panel. Treating a model/year as an observation, the total sample size is 735. Table 5 provides summary descriptive statistics of the variables in the dataset. The reported mean are weighted by sales.

Given the large number of products in each year, we map the matrix B down to be functions of car attributes as discussed in the specification B above. BLP-style instruments, weighted sums of distances between each product j and its rivals and their interaction are used in the estimation along with some additional cost shifters. Specifically, we use the price of aluminium and the one-year lag, and, each of these prices interacted with car characteristics like size and weight. In the estimation, we use the BLP-style moment conditions, \( G(\theta) = \frac{1}{n} Z^\prime \xi(\theta) \) so that the GMM estimator is \( \hat{\theta} = \arg\min_\theta G(\theta)^\prime W^{-1} G(\theta) \), where W is the optimal waiting matrix. Finally in the estimation, prices are divided by the nominal GDP per capita. The results are reported in table 6.

The estimates are all significant and have the expected sign. In particular, the price coefficient is precisely estimated and has the value −2.52. The parameters entering the matrix B are also well identified except for one.

Table 7 reports average own- and cross-price elasticities across different segments. The own-price elasticities are higher for larger and more expensive models. Moreover, table 7 shows that domestic cars have lower elasticities compared to foreign ones across all segments which suggest a higher degree of market power for the local manufacturer Fiat (Fiat Group’s share was about 50% in 1991).

Finally, table 8 reports the own- and cross-price elasticity for a selected sample of cars in 1991. It reads as follows: a one percentage increase in the price of the BMW 3 series raises the demand of the Citroen XM by 0.00032% and of the Fiat Uno by 0.008156% which was the most popular model in Italy.

7. Conclusions

In this paper, we develop the FC-MNL model of demand for differentiated products using aggregate data. FC-MNL relaxes the constraints imposed on own- and cross-price elasticities by popular analytic discrete choice models and yet does not require estimation via simulation; it is fully analytic. We develop a number of properties of the FC-MNL model. In particular, FC-MNL is shown to be a previously unexplored member of McFadden’s 1978 class of MEV discrete choice models. Hence, under testable parameter restrictions, FC-MNL is fully consistent with an underlying structural model of heterogeneous, utility maximizing, consumers. We provide a Monte-Carlo study to illustrate use of the model and to verify that the proposed estimators perform as anticipated in both the large T and large J cases. We notice that estimating the parameters which control own and cross-price effects requires systematically larger samples than those required to estimate the “linear” parameters. Finally, we illustrate how the model performs using a real-world

\(^{32}\) The data set has been used in Goldberg and Verboven (2001) and it is available upon request.

© RAND 2014
data set. By way of a closing remark, we note that the model could equally be estimated using individual choice data.

Appendix A

In Appendix A we provide the proofs to the various propositions, lemmata and corollaries stated in the main body of the paper.

Proof to Corollary 1. Part (i) follows since:

$$\frac{\partial \ln s_j}{\partial p_k} = \sum_{i=0}^{J} \left( \frac{\partial \ln s_i}{\partial r_i} \right) \left( \frac{\partial \ln r_i}{\partial p_k} \right) = \sum_{i=0}^{J} \left( \frac{\partial \ln s_i}{\partial \ln p_k} \right) = \frac{\partial \ln s_j}{\partial \ln p_k}$$

since $\frac{\partial \ln s_i}{\partial \ln p_k} = 0$ for all $i \neq k$. Defining, $r = (r_0, \ldots, r_J)$, we have $s_j(r) = r_j / \tau H_j(r)$ and hence $\frac{\partial \ln s_j}{\partial \ln p_k} = \frac{r_j H_j(r)}{\tau H_j(r)} - \frac{r_k H_k(r)}{\tau H_k(r)}$ and so for each inside good price, $k > 1$,

$$\frac{\partial \ln s_j}{\partial p_k} = \frac{\partial \ln s_j}{\partial \ln r_k} \frac{\partial \ln r_k}{\partial p_k} = \left( \frac{r_j H_j(r)}{\tau H_j(r)} - \frac{r_k H_k(r)}{\tau H_k(r)} \right) \frac{\partial \ln s_k}{\partial \ln p_k}$$

Part (ii) follows similarly since $\frac{\partial \ln s_j}{\partial \ln y} = \sum_{i=0}^{J} \left( \frac{\partial \ln s_i}{\partial \ln y} \right) \left( \frac{\partial \ln y}{\partial \ln p_k} \right)$ and so

$$\frac{\partial \ln s_j}{\partial \ln y} = \sum_{i=0}^{J} \left( I(i = j) + \frac{r_i H_i(r)}{H_j(r)} - \frac{r_j H_j(r)}{\tau H_j(r)} \right) \frac{\partial \ln s_i}{\partial \ln y} = \frac{\partial \ln s_j}{\partial \ln y} + \frac{1}{\tau H_j(r)} \sum_{i=0}^{J} \left( \frac{r_i H_i(r)}{H_j(r)} - \frac{r_j H_j(r)}{\tau H_j(r)} \right) \frac{\partial \ln s_i}{\partial \ln y}.$$

Q.E.D

Proof to Lemma 1. This essentially follows immediately from the expression $\frac{\partial \ln s_j}{\partial \ln p_k} = \frac{\partial \ln s_j}{\partial \ln y} = \frac{\partial \ln s_j}{\partial \ln p_k}$ provided exists a unique vector $r$ to ensure that $s_j(r) = s_j^*$. Then, for $j = k$, $\frac{\partial \ln s_j}{\partial \ln p_k} = \frac{\partial \ln s_j}{\partial \ln y}$ and so we need only to choose $\alpha_j = \frac{1}{\sigma_i = 0, \tau \sigma_i = 1, \sigma_i > 1}$ provided $\tau > 0$, $\sigma > 0$, $\tau \sigma_i = 1$, $\sigma_i > 1$ and $h_{j,k} \geq 0$ for all $j, k = 0, 1, \ldots, J$.

First, note that for all $r > 0$ the function $H(r; \theta_2, \gamma) = \sum_{j=0}^{J} \left( \frac{b_{j,k} + b_{j,k}}{2} \right) \left( \frac{r_j^2 + r_k^2}{2} \right)^{\gamma} + \frac{b_{j,k}^2 r_j^2}{2}$ is homogeneous of degree $\tau$ on $r \in \mathbb{R}_{j,k}^{+1}$. Next note that $H(r; \theta_2, \gamma) > 0$ on $r \in \mathbb{R}_{j,k}^{+1}$ if $h_{j,k} \geq 0$ for all $j, k = 0, 1, \ldots, J$.

Third, we must show that the cross derivatives are satisfied; namely that all first derivatives are non-negative while the second-order cross-derivative terms are non-positive. To do so, notice that:

1. $\frac{\partial H(r; \theta_2, \gamma)}{\partial r_j} = H_j(r; \theta_2, \gamma) = \tau \sum_{k,j} \left( \frac{b_{j,k} + b_{j,k}}{2} \right) \left( \frac{r_j^2 + r_k^2}{2} \right)^{\gamma-1} \frac{r_j^{\gamma-1} + \gamma r_j^2 r_j^{\gamma-1}}{2} \geq 0$

2. $\frac{\partial^2 H(r; \theta_2, \gamma)}{\partial r_j \partial r_k} = H_{j,k}(r; \theta_2, \gamma) = \tau \sum_{k,j} \left( \frac{b_{j,k} + b_{j,k}}{2} \right) \left( \frac{r_j^2 + r_k^2}{2} \right)^{\gamma-2} \frac{r_j^{\gamma-1} r_k^{\gamma-1} + \gamma r_j^2 r_k^{\gamma-1}}{2} \leq 0$

provided $\tau > 0$, $\sigma > 0$, $\tau \sigma_i = 1$, $\sigma_i > 1$ and $h_{j,k} \geq 0$ for all $j, k = 0, 1, \ldots, J$.

All third- and higher-order cross-derivative terms are zero by construction and hence $H$ satisfies the constraints that the cross derivatives alternate between non-negative and nonpositive. Although no sign restriction on the second own derivative is required, for completeness we note here that

$$H_{j,j}(r; \theta_2, \gamma) = \tau \sum_{k,j} \left( \frac{b_{j,k} + b_{j,k}}{2} \right) \left( \frac{r_j^2 + r_k^2}{2} \right)^{\gamma-2} \left( \frac{1}{\gamma-1} \right)$$

provided $\tau > 0$, $\sigma > 0$, $\tau \sigma_i = 1$, $\sigma_i > 1$ and $h_{j,k} \geq 0$ for all $j, k = 0, 1, \ldots, J$.

© RAND 2014.
\[
\sum_{j=0}^{J} \sum_{k=1}^{k_j} \left( \frac{a_{jk} + b_{jy}}{2} \right) \left( r_j^2 + r_j^* \right) 2^{r_j^* - 1} 
\]

Finally, we consider the limit condition \( \lim_{r_j \to +\infty} H(r_j; \theta_2; \mathcal{J}) = +\infty \) which is clearly satisfied since \( H(r_j; \theta_2; \mathcal{J}) = \sum_{j=0}^{J} \sum_{k=1}^{k_j} (a_{jk} + b_{jy}) r_j^2 + \tau (r_j - b_j^*) r_j^* - 2 \) for all \( r_j > 0 \) and \( b_j^* > 0 \).

\[Q.E.D\]

**Proof to Lemma 3.** We begin with the J+1 equations \( s_{j}^{\text{obs}} = f_j(r; \mathcal{J}) \), for \( j \in \mathcal{J} \) where \( f_j(r; \mathcal{J}) \) is homogenous of degree \( \tau > 0 \) in \( r \).

First notice that if \( s_j^{\text{obs}} = 0 \) for any good, \( j \), condition (i) on the \( f_j(r; \mathcal{J}) \) function ensures that setting \( r_j = 0 \) will solve the 0th equation exactly. Having set the \( r_j \)'s corresponding to products with zero market shares equal to zero, we can progress to consider the solution to the smaller set of equations \( s_j^{\text{obs}} = f_j(r_j; \mathcal{J}^+) \) for \( k \in \mathcal{J}^+ \), defined as in the lemma. Henceforth, we can work entirely with the functions for the reduced set of products \( \mathcal{J}^+ \). For brevity, since there's no ambiguity, we will drop the explicit dependence of functions on that set.

To prove the result, we will work with a function defined in terms of \( d_k \equiv \ln r_j \), namely define \( g_k(\delta) \equiv f_j(r_j) - s_j^{\text{obs}} \) for \( j \in \mathcal{J} \). Next, recall that a sufficient condition for uniqueness of a system of equations \( g_k(\delta) = 0 \) is that the Jacobian matrix of a function \( D_k g(\delta) \) is positive definite. (See, for example, Theorems 1.6 and 1.7 in Nagurney, 1999.) Recall also that a sufficient condition for positive definiteness is that there is a dominant diagonal, where the matrix \( D_k g(\delta) \) has a dominant diagonal if \( |d_{kk}(\delta)| > \sum_{k \neq k} |d_{kj}(\delta)| \) for each \( j \in \mathcal{J} \). Note that if \( |d_{kk}(\delta)| > 0 \) and \( |d_{kj}(\delta)| \leq 0 \) for all \( k \neq j \) with \( j, k \in \mathcal{J} \) then the dominant diagonal condition can be written in terms of the function as \( |d_{kk}(\delta)| > \sum_{j \neq k} (1 - |d_{kk}(\delta)|) \).

To establish that the dominant diagonal condition holds under the conditions in the lemma, note that for any \( j \in \mathcal{J}^+ \), since \( f_j(r) \) is homogenous of degree \( \tau > 0 \) in \( r \), by Euler's theorem, \( \tau f_j(r) = \sum_{k=0}^{r} \frac{\partial}{\partial r_j} f_j(r) \), so that \( \frac{\partial}{\partial r_j} f_j(r) = \sum_{k=0}^{r} \frac{\partial}{\partial r_j} f_j(r) \). Under the condition (iii) in the lemma, for each \( j \in \mathcal{J}^+ \), \( \tau f_j(r) > 0 \), we have \( |d_{kk}(\delta)| = \tau f_j(r) - \sum_{k \neq k} |d_{kj}(\delta)| \leq 0 \) for all \( k \neq j \) with condition (iv). Thus \( |d_{kk}(\delta)| > \sum_{j \neq k} (1 - |d_{kk}(\delta)|) \). This is of course exactly the same thing as saying that there is a unique \( \delta_j \) that solves the system of equations \( f_j(e^\delta; s) = s_j^{\text{obs}} \) for all \( j \in \mathcal{J}^+ \), i.e., that there are products with \( s_j^{\text{obs}} > 0 \). Thus there is a unique \( \delta_j \) that solves the system of equations \( f_j(e^\delta; s) = s_j^{\text{obs}} \) for all \( j \in \mathcal{J}^+ \), i.e., that there are products with \( s_j^{\text{obs}} > 0 \).

**For completeness, it is useful also to follow the logic in Berry (1994) and argue that all \( r_j \)'s will be finite provided \( s_j^{\text{obs}} > 1 \) for all \( j \in \mathcal{J} \). If so, then we need not worry about the potential unboundedness of the domain of the function \( f_j(r) \). Provided \( f_j(r) \) is continuous and increasing with \( \lim_{r_j \to +\infty} f_j(r) \geq 1 \), since \( s_j^{\text{obs}} < 1 \) we know that if a solution to the system of equations exists, then at the solution \( r_j < \infty \) for every \( j \in \mathcal{J} \). Then we need only consider whether a solution exists on a closed bounded set, \( r \in [0, M]^{J+1} \subset \mathbb{R}^{J+1}_+ \) for some sufficiently large \( M < \infty \). For any functions \( f_j(r) \) that are continuous on the set \([0, M]^{J+1} \subset \mathbb{R}^{J+1}_+ \), it also follows immediately that there is a solution by Brower's fixed point theorem on the set of the restricted domain, and hence on the unrestricted domain.

\[Q.E.D\]

**Proof to Corollary 2.** For brevity, we drop the explicit dependence of functions on the parameters, \( \theta_2 \). First, define

\[
f_j(r_j; \mathcal{J}^+) = f_j(r_j; 0, \ldots, 0; \mathcal{J}) = \tau \sum_{k=0}^{k_j} \frac{r_j^2 + r_j^*}{2} 2^{r_j^* - 1} + \sum_{k=0}^{k_j} b_{jk} r_j^2 + \tau (r_j - b_j^*) r_j^* = \tau \sum_{k=0}^{k_j} b_{jk} r_j^2 + \tau (r_j - b_j^*) r_j^*
\]

where \( b_j^* = 2^{1-\tau} \left( \sum_{k=0}^{k_j} \frac{r_j^2}{2} \right) + b_j^* \). We will apply Lemma 3 in order to establish that there is a solution \( \tau^* \) to the J+1 equations, \( s_j^{\text{obs}} = r_j N_j(r; \mathcal{J}) \), \( j \in \mathcal{J} \). Since \( 1 = \sum_{j=0}^{J} s_j^{\text{obs}} = \sum_{j=0}^{J} r_j N_j(r; \mathcal{J}) \), it will then follow immediately that \( \tau^* \) will also solve the market predicted equals actual market share equations, \( s_j^{\text{obs}} = \frac{\sum_{j=0}^{J} r_j N_j(r; \mathcal{J})}{\sum_{j=0}^{J} N_j(r; \mathcal{J})} \), \( j \in \mathcal{J} \). (Note that we are effectively imposing the normalization \( \sum_{j=0}^{J} r_j N_j(r; \mathcal{J}) = 1 \) instead of the more familiar normalization \( r_0 = 1 \).) 33 To

---

33 Imposing this alternative normalization also makes the analysis of some models particularly simple. For example, in the MNL model \( s_j(r) = r_j / \sum_{j=0}^{J} r_j \) so that the normalization \( H(r) = \sum_{j=0}^{J} r_j = 1 \) means we must find a solution to the entirely trivial set of equations: \( s_j^{\text{obs}} = r_j \) for all \( j = 0, 1, \ldots, J \). This normalization makes clear that the usual MNL model embodies a constraint that the market share of good \( j \) only depends on other goods characteristics because of
derive a solution with the usual normalization, just rescale the r’s being careful to use the full formula: \( s_j(r) = \frac{r_j \eta(r; \delta)}{\sum_{j} r_j \eta(r; \delta)} \) which is homogeneous of degree zero in \( r \).

Step 1. First, we establish condition (i) in Lemma 3. Notice that since \( r_j \eta_j(r_j; \delta) = -1 \) \( \sum_{j \in \mathcal{J}} b_{jk} \left( \frac{r_j^\tau + r_k^\tau}{2} \right)^{\tau-1} r_j^\tau + \tau b_j^\tau r_j^\tau \), setting any \( r_j = 0 \) ensures that \( r_j \eta_j(r_j; \delta) \leq 0 \) provided \( r_j > 0 \) and \( \tau \sigma - 1 < 0 \), since for \( \tau \sigma < 1 \), \( \frac{r_j^\tau}{r_k^\tau} \leq \frac{r_j}{r_k} \) which converges to zero as \( r_j \to 0 \) provided \( \tau > 0 \). In turn \( r_j \eta_j(r_j; \delta) \leq 0 \) at \( r_j = 0 \) ensures (i) that the predicted market share for that product is zero, and (ii) that when some substitute goods are missing, the parameter which controls the own-price elasticity of demand, \( b_j^\sigma \), is affected by a term \( \ln \sigma \) which depends on the extent of the substitution between products \( j \) and those products which are no longer available in the choice set.

Note that everything below must only hold on the set of products with positive market shares. In addition, if \( s_j^\alpha > 0 \), then the solution to \( s_j^\alpha = s_j(r) \) must involve setting \( r_j^\alpha(s) > 0 \); the predicted market share with \( r_j = 0 \) would be zero and so zero is the observed market share.

Step 2. Next, we establish condition (ii) of Lemma 3 holds, namely that \( \lim_{\tau \to -1} r_j^\alpha \frac{\delta f_j(r; \delta)}{\partial \ln r_j} \geq 0 \) for all \( \tau \in \mathcal{J} \). In fact, it will be algebraically easier and equivalent to work with \( \frac{\delta f_j(r; \delta)}{\partial \ln r_j} = \frac{\tau (\tau \sigma - 1)}{2} b_{jk} \left( \frac{r_j^\tau + r_k^\tau}{2} \right)^{\tau-1} \frac{r_j^\tau}{r_k^\tau} \leq 0 \).

proof to Proposition 1. We want to show that if we observe a vector of market shares \( s_j^\alpha \) for \( j \in \mathcal{J} \), the model can match that vector of market shares and also a true matrix of own- and cross-price elasticities \( \left( \frac{\partial \ln f_j}{\partial \ln r_k} \right)^\tau \) for any fixed values of \( \sigma > 0 \), \( \tau > 1 \) with \( \sigma \tau < 1 \). That is, to establish flexibility, we want to show that we can choose \( r, \alpha \) and the matrix of parameters \( B \) with \( jk \) element \( b_{jk} \), so that the model satisfies the following equations:

- \( s_j^\alpha = s_j(r; \theta_j) \) for \( j = 0, 1, \ldots, J \) \hspace{1cm} (5)
- \( \left( \frac{\partial \ln s_j}{\partial \ln p_k} \right)^\tau = \left( I(k = j) + \frac{r_k H_j(r; \theta_j)}{H_j(r; \theta_j)} - \tau s_j(r; \theta_j) \right) \frac{\partial \ln b_{jk}}{\partial \ln p_k} \) for \( j \in \mathcal{J}, k \in \mathcal{J}/0 \) \hspace{1cm} (6)

where \( \frac{\partial \ln b_{jk}}{\partial \ln p_k} = \alpha < 0 \).

Step 1

Equation (5) has an inverse.

We have already shown that under the conditions in Corollary 2 to Lemma 3, there exists a unique finite vector solution to equation (5) defined on the set of positive market shares. Following the approach taken in Lemma 3, it suffices to establish that there is a solution \( r^\ast \) to the \( J+1 \) equations, \( s_j^\ast = r_j, H_j(r; \theta_j) \), \( j \in \mathcal{J} \). Since \( 1 = \sum_{j=0}^J s_j^\ast = \sum_{j=0}^J r_j H_j(r; \theta_j) \)

the choice of normalization. Since \( r_j = e^{\theta_j} \), scaling the r’s by a constant \( \lambda > 0 \) is essentially adding the constant \( \ln \lambda \) to the vector \( \delta \), i.e., \( H(\lambda e^{\delta}) = H(e^{\delta + \ln \lambda}) \). It is for this reason, we can always rescale the r’s so to impose the normalization \( H(r) = 1 \).
which in turn ensures that \( \sum_{j=0}^{\ell_j} r_{j}^* H(r^*; \theta_2) = \tau H(r^*; \theta_2) = 1 \) at that solution since \( H \) is homogeneous of degree \( \tau \).

Under Corollary 2 then we can write the inverse of equations (5) as bounded vector function, \( 0 \leq r^*(\theta_2; \tau') < \infty \).

Equation (6) has an inverse.

We continue to work with the slightly amended market share equations, with \( s_j^* = r_j H_j(r; \theta_2) \) from (1), first. Notice that a vector \( r^* \geq 0 \) which solves the equations \( s_j^* = r_j H_j(r; \theta_2) \) \( j = 0, 1, \ldots, J \), we can rearrange equation (6) so to write each element \( b_{j\ell} \) as function of \( r^* \), \( \tau \) and the true shares and price elasticities given the fixed parameters \( \sigma \) and \( \tau \). Substituting in \( s_j^* = r_j H_j(r; \theta_2), \) \( \left( \tau \frac{\partial}{\partial p_k} \right)^{2} \) yields \( \frac{\partial}{\partial p_k} \) and rearranging simplifies our expression to give \( \frac{\partial}{\partial p_k} \left( \frac{1}{s_j^*} \right) \), which collapses to \( (\tau \frac{\partial}{\partial p_k} \left( \frac{1}{s_j^*} \right) \). Symmetry thus implies the constraint:

\[
\sum_{k=0}^{J} b_{j\ell}^* Y_k = \gamma_0 \left( s_j^* \right) \]

where \( \gamma_0 = \left( \frac{1}{s_j^*} \right) \). This implies the constraint thus reduces down to the Slutsky Symmetry restriction on the DGP that, \( \left( \frac{1}{s_j^*} \right) \), which is satisfied by assumption.

**Own Price Elasticities.** For own-price elasticities, our equation \( -r_j r_j H_j(r; \theta_2) = \gamma_j \) becomes \( -r_j r_j H_j(r; \theta_2) = \gamma_j \) for \( j \neq k \), so that we can choose

\[
b_{j\ell} = \frac{2 \sigma}{\tau (\tau - 1)} \left( \frac{r_j^* + r_k^*}{2} \right)^{2-\tau} \frac{r_j^*}{r_k^*} \frac{1}{\tau - 1 - \tau s_j^* s_k^*}
\]

which allows us to solve the \((J+1) (J+1)\) equations associated with all of the cross-price elasticities provided \( r_j > 0, \)

\[
r_k > 0, \tau > 0, \sigma > 0, \text{ and } (1 - \tau) > 0. \]

The requirement for \( r_j > 0, r_k > 0 \) always will be satisfied following Lemma 3 provided \( s_j^* > 0 \) and \( s_k^* > 0 \).

This means that we can write the inverse functions \( b_{j\ell}(r, \alpha, \sigma, \tau, \gamma_j) \) for all \( j \neq k \) and that it will be bounded so \( b_{j\ell}(r, \alpha, \sigma, \tau, \gamma_j) = M' < \infty \) for some fixed finite bound \( M' \) provided all of its components are bounded and, under our assumptions, we are not dividing anywhere by zero.

Non-negativity condition for cross-effects: Since we require a solution with \( b_{j\ell} > 0 \), and we have assumed \( \tau < 1 \) we require \( \gamma_j \equiv (1 - \tau) (s_j^* s_k^*) \gamma_j > 0 \) for every \( j \neq k \). Since we assume the DGP satisfies positive cross price effects, \( (\frac{1}{s_j^*}) > 0 \), for any fixed \( \tau > 0 \) we can always find an \( \alpha < 0 \) sufficiently small in magnitude to ensure this condition does indeed hold, for example set \( \alpha = -\tau \max_j s_j^* s_k^* (\frac{1}{s_j^*}) \). Thus we have established that a bounded inverse function exists \( 0 \leq b_{j\ell}(r, \alpha, \sigma, \tau, \gamma_j) < \infty \) for every \( j \neq k \).

The symmetry condition requires \( b_{j\ell} = b_{j\ell} \) for all \( j \neq k \). Symmetry thus implies the constraint:

\[
b_{j\ell} = \frac{2 \sigma}{\tau (\tau - 1)} \left( \frac{r_j^* + r_k^*}{2} \right)^{2-\tau} \frac{1}{\tau - 1 - \tau s_j^* s_k^*}
\]

which collapses to \( \left( \frac{1}{s_j^*} \right) \gamma_j \).
where $b_{ij} \geq 0$ for $j \neq k$ will be given by the solutions provided above. Since $rac{1}{\tau} \sum_{j} b_{ij} > 0$, provided $\tau > 1$ (while ensuring $\tau \sigma < 1$), by choosing $\alpha$ sufficiently small in magnitude, we can ensure that the term $\frac{1}{\tau} \sum_{j} b_{ij}$ dominates the expression for $-y_{ij}^\tau = \left(\frac{1}{\tau} \sum_{j} b_{ij}\right)\left(1 - \tau s_{ij}^\tau\right)$ and hence that the solution involves a $b_{ij}^\tau > 0$. Moreover, provided each $r$ is finite and nonzero (which we will ensure), the solution $b_{ij}^\tau(\alpha, \sigma, \tau, (\frac{1}{\tau} s_{ij}^\tau, s_{ij}^\tau))$ will be both positive and bounded — i.e., we have $0 < b_{ij}^\tau(\alpha, \sigma, \tau, (\frac{1}{\tau} s_{ij}^\tau, s_{ij}^\tau)) < M < \infty$.

Step 2

Fixed Point Argument.

Define $\omega$ to be the vector of the $(J+1)J/2$ unique elements of the symmetric matrix $B$, which includes $b_{ij}$ but not $b_{ij}$. Furthermore, define $\Delta_\omega = \{0, M\}^{J+1}J/2(\omega_1, \ldots, \omega_{J+1}/2)$ and $\Delta_B = \{0, M\}^{J+1}J/2(\omega_1, \ldots, \omega_{J+1}/2)$.

We have argued (in step 1) that there is a vector of bounded functions $\omega(r) : \Delta_B \to \Delta_\omega$ in the positive orthant such that $\omega = \omega(r, s^\tau, (\frac{1}{\tau} s_{ij}^\tau), \alpha, \tau, \sigma)$ where $\omega$ identifies all the elements of the matrix $B$ since it is symmetric.

Similarly, we have argued (following Corollary 2 and using $\omega$ to replace $B$) that there is a vector function $r(\omega) : \Delta_\omega \to \Delta_B$, defined as the inverse of the vector of equations defining that observed shares are equal to their predicted market shares for the avoidance of doubt, this vector function is made up of bounded functions in the positive orthant $r = (r(\omega), s^\tau, \alpha, \tau, \sigma)$.

Each of these functions is continuous and map from a nonempty, convex and compact set into a nonempty, convex and compact set. Thus we can define $\psi : \Delta_\omega \times \Delta_B \to \Delta_\omega \times \Delta_B$ by:

$$
\psi(r, \omega) = (r(\omega), \omega(r))
$$

which is by construction continuous and maps form a nonempty, convex and compact set into itself. We can therefore apply Brouwer’s theorem to establish the existence of a fixed point $(r^*, \omega^*) \in \psi(r^*, \omega^*)$.

Finally, for completeness, we note that at this fixed point, by construction we will have $\sum_{j=0}^J H_j(r^*; \omega^*) = \tau H(r^*; \omega^*) = 1$ and so the solution to our simplified equations is also a solution to our original equations equating market shares and own and cross-price elasticities. $Q.E.D$

Proof to Lemma 4. Part (i) Given some general distribution for heterogeneity in taste $P(v)$ for the additively separable RC-MNL model we can write, for any $j \neq k$, $\frac{\partial}{\partial t} \frac{\partial f_1}{\partial \tau} = 0$, only if $\frac{\partial}{\partial \tau} f_1$ is independent of option $k$.

Notice that the additively separable and linear assumption implies that $\frac{\partial}{\partial \tau} g(v_i) = \frac{\partial}{\partial \tau} g(v)$. Notice that the additively separable and linear assumption implies that $\frac{\partial}{\partial \tau} g(v_i) = \frac{\partial}{\partial \tau} g(v)$. Notice that the additively separable and linear assumption implies that $\frac{\partial}{\partial \tau} g(v_i) = \frac{\partial}{\partial \tau} g(v)$. Notice that the additively separable and linear assumption implies that $\frac{\partial}{\partial \tau} g(v_i) = \frac{\partial}{\partial \tau} g(v)$. Notice that the additively separable and linear assumption implies that $\frac{\partial}{\partial \tau} g(v_i) = \frac{\partial}{\partial \tau} g(v)$. Notice that the additively separable and linear assumption implies that $\frac{\partial}{\partial \tau} g(v_i) = \frac{\partial}{\partial \tau} g(v)$. Notice that the additively separable and linear assumption implies that $\frac{\partial}{\partial \tau} g(v_i) = \frac{\partial}{\partial \tau} g(v)$.

Part (ii) For any member of the MEV class of models $s_j = \frac{r^* j \partial_{ij}}{\partial \tau} \frac{\partial f_1}{\partial \tau}$, and for any $j \neq k$ we have $\frac{\partial}{\partial \tau} g(v_j) = \frac{\partial}{\partial \tau} g(v_k) = \frac{\partial}{\partial \tau} g(v)$. Notice that the additively separable and linear assumption implies that $\frac{\partial}{\partial \tau} g(v_i) = \frac{\partial}{\partial \tau} g(v)$. Notice that the additively separable and linear assumption implies that $\frac{\partial}{\partial \tau} g(v_i) = \frac{\partial}{\partial \tau} g(v)$. Notice that the additively separable and linear assumption implies that $\frac{\partial}{\partial \tau} g(v_i) = \frac{\partial}{\partial \tau} g(v)$. Notice that the additively separable and linear assumption implies that $\frac{\partial}{\partial \tau} g(v_i) = \frac{\partial}{\partial \tau} g(v)$. Notice that the additively separable and linear assumption implies that $\frac{\partial}{\partial \tau} g(v_i) = \frac{\partial}{\partial \tau} g(v)$.

More on Identification and Estimation

For identification, we further need that rank $\left(\frac{\partial^2}{\partial \tau^2} E(Z|x_0)\right)_{x_0=0}$ is $\dim(\delta_i) = J^2 + J$. Since the matrix $\frac{\partial^2}{\partial \tau^2} E(Z|x_0)\right)_{x_0=0}$ has full rank$^{34}$ which is equal to $n$ while $Z$ is $(n \times q)$ matrix with rank $q \geq \dim(\delta_i) > \dim(\delta_i)$, then identification essentially requires that $\text{rank}(\frac{\partial^2}{\partial \tau^2} E(Z|x_0)\right)_{x_0=0}) = \dim(\delta_i)$. As $s_{kj}(r_j, \delta_j, \delta_i)$, we can write $\frac{\partial^2}{\delta_i} g(v_j) = \frac{\partial^2}{\delta_i} g(v)$.

To simplify the algebra, define $\rho_{ij} = \frac{\partial^2}{\partial \tau^2} E(Z|x_0)\right)_{x_0=0}$, we can write $\frac{\partial^2}{\delta_i} g(v_j) = \frac{\partial^2}{\delta_i} g(v_k) = \frac{\partial^2}{\delta_i} g(v)$. Notice that the additively separable and linear assumption implies that $\frac{\partial}{\partial \tau} g(v_i) = \frac{\partial}{\partial \tau} g(v)$. Notice that the additively separable and linear assumption implies that $\frac{\partial}{\partial \tau} g(v_i) = \frac{\partial}{\partial \tau} g(v)$. Notice that the additively separable and linear assumption implies that $\frac{\partial}{\partial \tau} g(v_i) = \frac{\partial}{\partial \tau} g(v)$.

Specifically at $r_j$ which equates actual and predicted shares, $s_{kj} = s_k(r_j, \delta_j, \delta_i)$ for all $k$, we can use the expressions above to write for $j > 0$:

$$
\frac{\partial^2}{\delta_i} s_{kj}(r_j, \delta_j, \delta_i) = \frac{\partial}{\delta_i} \left(\frac{\partial}{\delta_i} s_{kj}(r_j, \delta_j, \delta_i)\right) = \left\{\begin{array}{ll}
\frac{\partial}{\delta_i} \left(\frac{\partial}{\delta_i} s_{kj}(r_j, \delta_j, \delta_i)\right) & \text{if } i = k, i = j \\
\frac{\partial}{\delta_i} \left(\frac{\partial}{\delta_i} s_{kj}(r_j, \delta_j, \delta_i)\right) & \text{if } i = k, i \neq j \\
\frac{\partial}{\delta_i} \left(\frac{\partial}{\delta_i} s_{kj}(r_j, \delta_j, \delta_i)\right) & \text{if } i \neq k, i = j \\
\frac{\partial}{\delta_i} \left(\frac{\partial}{\delta_i} s_{kj}(r_j, \delta_j, \delta_i)\right) & \text{if } i \neq k, i \neq j
\end{array}\right.
$$

and for $j = 0$, $s_{jk} = b_{jk}$.

$$
\frac{\partial^2}{\delta_i} s_{kj}(r_j, \delta_j, \delta_i) = \left\{\begin{array}{ll}
\frac{\partial}{\delta_i} \left(\frac{\partial}{\delta_i} s_{kj}(r_j, \delta_j, \delta_i)\right) & \text{if } i = k \\
\frac{\partial}{\delta_i} \left(\frac{\partial}{\delta_i} s_{kj}(r_j, \delta_j, \delta_i)\right) & \text{if } i \neq k
\end{array}\right.
$$

$^{34}$ Since it is block diagonal with each block invertible so is itself full rank and invertible.

$^{35}$ Please see the web-appendix for further details.
Thus the matrix \( \left( \frac{\partial \tilde{r}_j}{\partial b_i} \right) \) can then be written out with \( \theta_j^i \equiv (b_{i1}, b_{i2}, \ldots, b_{iJ}, b_{i0}, \ldots, b_{iJ}, b_{iJ}, b_{iJ}, b_{i0}) \)

\[
1 \left( \begin{array}{cccc}
-j_1 \rho^{j_1} - s_1 \rho^{j_1} & -s_1 \rho^{j_1} & \cdots & -s_1 \rho^{j_1} \\
-j_2 \rho^{j_2} & -j_2 \rho^{j_2} & \cdots & -j_2 \rho^{j_2} \\
\vdots & \vdots & \ddots & \vdots \\
-j_J \rho^{j_J} & -j_J \rho^{j_J} & \cdots & -j_J \rho^{j_J}
\end{array} \right) + \tau \left( \begin{array}{cccc}
-s_j \rho^{j_1} & -s_j \rho^{j_1} & \cdots & -s_j \rho^{j_1} \\
-s_2 \rho^{j_2} & -s_2 \rho^{j_2} & \cdots & -s_2 \rho^{j_2} \\
\vdots & \vdots & \ddots & \vdots \\
-s_J \rho^{j_J} & -s_J \rho^{j_J} & \cdots & -s_J \rho^{j_J}
\end{array} \right) = \tau \left( \begin{array}{cccc}
-s_J \rho^{j_J} & -s_J \rho^{j_J} & \cdots & -s_J \rho^{j_J} \\
-s_J \rho^{j_J} & -s_J \rho^{j_J} & \cdots & -s_J \rho^{j_J} \\
\vdots & \vdots & \ddots & \vdots \\
-s_J \rho^{j_J} & -s_J \rho^{j_J} & \cdots & -s_J \rho^{j_J}
\end{array} \right)
\]

It is immediate from its inspection to see that variation in market shares (and consequently in \( r \)) across products and periods guarantees that both the column and the row vectors in (4) are linearly independent. Therefore this along with variation among \( x \) and \( \rho \), and a sufficient number of instruments will provide the local identification for the parameters.

If parameters are mapped down to be functions of underlying characteristics, i.e., \( b_{jkt} = 1/d_{jk}(x_j, x_k; \alpha) \) if \( j \neq k \) \( \exp(x_j/x_k) \) otherwise, the latter follows by the assumption that \( \rho_j = \{1/p_i\} \) denotes the numerators of the market share function. Write the vector of equations, \( F(\delta) = \ln f(r(\delta)) - \ln x_{obs} \), in every column dimension of the matrix (4). Therefore, identification will require less data availability. Using such specification the parameters \( \alpha \) are identified up to their sign.

**Appendix B: Computational Methods used to Estimate the Model**

In Appendix B we consider methods that can be used in estimation of the model. In particular we first establish the properties of an amended BLP style algorithm. We then discuss an alternative algorithm based on Variational Inequalities. We also provide matrix forms for the key equations of the model.

**BLP style contraction mapping.** In this subsection, we establish that a BLP style algorithm is guaranteed to converge for this class of models. Specifically, we have argued previously it is often convenient to normalize the denominator of the market share function to 1. That is, with \( N_j(\rho; \alpha) = \frac{\sum_{i \in J} \rho_i^{\alpha_i}}{\rho_j} \), we may consider the J-dimensional equations \( F_j(\delta) = \ln f_j(r(\delta)) - \ln x_{obs} \) and \( J \neq j \) \( N_j(\rho; \alpha) \) denotes the numerators of the market share function. Write the vector of equations, \( F(\delta) = \ln f(r(\delta)) - \ln x_{obs} \). We will solve the J+1 equations, \( F(\delta) = 0 \).

Notice that at a solution point, \( x_{obs} = f(r(\delta)) \) and so \( x_{obs} = \sum_{j=1}^{J} f_j(r(\delta)) \), that is the denominator of each of the market share functions is exactly one. One advantage of using this normalization is that it saves a substantive amount of summation and division.

**Lemma (Amended BLP style contraction mapping algorithm).** A very slightly amended version of BLP’s contraction algorithm, with \( \hat{F}_j(\delta) = \delta_j + \rho \ln x_{obs} - \ln f_j(r(\delta)) \) defined for \( J = 0, 1, \ldots, J \) with \( 0 < \rho < \frac{1}{J} \) will converge.

**Proof.** We will apply the contraction mapping theorem in BLP to the function, \( F_j(\delta) = \delta_j + \rho \ln x_{obs} - \ln f_j(r(\delta)) \) \( J = 0, 1, \ldots, J \), noting that it is almost, but not quite, the function used by BLP (in particular it is defined over the J+1 dimensions of delta.)

To do so, note that \( \frac{\partial F_j(\delta)}{\partial \rho} = I(j = k) - \rho \frac{\partial \ln f_j(r(\delta))}{\partial \rho} = 0 \), where the inequality follows since \( \frac{\partial \ln f_j(r(\delta))}{\partial \rho} \leq 0 \) for \( j \neq k \), while \( \frac{\partial \ln f_k(r(\delta))}{\partial \rho} = \tau > 0 \), \( \frac{\partial \ln f_j(r(\delta))}{\partial \rho} = \tau - \sum_{j \neq k} \frac{\partial \ln f_j(r(\delta))}{\partial \rho} = 0 \) so \( I(j = k) - \rho \frac{\partial \ln f_j(r(\delta))}{\partial \rho} \geq 0 \) \( J = 0, 1, \ldots, J \) with \( 0 < \rho < \frac{1}{J} \) follows by the assumption that \( \rho \tau < 1 \).

We also require that \( \sum_{j \neq k} \frac{\partial \ln f_j(r(\delta))}{\partial \rho} = 1 - \rho \sum_{j \neq k} \frac{\partial \ln f_j(r(\delta))}{\partial \rho} = 1 - \rho \tau < 1 \) which follows immediately. \( Q.E.D \)

**Invertibility of the demand under the symmetry restrictions and the Variational Inequality algorithm.** In this section, we provide an alternative algorithm which solves the J dimensional vector of equations \( \delta(\tau; \tilde{\theta}) = \delta \) in every time period or market. We first establish that under symmetry these equations can be solved using convex programming methods. Under asymmetry, the equations must either be solved using BLP’s contraction algorithm or else we found that a Variational Inequality (VI) algorithm based on Nagurney’s (1999) General Iterative Scheme worked extremely well. After establishing the results for the symmetric model, we describe how to apply Nagurney’s algorithm to this context. We have not as yet proved a general convergence result for this algorithm for our problem, but it is a more natural generalization of the methods we discuss next for the symmetric case than the BLP contraction algorithm for reasons we also discuss below.

**Proposition 2:** Let \( V(\tau; \tilde{\theta}; \zeta) = \frac{1}{\tau} \ln H(\tau; \tilde{\theta}; \zeta) \), which is the MEV expected maximum utility function when evaluated at \( r_j = e^{\tilde{\theta}_j(\tau; \tilde{\theta}; r; \tilde{\rho})} \) for \( j = 0, 1, \ldots, J \), and let \( r_0 = 1 \). If for \( j > 0 \) and \( s_j = 0 \), set \( r_j = 0 \). If for all \( j \neq k \) with \( j, k \in \mathbb{N} \),

© RAND 2014
$\frac{\partial^2 V(r_j, \theta; \tau)}{\partial \theta_j \partial \theta_k} \leq 0$ and for each $k \in \mathcal{J}$, there is at least one $j_k \in \{1, \ldots, J\}$ with $j_k \neq k$ such that $\frac{\partial^2 V(r_j, \theta; \tau)}{\partial \theta_j \partial \theta_k} < 0$, then the solution to the inside good equations with positive market shares $s(r_j, \theta_j; 3^+)$ is $s^{obs}_{j,0}$ for $r_J > 0$ may be found as the unique solution to the strictly convex (and submodular) optimization problem:

$$\min_{[\ln r_j, J^{+}]^J} V(r; \theta; 3) - \sum_{j \in J^{+}} \ln r_j s^{obs}_j$$

Moreover, the $r$ solving this minimization problem can be found numerically using a Quasi-Newton method such as Davidson-Fletcher-Powell (DFP) with exact line search. Such an iterative algorithm provides a superlinear rate of convergence. When augmented with zeros for any products with zero market shares and $r_0 = 1$, the resulting $r_J$ will solve the full set of equations, $s(r, \theta_j; 3) = s^{obs}_j$.

Proof. McFadden (1981) establishes that the MEV class of models has an expected maximum utility function $V(r; \theta; 3) = 1/2 \ln H(r; \theta; 3)$ (with $r \in \mathcal{J}$) where $r_j = e^{(s_j(r; \theta; 3))}$ for $j \in \mathcal{J}$, the set of products. The MEV model has market share functions, $s_j = \frac{1}{2} \ln \left(\frac{\theta_j r_j}{\sum_j \theta_j r_j} \right)$ so that $s_j = \frac{\partial^2 V(r; \theta; 3)}{\partial \theta_j \partial \theta_k}$.

First notice that if $s^{obs}_j = 0$, setting $r_j^* = 0$ will solve the equation $s^{obs}_j = s_j(r; \theta; 3) = r \frac{\partial^2 V(r; \theta; 3)}{\partial \theta_j \partial \theta_k}$. Thus the solution involves $r_j^* = 0$ for any $j \in \mathcal{J} / \mathcal{J}^+$, where $\mathcal{J}^+ = \{ j \mid s^{obs}_j > 0, \, j \in \mathcal{J} \}$. Thus we will consider solving the equations only for goods with positive market shares, $j \in \mathcal{J}^+$. Define $V(r_j, \theta_j; 3^+) = V((r_j, 0, \ldots, 0); \theta_j; 3)$ when considering the reduced set of products so we may consider the equations $s^{obs}_j = s_j(r_j; \theta_j, 3^+) = \frac{-\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k}$ for $j \in \mathcal{J}^+/0$ are clearly the first order conditions to the following problem as stated in the lemma: $\min_{[\ln r_j, J^{+}]^J} V(r; \theta; 3) - \sum_{j \in J^{+}} \ln r_j s^{obs}_j$. For the rest of the proof for brevity we leave the dependency of all functions on the set $\mathcal{J}^+$ and the parameters $\theta_i$ implicit.

It remains to show that the program is convex and submodular. Convexity follows since $H(r)$ is a homogeneous degree $\tau$ function so that by Euler's equation:

$$\tau H(r) = \sum_{j \in \mathcal{J}^+} \frac{\sigma}{r_j} \frac{\partial V_{\mathcal{J}^+}}{\partial r_j}$$

w.r.t. $\ln r_j$ yields $0 = \sum_{j \in \mathcal{J}^+} \frac{\sigma}{r_j} \frac{\partial V_{\mathcal{J}^+}}{\partial r_j}$. Then $\frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k} = -\sum_{j \in \mathcal{J}^+} \frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k}$ so that provided $\frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k} \leq 0$ the Hessian of the function $V$ has a weakly dominant positive diagonal and $V$ is therefore convex in $\ln r$. If for each $k \in \mathcal{J}^+$ there exists at least one $j \in \mathcal{J}^+$ with $\frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k} < 0$ (strict) then $\frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k} = -\sum_{j \in \mathcal{J}^+} \frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k} < 0$ and a strict dominant diagonal condition in $\ln r$ is satisfied. Thus the objective function subtracts a concave (i.e., adds a convex (and in fact linear)) in $\ln r$ function, $\sum_{j \in J^{+}} \ln r_j s^{obs}_j$, from $V$ and so retains $V$'s convexity property.

Similarly, the condition that $\frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k} \leq 0$ implies $\frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k} = r \frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k}$ for $j \neq k$ (since $\frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k} = r \frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k}$) so that the function $V$ has weakly decreasing differences in $\ln r$ and hence so does the objective function. As well known, decreasing differences is a sufficient condition to ensure that a function is sub-modular. For the proof of the remaining elements of this proposition, see the proof in the main body of the text.

Once it is established that the objective function is strictly convex, any local minimum will also be a global minimum of the problem and hence the solution will also be unique. In addition, it is well known that suitably chosen Quasi-Newton methods will be globally convergent for convex problems. For example, Powell (1971, 1972) establishes that if an objective function is convex, then the DFP method with exact line search converges globally, with superlinear convergence. **Q.E.D**

Corollary 3. If for all $j, k, b_{jk} \geq 0$ and for each $k$, there exists at least one $j_k \in \{1, \ldots, J\}$ with $j_k \neq k$ so that $b_{jk} > 0$ then the model which sets $H(r; \theta_j) = \sum_{j \in \mathcal{J}} \sum_{k \neq k} b_{jk} \left( \frac{\theta_k}{\theta_k} \right)^{\alpha} + \sum_{k \neq k} b_{kk} r_k^\sigma$ satisfies the conditions of Proposition 2 and thus there exists a unique vector of $r$ which eafnes observed and predicted market shares. Moreover that vector may be computed using a Quasi-Newton algorithm such as the DFP method with exact line search.

Proof. $\frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k} \leq 0$ while $\frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k} = r \frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k}$ so that for $j \neq k$ and $r \in \mathcal{J}^{+}$ we have $\frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k} = \frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k}$.

In addition, $\frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k} = \frac{1}{2} \ln \frac{\theta_j r_j}{\sum_j \theta_j r_j} = \frac{1}{2} H(r) H(r)^{-1}$, and $\frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k} = \frac{1}{2} [H_j(r)(H(r)^{-1} - H(r) H_j(r) H(r)^{-1})]$ so sufficient conditions for $\frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k} \leq 0$ are that $H_j(r) \geq 0$ and $H_j(r) \leq 0$. Since in this model, $H_j(r; \theta_j; 3) = \frac{1}{2} (\sigma - 1) \ln \left( \frac{\theta_j r_j}{\sum_j \theta_j r_j} \right)^{\tau} \leq 0$ while $0 < \sigma < 1$, $\tau > 0$ and $\sigma \tau < 1$ for $j \neq k$ and $b_{jk} > 0$ is sufficient for $H_j(r; \theta_j; 3) < 0$ everywhere away from $r = 0$. **Q.E.D**

To build some intuition for this algorithm, note that within a neighbourhood of the solution, we will be able to use Newton's algorithm, which has a quadratic rate of convergence. (See, for example, Theorem 4.1.1. in Judd, 1998, or Fletcher, 1980.) Using Newton directly on this optimisation problem would involve using the recursion: $\ln r^{j+1} = \ln r^j - (D_{\ln r_j} V(r^j)) \cdot (D_{\ln r_j} V(r^j) - s)$, where $D_{\ln r_j} V(r) = (\frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k})$ and $D_{\ln r_j} V(r) = (\frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k})$ are analytic functions. In addition, since $\frac{\partial^2 V(r_j, \theta_j; 3^+)}{\partial \theta_j \partial \theta_k} = \frac{1}{2} \ln \left( \frac{\theta_j r_j}{\sum_j \theta_j r_j} \right) = s(r)$ and $s$ is the vector of observed market shares for the inside
goods with positive market shares, the term $D_{ss} V(r^n) - s$ is exactly the difference between the predicted and actual market shares. Essentially therefore this algorithm says decrease the $\tau$ associated with option $j$ if you are currently overpredicting its market share. The Hessian matrix provides a weighting matrix for the updates as usual, $D r^2 (r^n, V(r) = \frac{\partial}{\partial r_k}$, $\tau + \delta l$ of $(\tau D_s)$, which incorporates information about the curvature of the objective function. Quasi-Newton methods augment this procedure in a manner which ensures global convergence.

It is instructive to compare this algorithm with BLP's contraction mapping which is based on the iteration $\ln r^{n+1} = \ln r^n - \ln (D_{ss} V(r^n)/s)$, where the division is element by element. Evidently, Newton and Quasi-Newton algorithms bring in more information about the shape of the function than BLP's contraction method, and hence such methods should be expected to work substantially more efficiently (at least locally). In practice, one should use start with the BLP's contraction mapping until $||\ln r^{n+1} - \ln r^n|| < \varepsilon$ and then use the more efficient contraction discussed above.

We have observed that the best performance is obtained when $\varepsilon = 1e-2$ for the BLP contraction.

In general settings, we can say that the $(Jx1)$ vector equation $K(r) = s(r; \theta)$ will have a unique solution $K(r^n) = 0$ provided $D_s s(r; \theta)$ is positive definite. (See in particular pages 14–19 of Nagurney (1999) where her Theorems 1.4, 1.6, and 1.7 establish existence provided $K(r)$ is continuous over a compact set and uniqueness provided $D_s s(r; \theta)$ is positive definite.) Thus, it is therefore highly likely that for the asymmetric case, an algorithm based on Variational Inequality Algorithms will have sufficient efficiency properties relative to the BLP algorithm. We develop and use such an algorithm but leave a discussion of its convergence properties for future research. The Variational Inequality algorithm we use, which was found to have extremely good convergence properties in practice, is discussed below.

**Lemma 5.** Computation of the model is aided by noting the following matrix forms:

(i) $H(r; \theta) = (B - \text{diag}(B)), \cdot \left(\frac{\delta_1}{\delta_1^2} \cdot \delta_1\right) j_{1x1} + I_{jxj} \text{diag}(B) r^2$

(ii) $D r H(r; \theta) = \tau + \delta l \cdot \left((B - \text{diag}(B)) \cdot \left(\frac{\delta_1}{\delta_1^2} \cdot \delta_1\right) j_{1x1} + \tau (\text{diag}(B) r^2) \right)$

(iii) $(\frac{\delta_1}{\delta_1^2} \cdot \delta_1) \text{diag}(B) r^2 = \left((B - \text{diag}(B)) \cdot \left(\frac{\delta_1}{\delta_1^2} \cdot \delta_1\right) j_{1x1} \right)$

(iv) $D r V(r) = s(r) = r \cdot D H(r) \cdot I_{jxj}$

(v) $D r V(r) = (r^2) \cdot D H(r) \cdot I_{jxj}^{-1} + s(r) (I_{jxj}^{-1} - \tau s(r))$

Where, as before, $B$ is the $(Jx1 \times Jx1)$ matrix with $b_{jk}$ and we use $\text{diag}(B)$ to indicate a matrix of the same size as $B$, with the same diagonal and with all off diagonal elements set to zero. $I_{jxj}$ is a $(Jx1 \times 1)$ vector of ones and $\bullet$ is the Hadamard or element-by-element product.

**Proof.** Immediate from the definitions. Note that under symmetry, $\tau H(r) = \left(I_{jxj} \cdot (r \cdot D H(r) I_{jxj})^{-1}\right)$ whereas without symmetry, we must use the market share model $s(r) = r \cdot D H(r) \cdot \left(I_{jxj} \cdot (r \cdot D H(r) I_{jxj})^{-1}\right)$. If we use a nonsymmetric B matrix in $D H(r; \theta)$, it will still be positive definite.

Q.E.D

The importance of Lemma 5 is that the matrix expressions in it are all that must be programmed in order to compute predicted market shares for any given $(r, B, \tau, \sigma)$, and also the $r^*$ which equates predicted and actual market shares using the Quasi-Newton algorithm above. Notice that in applying a Quasi-Newton algorithm, we fix $r_0 = 1$ and use the expressions for $D r V(r)$ and $D r V(r)$ to compute the submatrix corresponding to the $r$'s for the Jx1 vector of inside goods. Thus, the core computation of the model for any given vector of taste parameters $\theta = (B, \tau, \sigma)$ requires little more than six lines of computer code -- the five lines in the lemma plus one for the Newton algorithm. Note that by programming the first derivative functions only in terms of $B$ rather than $(B + B')/2$ we can use the code to estimate the model even without the symmetry restrictions required to justify an underlying MEV model imposed.

**Variational Inequality Algorithms.** A finite dimensional Variational Inequality problem, $VI(E, K)$ is to determine a vector $\delta^* \in K \subset R^{n+1}$ such that $(F(\delta^*), \delta^* - \delta) \geq 0$ for all $\delta \in K$ where $\langle a, b \rangle$ denotes an inner product, $F$ is a given continuous function from $K$ to $R^{n+1}$ and $K$ is a given closed convex set. Both optimization problems and nonlinear equations can be solved using variational inequalities. Indeed, Proposition 1.1 of Nagurney (1999) establishes that if $K = R^{n+1}$ and $F : R^{n+1} \rightarrow R^{n+1}$ then a vector $\delta^* \in R^{n+1}$ solves $V(I, F, R^{n+1})$ if $F(\delta^*) = 0$.

Nagurney (1999) discusses a general iterative scheme which follows the following three steps:

(a) **Initialization.** Start with an $\delta_0 \in K$, set $k = 1$.

(b) **Construction and Computation.** Compute $\delta^k \in K$ by solving the variational subproblem $\langle g(\delta^k, \delta^k), \delta - \delta^k \rangle \geq 0$ for all $\delta \in K$.

(c) **Convergence verification.** If $|\delta^k - \delta^{k-1}| \leq \varepsilon$ for some $\varepsilon > 0$, a prespecified tolerance, the stop. Otherwise set $k = k+1$ and go to step (b).
The algorithm we worked with used Nagurney’s projection method, which chose the function
\[ g(\delta^k, \delta^{k-1}) = F(\delta^{k-1}) + \frac{1}{\rho} G \delta \]
for some fixed \( \rho > 0 \) and G a symmetric and positive definite matrix (e.g., the identity matrix.) As Nagurney describes, the special structure of the function \( g(\delta^k, \delta^{k-1}) \) means that we write the VI subproblem examined at step (b) as equivalent to a quadratic program. Note that follows because G is symmetric so that G is the VI solved at stage (b) and it is equivalent to solving the set of first order conditions for the optimization problem:
\[ \delta^k = \arg \min_{\delta \in \mathbb{R}^n} \frac{1}{2} \delta^T G \delta + (\rho F(\delta^{k-1}) - G \delta^{k-1}) \delta. \]

By differentiating, in the case where none of the constraints bind, it becomes clear that at the solution, \( \delta^k \), to this problem we will get \( \rho F(\delta^{k-1}) + G \delta^k = 0 \). (This equation can also be written, \( \delta^k = \delta^{k-1} - \rho G^{-1} F(\delta^{k-1}) \), from which it can be seen that we’re updating \( \delta \) by some multiple of the prediction error in log market shares.) Thus stated in terms of \( \delta \), the addition to the BLP algorithm is the presence of the positive definite matrix G and the “step-size” \( \rho > 0 \). The algorithm may be stated in terms of \( r \)‘s rather than delta’s which eases computation.

Quadratic program solvers are available in Gauss and Matlab and are exceedingly fast. Solving the problem in step (b) of this iterative procedure is very straightforward.

Formally to establish the conditions required for convergence we could use for example Nagurney’s Theorem 2.3.

References


