Peter Davis and Pasquale Schiraldi

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

Working paper

The Flexible Coefficient Multinomial Logit (FC-MNL) Model of Demand for Differentiated Products[1]

Peter Davis[†] ***Pasquale Schiraldi***[‡]

17th August, 2013

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 the use by estimating the demand for new automobiles in Italy.

[1] 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 re-circulated 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.

[†] Compass Lexecon and University College London Faculty of Laws. Email: pdavis@compasslexecon.com

[‡] London School of Economics and Political Science & CEPR. Email: p.schiraldi@lse.ac.uk

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).

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 & Pakes (2007) for a critique of the NMNL model.[2]

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 semi-parametric methods. This paper follows authors using the first of these three approaches, which we now discuss in turn.

[2] 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 (2010).

Recent papers proposing 'more' flexible parametric forms for discrete choice models include Ben-Akiva and Bielaire (1999), Bresnahan, Stern & Trajtenberg (1997), Chu (1989), Verboven & Brenkers (2002), Wen & Koppelman (2001), and Koppelman and Sethi (2010) 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 for example 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 have 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[3] (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

[3] 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 for example 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.

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.

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 & 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 and Bolduc 1996, Train 2003, Green at 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 & Mellman (1980) and Cardell & Dunbar (1980) for the first applications of the aggregate demand RC-MNL model and the more recent seminal contribution in Berry, Levinsohn & 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 & Train (2000) show that in principle, RC-MNL models are flexible in a far more general sense.[4] That said, we show that a large sub-class 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.[5] 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)

[4] 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 match arbitrary continuous functions of prices and indeed other product characteristics.

[5] 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.

generate instability and numerical errors which have been widely discussed in the literature in the context of RC-MNL estimation (see for example 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, e.g. those where computational constraints currently mean simulation based models are too costly to apply.[6] One such important arena involves the estimation of dynamic choice models. Recent papers in this area (see Gowrisankaran & Rysmann, 2009 and 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-NFXP algorithm turns out to be challenging from a computational point of view. Clearly, allowing for a flexible 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 semi-parametric and non-parametric discrete choice models, most notably in the seminal contributions from Matzkin (1992, 1993), discussed for example by Horowitz & 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 non-parametric analysis is easy to make in principle. However, as is well understood, the kinds of datasets we typically have in industrial organization or competition policy would not usually be sufficient to estimate fully non-parametric models. Briesch, Chintagunta & Matzkin (2002) provide an application of such techniques allowing some components of the model to be either semi- or non-parametric, and compare the results with MNL. Unfortunately, the estimation precision of non-

[6] 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 σ to vary with a consumer's characteristics z_i, so that we model $\sigma_i = F(\alpha + \gamma z_i)$ where now (α, γ) 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 and Sethi, 2005) who build upon Swait and Adamowicz (2001). Such models, designed for use with individual level datasets, 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 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.

parametric 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 semi-parametric 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 appendices.

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 $j \in \{0,1,..,J\} = \Im$. Each consumer is assumed to choose the option which provides the greatest conditional indirect utility, $\max_{j \in \Im} v_{ij}(w_j, y - p_j; \theta_1)$ where $w_j = (x_j, \xi_j)$ denotes the observed and unobserved characteristics of product j (respectively denoted x_j and ξ_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 θ_1 parameterizes the

conditional indirect utility functions.[7] We further suppose that $v_{ij}(w_j, y-p_j;\theta_1) = \delta_j(w_j, y-p_j;\theta_1) + \varepsilon_{ij}$ with ε_{ij} can be an option- and an individual-specific idiosyncratic component of conditional indirect utility.

In the familiar MNL model, ε_{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 $F(\varepsilon_{ij}) = \exp\{-\exp\{-\varepsilon_{ij}\}\}$. In that case, the aggregate market shares of each of the J choices are well known to have an analytic form: $s_j = \dfrac{e^{\delta_j(w_j, y-p_j;\theta_1)}}{\sum_{k=0}^{J} e^{\delta_k(w_k, y-p_k;\theta_1)}}$. It will sometimes be useful to define the functions $r_j = e^{\delta_j(w_j, y-p_j;\theta_1)}$ 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 τ functions due to Ben-Akiva and François (1983) for reasons that will become clear below.

Theorem 1 *(McFadden (1978); Ben-Akiva and François (1983)) Suppose* $H(r_0, r_1, ..., r_J; \theta_2, \Im)$ *is a non-negative, homogeneous of degree* $\tau > 0$*, function of* $(r_0, r_1, ..., r_J) \geq 0$*, where* θ_2 *is a vector of parameters and* $\Im = \{0,1,2,...J\}$*. Suppose (i)* $\lim_{r_j \to +\infty} H(r_0, r_1, ..., r_J; \theta_2, \Im) = +\infty$ *for* $j \in \Im$ *(ii) that for any distinct* $(i_1, ..., i_k)$ *from* $\Im$*,* $\dfrac{\partial^k H(r_0, ..., r_J; \theta_2, \Im)}{\partial r_{i_1} ... \partial r_{i_k}}$ *is non-negative if k is odd and non-positive if k is even. Then if the joint distribution of consumer heterogeneity has the form* $F(\varepsilon_0, ..., \varepsilon_J; \theta_2, \Im) = \exp\{-H(e^{-\varepsilon_0}, ..., e^{-\varepsilon_J}; \theta_2, \Im)\}$*, and consumers solve* $\max_{j=0,1,..,J} \delta_j(w_j, y-p_j;\theta_1) + \varepsilon_{ij}$ *then market shares have the analytic form:* $s_j = \dfrac{r_j \dfrac{\partial H(r_0, r_1, .., r_J; \theta_2, \Im)}{\partial r_j}}{\tau H(r_0, r_1, .., r_J; \theta_2, \Im)} = \dfrac{1}{\tau} \dfrac{\partial \ln H(r; \theta_2, \Im)}{\partial \ln r_j}$*, evaluated at* $r_j = e^{\delta_j(w_j, y, p_j;\theta_1)}$ *for all* $j \in \Im$*, and where* $r = (r_0, r_1, \ldots, r_J)$*.*

[7] 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 (2010), pages 462-467.

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 $\max_{j=0,1,..,J} \delta_j(w_j, y - p_j; \theta_1) + \varepsilon_{ij}$ - 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, $F(\varepsilon_0,...,\varepsilon_J; \theta_2, \Im) = \exp\left\{-H(e^{-\varepsilon_0},...,e^{-\varepsilon_J}; \theta_2, \Im)\right\}$.

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 Multinomial Nested Logit (MNNL) model (Williams, 1977; McFadden, 1978; Daly and Zachary, 1978) and subsequent modifications (see for example, the nest-specific parameters used by Verboven and Brenkers, 2002). Subsequent authors noted that the MNNL'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 and Bierlaire, 1999; Papola, 2004; Bielaire, 2006) and the Generalized Nested Logit (GNL) model (Wen and Koppelman, 2001) which Bielaire (2006) shows are equivalent.[8] The empirical industrial organization literature has also contributed to this branch of literature in the form of the Product Differentiation Logit (PDL) model (Breshanan *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 and Wen, 2000) and the Network GEV (NetGEV) model

[8] 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).

(Bierlaire, 2002; Daly and 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.

Table 1: Examples of members of the MEV class of models

	Author(s)	Description of H function	Parameters in the distribution of tastes
Multinomial Logit (MNL)	McFadden (1978)	$H(r_0, r_1, .., r_J; \Im) = \sum_{j \in \Im}^{J} r_j$	None
One Level Multinomial Nested Logit (MNNL)	McFadden (1978)	$H(r_0, r_1, .., r_J; \theta_2, \Im) = \sum_{g=1}^{G} \left(\sum_{j \in \Im_g}^{J} r_j^{\frac{1}{1-\sigma_g}} \right)^{1-\sigma_g}$	σ_g for g=1,..,G (G parameters)
Two Level Multinomial Nested Logit with nest specific parameters	Verboven and Brenkers (2002)	$H(r_0, r_1, .., r_J; \theta_2, \Im) = \sum_{g=1}^{G} \left(\sum_{h=1}^{H} \left(\left(\sum_{j \in \Im_{gh}}^{J} r_j^{\frac{1}{1-\sigma_{hg}}} \right)^{1-\sigma_{hg}} \right)^{\frac{1}{1-\sigma_g}} \right)^{1-\sigma_g}$	σ_{hg}, σ_g for $g = 1,..,G$ and $h = 1,..,H$. (G+G*H) parameters. $\Im_{gh}$ partition $\Im$.
Cross Nested Logit Model	*Vovsha (1997)*	$H(r_0, r_1, .., r_J; \theta_2, \Im) = \sum_{g=1}^{G} \left(\sum_{j \in \Im_g}^{J} a_{jg} r_j \right)^{\sigma}$	$0 \le a_{jg} \le 1$ for all j,g $0 \le \sum_g a_{jg} \le 1$ and $\sum_g a_{jg}^{\sigma} = 1$. $\Im_g$ do not partition $\Im$.
Product-Differentiation model (PD)	Bresnaha, Stern and Trajtenberg (1997)	$H(r_0, r_1, .., r_J; \theta_2, \Im) = \sum_{g=1}^{G} a_g \left(\sum_{h=1}^{H} \left(\sum_{j \in \Im_{gh}}^{J} r_j^{\frac{1}{1-\sigma_g}} \right)^{1-\sigma_g} \right)$ where $\sum_{g=1}^{G} a_g = 1$	a_g, σ_g for $g = 1,..,G$ (2G parameters) . $\Im_{gh}$ do not partition $\Im$.
Paired Combinatorial Logit model (PCL)	Chu (1989)	$H(r_0, r_1, .., r_J; \theta_2, \Im) = \sum_{i=0}^{J-1} \sum_{j=i+1}^{J} \left(r_i^{\frac{1}{1-\sigma_{ij}}} + r_j^{\frac{1}{1-\sigma_{ij}}} \right)^{1-\sigma_{ij}}$	$0 \le \sigma_{ij} < 1$ for all i,j pairs
Generalized Nested Logit (GNL) or Cross Nested Logit model (CNL)	Wen & Koppelman (2001) or Ben-Akiva and Bierlaire (1999); see Bierlaire (2006)	$H(r_0, r_1, .., r_J; \theta_2, \Im) = \sum_{g=1}^{G} \left(\sum_{j \in \Im_g} a_{jg} r_j^{\sigma_g} \right)^{\frac{\sigma}{\sigma_g}}$	$a_{jg} \ge 0$,, $\sum_{g=1}^{G} a_{jg} > 0$ '' $\sum_{g=1}^{G} a_{jg}^{\frac{\sigma}{\sigma_g}} = e^{-\gamma}$ where $\gamma = -0.5772$ is Euler's constant and $\sigma_g \ge \sigma > 0$ $\Im_g$ do not partition $\Im$.

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, $\Im_{hg}$ denotes the set of products that are in top level nest g and in second level nest h.

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 we will, for example, show later in the paper 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 discus 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

$$H_j(r;\theta_2,\Im)=\frac{\partial H(r;\theta_2,\Im)}{\partial r_j} \text{ and } H_{jk}(r;\theta_2,\Im)=\frac{\partial H_j(r;\theta_2,\Im)}{\partial r_k}.$$

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

$$\left(\frac{\partial \ln s_j}{\partial \ln p_k}\right)=\frac{\partial \ln s_j}{\partial \ln r_k}\frac{\partial \delta_k}{\partial \ln p_k}=\left(I(k=j)+\frac{r_k H_{jk}(r;\theta_2,\Im)}{H_j(r;\theta_2,\Im)}-\pi s_k(r)\right)\frac{\partial \delta_k}{\partial \ln p_k} \text{ for}$$

$j\in\Im=\{0,1,2,...J\}$, $k\in\Im/0=\{1,2,...J\}$ *(i.e., A/B denotes the set 'A not B') and with each element evaluated at* $r_j=e^{\delta_j(w_j,y-p_j;\theta_1)+\xi_j}$ *for* $j\in\Im$.

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

$$\left(\frac{\partial \ln s_j}{\partial \ln y}\right) = \frac{\partial \delta_j}{\partial \ln y} + \frac{1}{H_j(r;\theta_2,\Im)} \sum_{l=0}^{J} r_l H_{jl}(r;\theta_2,\Im) \frac{\partial \delta_l}{\partial \ln y} - \tau \sum_{l=0}^{J} s_l(r) \frac{\partial \delta_l}{\partial \ln y} \quad \text{for } j \in \Im.$$

Using the expressions provided in Corollary 1, it is possible to see why the MNL model, wherein $H(r_1,..,r_J;\Im) = \sum_{j\in\Im} r_j$ 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 criticised expression $\left(\frac{\partial \ln s_j}{\partial \ln p_k}\right) = \left(I(j=k) - s_k(r)\right)\left(\frac{\partial \delta_k}{\partial \ln p_k}\right) = \left(I(j=k) - s_k(r)\right)\left(\frac{\partial \delta_k}{\partial p_k}\right) p_k$. 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 (eg. Bresnahan, Stern & Trajtenberg, 1997; Verboven & Brenkers, 2002; Wen & Koppelman, 2001; Ben-Akiva and 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 non-zero for *j, k*.[9]

On the other hand, provided we make the last term parametric $\frac{\partial \delta_k}{\partial p_k} = \alpha_k$, 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 and

[9] Specifically, Wen & Koppelman's Table 1 suggests that allowing for a sufficient number of log-sum parameters the GNL could potentially match any *symmetric* price elasticities matrix. However, no proofs or particular discussions in this direction are provided.

$\left(\frac{\partial \ln s_j}{\partial \ln p_k}\right)^*$ 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,..,J.*

Lemma 1 *The MNL model with option specific price parameters can match any vector of own-price elasticities provided* $0 < s_k^* < 1$ and *provided there exists an r (or following BLP an underlying vector of unobserved product characteristics* $\xi = (\xi_1,...,\xi_J)$*) such that* $s_k(r) = s_k^*$ *with* $r_j = e^{\delta_j(x_j,\xi_j,y-p_j;\theta_1)}$.

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 $\left(\frac{\partial s_j}{\partial p_k}\right)^* = -s_k^* s_j^* \alpha_k$ for all $j \neq k$, while fitting the own-price elasticites involves setting $\alpha_k = \frac{1}{(1-s_k^*)s_k^*}\left(\frac{\partial s_j}{\partial p_k}\right)^*$. These relationships mean we can only fit both if the DGP satisfies the restriction that $\left(\frac{\partial s_j}{\partial p_k}\right)^* = -\frac{s_j^*}{(1-s_k^*)}\left(\frac{\partial s_k}{\partial p_k}\right)^*$ for all $j \neq k$, a set of restrictions for which there is no obvious theoretical motivation. For datasets 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:

$$H(r;\theta_2,\Im) = \sum_{j\in\Im}\sum_{k\neq j} b_{jk}\left(\frac{r_j^{\frac{1}{\sigma}} + r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma} + \sum_{j=0}^{J} b_{jj}^{\Im} r_j^{\tau}$$

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*)x(J+*1*) *matrix with jk*th *element* b_{jk}. The parameter $b_{jj}^{\Im}$ is indexed by the choice set $\Im$ 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_{jj} 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 $j \neq k$, $b_{jj} = 1$ 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 & Koppelman (2001) and Ben-Akiva and Bierlaire (1999). To see the mapping, recall that the CNL has a general MEV generating function $H(r_0, r_1, .., r_J; \theta_2, \Im) = \sum_{g=1}^{G}\left(\sum_{j\in\Im_g} a_{jg} r_j^{\mu_g}\right)^{\frac{\mu}{\mu_g}}$ which, if we (i)

put every pair of products into a nest so that nest g is made up of products $\Im_g = \{j,k\}$ for all combinations of products $j \neq k$ and in addition, (ii) put every product in its own nest so that we also have $\Im_{g'} = \{j\}$ for all products $j = 0,..,J$; and (iii) set $\mu = \tau$, $\mu_g = \frac{1}{\sigma}$ for all g. Doing so allows us to write the expression $\left(a_{jg} r_j^{\mu_g} + a_{kg} r_j^{\mu_g}\right)^{\frac{\mu}{\mu_g}}$ for each $j \neq k$ nest since $\left(a_{jk} r_j^{\mu_g} + a_{kj} r_j^{\mu_g}\right)^{\frac{\mu}{\mu_g}} = \left(2a_{jk}\right)^{\frac{\mu}{\mu_g}} \left(\frac{r_j^{\mu_g} + r_j^{\mu_g}}{2}\right)^{\frac{\mu}{\mu_g}}$. That is, we can set $b_{jk} = \left(2a_{jk}\right)^{\frac{\mu}{\mu_g}}$ for $j \neq k$ and $b_{jj}^{\Im} = \left(a_{jj}\right)^{\frac{\mu}{\mu_g}}$ 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* $H(r;\theta_2,\Im)$ *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,…J,* (ii) for each j there is a good indexed $m_j \in \Im$ 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 the annex 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 $\frac{\partial H(r;\theta_2,\Im)}{\partial r_j} = \tau \sum_{k \neq j} \left(\frac{b_{jk} + b_{kj}}{2}\right) \left(\frac{r_j^{\frac{1}{\sigma}} + r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma - 1} r_j^{\frac{1}{\sigma} - 1} + \tau b_{jj}^{\Im} r_j^{\tau - 1}$ 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, $s_j(r) = r_j N_j(r) \Big/ \sum_{l=0}^{J} r_l N_l(r)$ and $N_j(r;\theta_2) = \tau \sum_{k \neq j} b_{jk} \left(\frac{r_j^{\frac{1}{\sigma}} + r_k^{\frac{1}{\sigma}}}{2} \right)^{\tau\sigma - 1} r_j^{\frac{1}{\sigma}-1} + \tau b_{jj} r_j^{\tau-1}$ 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 $N_j(r) = \frac{\partial H(r;\theta_2)}{\partial r_j}$. To see that, notice that since $H(r;\theta_2)$ is homogeneous of degree τ, $\tau H(r;\theta_2) = \sum_{k=0}^{J} r_k \frac{\partial H(r;\theta_2)}{\partial r_k}$ and so the share function becomes one generated by an MEV model wherein $s_j(r) = r_j H_j(r) \Big/ \tau H(r)$. 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 sub-section 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 $j = 0,..,J$. 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, $\left(\frac{\partial s_j}{\partial p_k} \right)^* = \left(\frac{\partial s_k}{\partial p_j} \right)^*$ 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 utilities[10] associated with each option that make predicted and actual market shares equal for all the goods. More precisely, we show that we can always solve for the monotonic transformation of the mean utilities $r_j = e^{\delta_j(w_j, y, p_j, p_1; \theta_1)} \quad j \in \Im$. 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 Diewert (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 $(x_j, y, p_j; \theta_1)$, the equation $r_j^* = e^{\delta_j(x_j, \xi^*_j, y, p_j; \theta_1)}$ implicitly defines a solution ξ_j^* which exists and is recoverable provided $\frac{\partial \delta_j(x_j, \xi_j, y, p_j; \theta_1)}{\partial \xi_j} > 0$. Thus we may entirely equivalently think of picking the r_j^* or ξ_j^* to match market shares. Under either of these interpretations, finding r_j^* or ξ_j^* amounts essentially to finding the level of utility that must be associated with product j in order to explain its observed market share, where $\xi^* = (\xi_0{}^*, \xi_1{}^*, \ldots, \xi_J{}^*)$. An implication of Lemma 3 below therefore will be that we can use an estimation strategy based on the suggestion from

[10] Notice that the term "mean utility" (of product *j*) to denote δ_j is misleading in our context because it is not difficult to show that the above discussed MEV distribution implies $E[\varepsilon_{ij}] = \left(\ln\left(\left(\frac{1}{2} \right)^{\tau\sigma - 1} \sum_{k \neq j} b_{jk} + b_{jj}^{\Im} \right) + \gamma \right) / \tau$ where γ is the Euler constant (see Abbe at al., 2005, for the derivation of the result for the CNL model). Therefore, the mean utility of product *j* is actually equal to the sum of the previous two elements. However, we will keep referring to δ_j as mean utility for clarity of exposition and to simplify the comparison with the other papers in this literature. Finally observe that one can impose the constraints $\left(\left(\frac{1}{2} \right)^{\tau\sigma - 1} \sum_{k \neq j} b_{jk} + b_{jj}^{\Im} \right) = 1$ to avoid this feature of the model.

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

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

Lemma 3 *Let* $\Im \equiv \{0,1,...,J\}$ *be the set of products and* $\Im^+ \equiv \left\{ j \mid s_j^{obs} > 0, j \in \Im \right\}$ *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 $r \in \Re_+^{\#\Im}$ *and for any* $\Im^+ \subseteq \Im$ *and any* $r_{\Im^+} \in \Re_+^{\#\Im^+}$, *define* $f(r_{\Im^+};\Im^+) \equiv f(r_{\Im^+},0,..,0;\Im)$. Further suppose $f(r;\Im)$ has the following properties: (i) if $r_j = 0$ then $f_j(r;\Im) = 0$, (ii) $\lim_{r_j \to \infty} f_j(r;\Im) \geq 1$ for all $j \in \Im$, (iii) for each $j \in \Im^+$, *and any* $r_j > 0$, $f_j(r) > 0$ *and (iv) for each* $k \in \Im^+ / j$, $\frac{\partial f_j(r_{\Im^+};\Im^+)}{\partial r_k} \leq 0$. *Then there exists a finite vector of r's that solve the J+1 vector of equations* $s_j^{obs} = f_j(r;\Im)$ *for* $j \in \Im$. *If* $s_j^{obs} = 0$ *then a solution sets* $r_j^* = 0$ *for* $j \in \Im$. *Moreover, the solution to the subset of equations* $s_j^{obs} = f_j(r_{\Im^+};\Im^+)$ *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* $s_j(r;\theta_2,\Im) = \frac{r_j N_j(r;\theta_2,\Im)}{\sum_{l=0}^{J} r_l N_l(r;\theta_2,\Im)}$ and let $s_j^{obs} < 1$ *for all j. If (i)* $b_{jk} \geq 0$, *for all j,k=0,…J,* (ii) $b_{jj} > 0$ 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* $s_j^{obs} = s_j(r;\theta_2,\Im)$ *for* $j \in \Im$. *If* $j \in \Im / \Im^+$, *so* $s_j^{obs} = 0$, *then a solution sets* $r_j^* = 0$ *while if* $j \in \Im^+$, *so* $s_j^{obs} > 0$, *the solution sets* $r_j^* > 0$. *Moreover, the solution* $r_{\Im^+}^*$ *to the subset of equations* $s_j^{obs} = s_j(r_{\Im^+};\theta_2,\Im^+)$ *(i.e., with* $s_j^{obs} > 0$*) is unique.*

Notice that starting from a general framework where consumer preferences are represented by a random utility model, Berry & Haile (2012) and Berry Gandhi & Haile (2011) prove the invertibility of a demand system under mild conditions: (i) the index restriction and (ii) the

connected substitute assumptions.[11] Lemma 3 and Corollary 2 prove the uniqueness results in the present context also when the demand system is not necessarily generated by random utility models, i.e. $b_{jk} \neq b_{kj}$.[12]

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 *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.[13] 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 *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

[11] The index restriction requires that $x_{jt}^{(1)}$ (where $x_{jt}^{(1)}$ is at least an observed *exogenous* characteristic of product *j*) and ξ_{jt} affect the distribution of utilities only through an (potentially non-linear) index, for example $\tilde{\delta} = x_{jt}^{(1)}\beta + \xi_{jt}$ with the essential requirement that the index that is strictly monotonic in ξ_{jt} 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 $\tilde{\delta}$. 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 (2012) proves the uniqueness of the vector δ that rationalized the market shares in present context.

[12] Proposition 2 in appendix B provides an alternative approach to establish the invertibility of the demand system under symmetry condition.

[13] Typically, if symmetry is rejected when estimating say a NIDS 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.

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.[14]

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).

Proposition 1 *(Flexibility of the discrete choice model) Consider the MEV model* $H(r;\theta_2)=\sum_{j=0}^{J}\sum_{k\neq j}b_{jk}\left(\frac{r_j^{\frac{1}{\sigma}}+r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma}+\sum_{j=0}^{J}b_{jj}^{\Im}r_j^{\tau}$ *where* $r_j=e^{\delta_j(x_j,\xi_j y,p_j;\theta_1)}$ *with* $\frac{\partial\delta_j(x_j,\xi_j,y,p_j;\theta_1)}{\partial\xi_j}>0$ *for* $j\in\Im$ *with parameters* σ,τ *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* $j,k\in\Im$ *with* $b_{jj}>0$ *for all* $j\in\Im^{+}$, *and (iii)* $b_{jk}=b_{kj}$. *There exists a* (ξ,B,α) *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* $\left(\frac{\partial s_j}{\partial p_k}\right)^{*}>0$ *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,* $\left(\frac{\partial s_j}{\partial p_k}\right)^{*}=\left(\frac{\partial s_k}{\partial p_j}\right)^{*}$.

This proposition establishes that for any fixed $\tau>1$ and $0<\sigma<1$ with $\tau\sigma<1$ there are values of (ξ,B,α), so that the model can match an arbitrary vector of positive market shares and an arbitrary matrix of symmetric own- and cross-price elasticities.[15] 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,

[14] 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* may provide a flexible but relatively easy to estimate option for generating asymmetric models.

[15] 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.

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 & 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 $\left(\frac{\partial s_j}{\partial p_k}\right)^* > 0$, 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 $v(w_j, p_j, y_i, \alpha_i, \beta_i, \varepsilon_{ij}) = \bar{v}_j(w_j, p_j, y_i, \alpha_i, \beta_i) + \varepsilon_{ij}$ *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* $\bar{v}_j(w_j, p_j, y_i, \alpha_i, \beta_i) = f_j(w_j, y_i, \alpha_i, \beta_i) - g(y_i, \alpha_i, \beta_i)p_j$*, i.e. the models satisfy an additive separability condition in* (w_j, p_j) and linearity in p_j *then a symmetry restriction,* $\frac{\partial s_j}{\partial p_k} = \frac{\partial s_k}{\partial p_j}$*, 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 $\left(\frac{\partial \delta_k}{\partial p_k}\right)^{-1} \frac{\partial s_j}{\partial p_k} = \left(\frac{\partial \delta_j}{\partial p_j}\right)^{-1} \frac{\partial s_k}{\partial p_j}$*. If* $\frac{\partial \delta_j}{\partial p_j} = \alpha$ *for all j>0, then the MEV class of models will impose the symmetry restriction,* $\frac{\partial s_j}{\partial p_k} = \frac{\partial s_k}{\partial p_j}$*.*

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 $\frac{\partial s_k}{\partial x_j} / \frac{\partial s_j}{\partial x_j} = \frac{\partial s_k}{\partial y_j} / \frac{\partial s_j}{\partial y_j}$ 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 τ function where $H(r;\theta_2) = \sum_{j=0}^{J}\sum_{k\neq j} b_{jk}\left(\frac{r_j^{\frac{1}{\sigma}} + r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma} + \sum_{j=0}^{J} b_{jj}^{\Im} r_j^{\tau}$ which, with fixed (σ,τ) has at least (J+1)(J+2)/2 parameters in B in the distribution of consumer tastes as well as the parameters θ_1 in the transformed function $r_j = e^{\delta_j(x_j,\xi_j;p_j,y;\theta_1)}$ 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 $\frac{1}{\tau}\ln H(r;\theta,\Im) + \gamma$ 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 (1966)'s 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[16] by following the approach recently suggested by Pinkse et al. (2002).[17] 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: $b_{jkt} = \begin{cases} 1/d_{jk}(x_j, x_k; \alpha_1) & if \quad j \neq k \\ \exp(x_j'\alpha_2) & j = k \end{cases}$ where $d_{jk}(x_j, x_k; \alpha_1) = \left(\sum_{l=1}^{L} \alpha_{1l} (x_{lj} - x_{lk})^2 \right)^2$ measure the "*distance*" between products j and k in characteristics space and $(\alpha_{1l}, \alpha_{2l})$ are the parameters to estimate[18]. 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..[19, 20]

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 non-linear parameters,[21] those in the distribution of tastes, $\theta_2 = (B, \tau, \sigma)$. Specifically, for any given value of θ_2, the vector of

[16] See section 5 for an explicit application.

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

[18] Notice that the parameters α_1 are identified up to their sign.

[19] 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 $F(\varepsilon_0, ..., \varepsilon_J; \theta_2, \Im) = \exp\{-H(e^{-\varepsilon_0}, ..., e^{-\varepsilon_J}; \theta_2, \Im)\}$. 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.

[20] 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 sub-vector 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 sub-vector 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 well 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.

[21] Alternatively, a one-step estimator following Dube, Fox and Che-Lin Su (2011) could be used.

r_t's are chosen so that predicted market shares are equal to actual market shares in each period (or market) *t*. Since $r_{jt} = e^{\delta_{jt}(w_j, p_j, y; \theta)}$, defining $\delta_{jt} = \ln r_{jt}$ we can compute δ_{jt} for each product, run a Berry (1994) style regression based on the linear relationship $\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$ and then employ a generalized method of moments (GMM) estimator via a forming of conditional moment restrictions, $E[\xi_{jt}(\theta^*) \mid z_{jt}] = 0$, where $\theta = (\theta_1, \theta_2)$, $\xi_{jt}(\theta) = \delta_{jt}(\theta) - x'_{jt}\beta_j + \alpha p_{jt}$, and z_{jt} denotes the $(1 \mathrm{x} q)$ vector of available instruments. Specifically, we define $G_n(\theta) = \frac{1}{n}\sum_{t=1}^{T}\sum_{j=1}^{J}\tilde{\xi}_{jt}(\theta)\tilde{z}'_{jt}$ to formulate the GMM criterion function to minimize where $\tilde{\xi}_{jt} = \xi_{jt}\chi_{jt}$, $\tilde{z}_{jt} = (z_{1jt}\chi_{jt}, ..., z_{qjt}\chi_{jt})$, $n = \sum_{j=1}^{J}\sum_{t=1}^{T}\chi_{jt}$ and $\chi_{jt} = 1$ if product j is sold in period (or market) t and zero otherwise so that χ_{jt} provides a missing value indicator. [22]

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 datasets 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_{j0} 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 datasets, we do not observe such variation. A natural restriction appears to be to

[22] 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*.

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_{0j} for all j inside goods. In estimating the model, we therefore impose the symmetry conditions that $b_{0j} = b_{j0}$ 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 τ, σ 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 non-linear 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 τ above but close to 1, so we set $\tau = 1.1$ while $\sigma = 0.5$ appears to be a natural choice.[23]

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 $\left.\frac{\partial E[Z'\xi(\theta)]}{\partial \theta'}\right|_{\theta=\theta^*}$ has rank equal to its number of columns (see Rothenberg, 1971).[24] In our context, we have

$$\left.\frac{\partial}{\partial \beta'} E\left(Z'\xi(\theta)\right)\right|_{\theta=\theta^*} = -E\left(Z'X\right) \quad (1)$$

$$\left.\frac{\partial}{\partial \alpha} E\left(Z'\xi(\theta)\right)\right|_{\theta=\theta^*} = E\left(Z'p\right) \quad (2)$$

[23] 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.

[24] See also the survey paper by Newey and McFadden (1994). Non-parametric identification of demand in a general set-up is shown in Berry & Haile (2012).

$$\frac{\partial}{\partial \theta_2} E\left(Z'\xi(\theta)\right)\Big|_{\theta=\theta^*} = -E\left(Z'\left(\frac{\partial s(\delta, \theta_2, \Im^+)}{\partial \delta'}\right)^{-1}\left(\frac{\partial s(\delta, \theta_2, \Im^+)}{\partial \theta_2'}\right)\right)\Bigg|_{\theta=\theta^*} =$$

$$= -E\left(Z'\begin{pmatrix} \left(\frac{\partial s(\delta_1, \theta_2, \Im^+)}{\partial \delta_1'}\right)^{-1} & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & \left(\frac{\partial s(\delta_{T-1}, \theta_2, \Im^+)}{\partial \delta_{T-1}'}\right)^{-1} & 0 \\ 0 & 0 & 0 & \left(\frac{\partial s(\delta_T, \theta_2, \Im^+)}{\partial \delta_T'}\right)^{-1} \end{pmatrix}\begin{pmatrix} \left(\frac{\partial s(\delta_1, \theta_2, \Im^+)}{\partial \theta_2'}\right) \\ \vdots \\ \vdots \\ \left(\frac{\partial s(\delta_T, \theta_2, \Im^+)}{\partial \theta_2'}\right) \end{pmatrix}\right)\Bigg|_{\theta=\theta^*} \quad (3)$$

where $\xi(\theta) = (\xi_1(\theta)', ..., \xi_T(\theta)')$ and in equation (3) each term $\frac{\partial \xi_t(\theta)}{\partial \theta_2'} = \left(\frac{\partial s(\delta_t, \theta_2, \Im^+)}{\partial \delta_t'}\right)^{-1}\left(\frac{\partial s(\delta_t, \theta_2, \Im^+)}{\partial \theta_2'}\right)$ 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 θ_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. $\alpha = (\alpha_1, \alpha_2)$, 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 $b_{00t} = 1$. Please see appendix A for further derivation and discussion.

Berry & Haile (2012)'s 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 θ to minimize $G_n(\theta)'AG_n(\theta)$ possibly subject to the linear restrictions $R\theta = r$ then, under standard regularity conditions:

$$\sqrt{n}(\hat{\theta} - \theta_0) \sim N(0, D\Gamma AV_G\Gamma' D)$$

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 θ, $\Gamma = (\Gamma_{1n}, \Gamma_{2n}) = (\frac{\partial G_n(\theta)}{\partial \theta_1}, \frac{\partial G_n(\theta)}{\partial \theta_2})$ and $V_G = Var[G_n(\theta)]$ 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.[25]

Finally, to solve the *J*-dimensional vector of share equations to compute δ_j, we use a slightly amended BLP contraction mapping: $\delta_j^{k+1} = \delta_j^k + \rho\left(\ln s_j^{obs} - \ln\left(s_j\left(r\left(\delta^k\right)\right)\right)\right)$ for *j=0,1,..,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.

[25] See the web-appendix for a detailed discussion of the two cases.

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 underling 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 datasets involving a small number of products from a larger number of markets and a second specification of the model appropriate for datasets 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) non-price 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 $\beta_1^{DGP} = ... = \beta_{L-1}^{DGP} = (1,1..,1,-1)'$ and $\beta_{price}^{DGP} = -1$. 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 $\frac{1}{J}\sum_{k \neq j} \frac{(x_j - x_k)^2}{\left(\sum_{l=1}^{L}(x_{jl} - x_{kl})^2\right)^3}$, 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 we 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 ξ_{jt} 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,..,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.[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.[27] This specification sets the true data generating taste parameters b_{jk}^{DGP} to 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 $b_{00} \equiv 1$.[28]

	J=3, T=100		J=3, T=500		J=3, T=1,000	
Parameter	**Bias**	**MSE**	**Bias**	**MSE**	**Bias**	**MSE**
β_1	-0.06454	0.01019	-0.00683	0.0005	-0.00409	0.00022
β_2	-0.07707	0.01198	-0.00252	0.0007	-0.00157	0.00028
β_{price}	0.08093	0.01768	0.00528	0.0013	-0.00111	0.00073
$b_{00} \equiv 1$	n/a	n/a	n/a	n/a	n/a	n/a
b_{10}	0.596	0.835	0.038	0.070	0.034	0.031
b_{21}	0.439	0.831	0.067	0.060	0.007	0.024
b_{01}	-0.083	0.454	0.009	0.011	-0.006	0.002
b_{11}	-0.399	0.372	-0.049	0.023	-0.004	0.012
b_{21}	0.101	0.861	0.023	0.028	-0.004	0.023
b_{02}	0.066	0.314	-0.021	0.010	0.001	0.002
b_{12}	-0.215	0.388	-0.001	0.026	0.001	0.012
b_{22}	-0.231	0.466	-0.019	0.040	0.005	0.011

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.

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

[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.
[27] See Coublucq (2010) for a first attempt to address this issue.
[28] We also included the constraints that each b_{jk} 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.

significant small sample biases in the estimates of b_{21} - but these reduce dramatically as sample sizes grow. Second, the 'linear' parameters in β 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 3 exogenous characteristics (L=4) drawn form $N(1,1)$, we allow for two additional characteristics that only shift costs, $w_{jt}^{DGP} + 1.1\sum_l |x_{jlt}^{DGP}|$, where each element of w_{jt}^{DGP} is independently drawn from a uniform distribution U(0, 2). The unobserved demand and cost characteristics are drawn from a bivariate normal distribution

$$\begin{bmatrix} \xi_{jt} \\ \omega_{jt} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.5 & 0.7 \\ 0.7 & 0.5 \end{bmatrix} \right)$$

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

[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 $b_{10} = b_{01}$ or $b_{20} = b_{02}$ 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 $\partial s_1 / \partial p_j$ to identify the (unobserved) variation in $\partial s_j / \partial p_1$. Unreported results show that when the restrictions $b_{10} = b_{01}$ and $b_{20} = b_{02}$ 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_{12} and b_{22} fall to 0.113 and 0.006 respectively. The variance and MSE's of the estimated B parameters also fall. The small sample bias in β also falls considerably (to -0.00732, -0.02196 and 0.02808 if T=100.)

	J=35, T=3		J=70, T=3		J=120, T=3		J=35, T=10		J=35, T=20	
Parameters	**Bias**	**MSE**	**Bias**	**MSE**	**Bias**	**MSE**	**Bias**	**MSE**	**Bias**	**MSE**
β_1	0.0463	0.0360	0.0056	0.0097	0.0076	0.0029	-0.0196	0.0122	-0.0034	0.0052
β_2	0.0344	0.0323	-0.0047	0.0101	-0.0014	0.0031	0.0145	0.0082	-0.0012	0.0069
β_3	-0.0869	0.0267	0.0102	0.0068	-0.0037	0.0034	0.0134	0.0121	-0.0073	0.0057
β_{price}	-0.0782	0.0070	0.0028	0.0003	0.0012	0.0001	-0.0009	0.0002	0.0019	0.0001
α_{11}	0.0463	0.0786	0.0065	0.0256	0.0029	0.0114	0.0373	0.0247	0.0119	0.0123
α_{12}	0.0344	0.0682	0.0192	0.0298	0.0237	0.0143	-0.0175	0.0188	0.0025	0.0133
α_{13}	0.0229	0.0690	-0.0045	0.0225	0.0146	0.0172	-0.0139	0.0247	0.0111	0.0121
α_{21}	-0.0869	0.1176	-0.0537	0.0518	-0.0076	0.0404	0.0768	0.0556	-0.0318	0.0238
α_{22}	-0.0782	0.1284	-0.0426	0.0512	-0.0224	0.0366	-0.0391	0.0386	-0.0172	0.0237
α_{23}	-0.0307	0.118	0.0151	0.0587	-0.0431	0.0413	-0.0123	0.04	-0.0023	0.0316

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

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 α_1 and α_2 requires systematically larger samples than is required to estimate the 'linear' parameters, $(\beta_1, \beta_2, \beta_3, \beta_{price})$.[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 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

[30] Examination of the underlying Monte-Carlo runs makes clear that the immediate cause of the reported small sample bias in α_1 and α_2 is a tendency on occasion for the optimization routine to push an individual parameter estimate in α_1 or α_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 α_1 and α_2 does not appear to contaminate the estimates of β.

$\beta_{price,it} = \frac{\beta_{price}}{y_{it}}$ where $\beta_{price} = -1$ and y_{it} is drawn from a *log-N(μ_t ,σ_t)* where (μ_t ,σ_t) are market specific and drawn from U(0,1). We simulate the integral in the market share equation with ns = 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 α_{1l} as before.[31]

	Own – price elasticity			Cross – price elasticity		
	Bias	**MSE**	**Truth**	**Bias**	**MSE**	**Truth**
T= 3, J=35	-0.0613	0.3453	-3.8869	-0.0091	0.0012	0.0692
T=10, J=35	-0.0630	0.2838	-3.8951	-0.0087	0.0010	0.0693
T=10, J=25	-0.1570	0.3998	-3.8807	-0.0099	0.0020	0.0918

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

Table 4 confirms the ability of the model in matching quite well, the matrix of own- and cross-price elasticities arising from the RC-MNL models. 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 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.

6 Italian Automobile market 1991-1999

We use the model studied in section 3 to estimate the demand for new automobiles in Italy between 1991 and 1999. [32] 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)

[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.
[32] The data set has been used in Goldberg and Verboven (2001) and it is available upon request.

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.

years	no. of models	quantity '000	price in €	size	l x 100 km	engine size(cc)	weight	Nominal GDP per capita €	inflation
1991	77	2,233	7,810	632.56	7.83	1184.17	866.46	12,992	1.0000
1992	76	2,273	8,740	634.76	8.03	1271.61	887.05	13,647	1.0466
1993	78	1,775	9,426	634.93	8.07	1261.55	898.30	14,034	1.0925
1994	77	1,613	9,981	640.90	8.17	1259.05	923.83	14,795	1.1302
1995	80	1,556	11,097	645.01	8.28	1272.85	940.14	15,974	1.1875
1996	84	1,518	12,249	659.29	8.39	1296.22	960.56	16,855	1.2465
1997	85	2,166	11,500	644.91	8.24	1253.70	934.46	17,515	1.2789
1998	87	2,065	12,137	655.54	9.09	1269.25	958.02	18,268	1.2946
1999	91	1,874	12,408	655.81	9.14	1253.60	989.47	19,110	1.3164

Table 5: Descriptive Statistics.

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, $G(\theta) = \frac{1}{n} Z' \xi(\theta)$ so that the GMM estimator is $\hat{\theta} = \arg\min_{\theta} G(\theta)' W^{-1} G(\theta)$, 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.

Parameters	Variables	parameter estimate	standard Error
linear parameters in δ	Constant	-13.64	3.28
	Fuel Consumption (L/100 Km)	-2.88	0.50
	Log(size)	3.95	2.15
	Log(cc)	4.95	1.22
	Price	-2.52	0.35
	Manufacturer Location dummies	Yes	-
parameters in B			
diagonal α_2			
	Fuel	3.55	0.55
	Log(size)	-6.10	2.62
	Log(cc)	-4.63	1.20
off-diagonal α_1			
	Fuel	4.27	7.73
	Log(size)	4.82	0.06
	Log(cc)	2.23	0.30

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

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).

Own Price Elasticity			Within-Market Segment Cross-Price Elasticity
Subcompact		-1.59	0.0066
	Domestic	-1.67	
	Foreign	-1.48	
Compact		-2.54	0.0050
	Domestic	-2.62	
	Foreign	-2.36	
Mid-Size		-3.32	0.0143
	Domestic	-3.66	
	Foreign	-2.87	
Sedan		-4.39	0.0015
	Domestic	-4.58	
	Foreign	-3.99	

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

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.

	BMW 3	Citroen XM	Fiat Uno	Honda Civic	Mitsubishi Lancer	Opel Vectra	Rover Montego	Suzuki Swift
BMW 3	-3.179536	0.000320	0.008156	0.000069	0.000016	0.000815	0.000020	0.000052
Citroen XM	0.001182	-4.843053	0.007156	0.000069	0.000117	0.000815	0.000020	0.000052
Fiat Uno	0.000020	0.000030	-1.199981	0.000069	0.000016	0.000815	0.000020	0.000058
Honda Civic	0.001082	0.000320	0.007174	-2.442549	0.000017	0.000817	0.000020	0.001378
Mitsubishi lancer	0.001182	0.003320	0.007156	0.000070	-2.931113	0.000822	0.000086	0.000052
Opel Vectra	0.002182	0.000320	0.007156	0.000069	0.000017	-2.060492	0.000047	0.000052
Rover Montego	0.001183	0.000320	0.007156	0.000069	0.000017	0.001930	-2.537783	0.000052
Suzuki Swift	0.001182	0.000321	0.007917	0.001814	0.000016	0.000815	0.000020	-2.099886

Table 8: Sample of own- and cross-price elasticities.

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 real-world dataset. By way of a closing remark, we note that the model could equally be estimated using individual choice data.

Appendix A: Proofs.

Proof to Corollary 1.

Part (i) follows since:

$$\left(\frac{\partial \ln s_j}{\partial \ln p_k}\right)=\sum_{l=0}^{J}\left(\frac{\partial \ln s_j}{\partial \ln r_l}\right)\left(\frac{\partial \ln r_l}{\partial \ln p_k}\right)=\sum_{l=0}^{J}\left(\frac{\partial \ln s_j}{\partial \ln r_l}\right)\left(\frac{\partial \delta_l}{\partial \ln p_k}\right)=\frac{\partial \ln s_j}{\partial \ln r_k}\frac{\partial \delta_k}{\partial \ln p_k}$$

since $\dfrac{\partial \delta_l\,(w_l, y-p_l 1(l>0);\theta_1)}{\partial \ln p_k}=0$ for all $l\neq k$. Defining, $r=(r_0,..,r_J)$, we have $s_j(r)=\dfrac{r_j H_j(r)}{\tau H(r)}$, so $\ln s_j(r)=\ln r_j+\ln H_j(r)-\ln \tau H(r)$ and hence

$\dfrac{\partial \ln s_j}{\partial \ln r_k}=I(k=j)+\dfrac{r_k H_{jk}(r)}{H_j(r)}-\tau\dfrac{r_k H_k(r)}{\tau H(r)}$ and so for each inside good price, $k>1$,

$$\left(\frac{\partial \ln s_j}{\partial \ln p_k}\right)=\frac{\partial \ln s_j}{\partial \ln r_k}\frac{\partial \delta_k}{\partial \ln p_k}=\left(I(k=j)+\frac{r_k H_{jk}(r)}{H_j(r)}-\tau\frac{r_k H_k(r)}{\tau H(r)}\right)\frac{\partial \delta_k}{\partial \ln p_k}$$

$$=\left(I(k=j)+\frac{r_k H_{jk}(r)}{H_j(r)}-\tau s_k(r)\right)\frac{\partial \delta_k}{\partial \ln p_k}$$

Part (ii) follows similarly since $\left(\dfrac{\partial \ln s_j}{\partial \ln y}\right)=\sum_{l=0}^{J}\left(\dfrac{\partial \ln s_j}{\partial \ln r_l}\right)\left(\dfrac{\partial \delta_l}{\partial \ln y}\right)$ and so

$$\left(\frac{\partial \ln s_j}{\partial \ln y}\right)=\sum_{l=0}^{J}\left(I(j=l)+\frac{r_l H_{jl}(r)}{H_j(r)}-\tau\frac{r_l H_l(r)}{\tau H(r)}\right)\frac{\partial \delta_l}{\partial \ln y}$$

$$=\left(\frac{\partial \delta_j}{\partial \ln y}+\frac{1}{H_j(r)}\sum_{l=0}^{J} r_l H_{jl}(r)\frac{\partial \delta_l}{\partial \ln y}-\tau\sum_{l=0}^{J} s_l(r)\frac{\partial \delta_l}{\partial \ln y}\right). \quad \blacksquare$$

Proof to Lemma1:

This essentially follows immediately from the expression $\left(\dfrac{\partial \ln s_j}{\partial \ln p_k}\right)=\left(I(j=k)-s_k(r)\right)\alpha_k p_k$ provided exists a unique vector r to ensure that $s_k(r)=s_k^*$. Then, for $j=k$, $\left(\dfrac{\partial \ln s_j}{\partial \ln p_j}\right)^*=\left(1-s_j^*\right)\alpha_j p_j$ and so we need only to choose

$\alpha_j=\dfrac{1}{\left(1-s_j^*\right)}p_j\left(\dfrac{\partial \ln s_j}{\partial \ln p_j}\right)^*=\dfrac{1}{\left(1-s_j^*\right)s_j^*}\left(\dfrac{\partial s_j}{\partial p_j}\right)^*$. Notice that since $s_j^*(1-s_j^*)\geq 0$, $sign(\alpha_j)=sign\left(\left(\dfrac{\partial s_j}{\partial p_j}\right)^*\right)$. $\blacksquare$

Proof to Lemma 2:

First, note that for all $\tau > 0$ the function $H(r;\theta_2,\Im) = \sum_{j=0}^{J}\sum_{k\neq j}\left(\frac{b_{jk}+b_{kj}}{2}\right)\left(\frac{r_j^{\frac{1}{\sigma}}+r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma} + \sum_{j=0}^{J} b_{jj}^{\Im} r_j^{\tau}$ is homogeneous of degree τ on $r \in \Re_+^{J+1}$. Next note that $H(r;\theta_2,\Im) \geq 0$ on $r \in \Re_+^{J+1}$ if $b_{jk} \geq 0$ for all $j,k = 0,1,...,J$. Third, we must show that the cross derivative properties 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. $\frac{\partial H(r;\theta_2,\Im)}{\partial r_j} = H_j(r;\theta_2,\Im) = \tau\sum_{k\neq j}\left(\frac{b_{jk}+b_{kj}}{2}\right)\left(\frac{r_j^{\frac{1}{\sigma}}+r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-1} r_j^{\frac{1}{\sigma}-1} + \tau b_{jj}^{\Im} r_j^{\tau-1} \geq 0$

2. $\frac{\partial^2 H(r;\theta_2,\Im)}{\partial r_j \partial r_k} = H_{jk}(r;\theta_2,\Im) = \frac{\tau}{2\sigma}(\tau\sigma-1)\left(\frac{b_{jk}+b_{kj}}{2}\right)\left(\frac{r_j^{\frac{1}{\sigma}}+r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-2} r_j^{\frac{1}{\sigma}-1} r_k^{\frac{1}{\sigma}-1} \leq 0$

provided $\tau \geq 0$, $\sigma > 0$, $\tau\sigma \leq 1$ and $b_{jk} \geq 0$ for all $j,k \in \Im$.

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 non-positive. Although no sign restriction on the second own-derivative is required, for completeness we note here that

$$H_{jj}(r_{\Im};\theta_2,\Im) = \tau\sum_{k\neq j}\left(\frac{b_{jk}+b_{kj}}{2}\right)\left(\frac{r_j^{\frac{1}{\sigma}}+r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-1} r_j^{\frac{1}{\sigma}-2}\left(\frac{1}{\sigma}-1\right) +$$

$$+\tau(\tau\sigma-1)\sum_{k\neq j}\left(\frac{b_{jk}+b_{kj}}{2}\right)\left(\frac{r_j^{\frac{1}{\sigma}}+r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-2} r_j^{\frac{1}{\sigma}-1} r_j^{\frac{1}{\sigma}-1} \frac{1}{2\sigma} + \tau(\tau-1) b_{jj}^{\Im} r_j^{\tau-2}$$

$$= \tau\sum_{k\neq j}\left(\frac{b_{jk}+b_{kj}}{2}\right)\left(\frac{r_j^{\frac{1}{\sigma}}+r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-2}\left(\left(\frac{r_j^{\frac{1}{\sigma}}+r_k^{\frac{1}{\sigma}}}{2}\right)\left(\frac{1}{\sigma}-1\right)+\frac{(\tau\sigma-1)}{2\sigma} r_j^{\frac{1}{\sigma}}\right) r_j^{\frac{1}{\sigma}-2} + \tau(\tau-1) b_{jj}^{\Im} r_j^{\tau-2}$$

$$= \frac{\tau}{2\sigma}\sum_{k\neq j}\left(\frac{b_{jk}+b_{kj}}{2}\right)\left(\frac{r_j^{\frac{1}{\sigma}}+r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-2}\left(\left(r_j^{\frac{1}{\sigma}}+r_k^{\frac{1}{\sigma}}\right)(1-\sigma)+(\tau\sigma-1)\, r_j^{\frac{1}{\sigma}}\right) r_j^{\frac{1}{\sigma}-2} + \tau(\tau-1) b_{jj}^{\Im} r_j^{\tau-2}$$

Finally, we consider the limit condition $\lim_{r_j \to +\infty} H(r;\theta_2,\Im) = +\infty$ which is clearly satisfied since $H(r;\theta_2,\Im) = \sum_{j=0}^{J}\sum_{k\neq j}\left(\frac{b_{jk}+b_{kj}}{2}\right)\left(\frac{r_j^{\frac{1}{\sigma}}+r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma} + \sum_{j=0}^{J} b_{jj}^{\Im} r_j^{\tau} \geq \sum_{j=0}^{J} b_{jj}^{\Im} r_j^{\tau}$ while $\lim_{r_j \to +\infty} \sum_{j=0}^{J} b_{jj}^{\Im} r_j^{\tau} = +\infty$ provided $\tau > 0$ and $b_{jj}^{\Im} > 0$. ■

Proof to Lemma 3.

We begin with the J+1 equations $s_j^{obs} = f_j(r;\Im)$, for $j \in \Im$ where $f_j(r;\Im)$ is homogenous of degree $\tau > 0$ in r.

First notice that if $s_j^{obs} = 0$ for any good, j, condition (i) on the $f_j(r;\Im)$ function ensures that setting $r_j = 0$ will solve the j[th] equation exactly. Having set the r's corresponding to products with zero market shares equal to zero, we can progress to consider the solution to the smaller set of equations $s_k^{obs} = f_k(r_{\Im^+};\Im^+)$ for $k \in \Im^+$, defined as in the lemma. Henceforth, we can work entirely with the functions for the reduced set of products $\Im^+$. 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 $\delta_k \equiv \ln r_k$, namely define $g_j(\delta) \equiv f_j(r(\delta)) - s_j^{obs} = f_j(e^{\delta}) - s_j^{obs}$ for $j \in \Im$. Next, recall that a sufficient condition for uniqueness of a system of equations $g(\delta) = 0$ is that the Jacobian matrix of a function $D_\delta\, g(\delta)$ 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 $D_\delta\, g(\delta)$ has a dominant diagonal if $\left|\frac{\partial g_j(\delta)}{\partial \delta_j}\right| > \sum_{k\neq j}\left|\frac{\partial g_j(\delta)}{\partial \delta_k}\right|$ for each $j \in \Im$. Note that if $\frac{\partial g_j(\delta)}{\partial \delta_j} \geq 0$ and $\frac{\partial g_j(\delta)}{\partial \delta_k} \leq 0$ for all $k \neq j$ with $j,k \in \Im$ then the dominant diagonal condition can be written in terms of the function as $\frac{\partial g_j(\delta)}{\partial \delta_j} > \sum_{k\neq j}\left(-1 * \frac{\partial g_j(\delta)}{\partial \delta_k}\right)$.

To establish that the dominant diagonal condition holds under the conditions in the lemma, note that for any $j \in \Im^+$, since $f_j(r)$ is homogenous of degree $\tau > 0$ in r, by Euler's theorem, $\tau f_j(r) = \sum_{k\in\Im^+} r_k \frac{\partial f_j(r)}{\partial r_k}$, so that $\frac{\partial f_j(r)}{\partial \ln r_j} = \tau f_j(r) - \sum_{k\neq j}\frac{\partial f_j(r)}{\partial \ln r_k}$ or $\frac{\partial f_j(e^{\delta})}{\partial \delta_j} = \tau f_j(e^{\delta}) - \sum_{k\neq j}\frac{\partial f_j(e^{\delta})}{\partial \delta_k}$. Under the condition (iii) in the lemma, for each $j \in \Im^+$, $\tau f_j(e^{\delta}) > 0$, we have $\frac{\partial f_j(e^{\delta})}{\partial \delta_j} > -\sum_{k\neq j}\frac{\partial f_j(e^{\delta})}{\partial \delta_k}$ with $\frac{\partial f_j(e^{\delta})}{\partial \delta_k} \leq 0$ for $k \in \Im^+ / j$ by condition

(iv). Thus $\frac{\partial f_j(r)}{\partial \delta_j} > 0$. This is of course exactly the dominant diagonal condition for the function, $g_j(\delta; s_j^{obs}) = f_j(e^{\delta}) - s_j^{obs}$ for all $j \in \Im^+$, i.e., on the set of products with $s_j^{obs} > 0$. Thus there is a unique $\delta_{\Im^+}$ that solves the system of equations $f_j(e^{\delta}; s^*) = s_j^* \quad j \in \Im^+$, i.e., that equates predicted and actual market shares for products with strictly positive observed market shares. Since there is a unique one-to-one mapping between $\delta_{\Im^+}$ and $r_{\Im^+}$, this in turn implies that there is a unique $r_{\Im^+}$ that solves the equations. We have already noted that for those products with zero shares, setting their respective r's to zero solves their equations. Thus, as the lemma claims, there is (i) a solution to the full problem and (ii) a unique solution to the reduced problem with zero market share goods having had their r's set to zero (or equivalently their delta's – utility - set to minus infinity.)

For completeness, is useful also to follow the logic in Berry (1994), and argue that all r's will be finite provided $s_j^{obs} < 1$ for all $j \in \Im$. 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 $s_j^{obs} < 1$ we know that if a solution to the system of equations exists, then at the solution $r_j < \infty$ for every $j \in \Im$. Thus we need only consider whether a solution exists on a closed bounded set, $r \in [0, M]^{J+1} \subset \Re_+^{J+1}$ for some sufficiently large $M < \infty$. For any functions $f_j(r)$ that are continuous on the set $[0, M]^{J+1} \subset \Re_+^{J+1}$, it also follows immediately that there is a solution by Brower's fixed point theorem to the set of equations on the restricted domain, and hence on the unrestricted domain.■

Proof to Corollary 2.

For brevity, we drop the explicit dependence of functions on the parameters, θ_2. First, define

$$f_j(r_{\Im^+}; \Im^+) \equiv f_j(r_{\Im^+}, 0, .., 0; \Im) = \tau \sum_{k \neq j, k \in \Im^+} b_{jk} \left(\frac{r_j^{\frac{1}{\sigma}} + r_k^{\frac{1}{\sigma}}}{2} \right)^{\tau\sigma - 1} r_j^{\frac{1}{\sigma}} + \tau \sum_{k \neq j, j, k \in \Im / \Im^+} b_{jk} \left(\frac{r_j^{\frac{1}{\sigma}} + 0}{2} \right)^{\tau\sigma - 1} r_j^{\frac{1}{\sigma}} + \tau b_{jj}^{\Im} r_j^{\tau}$$

$$= \tau \sum_{k \neq j, j, k \in \Im^+} b_{jk} \left(\frac{r_j^{\frac{1}{\sigma}} + r_k^{\frac{1}{\sigma}}}{2} \right)^{\tau\sigma - 1} r_j^{\frac{1}{\sigma}} + \tau b_{jj}^{\Im^+} r_j^{\tau}$$

where $b_{jj}^{\Im^+} \equiv 2^{1-\tau\sigma} \left(\sum_{k \neq j, k \in \Im / \Im^+} b_{jk} \right) + b_{jj}^{\Im}$. We will apply Lemma 3 in order to establish that there is a solution r* to the J+1 equations, $s_j^{obs} = r_j N_j(r; \Im)$, $j \in \Im$. Since $1 = \sum_{j=0}^{J} s_j^{obs} = \sum_{j=0}^{J} r_j N_j(r; \Im)$, it will then follow immediately that r* will also solve the market predicted equals actual market share equations, $s_j^{obs} = \frac{r_j N_j(r; \Im)}{\sum_{l=0}^{J} r_l N_l(r; \Im)}$, $j \in \Im$. (Note that we

are effectively imposing the normalization $\sum_{l=0}^{J} r_l N_l(r;\Im) = 1$ instead of the more familiar normalization $r_0 = 1$.[33] To derive a solution with the usual normalization, just rescale the r's being careful to use the full formula: $s_j(r) = \dfrac{r_j N_j(r;\Im)}{\sum_{l=0}^{J} r_l N_l(r;\Im)}$ which is homogeneous of degree zero in r.)

Step 1. First, we establish condition (i) in Lemma 3. Notice that since $r_j N_j(r_\Im;\Im) = \tau \sum_{k \neq j} b_{jk} \left(\dfrac{r_j^{\frac{1}{\sigma}} + r_k^{\frac{1}{\sigma}}}{2} \right)^{\tau\sigma - 1} r_j^{\frac{1}{\sigma}} + \tau b_{jj}^{\Im} r_j^{\tau}$, setting any $r_j = 0$ ensures that $r_j N_j(r_\Im;\Im) = 0$ provided $\sigma > 0$, $\tau > 0$ and $\tau\sigma - 1 < 0$, since for $\sigma\tau < 1$, $\left(\dfrac{2}{r_j^{\frac{1}{\sigma}} + r_k^{\frac{1}{\sigma}}} \right)^{1-\tau\sigma} r_j^{\frac{1}{\sigma}} \leq \left(\dfrac{2}{r_j^{\frac{1}{\sigma}}} \right)^{1-\tau\sigma} r_j^{\frac{1}{\sigma}} = 2^{1-\tau\sigma} r_j^{\frac{1}{\sigma} - \left(\frac{1-\tau\sigma}{\sigma}\right)} = 2^{1-\tau\sigma} r_j^{\frac{1}{\sigma} - \left(\frac{1-\tau\sigma}{\sigma}\right)} = 2^{1-\tau\sigma} r_j^{\tau}$ which converges to zero as $r_j \to 0$ provided $\tau > 0$. In turn $r_j N_j(r;\Im) = 0$ 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, $b_{jj}^{\Im}$, is affected by a term 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 $s_j^{obs} > 0$, then the solution to $s_j^{obs} = s_j(r)$ must involve setting $r_j^*(s) > 0$; the predicted market share with $r_j = 0$ would be zero and so smaller than the observed market share.

[33] Imposing this alternative normalization also makes the analysis of some models particularly simple. For example, in the MNL model $s_j(r) = r_j \Big/ \sum_{l=0}^{J} r_l$ so that the normalization $H(r) = \sum_{l=0}^{J} r_l = 1$ means we must find a solution to the entirely trivial set of equations: $s_j^{obs} = r_j$ for all $j = 0,1,..,J$. 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 the choice of normalization. Since $r_j = e^{\delta_j}$, scaling the r's by a constant $\lambda > 0$ is essentially adding the constant $\ln\lambda$ to the vector δ, 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 $H(r) = 1$.

Step 2. Next, we establish condition (ii) of Lemma 3 holds, namely that $\lim_{r_j\to\infty} f_j(r;\Im)\geq 1$.

Recall $f_j(r;\Im)=\tau\sum_{k\neq j,k\in\Im} b_{jk}\left(\frac{r_j^{\frac{1}{\sigma}}+r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-1} r_j^{\frac{1}{\sigma}}+\tau b_{jj}^{\Im} r_j^{\tau}$, so that

$$\lim_{r_j\to\infty} f_j(r;\Im)=\lim_{r_j\to\infty}\left(\tau\sum_{k\neq j,k\in\Im} b_{jk}\left(\frac{r_j^{\frac{1}{\sigma}}+r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-1} r_j^{\frac{1}{\sigma}}+\tau b_{jj}^{\Im} r_j^{\tau}\right)\geq \lim_{r_j\to\infty}\left(\tau b_{jj}^{\Im} r_j^{\tau}\right)=\infty$$

where the inequality follows since the first term is weakly positive and the limit provided $\tau>0$ and $b_{jj}^{\Im}>0$ for each $j\in\Im$.

Step 3. Condition (iii) of Lemma 3, $f_j(r;\Im)>0$ when $r_j>0$, holds immediately since $f_j(r;\Im)=\tau\sum_{k\neq j,k\in\Im} b_{jk}\left(\frac{r_j^{\frac{1}{\sigma}}+r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-1} r_j^{\frac{1}{\sigma}}+\tau b_{jj}^{\Im} r_j^{\tau}\geq 0$, strict if $b_{jj}^{\Im}>0$ and $\tau>0$ when $r_j>0$.

Step 4. Finally, we establish condition (iv) of Lemma 3 holds, $\frac{\partial f_j(r;\Im^+)}{\partial \ln r_k}\leq 0$ for all $k\in\Im^+/j$. In fact, it will be algebraically easier and equivalent to work with $\frac{\partial f_j(r;\Im^+)}{\partial \ln r_k}=r_k\frac{\partial f_j(r;\Im^+)}{\partial r_k}$.

$f_j(r;\Im^+)=\tau\sum_{k\neq j,k\in\Im^+} b_{jk}\left(\frac{r_j^{\frac{1}{\sigma}}+r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-1} r_j^{\frac{1}{\sigma}}+\tau b_{jj}^{\Im^+} r_j^{\tau}$, so for $k\in\Im^+/j$,

$$\frac{\partial f_j(r;\Im^+)}{\partial \ln r_k}=r_k\frac{\partial f_j(r;\Im^+)}{\partial r_k}=r_k\frac{\tau(\tau\sigma-1)}{2\sigma}b_{jk}\left(\frac{r_j^{\frac{1}{\sigma}}+r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-2} r_j^{\frac{1}{\sigma}}r_k^{\frac{1}{\sigma}-1}$$

$$=\frac{\tau(\tau\sigma-1)}{2\sigma}b_{jk}\left(\frac{r_j^{\frac{1}{\sigma}}+r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-2} r_j^{\frac{1}{\sigma}}r_k^{\frac{1}{\sigma}}\leq 0$$

provided $\tau\sigma-1\leq 0$, $\sigma>0$ and $\tau>0$. ■

Proof to Proposition 1.

We want to show that if we observe a vector of market shares s_j^* for $j \in \Im$, the model can match that vector of market shares and also a true matrix of own- and cross-price elasticities $\left(\frac{\partial \ln s_j}{\partial \ln p_k}\right)^*$ 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*, α and the matrix of parameters B with jk[th] element b_{jk}, so that the model satisfies the following equations:

- $s_j^* = s_j(r;\theta_2)$ for *j=0,1,…,J* (5)

- $\left(\frac{\partial \ln s_j}{\partial \ln p_k}\right)^* = \left(I(k=j) + \frac{r_k H_{jk}(r;\theta_2)}{H_j(r;\theta_2)} - \tau s_k(r;\theta_2)\right)\frac{\partial \delta_k}{\partial \ln p_k}$ for $j \in \Im$, $k \in \Im/0$ (6)

where $\frac{\partial \delta_k}{\partial p_k} = \alpha < 0$.

Step 1:

Equations (5) have 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* to the J+1 equations, $s_j^* = r_j H_j(r;\theta_2)$, $j \in \Im$. Since $1 = \sum_{j=0}^{J} s_j^* = \sum_{j=0}^{J} r_j H_j(r;\theta_2)$ which in turn ensures that $\sum_{j=0}^{J} r_j^* H_j(r^*;\theta_2) = \tau H(r^*;\theta_2) = 1$ at that solution since *H* is homogeneous of degree τ. Under Corollary 2 then we can write the inverse of equations (5) as bounded vector function, $0 \le r^*(\theta_2; s^*) < \infty$.

Equations (6) have an inverse

We continue to work with the slightly amended market share equations, with $s_j^* = r_j H_j(r;\theta_2)$. First, notice that given a vector $r^* \ge 0$ which solves the equations $s_j^* = r_j H_j(r;\theta_2)$ j=0,1,..J, we can rearrange equation (6) so to write each element b_{jk} as function of r^*, α and the true shares and price elasticities given the fixed parameters σ and τ. Substituting in $s_j^* = r_j H_j(r;\theta_2)$, $\left(\frac{\partial \ln s_j}{\partial \ln p_k}\right)^* = \frac{p_k}{s_j^*}\left(\frac{\partial s_j}{\partial p_k}\right)^*$, $\frac{\partial \delta_k}{\partial \ln p_k} = \alpha$ and rearranging simplifies our expression to give $\frac{p_k}{s_j^*}\left(\frac{\partial s_j}{\partial p_k}\right)^* = \left(\frac{r_j r_k H_{jk}(r;\theta_2)}{s_j^*} + I(k=j) - \tau s_k^*\right)\frac{\partial \delta_k}{\partial \ln p_k}$ and

hence $\left(\frac{\partial s_j}{\partial p_k}\right)^* = \left(r_j r_k H_{jk}(r;\theta_2) + \left(I(k=j) - \tau s_k^*\right)s_j^*\right)\frac{\partial \delta_k}{\partial p_k}$ or equivalently $-r_j r_k H_{jk}(r;\theta_2) = \gamma_{jk}$ where $\gamma_{jk} \equiv \left(-\left(\frac{\partial s_j}{\partial p_k}\right)^* \frac{1}{\alpha} + \left(I(k=j) - \tau s_k^*\right)s_j^*\right)$ does not depend on any parameters in B.

Cross Price Elasticities**.** For our particular member of the MEV class of models under symmetry, we have

$$\frac{\partial^2 H(r;\theta_2,\Im)}{\partial r_j \partial r_k} = H_{jk}(r;\theta_2,\Im) = \frac{\tau}{2\sigma}(\tau\sigma - 1)b_{jk}\left(\frac{r_j^{\frac{1}{\sigma}} + r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-2} r_j^{\frac{1}{\sigma}-1} r_k^{\frac{1}{\sigma}-1} \leq 0$$

so that $-r_j r_k H_{jk}(r;\theta_2) = \gamma_{jk}$ becomes

$$-r_j r_k H_{jk}(r;\theta_2) = \frac{-\tau}{2\sigma}(\tau\sigma - 1)b_{jk}\left(\frac{r_j^{\frac{1}{\sigma}} + r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-2} r_j^{\frac{1}{\sigma}} r_k^{\frac{1}{\sigma}} = \gamma_{jk}$$

for $j \neq k$, so that we can choose

$$b_{jk} = -\gamma_{jk}\frac{2\sigma}{\tau(\tau\sigma-1)}\left(\frac{r_j^{\frac{1}{\sigma}} + r_k^{\frac{1}{\sigma}}}{2}\right)^{2-\tau\sigma} r_j^{-\frac{1}{\sigma}} r_k^{-\frac{1}{\sigma}} \quad (7)$$

which allows us to solve the $(J+1)^2$-$(J+1)$ equations associated with all of the cross-price elasticities provided $r_j > 0$, $r_k > 0$, $\tau > 0$, $\sigma > 0$ and $(1-\tau\sigma) > 0$. The requirement for $r_j > 0$, $r_k > 0$ will always be satisfied following Lemma 3 provided $s_j^* > 0$ and $s_k^* > 0$.

This means that we can write the inverse functions $b_{jk}(r, \alpha, \sigma, \tau, \left(\frac{\partial s_j}{\partial p_k}\right)^*, s_k^*, s_j^*)$ for all for $j \neq k$ and that it will be bounded so $b_{jk}(r, \alpha, \sigma, \tau, \left(\frac{\partial s_j}{\partial p_k}\right)^*, s_k^*, s_j^*) < M' < \infty$ 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 $b_{jk} \geq 0$, and we have assumed $\tau\sigma < 1$ we require $\gamma_{jk} \equiv \left(-\left(\frac{\partial s_j}{\partial p_k}\right)^* \frac{1}{\alpha} - \tau s_k^* s_j^*\right) \geq 0$ for every $j \neq k$. Since we assume the DGP satisfies positive cross price effects, $\left(\frac{\partial s_j}{\partial p_k}\right)^* > 0$, 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 $\alpha = -\tau \max_{j,k} s_k^* s_j^* \left(\frac{\partial s_j^*}{\partial p_k} \right)^{-1}$. Thus we have established that a bounded inverse function exists $0 \le b_{jk}(r\ , \alpha, \sigma, \tau, \left(\frac{\partial s_j}{\partial p_k} \right)^*, s_k^*, s_j^*) < M < \infty$ for every $j \neq k$.

<u>The symmetry condition</u> requires $b_{jk} = b_{kj}$ for any $j \neq k$. Symmetry thus implies the constraint:

$$b_{jk} = \frac{2\sigma}{(1-\tau\sigma)} \left(\frac{r_j^{\frac{1}{\sigma}} + r_k^{\frac{1}{\sigma}}}{2} \right)^{2-\tau\sigma} r_j^{\frac{1}{\sigma}} r_k^{\frac{1}{\sigma}} \left(-\left(\frac{\partial \delta_k}{\partial p_k} \right)^{-1} \left(\frac{\partial s_j}{\partial p_k} \right)^* - \tau s_k^* s_j^* \right)$$

$$= \frac{2\sigma}{(1-\tau\sigma)} \left(\frac{r_k^{\frac{1}{\sigma}} + r_j^{\frac{1}{\sigma}}}{2} \right)^{2-\tau\sigma} r_k^{\frac{1}{\sigma}} r_j^{\frac{1}{\sigma}} \left(-\left(\frac{\partial \delta_j}{\partial p_j} \right)^{-1} \left(\frac{\partial s_k}{\partial p_j} \right)^* - \tau s_j^* s_k^* \right) = b_{kj}$$

which collapses to $\left(\frac{\partial \delta_k}{\partial p_k} \right)^{-1} \left(\frac{\partial s_j}{\partial p_k} \right)^* = \left(\frac{\partial \delta_j}{\partial p_j} \right)^{-1} \left(\frac{\partial s_k}{\partial p_j} \right)^*$. With $\left(\frac{\partial \delta_k}{\partial p_k} \right) = \left(\frac{\partial \delta_j}{\partial p_j} \right) = \alpha$, this constraint thus reduces down to the 'Slutsky Symmetry' restriction on the DGP that, $\left(\frac{\partial s_j}{\partial p_k} \right)^* = \left(\frac{\partial s_k}{\partial p_j} \right)^*$, which is satisfied by assumption.

<u>Own Price Elasticities.</u> For own-price elasticities, our equation $-r_j r_k H_{jk}(r;\theta_2) = \gamma_{jk}$ becomes $-r_j r_j H_{jh}(r;\theta_2) = \gamma_{jj}$ where $\gamma_{jj} \equiv \left(-\left(\frac{\partial s_j}{\partial p_j} \right)^* \frac{1}{\alpha} + (1 - \tau s_j^*) s_j^* \right)$.

We have previously established that

$$H_{jj}(r_\Im;\theta_2,\Im) = \frac{\tau}{2\sigma} \sum_{k \neq j} b_{jk} \left(\frac{r_j^{\frac{1}{\sigma}} + r_k^{\frac{1}{\sigma}}}{2} \right)^{\tau\sigma-2} \left(r_k^{\frac{1}{\sigma}} (1-\sigma) + \sigma(\tau-1)\ r_j^{\frac{1}{\sigma}} \right) r_j^{\frac{1}{\sigma}-2} + \tau(\tau-1) b_{jj}^{\Im} r_j^{\tau-2}$$

.

So that we can write our equations as:

$$-r_j r_j H_{jj}(r_\Im;\theta_2,\Im) = \frac{-\tau}{2\sigma} \sum_{k \neq j} b_{jk} \left(\frac{r_j^{\frac{1}{\sigma}} + r_k^{\frac{1}{\sigma}}}{2} \right)^{\tau\sigma-2} \left(r_k^{\frac{1}{\sigma}} (1-\sigma) + \sigma(\tau-1)\ r_j^{\frac{1}{\sigma}} \right) r_j^{\frac{1}{\sigma}} - \tau(\tau-1) b_{jj}^{\Im} r_j^{\tau} = \gamma_{jj}$$

And so we can write:

$$b_{jj}^{\Im}=\left(-\gamma_{jj}-\frac{1}{2\sigma}\sum_{k\neq j}b_{jk}\left(\frac{r_j^{\frac{1}{\sigma}}+r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-2}\left(r_k^{\frac{1}{\sigma}}(1-\sigma)+\sigma(\tau-1)\ r_j^{\frac{1}{\sigma}}\right)r_j^{\frac{1}{\sigma}}\right)\frac{\left(r_j^*\right)^{-\tau}}{\tau(\tau-1)} \qquad (8)$$

where $b_{jk}\geq 0$ for $j\neq k$ will be given by the solutions provided above. Since $\frac{1}{\alpha}\left(\frac{\partial s_j}{\partial p_j}\right)^*>0$, provided $\tau>1$ (while ensuring $\tau\sigma<1$), by choosing α sufficiently small in magnitude, we can ensure that the term $\frac{1}{\alpha}\left(\frac{\partial s_j}{\partial p_j}\right)^*$ dominates the expression for $-\gamma_{jj}=\left(\left(\frac{\partial s_j}{\partial p_j}\right)^*\frac{1}{\alpha}-\left(1-\tau s_j^*\right)s_j^*\right)$ and hence that the solution involves a $b_{jj}^{\Im}>0$.

Moreover, provided each r is finite and non-zero (which we will ensure), the solution $b_{jj}^{\Im}(\alpha,\sigma,\tau,\left(\frac{\partial s}{\partial p}\right)^*,s^*)$ will be both positive and bounded – i.e., we have $0<b_{jj}^{\Im}(\alpha,\sigma,\tau,\left(\frac{\partial s}{\partial p}\right)^*,s^*)<M<\infty$.

Step 2

Fixed Point Argument.

Define ω to be the vector of the (J+1)(J+2)/2 unique elements of the symmetric matrix B, which includes b_{jk} but not b_{kj}. Furthermore, define $\Delta_\Omega=[0,M']^{(J+1)(J+2)/2}\subset\Re_+^{(J+1)(J+2)/2}$ and $\Delta_R=[0,M]^{J+1}\subset\Re_+^{J+1}$.

We have argued (in step 1) that there is a vector of bounded functions $\omega(r):\Delta_R\to\Delta_\Omega$ in the positive orthant such that $\omega=\omega(r;s^*,\left(\frac{\partial s}{\partial p}\right)^*,\alpha,\tau,\sigma)$ where ω identifies all the elements of the matrix B since it is symmetric.

Similarly, we have argued (following Corollary 2 and using ω to replace B) that there is a vector function $r(\omega):\Delta_\Omega\to\Delta_R$, 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 $r=r(\omega;s^*,\alpha,\tau,\sigma)$.

Each of these functions is continuous and map from a non-empty, convex and compact set into a non-empty, convex and compact set. Thus we can define $\psi:\Delta_\Omega\times\Delta_R\to\Delta_\Omega\times\Delta_R$ by:

$$\psi(r,\omega)=\left(r(\omega),w(r)\right)$$

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

Finally, for completeness, we note that at this fixed point, by construction we will have $\sum_{j=0}^{J} r_j^* H_j(r^*;\theta_2)=\tau H\ (r^*;\theta_2)=1$ 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. ■

Proof to Lemma 4

Part (i) Given some general distribution for heterogeneity in taste $P(v_i)$ for the additively separable RC-MNL model we can write, for any $j\neq k$, $\frac{\partial s_j}{\partial p_k}=\int\frac{\partial \delta_k}{\partial p_k}s_{ij}s_{ik}dP(v_i).\ \frac{\partial s_j}{\partial p_k}=\frac{\partial s_k}{\partial p_j}$ only if $\frac{\partial \delta_k}{\partial p_k}$ is independent of option *k*. Notice that the additively separable and linear assumption implies that $\frac{\partial \delta_k}{\partial p_k}=g(v_i)$ which is indeed independent of option *k.*

Part (ii) For any member of the MEV class of models $s_j=\frac{r_jH_j}{\tau H}$, and for any $j\neq k$ we have $\frac{\partial s_j}{\partial p_k}=\frac{\partial s_j}{\partial r_k}\frac{\partial r_k}{\partial p_k}=\frac{\partial s_j}{\partial r_k}r_k\frac{\partial \delta_k}{\partial p_k}=\frac{1}{\tau}\left[r_kr_jH_{jk}H^{-1}-r_kr_jH_jH_kH^{-2}\right]\frac{\partial \delta_k}{\partial p_k}$ and similarly $\frac{\partial s_k}{\partial p_j}=\frac{1}{\tau}\left[r_kr_jH_{kj}H^{-1}-r_kr_jH_jH_kH^{-2}\right]\frac{\partial \delta_j}{\partial p_j}$. Since $H_{jk}=H_{kj}$ then $\left(\frac{\partial \delta_k}{\partial p_k}\right)^{-1}\frac{\partial s_j}{\partial p_k}=\left(\frac{\partial \delta_j}{\partial p_j}\right)^{-1}\frac{\partial s_k}{\partial p_j}$. ■

More on Identification and Estimation

For identification, we further need that rank$\left(\frac{\partial}{\partial \theta_2'}E\left(Z'\xi(\theta)\right)\Big|_{\theta=\theta^*}\right)=\dim\left(\theta_2\right)=J^2+J$. Since the matrix $\left(\frac{\partial s(\delta\ ,\theta_2^*,\Im^+)}{\partial \delta\ '}\right)^{-1}$ has full rank[34] which is equal to *n* while Z is (*nxq*) matrix with rank $q\geq\dim(\theta)>\dim(\theta_2)$, then identification essentially requires that

[34] Since it is block diagonal with each block invertible so is itself full rank and invertible.

$rank\left(\frac{\partial s(\delta,\theta_2^*,\Im^+)}{\partial\theta_2'}\right)=\dim(\theta_2)$. As $s_{kt}(r_t,\theta_2^*,\Im^+)=\frac{r_{kt}H_k(r_t;\theta_2^*,\Im)}{\tau H(r_t;\theta_2^*,\Im)}$, we can write

$$\frac{\partial s_k(r_t,\theta_2^*,\Im^+)}{\partial b_{ij}}=\frac{r_{kt}}{\tau}\left(\frac{1}{H(r_t;\theta_2^*,\Im)}\frac{\partial H_k(r_t;\theta_2^*,\Im)}{\partial b_{ij}}-\frac{H_k(r_t;\theta_2^*,\Im)}{H(r_t;\theta_2^*,\Im)^2}\frac{\partial H(r_t;\theta_2^*,\Im)}{\partial b_{ij}}\right).$$ To

simplify the algebra, define $\rho_{kjt}=\rho_{jkt}=\frac{r_{kt}^{\frac{1}{\sigma}}+r_{jt}^{\frac{1}{\sigma}}}{2}$ for all *j*,*k* and normalize the mean utilities

in each period such that at the true parameter values, θ^*, $\tau H(r_t;\theta^*)=1$. The latter allows

us to rewrite $s_{kt}(r_t,\theta_2^*,\Im^+)=r_{kt}H_k(r_t;\theta_2^*,\Im)$ and

$$\left.\frac{\partial s_{kt}(r_t,\theta_2^*,\Im^+)}{\partial b_{ij}}\right|_{\theta=\theta^*}=r_{kt}\frac{\partial H_k(r_t;\theta_2^*,\Im)}{\partial b_{ij}}-\tau s_{kt}(r_t,\theta_2^*,\Im^+)\frac{\partial H(r_t;\theta_2^*,\Im)}{\partial b_{ij}}.$$ Specifically at r_t

which equates actual and predicted shares, $s_{kt}=s_{kt}(r_t,\theta_2^*,\Im^+)$ for all *k, t*, we can use the

expressions above to write for j>0[35]:

$$\frac{\partial s_k(r_t,\theta_2,\Im^+)}{\partial b_{ij}}=\begin{cases}\tau r_{kt}^{\tau}(1-s_{kt}) & if\quad i=k,i=j\ \text{so}\ j=k\\ \tau\left(\left(\rho_{jkt}\right)^{\tau\sigma-1}r_{kt}^{\frac{1}{\sigma}}-s_{kt}\left(\rho_{kjt}\right)^{\tau\sigma}\right) & if\quad i=k,i\neq j\ \text{so}\ k\neq j\\ -\tau s_{kt}r_{jt}^{\tau} & if\quad i\neq k,i=j\ \text{so}\ j\neq k\\ -\tau s_{kt}\left(\rho_{ijt}\right)^{\tau\sigma} & if\quad i\neq k,i\neq j\end{cases}$$

and for j=0, since $b_{k0}=b_{0k}$,

$$\frac{\partial s_k(r_t,\theta_2,\Im^+)}{\partial b_{i0}}=\begin{cases}\tau\left(\left(\rho_{k0t}\right)^{\tau\sigma-1}r_{kt}^{\frac{1}{\sigma}}-2s_{kt}\left(\rho_{k0t}\right)^{\tau\sigma}\right) & if\quad i=k\\ -\tau 2s_{it}\left(\rho_{i0t}\right)^{\tau\sigma} & if\quad i\neq k\end{cases}$$

[35] Please see the web-appendix for further details

Thus the matrix $\left(\frac{\partial s(\delta,\theta_2,\Im^+)}{\partial \theta_2'}\right)$ can then be written out with

$$\theta_2' \equiv (b_{11}, b_{12}, ..., b_{1j}, ..., b_{1J}, b_{10}, ..., b_{k1}, ..., b_{kj}, ..., b_{kk}, ..., b_{kJ}, b_{k0}, .., b_{J1}, ..., b_{JJ}, b_{J0})$$

$$\begin{array}{c} \\ 1 \\ \\ \\ \tau \\ t \\ \\ \\ T \end{array} \begin{pmatrix} r_{11}^{\tau}(1-s_{11}) & \cdots & r_{11}^{\frac{1}{\sigma}}\rho_{1j1}^{\tau\sigma-1} - s_{11}\rho_{1j1}^{\tau\sigma} & \cdots & r_{11}^{\frac{1}{\sigma}}\rho_{101}^{\tau\sigma-1} - 2s_{11}\rho_{101}^{\tau\sigma} & \cdots & -s_{11}\rho_{kj1}^{\tau\sigma} & \cdots & -s_{11}r_{k1}^{\tau} & \cdots & -s_{11}\rho_{jk1}^{\tau\sigma} & \cdots \\ \cdots & & \cdots & & \cdots & & \cdots & & \cdots & & \cdots & \\ -s_{j1}r_{11}^{\tau} & \cdots & -s_{j1}\rho_{1j1}^{\tau\sigma} & \cdots & -2s_{j1}\rho_{101}^{\tau\sigma} & \cdots & -s_{j1}\rho_{kj1}^{\tau\sigma} & \cdots & -s_{j1}r_{k1}^{\tau} & \cdots & r_{j1}^{\frac{1}{\sigma}}\rho_{jk1}^{\tau\sigma-1} - s_{j1}\rho_{jk1}^{\tau\sigma} & \cdots \\ \cdots & & \cdots & & \cdots & & \cdots & & \cdots & & \cdots & \\ \cdots & & \cdots & & \cdots & & \cdots & & \cdots & & \cdots & \\ -s_{kt}r_{1t}^{\tau} & \cdots & -s_{kt}\rho_{1jt}^{\tau\sigma} & \cdots & -2s_{kt}\rho_{10t}^{\tau\sigma} & \cdots & r_{kt}^{\frac{1}{\sigma}}\rho_{kj1}^{\tau\sigma-1} - s_{kt}\rho_{kjt}^{\tau\sigma} & \cdots & r_{kt}^{\tau}(1-s_{kt}) & \cdots & -s_{kt}\rho_{jkt}^{\tau\sigma} & \cdots \\ \cdots & & \cdots & & \cdots & & \cdots & & \cdots & & \cdots & \\ \cdots & & \cdots & & \cdots & & \cdots & & \cdots & & \cdots & \\ -s_{JT}r_{1T}^{\tau} & \cdots & -s_{JT}\rho_{1jT}^{\tau\sigma} & & -2s_{JT}\rho_{10T}^{\tau\sigma} & \cdots & -s_{JT}\rho_{kjT}^{\tau\sigma} & & -s_{JT}r_{kt}^{\tau} & & -s_{JT}\rho_{jkt}^{\tau\sigma} & \cdots \end{pmatrix} \quad (4)$$

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 p, 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. $b_{jkt} = \begin{cases} 1/d_{jk}(x_j, x_k; \alpha_1) & if \quad j \neq k \\ \exp(x_j'\alpha_2) & j = k \end{cases}$ where $d_{jk}(x_j, x_k; \alpha_1) = \left(\sum_{l=1}^{L} \alpha_{1l}(x_{lj} - x_{lk})^2\right)^2$ then

$$\frac{\partial H_k(r_t;\theta_2)}{\partial \alpha_1} = \tau \sum_{j \neq k} \frac{\partial b_{jkt}(x_{jt}, x_{kt}; \alpha_1)}{\partial \alpha_1} \left(\frac{r_{jt}^{\frac{1}{\sigma}} + r_{kt}^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-1} r_{kt}^{\frac{1}{\sigma}-1} \quad \text{and}$$

$$\frac{\partial H_k(r_t;\theta_2)}{\partial \alpha_2} = \tau \frac{\partial b_{kk}^{\Im}(x_{kt}, \alpha_2)}{\partial \alpha_2} r_{kt}^{\tau-1} \quad \text{with} \quad \frac{\partial b_{kk}^{\Im}(x_{kt};\alpha_2)}{\partial \alpha_{2l}} = x_{lkt} e^{x_{kt}'\alpha_2} \quad \text{and}$$

$$\frac{\partial b_{jkt}(x_{jt}, x_{kt}; \alpha_1)}{\partial \alpha_{1l}} = -\frac{1}{2}\left(\sum_{l=1}^{L}(x_{ljt} - x_{lkt})^2 \alpha_{1l}\right)^{-3} (x_{ljt} - x_{lkt})^2$$ which reduces the column dimension of the matrix (4). Therefore, identification will require less data availability. Using such specification the parameters α_1 are identified up to their sign.

Appendix B:

BLP style contraction mapping. In this section, 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 $s_j(r) = \frac{r_j N_j(r;\Im)}{\sum_{l=0}^{J} r_l N_l(r;\Im)}$, we may consider the J+1 dimensional equations $F_j(\delta) = \ln\left(f_j\left(r(\delta)\right)\right) - \ln s_j^{obs}$, j=0,1,..,J where $f_j\left(r(\delta)\right) = r_j N_j(r;\Im)$ denotes the numerators of the market share function. Write the vector of equations, $F(\delta) = \ln(f(r(\delta))) - \ln(s^{obs})$. We will solve the J+1 equations, $F(\delta) = 0$.

Notice that at a solution point, $s_j^{obs} = f_j\left(r(\delta)\right)$ and so $1 = \sum_{j=0}^{J} s_j^{obs} = \sum_{j=0}^{J} f_j\left(r(\delta)\right)$, 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 $\tilde{F}_j(\delta) = \delta_j + \rho\left(\ln s_j^{obs} - \ln\left(f_j\left(r(\delta)\right)\right)\right)$ defined for J=0,1,..,J with $0 < \rho < \frac{1}{\tau}$ will converge.

Proof: We will apply the contraction mapping theorem in BLP to the function, $F_j(\delta) = \delta_j + \rho\left(\ln s_j^{obs} - \ln\left(f_j\left(r(\delta)\right)\right)\right)$ J=0,1,…,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 $\frac{\partial F_j(\delta)}{\partial \delta_k} = I(j=k) - \rho \frac{\partial \ln f_j(r(\delta))}{\partial \delta_k} = I(j=k) - \rho \frac{\partial \ln f_j(r(\delta))}{\partial \ln r_k} \geq 0$, where the inequality follows since $\frac{\partial \ln f_j(r(\delta))}{\partial \ln r_k} \leq 0$ for $j \neq k$, while $\frac{\partial \ln f_j(r)}{\partial \ln r_j} + \sum_{k \neq j} \frac{\partial \ln f_j(r)}{\partial \ln r_k} = \tau > 0 \Leftrightarrow \frac{\partial \ln f_j(r)}{\partial \ln r_j} = \tau - \sum_{k \neq j} \frac{\partial \ln f_j(r)}{\partial \ln r_k} \geq \tau$ so $I(j=k) - \rho \frac{\partial \ln f_j(r(\delta))}{\partial \ln r_k} \geq I(j=k) - \rho\tau > 0$ where the latter follows by the assumption that $\rho\tau < 1$.

We also require that $\sum_k \frac{\partial F_j(\delta)}{\partial \delta_k} = 1 - \rho \sum_k \frac{\partial \ln f_j(r)}{\partial \ln r_k} = 1 - \rho\tau < 1$ which follows immediately. ■

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 $s(r;\theta_2)=s$ 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 discus next for the symmetric case than the BLP contraction algorithm for reasons we also discus below.

Proposition 2: *Let* $V(r;\theta_2;\Im)=\frac{1}{\tau}\ln H(r;\theta_2,\Im)$, *which is the MEV expected maximum utility function when evaluated at* $r_j=e^{\delta_j(x_j,\xi_j,y,p_j;\theta_1)}$ *for* $j=0,1,..,J$, *and let* $r_0=1$. *If* for j>0 and $s_j^{obs}=0$, set $r_j^*=0$. *If for all* $j\neq k$ *with* $j,k\in\Im$, $\frac{\partial^2 V(r;\theta_2,\tau;\Im)}{\partial r_j\partial r_k}\leq 0$ *and for each* $k\in\Im$, *there is at least one* $j_k\in\{1,..,J\}$ *with* $j_k\neq k$ *such that* $\frac{\partial^2 V(r;\theta_2,\tau;\Im)}{\partial r_{j_k}\partial r_k}<0$, *then the solution to the inside good equations with positive market shares* $s(r_{\Im^+},\theta_2;\Im^+)=s_{\Im^+/0}^{obs}$ *for* $r_{\Im^+/0}$ *may be found as the unique solution to the strictly convex (and sub-modular) optimization problem:*

$$\min_{\{\ln r_j;\, j\in\Im^+/0\}} V(r;\theta_2;\Im)-\sum_{j\in\Im^+/0}(\ln r_j)s_j^{obs}$$

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 super-linear rate of convergence. When augmented with zeros for any products with zero market shares and* $r_0=1$, *the resulting* $r_{\Im}^*$ *will solve the full set of equations,* $s(r,\theta_2;\Im)=s_{\Im}^{obs}$.

Proof: Mcfadden(1981) establishes that the MEV class of models has an expected maximum utility function $V(r;\theta_2;\Im)=\frac{1}{\tau}\ln H(r;\theta_2,\Im)$ (with $\tau\in\theta_2$) where $r_j=e^{\delta_j(w_j,y,p_j;\theta_1)}$ for $j\in\Im$, the set of products. The MEV model has market share functions, $s_j=\frac{1}{\tau}\frac{\partial\ln H(r;\theta_2,\Im)}{\partial\ln r_j}$ so that $s_j=\frac{\partial V(r;\theta_2,\Im)}{\partial\ln r_j}=r_j\frac{\partial V(r;\theta_2,\Im)}{\partial r_j}$.

First notice that if $s_j^{obs}=0$, setting $r_j^*=0$ will solve the equation $s_j^{obs}=s_j(r;\theta_2,\Im)=r_j\frac{\partial V(r;\theta_2,\Im)}{\partial r_j}$. Thus the solution involves $r_j^*=0$ for any $j\in\Im\ /\Im^+$, where $\Im^+=\{j\mid s_j^{obs}>0,\ j\in\Im\}$. Thus we will consider solving the equations only for goods with positive market shares, $j\in\Im^+$. Define $V(r_{\Im^+};\theta_2;\Im^+)=V((r_{\Im^+},0,..,0);\theta_2;\Im)$ when considering the reduced set of products so we may consider the equations

$s_j^{obs} = s_j(r_{\Im^+};\theta_2,\Im^+) = \dfrac{\partial \ln V(r_{\Im^+};\theta_2,\Im^+)}{\partial \ln r_j}$ which for $j \in \Im^+/0$ are clearly the first order conditions to the following problem as stated in the lemma: $\min_{\{\ln r_j;\, j\in\Im^+/0\}} V(r;\theta_2;\Im) - \sum_{j\in\Im^+}(\ln r_j)s_j^{obs}$.For the rest of the proof for brevity we leave the dependency of all functions on the set $\Im^+$ and the parameters θ_2 implicit.

It remains to show that the program is convex and sub-modular. Convexity follows since H(r) is a homogeneous degree τ function so that by Euler's equation:

$\tau H(r) = \sum_{j\in\Im^+} r_j H_j(r)$ and hence $1 = \sum_{j\in\Im^+} \dfrac{r_j H_j(r)}{\tau H(r)} = \sum_{j\in\Im^+} s_j(r) = \sum_{j\in\Im^+} \dfrac{\partial V(r)}{\partial \ln r_j}$. Differentiating both sides w.r.t. $\ln r_k$ yields $0 = \sum_{j\in\Im^+} \dfrac{\partial^2 V(r)}{\partial \ln r_j \partial \ln r_k}$. Then $\dfrac{\partial^2 V(r)}{\partial \ln r_k \partial \ln r_k} = -\sum_{j\neq k,\, j,k\in\Im^+} \dfrac{\partial^2 V(r)}{\partial \ln r_j \partial \ln r_k}$ so that provided $\dfrac{\partial^2 V(r)}{\partial \ln r_j \partial \ln r_k} \leq 0$ the Hessian of the function V has a weakly dominant positive diagonal and V is therefore convex in $\ln r$. If for each $k \in \Im^+$ there exists at least one $j \in \Im^+$ with $\dfrac{\partial^2 V(r)}{\partial \ln r_j \partial \ln r_k} < 0$ (strict) then $\dfrac{\partial^2 V(r)}{\partial \ln r_k \partial \ln r_k} = -\sum_{j\neq k} \dfrac{\partial^2 V(r)}{\partial \ln r_j \partial \ln r_k} > 0$ and a strict dominant diagonal condition in $\ln r$ is satisfied. Thus, the objective function subtracts a concave (ie., adds a convex (and in fact linear)) in $\ln r$ function, $\sum_{j\in\Im^+/0}(\ln r_j)s_j^{obs}$, from V and so retains V's convexity property.

Similarly, the condition that $\dfrac{\partial^2 V(r)}{\partial r_j \partial r_k} \leq 0$ implies $\dfrac{\partial^2 V(r)}{\partial \ln r_j \partial \ln r_k} = r_k r_j \dfrac{\partial^2 V(r)}{\partial r_j \partial r_k} \leq 0$ for $j \neq k$ (since $\dfrac{\partial V(r)}{\partial \ln r_k} = r_k \dfrac{\partial V(r)}{\partial r_k}$) so that the function V has weakly decreasing differences in $\ln r$ and hence so does the objective function. As is 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 super-linear convergence.■

Corollary 3

If for all j,k, $b_{jk} \geq 0$ *and for each k, there exists at least one* $j_k \in \{1,..,J\}$ *with* $j_k \neq k$ *so that* $b_{jk} > 0$ *then the model which sets* $H(r;\theta_2) = \sum_{j=0}^{J}\sum_{k\neq j} b_{jk}\left(\dfrac{r_j^{\frac{1}{\sigma}} + r_k^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma} + \sum_{j=0}^{J} b_{jj}^{\Im} r_j^{\tau}$

satisfies the conditions of Proposition 2 and thus there exists a unique vector of r's which equate 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: $\frac{\partial V(r)}{\partial \ln r_j} = r_j \frac{\partial V(r)}{\partial r_j}$ while $\frac{\partial^2 V(r)}{\partial \ln r_j \partial \ln r_k} = r_k\left(I(j=k)\frac{\partial V(r)}{\partial r_j} + r_j \frac{\partial^2 V(r)}{\partial r_j \partial r_k} \right)$ so that for $j \neq k$ and $r \in \Re_+^{J+1}$ we have $sign\left(\frac{\partial^2 V(r)}{\partial \ln r_j \partial \ln r_k} \right) = sign\left(\frac{\partial^2 V(r)}{\partial r_j \partial r_k} \right)$.

In addition, $\frac{\partial V(r)}{\partial r_j} = \frac{1}{\tau}\frac{\partial \ln H(r)}{\partial r_j} = \frac{1}{\tau} H_j(r) H(r)^{-1}$, and $\frac{\partial^2 V(r)}{\partial r_j \partial r_k} = \frac{1}{\tau}\left[H_{jk}(r)H(r)^{-1} - H_j(r)H_k(r)H(r)^{-2} \right]$ so sufficient conditions for $\frac{\partial^2 V(r)}{\partial r_j \partial r_k} \leq 0$ are that $H_j(r) \geq 0$ and $H_{jk}(r) \leq 0$. Since in this model, $H_{jk}(r;\theta_2) = \frac{\tau}{2\sigma}(\tau\sigma - 1) b_{jk} \left(\frac{r_j^{\frac{1}{\sigma}} + r_k^{\frac{1}{\sigma}}}{2} \right)^{\tau\sigma - 2} r_j^{\frac{1}{\sigma}-1} r_k^{\frac{1}{\sigma}-1}$ while $0 < \sigma < 1$, $\tau > 0$ and $\sigma\tau < 1$ for $j \neq k$ and $b_{jk} > 0$ is sufficient for $H_{jk}(r;\theta_2) < 0$ everywhere away from r=0.■

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: $\ln r^{n+1} = \ln r^n - \left(D^2_{\ln r} V(r^n) \right)^{-1} \left(D_{\ln r} V(r^n) - s \right)$, where $D^2_{\ln r} V(r) = \left(\frac{\partial^2 V(r^n)}{\partial \ln r_j \partial \ln r_k} \right)_{j,k=1}^{J}$ and $D_{\ln r} V(r) = \left(\frac{\partial V(r)}{\partial \ln r_j} \right)_{j=1}^{J} = s(r)$ are analytic functions. In addition, since $\frac{\partial V(r)}{\partial \ln r_j} = \frac{1}{\tau}\frac{\partial \ln H(r)}{\partial \ln r_j} = s_j(r)$ and *s* is the vector of observed market shares for the inside goods with positive market shares, the term $D_{\ln r} V(r^n) - s$ is exactly the difference between the predicted and actual market shares. Essentially therefore this algorithm says decrease the r_j associated with option j if you are currently over-predicting its market share. The Hessian matrix provides a weighting matrix for the updates as usual, $D^2_{\ln r} V(r) = \left(\frac{\partial s_k(r)}{\partial \ln r_j} \right)_{j,k=1}^{J}$ where $\frac{\partial s_k(r)}{\partial \ln r_j} = \frac{r_j r_k H_{jk}(r)}{\tau H(r)} + s_j(r)(I(j=k) - \tau s_k(r))$, 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 that $\ln r^{n+1} = \ln r^n - \ln\left(D_{\ln r} V(r^n) / s \right)$, 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 $K(r) = s(r;\theta_2) - s$ will have a unique solution $K(r^*) = 0$ provided $D_r s(r;\theta_2)$ 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 $D_r s(r;\theta_2)$ is positive definite.) Thus, it is therefore highly likely that for the asymmetric case, an algorithm based on Variational Inequalities will have substantial 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 practise, is discussed below.

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

(i) $$H(r;\theta_2) = l_{J+1}^{'}\left((B - diag(B)) \bullet \left(\frac{r^{\frac{1}{\sigma}} + r^{\frac{1}{\sigma}'}}{2} \right)^{\tau\sigma} \right) l_{J+1} + l_{J+1}^{'} diag(B) r^{\tau}$$

(ii) $$D_r H(r;\theta_2) = \tau r^{\frac{1}{\sigma}-1} \bullet \left(\left((B - diag(B)) \bullet \left(\frac{r^{\frac{1}{\sigma}} + r^{\frac{1}{\sigma}'}}{2} \right)^{\tau\sigma-1} \right) l_{J+1} + \tau(diag(B) r^{\tau-1}) \right)$$

$$D_r^2 H(r;\theta_2) = \frac{\tau(\tau\sigma-1)}{2\sigma} r^{\frac{1}{\sigma}-1} \left(r^{\frac{1}{\sigma}-1} \right)' \bullet (B - diag(B)) \bullet \left(\frac{r^{\frac{1}{\sigma}} + r^{\frac{1}{\sigma}'}}{2} \right)^{\tau\sigma-2} + \tau(\tau-1)(diag(B) r^{\tau-2})$$

(iii)
$$+ \left(\frac{\tau(1-\sigma)}{\sigma} \right) diag\left(r^{\frac{1}{\sigma}-2} \bullet \left((B - diag(B)) \bullet \left(\frac{r^{\frac{1}{\sigma}} + r^{\frac{1}{\sigma}'}}{2} \right)^{\tau\sigma-1} \right) l_{J+1} \right)$$

$$+ \frac{\tau(\tau\sigma-1)}{2\sigma} diag\left(r^{\frac{2}{\sigma}-2} \bullet \left((B - diag(B)) \bullet \left(\frac{r^{\frac{1}{\sigma}} + r^{\frac{1}{\sigma}'}}{2} \right)^{\tau\sigma-2} \right) l_{J+1} \right)$$

(iv) $$D_{\ln r} V(r) = s(r) = r \bullet D_r H(r) \bullet \left(l_{J+1}^{'} \left(r \bullet D_r H(r) \right) l_{J+1} \right)^{-1}$$

(v) $$D_{\ln r}^2 V(r) = (rr') \bullet D_r^2 H(r) \bullet \left(l_{J+1}^{'} \left(r \bullet D_r H(r) \right) l_{J+1} \right)^{-1} + s(r)(l_{J+1} - \tau s(r))'$$

Where, as before, B is the J+1xJ+1 matrix with jk[th] element b_{jk} *and we use 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.* l_{J+1} *is a (J+1x1) vector of ones and* $\bullet$ *is the Hadamard or element-by-element 'dot' product.*

Proof. Immediate from the definitions. Note that under symmetry, $\tau H(r) = \left(l_{J+1}^{'} \left(r \bullet D_r H(r) \right) l_{J+1} \right)$ whereas without symmetry, we must use the market share

model $s(r) = r \bullet D_r H(r) \bullet \left(l_{J+1}^{'}\left(r \bullet D_r H(r)\right) l_{J+1}\right)^{-1}$. If we use a non-symmetric B matrix in $D_r^2 H(r;\theta_2)$, it will still be positive definite.■

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, τ, σ), 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 $D_{\ln r}^2 V(r)$ and $D_{\ln r} V(r)$ to compute the sub-matrix 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 $\theta_2 = (B, \tau, \sigma)$ 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(F,K) is to determine a vector $\delta^* \in K \subset R^{J+1}$ such that $\left\langle F(\delta^*)', \delta - \delta^* \right\rangle \geq 0$ for all $\delta \in K$ where <a,b> denotes an inner product, F is a given continuous function from K to R^{J+1} and K is a given closed convex set. Both optimisation problems and non-linear equations can be solved using variational inequalities. Indeed, Proposition 1.1 of Nagurney (1999) establishes that if $K = R^{J+1}$ and $F : R^{J+1} \mapsto R^{J+1}$ then a vector $\delta^* \in R^{J+1}$ solves $VI(F, R^{J+1})$ iff $F(\delta^*) = 0.$

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 sub-problem

$$\left\langle g(\delta^k, \delta^{k-1})', \delta - \delta^k \right\rangle \geq 0 \text{ for all } \delta \in K$$

(c) Convergence verification. If $\left|\delta^k - \delta^{k-1}\right| \leq \varepsilon$ 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

$$g(\delta^k, \delta^{k-1}) = F(\delta^{k-1}) + \frac{1}{\rho} G * (\delta^k - \delta^{k-1})$$

for some fixed $\rho > 0$ and G a symmetric and positive definite matrix (eg., the identity matrix.) As Nagurney describes, the special structure of the function $g(\delta^k, \delta^{k-1})$ means that we write the VI sub-problem 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:

$$\delta^k = \arg\min_{\delta \in K} \quad \frac{1}{2}\delta' G\delta + \left(\rho F(\delta^{k-1}) - G*\delta^{k-1}\right)\delta .$$

By differentiating, in *the case where none of the constraints bind*, it becomes clear that at the solution, δ^k, to this problem we will get $\rho F(\delta^{k-1}) + G*(\delta^k - \delta^{k-1}) = 0$. (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 δ by some multiple of the prediction error in log market shares.) Thus stated in terms of δ, 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 straight forward.

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

References

Abbe, E., Bierlaire, M., and Toledo T. “Normalization and correlation of cross-nested logit models”, *Transportation Research Part B: Methodological*, Vol. 41 (2005), pp. 795-808

Banks, J., Blundell, R. and Lewbel, A. “Quadratic Engel Curves and Consumer Demand”, *Review of Economics and Statistics*, Vol. 4 (1996), pp. 527-539.

Ben-Akiva, M.E. and Bierlaire M. “Discrete choice methods and their applications to short-term travel decisions”, in R. Hall (ed.), Handbook of Transportation Science, Kluwer, 1999.

Ben-Akiva, M., Bolduc, D. (1996) “Multinomial probit with a logit kernel and a general parametric specification of the covariance structure”, working paper, Department of Civil and Environmental Engineering, MIT, Cambridge .

Ben-Akiva M. and François B. “μ homogeneous generalized extreme value model”, working paper, Department of Civil Engineering, MIT, Cambridge, Ma, 1983.

Berry, S. “Estimating Discrete Choice Models of Product Differentiation”, *RAND Journal of Economics*, Vol. 25 (1994), pp. 242-262.

Berry, S., Gandhi, A., and Haile P. “Connected Substitutes and Invertibility of Demand”, *Econometrica*, forthcoming.

Berry, S. and Haile P. “Identification in Differentiated Products Markets Using Market Level Data”, working paper, Yale University, 2012.

Berry, S., Levinsohn, J. and Pakes, A. “Automobile Prices in Market Equilibrium”, *Econometrica*, Vol. 63 (1995), pp. 841-890.

Berry, S. and Pakes, A. “The Pure Characteristics Discrete Choice Model”, *International Economic Review,* Vol .48 (2007), pp. 1193-1225

Bhat, C.R. “Covariance heterogeneity in nested logit models: Econometric structure and application to intercity travel”, *Transportation Research Part B*, Vol. 31 (1997), pp. 11-21.

Bhat, C.R. “An analysis of travel mode and departure time choice for urban shopping trips”, *Transportation Research Part B*, Vol. 32 (1998), pp. 361-371.

Bierlaire, M. “The network GEV model”, presented at the 2002 Swiss Transport Research Conference (STRC), Monte Verita, 2002.

Bierlaire, M. “A theoretical analysis of the nested logit model”, Annals of Operations Research, Vol. 1 (2006), pp. 287-300.

Bresnahan, T. Stern, S. and Trajtenberg, M. “Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980's”, *RAND Journal of Economics*, Special Issue in Honor of Richard E. Quandt. Vol. 28 (1997).

Briesch, R., Chintagunta, P. and Matzkin, R. L. “Semi-parametric Estimation of Choice Brand Behavior”, *Journal of the American Statistical Association*, Vol. 97 (2002), pp. 973-982.

Boyd, J. and Mellman R. E. “The Effect of Fuel Economy Standards on the U.S. Automotive Market: A Hedonic Analysis”, *Transportation Research*, Vol. 14A (1980), pp. 357-378.

Cardell, N. and Dunbar, F. “Measuring the Societal Impacts of Automobile Downsizing”, *Transportation Research*, 14A (1980), pp. 423-434.

Christensen, L. R., Jorgenson, D. W. and Lau, L. J. “Transcendental Logarithmic Utility Functions”, *American Economic Review*, Vol. 65 (1975), pp. 367-83.

Chu, C. “A paired combinational logit model for travel demand analysis”, in *Transport Policy, Management and Technology Towards 2001: Selected Proceedings of the Fifth World Conference on Transport Research*, Vol. 4, Western Periodicals, Ventura, CA, 1989.

Coublucq, D. “Demand Model and Selection Bias due to Exit/Mergers: A Dynamic Approach with an Application to the U.S. Railroad Industry” mimeo, Toulouse School of Economics, 2010.

Daly, A. and Bierlaire M. “A general and operational representation of generalised extreme value models”, *Transportation Research Part B: Methodological*, Vol. 40 (2006), pp. 285-305

Daly, A. and Zachary S. “Improved multiple choice models”, in D.A. Hensher and M.Q. Dalvi (eds.) *Determinants of Travel Choice*, Saxon House, Westmead, 1978.

Davis, P. “Empirical Models of Demand for Differentiated Products”, *European Economic Review (papers and proceedings)*, Vol. 44 (2000), pp. 993-1005.

Davis, P. “Estimating Multi-way Error Components Models with Unbalanced Data Structures”, *Journal of Econometrics*, Vol. 106 (2002), pp. 67-95.

Davis, P. “Spatial Competition in Retail Markets: Movie Theaters”, *RAND Journal of Economics*, vol. 37 (2006a), pp. 964-982

Davis, P. “Constraints on own- and cross-price Elasticities Obtained from Random Coefficient Multinomial Logit Models.” mimeo, LSE, 2006b.

Davis, P. and Garces E. *Quantitative Techniques for Competition and Antitrust Analysis*, Princeton University Press, 2009.

Deaton, A. and Muelbauer J. “An Almost Ideal Demand System”, *American Economic Review*, Vol. 70 (1980), pp. 312-326.

Diewert, E. “Applications of Duality Theory” in *Frontiers of Quantitative Economics*, edited by M. Intriligator and D. A. Kendrick, North Holland, 1974.

Diewert, E. “Generalized Slutsky Conditions for Aggregate Consumer Demand Functions,” *Journal of Economic Theory,* Vol. 15 (1977), pp. 353-362.

Diewert, E. “Symmetry Conditions for Market Demand Functions,” *The Review of Economic Studies, Vol.* 47 (1980), pp. 595-601.

Dube, J. P., Fox, J. and Che-Lin Su, “Improving the Numerical Performance of Discrete Choice Random Coefficients Demand Estimation”, *Econometrica*. 80 (2009), pp. 2231-2267.

Epstein, R. and Rubinfeld, D., "Merger Simulation: A Simplified Approach with New Applications", *Antitrust Law Journal,* Vol. 69 (2001), pp. 883-919.

Fletcher, R. “*Practical Methods of Optimisation: Volume 1: Unconstrained Optimization*”, John Wiley & Sons, Chichester, UK, 1980.

Freyberger J. “*Asymptotic theory for differentiated products demand models with many markets*” mimeo Northwestern University, 2012.

Goldberg, P. K. and Verboven, F. "The Evolution of Price Dispersion in the European Car Market", *Review of Economic Studies*, Vol. 68 (2001), pp. 811-48.

Goolsbee, A. and Petrin, A. "The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV", *Econometrica* , Vol. 72 (2004), pp.351-381.

Gowrisankaran, G. and Rysman, M. “Dynamics of Consumer Demand for New Durable Goods.” *Journal of Political Economy*, forthcoming.

Greene, W.H., Hensher, D.A., Rose, J.M. “Accounting for heterogeneity in the variance of unobserved effects in mixed logit models”. Transp. Res. B., Vol. 40 (2006), pp. 75–92

Grigolon, L. A “Discrete choice model for ordered nests”, mimeo, 2012.

Hansen, L. P. "Large Sample Properties of Generalized Method of Moments Estimation," *Econometrica,* Vol**.** 50 (1982), pp. 1029-1054

Hausman, J., Leonard G., and Zona J. D. “Competitive Analysis with Differentiated Products”, *Annales d’Economie et de Statistique,* Vol. 34 (1994), pp. 159-180.

Hausman, J. and Wise, D. “A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognising Interdependence and Heterogeneous Preferences”, *Econometrica*, Vol. 46 (1978), pp. 403-426.

Hess, S., Bolduc D. and Polak J.W. “Random Covariance Heterogeneity in Discrete Choice Models”, Transportation, Vol. 37 (2010), pp. 391-411.

Horowitz, J. and Savin, N. E. “Binary Response Models: Logits, Probits, and Semi-parametrics” *Journal of Economic Perspectives*, Vol. 15 (2001), pp. 43-56.

Ivaldi M. and Verboven F. “Quantifying the effects from horizontal mergers on European competition policy”, *International Journal of Industrial Organization,* Vol. 23 (2005), pp. 669-702

Joe, H. “*Multivariate Models and Dependence Concepts*” Monographs on Statistics and Applied Probability, Chapman and Hall, New York, 1996.

Judd, K. “*Numerical Methods in Economics*” MIT Press, Cambridge, MA, 1999.

Judd, K. and Skrainka B. "High performance quadrature rules: how numerical integration affects a popular model of product differentiation", CeMMAP working papers CWP03/11, Centre for Microdata Methods and Practice, Institute for Fiscal Studies, 2011.

Knittel, C. and Metaxoglou K. "Challenges in Merger Simulation Analysis", *The American Economic Review, Papers & Proceedings*, Vol. 101 (2011), pp. 56-59.

Knittel, C. and Metaxoglou K. “Estimation of Random Coefficient Demand Models: Challenges, Difficulties and Warnings,” Mimeo Sloan MIT, 2012.

Koppelman, F. and Sethi V. “Closed-Form Discrete Choice Model”, in *Handbook of Transport Modelling,* Chapter 13, edit by D.A. Hensher and K.J. Button, Elsevier Science Ltd, 2000.

Kotz, S. Balakrishnan, N. and Johnson, N. L. “*Continuous Multivariate Distributions*”, Wiley Series in Probability and Statistics, Volume 1, John Wiley &Sons, Inc. New York, 2000.

Lancaster K. “A new approach to consumer theory”, *Journal of Political Economy,* Vol. 74 (1996), pp. 132-157.

Matzkin, R. L. “Non-parametric and Distribution Free Estimation of the Binary Choice and Threshold crossing Models”, *Econometrica*, Vol. 60 (1992), pp. 239-270.

Matzkin, R. L. “Non-parametric identification and Estimation of Polychotomous Choice Models.” *Journal of Econometrics*, Vol. 58 (1993), pp. 137-168.

McFadden, D. "Modelling the Choice of Residential Location”, in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull (eds.), *Spatial Interaction Theory and Planning Models*, North Holland: Amsterdam, 1978.

McFadden, D. “Structural Discrete Probability Models Derived from Theories of Choice” Chapter 5, in “Structural Analysis of Discrete Data and Econometric Applications” D. McFadden and C. Manski (eds.) MIT Press, 1981.

McFadden, D. “*Generalized Method of Moments*” Chapter 3 from unpublished lecture notes available from: http://elsa.berkeley.edu/users/powell/e240b_f02/e240b.html , 1999.

McFadden, D. and Train K. "Mixed MNL Models of Discrete Response", *Journal of Applied Econometrics*, Vol. 15 (2000), pp. 447-470.

Nagurney, A. *Network Economics: A Variational Inequality Approach*. Second Revised Edition. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999.

Nakanishi, M. and Cooper, L. “Parameter Estimates for Multiplicative Interactive Choice Model: Least Squares Approach.” *Journal of Marketing Research*, Vol. 11 (1974), pp. 303-311.

Nevo, A. “Measuring Market Power in the Ready-To-Eat Cereal Industry”, *Econometrica*, Vol. 69 (2001), pp. 307-342.

Newey, W. and McFadden, D. “Large Sample Estimation and Hypothesis Testing” Chapter 36, Handbook of Econometrics, Volume IV (Eds. R. F. Engle and D. L. McFadden.) Elsevier, 1994.

Papola, A. “Some developments on the cross-nested Logit model”, *Transportation Research Part B: Methodological,* Vol. 38 (2004), pp. 833–854.

Petrin, A. "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy*, Vol. 110 (2002), pp. 705-29.

Pinkse, J., Slade, M. and Brett, C. “Spatial Price Competition: a Semiparametric Approach.” *Econometrica*, Vol. 70 (2002), pp. 1111-1153.

Pollak, R. A. and Wales, T. J. “*Demand System Specification and Estimation*” Oxford University Press, UK., 1995.

Powell, M. J. D. “On the convergence of the variable metric algorithm.” *J. Inst. Maths Applns*., Vol. 7 (1971), pp. 21-36.

Powell, M. J. D. “Some properties of the variable metric algorithm” AERE Harwell Report, TP483, 1972.

Rothenberg, T. J. “Identification in parametric models”. *Econometrica,* Vol. 39 (1971), pp. 577-591

Rust, J. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher", *Econometrica,* Vol. 53 (1987), pp. 783-805.

Schiraldi, P. “Automobile replacement: A dynamic structural approach”, *RAND Journal of Economics*, Vol. 42 (2011), pp. 266-291.

Small, K. “A discrete choice model for ordered alternatives”, *Econometrica*, Vol. 55 (1987), pp. 409-424.

Steenburgh, T. J. "The Invariant Proportion of Substitution Property (IPS) of Discrete-Choice Models." *Marketing Science,* Vol. 27 (2008), pp. 300-307

Swait, J. and Adamowicz W.“Choice Environment, Market Complexity, and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice”, *Organizational Behavior and Human Decision Processes*, Elsevier, Vol. 86 (2001), pp. 141-167, November.

Train, K. “*Discrete Choice Methods with Simulation”*, Cambridge University Press, Cambridge, MA, 2003.

Verboven F. and Brenkers R. “Liberalizing a distribution system: The European car market”, DTEW Research Report 0247, K.U.Leuven, 2002.

Vovsha, P. “Application of cross-nested logit model to mode choice in Tel Aviv, Israel, metropolitan area”, *Transportation Research Record*, Vol. 1607 (1997), pp. 6-15.

Wen, C. H. & Koppelman, F. S. ”The Generalized Nested Logit Model”, *Transportation Research Part B: Methodological,* Vol. 35 (2001), pp. 627–641.

Werden, G. J. & Froeb, L. M (1994) "The Effects of Mergers in Differentiated Products Industries: Logit Demand and Merger Policy", *Journal of Law, Economics and Organization,* Oxford University Press, 10 (2), pp. 407-26.

Whelan, G., R. Batley, T. Fowkes, and Daly A. “Flexible models for analysing route and departure time choice”, European Transport Conference Proceedings, Association for European Transport, Cambridge, MA, 2002

Williams, H. “On the formation of travel demand models and economic evaluation measures of user benefits”, *Environment and Planning A*, Vol. 9 (1977), pp. 285–344.