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 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. 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 Correlation structure in the unobserved product characteristics

The structure of the matrix $V_G = Var[G_n(\theta)]$ of the moment conditions will depend on our assumptions about the covariance structure of the unobserved product characteristics. We may for example assume that they are independent across markets but allow for correlation across products within a given market, or alternatively we may prefer the BLP style assumption that they are correlated across time and independent across markets. We discuss each case separately below.

Case 1: If we wish to allow for arbitrary correlation in the unobserved product characteristics across products and are prepared to assume independence across time, we can define $\tilde{Z} = \begin{bmatrix} \tilde{z}_{.1} \\ \vdots \\ \tilde{z}_{.t} \\ \vdots \\ \tilde{z}_{.T} \end{bmatrix}$, $\tilde{z}_{.t} = \begin{bmatrix} \tilde{z}_{1t} & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \tilde{z}_{jt} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \tilde{z}_{.Jt} \end{bmatrix}$ with $\tilde{z}'_{jt}$ which are respectively of size (TJxJq) (JxJq) and (1xq). Doing so allows us to write, $V_G = Var[G_n(\theta) \mid \tilde{Z}] = Var[\frac{1}{n}\sum_{t=1}^{T}\sum_{j=1}^{J}\tilde{\xi}_{jt}(\theta)\tilde{z}_{jt}' \mid \tilde{Z}] = \frac{1}{n^2}E\left[\tilde{Z}'\tilde{\xi}\tilde{\xi}'\tilde{Z} \mid \tilde{Z}\right] = \frac{1}{n^2}\tilde{Z}'(I_T \otimes \tilde{\Sigma})\tilde{Z}$

where $\tilde{\xi}' = (\tilde{\xi}'_{1,\dots,}\tilde{\xi}'_{.T})$ is stacked by time period (and product within time period) and where we assume that the errors are iid across t but not necessarily across j. We can then form an estimate of $\hat{V}_G = \frac{1}{n^2}\tilde{Z}'(I_T \otimes \hat{\Sigma})\tilde{Z}$ by plugging in an estimate of the JxJ matrix $\hat{\Sigma}$ with representative element $\hat{\sigma}_{jk} = \frac{1}{T_{jk}}\sum_{t=1}^{T}\hat{\tilde{\xi}}_{jt}\chi_{jt}\hat{\tilde{\xi}}_{kt}\chi_{kt}$ where $T_{jk} = \sum_{t=1}^{T}\chi_{jt}\chi_{kt}$ counts the number of observations (periods or markets) where products j and k are both observed.

Case 2: Alternatively we may wish to allow for arbitrary correlation in the unobserved product characteristics across time. To do so we instead stack the data by product (and time periods within product, indicating this alternative permutation of the rows of $\tilde{Z}$ and $\tilde{\xi}$ by $\tilde{Z}_\pi$ and $\tilde{\xi}_\pi$ respectively so that

$$V_G = Var[G_n(\theta) \mid \tilde{Z}] = Var[\frac{1}{n}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\tilde{\xi}_{jt}(\theta)\tilde{z}_{jt}' \mid \tilde{Z}] = \frac{1}{n^2}E\left[\tilde{Z}_\pi'\tilde{\xi}_\pi\tilde{\xi}_\pi'\tilde{Z}_\pi \mid \tilde{Z}\right] = \frac{1}{n^2}\tilde{Z}_\pi'(I_J \otimes \tilde{\Sigma})\tilde{Z}_\pi$$

which follows the BLP style assumption that the errors are i.i.d. across j but not necessarily across t. We can then form an estimate of $\hat{V}_G = \frac{1}{n^2}\tilde{Z}'(I_T \otimes \hat{\Sigma})\tilde{Z}$ by plugging in an estimate of the TxT matrix $\hat{\Sigma}$ with representative element $\hat{\sigma}_{st} = \frac{1}{J_{st}}\sum_{j=1}^{J}\hat{\tilde{\xi}}_{jt}\chi_{jt}\hat{\tilde{\xi}}_{js}\chi_{js}$ where $J_{st} = \sum_{j=1}^{J}\chi_{jt}\chi_{js}$ counts the number of observations (products) observed in both time periods (s and t).

In each case, the resulting estimators can be used with the unbalanced data sets typical in the demand context.

2 GMM Estimation details

Case 1: ***(Small J, large T)*** Define, $\theta_1 = \left(\beta_1',..,\beta_j',...,\beta_J'\right)'$ and $\theta_2 = (vec(B)',\sigma,\tau)'$ with $\theta = (\theta_1,\theta_2)'$ let $\xi_{jt}(\theta) = \delta_{jt}(\theta_2) - x_{jt}'\beta_j$ and let the sample moment condition be

$$G_n(\theta) = \frac{1}{n}\sum_{t=1}^{T}\sum_{j=1}^{J}\tilde{\xi}_{jt}(\theta)z_{jt}' = \frac{1}{n}\tilde{Z}'\tilde{\xi}(\theta) \quad \text{where} \quad \tilde{\xi} = \begin{bmatrix}\tilde{\xi}_{.1}\\ \vdots \\ \tilde{\xi}_{.t} \\ \vdots \\ \tilde{\xi}_{.T}\end{bmatrix}, \tilde{\xi}_{.t} = \begin{bmatrix}\tilde{\xi}_{1t}\\ \vdots \\ \tilde{\xi}_{jt} \\ \vdots \\ \tilde{\xi}_{JT}\end{bmatrix}, \tilde{\xi}_{jt} = \xi_{jt}\chi_{jt},$$

$$\tilde{Z} = \begin{bmatrix}\tilde{z}_{.1}\\ \vdots \\ \tilde{z}_{.t} \\ \vdots \\ \tilde{z}_{.T}\end{bmatrix}, \tilde{z}_{.t} = \begin{bmatrix}\tilde{z}_{1t} & 0 & \cdots & \cdots & 0\\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \tilde{z}_{jt} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \tilde{z}_{Jt}\end{bmatrix}, \quad \text{where} \quad \tilde{z}_{jt} = (z_{1jt}\chi_{jt},...,z_{qjt}\chi_{jt}), \quad \tilde{z}_{1t} = \underline{0},$$

$n = \sum_{j=1}^{J}\sum_{t=1}^{T}\chi_{jt}$ and $\chi_{jt} = 1$ if product j is sold in period t and zero otherwise so that χ_{jt} provides a missing value indicator. Following Hansen (1982) we choose $\theta = (\theta_1,\theta_2)'$ to minimize $G_n(\theta)'AG_n(\theta)$ where in practise this problem can be solved in two stages because the moments are linear in $\theta_1 = \beta$ so that the first order conditions with respect to these parameters $\frac{G_n(\theta)}{\partial\theta_1}'AG_n(\theta) = 0$ have an analytic solution. To see that, define

$$\tilde{\tilde{X}} = \begin{bmatrix}\tilde{\tilde{x}}_{.1}\\ \vdots \\ \tilde{\tilde{x}}_{.t} \\ \vdots \\ \tilde{\tilde{x}}_{.T}\end{bmatrix}, \tilde{\tilde{x}}_{.t} = \begin{bmatrix}\tilde{\tilde{x}}_{1t} & 0 & \cdots & \cdots & 0\\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \tilde{\tilde{x}}_{jt} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \tilde{\tilde{x}}_{Jt}\end{bmatrix}, \tilde{\tilde{x}}_{jt} = (x_{1jt}\chi_{jt},...,x_{kjt}\chi_{jt}), \quad \text{we can write}$$

$G_n(\theta) = \frac{1}{n}\tilde{Z}'\tilde{\xi}(\theta) = \frac{1}{n}\left(\tilde{Z}'\tilde{\delta}(\theta_2) - \tilde{Z}'\tilde{\tilde{X}}\theta_1\right)$ and hence $\Gamma_{1n} = \frac{\partial G_n(\theta)}{\partial\theta_1} = \frac{-1}{n}\tilde{Z}'\tilde{\tilde{X}}$.[2]

where

[2] In the traditional restricted case where $\beta_2 = \beta_3 = \cdots\beta_J = \beta$ so that $\theta_1 = \left(1_{J-1} \otimes \beta\right)$, it is sometimes useful to estimate the unrestricted model and put $\Gamma_{1n} = \frac{\partial G_n(\theta)}{\partial\theta_1} = -\tilde{Z}'\tilde{\tilde{X}}\left(1_{J-1}\otimes I_K\right)$ which follows since for any A matrices and $(m\text{x}n)$ matrix B, $\frac{\partial\left(vec(A)\otimes vec(B)\right)}{\partial vec(B)'} = vec(A)\otimes I_{mn}$. (See for example Lütkepohl(1996) page 184.)

$$\tilde{Z}'\tilde{\tilde{X}} = \begin{bmatrix} \tilde{\tilde{z}}_{.1}' & \cdots & \tilde{\tilde{z}}_{.t}' & \cdots & \tilde{\tilde{z}}_{.T}' \end{bmatrix} \begin{bmatrix} \tilde{\tilde{x}}_{.1} \\ \vdots \\ \tilde{\tilde{x}}_{.t} \\ \vdots \\ \tilde{\tilde{x}}_{.T} \end{bmatrix} = \sum_{t=1}^{T} \tilde{\tilde{z}}_{.t}'\tilde{\tilde{x}}_{.t} = \sum_{t=1}^{T} \begin{bmatrix} \tilde{\tilde{z}}_{1t}' & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \tilde{\tilde{z}}_{jt}' & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \tilde{\tilde{z}}_{Jt}' \end{bmatrix} \begin{bmatrix} \tilde{\tilde{x}}_{1t} & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \tilde{\tilde{x}}_{jt} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \tilde{\tilde{x}}_{Jt} \end{bmatrix}$$

$$= \sum_{t=1}^{T} \begin{bmatrix} \tilde{\tilde{z}}_{1t}'\tilde{\tilde{x}}_{1t} & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \tilde{\tilde{z}}_{jt}'\tilde{\tilde{x}}_{jt} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \tilde{\tilde{z}}_{Jt}'\tilde{\tilde{x}}_{Jt} \end{bmatrix}$$

In which case,
$\frac{G_n(\hat{\theta}_1,\theta_2)}{\partial\theta_1}' A G_n(\hat{\theta}_1,\theta_2) = 0$ and so $(\tilde{Z}'\tilde{\tilde{X}})' A(\tilde{Z}'\tilde{\delta} - \tilde{Z}'\tilde{\tilde{X}}\hat{\theta}_1) = 0$ or $\hat{\theta}_1 = \left(\tilde{\tilde{X}}'\tilde{Z}A\tilde{Z}'\tilde{\tilde{X}}\right)^{-1}\tilde{\tilde{X}}'\tilde{Z}A\tilde{Z}'\tilde{\delta}$. In general it is useful to allow for the estimator to be subject to linear constraints, $\begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix}\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}$ which for a fixed value of θ_2 collapse to $R_1\theta_1 = r_1$ and so we define the restricted estimator $\hat{\theta}_1^{\mathrm{Re}str} = \hat{\theta}_1^{un} - C^{-1}R_1'(R_1C^{-1}R_1')^{-1}(R_1'\hat{\theta}_1^{un} - r_1)$ where $C^{-1} = \left(\tilde{\tilde{X}}'\tilde{Z}A\tilde{Z}'\tilde{\tilde{X}}\right)^{-1}$. (See for example, Mcfadden, 1999).

The non-linear first order conditions must be solved numerically by maximizing the function subject to the constraints $R_2\theta_2 = r_2$ or we can solve $G_n(\theta_1(\hat{\theta}_2),\hat{\theta}_2)' A \frac{G_n(\theta_1(\hat{\theta}_2),\hat{\theta}_2)}{\partial\theta_2} = 0$ where $\hat{\theta}_1(\theta_2)$ describes the optimal value of the linear parameters given a value for the non-linear parameters $\hat{\theta}_2$. In order to calculate standard errors, we must also calculate $\Gamma_{2n} = \frac{\partial G_n(\theta)}{\partial\theta_2} = \frac{1}{n}\tilde{Z}' D_{\theta_2}\tilde{\delta}(\theta_2)$ which is generally most easily computed numerically but may also be computed using analytically using the implicit function theorem.

Case 2: **(Large J)**

In this case, define, $\theta_1 = \left(\beta_2',..,\beta_j',...,\beta_J'\right)'$ and $\theta_2 = (vec(\alpha_1)', vec(\alpha_2)', \sigma, \tau)'$ with $\theta = (\theta_1,\theta_2)'$ where $b_{jkt} = \begin{cases} 1/d_{jk}(x_{jt}, x_{kt}; \alpha_1) & if \quad j \neq k \\ \exp(x_{jt}'\alpha_2) & \quad j = k \end{cases}$ with $d_{jk}(x_{jt}, x_{kt}; \alpha_1) = \left(\sum_{l=1}^{L} \alpha_{1l}(x_{ljt} - x_{lkt})^2\right)^2$ as described in the text. In this case, we estimate using the BLP style moment conditions, $G_n(\theta) = \frac{1}{n}\sum_{t=1}^{T}\sum_{j=1}^{J}\xi_{jt}(\theta)z_{jt}^{'} = \frac{1}{n}Z'\xi(\theta)$, which will allow us to have arbitrary correlation in $\xi_{jt}(\theta)$ overtime à la BLP. Identification in this case works from the assumption that the random vectors $\left(\xi_{j1}(\theta),..,\xi_{jT}(\theta)\right)$ are i.i.d. over products, j. We may use the matrix algebra above wherever the matrix expressions capture unweighted sums over time and products except for the redefinition: $z_{jt} = (z_{1jt},...,z_{qjt})$. In addition, when calculating the standard errors of parameters we must calculate $Var\left[G_n(\theta)\right] = \frac{1}{n^2}E\left[Z'\xi(\theta)\xi(\theta)'Z\right]$. In doing so, it is easiest to use a line of code to re-sort the Z matrix so that the time-periods for a given product are collected together, Z_π (with subscript π denoting Z re-ordered; now stacked by T then J). Doing so allows us to exploit Kronecker product in the expression:

$$Var\left[G_n(\theta)\,|\,Z\right] = \frac{1}{n^2}E\left[Z'\xi(\theta)\xi(\theta)'Z\,|\,Z\right] = \frac{1}{n^2}E\left[Z_\pi{}'\xi_\pi(\theta)\xi_\pi(\theta)'Z_\pi\,|\,Z_\pi\right]$$
$$= \frac{1}{n^2}Z_\pi{}'(I_J \otimes \Sigma_T)Z_\pi.$$

where the elements of the (TxT) matrix Σ_T can be estimated using $\hat{\sigma}^2 = \frac{1}{J}\sum_{j=1}^{J}\xi_{jt}\xi_{js}$. Standard GMM asymptotic theory is then easily applied.

3 Gradients and Efficient Instruments

Following Chamberlin (1986) and most directly the appendix to BLP (1999) the efficient set of instruments when we only have conditional moment restrictions is:

$I(z) = E\left[\frac{\partial \xi_j}{\partial \theta_2}\,|\,z_j\right]T(z_j) \equiv D(z_j)T(z_j)$ where $T(z_j)$ is the matrix which normalizes the error matrix, $T(z)'T(z) = \Omega(z)^{-1} \equiv E[\xi\ \xi\ '|\,z]^{-1}$.

Now in this model, since $r_{jt} = e^{\delta_{jt}} = e^{x_{jt}\beta + \xi_{jt}}$ we can write $\ln r_{jt} = x_{jt}\beta + \xi_{jt}$ and hence $\frac{\partial \xi_{jt}}{\partial \beta} = -x_{jt}$ and $\frac{\partial \xi_{jt}}{\partial \theta_2} = \frac{\partial \ln r_{jt}}{\partial \theta_2} = \frac{1}{r_{jt}}\frac{\partial r_{jt}}{\partial \theta_2}$. Furthermore, since the vector function $s_j(r_t;\theta_2,\Im^+) = s_{jt}^{obs}\quad j \in \Im^+$ where $\Im^+ \equiv \left\{j\,|\,s_{jt}^{obs} > 0, j \in \Im\right\}$ defines the implicit vector function $r\,(\theta_2; s_t^{obs},\Im^+)$ where $r \in \Re_+^{\#\Im}$ with gradient:

$$\frac{\partial r(\theta_2; s_t^{obs},\Im^+)}{\partial \theta_2{}'} = \left(-\frac{\partial s(r_t,\theta_2,\Im^+)}{\partial r_t{}'}\right)^{-1}\left(\frac{\partial s(r_t\ ,\theta_2,\Im^+)}{\partial \theta_2{}'}\right)$$

We may write

$$\frac{\partial\xi_t}{\partial\theta_2}=\frac{1}{r_t}\bullet\left(-\frac{\partial s(r_t,\theta_2,\Im^+)}{\partial r_t'}\right)^{-1}\left(\frac{\partial s(r_t,\theta_2,\Im^+)}{\partial\theta_2'}\right)=\left(-r_t\bullet\frac{\partial s(r_t,\theta_2,\Im^+)}{\partial r'}\right)^{-1}\left(\frac{\partial s(r_t,\theta_2,\Im^+)}{\partial\theta_2'}\right)$$
$$=\left(-\frac{\partial s(r_t,\theta_2,\Im^+)}{\partial\ln r_t'}\right)^{-1}\left(\frac{\partial s(r_t,\theta_2,\Im^+)}{\partial\theta_2'}\right)$$

where $\frac{1}{r_t}$ denotes the Jx1 vector with jth element r_{jt}^{-1} and we have previously established that: $\frac{\partial s_k(r_t)}{\partial\ln r_{jt}}=\frac{r_{jt}r_{kt}H_{jk}(r_t)}{\tau H(r_t)}+s_j(r_t)(I(j=k)-\tau s_k(r_t))$.

Since $s_k(r_t,\theta_2^*,\Im^+)=\frac{r_{kt}H_k(r_t;\theta_2^*,\Im)}{\tau H(r_t;\theta_2^*,\Im)}$, we can write:

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

Given that $H(r_t;\theta_2)$ is homogenous of degree τ in r_t, we can always rescale the (J+1)x1 vector of r_t's to ensure an equivalent solution with the normalization imposed as $\tau H(r_t^*;\theta_2)=1$ which proves algebraically easier to work with.[3]
In such case the previous expression simplifies to $\left.\frac{\partial s_k(r_t,\theta_2,\Im^+)}{\partial b_{ij}}\right|_{\theta=\theta^*}=r_{kt}\frac{\partial H_k(r_t;\theta_2^*,\Im)}{\partial b_{ij}}-\tau s_k(r_t,\theta_2^*,\Im^+)\frac{\partial H(r_t;\theta_2^*,\Im)}{\partial b_{ij}}$. Now since

$H(r_t;\theta_2,\Im)=\sum_{j\in\Im}\sum_{k\neq j}b_{jk}\left(\frac{r_{jt}^{\frac{1}{\sigma}}+r_{kt}^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma}+\sum_{j=0}^{J}b_{jj}^{\Im}r_{jt}^{\tau}$, then $\frac{\partial H(r_t;\theta_2^*,\Im)}{\partial b_{ij}}=\left(\frac{r_{it}^{\frac{1}{\sigma}}+r_{jt}^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma}1(i\neq j)+r_{jt}^{\tau}1(i=j)$

for j>0 and $\frac{\partial H(r_t;\theta_2,\Im)}{\partial b_{i0}}=2\left(\frac{r_{it}^{\frac{1}{\sigma}}+r_{0t}^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma}1(i\neq 0)+r_{0t}^{\tau}1(i=0)$ given that $b_{k0}=b_{0k}$, and since

$H_k(r_t;\theta_2)=\tau\sum_{j\neq k}b_{kj}\left(\frac{r_{jt}^{\frac{1}{\sigma}}+r_{kt}^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-1}r_{kt}^{\frac{1}{\sigma}-1}+\tau b_{kk}^{\Im}r_{kt}^{\tau-1}$, we can write

---

[3] For example, for the MNL model following Berry (1994) we ordinarily normalize $\delta_{0t}=0$ and then set $\delta_{jt}=\ln s_{jt}-\ln s_{0t}$ to ensure that the models utility levels match observed shares exactly. Since $r_{jt}\equiv\exp(\delta_{jt})$, this corresponds to normalizing $r_{0t}=1$ and setting $r_{jt}=s_{jt}/s_{0t}$. Since $H(r_t)=\sum_{j=0}^{J}r_{jt}$, we have $H(r_t)=1+\sum_{j>0}\frac{s_{jt}}{s_{0t}}=1+\frac{1-s_{0t}}{s_{0t}}=\frac{1}{s_{0t}}$. If instead we add $\ln s_{0t}$ to every utility level so that we normalize the utility of the outside good to be $\delta_{0t}=\ln s_{0t}$ then equating observed shares to predicted shares in the MNL model would involve simply setting $\delta_{jt}=\ln s_{jt}$ for all j=0,..,J. Under this normalization therefore, $H(r_t)=\sum_{j=0}^{J}r_{jt}=\sum_{j=0}^{J}s_{jt}=1$. In this specific case, the normalization required to set $H(r_t;\theta^*)=1$ does not depend on the parameters of the distribution of tastes- since there are none - but in richer models it will.

$$\frac{\partial H_k(r_t;\theta_2)}{\partial b_{kj}} = \tau\left(\frac{r_{jt}^{\frac{1}{\sigma}} + r_{kt}^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-1} r_{kt}^{\frac{1}{\sigma}-1} 1(j\neq k) + \tau r_{kt}^{\tau-1} 1(j=k) \;\; \forall \;\; j \text{ and } \frac{\partial H_k(r_t;\theta_2)}{\partial b_{ij}} = 0 \;\; \forall \;\; i\neq k$$

for the ‘small J’ model or, if parameters are mapped down to be functions of underlying characteristics so that we can write:

$$H_k(r_t;\theta_2) = \tau\sum_{j\neq k} b_{jk}(x_{jt},x_{kt};\alpha_1)\left(\frac{r_{jt}^{\frac{1}{\sigma}} + r_{kt}^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-1} r_{kt}^{\frac{1}{\sigma}-1} + \tau b_{kk}^{\Im}(x_{kt};\alpha_2) r_{kt}^{\tau-1} \text{ and hence}$$

$$\frac{\partial H_k(r_t;\theta_2)}{\partial \alpha_1} = \tau\sum_{j\neq k}\frac{\partial b_{jk}(x_{jt},x_{kt};\alpha_1)}{\partial\alpha_1}\left(\frac{r_{jt}^{\frac{1}{\sigma}} + r_{kt}^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-1} r_{kt}^{\frac{1}{\sigma}-1} \text{ and } \frac{\partial H_k(r_t;\theta_2)}{\partial \alpha_2} = \tau\frac{\partial b_{kk}^{\Im}(x_{jt},\alpha_2)}{\partial\alpha_2} r_{kt}^{\tau-1}.$$

Now, if $b_{jk}(x_{jt},x_{kt};\alpha_1) = \left(\sum_{l=1}^{L}(x_{jlt}-x_{klt})^2\alpha_1\right)^{-2}$ and $b_{jj}(x_{jt};\alpha_2) = \exp(x_{jt}'\alpha_2)$ then

$$\frac{\partial b_{jk}(x_{jt},x_{kt};\alpha_1)}{\partial\alpha_1} = -\frac{1}{2}\left(\sum_{l=1}^{L}(x_{jlt}-x_{klt})^2\alpha_1\right)^{-3}(x_{jt}-x_{kt})^2 \text{ and } \frac{\partial b_{kk}^{\Im}(x_{jt};\alpha_2)}{\partial\alpha_2} = x_{kt}e^{x_t'\alpha_2}.$$

The form of this expression suggests that instruments based on the expression

$$\frac{\partial H_k(r_t;\theta_2)}{\partial \alpha_1} = -\frac{1}{2}\tau\sum_{k\neq j}\frac{(x_{jt}-x_{kt})^2}{\left(\sum_{l=1}^{L}(x_{jlt}-x_{klt})^2\alpha_1\right)^3}\left(\frac{r_{jt}^{\frac{1}{\sigma}} + r_{kt}^{\frac{1}{\sigma}}}{2}\right)^{\tau\sigma-1} r_{kt}^{\frac{1}{\sigma}-1}$$

May be useful and in particular we chose to use an initial set of instruments based on the ‘equal weighting’ metric: $z_{jt} = \sum_{k\neq j}\frac{(x_{jt}-x_{kt})^2}{\left(\sum_{l=1}^{L}(x_{jlt}-x_{klt})^2\right)^3}$ using only those product characteristics deemed exogenous for the particular application.

Moving to our slightly more complex moment condition, involving both time periods and products, we can simply write $G_n(\theta) = \frac{1}{n}\sum_{t=1}^{T}\sum_{j=1}^{J}\tilde{\xi}_{jt}(\theta)z_{jt}' = \frac{1}{n}\tilde{Z}'\tilde{\xi}(\theta)$ which implies $D_\theta G_n(\theta) = \frac{1}{n}\sum_{t=1}^{T}\sum_{j=1}^{J}D_\theta\tilde{\xi}_{jt}(\theta)z_{jt}'$ where $D_\beta\tilde{\xi}_{jt}(\theta) = -x_{jt}$ and $D_\alpha\tilde{\xi}_{jt}(\theta)$ follows from the expressions developed above.

References

Berry, S., Levinsohn J. and Pakes A. "Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy." *American Economic Review*, Vol. 89 (1999), pp. 400-430.

Chamberlain, G. “Asymptotic efficiency in semiparametric model with censoring”, *Journal of Econometrics,* Vol. 32 (1986) , pp. 189–218.

Lütkepohl, H. “*Handbook of Matrices*” John Wiley and Sons, Chichester, UK, 1996.