EASI made Easier

Krishna Pendakur*

December 2008

Abstract

Lewbel and Pendakur (2008) develop the idea of implicit Marshallian demands. Implicit Marshallian demand systems allow the incorporation of both unobserved preference heterogeneity and complex Engel curves into consumer demand analysis, circumventing the standard problems associated with combining rationality with either (or both) unobserved heterogeneity and high rank in demand. They also develop the Exact Affine Stone Index (EASI) implicit Marshallian demand system wherein much of the demand system is linearised and thus relatively easy to implement and estimate. The current paper offers a less technical introduction to implicit Marshallian demands in general and to the EASI demand system in particular. I show how to implement the EASI demand system, paying special attention to tricks which allow the investigator to further simplify the problem without sacrificing too much in terms of model flexibility. STATA code to implement the simplified models is included throughout the text and in an appendix.

Keywords: consumer demand analysis; EASI; implicit Marshallian demands; complex Engel curves; unobserved preference heterogeneity; rationality

JEL Codes: D11, D12, C31, C33, C51

1 Introduction

Empirical work with large consumer expenditure data sets finds Engel curves (budget-share functions over expenditure, holding prices constant) that are quite different across goods. For example, some goods have Engel curves that are close to linear or quadratic, while others are more S-shaped (see, Blundell, Chen and Kristensen (2007). Typical parametric demand models cannot encompass this variety of shapes. Demand models whose Engel curves are additive in functions of expenditure (for example, polynomials in log-expenditure) are constrained by Gorman's (1981) rank restriction: no matter how many Engel curves are in the model, they must be expressed as linear combinations of at most 3 functions of expenditure. This is of course satisfied by budget-shares which are quadratic in log-expenditure: the 3 functions are a constant, the log of expenditure and its square. However, it is hard to see an *a priori* reason that, for example, all 100 Engel curves in a 100-good demand system could be reduced to linear combinations of just 3 shapes (functions of expenditure).

*Tel.: 1-778-782-5501; fax: 1-778-782-5944.

E-mail address: pendakur@sfu.ca. *www:* www.sfu.ca/~pendakur.

Other current research shows the importance of allowing for unobserved preference heterogeneity in demand systems, and the difficulty of doing so in a coherent fashion. In typical consumer demand models, observables like prices, expenditure and household demographics explain no more than half the variation in budget shares. The rest is left to the usual suspects, including measurement error and unobserved heterogeneity in the preferences of consumers. Unfortunately, in most empirical models of consumer demand, model error terms cannot be interpreted as random utility parameters representing unobserved heterogeneity.

It is easy to see how unoberved preference heterogeneity complicates things via a simple example. If a consumer has a particularly high unobserved preference parameter for, say, food, then she will allocate a large budget share to food. This will manifest as a large positive 'error term' in the food share for her. However, this same consumer will be more affected by food price increases than other consumers, because she spends more on food. So, the income effect of the price change will be large for her. Thus, income effects are tied to unobserved preference heterogeneity. If budget shares themselves are nonlinear in income, then this chain of effects induces nonlinearity in the effects of the unobserved parameter on the consumer's chosen budget shares, which makes estimation hard. This is essentially the same argument as that presented for *observed* preference heterogeneity in Blundell, Duncan and Pendakur (1998) and Pendakur (1999). These ideas are developed at length in Brown and Walker (1989), McFadden and Richter (1990), Brown and Matzkin (1998), Lewbel (2001), and Beckert and Blundell (2004).

Although *a priori* we have no reason to think that Engel curves lack variety in shapes, or that observable variables capture all the varation in preferences, the Almost Ideal Demand (AID) model (Deaton and Muellbauer 1980), which has linear Engel curves for all goods and does not incorporate unobserved heterogeneity, remains very popular. This popularity is at least partly because alternative models involve nonlinear functions of many prices and parameters, which are often numerically difficult or intractible to implement. In addition, the AID model has a very convenient approximate form which may be estimated by linear methods.

Lewbel and Pendakur (2008) develop an approach to the specification and estimation of consumer demands that addresses the above issues while maintaining the simplicity of the AID model. Their contribution hinges on the development of *implicit Marshallian demands* which, in contrast to explicit (or, 'normal') Marshallian demands, express budget shares as an *implicit* function of observable prices, expenditures and demographic characteristics. Econometrically, implicit Marshallian demands have the dependent variable on both sides of the equation: that is, implicit Marshallian demand systems suffer from endogeneity. This econometric problem is easily solved via instrumental variables. However, relaxing the restriction that Marshallian budget-share equations have an explicit solution allows us to solve the problems described above.

2 An Extended Example

Consider a consumer with nominal total expenditures x that faces the $J-$vector of prices $\mathbf{p} = [p^1, ..., p^J]$. Assume she chooses a bundle of goods, described by the $J-$ vector of budget shares $\mathbf{w} = [w^1, ..., w^J]$, to maximize utility given her linear budget constraint. Let $x = C(\mathbf{p}, u)$ be her cost function giving the minimum nominal total expenditure to attain a utility level, u, given prices $\mathbf{p}$.

Suppose that we can write budget-shares as a function of prices, expenditures and budget-shares themselves, that is, we can write:

$$w^j = \psi^j(p^1, ..., p^J, x, w^1, ..., w^J),$$

for all $j = 1, ..., J$. This is an implicit Marshallian demand system. We say that it is *implicit* because budget-shares appear on both the left- and the right-hand sides. It is a *Marshallian* (uncompensated), rather than a Hicksian (compensated), demand system because it is expressed in terms of expenditure, x, and is not expressed in terms of utility, u. If $w^1, ..., w^J$ were not present on the right-hand side, this would reduce to an explicit (or, 'normal') Marshallian demand system. Thus, implicit Marshallian demand systems are more general than Marshallian demand systems, because Marshallian demand systems can be seen as imposing a restriction on implicit Marshallian demand systems. It is this increased generality that allows us to solve the problems identified above.

The value of implicit Marshallian demands is most easily seen by example. A simple example will be extended step-by-step to build up to an empirical model that can capture everything captured in other parametric models of demand. In addition, the final empirical model will be one that can accomodate arbitrary variation in observable demographic characteristics, arbitrarily complex Engel curves and additive unobserved preference heterogeneity. However, it is best to start with a very simple example.

Let $\boldsymbol{\omega}(\mathbf{p}, u) = \left[\omega^1(\mathbf{p}, u), ..., \omega^J(\mathbf{p}, u)\right]$ be the Hicksian (or, compensated) budget-share functions associated with the utility function. By Shephard's Lemma, these are equal to the price elasticity of the cost function: $\omega^j(\mathbf{p}, u) = \partial \ln C(\mathbf{p}, u)/\partial \ln p^j$. Here, budget-shares are expressed as functions of the price vector $\mathbf{p}$ and attained the utility level u, and can easily be specified to have many desirable properties. Unfortunately, since they depend on unobserved utility, u, they are not typically used in demand analysis. However, with implicit Marshallian demands, one can exploit the nice features of Hicksian demands while maintaining dependence of demands on observable variables only. Basically, the strategy is to define Hicksian budget-share functions that 'look right' and find an observable function of prices, expenditure and budget-shares that equals utility, and substitute that function into the Hicksian demands.

Say we wanted the Hicksian budget-share functions to be completely unrelated across goods and given by $m^j(u)$ for $j = 1, ..., J$. Working backwards through Shephard's Lemma, this implies a cost function

$$\ln C(\mathbf{p}, u) = u + \sum_{j=1}^{J} m^j(u) \ln p^j, \tag{1}$$

with Hicksian budget-share functions

$$\omega^j(\mathbf{p}, u) = m^j(u) \tag{2}$$

for $j = 1, ..., J$. This Hicksian budget-share system has one very nice characteristic: the budget-share functions $m^j(u)$ are completely unrestricted and unrelated across budget-shares $j = 1, ..., J$. However, like all Hicksian demands, they depend on utility rather than on an observable like expenditure.

Now, assume that budget-shares, $w^j = \omega^j(\mathbf{p}, u)$, are observable in the data. In this case, knowledge of budget shares allows us to express utility in terms of observables. Manipulating (1), and subsitituting x for $C(\mathbf{p}, u)$ and w^j for $\omega^j(\mathbf{p}, u)$, gives

$$u = \ln x - \sum_{j=1}^{J} w^j \ln p^j. \tag{3}$$

Here, utility is expressed in terms of observables: expenditure, x, prices, $p^1, ..., p^J$ and budget-shares, $w^1, ..., .w^J$.

Next, substituting (3) for u in (2) gives implicit Marshallian demands

$$w^j = m^j(\ln x - \sum_{j=1}^{J} w^j \ln p^j), \tag{4}$$

or, equivalently,

$$w^j = m^j(y), \tag{5}$$

where y is 'implicit utility' given by

$$y = \ln x - \sum_{j=1}^{J} w^j \ln p^j. \tag{6}$$

The presence of budget-shares on both sides of (4) means that budget-shares are implicitly defined. But, the absence of utility, u, means that it is implicitly defined in terms of observables. These implicit Marshallian budget-share functions can have any shapes at all over y. That is, they are not constrained by Gorman's rank restrictions.

This general feature of these models is important for at least two reasons. First, for a 100 good demand system, there could be 100 distinct shapes for the 100 Engel curves. That is, the demand system may have any rank. Second, the researcher need not know the exact parametric structure of budget-share functions beforehand. That is, there is room to let the data do the talking. For example, one could estimate the budget-share system as a 10th order polynomial in y. If all the orders matter, the data have spoken. If not, the data have still spoken. One cannot do this with explicit Marshallian demand systems that are polynomial in expenditure. Given Gorman's finding that at most 3 terms could matter, what would one do if 4 (or more) of the terms were statistically significant in at least one budget-share function?

Given a functional form for $m^j(y)$, this implicit Marshallian budget-share system is easy to estimate via instrumental variables. For example, if $m^j(y)$ is a 5th order polynomal in y, one could estimate (5) via two-stage least squares the linear regression of w^j on a constant plus 5 powers of y. The choice of instruments is not a difficult one because the model gives the structural equation for the endogenous regressor: equation (6) says that y depends on exogenous $\ln x$ and $\ln p^j$. Thus, any functions of these exogenous variables are allowable instruments. Thus, if y1-y5 were powers of y, lnx1-lnx5 were powers of $\ln x$ and lnp1-lnpJ were the the logged price vector, then one could estimate the model with the STATA code:

```
ivregress 2sls w1 (y1-y5=lnx1-lnx5 lnp1-lnpJ)
...
```

```
ivregress 2sls wJ (y1-y5=lnx1-lnx5 lnp1-lnpJ)
```

Here, the J'th equation does not add any information, since its parameters could be obtained via the adding up restriction ($\sum_{j=1}^{J} w^j = 1$). However, since the equations are estimated separately, choosing to estimate rather than calculate the parameters of the last equation is immaterial to the resulting estimates.

Now consider adding unobserved preference heterogeneity to the model. Let $\boldsymbol{\varepsilon} = \left[\varepsilon^1, ..., \varepsilon^J\right]$ be a vector of unobserved preference heterogeneity parameters for the consumer, and let $E\left[\boldsymbol{\varepsilon}\right] = \mathbf{0}_J$. We want $\boldsymbol{\varepsilon}$ to come in to budget-share functions as additive error terms. Adding the argument to the cost function, we let $C(\mathbf{p}, u, \boldsymbol{\varepsilon})$ be the minimum total expenditure for a consumer with unobserved heterogeneity parameters $\boldsymbol{\varepsilon}$ to attain a utility level u when facing prices $\mathbf{p}$. Again, it is easiest to consider the Hicksian budget-share system first. If we want ε^j to enter the Hicksian budget-share function as an additive component, then it must multiply $\ln p^j$ in an additive component in the (logged) cost function. Thus, we write the cost function

$$\ln C(\mathbf{p}, u, \boldsymbol{\varepsilon}) = u + \sum_{j=1}^{J} m^j(u) \ln p^j + \sum_{j=1}^{J} \varepsilon^j \ln p^j, \tag{7}$$

which gives (by Sheppard's Lemma) Hicksian budget-share functions

$$\omega^j(\mathbf{p}, u, \boldsymbol{\varepsilon}) = m^j(u) + \varepsilon^j. \tag{8}$$

Proceeding as before, manipulating (7), and subsitituting x for $C(\mathbf{p}, u, \boldsymbol{\varepsilon})$ and w^j for $\omega^j(\mathbf{p}, u, \boldsymbol{\varepsilon})$, gives implicit utility $y = u = \ln x - \sum_{j=1}^{J} w^j \ln p^j$ as before. Thus, implicit Marshallian budget shares are given by

$$w^j = m^j(y) + \varepsilon^j$$

and, as before, implicit utility is given by

$$y = \ln x - \sum_{j=1}^{J} w^j \ln p^j.$$

Estimation of this model proceeds exactly as in the model above which lacks unobserved preference heterogeneity, and could use the same STATA code. Here, the 'error terms' in the budget-share equations are interpreted as unobserved preference heterogeneity parameters. Because these parameters show up in both budget-share functions and the cost function, they are relevant factors in both predicting demand and in assessing the cost of living as prices change (these are considered in Section 6).

From its definition in equation (6), one can see that if the price vector were $\mathbf{1}_J$, so that the log-price vector were $\mathbf{0}_J$, then $y = \ln x$. So, implicit utility y is a log-money-metric representation of utility for a unit price vector (for more on this, see Pendakur and Sperlich 2008). Thus, we could equivalently call y as log real-expenditures. Lewbel and Pendakur (2008) call the cost function (7) and associated implicit Marshallian demand system as *Exact Stone Index* cost and demands. This is because the stone index (Stone 1954), given by $\Pi_{j=1}^{J} \left(p^j\right)^{w^j}$, is the exact deflator which converts nominal expenditures x into real expenditures $\exp(y)$ (exponentiate equation (6) to see this).

Exact Stone Index (ESI) implicit Marshallian demands have a lot going for them as an empirical approach: (1) budget-share functions can have any degree of variety of shapes across goods; (2) unobserved preference heterogeneity is incorporated in a simple and intuitive fashion, and, because it is embedded in the cost function, is integrated into welfare analysis. However, ESI demands have one very large drawback: they incorporate prices into demands solely through implicit utility y. That is, Hicksian budget-share functions do not respond to prices at all. Thus, although pleasing to the eye, ESI demands are not suitable for demand analysis in the real world. However, a simple modification, which Lewbel and Pendakur (2008) call *Exact Affine Stone Index* cost and demands, incorporates price effects in a simple and tractable way.

3 The EASI Demand System

Now, consider a modification of the ESI cost function that incorporates both prices and observable demographic characteristics. Let $\mathbf{z} = [z_1, ..., z_T]$ be a vector of demographic characteristics of the consumer and let the first element of $\mathbf{z}$ be a constant, so that $z_1 = 1$. Let $\overline{\mathbf{z}} = [1, 0, ..., 0]$ be a vector of zeroes with a leading 1, and let $\overline{\mathbf{z}}$ be the value of $\mathbf{z}$ for a reference type of consumer. Add this new argument to the cost function so that $C(\mathbf{p}, u, \mathbf{z}, \boldsymbol{\varepsilon})$ is the minimum total expenditure for a consumer with observed characteristics $\mathbf{z}$ and unobserved characteristics $\boldsymbol{\varepsilon}$ to attain a utility level u when facing prices $\mathbf{p}$. Essentially, we incorporate prices by modifying (7) to include quadratic form in log-prices, and we include demographic characteristics $\mathbf{z}$ in the m^j functions. This results in an Exact Affine Stone Index (EASI) cost function of the form

$$\ln C(\mathbf{p}, u, \mathbf{z}, \boldsymbol{\varepsilon}) = u + \sum_{j=1}^{J} m^j(u, \mathbf{z}) \ln p^j + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} a^{jk}(\mathbf{z}) \ln p^j \ln p^k + \sum_{j=1}^{J} \varepsilon^j \ln p^j. \tag{9}$$

Sheppard's Lemma gives Hicksian budget-share functions as

$$\omega^j(\mathbf{p}, u, \mathbf{z}, \boldsymbol{\varepsilon}) = m^j(u, \mathbf{z}) + \sum_{k=1}^{J} a^{jk}(\mathbf{z}) \ln p^k + \varepsilon^j, \tag{10}$$

where $a^{jk}(\mathbf{z}) = a^{kj}(\mathbf{z})$ for all j, k. Note that

$$\sum_{j=1}^{J} w^j \ln p^j = \sum_{j=1}^{J} m^j(u, \mathbf{z}) \ln p^j + \sum_{j=1}^{J} \sum_{k=1}^{J} a^{jk}(\mathbf{z}) \ln p^j \ln p^k,$$

which is missing the $\frac{1}{2}$ multiplying the quadratic form in (9). Thus, implicit utility is given by

$$y = u = \ln x - \sum_{j=1}^{J} w^j \ln p^j + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} a^{jk}(\mathbf{z}) \ln p^j \ln p^k. \tag{11}$$

Here, the log of the deflator that *exact*ly converts nominal expenditures into real expenditures is $\sum_{j=1}^{J} w^j \ln p^j - \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} a^{jk}(\mathbf{z}) \ln p^j \ln p^k$, which is *affine* in the Stone Index (hence the name).

Implicit Marshallian budget shares are obtained by substituting y (equation (11)) for u in the Hicksian budget-share functions (10):

$$w^j = m^j(y, \mathbf{z}) + \sum_{k=1}^{J} a^{jk}(\mathbf{z}) \ln p^k + \varepsilon^j, \tag{12}$$

where $a^{jk}(\mathbf{z}) = a^{kj}(\mathbf{z})$ for all j, k.

This EASI implicit Marshallian demand system has several features in common with traditional demand systems, such as the popular Quadratic Almost Ideal (QAI) demand system of Banks, Blundell and Lewbel (1997). First, it is easy to estimate via iterative linear methods, which we describe below. Second, there are linear price effects which may depend on observable characteristics ($a^{jk}(\mathbf{z})$). Third, the functions $m^j(y, \mathbf{z})$ can be independent of y as in homothetic demand systems, linear in y as in the Almost Ideal demand system, or quadratic in y as in the QAI demand system.

In addition, this EASI implicit Marshallian demand system has several clear advantages over traditional demand systems. First, the functions $m^j(y, \mathbf{z})$ are completely unrestricted in their dependence on implicit utility y and observable demographic characteristics $\mathbf{z}$. Thus, Engel curves may have any shape and any degree of variety across goods. Nothing about the shape of Engel curves need be known in advance. Second, unobserved preference heterogeneity is captured through the parameters ε^j. These parameters show up as 'error terms' in the estimating equation and as cost shifters in the cost function.

4 EASI Estimation

Estimation of (12) is complicated by two factors: (1) is it slightly nonlinear; and (2) the equation system is endogenous due to the presence of w^j on both sides. The nonlinearity is due to the fact that $m^j(y, \mathbf{z})$ may be nonlinear in y and that y is itself a function of the vectors $\mathbf{w}$, $\mathbf{p}$ and $\mathbf{z}$. For the purposes of showing how to implement estimation, we will parameterise $m^j(y, \mathbf{z})$ and $a^{jk}(\mathbf{z})$ with simple additive structures. Consider $m^j(y, \mathbf{z})$ additively separable in $y, \mathbf{z}$, linear in $\mathbf{z}$ and polynomial in y, given by

$$m^j(y, \mathbf{z}) = \sum_{r=1}^{R} b_r^j y^r + \sum_{t=1}^{T} g_t^j z_t, \tag{13}$$

and $a^{jk}(\mathbf{z})$ given by

$$a^{jk}(\mathbf{z}) = a^{jk} \tag{14}$$

where $a^{jk} = a^{kj}$ for all j, k. Note that because z_1 is equal to 1, we have for the reference consumer (with $\mathbf{z} = [1, 0, ..., 0]$)

$$m^j(y, [1, 0, ..., 0]) = g_1^j + \sum_{r=1}^{R} b_r^j y^r.$$

These choices results in implicit utility given by

$$y = \ln x - \sum_{j=1}^{J} w^j \ln p^j + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} a^{jk} \ln p^j \ln p^k \tag{15}$$

Substituting (15) into (13) and substituting (13) and (14) into (12) yields estimating equations given by

$$w^j = \sum_{r=1}^{R} b_r^j (y)^r + \sum_{t=1}^{T} g_t^j z_t + \sum_{k=1}^{J} a^{jk} \ln p^k + \varepsilon^j, \tag{16}$$

or, equivalently,

$$w^j = \sum_{r=1}^{R} b_r^j \left(\ln x - \sum_{j=1}^{J} w^j \ln p^j + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} a^{jk} \ln p^j \ln p^k \right)^r + \sum_{t=1}^{T} g_t^j z_t + \sum_{k=1}^{J} a^{jk} \ln p^k + \varepsilon^j. \tag{17}$$

Here, the parameters b_r^j control the shape of the Engel curve. The only restriction is that $R < J$ (see Lewbel 1991): otherwise, Engel curves can be arbitrarily complex. The parameters g_t^j allow for demographic shifters in budget shares, and the parameters a^{jk} govern price effects. Finally, the parameters ε^j (or, error terms) incorporate unobserved preference heterogeneity into budget-shares and the cost function.

This EASI equation system is nonlinear and endogenous. The nonlinearity in the parameters is driven solely by the fact that b_r multiplies (a power of) a^{jk}. The endogeneity arises from the fact that budget-shares w^j, $j = 1, ..., J$ are expressed implicitly and are thus on both sides of the system of equations. Endogenous nonlinear systems may be estimated efficiently via the Hansen's (1982) Generalised Method of Moments (GMM).

Familiar software, such as GAUSS, MATLAB, R and SAS, allow the estimation by GMM of nonlinear endogenous systems of equations. However, such estimation can be cumbersome and there are legitimate problems with overidentified GMM models in small samples. For these reasons, the next four subsections consider how to estimate EASI models via more familiar linear methods. In particular, since the nonlinearity in (17) is similar to the type of nonlinearity discussed in Blundell and Robin (1999), an iterated linear estimator similar to theirs is discussed below.

4.1 Approximate Models

Before turning to iterated linear estimation to estimate the EASI demand system, I note that Lewbel and Pendakur (2008) provide some evidence that both the nonlinearity and endogeneity are relatively small issues in practise. They discuss an 'approximate model' which replaces y with

$$\widetilde{y} = \ln x - \sum_{j=1}^{J} w^j \ln p^j$$

and estimates via OLS. Here, $\widetilde{y}$ is the log of Stone-index deflated nominal expenditures. Applied to (17), this gives

$$w^j = \sum_{r=1}^{R} b_r^j (\widetilde{y})^r + \sum_{t=1}^{T} g_t z_t + \sum_{k=1}^{J} a^{jk} \ln p^k + \varepsilon^j.$$

If `ỹ1-ỹR` were R powers of $\widetilde{y}$ and `z1-zT` were the a constant plus $T-1$ other demographic characteristics, one would estimate the approximate model with the following STATA code:

```
regress w1 ỹ1-ỹR z1-zT p1-pJ, noconst
...
regress wJ ỹ1-ỹR z1-zT p1-pJ, noconst
```

Here, the `noconst` option is used because `z1` is the constant term. In their empirical work with Canadian price and expenditure data, they find that $\widetilde{y}$ is so highly correlated with y that inefficient, endogeneity-polluted and linearised OLS regression performs almost as well as fully efficient endogeneity-corrected nonlinear GMM estimation.

4.2 Iterated Linear Estimation

Blundell and Robin (1999) show that the QAI can be estimated by iterated linear methods. The QAI is very similar to (17) with $R = 2$. The only two differences are that, in the QAI, $\sum_{j=1}^{J} w^j \ln p^j$ does not appear on the right hand side, and b_2^j gets divided by a price index depending on b_1^j. An iterated linear strategy one could use to estimate (17) is as follows:

- Let a_0^{jk} denote initial values for a^{jk}.
- Compute
$$y_0 = \ln x - \sum_{j=1}^{J} w^j \ln p^j + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} a_0^{jk} \ln p^j \ln p^k.$$
- let *tol*=some small value; let $n = 1$
- loop while *crit<tol:*

  1. Estimate the linear model (subscripts for individual observations are suppressed):
  $$w^j = \sum_{r=1}^{R} b_r^j \left(y_{n-1}\right)^r + \sum_{t=1}^{T} g_t^j z_t + \sum_{k=1}^{J} a^{jk} \ln p^k + \varepsilon^j.$$
  2. denote estimated values of a^{jk} as a_n^{jk}, and compute
  $$y_n = \ln x - \sum_{j=1}^{J} w^j \ln p^j + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} a_n^{jk} \ln p^j \ln p^k.$$
  3. compute a criterion assessing the change $y_n - y_{n-1}$, such as the max of this change over all the individuals.
  4. let $n = n + 1$

- Retain the final estimates b_r^j, g_t and a^{jk} at convergence of y_n.

Note that the linear model to be estimated at each linear iteration is an endogenous model. The next subsection considers exactly what instruments can and should be used.

4.3 Instrumental Variables Estimation

The iterated linear estimation described above has an endogenous regressor whose structure is given by the model. The endogenous regressors are R powers of y_n, and y_n is a function of exogenous $\ln x$, z_t and $\ln p^j$ (as well as endogenous w^j). If $\ln x$, z_t and $\ln p^j$ are exogenous, then any functions of them that are correlated with y are allowable as instruments. STATA code for each instrumental variables linear regression is easy to construct. For example, if y1-yR were R powers of y_n, then the following would implement the regression for each iteration:

```
ivregress 2sls w1 z1-zT p1-pJ (y1-yR=lnx1-lnxR p1-pJ)
...
ivregress 2sls wJ z1-zT p1-pJ (y1-yR=lnx1-lnxR p1-pJ)
```

Although these instruments satisfy exogeneity and are correlated with the endogenous regressor y, they may not have the maximum possible correlation with y. Because the structure of y is completely known, it is possible to improve on these instruments, in the sense of increasing their correlation with y. In particular, given a set of fixed exogenous parameter values $\overline{a}^{jk}$ and fixed exogenous budget-shares $\overline{w}^j$, one could substitute these into (15) to generate an instrument $\overline{y}$:

$$\overline{y} = \ln x - \sum_{j=1}^{J} \overline{w}^j \ln p^j + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} \overline{a}^{jk} \ln p^j \ln p^k. \tag{18}$$

The exogenous budget-shares $\overline{w}^j$ could be the sample average budget-share vector. The values $\overline{a}^{jk}$ could be estimated values from an initial estimator. Note that in the iterated estimation, the instrument would not be updated at each iteration. If ȳ1-ȳR were R powers of $\overline{y}$, then each iteration could be estimated with the following STATA code:

```
ivregress 2sls w1 z1-zT p1-pJ (y1-yR=ȳ1-ȳR)
...
ivregress 2sls wJ z1-zT p1-pJ (y1-yR=ȳ1-ȳR)
```

4.4 Imposing Symmetry with Linear System Estimation

Up to now, estimation has been equation-by-equation. This is consistent with the possibility that Slutsky symmetry holds, but does not impose Slutsky symmetry. That is, it allows $a^{jk} = a^{kj}$ for all j, k, but does not impose that restriction. Imposition of symmetry requires the use of system methods. In general one could use nonlinearly restricted GMM methods. However, the iterative linear procedure described above works here, too. It is relatively easy to embed three stage least squares linear endogenous system estimation in the estimation step (step 1) of the iterative procedure. Three stage least squares would proceed by specifying the instruments as above and imposing the linear restrictions $a^{jk} = a^{kj}$ for all j, k.

In STATA, one would set up global macros for each equation and for the cross-equation restrictions prior to iteration:

```
global eq1 "w1 y1-yR z1-zT p1-pJ"
...
global eq{J-1} "w{J-1} y1-yR z1-zT p1-pJ"
constraint 12 [w1]p2=[w2]p1
...
constraint {J-2}{J-1} [w{J-2}]p{J-1}=[w{J-1}]p{J-2}
```

Then, at each iteration's estimation step, one would implement three stage least squares with the STATA command:

```
reg3 $eq1 ... $eq{J-1}, endog(y1-yR) exog(ȳ1-ȳR) constraints(12...{J-2}{J-1})
```

Here, curly brackets indicate indices over goods (prices). Because this implementation estimates the whole equation system, the redundant final equation must be dropped, as well as all restrictions associated with it. As noted above, Lewbel and Pendakur (2008) find that the approximate model performs tolerably well. However, they note that its weakest point in their empirical exercise was the failure of equation-by-equation OLS to impose symmetry. That is, whereas dealing with the nonlinearity and endogeneity of the EASI estimating equations doesn't change the estimates much in practise, imposing Slutsky symmetry via cross-equation restrictions does affect the resulting estimates in important ways.

5 EASI Extensions

The EASI estimation model given in equation (16) allows for additively separable effects of implicit utility, y, demographic characteristics, $\mathbf{z}$, prices, $\mathbf{p}$, and unobserved preference heterogeneity, $\boldsymbol{\varepsilon}$. This approach can also easily accomodate all two-way interactions among y, $\mathbf{z}$ and $\mathbf{p}$. Consider another cost function in the EASI class:

$$\ln C(\mathbf{p},u,\mathbf{z},\boldsymbol{\varepsilon}) = u+\sum_{j=1}^{J} m^j(u,\mathbf{z})\ln p^j+\frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} a^{jk}(\mathbf{z})\ln p^j \ln p^k+\frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} b^{jk}\ln p^j \ln p^k u+\sum_{j=1}^{J}\varepsilon^j \ln p^j. \tag{19}$$

Sheppard's Lemma gives Hicksian budget-share functions as

$$\omega^j(\mathbf{p},u,\mathbf{z},\boldsymbol{\varepsilon}) = m^j(u,\mathbf{z})+\sum_{k=1}^{J} a^{jk}(\mathbf{z})\ln p^k+\sum_{k=1}^{J} b^{jk}\ln p^k u+\varepsilon^j, \tag{20}$$

where $a^{jk}(\mathbf{z}) = a^{kj}(\mathbf{z})$ and $b^{jk} = b^{kj}$ for all j,k. A little algebra reveals that implicit utility is given by

$$y = u = \frac{\ln x - \sum_{j=1}^{J} w^j \ln p^j + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} a^{jk}(\mathbf{z})\ln p^j \ln p^k}{1-\frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} b^{jk}\ln p^j \ln p^k}, \tag{21}$$

which is affine in the log of stone-index deflated nominal expenditures, $\ln x - \sum_{j=1}^{J} w^j \ln p^j$.

To parameterise, consider modifying $m^j(y,\mathbf{z})$ to include an interaction between y and $\mathbf{z}$:

$$m^j(y,\mathbf{z}) = \sum_{r=1}^{R} b_r^j y^r + \sum_{t=1}^{T} g_t^j z_t + \sum_{t=2}^{T} h_t^j z_t y.$$

Note that since the first element of z is 1, the trailing summation is from $2,..,T$ (the h_1^j effect is captured by b_1^j). Next, allow for an interaction between $\mathbf{p}$ and $\mathbf{z}$ by letting $a^{jk}(\mathbf{z})$ be linear in z:

$$a^{jk}(\mathbf{z}) = \sum_{t=1}^{T} a^{jkt} z_t$$

Substituting these into (21) and into (20) yields the following implicit Marshallian demand system:

$$w^j = \sum_{r=1}^{R} b_r^j y^r + \sum_{t=1}^{T} g_t^j z_t + \sum_{t=2}^{T} h_t^j z_t y + \sum_{k=1}^{J}\sum_{t=1}^{T} a^{jkt} z_t \ln p^k + \sum_{k=1}^{J} b^{jk}\ln p^k y + \varepsilon^j, \tag{22}$$

where

$$y = \frac{\ln x - \sum_{j=1}^{J} w^j \ln p^j + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J}\sum_{t=1}^{T} a^{jkt} z_t \ln p^j \ln p^k}{1-\frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} b^{jk}\ln p^j \ln p^k}.$$

The demand system (22) has additively separable effects in y, $\mathbf{z}$ and $\ln \mathbf{p}$ (the direct effect goes through $a^{jk1}z_1 = a^{jk1}$). In addition, it has additive two-way interactions for $z_t y$, $z_t \ln p^k$ and $\ln p^k y$. Iterative linear estimation proceeds as outlined above. STATA code to estimate EASI models with and without two-way interactions is provided in the appendix and online at www.sfu.ca/~pendakur and online in the American Economic Review Electronic Archive.

6 EASI-to-Use

Because the EASI demand systems are dual to cost functions, they are very easy to use for consumer surplus estimation. Consider evaluating the cost to an individual of a price change. A consumer surplus measure for the price change from $\mathbf{p}_0$ to $\mathbf{p}_1$ is the log cost of living index, which for the cost function (19) is given by

$$\ln\left[\frac{C(\mathbf{p}_1,u,\mathbf{z},\boldsymbol{\varepsilon})}{C(\mathbf{p}_0,u,\mathbf{z},\boldsymbol{\varepsilon})}\right] = \sum_{j=1}^{J} m^j(u,\mathbf{z})(\ln p_1^j - \ln p_0^j) + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} a^{jk}(\mathbf{z})\ln p_1^j \ln p_1^k - \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} a^{jk}(\mathbf{z})\ln p_0^j \ln p_0^k + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} b^{jk}\ln p_1^j \ln p_1^k u - \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} b^{jk}\ln p_0^j \ln p_0^k u + \sum_{j=1}^{J}\varepsilon^j(\ln p_1^j - \ln p_0^j)$$

If $C(\mathbf{p}_0,u,\mathbf{z},\boldsymbol{\varepsilon})$ is the cost function of a household that has budget shares $\mathbf{w}_0$ and implicit utility level y, and if we use the parameterised model (22), then this expression can be rewritten in terms of observables and parameters as

$$\ln\left[\frac{C(\mathbf{p}_1,u,\mathbf{z},\boldsymbol{\varepsilon})}{C(\mathbf{p}_0,u,\mathbf{z},\boldsymbol{\varepsilon})}\right] = \sum_{j=1}^{J} w_0^j(\ln p_1^j - \ln p_0^j) + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J}\left(\sum_{t=1}^{T} a^{jkt}z_t + b^{jk}y\right)\left(\ln p_1^j - \ln p_0^j\right)\left(\ln p_1^j - \ln p_0^j\right),$$

or, in the case without two-way interactions,

$$\ln\left[\frac{C(\mathbf{p}_1,u,\mathbf{z},\boldsymbol{\varepsilon})}{C(\mathbf{p}_0,u,\mathbf{z},\boldsymbol{\varepsilon})}\right] = \sum_{j=1}^{J} w_0^j(\ln p_1^j - \ln p_0^j) + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} a^{jk}\left(\ln p_1^j - \ln p_0^j\right)\left(\ln p_1^j - \ln p_0^j\right).$$

The first term in this cost of living index is the Stone index for the price change. Such indices are commonly used on the grounds that they are appropriate for small price changes and that they allow for unobserved preference heterogeneity across households. The presence of the second term allows one to explicitly model substitution effects, and so consider large price changes, while also accounting for the behavioral importance of both observed and unobserved heterogeneity.

Demand elasticities are also easy to compute in this framework. Define semielasticities to be derivatives of budget shares with respect to log prices, $\ln \mathbf{p}$, implicit utility, y, and demographic characteristics, $\mathbf{z}$. Note that derivatives with respect to implicity utility, y, are attained by differentiating Hicksian budget shares with respect to utility, u, and these are independent of monotone transformations of utility. The semielasticity of a budget share can be converted into an ordinary elasticity of budget share by dividing by that budget share. Hicksian (compensated) price semielasticities for the EASI cost function (19) and implicit Marshallian demand system (22) are given by

$$\frac{\partial\omega^j(\mathbf{p},u,\mathbf{z},\boldsymbol{\varepsilon})}{\partial \ln p^k} = a^{jk}(\mathbf{z}) + b^{jk}u = a^{jk}(\mathbf{z}) + b^{jk}y.$$

Similarly, derivatives with respect to y, interpretable as real expenditure semi-elasticities, are given by

$$\frac{\partial\omega^j(\mathbf{p},u,\mathbf{z},\boldsymbol{\varepsilon})}{\partial u} = \frac{\partial\omega^j(\mathbf{p},u,\mathbf{z},\boldsymbol{\varepsilon})}{\partial y} = \frac{\partial m^j(\mathbf{p},u,\mathbf{z},\boldsymbol{\varepsilon})}{\partial y} + \sum_{k=1}^{J} b^{jk}\ln p^k,$$

and semielasticities with respect to observable demographics $\mathbf{z}$ are

$$\frac{\partial\omega^j(\mathbf{p},u,\mathbf{z},\boldsymbol{\varepsilon})}{\partial z_t} = \frac{\partial m^j(\mathbf{p},u,\mathbf{z},\boldsymbol{\varepsilon})}{\partial z_t} + \sum_{k=1}^{J}\frac{\partial a^{jk}(\mathbf{z})}{\partial z_t}\ln p^k.$$

Substituting the relevant parameters from the parameterised demand system (22), we have

$$\begin{aligned} \frac{\partial \omega^j(\mathbf{p}, u, \mathbf{z}, \varepsilon)}{\partial \ln p^k} &= \sum_{t=1}^{T} a^{jkt} z_t + b^{jk} y, \\ \frac{\partial \omega^j(\mathbf{p}, u, \mathbf{z}, \varepsilon)}{\partial u} &= \sum_{r=1}^{R} b_r^j r y^{r-1} + \sum_{t=2}^{T} h_t^j z_t + \sum_{k=1}^{J} b^{jk} \ln p^k, \end{aligned} \tag{23}$$

and

$$\frac{\partial \omega^j(\mathbf{p}, u, \mathbf{z}, \varepsilon)}{\partial z_t} = g_t^j + h_t^j y + \sum_{k=1}^{J} a^{jkt} \ln p^k.$$

7 Conclusions

The EASI demand system recently proposed by Lewbel and Pendakur (2008) aims to solve several longstanding problems in consumer demand estimation: (1) it allows for arbitrarily complex Engel curves which are arbitrarily varied across goods; (2) it allows for the incorporation of unobserved preference heterogeneity; and (3) an approximate model can be estimated by linear methods with the exact model being estimable by iterative linear methods.

8 Appendix: STATA Code

8.1 EASI with No Interactions

```
* Tricks with Hicks: The EASI demand system
    * Arthur Lewbel and Krishna Pendakur
    * 2008, American Economic Review
    * Herein, find Stata code to estimate a demand system with neq equations, nprice prices,
    *      ndem demographic characteristics and npowers powers of implicit utility
    set more off
    macro drop _all
    use "C:\projects\hixtrix\revision\hixdata.dta", clear
    * set number of equations and prices and demographic characteristics and convergence criterion
    global neqminus1 "7"
    global neq "8"
    global nprice "9"
    global ndem 5
    global npowers "5"
    global conv_crit "0.000001"
    *data labeling conventions:
    * budget shares: s1 to sneq
    * prices: p1 to nprice
    * implicit utility: y, or related names
    * demographic characteristics: z1 to zTdem
```

```
g s1=sfoodh
g s2=sfoodr
g s3=srent
g s4=soper
g s5=sfurn
g s6=scloth
g s7=stranop
g s8=srecr
g s9=spers
g p1=pfoodh
g p2=pfoodr
g p3=prent
g p4=poper
g p5=pfurn
g p6=pcloth
g p7=ptranop
g p8=precr
g p9=ppers
* normalised prices are what enter the demand system
* generate normalised prices, backup prices (they get deleted), and Ap
forvalues j=1(1)$neq {
    g np`j'=p`j'-p$nprice
}
forvalues j=1(1)$neq {
    g np`j'_backup=np`j'
    g Ap`j'=0
}
g pAp=0
*list demographic characteristics: fill them in, and add them to zlist below
g z1=age
g z2=hsex
g z3=carown
g z4=time
g z5=tran
global zlist "z1 z2 z3 z4 z5"
*make y_stone=x-p'w, and gross instrument, y_tilda=x-p'w^bar
g x=log_y
g y_stone=x
g y_tilda=x
```

```
forvalues num=1(1)$nprice {
    egen mean_s`num'=mean(s`num')
    replace y_tilda=y_tilda-mean_s`num'*p`num'
    replace y_stone=y_stone-s`num'*p`num'
}
*list of functions of (implicit) utility, y: fill them in, and add them to ylist below
*alternatively, fill ylist and yinstlist with the appropriate variables and instruments
g y=y_stone
g y_inst=y_tilda
global ylist ""
global yinstlist ""
forvalues j=1(1)$npowers {
    g y`j'=y^`j'
    g y`j'_inst=y_inst^`j'
    global ylist "$ylist y`j'"
    global yinstlist "$yinstlist y`j'_inst"
}
*set up the equations and put them in a list
global eqlist ""
forvalues num=1(1)$neq {
    global eq`num' "(s`num' $ylist $zlist np1-np$neq)"
    macro list eq`num'
    global eqlist "$eqlist \$eq`num'"
}
*create linear constraints and put them in a list, called conlist
global conlist ""
forvalues j=1(1)$neq {
    local jplus1=`j'+1
    forvalues k=`jplus1'(1)$neq {
        constraint `j'`k' [s`j']np`k'=[s`k']np`j'
        global conlist "$conlist `j'`k'"
    }
}
*first get a pre-estimate to create the instrument:
*run three stage least squares on the model with no py, pz or yz interactions, and then iterate to convergence
* note that the difference in predicted values between p and p=0 is Ap
replace y=y_stone
g y_old=y_stone
g y_change=0
```

```
scalar crit_test=1
while crit_test>$conv_crit {
    quietly reg3 $eqlist, constr($conlist) endog($ylist) exog($yinstlist)
    quietly replace pAp=0
    replace y_old=y
    forvalues j=1(1)$neq {
        quietly predict s`j'hat, equation(s`j')
    }
    forvalues j=1(1)$neq {
        quietly replace np`j'=0
    }
    forvalues j=1(1)$neq {
        quietly predict s`j'hat_p0, equation(s`j')
    }
    forvalues j=1(1)$neq {
        quietly replace np`j'=np`j'_backup
        quietly replace Ap`j'=s`j'hat-s`j'hat_p0
        quietly replace pAp=pAp+np`j'*Ap`j'
        quietly drop s`j'hat s`j'hat_p0
    }
    replace pAp=int(1000000*pAp+0.5)/1000000
    summ pAp
    quietly replace y=y_stone+0.5*pAp
    forvalues j=1(1)$npowers {
        quietly replace y`j'=y^`j'
    }
    quietly replace y_change=abs(y-y_old)
    summ y_change
    scalar crit_test=r(max)
    display `k'
    scalar list crit_test
    summ y_stone y y_old
}
*now, create the instrument
quietly replace y_inst=y_tilda+0.5*pAp
forvalues j=1(1)$npowers {
    quietly replace y`j'_inst=y_inst^`j'
}
*run three stage least squares on the model with no py, pz or yz interactions, and then iterate to convergence
```

```
* note that the difference in predicted values between p and p=0 is Ap
*reset the functions of y
replace y=y_stone
forvalues j=1(1)$npowers {
    quietly replace y`j'=y^`j'
}
replace y_old=y_stone
replace y_change=0
scalar crit_test=1
while crit_test>$conv_crit {
    quietly reg3 $eqlist, constr($conlist) endog($ylist) exog($yinstlist)
    quietly replace pAp=0
    replace y_old=y
    forvalues j=1(1)$neq {
        quietly predict s`j'hat, equation(s`j')
    }
    forvalues j=1(1)$neq {
        quietly replace np`j'=0
    }
    forvalues j=1(1)$neq {
        quietly predict s`j'hat_p0, equation(s`j')
    }
    forvalues j=1(1)$neq {
        quietly replace np`j'=np`j'_backup
        quietly replace Ap`j'=s`j'hat-s`j'hat_p0
        quietly replace pAp=pAp+np`j'*Ap`j'
        quietly drop s`j'hat s`j'hat_p0
    }
    replace pAp=int(1000000*pAp+0.5)/1000000
    summ pAp
    quietly replace y=y_stone+0.5*pAp
    forvalues j=1(1)$npowers {
        quietly replace y`j'=y^`j'
    }
    quietly replace y_change=abs(y-y_old)
    summ y_change
    scalar crit_test=r(max)
    display `k'
    scalar list crit_test
```

```
    summ y_stone y y_old
}
*note that reported standard errors are wrong for iterated estimates
reg3 $eqlist, constr($conlist) endog($ylist) exog($yinstlist)
```

8.2 EASI with Two-Way Interactions

```
* Tricks with Hicks: The EASI demand system
* Arthur Lewbel and Krishna Pendakur
* 2008, American Economic Review
* Herein, find Stata code to estimate a demand system with neq equations, nprice prices,
*      ndem demographic characteristics and npowers powers of implicit utility
* This Stata code estimates Lewbel and Pendakur's EASI demand system using approximate
* OLS estimation and iterated linear 3SLS estimation. Note that iterated linear 3SLS is
* not formally equivalent to fully nonlinear 3SLS (which does not exist in Stata).
* However, in some contexts they are asymptotically equivalent (see, e.g., Blundell and
* Robin 1999 and Dominitz and Sherman 2005), and we have verified in our data that
* coefficients estimated using iterated linear 3SLS are within 0.001 of those
* estimated using fully nonlinear 3SLS.
* Code to estimate the fully nonlinear 3SLS/GMM version in TSP is available on request
* from the authors.
* This model includes pz,py,zy interactions. See 'iterated 3sls without pz,py,zy.do' for
* shorter code to estimate the model without interactions.
set more off
macro drop _all
use "C:\projects\hixtrix\revision\hixdata.dta", clear
* set number of equations and prices and demographic characteristics and convergence criterion
global neqminus1 "7"
global neq "8"
global nprice "9"
global ndem 5
global npowers "5"
* set a convergence criterion and choose whether or not to base it on parameters
global conv_crit "0.000000000000001"
scalar conv_param=1
scalar conv_y=0
*note set the matrix size big enough to do constant,y,z,p,zp,yp,yz
global matsize_value=100+$neq*(1+$npowers+$ndem+$neq*(1+$ndem+1)+$ndem)
set matsize $matsize_value
*data labeling conventions:
```

```
* data weights: wgt (replace with 1 if unweighted estimation is desired)
* budget shares: s1 to sneq
* prices: p1 to nprice
* log total expenditures: x
* implicit utility: y, or related names
* demographic characteristics: z1 to zndem
g obs_weight=wgt
g s1=sfoodh
g s2=sfoodr
g s3=srent
g s4=soper
g s5=sfurn
g s6=scloth
g s7=stranop
g s8=srecr
g s9=spers
g p1=pfoodh
g p2=pfoodr
g p3=prent
g p4=poper
g p5=pfurn
g p6=pcloth
g p7=ptranop
g p8=precr
g p9=ppers
* polynomial systems are easier to estimate if you normalise the variable in the polynomial
g x=log_y
*egen mean_log_y=mean(log_y)
*replace x=log_y-mean_log_y
* normalised prices are what enter the demand system
* generate normalised prices, backup prices (they get deleted), and pAp, pBp
global nplist ""
forvalues j=1(1)$neq {
    g np`j'=p`j'-p$nprice
    global nplist "$nplist np`j'"
}
forvalues j=1(1)$neq {
    g np`j'_backup=np`j'
    g Ap`j'=0
```

```
    g Bp`j'=0
}
g pAp=0
g pBp=0
*list demographic characteristics: fill them in, and add them to zlist below
g z1=age
g z2=hsex
g z3=carown
g z4=tran
g z5=time
global zlist "z1 z2 z3 z4 z5"
*make pz interactions
global npzlist ""
forvalues j=1(1)$neq {
    forvalues k=1(1)$ndem {
        g np`j'z`k'=np`j'*z`k'
        global npzlist "$npzlist np`j'z`k'"
    }
}
*make y_stone=x-p'w, and gross instrument, y_tilda=x-p'w^bar
g y_stone=x
g y_tilda=x
forvalues num=1(1)$nprice {
    egen mean_s`num'=mean(s`num')
    replace y_tilda=y_tilda-mean_s`num'*p`num'
    replace y_stone=y_stone-s`num'*p`num'
}
* make list of functions of (implicit) utility, y: fill them in, and add them to ylist below
* alternatively, fill ylist and yinstlist with the appropriate variables and instruments
g y=y_stone
g y_inst=y_tilda
global ylist ""
global yinstlist ""
global yzlist ""
global yzinstlist ""
global ynplist ""
global ynpinstlist ""
forvalues j=1(1)$npowers {
    g y`j'=y^`j'
```

```
	g y‘j’_inst=y_inst^‘j’
	global ylist "$ylist y‘j’"
	global yinstlist "$yinstlist y‘j’_inst"
}
forvalues k=1(1)$ndem {
	g yz‘k’=y*z‘k’
	g yz‘k’_inst=y_inst*z‘k’
	global yzlist "$yzlist yz‘k’"
	global yzinstlist "$yzinstlist yz‘k’_inst"
}
forvalues k=1(1)$neq {
	g ynp‘k’=y*np‘k’
	g ynp‘k’_inst=y_inst*np‘k’
	global ynplist "$ynplist ynp‘k’"
	global ynpinstlist "$ynpinstlist ynp‘k’_inst"
}
*set up the equations and put them in a list
global eqlist ""
forvalues num=1(1)$neq {
	global eq‘num’ "(s‘num’ $ylist $zlist $yzlist $nplist $ynplist $npzlist)"
	macro list eq‘num’
	global eqlist "$eqlist \$eq‘num’"
}
*create linear constraints and put them in a list, called conlist
global conlist ""
forvalues j=1(1)$neq {
	local jplus1=‘j’+1
	forvalues k=‘jplus1’(1)$neq {
		constraint ‘j’‘k’ [s‘j’]np‘k’=[s‘k’]np‘j’
		global conlist "$conlist ‘j’‘k’"
	}
}
*add constraints for yp interactions
forvalues j=1(1)$neq {
	local jplus1=‘j’+1
	forvalues k=‘jplus1’(1)$neq {
		constraint ‘j’‘k’0 [s‘j’]ynp‘k’=[s‘k’]ynp‘j’
		global conlist "$conlist ‘j’‘k’0"
	}
```

```
}
* add constraints for pz interactions
forvalues h=1(1)$ndem {
    forvalues j=1(1)$neq {
        local jplus1=`j'+1
        forvalues k=`jplus1'(1)$neq {
            constraint `j'`k'`h' [s`j']np`k'z`h'=[s`k']np`j'z`h'
            global conlist "$conlist `j'`k'`h'"
        }
    }
}
*an approximate model would use one of:
*reg3 $eqlist [aweight=obs_weight], constr($conlist) endog($ylist $ynplist $yzlist) exog($yinstlist $ynpinstlist $yzinstlist)
*sureg $eqlist, constr($conlist)
*sureg $eqlist
*the exact model requires two steps: step 1) get a pre-estimate to construct the intrument, step 2) use the instrument to estimate the model
*first get a pre-estimate to create the instrument:
*run three stage least squares on the model with py, pz or yz interactions, and then iterate to convergence,
* constructing y=(y_stone+0.5*p'A(z)p)/(1-0.5*p'Bp) at each iteration
* note that the difference in predicted values for y=1 between p and p=0 is A(z)p, and
* that the difference in difference in predicted values for y=1 vs y=0 between p and p=0 is Bp
replace y=y_stone
g y_backup=y_stone
g y_old=y_stone
g y_change=0
scalar crit_test=1
scalar iter=0
while crit_test>$conv_crit {
    scalar iter=iter+1
    quietly reg3 $eqlist [aweight=obs_weight], constr($conlist) endog($ylist $ynplist $yzlist) exog($yinstlist $ynpinstlist $yzinstlist)
    if (iter>1) {
        matrix params_old=params
    }
    matrix params=e(b)
    quietly replace pAp=0
    quietly replace pBp=0
```

```
quietly replace y_old=y
quietly replace y_backup=y
*predict with y=1
*generate rhs vars,interactions with y=1
forvalues j=1(1)$npowers {
    quietly replace y`j'=1
}
forvalues j=1(1)$neq {
    quietly replace ynp`j'=np`j'
}
forvalues j=1(1)$ndem {
    quietly replace yz`j'=z`j'
}
*generate predicted values
forvalues j=1(1)$neq {
    quietly predict s`j'hat_y1, equation(s`j')
}
*set all p, pz, py to zero
foreach yvar in $nplist $ynplist $npzlist {
    quietly replace `yvar'=0
}
forvalues j=1(1)$neq {
    quietly predict s`j'hat_y1_p0, equation(s`j')
}

*refresh p,pz
forvalues j=1(1)$neq {
    quietly replace np`j'=np`j'_backup
    forvalues k=1(1)$ndem {
        quietly replace np`j'z`k'=np`j'_backup*z`k'
    }
}

*generate rhs vars,interactions with y=0
foreach yvar in $ylist $ynplist $yzlist {
    quietly replace `yvar'=0
}
*generate predicted values
forvalues j=1(1)$neq {
```

```
    quietly predict s`j'hat_y0, equation(s`j')
}
*set all p, pz, py to zero
foreach yvar in $nplist $ynplist $npzlist {
    quietly replace `yvar'=0
}
forvalues j=1(1)$neq {
    quietly predict s`j'hat_y0_p0, equation(s`j')
}
*refresh p only
forvalues j=1(1)$neq {
    quietly replace np`j'=np`j'_backup
}

*fill in pAp and pBp
forvalues j=1(1)$neq {
    quietly replace Ap`j'=s`j'hat_y0-s`j'hat_y0_p0
    quietly replace pAp=pAp+np`j'*Ap`j'
    quietly replace Bp`j'=(s`j'hat_y1-s`j'hat_y1_p0)-(s`j'hat_y0-s`j'hat_y0_p0)
    quietly replace pBp=pBp+np`j'*Bp`j'
    quietly drop s`j'hat_y0 s`j'hat_y0_p0 s`j'hat_y1 s`j'hat_y1_p0
}
*round pAp and pBp to the nearest millionth, for easier checking
quietly replace pAp=int(1000000*pAp+0.5)/1000000
quietly replace pBp=int(1000000*pBp+0.5)/1000000
*recalculate y,yz,py,pz
quietly replace y=(y_stone+0.5*pAp)/(1-0.5*pBp)
forvalues j=1(1)$npowers {
    quietly replace y`j'=y^`j'
}
forvalues j=1(1)$ndem {
    quietly replace yz`j'=y*z`j'
}
*refresh py,pz
forvalues j=1(1)$neq {
    quietly replace ynp`j'=y*np`j'_backup
    forvalues k=1(1)$ndem {
        quietly replace np`j'z`k'=np`j'_backup*z`k'
    }
```

```
        }
        if (iter>1 & conv_param==1) {
            matrix params_change=(params-params_old)
            matrix crit_test_mat=(params_change*(params_change'))
            svmat crit_test_mat, names(temp)
            scalar crit_test=temp
            drop temp
        }
        quietly replace y_change=abs(y-y_old)
        quietly summ y_change
        if(conv_y==1) {
            scalar crit_test=r(max)
        }
        display "iteration " iter
        scalar list crit_test
        summ y_change y_stone y y_old pAp pBp
    }
    *now, create the instrument, and its interactions yp and yz
    quietly replace y_inst=(y_tilda+0.5*pAp)/(1-0.5*pBp)
    forvalues j=1(1)$npowers {
        quietly replace y`j'_inst=y_inst^`j'
    }
    forvalues j=1(1)$neq {
        replace ynp`j'_inst=y_inst*np`j'
    }
    forvalues j=1(1)$ndem {
        replace yz`j'_inst=y_inst*z`j'
    }
    *with nice instrument in hand, run three stage least squares on the model, and then iterate to convergence
    replace y_old=y
    replace y_change=0
    scalar iter=0
    scalar crit_test=1
    while crit_test>$conv_crit {
        scalar iter=iter+1
        quietly reg3 $eqlist [aweight=obs_weight], constr($conlist) endog($ylist $ynplist $yzlist) exog($yinstlist
$ynpinstlist $yzinstlist)
        if (iter>1) {
            matrix params_old=params
```

```
}
matrix params=e(b)
quietly replace pAp=0
quietly replace pBp=0
quietly replace y_old=y
quietly replace y_backup=y
*predict with y=1
*generate rhs vars,interactions with y=1
forvalues j=1(1)$npowers {
    quietly replace y`j'=1
}
forvalues j=1(1)$neq {
    quietly replace ynp`j'=np`j'
}
forvalues j=1(1)$ndem {
    quietly replace yz`j'=z`j'
}
*generate predicted values
forvalues j=1(1)$neq {
    quietly predict s`j'hat_y1, equation(s`j')
}
*set all p, pz, py to zero
foreach yvar in $nplist $ynplist $npzlist {
    quietly replace `yvar'=0
}
forvalues j=1(1)$neq {
    quietly predict s`j'hat_y1_p0, equation(s`j')
}

*refresh p,pz
forvalues j=1(1)$neq {
    quietly replace np`j'=np`j'_backup
    forvalues k=1(1)$ndem {
        quietly replace np`j'z`k'=np`j'_backup*z`k'
    }
}

*generate rhs vars,interactions with y=0
foreach yvar in $ylist $ynplist $yzlist {
```

```
	quietly replace `yvar'=0
}
*generate predicted values
forvalues j=1(1)$neq {
	quietly predict s`j'hat_y0, equation(s`j')
}
*set all p, pz, py to zero
foreach yvar in $nplist $ynplist $npzlist {
	quietly replace `yvar'=0
}
forvalues j=1(1)$neq {
	quietly predict s`j'hat_y0_p0, equation(s`j')
}
*refresh p only
forvalues j=1(1)$neq {
	quietly replace np`j'=np`j'_backup
}

*fill in pAp and pBp
forvalues j=1(1)$neq {
	quietly replace Ap`j'=s`j'hat_y0-s`j'hat_y0_p0
	quietly replace pAp=pAp+np`j'*Ap`j'
	quietly replace Bp`j'=(s`j'hat_y1-s`j'hat_y1_p0)-(s`j'hat_y0-s`j'hat_y0_p0)
	quietly replace pBp=pBp+np`j'*Bp`j'
	quietly drop s`j'hat_y0 s`j'hat_y0_p0 s`j'hat_y1 s`j'hat_y1_p0
}
*round pAp and pBp to the nearest millionth, for easier checking
quietly replace pAp=int(1000000*pAp+0.5)/1000000
quietly replace pBp=int(1000000*pBp+0.5)/1000000
*recalculate y,yz,py,pz
quietly replace y=(y_stone+0.5*pAp)/(1-0.5*pBp)
forvalues j=1(1)$npowers {
	quietly replace y`j'=y^`j'
}
forvalues j=1(1)$ndem {
	quietly replace yz`j'=y*z`j'
}
*refresh py,pz
forvalues j=1(1)$neq {
```

```
        quietly replace ynp`j'=y*np`j'_backup
        forvalues k=1(1)$ndem {
            quietly replace np`j'z`k'=np`j'_backup*z`k'
        }
    }
    if (iter>1 & conv_param==1) {
        matrix params_change=(params-params_old)
        matrix crit_test_mat=(params_change*(params_change'))
        svmat crit_test_mat, names(temp)
        scalar crit_test=temp
        drop temp
    }
    quietly replace y_change=abs(y-y_old)
    quietly summ y_change
    if(conv_y==1) {
        scalar crit_test=r(max)
    }
    display "iteration " iter
    scalar list crit_test
    summ y_change y_stone y y_old pAp pBp
}
*note that reported standard errors are wrong for iterated estimates
reg3 $eqlist [aweight=obs_weight], constr($conlist) endog($ylist $ynplist $yzlist) exog($yinstlist $ynpinstlist
$yzinstlist)
```

References

[1] **Banks, James; Blundell, Richard and Lewbel, Arthur.** "Quadratic Engel curves and Consumer Demand," Review of Economics and Statistics, 1997, 79, 527–539.

[2] **Beckert, Walter and Blundell, Richard.** "Invertibility of Nonparametric Stochastic Demand Functions," 2004, Birkbeck Working Papers in Economics and Finance No. 406.

[3] **Blundell, Richard; Chen, Xiaohong and Kristensen, Dennis.** "Nonparametric IV Estimation of Shape-Invariant Engel curves," Econometrica, 2007, 75, 1613–1669.

[4] **Blundell, Richard; Duncan, Alan and Pendakur, Krishna.** "Semiparametric Estimation of Consumer Demand", Journal of Applied Econometrics, 1998, 13, 435-461.

[5] **Blundell, Richard and Robin, Jean-Marc.** "Estimation in Large and Disaggregated Demand Systems: An Estimator for Conditionally Linear Systems," Journal of Applied Econometrics, 1999, 14, 209-232.

[6] **Brown, Donald J. and Matzkin, Rosa L.** "Estimation of Nonparametric Functions in Simultaneous Equations Models, with an Application to Consumer Demand," 1998, Cowles Foundation Discussion Papers 1175, Cowles Foundation, Yale University.

[7] **Brown, Bryan W. and Walker, Mary Beth.** "The Random Utility Hypothesis and Inference in Demand Systems" Econometrica, 1989, 57, 815-829

[8] **Deaton, Angus S., and Muellbauer, John.** "An Almost Ideal Demand System," American Economic Review, 1980, vol. 70, 312-326.

[9] **Gorman, William M.** "Some Engel Curves," 1981, in Essays in the Theory and Measurement of Consumer Behaviour in Honor of Sir Richard Stone, ed. by Angus Deaton, Cambridge: Cambridge University Press.

[10] **Hansen, Lars.** "Large Sample Properties of Generalized Method of Moments Estimators," Econometrica, 1982, 50, 1029-1054.

[11] **Lewbel, Arthur.** "The Rank of Demand Systems: Theory and Nonparametric Estimation," Econometrica, 1991, 59, 711-730.

[12] **Lewbel, Arthur.** "Demand Systems with and without Errors," American Economic Review, 2001, 91, 611-618.

[13] **Matzkin, Rosa L.** "Heterogeneous Choice," 2005, in Advances in Economics and Econometrics, edited by Richard Blundell, Whitney Newey, and Torsten Persson, Cambridge University Press; presented at the Invited Symposium on Modeling Heterogeneity, World Congress of the Econometric Society, London, U.K.

[14] **McFadden, Daniel. and Richter, Marcel K.** "Stochastic Rationality and Revealed Stochastic Preference," 1990, in J. Chipman, D. McFadden, M. K. Richter (eds) Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz, 161-186, Westview Press: Boulder, CO.

[15] **Pendakur, Krishna.** "Estimates and Tests of Base-Independent Equivalence Scales," Journal of Econometrics, 1999, 88, 1-40.

[16] **Pendakur, Krishna.** "Taking Prices Seriously in the Measurement of Inequality," Journal of Public Economics, 2002, 86, 47-69.

[17] **Pendakur, Krishna and Sperlich, Stefan.** "Semiparametric Estimation of Consumer Demand Systems in Real Expenditure," 2008, forthcoming, Journal of Applied Econometrics.