Integrability and the Semiparametric Estimation of Consumer Demand Systems

Krishna Pendakur, Economics, Simon Fraser University
Burnaby, BC, CANADA, V5A 1S6, pendakur@sfu.ca

Received: _______________; Accepted: _______________

Abstract

This paper uses local polynomial modelling to estimate expenditure share equations over prices and expenditure levels, which yields local derivatives that are easy to estimate. Since the Slutsky matrix is a function of levels and derivatives of share equations, this permits the easy estimation of the Slutsky matrix which is useful in the measurement of consumer surplus. This framework also allows the researcher to restrict the expenditure share equations to locally satisfy homogeneity and the Slutsky matrix to locally satisfy symmetry, and yields Wald-like tests of local homogeneity and local symmetry. Procedures are also developed to measure exact consumer surplus, incorporate demographic effects and simulate confidence bands. Consumer demand systems are estimated with Canadian price and expenditure data for the period 1969 to 1999. Estimated expenditure share equations, Slutsky matrices, and consumer surplus estimates using homogeneity- and symmetry-restricted estimates are presented.

I acknowledge Doug Allen, Orazio Attanasio, James Banks, Richard Blundell, Martin Browning, Andrew Chesher, Ian Crawford, Tom Crossley, David Donaldson, Hide Ichimura, Brian Krauth and Joris Pinkse, Ian Preston, Clyde Reed, Abijit Sengupta and numerous other patient colleagues for their help on this work. I also gratefully acknowledge the support of the Social Sciences and Humanities Research Council of Canada through their Standard Grant Program, and through the Equality, Security and Community collaborative research initiative.
1. Introduction

Consumer demand equations over prices and expenditure must satisfy homogeneity and have a Slutsky matrix which satisfies symmetry and concavity in order to satisfy integrability. However, previous non- and semi-parametric approaches to the estimation of consumer demand do not give estimates restricted to satisfy these conditions. The current paper fills this gap.

Integrable demand systems are desirable because they can be used for welfare analysis, whereas demand systems that are not integrable cannot be so used. In particular, although demand systems that are not integrable may be used for prediction purposes, they cannot be used to measure consumer surplus or to adjust for price differences in inequality and poverty measurement. Integrable demand systems allow for the estimation of exact consumer surplus over a price change for each family or individual in a population. Micro-level consumer surplus measured with an integrable demand system will satisfy path-independence, that is, will not depend on the sequence of small price changes used to approximate the large price change (see Vartia 1983). These micro-level consumer surpluses may then be “added up” with a social welfare function or other aggregator to generate measures of “social surplus” which can be used for policy evaluation. Similarly, integrable demand systems allow for the estimation of exact cost-of-living indices for each family or individual in a population. Such micro-level price indices can be used to adjust for price differences across families or individuals when measuring economic inequality or poverty.

The primary aim of this paper is to develop a methodology for the semiparametric estimation of consumer demand systems under the restrictions of integrability and for the
semiparametric estimation of path-independent consumer surplus measures. The key innovation in this paper is the use of local polynomial modelling to estimate expenditure share equations over prices and expenditure levels. The main advantage of local polynomial modelling is that local derivatives are easy to estimate, and because the Slutsky matrix is a function of levels and derivatives of share equations, it, too, is easy to estimate. If the expenditure share equations satisfy homogeneity and the Slutsky matrix satisfies symmetry, then they may be used in consumer surplus measurement.

The local polynomial framework also allows the researcher to easily restrict the expenditure share equations to locally satisfy homogeneity and to restrict the Slutsky matrix to locally satisfy symmetry. Further, the local polynomial framework yields simple wald-like tests of local homogeneity and local symmetry.

One need not take the local polynomial estimator to a very high degree. In particular, since the Slutsky matrix only uses first order derivatives, I use a local quadratic model in this paper to estimate expenditure share equations and Slutsky terms, and to test homogeneity and symmetry. Restricted local quadratic estimates which satisfy homogeneity and symmetry are also presented.

Given local quadratic estimates that satisfy symmetry, a second-order Taylor approximation of the log-cost function may be used to obtain exact consumer surplus measures for large price changes. This is in the spirit of work by Vartia (1983), Hausman and Newey (1995) and Breslaw and Smith (1995) which uses Taylor approximations of the cost function to estimate exact consumer surplus. However, the current approach ensures that the surplus measures are locally path-independent due to the imposition of symmetry on the estimated local Slutsky Matrix.

I also present procedures for incorporating demographic effects into the model in such
a way that the estimated demand system locally satisfies homogeneity and symmetry for all possible vectors of demographic characteristics. I also develop procedures to estimate confidence intervals for estimated demand systems, Slutsky terms and homogeneity and symmetry test statistics via an extension of Gozalo’s (1997) Smooth Conditional Moment Bootstrap. The extension allows for cross equation correlations in the structure of simulated errors, which is important in this context because Slutsky symmetry involves cross-equation restrictions.

The paper proceeds as follows. First, I briefly discuss the nonparametric estimation of a consumer demand system. Then I discuss the testing and imposition of homogeneity, symmetry and (quite briefly) concavity in nonparametric local quadratic models. With homogeneity- and symmetry-satisfying estimates in hand, I discuss the estimation consumer surplus measures, focusing on the cost-of-living index, and the incorporation of demographic effects. Following this, the model is implemented using Canadian price and expenditure data from 1969 to 1999. I present estimated expenditure share equations and Slutsky terms using various kinds of local models and compare these estimates with those from a global (fully parametric) model.

2. Nonparametric Estimation of Consumer Demand

Parametric approaches have hitherto dominated the estimation of consumer demand systems for at least three reasons: (1) parametric models use global structure to reduce the number of observations required to precisely estimate models; (2) such models are easy to implement; and (3) such models allow the imposition of integrability—which is required for the estimation of consumer surplus—with relative ease. With the advent of very large micro-level expenditure data, nonparametric methods have been brought to bear on consumer demand
problems. A few papers in this literature have focussed on maintaining integrability in the nonparametric estimation of expenditure share equations (see, e.g., Blundell, Duncan and Pendakur 1998; Pendakur 1999), but these papers have not estimated full consumer demand systems. Rather, they estimate expenditure shares one-by-one at a single price vector rather than attacking the full demand system which estimates many expenditure shares at many price vectors imposing all the within- and cross-equation restrictions of integrability.

Several papers have developed nonparametric approaches to the estimation of the Slutsky matrix and consumer surplus. Vartia (1983), Hausman and Newey (1995) and Breslaw and Smith (1995) use taylor approximations to the local cost function to estimate consumer surplus over large price changes. Other researchers, such as Lewbel (1995) and Hausman and Newey (1995) use nonparametric methods to test the symmetry restrictions. However, none of these papers offer methods to estimate subject to the restrictions of homogeneity, symmetry and concavity which together make up integrability.

2.1 Local Quadratic Estimation

Define $P^j$ as the price of the j’th good, denote the log-price of the j’th good as $p^j = \ln(P^j)$, and denote $p = [p^1, ..., p^M]'$. Define $X$ as total expenditure, and denote log-expenditure as $x = \ln(X)$. Let $p_i = [p_i^1, ..., p_i^M]'$ denote the log-price vector, $x_i$ denote the log-expenditure and $w_i = [w_i^1, ..., w_i^M]'$ the expenditure share vector of the i’th expenditure unit (eg, a family) where $i = 1, ..., N$. Consider a local quadratic (LQ) model where in the neighbourhood of a price vector, $p$, and log total expenditure level, $x$,

$$w_i^j = \alpha^j + \beta_{k}^j (p^k - p_i^k) + \beta_{x}^j (x - x_i) + \frac{1}{2} \sum_{k=1}^{M} \sum_{l=1}^{M} \gamma_{kl}^j (p^k - p_i^k)(p^l - p_i^l) + \sum_{k=1}^{M} \gamma_{kk}^j (p^k - p_i^k)(x - x_i) + \sum_{k=1}^{M} \gamma_{xx}^j (x - x_i)^2 + \epsilon_i^j. (1)$$
The parameters of this local model may be estimated by minimising

$$
\min_{\alpha_j, \beta_{jk}, \gamma_{jk}, \Omega} \sum_{i=1}^{N} \Omega_i \left( \epsilon_i^2 \right)
$$

for each independent expenditure share equation \( j = 1, \ldots, M-1 \). I note that for any \( M \) good demand system, there are only \( M-1 \) independent equations due to adding-up restrictions.

Here, \( \Omega \) is a multivariate kernel weighting function which penalises the distance of \((p_i, x_i)\) from \((p, x)\). I specify this weighting function as

$$
\Omega_i = \prod_{t=1}^{M+1} K(u_t^{i})
$$

where \( K \) is the gaussian kernel

$$
K(u_i) = \frac{\exp(-u_i^2/2)}{\sqrt{2\pi}}
$$

and the \( M+1 \) vector \( u_i = [u_1^i, \ldots, u_{M+1}^i]^T \) is given by

$$
u_i = H^{-1} \overline{u}_i
$$

where \( H \) is an \( M+1 \) by \( M+1 \) bandwidth matrix and \( \overline{u}_i \) is the distance of \((p_i, x_i)\) from the point of estimation \((p, x)\):

$$
\overline{u}_i = \begin{bmatrix} p_1^i \\ \vdots \\ p_M^i \\ x_i \end{bmatrix} - \begin{bmatrix} p^1 \\ \vdots \\ p^M \\ x \end{bmatrix}
$$

Here, \( \overline{u}_1, \ldots, \overline{u}_M \) give the distances of log-prices and \( \overline{u}_{M+1} \) gives the distance of log total expenditure.

Given the bandwidth matrix \( H \), estimation of the parameters can proceed by locally weighted ordinary least squares. That is, at each point of interest in \((p, x)\), the researcher
regresses $w^j$ on a constant, $(p_i - p)$, $x_i - x$, and their unique interaction terms using $\Omega_i$ as weights.

Local quadratic estimation gives nice boundary characteristics in comparison with local mean estimation (see Bowman and Azzalani 1999, or Wand and Jones 1995). In general finite-order local polynomial estimation is biased near the boundaries of the data, but this bias declines with the degree of the polynomial (although the bias decreases incrementally only for odd-ordered polynomials).

I note that this LQ model nests a local almost ideal model given by

$$w^j_i = \alpha^j + \sum_{k=1}^{M} \beta^j_k (p^k - p^k_i) + \beta^j_x (x - \ln a(p - p_i)) + \epsilon^j_i \quad (7)$$

where $a$ is a translog such that

$$\ln a(p - p_i) = \sum_{j=1}^{M} \alpha^j_i (p^j - p^j_i) + \frac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{M} \beta^j_k (p^j - p^j_i)^2 (p^k - p^k_i)$.

The parameters may be different at every $(p, x)$.

Although the LQ model nests the local almost ideal, the latter model is restrictive compared to the local quadratic. For example, the second order price terms in the local almost ideal are entirely determined by the intercepts and first order terms. This is because the first order terms are part of the function $a$ which shows up in the $x$ term.

It is also possible develop a local quadratic almost ideal (LQAI) model which has as much flexibility as, but is not nested within, the LQ model. This model is given by

$$w^j_i = \alpha^j + \sum_{k=1}^{M} \beta^j_k (p^k - p^k_i) + \frac{1}{2} \sum_{k=1}^{M} \sum_{l=1}^{M} \gamma^j_{kl} (p^k - p^k_i)(p^l - p^l_i) + \gamma^j_x (x - \ln a(p - p_i)) + \epsilon^j_i \quad (9)$$

1 The LQ model also nests a local extended almost ideal model (see Dickens, Fry and Pashardes 1993) given by

$$w^j_i = \frac{\partial \ln a(\cdot, \cdot, \cdot)}{\partial p^j} - \frac{\partial \ln b(\cdot, \cdot, \cdot)}{\partial p^j} (x - \ln a(\cdot, \cdot, \cdot)) + \frac{\partial q(\cdot, \cdot, \cdot)}{\partial p^j} (x - \ln a(\cdot, \cdot, \cdot))^2 \quad (8)$$

where $a$ is cobb-douglas and homogeneous of degree one in prices, $b$ is cobb-douglas and homogeneous of degree zero in prices and $q$ is proportional to $b$. The functions $a$, $b$ and $q$ may be different at every $(p, x)$. 

7
\[
\left( \beta^j_x + \sum_{k=1}^{M} \gamma^j_{k,x} (p^k - p_i^k) \right) (x - \ln a(\cdot)) + \frac{\gamma^j_{x,x}}{b(\cdot)} (x - \ln a(\cdot))^2 + c^j
\]

where \(a\) is a third-order translog that may be different at each point,

\[
\ln a(p - p_i) = \sum_{j=1}^{M} \alpha^j (p^j - p_i^j) + \frac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{M} \beta^j_k (p^j - p_i^j)(p^k - p_i^k) +
\]

\[
\frac{1}{6} \sum_{j=1}^{M} \sum_{k=1}^{M} \sum_{l=1}^{M} \gamma^j_{k,l} (p^j - p_i^j)(p^k - p_i^k)(p^l - p_i^l),
\]

and \(b\) is a translog that may be different at each point,

\[
\ln b(p - p_i) = \sum_{j=1}^{M} \beta^j_x (p^j - p_i^j) + \frac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{M} \gamma^j_{k,x} (p^j - p_i^j)(p^k - p_i^k).
\]

Adding-up implies \(\alpha^M = 1 - \sum_{j=1}^{M-1} \alpha^j\), \(\beta^M_k = -\sum_{j=1}^{M-1} \beta^j_k\), \(\gamma^M_{k,l} = -\sum_{j=1}^{M-1} \gamma^j_{k,l}\), \(\beta^M_x = -\sum_{j=1}^{M-1} \beta^j_x\), and \(\gamma^M_{k,x} = -\sum_{j=1}^{M-1} \gamma^j_{k,x}\). This specification is more complex than that typically used in the global parametric estimation of consumer demand (see, eg, Banks, Blundell and Lewbel 1997), and gives expenditure shares characterised by second-order translogs in prices and expenditure, plus other more complicated terms involving the functions \(a\) and \(b\).

The LQAII model may be estimated by minimising \(\sum_{i=1}^{N} \Omega_i \left( c_i^j \right)^2 \) via nonlinear estimation, but Browning and Meghir (1991) introduce an iterated linearised estimation strategy for this type of model which is much easier to implement. In particular, they suggest an iterative approach wherein the researcher treats \(a\) and \(b\) as fixed in each iteration. This gives a linear problem within iterations, for which one may use weighted (by \(\Omega_i\)) ordinary least squares to identify the parameters. Then, since these are the only parameters used in \(a\) and \(b\), \(a\) and \(b\) may be updated in the next iteration, and so on. The researcher iterates until \(a\) and \(b\) stop changing, which is usually in a matter of a three to six iterations. Blundell and Robin (1999) show that the parameter estimates resulting from this approach are asymptotically equivalent to those that might be gotten via nonlinear estimation.
An important feature of local polynomial models is that the estimated values and derivatives are easy to compute locally, that is, at the point of estimation. In particular, the estimated local values of \( w^j \) are given by the estimated intercept terms and the estimated local first derivatives of \( w^j \) with respect to \((p, x)\) are the estimated local linear parameters (see, eg, Pagan and Ullah 1999, Wand and Jones 1995; Simonoff 1995). In the LQ model, the estimated local level of each expenditure share is,

\[
\hat{w}^j(p, x) = \alpha^j, \tag{12}
\]

and the estimated local derivatives of the expenditure shares are

\[
\frac{\partial \hat{w}^j(p, x)}{\partial p_k} = \beta^j_k \tag{13}
\]

and

\[
\frac{\partial \hat{w}^j(p, x)}{\partial x} = \beta^j_x. \tag{14}
\]

In the LQAI model, the estimated local level of each expenditure share is,

\[
\hat{w}^j(p, x) = \alpha^j, \tag{15}
\]

and the estimated local derivatives of the expenditure shares are

\[
\frac{\partial \hat{w}^j(p, x)}{\partial p_k} = \beta^j_k - \beta^j_x \alpha^k \tag{16}
\]

and

\[
\frac{\partial \hat{w}^j(p, x)}{\partial x} = \beta^j_x. \tag{17}
\]

The simplicity of these estimated expenditure shares and expenditure share derivatives with respect to log-prices and log-expenditure allows a straightforward approach to restricting the demand system to satisfy homogeneity, estimating Slutsky terms and restricting the estimated Slutsky matrix to satisfy symmetry.
2.2 Local Homogeneity

Homogeneity requires that the demand system be independent of scalings of prices and total expenditure. In particular, for any function homogeneity of degree zero requires that its derivatives with respect to log arguments sum to zero. Thus, for each expenditure share equation, we require

$$\sum_{k=1}^{M} \frac{\partial w^j(p, x)}{\partial p^k} + \frac{\partial w^j(p, x)}{\partial x} = 0.$$  \hspace{1cm} (18)

Substituting the estimated derivatives, we see that the LQ model locally satisfies homogeneity if

$$\sum_{s=1}^{M} \beta^j_s - \beta^j_x = 0$$ \hspace{1cm} (19)

for all $j = 1, \ldots, M - 1$.

This restriction is easily tested because it is a linear restriction and the parameters of LQ models (indeed all local polynomials) are distributed asymptotically normal (Pagan and Ullah 1999). Defining $\Psi$ as the list of all parameters,

$$\tau^H_j = \sum_{s=1}^{M} \beta^j_s - \beta^j_x,$$  \hspace{1cm} (20)

one constructs the Wald-like test statistic

$$\tau^{H, om} = \begin{bmatrix} \tau^H_j \\ \vdots \\ \tau^H_{M-1} \end{bmatrix}^T \begin{bmatrix} \partial \tau^H_j / \partial \Psi \\ \vdots \\ \partial \tau^H_{M-1} / \partial \Psi \end{bmatrix} V(\Psi) \begin{bmatrix} \partial \tau^H_j / \partial \Psi \\ \vdots \\ \partial \tau^H_{M-1} / \partial \Psi \end{bmatrix}^{-1} \begin{bmatrix} \tau^H_j \\ \vdots \\ \tau^H_{M-1} \end{bmatrix}$$  \hspace{1cm} (21)

and asks whether or not it is large compared to a $\chi^2_{M-1}$. Since the derivatives in $\tau^{H, om}$ are all constants this is easy to implement.

Obviously, the crucial element of the test statistic $\tau^{H, om}$ is the covariance matrix for the estimated parameters $V(\Psi)$. One can use the asymptotic covariance matrix (eg, see Pagan
and Ullah 1999), or one might estimate it by simulation. In this paper, I use the latter method.

The homogeneity restriction is easily imposed in LQ context by estimating the following model:

\[
\begin{align*}
w_i^j &= \alpha_i^j + \sum_{k=1}^{M} \beta_{i k}^j (\hat{p}_k^k - \hat{\tau}_k^i) + \frac{1}{2} \sum_{k=1}^{M} \sum_{l=1}^{M} \gamma_{k l}^j (p_{k}^k - p_{i}^k) (p_{l}^l - p_{i}^l) + \\
&\quad \sum_{k=1}^{M} \gamma_{k x}^j (p_{k}^k - p_{i}^k) (x_i - x_i) + \sum_{k=1}^{M} \gamma_{x x}^j (x_i - x_i)^2 + \epsilon_i^j,
\end{align*}
\]

where \(\hat{p}_k^k\) is the \(k\)'th log normalised price, defined as

\[
\hat{p}_k^k = p_k^k - x.
\]

Homogeneity may be similarly imposed in the LQAI model as a set of linear (within iterations) restrictions. Substituting the LQAI estimates into (18) gives

\[
\sum_{k=1}^{M} \beta_{i k}^j - \beta_{i x}^j \alpha_k^k + \beta_{x x}^j = 0,
\]

and noting that \(\sum_{k=1}^{M} \alpha_k^k = 1\), we get

\[
\sum_{k=1}^{M} \beta_{i k}^j = 0,
\]

and the analogous test statistic

\[
T_{H,LQAI}^H = \sum_{s=1}^{M} \beta_{i s}^j
\]

This set of restrictions may be imposed easily in the LQAI via linearly restricted OLS within iterations.

For both the LQ and LQAI models, imposition of the linear restriction required for homogeneity still leaves one with a locally weighted OLS problem which is computationally quick and easy to do. It is also worth noting that the imposition of homogeneity reduces by one order the curse of dimensionality faced by the researcher.
2.3 Local Symmetry

Define the cost function, $C(p, u)$, to give the minimum cost to attain utility level $u$ facing log-prices $p$. The Slutsky Matrix $S$ with elements $S_{jk}$ is defined as the Hessian of the cost function with respect to (unlogged) prices, and may be expressed in terms of log-price derivatives as follows (see Pollak and Wales 1991, p 58):

$$S_{jk} = \frac{\partial w^j}{\partial p^k} + \frac{\partial w^j}{\partial x} w^k + w^j w^k - d^j w^j$$  \hspace{1cm} (26)

where $d^j$ is an indicator of a diagonal element of the Slutsky matrix.

Slutsky symmetry is satisfied if and only if

$$S_{jk} = S_{kj}.$$  \hspace{1cm} (27)

If this condition holds continuously over a region of the $(p, x)$ space, then Young’s Theorem guarantees that there exists a cost function whose derivatives could produce the observed demand system over this region.

Substituting the estimates (12), (13) and (14) into (26), we get the local estimate of the elements of the Slutsky matrix for the LQ model:

$$\hat{S}_{jk} = \beta^j_k + \beta^j_x \alpha^k + \alpha^j \alpha^k - d^j \alpha^j.$$  \hspace{1cm} (28)

Setting $\hat{S}_{jk} = \hat{S}_{kj}$, local Slutsky symmetry requires that

$$\left( \beta^j_k + \beta^j_x \alpha^k \right) - \left( \beta^k_j + \beta^k_x \alpha^j \right) = 0$$  \hspace{1cm} (29)

for all $j \neq k$.

If homogeneity is also imposed, then the $\beta^k_x$ terms are computed from (19), giving

$$\left( \beta^j_k - \sum_{s=1}^{M} \beta^j_s \alpha^k \right) - \left( \beta^k_j - \sum_{s=1}^{M} \beta^k_s \alpha^j \right) = 0.$$  \hspace{1cm} (30)
This framework yields a Wald-like test of symmetry. Under the null that the estimated Slutsky Matrix is symmetric, difference between $S_{jk}$ and $S_{kj}$ is zero, and under the alternative, it is not. Thus, we may test symmetry of the estimated Slutsky Matrix by defining

$$
\tau^S_{j,k} = \left( \beta^j_k + \beta^j_x \alpha^k \right) - \left( \beta^k_j + \beta^k_x \alpha^j \right)
$$

and constructing the following test statistic

$$
\tau_{\text{Sym}} = \begin{bmatrix} 
\tau^S_{j,k} \\
\vdots \\
\tau^S_{M-2,M-1}
\end{bmatrix}^T \begin{bmatrix}
\partial \tau^S_j / \partial \Psi \\
\vdots \\
\partial \tau^S_{M-2,M-1} / \partial \Psi
\end{bmatrix} V(\Psi) \begin{bmatrix}
\partial \tau^S_j / \partial \Psi \\
\vdots \\
\partial \tau^S_{M-2,M-1} / \partial \Psi
\end{bmatrix}^{-1} \begin{bmatrix}
\tau^S_j \\
\vdots \\
\tau^S_{M-2,M-1}
\end{bmatrix}.
$$

(32)

Under the null hypothesis, $\tau_{\text{Sym}}$ is distributed chi-squared with $(M-1)(M-2)/2$ degrees of freedom. So, to test symmetry, we ask whether or not the observed value of $\tau_{\text{Sym}}$ is large compared to this distribution.²

The restrictions (29) and (30) for the LQ model can be quite complicated in practice. For example, with 4 prices and 3 independent expenditure share equations the following 3 restrictions could be imposed to satisfy (30):

$$
\beta_1^2 - \frac{-\beta_2 + \alpha^2 \beta_1 + \alpha^2 \beta_2 + \alpha^2 \beta_1 + \alpha^2 \beta_4 - \alpha_1 \beta_2 - \alpha_1 \beta_3 - \alpha_1 \beta_4}{-1 + \alpha^1} = 0,
$$

$$
\beta_3^3 - \frac{-\alpha_3 \beta_1 - \alpha_3 \beta_2 - \alpha_3 \beta_3 + \alpha_3 \beta_1 + \alpha_3 \beta_2 + \alpha_3 \beta_4 - \alpha_3 \beta_1 + \alpha_3 \beta_2 + \alpha_3 \beta_3}{1 - \alpha^2 - 2 \alpha^1 + \alpha^2 \alpha^1 + \alpha^1 \alpha^1} = 0,
$$

² Symmetry restricts functions of levels and derivatives of expenditure share equations (unlike homogeneity which restricts only derivatives). Since in nonparametric estimation, levels converge faster than derivatives, in the tests the level terms are treated as constants. Thus, $\tau_{\text{Sym}}$ depends only on the covariance matrix of the derivatives of the expenditure share equations.
and
\[ \beta_2^2 - \frac{\alpha^1 \beta_3^2 - \beta_3^2 + \alpha^3 \beta_2^2 + \alpha^3 \beta_3^2 + \alpha^1 \beta_4^2 - \alpha^2 \beta_3^2 - \alpha^2 \beta_4^2 + \alpha^3 \beta_2^1}{-1 + \alpha^2 + \alpha^1} = 0. \] (35)

Since the local symmetry restrictions (29) for the LQ model are nonlinear, we cannot impose them via linearly restricted OLS as we could with homogeneity. However, one might estimate under the restrictions (29) via nonlinear methods, for example, locally weighted seemingly unrelated regression estimation with nonlinear restrictions.

For the LQAI model, the symmetry restrictions are much easier to impose. Substitution of the LQAI estimated levels and derivatives into (26) yields the following symmetry restrictions:
\[ \beta_j^k = \beta_k^j \] (36)
for all \( j, k \). The analogous test statistic may be constructed as
\[ \tau_{S,LQAI}^{j,k} = \beta_j^k - \beta_k^j, \] (37)
and these cross-equation linear restrictions may be easily imposed in the linearised model.

In this paper, I estimate homogeneity- and symmetry-restricted LQ and LQAI models. For comparison, I also estimate a global quadratic almost ideal (GQAI) model with \( a \) given by a third order translog and \( b \) given by a translog. In the global model, the parameters are the same everywhere so that there is only one ‘point of estimation’. In the GQAI, global homogeneity requires the additional restrictions that \( \sum_{r=1}^{M} \gamma_{kr}^j = 0 \) for all \( j, k \) and \( \sum_{r=1}^{M} \gamma_{rk}^j = 0 \) for all \( j \), and global symmetry requires the additional restrictions that \( \gamma_{kx}^j = \gamma_{jx}^k \) for all \( j, k \) and \( \gamma_{kl}^j = \gamma_{lk}^j = \gamma_{jk}^i = \gamma_{kj}^i \) for all \( j, k, l \).
2.3.1 Objections

One might argue that since the $\gamma_{kl}^j$ parameters in all these local models are not constrained by the restrictions, local linear estimation would suffice. However, several researchers (see Fan and Gijbels 1996) have noted that for derivative estimation, local polynomials perform better if the polynomial is at least one degree higher than that of the derivative of interest. Thus, for this application, local quadratics are appropriate because the derivatives of interest are given by the linear terms.

One might also argue that since Young's Theorem only gives integrability when the Slutsky Matrix is continuous on the domain of prices and expenditure, nonlinearly restricted estimation might be problematic. However, since the unrestricted estimated coefficients are just locally weighted least squares estimates, they are continuous by construction. This implies that the unconstrained (possibly inhomogeneous and asymmetric) Slutsky Matrix is continuous in log-prices and log-expenditure. Under the null that homogeneity and symmetry are true, the restricted estimates should not stray 'too far' from the unrestricted estimates, and so the homogeneity- and symmetry-restricted estimated Slutsky Matrix should also be continuous in prices and expenditure.

2.4 Local Concavity

If estimated expenditure share equations satisfy homogeneity and the estimated Slutsky Matrix satisfies symmetry, then estimated consumer surplus measures will satisfy path-independence and the weak axiom of revealed preference. However, to satisfy the strong axiom of revealed preference, the estimated Slutsky Matrix must also satisfy concavity.

Local concavity is satisfied if the Slutsky Matrix is a negative semidefinite matrix. One may use the Ryan and Wales (1999, 2000) strategy of locally estimating the Slutsky Matrix
as the negative of the inner product of a triangular matrix and its transpose. For the LQ, this requires that

\[ \beta^j_k + \beta^j_x \alpha^k + \alpha^j \alpha^k - \alpha^j d^j = -(\Lambda \Lambda')_{jk} \]  \hspace{1cm} (38)

for some lower triangular matrix \( \Lambda \).

Note that by construction \( \Lambda \Lambda' \) is symmetric. This means that imposition of concavity via restrictions (38) implies the symmetry restrictions (29). Beyond the symmetry restrictions (29), the concavity restrictions (38) impose only inequality constraints—for example, the diagonal elements of \( S \) must be weakly negative—and these are embodied in (38).

Estimation of the restricted parameters given the restrictions (38) may proceed via expressing \( -(\Lambda \Lambda')_{jk} \) in terms of the elements \( \lambda_{jk} \) of \( \Lambda \). Then, the researcher estimates \( \lambda_{jk} \), and computes \( \beta_k^j \) as functions of \( \Lambda \). In the local quadratic case, this is quite cumbersome. However, in the local quadratic almost ideal (LQAI) case, sufficient conditions for concavity restrictions are much easier to impose.

The local log-cost function for the LQAI demand system is given by

\[ \ln C(p_i, u_i) = \ln a(p_i - p) + \frac{b(p_i - p)u_i}{1 + q(p_i - p)u_i}. \]  \hspace{1cm} (39)

At the point of estimation, \( p_i = p \) and \( x_i - x \), so that \( u_i = 0 \). Thus, the local estimate of the Slutsky Matrix is

\[ S_{jk} = \frac{\partial^2 a(p_i - p)}{\partial P_j \partial P_k}, \]

so, if \( a \) is locally concave in unlogged prices, then \( C \) is locally concave. A sufficient condition for concavity of \( a \) in \( P \) is to restrict \( \ln a \) to be concave in \( p = \ln P \) (although this imposes ‘too much’ concavity). The symmetry-restricted estimates presented below locally satisfy concavity\(^3\), so concavity-restricted estimates are not presented in this paper.

\(^3\) The paper focuses on the estimated expenditure shares and Slutsky terms for observations between the 5th and 95th percentiles of the total expenditure distribution in a particular region and period, Ontario
3. Dealing with Endogeneity

The estimation procedures suggested thus far have assumed that all prices and expenditure are exogenous, and thus uncorrelated with the error terms in the nonparametric equation system. Many researchers have pointed out that total expenditure is typically endogenous, and so correction for the possible endogeneity of total expenditure is desirable. In this section, I consider one possible strategy for dealing with endogeneity following the lead of Blundell and Powell (2001).

Blundell and Powell (2001) discuss the ‘control function’ approach to correcting for endogeneity in nonparametric and semiparametric models. Assume that the LQ model (1) is the ‘structural function’ in which we are interested, but its error terms $\epsilon^j$ are correlated with log total-expenditure $x$. In this case, one can add a ‘control variable’ $v$ to the right hand side of (1) which ‘controls for’ this endogeneity. The control variable $v$ can be the residual of $x$ from its conditional mean given a set of exogenous variables $\chi$, which would include log-prices, $p$, and other instruments correlated with $x$ but not with $\epsilon^j$. Thus, $v$ may be estimated as the residual in a nonparametric regression of $x$ on $\chi$ as follows:

$$v_i = x_i - E[x_i|\chi_i]. \quad (40)$$

The control function approach restricts the stochastic dependence of $\epsilon^j$, $v$, $\chi$ and $x$ to allow the researcher to estimate the structural function.

Newey, Powell and Vella (1999) develop this approach for environments where the structural function has additive errors and the structural error terms $\epsilon^j$ are conditionally mean-independent of the endogenous variable(s), so that

$$E[\epsilon^j|x,\chi] = E[\epsilon^j|x,v] = E[\epsilon^j|v]. \quad (41)$$

1986. The estimated Slutsky matrix for symmetry-restricted LQ and LQAI models is negative semidefinite for these observations, so concavity is locally satisfied.
Because by assumption $\epsilon^j$ and $\chi$ are uncorrelated and by construction $v$ depends only on $x$ and $\chi$, $E[\epsilon^j|\chi] = E[\epsilon^j|v]$. The stochastic restriction is that this expectation is independent of $x$.

To simplify the discussion, consider with no loss of generality a local polynomial regression with just a single term $\alpha_j$ for the local mean—that is, a zero-order local polynomial—and show how it varies across $p, x$ by writing it as a function $\alpha^j(p, x)$:

$$w_i^j = \alpha^j(p, x) + \epsilon_i^j.$$  \hspace{1cm} (42)

In this case, Newey, Powell and Vella (1999) show that one can simply insert an additive component in $v$ to the nonparametric regression equation, giving

$$w_i^j = \alpha^j(p, x) + \eta_i v_i + \epsilon_i^j.$$  \hspace{1cm} (43)

We are interested in the average value of the structural function at exogenously given values of $p$ and $x$. This may be calculated by integrating over $v_i$ at those values. Imposing the normalisation $E[\eta^j v_i] = 0$, the average value of the structural function is given by $\alpha^j(p, x)$.

Blundell, Duncan and Pendakur (1998) used this approach in their estimation of nonparametric Engel curves. However, a model which is additive in $v$ requires that $v$ has the same effect at every point of estimation. This is cumbersome to implement in the local polynomial environment developed above.

Blundell and Powell (2001) also discuss the control function approach for nonadditive environments. In this environment, the stochastic restriction must be strengthened:

$$\epsilon^j|_{x, \chi} \sim \epsilon^j|_{x, v} \sim \epsilon^j|_v.$$  \hspace{1cm} (44)

Here, the stochastic restriction is that the structural error terms are not only conditionally mean-independent of the endogenous variable $x$, they are distributed conditionally identically
for all values of $x$. However, with this strengthened stochastic restriction, we can identify the structural model with a nonadditive local polynomial:

$$u^j_i = \alpha^j(p, x) + \eta^j(p, x)v_i + \epsilon^j_i$$  \hspace{1cm} (45)$$

To estimate the average value of the structural function at a given exogenous value of $(p, x)$, one proceeds by marginal integration over $v$:

$$\int (\alpha^j(p, x) + \eta^j(p, x)v) f(v)|_{p,x} \partial v = \alpha^j(p, x) + \eta^j(p, x) \int v f(v)|_{p,x} \partial v = \alpha^j(p, x) + \eta^j(p, x) E[v|_{p,x}].$$  \hspace{1cm} (46)$$

Thus, the average structural function is given by $\alpha^j(p, x)$ plus $\eta^j(p, x)$ multiplied by the local expectation of $v$.

One may implement this estimator easily with the following trick. Estimate a local polynomial given by

$$w^j_i = \alpha^j(p, x) + \eta(p, x)\tilde{v}_i + \epsilon^j_i$$

where

$$\tilde{v}_i = v_i - E[v|_{p,x}].$$

The normalisation $\tilde{v}_i$ makes the estimation of the average structural function easier because it loads the effect of $E[v|_{p,x}]$ on to the constant term $\alpha^j(p, x)$. Note that $\tilde{v}_i$ are different at every $p, x$ because $E[v|_{p,x}]$ is different at every $p, x$. Marginal integration over $\tilde{v}$ gives:

$$\int (\alpha^j(p, x) + \eta^j(p, x)\tilde{v}) f(\tilde{v})|_{p,x} \partial \tilde{v} = \alpha^j(p, x) + \eta^j(p, x) \int \tilde{v} f(\tilde{v})|_{p,x} \partial \tilde{v} = \alpha^j(p, x)$$

Since $E[\tilde{v}|_{p,x}] = 0$, the average structural function is given by $\alpha^j(p, x)$.

This argument extends naturally to the local polynomial case, since it depends only on the fact that $\tilde{v}_i$ comes in linearly. Dropping the explicit dependence of $\alpha^j$ and $\eta^j$ on $p, x$, we
estimate an augmented local quadratic model in the neighbourhood of \( p, x \) as follows:

\[
w_i^j = \alpha_j + \sum_{k=1}^{M} \beta^j_k (p^k - p^j_i) + \beta^j_2 (x - x_i) + \frac{1}{2} \sum_{k=1}^{M} \sum_{l=1}^{M} \gamma^j_{kl} (p^k - p^j_i) (p^l - p^j_i) + \sum_{k=1}^{M} \gamma^j_{kx} (p^k - p^j_i) (x - x_i) + \sum_{k=1}^{M} \gamma^j_{xx} (x - x_i)^2 + \eta^j v_i + \epsilon^j_i.
\] (47)

The estimated local levels, log-price derivatives and log-expenditure derivatives of the structural expenditure share equations are given, as before, by (12), (13) and (13). The logic for augmenting the LQAI model is exactly analogous.

In this paper, I use a set of instruments for log total-expenditure to nonparametrically estimate the control variable \( \tilde{v} \), and augment the LQ and LQAI models with linear effects in \( \tilde{v} \) to control for the endogeneity of log total-expenditure. The instruments used are log gross-income and log current-consumption (the excluded instruments) plus log-prices. These instruments explain about half the variation in log total-expenditure and the control variable \( \tilde{v} \) is usually statistically significant in the augmented local polynomials. All the estimation presented below controls for endogeneity, and estimation which does not control for endogeneity yields quantitatively similar and qualitatively identical results.

4. Exact Consumer Surplus Measurement

4.1 Small Price Changes

Consumer surplus measurement from estimated demand systems is appropriate only if the estimated Slutsky Matrix is symmetric. Exact consumer surplus for a small price change from \( p_0 \) to \( p_1 \) can be estimated by evaluation of a Taylor expansion of \( \ln C \) around \( p_0 \) to approximate \( \ln C(p_1, u) \). It is convenient to consider the cost-of-living index, which is a consumer surplus measure giving the ratio of expenditure needs between price situations \( p_0 \) and \( p_1 \). Consider a second order Taylor expansion of the log-cost function, \( \ln C(p, u) \), around
a price vector $p_0 = (p_0^1, \ldots, p_0^M)$. The log of the cost-of-living index is given by:

$$\ln \frac{C(p_1, u)}{C(p_0, u)} = \sum_{j=1}^{M} \frac{\partial \ln C(p_0, u)}{\partial p^j}(p^j_1 - p^j_0) + \frac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{M} \frac{\partial^2 \ln C(p_0, u)}{\partial p^j \partial p^k}(p^j_1 - p^j_0)(p^k_1 - p^k_0) + R,$$

where $R$ is the residual due to higher order terms.

Applying Sheppard’s Lemma, the derivatives and Hessian of the log-cost function with respect to log-prices are locally given by

$$\frac{\partial \ln C(p_0, u)}{\partial p^j} = w^j(p_0, x)$$

and

$$\frac{\partial^2 \ln C(p_0, u)}{\partial p^j \partial p^k} = \frac{\partial w^j w^k(p_0, x)}{\partial p^k} + \frac{\partial w^j w^k(p_0, x)}{\partial x} w^j w^k(p_0, x).$$

Substituting in LQ estimates at $(p_0, x)$, the second order Taylor approximation of the cost-of-living index for a small price change is given by

$$\ln \frac{C(p_1, u)}{C(p_0, u)} \approx \sum_{j=1}^{M} \alpha^j (p^j_1 - p^j_0) + \frac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{M} \left( \beta^j_k + \beta^j_k x \beta^k_j \right) (p^j_1 - p^j_0)(p^k_1 - p^k_0).$$

For the LQAII model, the approximate cost-of-living index is somewhat simpler and is given by

$$\ln \frac{C(p_1, u)}{C(p_0, u)} \approx \sum_{j=1}^{M} \alpha^j (p^j_1 - p^j_0) + \frac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{M} \beta^j_k (p^j_1 - p^j_0)(p^k_1 - p^k_0).$$

Previous work on the semi- and non-parametric estimation of consumer surplus (see, eg, Vartia 1983 and Hausman and Newey 1995) has not offered estimates that satisfy local symmetry or local path-independence. The use of symmetry-restricted parameter estimates in the above Taylor approximation fills this gap in the literature.
4.2 Large Price Changes

Following Vartia (1983), any large price change can be broken into a sequence of small price changes if the estimated Slutsky Matrix is everywhere continuous and symmetric. Symmetry guarantees that the price path taken will not affect the exact surplus estimate. If symmetry restricted estimates are continuous, then any sequence of (sufficiently) small price changes which sum to the large price change will "add up" to the same cost-of-living index for the large price change. In this paper, I use this method to estimate the cost-of-living index for the poverty line level of utility over the large relative and absolute price change from 1986 to 1999.

5. Demographic Effects

Define \( Z_i = Z_{i1}, ..., Z_{iT} \) as a \( T \times \) vector of demographic variables for the \( i \)th family, with \( Z_i = 0_T \) for the reference family type. Consider a demographically extended LQ model for which expenditure share equations are locally given by

\[
\begin{align*}
\hat{w}_i^j &= \alpha^j + \sum_{t=1}^{T} \alpha^{tj} Z_{it}^t + \sum_{k=1}^{M} \beta^j_k (p^k - p_i^k) + \beta^j_x (x - x_i) + \sum_{t=1}^{T} \sum_{k=1}^{M} \beta^{tj}_k Z_{it}^t (p^k - p_i^k) + \\
&\quad \sum_{t=1}^{T} \beta^j_x Z_{it}^t (x - x_i) + \frac{1}{2} \sum_{k=1}^{M} \sum_{l=1}^{M} \gamma^j_{kl} (p^k - p_i^k) (p^l - p_i^l) + \\
&\quad \sum_{k=1}^{M} \gamma^j_{xx} (p^k - p_i^k) (x - x_i) + \sum_{k=1}^{M} \gamma^j_{xx} (x - x_i)^2 + e_i^j.
\end{align*}
\]

(53)

Here, the demographic variables are interacted with the intercepts, log-prices and log-expenditure, but not with the quadratic terms. Again, the solutions are given by minimising

\[
\min_{\alpha^j, \alpha^{tj}, \beta^j_k, \beta^{tj}_k, \beta^j_x, \beta^{tj}_x, \gamma^j_{kl}, \gamma^j_{xx}} \sum_{i=1}^{N} \Omega_i (e_i^j)^2,
\]

which is just locally weighted least squares. Clearly, this is an econometrically easy way to include demographic effects. The question is how to constrain the parameters so that
The local estimates and estimated derivatives for the demographically extended LQ are given by

\[ \hat{w}^j = \alpha^j + \sum_{t=1}^{T} \alpha^{tj} Z_t^j, \quad (55) \]

\[ \frac{\partial \hat{w}^j}{\partial p^k} = \beta_j^k + \sum_{t=1}^{T} \beta^{ktj} Z_t^k, \quad (56) \]

and

\[ \frac{\partial \hat{w}^j}{\partial p^k} = \beta_j^x + \sum_{t=1}^{T} \beta^{xtj} Z_t^x. \quad (57) \]

Assuming that the homogeneity restrictions (19) hold for the reference family type, local homogeneity requires the additional restrictions

\[ \sum_{s=1}^{M} \beta^{jt}_s - \beta^{jt}_a = 0 \quad (58) \]

for all \( j, t \). These linear restrictions are easy to impose and test in the LQ model.

Restricting the demographically extended LQ for symmetry is more difficult. Assuming that the symmetry restrictions for the reference family type (29) hold, substituting the estimated levels and derivatives into the definition of a Slutsky term (26), and re-arranging, we get

\[ \sum_{t=1}^{T} Z_t^j \left[ \beta_{jt}^k + \beta_{xt}^j \alpha^k + \sum_{s=1}^{T} \beta_{zs}^j Z_s^z \alpha^s \right] = \sum_{t=1}^{T} Z_t^j \left[ \beta_{jt}^k + \beta_{xt}^j \alpha^k + \sum_{s=1}^{T} \beta_{zs}^j Z_s^z \alpha^{tj} + \beta_{xz}^j \alpha^j \right] \quad (59) \]

is symmetric. One might suspect that restricting \( \alpha^{tj} \) and \( \beta_{jt}^k \) in a way analogous to the restrictions (29) on \( \alpha^j \) and \( \beta_{jt}^k \) would give symmetry, but since \( Z \) shows up inside the square brackets, this does not work.

Consider imposing the restriction

\[ \beta_{jt}^k = 0 \quad (60) \]
which forces the log-expenditure effect to be invariant to demographics. Given this, \( Z \) drops out of the square brackets, so that \( \hat{S}_{jk} = \hat{S}_{kj} \) implies

\[
\sum_{t=1}^{T} Z_t^j \left( \beta_{tk}^j + \beta_{x}^{jk} \alpha_{tk}^j \right) = \sum_{t=1}^{T} Z_t^k \left( \beta_{tk}^k + \beta_{x}^{kt} \alpha_{tk}^k \right).
\]  

(61)

Since this must hold for all \( Z_t \), we get

\[
\beta_{tk}^j + \beta_{x}^{jk} \alpha_{tk}^j - \beta_{tk}^k - \beta_{x}^{kt} \alpha_{tk}^k = 0
\]  

or all \( j, k, t \).

Although symmetry may be achieved in the demographically extended LQ model via the restrictions (29), (60) and (62), these are quite cumbersome to impose in practise.

The incorporation of demographics satisfying homogeneity and symmetry is much easier in the local quadratic almost ideal (LQAI) case, because one can apply locally the strategies from global parametric models. Following, for example, Blundell and Robin (1999), one may create a demographically extended LQAI as follows:

\[
\sum_{t=1}^{T} Z_t^j \left( \beta_{tk}^j + \beta_{x}^{jk} \alpha_{tk}^j \right) (p^j - p_i^j) + \frac{1}{2} \sum_{k=1}^{M} \sum_{l=1}^{M} \gamma_{kl}^j (p^k - p_i^k) (p^k - p_i^k)
\]  

(63)

where

\[
\ln a(p - p_i) = \sum_{j=1}^{M} \left( \alpha^j + \sum_{t=1}^{T} \alpha^{jt} Z_t^j \right) (p^j - p_i^j) + \frac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{M} \left( \beta_{tk}^j + \sum_{t=1}^{T} \beta_{tk}^{jt} Z_t^j \right) (p^j - p_i^j) (p^k - p_i^k) + \frac{1}{6} \sum_{j=1}^{M} \sum_{k=1}^{M} \sum_{l=1}^{M} \gamma_{kl}^j (p^j - p_i^j) (p^k - p_i^k) (p^k - p_i^k). \]

(64)

This demand system may be estimated locally with the same iterated linearised methods as those discussed above.
Homogeneity and symmetry are easily imposed in this environment. Homogeneity requires the linear restrictions
\[ \sum_{k=1}^{M} \beta_{jk}^{tk} = 0 \]
for all \( t, j, k \). Symmetry is achieved in the demographically extended LQAI via the linear restrictions
\[ \beta_{jk}^{tk} = \beta_{tk}^{jk} \]
for all \( j, k, t \).

The simple imposition of symmetry for the demographically extended LQAI is a big advantage compared to the demographically extended LQ. In this paper, I present homogeneity- and symmetry-restricted demographically extended LQAI estimates.

6. Estimation

6.1 The Data


The grouping of individuals is typically constrained by the data, and in Canadian expenditure data, consumption data are for ‘spending units’ prior to 1992 and for households thereafter. Thus, I use consumption data for families, and restrict attention to families that were spending units alone in their household prior to 1992 and househoulds consisting of a single family after 1992.

Table 1 gives summary statistics for the 6952 observations of rental-tenure single-member
families used in the analysis without demographic effects, and for the 19276 families used in the analysis with demographic effects. The empirical analysis uses four expenditure categories yielding three independent expenditure share equations which depend on 4 log-prices, log-expenditure, and for some estimation, one demographic characteristic, the log of the number of family members.

<table>
<thead>
<tr>
<th>Table 1: The Data</th>
<th>Single Member Families</th>
<th>All Families</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expenditure Shares</td>
<td>Min</td>
<td>Max</td>
</tr>
<tr>
<td>Food at Home</td>
<td>0.02</td>
<td>0.84</td>
</tr>
<tr>
<td>Food Out</td>
<td>0.00</td>
<td>0.75</td>
</tr>
<tr>
<td>Rent</td>
<td>0.01</td>
<td>0.97</td>
</tr>
<tr>
<td>Clothing</td>
<td>0.00</td>
<td>0.61</td>
</tr>
<tr>
<td>Total Expenditure</td>
<td>640</td>
<td>40270</td>
</tr>
<tr>
<td>Log Total Expenditure</td>
<td>6.46</td>
<td>10.60</td>
</tr>
<tr>
<td>Number of Members</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(All Families)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log-Prices</td>
</tr>
<tr>
<td>Food at Home</td>
</tr>
<tr>
<td>Food Out</td>
</tr>
<tr>
<td>Rent</td>
</tr>
<tr>
<td>Clothing</td>
</tr>
</tbody>
</table>

6.2 Choice of Bandwidth Matrix

There is no consensus as to how to choose the bandwidth matrix $H$ in local polynomial regression estimation with multiple predictors. However, in the context of local polynomial density estimation with multiple predictors, some researchers have used a strategy called
‘sphering’. Here, the bandwidth matrix used is given by

$$H = h\Sigma^{1/2}$$ (65)

where $\Sigma$ is the estimated covariance matrix of $\tilde{p}_k$ and $h$ is a scaling parameter. For multivariate local polynomial regression, the optimal bandwidth, $h_{opt}$, is proportional to $N^{-1/(q+2r+2s)}$ where $q$ is the number of dimensions, $r$ is the order of the kernel, and $s$ is the order of the derivative of interest. In this paper, $q = 5$ (4 prices and expenditure), $r = 3$ (because the local quadratic is equivalent to a third order kernel for bias reduction, see Pagan and Ullah, p 106), and $s = 1$ (because the first derivatives are required to estimate the Slutsky Matrix).

Thus,

$$h_{opt} = kN^{-1/13}.$$ 

In this paper, I find $k$ by cross-validation, which yields $k = 2.77$. In addition, to assess the dependence of the estimates on the bandwidth, I use $k = 2$, which is chosen by ‘eyeballing’.

6.3 Estimated Expenditure Share Equations

Figures 1 to 3 show estimated expenditure share equations with bandwidth parameter $k$ selected by cross-validation for Food at Home, Food Out and Rent for the sample of 6952 rental-tenure single-member families. These estimates control for endogeneity as described above. The models are estimated locally at the 192 data points for observations in Ontario 1986, the price regime with the largest data subsample. The prices used are those facing consumers in Ontario 1986. These estimated regression curves are just the $a^j$ parameters for the local quadratic (LQ) and local quadratic almost ideal (LQAI) models at each of the 192 points in the price, expenditure space. The thin black line is the unrestricted LQ, the thick black line is the homogeneity-restricted LQ, the light gray line is the symmetry- and homogeneity-restricted LQ, and the dark gray line is the symmetry- and homogeneity-
restricted LQAI. I also provide global quadratic almost ideal (GQAI) estimates with a dotted line. The global estimates feature expenditure share levels that are quadratic in the log of total expenditure.

For each expenditure share, I include pointwise 90% confidence intervals for the homogeneity-restricted estimates (thick black lines) represented by the x’s in the figures. Confidence intervals are of similar size for other estimators, but the graph gets excessively cluttered if all confidence bands are drawn on. These pointwise intervals are estimated at 19 points for every 5th percentile from the 5th to 95th percentiles of the distribution of total expenditure in Ontario 1984. They are computed by simulating the distribution of the estimates under the null that they are actually equal to the estimated curves shown with thick lines. The simulation methodology uses the Smooth Conditional Moment System (SCMS) bootstrap technique described in the appendix, with 1000 bootstrap iterations and the confidence bands shown are 5th and 95th percentiles of the distribution of the simulated regression curves at each point in the domain. I note that the simulated marginal distributions of the $\alpha^j$ parameters are very close to normal. Thus, using +/- 1.605 times the standard error of $\alpha^j$ yields essentially identical confidence bands.

Looking first at Figure 1, the food-at-home expenditure share equation, it is clear that all the local estimates are similar to each other. The locally weighted estimation method and restrictions do not much affect the estimated levels of expenditure share equations. The unrestricted and homogeneity-restricted estimates are almost identical, which suggests that the imposition of local homogeneity is not very costly. Below, we will see that this claim is not supported by statistical testing.

The homogeneity- and symmetry-restricted LQ and LQAI estimates are essentially identical (visible differences will only emerge in Figures 4-9), which suggests that the exact
method used to impose symmetry doesn’t matter. Since the LQAI model is about 20 times easier to estimate\(^4\), I prefer the LQAI estimator. Further, the symmetry-restricted estimates are very close to unrestricted estimates, which suggests that the symmetry restrictions might not be binding. Below, we will see that this claim is not supported by statistical testing.

In Figure 1, below the median level of total expenditure, all the local estimates are quite similar to the global QAI estimator given by the dotted line. Above the median, the local estimates diverge a little bit, showing higher food-at-home shares.

Figure 2 presents the food-out equation, and here again the local estimates lie almost on top of each other. However, they are quite different from the global QAI estimate which is steeper than the local estimates. Figure 3 presents the rent share equation. In this equation, the locally weighted estimates are not as similar to each other as they are in Figures 1 and 2, nor is the global QAI estimate similar to the local estimates.

Considering the estimated levels of expenditure shares all together, it seems that the locally weighted LQ and LQAI estimators are quite similar to each other, which suggests that the particular local estimation method and restrictions used are not all that important in estimating the levels of expenditure shares. Further, the global quadratic almost ideal model captures some, but not all, of the the variation in expenditure shares over total expenditure. This latter finding is consistent with the literature on the nonparametric estimation of engel curves (see Banks, Blundell and Lewbel 1997; Blundell, Duncan and Pendakur 1998) which finds that engel curves over log total expenditure alone ‘look quadratic’.

\(^4\) I estimate a 3 equation system with 3 homogeneity restrictions and 3 symmetry restrictions on 6952 observations using TSP on a 1.3 GH Pentium 4 PC. The nonlinear LQ estimation takes about 60 seconds. The iterated linearised LQAI estimation takes about 3 seconds.
6.4 Estimated Own-Price Slutsky Terms

Figures 4 to 6 show estimated own-price Slutsky terms. If concavity is true, these must be everywhere negative. Ninety percent pointwise confidence intervals are provided for the homogeneity-restricted LQ estimates from the 5th to 95th percentiles of the total expenditure distribution. Again, GQAI estimates are presented with dotted lines. The GQAI model features Slutsky terms that are (globally) fourth order polynomials in the log of total expenditure. In Figures 4 and 6, the dark grey lines for the LQAI estimates look like they have ‘shadows’. The shadows are the light grey lines for the homogeneity- and symmetry-restricted LQ estimates. Thus in contrast to what we saw in for the estimation of levels, for the estimation of derivatives, slight differences between the LQ and LQAI estimators arise.

The estimated own-price Slutsky terms are almost everywhere negative for food-at-home and rent for all estimates over the whole total expenditure distribution. The estimated own-price Slutsky term for food-out is negative for all homogeneity-restricted estimates above the 5th percentile of expenditure. In addition, the estimated Slutsky matrix is negative semidefinite for both sets of homogeneity- and symmetry-restricted estimates for the 172 observations between the 5th and 95th percentiles of the total expenditure distribution. Since the concavity restriction is not binding in these data, I do not present concavity-restricted estimates.

Figure 4 gives the own-price Slutsky term for food-at-home. Here, the estimated 90% confidence intervals are much larger than those in the expenditure share equations. This is because derivatives are estimated less precisely than levels in nonparametric models. The two symmetry-unrestricted local estimates (black lines) are quite similar to each other, but are quite different from the two symmetry-restricted local estimates (gray lines). In particular, the symmetry-restricted LQ and LQAI estimates are less steeply upward sloping. The big
difference, though, is between the local estimates and the global estimate. The estimated own-price Slutsky term for food-at-home in the global model is negative, but is downward sloping rather than upward sloping.

The results shown for other own-price Slutsky terms are similar, though less dramatic. The estimated own-price Slutsky terms for food-out shown in Figure 5 are mostly positive for the unrestricted LQ, but almost everywhere negative for the other estimators. The homogeneity- and symmetry-restricted LQ and LQAI estimators are essentially identical to each other and downward sloping. The unrestricted LQ is something of an outlier—it is not everywhere negative, it is not close to any of the other estimates, and it becomes upward sloping above the median of expenditure. The GQAI estimates are downward sloping, but a little smaller in absolute value, and they do not capture the change in the slope above the 80th percentile of expenditure.

Figure 6 shows the own-price Slutsky terms for rent. Here, the local estimates are a more precisely estimated, and are comparatively close together, showing own-price Slutsky terms for rent that are negative, decreasing in absolute value, and whose slope changes around the 20th percentile of expenditure. The global estimates are significantly smaller in absolute value (they lie outside the confidence band), and do not show the change in slope.

There are three features of the own-price Slutsky terms that are common across Figures 4 to 6. First, with the exception of the unrestricted LQ for food-out, they are everywhere negative, which is consistent with concavity. That the unrestricted and homogeneity-restricted LQ models give different estimates suggests that the intuition of no-effect from Figures 1 to 3 is misleading. The homogeneity restriction is probably binding on estimates of derivatives, which is what should be expected for a restriction on derivatives. However, imposing homogeneity is sufficient to get negative own-price Slutsky terms with these Canadian expen-
diture data. Second, the symmetry-restricted LQ and LQAI models give essentially identical estimates of the own-price Slutsky terms, so that the two strategies offered to maintain symmetry give identical results. Since the LQAI is much easier to estimate, it may be preferred in practice. Third, the global quadratic almost ideal model does not give estimated own-price Slutsky terms that are similar to the locally weighted estimates. Thus, although the global parametric approach does tolerably well in estimating expenditure share equations, it does not do nearly as well in estimating price effects such as own-price Slutsky terms.

6.5 Estimated Cross-Price Slutsky Terms

Figures 7 to 9 give cross price Slutsky terms. To keep down the clutter, confidence bands are not provided, but they are similar in magnitude to those for the own-price Slutsky terms. Ninety percent confidence intervals are on the order of 0.3 wide between the 5th and 95th percentiles. Here, there are two thin black lines and two thick black lines corresponding to the LQ estimates which do not impose symmetry (unrestricted and homogeneity-restricted). Symmetry requires that the cross-price Slutsky terms be the same.

None of these cross-price Slutsky terms look like they satisfy symmetry. In Figure 7, we see that the food-at-home, food-out cross-price terms seem quite asymmetric for both the unrestricted (thin black lines) and homogeneity-restricted (thick black lines) estimates. In fact, for both types of estimates, one Slutsky term is largely positive while the other is negative. These features seem common across all three cross-price Slutsky terms, and the violation of symmetry seems most egregious for high-expenditure observations in the food-at-home, rent cross-price terms. As with the homogeneity restriction, the symmetry restriction seems to bind in the estimation of derivatives.

In Figures 7-9, the symmetry-restricted LQ and LQAI estimates are very close to each other, and lie in between the estimates that do not impose symmetry.
6.6 Sensitivity to Bandwidth

Figures 1 to 9 use the cross-validated value of 2.77 for the bandwidth parameter $k$. However, in the multidimensional covariate context, bandwidths which optimise cross-validation criteria converge only very slowly to the mean integrated squared error minimising bandwidth. Further, in this environment (and many others), nonparametric regression estimates are sensitive to the bandwidth choice. Figures 10 and 11 present homogeneity- and symmetry-restricted LQAI estimates of expenditure shares and selected own-price Slutsky terms using the $k = 2$, and the cross-validated value of 2.77. In this 5-dimension problem, a bandwidth parameter value of $k = 2.77$ gives approximately 10% of observations less than 10% of the maximum possible weight in the locally weighted estimation. In contrast, a bandwidth parameter value of $k = 2$ gives approximately 40% of observations less than 10% of the maximum possible weight in the locally weighted estimation.

Looking first at Figure 10 which give the estimated expenditure share equations, it is clear that the change in bandwidth does not much affect the estimated LQAI expenditure shares. Although not shown, this is also true for all the other local estimators.

Figure 11 gives estimated own-price Slutsky terms for food-out and rent using the two bandwidths. Here, bandwidth choice makes a pretty big difference. Whereas with the cross-validated bandwidth, the food-out Slutsky term is essentially monotonically downward sloping, with the smaller bandwidth, it looks more U-shaped, with the bottom of the U around the 65th percentile. The rent own-price Slutsky term is also quite sensitive to the bandwidth chosen. It is monotonically increasing given the cross-validated bandwidth, but essentially flat above the median expenditure level given the smaller bandwidth.

Since the object of interest for consumer surplus measurement is the Slutsky terms, the dependence on bandwidth is troubling.
6.7 Incorporating Family Size

Because it requires only iterated linear estimation, it is easy to incorporate demographic effects into the local quadratic almost ideal model. Figure 12 presents estimated expenditure shares for the homogeneity- and symmetry-restricted demographically extended LQAI model using 19276 observations of rental-tenure families evaluated at the same set of points used in Figures 1 to 11. I use a single demographic characteristic, the log of family size \( fsize \), so that \( T = 1 \), and \( Z^i_t = Z_t = \ln(fsize_i) \). Figure 12 shows expenditure share equations for the reference family type of single adults with a thick grey line, and for a family with two members with a thick dotted line. The downward sloping share equations are for food-at-home, the upward sloping share equations are for food-out, and the humped share equations are for rent.

The inclusion of demographic effects in this model is more general than the partially linear way that demographic effects are typically included in nonparametric estimation of Engel curves (eg, Blundell, Duncan and Pendakur 1998). With partially linear covariates, the effect of demographics is everywhere the same. In Figure 12, it is clear that family size affects expenditure patterns differently for rich and poor. In particular, rent shares for couples are about 8 percentage points lower than for singles at low expenditure, but only about 4 percentage points lower at high expenditure. This suggests that shelter might be 'less shareable' for rich than for poor families.

Figure 13 presents estimated own-price Slutsky terms for the same LQAI model. Demographics seem to have a fairly large effect on estimated own-price Slutsky terms, especially for food-at-home. These estimates are all restricted to satisfy homogeneity and symmetry. Further, the estimated Slutsky matrix is negative semidefinite for all family sizes less than 10 at all the points of estimation above the 5th percentile of total expenditure. The easy
inclusion of demographic effects in a homogeneity-and-symmetry satisfying way is a strong selling point for the LQAI model.

6.8 Consumer Surplus Estimates

Estimated demand systems which satisfy symmetry allow the path-independent estimation of consumer surplus, including cost-of-living indices. Table 2 presents estimated cost-of-living indices for the poverty level single-member family living in Ontario for the period 1986 to 1999. During this period, the prices of food-at-home, food-out, rent and clothing rose 27%, 54%, 44% and 39%, respectively. The poverty level is defined as the level of expenditure on the 4 included goods which corresponds to the poverty level of total expenditure for Ontario 1986 estimated in Pendakur (2001). This poverty level is equal to $5096, so that \( x = 8.54 \), which is approximately the 10th percentile of log-expenditure on the 4 included goods in Ontario 1986. Table 2 shows estimated cost-of-living indices using 4 steps to get from initial to final prices, and using 13 steps (one for each year) to get from initial to final prices. Two paths from initial to final prices are taken. The ‘straight’ path is comprised of equispaced points on a linear combination of initial and final log-prices. The ‘actual’ path is the set of annual log-prices for Ontario over 1986 to 1999. The estimated cost-of-living index using the global quadratic almost ideal model is 1.400.

<table>
<thead>
<tr>
<th>Path</th>
<th>Straight</th>
<th>Actual Prices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4 steps</td>
<td>13 steps</td>
</tr>
<tr>
<td>LQAI 2</td>
<td>1.400</td>
<td>1.400</td>
</tr>
<tr>
<td>LQAI 2.77</td>
<td>1.399</td>
<td>1.399</td>
</tr>
</tbody>
</table>

Four features emerge from Table 2. First, the estimated cost-of-living index is almost the
same whether 4 or 13 steps are taken. This is consistent with what Breslaw and Smith (1995) found using second order taylor approximations of the cost function. Second, the path of prices taken does not seem to matter, especially if 13 steps are taken. This is comforting, and consistent with Vartia (1983) who shows that given symmetry, this way of estimating consumer surplus satisfies path-independence. Third, the bandwidth chosen does not much affect the estimated cost-of-living index. Fourth, although the global quadratic almost ideal model does not capture all the variation over total expenditure of the Slutsky matrix, it does a very good job of estimating the cost-of-living index for the poverty line.

6.9 Testing the Homogeneity and Symmetry Restrictions

A key feature of Figures 4 to 9 is that the unrestricted LQ is often different from the other locally weighted estimates, all of which impose local homogeneity. The question is, are these differences statistically significant? Figures 7 to 9 show that the symmetry-unrestricted estimates do not seem like they satisfy symmetry, but as noted in the discussion of Figures 4-6, the confidence bands around estimated Slutsky terms are large. An interesting question is whether or not these differences are statistically significant.

I use the test statistics $\tau^\text{Hom}$ and $\tau^\text{Sym}$ given by equations (21) and (32). The parameter covariance matrices used in the test statistics $\tau^\text{Hom}$ and $\tau^\text{Sym}$ are simulated via the bootstrap as described in the appendix. The hypothesis of homogeneity is tested using estimates from the unrestricted LQ model, and the hypothesis of symmetry is tested using estimates from the homogeneity-restricted LQ model. In this case, with three independent expenditure shares, the test statistics are both distributed chi squared with 3 degrees of freedom. Figure 14 shows the value of these test statistics at every 5th percentile from the 5th to 95th percentile of the total expenditure distribution. The 1% critical value for this chi-squared test statistic with three degrees of freedom is 11.3, so homogeneity and symmetry are both
rejected above the 20th percentile of expenditure, but not below it.

7. Conclusions

This paper uses local polynomial modelling to estimate a consumer demand system, which allows easy estimation of the derivatives of expenditures shares with respect to prices and expenditure. This permits the easy estimation of the Slutsky matrix which is useful in the measurement of consumer surplus, and allows the researcher to restrict the expenditure share equations to locally satisfy homogeneity and the Slutsky matrix to locally satisfy symmetry. It also yields Wald-like tests of local homogeneity and local symmetry. Procedures are also developed to measure exact consumer surplus, incorporate demographic effects and simulate confidence bands.

Consumer demand systems are estimated with Canadian price and expenditure data for the period 1969 to 1999. Estimated expenditure share equations, Slutsky matrices, and consumer surplus estimates using homogeneity- and symmetry-restricted estimates are presented. Among the four local models used, the locally weighted quadratic almost ideal model is favoured because it is particularly easy to restrict for homogeneity and symmetry, offers easy incorporation of demographic effects, and is estimable via iterated linear methods. Local estimation is shown to give different estimates of expenditure shares and Slutsky terms compared to global parametric estimation.

REFERENCES


8. Simulating Confidence Bands and Test Statistics

8.1 Smooth Conditional Moment System Methods

Gozalo’s (1997) Smooth Conditional Moment (SCM) bootstrap allows the researcher to reproduce the conditional heteroskedasticity of the data in the simulation of confidence bands and test statistic sampling distributions for a single equation model. However, the SCM bootstrap is suitable for systems of equations only when there is no cross-equation correlation. Here, I develop a simple modification of the SCM bootstrap that allows the researcher to reproduce the conditional cross-equation correlation in the data.

The following algorithm generates a multivariate distribution of errors conditional on an independent variable vector \((p, x)\). The algorithm generates error vectors of length \(M - 1\) for families indexed \(i = 1...N\). (There are only \(M - 1\) error terms for each family because the last error term is given by the restriction that the errors sum to zero.) These vectors, \(\epsilon_i(p, x) = [\epsilon_i^1(p, x), ..., \epsilon_i^{M-1}(p, x)]'\), are constructed so that \(E[\epsilon_i(p, x)] = 0_{M-1}\), \(E[\epsilon_i(p, x)\epsilon_i(p, x)'] = \Sigma(p, x)\), and \(E[\epsilon_i^2(p, x)\epsilon_i^j(p, x)] = \mu_j(p, x)\). Gozalo’s (1997) SCM Bootstrap does...
not allow for cross-equation correlations, and so requires that \( \Sigma(p,x) \) be a diagonal matrix; here I relax that restriction. I call this method the Smooth Conditional Moment System (SCMS) bootstrap.

Since the data are not assumed to be identically distributed, these second and third moments depend on the independent variables \((p,x)\). Following Gozalo (1997), one may estimate \( \mu^j(p,x) \) and the elements of \( \Sigma(p,x) \) from kernel regressions. In particular, given \( M - 1 \) independent equations, one might construct the empirical error distribution, \( \hat{\epsilon}_i \), and estimate the \( \Sigma(p,x) \) matrix by nonparametric regression of \( (\hat{\epsilon}_i^j \hat{\epsilon}_i^k) \) over \((p,x)\), and estimate \( \mu^j(p,x) \) by nonparametric regression of \( (\hat{\epsilon}_i^j)^3 \) over \((p,x)\). The strategy is to construct a distribution of uncorrelated errors, \( \nu_i = [\nu_1^i, ..., \nu_{M-1}^i]' \), and to generate \( \varepsilon_i \) satisfying the above moment conditions as a linear combination of \( \nu_i \).

Suppress the \( i \) subscript on all functions, vectors and matrices, and consider the generation of a set of bootstrap errors for a single family. Since \( \Sigma(p,x) \) is a covariance matrix evaluated at each point on \((p,x)\), it is everywhere positive semidefinite. Define \( \sigma(p,x) \) such that \( \sigma(p,x)' \sigma(p,x) = \Sigma(p,x) \) (for example, using the Choleski Decomposition of \( \Sigma(p,x) \) at each point in \((p,x)\)). Write the elements of \( \sigma(p,x) \) as \( \sigma^{jk}(p,x) \). Define \( \Gamma(p,x) \) as a matrix containing the cubed elements of \( \sigma(p,x) \), so that \( \Gamma(p,x) = \{ (\sigma^{jk}(p,x))^3 \} \). Define a vector \( \mu(p,x) = [\mu^1(p,x), ..., \mu^{M-1}(p,x)]' \). Define a vector \( \lambda(p,x) = \Gamma^{-1}(p,x) \mu(p,x) \), and denote its elements as \( \lambda^j(p,x) \). All of these matrices and vectors are conditioned on \((p,x)\).

Define an independent \( M - 1 \) vector random variables \( \nu = [\nu^1, ..., \nu^{M-1}]' \). The following set of two-point distributions for \( \nu^j \) satisfies \( \mathbb{E}[\nu_j] = 0_{M-1} \), \( \mathbb{E}[\nu \nu'] = 1_{M-1} \), \( \mathbb{E}[(\nu^j)^3] = \lambda^j \) (see Gozalo 1997 for details):

\[
\widetilde{F}^j(p,x) = \pi^j(p,x) \theta_{\nu^j(p,x)} + (1 - \pi^j(p,x)) \theta_{\nu^{-j}(p,x)}
\]  

(66)
where \( \theta_K \) denotes a probability measure that puts mass one at \( K \), \( \pi^j(p, x) \) in \([0, 1]\),

\[
\pi^j(p, x) = \left[ 1 + \lambda^j(p, x) / T^j(p, x) \right] / 2,
\]

(67)

\[
a^j(p, x) = \left( \lambda^j(p, x) - T^j(p, x) \right) / 2,
\]

(68)

\[
b^j(p, x) = \left( \lambda^j(p, x) + T^j(p, x) \right) / 2
\]

(69)

and

\[
T^j(p, x) = \sqrt{\lambda^j(p, x)^2 + 4}.
\]

(70)

Generate \( \varepsilon \) as follows:

\[
\varepsilon = \sigma' \nu
\]

(71)

\( \varepsilon \) constructed in this manner satisfies \( E[\varepsilon(p, x)] = 0 \), \( E[\varepsilon(p, x)\varepsilon(p, x)'] = \Sigma(p, x) \), \( E[\varepsilon_i(p, x)^3] = \mu_i(p, x) \).

One can also obtain a continuous rather than discrete distribution by multiplying \( \nu^j \) by a normal random variate \( \rho^j \) with mean \( \eta = (\sqrt{3} - 1)/2 \) and variance \( 1 - \eta^2 \) (see Gozalo 1997). In this paper, I use this continuous distribution.

Proof: Suppress the dependence of all moments and functions on \( (p, x) \). Since \( E[\nu] = 0 \) and \( \varepsilon \) is linear in \( \nu \), \( E[\varepsilon] = E[\sigma' \nu] = 0_K \). Noting that \( E[\nu \nu'] = I_K \), \( E[\varepsilon \varepsilon'] = E[\sigma' \nu \nu' \sigma] = \sigma' \sigma = \Sigma \).

Noting that \( \varepsilon = S \nu \) and \( E[\nu \nu'] = I_K \), and that all of the cross-products in the third moments of \( \nu_i \) have expectation zero, \( E[\varepsilon^3] = \Sigma_{j=1}^{M-1} \sigma^3 \varepsilon^3_{j\lambda} \). Denoting \( \varepsilon^3 = [\varepsilon^3_1, ..., \varepsilon^3_K]' \), this condition may be expressed in matrix form as \( E[\varepsilon^3] = \Gamma \lambda \). Substituting in for \( \lambda \) gives \( E[\varepsilon^3] = \Gamma \Gamma^{-1} \mu = \mu \).

QED.

8.2 Simulation of Confidence Bands and Parameter Covariance Matrices

Confidence bands for nonparametric local polynomial estimated expenditure share equations may be simulated as follows. Define a set of points of estimation \( (p, x) \) as \( \pi_\delta \), \( \delta = 1, ..., \Delta \).
These may or may not be the same as the observations of data. Estimate the model at every point in \( \pi \), denoting the estimates as \( \hat{w}_i^j \).

**Step 1:** Estimate the model at every point in the data for observations \( w_i \) and \( p_i \), \( i = 1, \ldots, N \). Denote estimated expenditure shares (given by the LQ parameters \( \alpha^j \)) as \( \hat{w}_i^j \). Construct empirical errors as \( \hat{\epsilon}_i^j = w_i^j - \hat{w}_i^j \). Estimate \( \Sigma(p, x) \) and \( \mu^j(p, x) \) by LQ estimation.

**Step 2:** Construct an \( M - 1 \) bootstrap error vector \( \varepsilon_i \) via SCMS bootstrap for every observation. Construct the bootstrap expenditure shares as \( \overline{w}_i^j = \hat{w}_i^j + \varepsilon_i^j \). Estimate the model at each point of estimation in \( \pi \). Denote the bootstrap expenditure shares (given by the local quadratic parameters \( \alpha^j \)) as \( \overline{w}_i^j \).

**Step 3:** Repeat Step 2 \( B \) times, and use the distribution of \( \overline{w}_i^j \) to estimate the confidence band for the estimated expenditure shares \( \hat{w}_i^j \). For example, the 90% confidence band for \( \hat{w}_i^j \) may be given by the 5'th and 95'th percentiles of the distribution of \( \overline{w}_i^j \). In this paper, \( B = 1000 \).

**Step 4:** To estimate the covariance matrix of the estimated parameters at each point in \( \pi \), use the covariance matrix of the \( B \) bootstrap estimates of the parameters. This covariance matrix can then be used to compute the test statistics \( \tau^{Hom} \) and \( \tau^{Sym} \) at each point of estimation.
Figure 1: Expenditure Shares: Food At Home
Figure 2: Expenditure Shares: Food Out

![Expenditure Shares: Food Out](image-url)
Figure 2: Expenditure Shares: Food Out
Figure 3: Expenditure Shares: Rent
Figure 4: Own-Price Slutsky Terms: Food At Home

[Graph showing various lines and markers representing different series and indicating the own-price Slutsky terms for food at home.]
Figure 5: Own-Price Slutsky Terms: Food Out
Figure 6: Own-Price Slutsky Terms: Rent
Figure 8: Cross-Price Slutsky Terms: Food At Home, Rent
Figure 9: Cross-Price Slutsky Terms: Food Out, Rent
Figure 10: LQAI Expenditure Shares: Sensitivity to Bandwidth
Figure 11: LQAI Own-Price Terms: Sensitivity to Bandwidth
Figure 12: Expenditure Shares with Demographic Effects
Figure 13: Own-Price Slutsky Terms with Demographic Effects
Figure 14: Tests of Homogeneity and Symmetry

The graph shows the test statistic values for both homogeneity and symmetry tests plotted against log total expenditure. The data points are indicated with different symbols: circles for homogeneity test statistic and crosses for symmetry test statistic.