Semiparametric Indirect Utility and Consumer Demand

Krishna Pendakur[a], Michael Scholz[b], Stefan Sperlich[*,b]

[a] *Department of Economics, Simon Fraser University, Burnaby, Canada*
[b] *Institut für Statistik und Ökonometrie, Georg-August Universität, Göttingen, Germany*

Abstract

A semiparametric model of consumer demand is considered. In the model, the indirect utility function is specified as a partially linear, where utility is nonparametric in expenditure and parametric (with fixed- or varying-coefficients) in prices. Because the starting point is a model of indirect utility, rationality restrictions like homogeneity and Slutsky symmetry are easily imposed. The resulting model for expenditure shares (as functions of expenditures and prices) is locally given by a fraction whose numerator is partially linear, but whose denominator is nonconstant and given by the derivative of the numerator. The basic insight is that given a local polynomial model for the numerator, the denominator is given by a lower-order local polynomial. The model can thus be estimated using modified versions of local polynomial modeling techniques. For inference, a new asymmetric version of the wild bootstrap is introduced. Monte Carlo evidence that the proposed techniques work is provided as well as an implementation of the model on Canadian consumer expenditure and price micro-data.

Key words: Consumer Demand, Engel Curves, Semiparametric Econometrics, Wild Bootstrap With Asymmetric Errors

[*] Corresponding author: Platz der Göttinger Sieben 5, 37073 Göttingen, Tel.: +49 551 39 7350
Email address: stefan.sperlich@wiwi.uni-goettingen.de (Stefan Sperlich)

1. Introduction

The specification and estimation of consumer demand systems, defined as the relationship between quantity demands, prices and total expenditures, presents many long-standing problems in econometric theory. Recent work has focused on the inclusion of highly nonlinear relationships between quantity demands (or expenditure shares) and total expenditures into empirical models of consumer demand. Since typical consumer demand microdata have a large amount of variation in total expenditures across consumers, complex relationships between demands and expenditure might potentially be identified. Unfortunately, because consumer demand models must satisfy a set of nonlinear cross-equation rationality restrictions (see, e.g., Deaton and Muellbauer (1980), and Varian (1978)), known as the Slutsky symmetry restrictions, such complex relationships have been hard to incorporate into semi- and non-parametric approaches.

This paper presents a semiparametric approach to the consumer demand problem which allows for the imposition of the Slutsky symmetry restrictions. We use a flexible nonparametric estimation method in the total expenditure direction – where the data provide a lot of information – to get demands which are arbitrarily flexible in total expenditure (ie., arbitrarily flexible Engel curves). However, in the price directions – where the data are less rich – we propose a parametric structure.

Like most models for consumer demand, our model uses the vector of expenditure shares commanded by each good as the dependent variable. In this paper, we introduce a wild bootstrap that accounts for the fact that expenditure shares lie in [0,1]. The idea is to draw bootstrap residuals from a local adaptive distribution that respects the boundaries via asymmetry.

Nonparametric approaches to consumer demand started by considering Engel curves, defined as the relationship between expenditure shares and the total expenditures of the consumer, at a fixed vector of prices. That is, they considered only 1 nonparametric direction and held the others fixed. Work by Blundell et al. (1998, 2003) revealed considerable complexity in the shapes of Engel curves. A fully nonparametric approach which considers both price and expenditure directions together and which allows for the imposition of rationality restrictions, has been developed by Haag et al. (2009). Here, the shapes of the demand equations are not restricted, but curse of dimensionality rears its head: with M price directions and 1 expenditure direction, the researcher faces an $M+1$ dimensional problem. Even if homogeneity, another rationality condition, is imposed, the researcher still faces an M dimensional problem, which is still very high in typical applications.

Parametric approaches like the popular Almost Ideal (Deaton and Muellbauer, 1980), dynamic Almost Ideal (Mazzocchi, 2006), Translog (Jorgensen et al., 1980) and Quadratic Almost Ideal (Banks et al., 1997) demand models typically impose strict limits on the functional complexity of Engel curves. In these cases, they must be linear, nearly linear, or quadratic, respectively, in the log of total expenditure. This lack of complexity is driven by the need for these parametric models to satisfy the Slutsky symmetry restrictions.

A major use of consumer demand systems is in policy analysis: demand systems are used to assess whether or not indirect tax changes are desirable, and are used to assess changes in the cost-of-living. In this regard, lack of complexity has costs: in particular, if the Engel curve is wrong, then all consumer surplus calculations (including cost-of-living calculations) are also wrong. For example, Banks et al. (1997) and Lewbel and Pendakur (2009) show that the false imposition of linear and quadratic Engel curves, respectively, can lead to very misleading estimates of behavioural and welfare responses to indirect tax changes.

In between the fully nonparametric and the fully parametric approaches, we have the realm of semiparametric econometrics. Two recent papers have explored this area. Lewbel and Pendakur (2009) propose a fully parametric approach which satisfies rationality restrictions and for which Engel curves can be arbitrarily complex. Because their model allows for arbitrarily complex Engel curves but parametrically restricted dependence of expenditure shares on prices, it may be interpreted as semiparametric. However, their approach relies critically on a particular interpretation of the 'error term' in the regression: it must represent unobserved preference heterogeneity, and thus cannot be measurement error or any other deviations from optimal choice on the consumer's part. Further, Lewbel and Pendakur (2009) do not allow for a varying-coefficients structure for price effects.

Pendakur and Sperlich (2010) propose a semiparametric model which allows for these latter interpretations of the role of the error term, does not restrict the shape of Engel curves, and incorporates price effects either parametrically or semiparametrically (through fixed- or varying-coefficients, cf. Sarmiento (2005)). Pendakur and Sperlich (2010) propose a model in which expenditure-shares are nonparametric in utility – an unobserved regressor – and (semi-)parametric in log-prices. The familiarity of this partially linear form makes the model appealing, but the unobserved regressor (utility) must be constructed under the model via numerical inversion of the (unknown) cost function. In the present paper, we propose a model in which utility is nonparametric in log-expenditure and parametric in log-prices. This results in a model of expenditure-shares which is locally nonlinear but has no unobserved or generated regressors. All of these semiparametric approaches address the curse of dimensionality: they each have just 1 nonparametric dimension.

The local nonlinearity of our model of expenditure-shares is driven by the fact that we start by modeling indirect utility as partially linear, and since Roy's Identity (Roy (1947)) gives expenditure shares as the ratio of derivatives of indirect utility, expenditure shares in our model are given by a ratio. This ratio has nonparametric functions in the numerator and their derivatives in the denominator. Our basic insight is if one models the numerator as a local polynomial, the denominator – which is comprised of derivatives of the numerator – is just a lower-order local polynomial. This fact suggests a natural iterative procedure to estimate the model. Our algorithm is computationally efficient and numerically robust such that large data sets can be handled in acceptable time, and the results are readily interpreted.

In Section 2 we introduce the model. In Section 3 we discuss the basic estimation idea, give the associated algorithm and describe the bootstrap inference. For the nonparametric part of the model we use an univariate local linear smoother on transformed data, a method that can be easily applied in empirical research. To accommodate the parametric part of the model, we use a restricted (to satisfy the Slutsky symmetry restrictions) least squares estimator. For inference, we designed an asymmetric version of the wild bootstrap. To fulfill the constraints that the (bootstrap) responses must be in $[0, 1]$, we propose a local adaptive χ^2-distribution for the bootstrap errors. A nice feature of our approach is that confidence intervals created this way are narrower than those based on standard wild bootstrap.

In Section 4 we evaluate our proposed methods and the accuracy of the bootstrap procedure in a small simulation study, and then we implement the model with Canadian price and expenditure data. Empirically, we find that some expenditure share equations show quite a lot of nonlinearity. Section 5 concludes and discusses extensions.

2. A Semiparametric Model for Indirect Utility

Define the indirect utility function $V(\mathbf{p}, x)$ to give the maximum utility attained by a consumer when faced with a vector of log–prices $\mathbf{p} = (p^1, \ldots, p^M)$ and log–total expenditure x. Let the expenditure share of a good be defined as the expenditure on that good divided by the total expenditure available to the consumer. Denote $\mathbf{w} = (w^1, \ldots, w^M)$ as the vector of expenditure share functions and note that since expenditure shares sum to 1, $w^M = 1 - \sum_{j=1}^{M-1} w^j$. Let $\{W_i^1, \ldots, W_i^M, P_i^1, \ldots, P_i^M, X_i\}_{i=1}^N$ be a random vector giving the expenditure shares, log–prices and log–total expenditure of a population of N individuals. Note that, as commonly done in the literature of demand systems, we use the superscript notation for single elements of vectors or matrices, i.e. for single goods or commodities, and the subscript for individuals.

2.1. A Partial Linear and Varying Coefficient Model for Indirect Utility

We consider two semiparametric specifications of the indirect utility function. First, we consider a partially linear (or, fixed-coefficient) specification of the form

$$V(\mathbf{p}, x) = x - \sum_{k=1}^{M} f^k(x)p^k - \frac{1}{2}\sum_{k=1}^{M}\sum_{l=1}^{M} a^{kl}p^k p^l, \tag{1}$$

or, in matrix notation,

$$V(\mathbf{p}, x) = x - \mathbf{f}(x)'\mathbf{p} - \frac{1}{2}\mathbf{p}'\mathbf{A}\mathbf{p}, \tag{2}$$

where $\mathbf{f} = (f^1, \ldots, f^M)'$ are unknown differentiable functions of log–total expenditure and $\mathbf{A} = \{a^{kl}\}_{k,l=1}^M$ are parameters. We impose the normalisation that $a^{kl} = a^{lk}$, or, equivalently, $\mathbf{A} = \mathbf{A}'$. This is not a restriction: since $p^k p^l = p^l p^k$, there is a symmetric version of $\mathbf{A}$ that yields the same V as any asymmetric version. Second, we consider the varying-coefficient extension of this model:

$$V(\mathbf{p}, x) = x - \sum_{k=1}^{M} f^k(x)p^k - \frac{1}{2}\sum_{k=1}^{M}\sum_{l=1}^{M} a^{kl}(x)p^k p^l, \tag{3}$$

or, in matrix notation,

$$V(\mathbf{p}, x) = x - \mathbf{f}(x)'\mathbf{p} - \frac{1}{2}\mathbf{p}'\mathbf{A}(x)\mathbf{p}, \tag{4}$$

where $a^{kl}(x) = a^{lk}(x)$ for all k, l, or, equivalently, $\mathbf{A}(x) = \mathbf{A}(x)'$.

Expenditure shares are functions of total expenditure and all prices. Roy's Identity relates the expenditure share for good j, $w^j(\mathbf{p}, x)$, to derivatives of the indirect utility function: $w^j(\mathbf{p}, x) = -\left[\partial V(\mathbf{p}, x)/\partial p^j\right] / \left[\partial V(\mathbf{p}, x)/\partial x\right]$. Application of Roy's Identity to the fixed-coefficients model yields

$$w^j(\mathbf{p}, x) = \frac{f^j(x) + \sum_{k=1}^M a^{jk}p^k}{1 - \sum_{k=1}^M \nabla_x f^k(x)p^k},$$

with ∇_x indicating the derivative (here of $f^k(x)$) with respect to x; or, in matrix notation,

$$\mathbf{w}(\mathbf{p}, x) = \frac{\mathbf{f}(x) + \mathbf{A}\mathbf{p}}{1 - \nabla_x \mathbf{f}(x)'\,\mathbf{p}}.$$

For the varying-coefficients model we get

$$w^j(\mathbf{p},x) = \frac{f^j(x) + \sum_{k=1}^M a^{jk}(x)p^k}{1 - \sum_{k=1}^M \nabla_x f^k(x)p^k - \frac{1}{2}\sum_{k=1}^M \sum_{l=1}^M \nabla_x a^{kl}(x)p^k p^l},$$

or, in matrix notation,

$$\mathbf{w}(\mathbf{p},x) = \frac{\mathbf{f}(x) + \ \mathbf{A}(x)\mathbf{p}}{1 - \nabla_x \mathbf{f}(x)'\mathbf{p} + \frac{1}{2}\mathbf{p}'\nabla_x \mathbf{A}(x)\mathbf{p}}.$$

We describe how to estimate these expenditure share equations in Section 3.

The motivation for these models is as follows. In real-world applications, there is typically a large amount of observed variation in total expenditures, so one may reasonably hope to identify a nonparametric component in that direction. However, typical micro-data sources do not have nearly as much variation in the price directions, which suggests that partially linear modelling might describe these effects sufficiently well. If in addition, the researcher feels that more may be identified on the strength of observed price variation, the varying-coefficients model allows price effects in the model (3) to be different at different expenditure levels. This would seem to be a pure advantage of the varying coefficients approach. However, in practise, this extension seriously increases the variance and computational cost of the estimates. In particular, the algorithm for model (3) is about five times slower than the one for model (1). The important feature here is that nonparametric dimensionality is kept to 1 in both models.

2.1.1. Rationality Restrictions

Rationality is comprised of three conditions: homogeneity, symmetry and concavity. Here we will deal only with symmetry and homogeneity (concavity is a topic of its own, investigated, e.g. in Millimet and Tchernis (2008)). Slutsky symmetry relates to the fact that expenditure share equations are derived in terms of the derivatives of indirect utility, V. Slutsky symmetry (see, e.g., Mas-Colell et al. (1995)) gives minimal restrictions under which expenditure share equations lead to a unique indirect utility function. In our context, it is satisfied if and only if $\mathbf{A} = \mathbf{A}'$ in the expenditure share equations (or, in the varying-coefficients case, if $\mathbf{A}(x) = \mathbf{A}(x)'$). In the indirect utility function, restricting these matrices to symmetry is merely a normalisation. However, in the expenditure share equations, this restriction has bite. In particular, because each expenditure share equation could be estimated separately, the estimated matrix could be asymmetric. In our estimation section below, we use an algorithm which maintains symmetry, and which is the semiparametric analog to a linearly restricted Seemingly Unrelated Regression (SUR) estimator.

Homogeneity is sometimes referred to as 'no money illusion'. If consumers do not suffer from money illusion, then scaling prices and expenditures by the same factor cannot affect utility. This requires that indirect utility is homogeneous of degree zero in (unlogged) prices and expenditure. This can be achieved by dividing all prices and expenditure by the price of the M-th expenditure category. Note that we use logarithms, so we subtract p^M from each log-price and from log-expenditure in the indirect utility function. For the fixed-coefficients case, this yields

$$V(\mathbf{p},x) = (x-p^M) - \sum_{k=1}^{M-1} f^k(x-p^M)\cdot(p^k-p^M) - \frac{1}{2}\sum_{k=1}^{M-1}\sum_{l=1}^{M-1} a^{kl}(p^k-p^M)(p^l-p^M),$$

in model (1) and analogously in model (3). The sums go only to $M-1$ because the M-th element of each sum (which multiplies $p^M - p^M$) is zero. Denoting $\tilde{x} = x - p^M$, $\tilde{p}^j = p^j - p^M$ and $\tilde{\mathbf{p}} = (\tilde{p}^1, \ldots, \tilde{p}^{M-1})$ we may write this more compactly as

$$V(\tilde{\mathbf{p}}, \tilde{x}) = \tilde{x} - \sum_{k=1}^{M-1} f^k(\tilde{x}) \cdot \tilde{p}^k - \frac{1}{2} \sum_{k=1}^{M-1} \sum_{l=1}^{M-1} a^{kl} \tilde{p}^k \tilde{p}^l, \tag{5}$$

with a^{kl} depending on $\tilde{x}$ in the varying-coefficients case. We thus estimate only the first $(M-1)$ elements of $\mathbf{f}$ and $\mathbf{w}$, and the first $(M-1)$ rows and columns of $\mathbf{A}$. In matrix notation, this may be written as $\mathbf{f} = (f^1, \ldots, f^{M-1})'$ and $\mathbf{A} = \{a^{kl}\}_{k,l=1}^{M-1}$ as

$$V(\tilde{\mathbf{p}}, \tilde{x}) = \tilde{x} - \mathbf{f}(\tilde{x})'\tilde{\mathbf{p}} - \tilde{\mathbf{p}}'\mathbf{A}\tilde{\mathbf{p}},$$

for the fixed-coefficient case and

$$V(\tilde{\mathbf{p}}, \tilde{x}) = \tilde{x} - \mathbf{f}(\tilde{x})'\tilde{\mathbf{p}} - \tilde{\mathbf{p}}'\mathbf{A}(\tilde{x})\tilde{\mathbf{p}},$$

for the varying coefficient case. Once again, since expenditures sum to 1 by construction, we have $w^M(\tilde{\mathbf{p}}, \tilde{x}) = 1 - \sum_{i=1}^{M-1} w^i(\tilde{\mathbf{p}}, \tilde{x})$, and we need only consider the first $(M-1)$ expenditure share equations.

As before, we get the expenditure share equations

$$\mathbf{w}(\tilde{\mathbf{p}}, \tilde{x}) = \frac{\mathbf{f}(\tilde{x}) + \mathbf{A}\tilde{\mathbf{p}}}{1 - \nabla_{\tilde{x}}\mathbf{f}(\tilde{x})'\,\tilde{\mathbf{p}}}, \tag{6}$$

for the fixed-coefficients model (1), and

$$\mathbf{w}(\tilde{\mathbf{p}}, \tilde{x}) = \frac{\mathbf{f}(\tilde{x}) + \mathbf{A}(\tilde{x})\tilde{\mathbf{p}}}{1 - \nabla_{\tilde{x}}\mathbf{f}(\tilde{x})'\tilde{\mathbf{p}} - \frac{1}{2}\tilde{\mathbf{p}}'\nabla_{\tilde{x}}\mathbf{A}(\tilde{x})\tilde{\mathbf{p}}}, \tag{7}$$

for the varying-coefficients model (3). Here, $\nabla_{\tilde{x}}\mathbf{f}(\tilde{x})$ is the $(M-1)$ vector of the derivatives of $\mathbf{f}(\tilde{x})$ with respect to $\tilde{x}$, and $\nabla_{\tilde{x}}\mathbf{A}(\tilde{x})$ is the $(M-1) \times (M-1)$ matrix function equal to the derivatives of $\mathbf{A}$ with respect to $\tilde{x}$.

These expressions for budget shares have a nice feature in comparison to Pendakur and Sperlich (2010). Whereas their model for expenditure shares uses a nonparametric function of a generated regressor which must be constructed under the model using numerical inversion of the unknown cost function, the expression above uses only observed regressors. However, in comparison to Pendakur and Sperlich (2010), which is a partially linear model, the above expression is partially linear only in the numerator. The presence of the denominator seems to complicate things. However, as we show below, with the use of local polynomials this problem becomes manageable.

3. Estimation of the Models

In the following sections, we show how to estimate the $(M-1)$ vector $\mathbf{w}(\tilde{\mathbf{p}}, \tilde{x})$ under the model. Such estimates satisfy "adding-up" by construction, since $w^M(\tilde{\mathbf{p}}, \tilde{x}) = 1 - \sum_{i=1}^{M-1} w^i(\tilde{\mathbf{p}}, \tilde{x})$. Also they satisfy homogeneity ("no money illusion") by construction, due to the use of normalised prices and expenditures as regressors. Finally, they can satisfy Slutsky symmetry, because the matrix $\mathbf{A}$ (or $\mathbf{A}(\tilde{x})$) can easily be restricted to be a symmetric matrix, e.g. with Deschamps (1988).

A more difficult problem is to restrict the estimated budget shares to be everywhere in the range $[0, 1]$. This problem is referred to as the "global regularity" problem in the literature on consumer demand. Roughly speaking, demand systems that are not homothetic – that is, those which have budget shares which respond to total expenditure – cannot typically be globally regular without restricting either the domain of $\mathbf{p}, x$ or the domain of model error terms in ad hoc ways. See Pollack and Wales (1991) for a discussion of the former, and Lewbel and Pendakur (2009) for discussion of the latter. We will judge our estimates in terms of "local regularity", that is, in terms of whether or not estimated budget shares are in the range $[0, 1]$ in a $\mathbf{p}, x$ domain of interest. In particular, under homogeneity and when $\mathbf{p} = \mathbf{0}_M$, in both the fixed-coefficients and varying-coefficients model, we have

$$\mathbf{w}(\mathbf{p}, x) = \mathbf{w}(\tilde{\mathbf{p}}, \tilde{x}) = \mathbf{f}(\tilde{x}) = \mathbf{f}(x).$$

Thus, the estimated functions $\mathbf{f}(x)$ characterise budget shares over a domain spanned by x with log-prices fixed at $\mathbf{0}_M$. If these estimated functions lie within $[0, 1]$, then we say that our estimates are "locally regular" in this sense. Note also that the vast majority of the literature on estimating expenditure systems does not tackle this problem due to its complexity (an exception is Moral-Arce et al. (2007)).

3.1. Basic Ideas

The basic idea of estimating the unknown nonparametric functions f^j and the (potentially varying) coefficients a^{jk}, $j, k = 1, \ldots, M-1$, consists of iteratively solving minimization problems, where the iteration is necessary only for the nonparametric part of the model. We use kernel smoothing for the nonparametric part, and, in case of the fixed-coefficients model (1), least squares for the parametric coefficients. Obtaining estimates consistent with Slutsky symmetry is via the use of (linearly) restricted least squares for the parametric part.

Keeping the dependence on $\tilde{x}$, we may approximate

$$\mathbf{f}(t) \approx \mathbf{f}(\tilde{x}) + \nabla_{\tilde{x}}\mathbf{f}(\tilde{x})(t - \tilde{x}) \approx \boldsymbol{\alpha}(\tilde{x}) + \boldsymbol{\beta}(\tilde{x})(t - \tilde{x}), \tag{8}$$

where $\boldsymbol{\alpha}(\tilde{x})$ and $\boldsymbol{\beta}(\tilde{x})$ are the local level and derivative of $\mathbf{f}(t)$. Then, for the partial linear model the local problem is

$$\min_{\boldsymbol{\alpha}(\tilde{x}), \boldsymbol{\beta}(\tilde{x}), \mathbf{A}} \sum_{i=1}^{N} \mathbf{e}_i' \boldsymbol{\Omega} \mathbf{e}_i, \text{ with}$$
$$\mathbf{e}_i \equiv \mathbf{w}_i - \frac{\boldsymbol{\alpha}(\tilde{x}) + (\tilde{x}_i - \tilde{x})\,\boldsymbol{\beta}(\tilde{x}) + \mathbf{A}\tilde{\mathbf{p}}_i}{1 - \boldsymbol{\beta}(\tilde{x})'\tilde{\mathbf{p}}_i},$$

where $\boldsymbol{\Omega}$ is an $(M-1) \times (M-1)$ weighting matrix.

Similarly, for the varying coefficient model (7), the local problem in the neighbourhood of each given $\tilde{x}$ is

$$\min_{\boldsymbol{\alpha}(\tilde{x}), \boldsymbol{\beta}(\tilde{x}), \boldsymbol{\Gamma}(\tilde{x}), \boldsymbol{\Delta}(\tilde{x})} \sum_{i=1}^{N} \mathbf{e}_i' \boldsymbol{\Omega} \mathbf{e}_i, \text{ with}$$
$$\mathbf{e}_i \equiv \mathbf{w}_i - \frac{\boldsymbol{\alpha}(\tilde{x}) + (\tilde{x}_i - \tilde{x})\,\boldsymbol{\beta}(\tilde{x}) + \boldsymbol{\Gamma}(\tilde{x})\tilde{\mathbf{p}}_i + (\tilde{x}_i - \tilde{x})\,\boldsymbol{\Delta}(\tilde{x})\tilde{\mathbf{p}}_i}{1 - \boldsymbol{\beta}(\tilde{x})'\tilde{\mathbf{p}}_i - \frac{1}{2}\tilde{\mathbf{p}}_i'\boldsymbol{\Delta}(\tilde{x})\tilde{\mathbf{p}}_i},$$

where $\boldsymbol{\Omega}$ is now a different $(M-1) \times (M-1)$ weighting matrix and $\boldsymbol{\Gamma}(\tilde{x})$ and $\boldsymbol{\Delta}(\tilde{x})$ are the local level and derivative, respectively, of the price coefficients.

Here, the imposition of homogeneity is via the use of normalised prices and expenditures (i.e. $\tilde{x}$ instead of x etc.). The imposition of Slutsky symmetry is via the restriction that $\mathbf{A} = \mathbf{A}'$, or in the varying-coefficients case, that $\mathbf{A}(x) = \mathbf{A}(x)'$ which is achieved by restricting $\mathbf{\Gamma}(\tilde{x}) = \mathbf{\Gamma}(\tilde{x})'$ and $\mathbf{\Delta}(\tilde{x}) = \mathbf{\Delta}(\tilde{x})'$. This local linear approach could easily be extended to higher order local polynomials, but for this we would need stronger assumptions on the data and model.

3.2. The Estimation Algorithm

Denote $\Delta_i = \tilde{X}_i - \tilde{x}$, $K_i = K((\tilde{X}_i - \tilde{x})/h)/h$, where K is some symmetric kernel function with the usual properties and h a bandwidth that controls the smoothness of the estimate. We omit an extra subscript h in K_i for the sake of notation.

Let us start with the minimization problem for the partial linear model (1). As above, the α^j are related to the functions f^j at point $\tilde{x}$ and the parameters β^j to its first derivatives, while the parameters a^{jk} are fixed for all $\tilde{x}$:

$$\min_{\alpha^j,\beta^j} \sum_{j=1}^{M-1} \sum_{i=1}^{N} \left(W_i^j - \frac{\alpha^j + \Delta_i \beta^j + \sum_{k=1}^{M-1} a^{jk} \tilde{P}_i^k}{1 - \sum_{k=1}^{M-1} \beta^k \tilde{P}_i^k} \right)^2 K_i. \tag{9}$$

In order to minimize, we set the first derivative equal to zero. Taking the derivative of (9) with respect to α^j, and using the notations $S_i = 1 - \sum_{k=1}^{M-1} \beta^k \tilde{P}_i^k$ and $T_i^j = \sum_{k=1}^{M-1} a^{jk} \tilde{P}_i^k$, we solve

$$0 = \sum_{i=1}^{N} \left(W_i^j - \frac{\alpha^j + \Delta_i \beta^j + T_i^j}{S_i} \right) \frac{K_i}{S_i}. \tag{10}$$

This gives immediately (for $j = 1, \ldots, M-1$)

$$\alpha^j = \left[\sum_{i=1}^{N} W_i^j K_i / S_i - \beta^j \sum_{i=1}^{N} K_i \Delta_i / S_i^2 - \sum_{i=1}^{N} K_i T_i^j / S_i^2 \right] \left[\sum_{i=1}^{N} K_i / S_i^2 \right]^{-1}. \tag{11}$$

On the other hand, by differentiating (9) with respect to β^j (again for $j = 1, \ldots, M-1$), we get the equations

$$\begin{aligned} 0 \;=\; & \sum_{i=1}^{N} \left(W_i^1 - \frac{\alpha^1 + \Delta_i \beta^1 + T_i^1}{S_i} \right) K_i \cdot \frac{(\alpha^1 + \Delta_i \beta^1 + T_i^1) \tilde{P}_i^j}{S_i^2} + \cdots + \\ & \sum_{i=1}^{N} \left(W_i^j - \frac{\alpha^j + \Delta_i \beta^j + T_i^j}{S_i} \right) K_i \cdot \frac{\Delta_i S_i + (\alpha^j + \Delta_i \beta^j + T_i^j) \tilde{P}_i^j}{S_i^2} + \cdots + \\ & \sum_{i=1}^{N} \left(W_i^{M-1} - \frac{\alpha^{M-1} + \Delta_i \beta^{M-1} + T_i^{M-1}}{S_i} \right) K_i \frac{(\alpha^{M-1} + \Delta_i \beta^{M-1} + T_i^{M-1}) \tilde{P}_i^j}{S_i^2}. \end{aligned}$$

This is equivalent to

$$\begin{aligned} 0 \;=\; & \sum_{k=1}^{M-1} \sum_{i=1}^{N} \left(W_i^k - \frac{\alpha^k + \Delta_i \beta^k + T_i^k}{S_i} \right) K_i \cdot \frac{(\alpha^k + \Delta_i \beta^k + T_i^k) \tilde{P}_i^j}{S_i^2} + \\ & \sum_{i=1}^{N} \left(W_i^j - \frac{\alpha^j + \Delta_i \beta^j + T_i^j}{S_i} \right) K_i \frac{\Delta_i}{S_i}. \end{aligned} \tag{12}$$

Certainly, we can not solve equation (12) analytically for β^j. But, for our iterative purpose it is enough to consider the following implicit representation:

$$\beta^j = \Big[\sum_{k=1}^{M-1}\sum_{i=1}^{N}\Big(W_i^k - \frac{\alpha^k + \Delta_i\beta^k + T_i^k}{S_i}\Big) K_i \cdot \frac{(\alpha^k + \Delta_i\beta^k + T_i^k)\tilde{P}_i^j}{S_i^2} + \sum_{i=1}^{N}\Big(W_i^j - \frac{\alpha^j + T_i^j}{S_i}\Big) K_i \frac{\Delta_i}{S_i}\Big] \Big/ \sum_{i=1}^{N}\frac{K_i\Delta_i^2}{S_i^2}. \tag{13}$$

We use the implicit representation (13) to calculate new values for β^j. With them we get new S_i, so that we can find new α^j:

$$\beta_{new}^j = \Big[\sum_{k=1}^{M-1}\sum_{i=1}^{N}\Big(W_i^k - \frac{\alpha_{old}^k + \Delta_i\beta_{old}^k + T_{i,old}^k}{S_{i,old}}\Big) K_i \frac{(\alpha_{old}^k + \Delta_i\beta_{old}^k + T_{i,old}^k)\tilde{P}_i^j}{S_{i,old}^2} + \sum_{i=1}^{N}\Big(W_i^j - \frac{\alpha_{old}^j + T_{i,old}^j}{S_{i,old}}\Big) K_i \frac{\Delta_i}{S_{i,old}}\Big] \Big/ \sum_{i=1}^{N}\frac{K_i\Delta_i^2}{S_{i,old}^2}, \tag{14}$$

$$S_{i,new} = 1 - \sum_{k=1}^{M-1}\beta_{new}^k\tilde{P}_i^k, \text{ and} \tag{15}$$

$$\alpha_{new}^j = \frac{\sum_{i=1}^{N} W_i^j K_i/S_{i,new} - \beta_{new}^j\sum_{i=1}^{N}K_i\Delta_i/S_{i,new}^2 - \sum_{i=1}^{N}K_i T_{i,old}^j/S_{i,new}^2}{\sum_{i=1}^{N}K_i/S_{i,new}^2}.$$

We repeat these steps until convergence. The optimal $\mathbf{A}$ will be the one that minimizes the least squares problem. In practice, at the end of each iteration step, we solve the restricted least squares problem resulting from equation (6). With some algebra, the problem is given by

$$W_i^j \cdot (1 - \sum_{k=1}^{M-1}\beta_i^k\tilde{P}_i^k) - \alpha_i^j = \sum_{k=1}^{M-1} a^{jk}\tilde{P}_i^k. \tag{16}$$

The modification of the algorithm to take the varying coefficients $\mathbf{A}(\tilde{x})$ into account is one along ideas of Fan and Zhang (1999), though it is more complex in our context. With the same local linear approximation arguments as above, we get the local problem in the neighbourhood of $\tilde{x}$ as

$$\min_{\boldsymbol{\theta}} \sum_{j=1}^{M-1}\sum_{i=1}^{N}\left(W_i^j - \frac{\alpha^j + \Delta_i\beta^j + \sum_{k=1}^{M-1}(\gamma^{jk} + \Delta_i\delta^{jk})\tilde{P}_i^k}{1 - \sum_{k=1}^{M-1}\beta^k\tilde{P}_i^k - \frac{1}{2}\sum_{k=1}^{M-1}\sum_{l=1}^{M-1}\delta^{kl}\tilde{P}_i^k\tilde{P}_i^l}\right)^2 K_i, \tag{17}$$

with $\boldsymbol{\theta}$ denoting $\alpha^j, \beta^j, \gamma^{jk}$ and δ^{jk}. Note that γ^{jk} and δ^{jk} are symmetric since we consider a symmetric matrix of functions $a^{kl}(\tilde{x})$. The minimization of (17) in the usual way gives the extended algorithm in analogy to the first step of 3.2. For α^j and β^j we proceed as before but with $S_i = 1 - \sum\beta^k\tilde{P}_i^k - 1/2\sum\sum\delta^{kl}\tilde{P}_i^k\tilde{P}_i^l$ and $T_i^j = \sum(\gamma^{jk} + \Delta_i\delta^{jk})\tilde{P}_i^k$. Furthermore, we obtain

$$\gamma^{st} = \frac{\sum_{i=1}^{N}\left[\left(W_i^s - \frac{C_i^s}{S_i}\right)\tilde{P}_i^t + \left(W_i^t - \frac{C_i^t}{S_i}\right)\tilde{P}_i^s\,\mathbb{1}_{s\neq t}\right]\frac{K_i}{S_i}}{\sum_{i=1}^{N}\left[(\tilde{P}_i^t)^2 + (\tilde{P}_i^s)^2\mathbb{1}_{s\neq t}\right]\frac{K_i}{S_i^2}},$$

with $C_i^s = \alpha^s + \Delta_i \beta^s + T_i^s - \gamma^{st} \tilde{P}_i^t$ and, defining $T_i^{s,-t} = T_i^s - \Delta_i \delta^{st} \tilde{P}_i^t$,

$$\delta^{st} = \left[\sum_{k=1}^{M-1} \sum_{i=1}^{N} \left(W_i^k - \frac{\alpha^k + \Delta_i \beta^k + T_i^k}{S_i} \right) \frac{K_i}{S_i^2} (\alpha^k + \Delta_i \beta^k + T_i^k) \tilde{P}_i^t \tilde{P}_i^s + \right.$$
$$\left. \sum_{i=1}^{N} \left\{ \left(W_i^s - \frac{\alpha^s + \Delta_i \beta^s + T_i^{s,-t}}{S_i} \right) \tilde{P}_i^t + \left(W_i^t - \frac{\alpha^t + \Delta_i \beta^t + T_i^{t,-s}}{S_i} \right) \tilde{P}_i^s \right\} \frac{K_i \Delta_i}{S_i} \right]$$
$$\times \left[\sum_{i=1}^{N} \left\{ (\tilde{P}_i^t)^2 + (\tilde{P}_i^s)^2 \mathbb{1}_{s \neq t} \right\} \frac{\Delta_i^2 K_i}{S_i^2} \right]^{-1} .$$

3.3. Bootstrap Inference

The "wild bootstrap" draws bootstrap responses based on the estimated model (1) with given sample $\{W_i, \tilde{X}_i, \tilde{P}_i\}_{i=1}^N$ and estimates $\hat{\alpha}^j$, $\hat{\beta}^j$ and $\hat{a}^{jk}$, $k, j = 1, \ldots, M-1$. Denote a prior bandwidth g with $O(g) > O(h)$ (obeying the needs of asymptotic theory, cf. Härdle and Marron (1991)), and let h be the bandwidth giving the desired smoothness in the original sample. The basic idea is now to use the estimated residuals from an estimate with bandwidth g,

$$\hat{\varepsilon}_i^j = W_i^j - \frac{\hat{\alpha}^j(\tilde{X}_i) + \sum_{k=1}^{M-1} \hat{a}^{jk} \tilde{P}_i^k}{1 - \sum_{k=1}^{M-1} \hat{\beta}^k(\tilde{X}_i) \tilde{P}_i^k}, \tag{18}$$

to get wild bootstrap residuals ε_i^{j*}. Given them we create bootstrap samples $\{W_i^*, \tilde{X}_i, \tilde{P}_i\}_{i=1}^N$ by

$$W_i^{j*} = \frac{\hat{\alpha}^j(\tilde{X}_i) + \sum_{k=1}^{M-1} \hat{a}^{jk} \tilde{P}_i^k}{1 - \sum_{k=1}^{M-1} \hat{\beta}^k(\tilde{X}_i) \tilde{P}_i^k} + \varepsilon_i^{j*}, \tag{19}$$

for $i = 1, \ldots, N$ and $j = 1, \ldots, M-1$. Here, ε_i^{j*} are bootstrap residuals that replicate desired properties of the distribution(s) of $\hat{\varepsilon}_i^j$. The W_i^{M*} are generated via $\sum_{j=1}^M W_i^{j*} = 1$. Repeating this many times, we get estimates (for $\mathbf{f}$ and $\mathbf{A}$) for each bootstrap sample and can use the bootstrap quantiles to construct point wise confidence bands for our estimates.

There exists several strategies to obtain bootstrap residuals ε_i^{j*}. Typically, when no restriction is faced, one may use $\varepsilon_i^{j*} = u_i \cdot \hat{\varepsilon}_i^j$, where u_i is a standard normal random scalar. Under the additional assumption of homoscedasticity, this can even be simplified to $\varepsilon_i^{j*} = u_i \cdot \hat{\sigma}_\varepsilon^j$, where σ_ε^j is estimated from the residuals (18).

In our case, one could argue that such bootstrap errors could cause the bootstrap values of W_i^{j*} to lie outside the admissible range of $[0, 1]$ for budget shares. On the one hand, this may not matter because the estimation algorithm does not control the constraint that $\hat{W}_i^j \in [0, 1]$. However, given that actual expenditure shares are bounded, the bootstrap residuals may poorly reflect the true error distribution and misrepresent the confidence intervals, for example putting them outside $[0, 1]$.

To address the possibility that inference is hampered by bootstrap budget-shares lying outside $[0, 1]$, we introduce an alternative formulation of the wild bootstrap. Because there are many expenditure shares, the main bounding problem is the lower bound at 0, and this is the problem we deal with. Thus, we are faced with a conditionally asymmetric (to the right) error distribution. We thus consider an asymmetric distribution for ε_i^{j*} given $\hat{\varepsilon}_i^j$ as follows. Generate bootstrap errors via

$$\frac{\chi_k^2}{\sqrt{k}} \cdot \frac{|\hat{\varepsilon}_i^j|}{\sqrt{2}} - \frac{|\hat{\varepsilon}_i^j|}{\sqrt{2}} \cdot \sqrt{k} \leq |\hat{W}_i^j|, \tag{20}$$

where $k \leq \lfloor (W_i^j/\hat{\varepsilon}_i^j)^2 \cdot 2 \rfloor$. In the case that k is less than one, we draw the bootstrap residual ε_i^{j*} from $\chi_1^2 \cdot |\hat{W}_i^j| - |\hat{W}_i^j|$. Note that this fulfills $E[\varepsilon_i^{j*}] = 0$ and $E[(\varepsilon_i^{j*})^2] = E[(\hat{\varepsilon}_i^j)^2]$ for all i and j. From (20), we have that, for positive $\hat{W}_i^j$, its bootstrap analog is always positive, too. This method leads automatically to confidence bands that lie almost fully inside $[0, 1]$ and are consequently narrower than those based on a simple normal bootstrap described above. In the simulation study below, we present empirical evidence that the above introduced asymmetric bootstrap is accurate.

3.4. Practical Considerations

One issue in such iterative procedures is the question of adequate initial values for the nonparametric part. Here we have a convenient model feature to exploit: when we normalise prices in the sample such that $\tilde{\mathbf{p}}_i = (0, \dots, 0)$ for some group of consumers, say N_0, equation (9) reduces to the well-known local linear case. That is, since for the denominator term we have $S_i = 1$, we get the objective function

$$\min_{\alpha^j, \beta^j} \sum_{j=1}^{M-1} \sum_{i=1}^{N_0} \left(W_i^j - \alpha^j + \Delta_i \beta^j \right)^2 K_i. \tag{21}$$

Solving this problem on the sample of consumers where $\tilde{\mathbf{p}}_i = (0, \dots, 0)$ gives us consistent estimates (though with a possibly large variance depending on the subsample size N_0) which be used as starting values for α^j and β^j. For the varying coefficient model we also need starting values for the γ^{jk} and δ^{jk}. As a natural choice for the γ^{jk} we use the results of the algorithm in Section 3.2 and zero for all δ^{jk} (i.e. starting in the first iteration with a simpler model).

For the bandwidth choice, we recommend using the same bandwidth h for all expenditure categories because: (a) the functions refer to the same expenditure data in all equations; and (b) the economic theory does not suggest that the shares of some goods would be smoother than those of others. Plug-in bandwidths could be derived from the asymptotics of $\widehat{\mathbf{f}}$, or, alternatively, one could construct a risk estimate similar to cross-validation but this time jointly for all elements of $\widehat{\mathbf{f}}$. It is clear, however, that the first depends on derivatives of the unknown $\widehat{\mathbf{f}}$ and the density of expenditures, while the second approach would be quite computationally costly. A rough idea of a bandwidth size to start with may be derived from (21). Run a leave-one-out cross validation for (21) with the subsample of individuals fulfilling $\tilde{\mathbf{p}}_i = (0, \dots, 0)$, and correct the obtained bandwidth h_0 for the size of the full sample, i.e. $h = N^{-1/5} h_0 N_0^{1/5}$.

In the fixed-coefficients model (1), we recommend running the estimation algorithm twice: first with an undersmoothing bandwidth in the nonparametric part to keep the possible smoothing bias small. The resulting estimate for the coefficient matrix $\mathbf{A}$ is kept, and used in the second run which uses a larger bandwidth for the nonparametric part to get reasonably smooth $\widehat{\mathbf{f}}$. This is unnecessary in the varying coefficients model (3) where we face only nonparametric functions.

Recall that the M^{th} equation and its Engel curve is simply a result of the homogeneity and the adding-up condition $\sum_{j=1}^{M} W_i^j = 1$. In practice one might choose the item with the least variation in expenditure shares across the households.

4. Empirical Analysis

4.1. A Simulation Study

First, to generate some artificial data, we generated 33 distinct price vectors, normally distributed in each dimension, for each of 6 expenditure categories (i.e. we have 6 items with different prices in 33 regions). As in typically observed micro-data, we did not allow for a wide price variety, see Lewbel (2000). Summary values for these price vectors can be found in Table 1.

Table 1: Summary of used price vectors in simulation

	1	2	3	4	5
Min	3.905	3.449	3.763	0.880	2.794
Max	4.130	3.585	3.919	1.121	3.010
Mean	4.018	3.517	3.841	1.002	2.901
Std.	0.030	0.020	0.020	0.030	0.030

For 32 regions (i.e. price vectors) we uniformly draw 30 log–total expenditure values from the interval $[1, 2]$. For the reference region (number 33) we draw 40 uniformly distributed values between one and two. In total, this gives us $N = 1000$ observations. These are used to generate expenditure shares using the expenditure functions shown in Figure 1 (solid lines), price parameters given in Table 2, and normal error terms with mean zero and standard deviation 0.01. In order to get shares which fulfill the conditions $W^j \in [0, 1]$ and $\sum_j W^j = 1$ we applied the rejection method (that is, we dropped and replaced values outside $[0, 1]$).

Next, we estimate the functions α^j and the price parameters using our estimation algorithm introduced in Section 3.2. This is repeated 250 times (using the same functions, price parameters, and range of log–total expenditure values) to get an idea of the mean squared errors of our estimators. For estimating $\mathbf{f}$ we used the Gaussian kernel with $h \approx 0.034$, the smallest bandwidth giving smooth estimates.

In Figure 1 we have plotted the true functions (solid lines) together with intervals of 90% coverage probabilities for the estimates (dotted lines) as a result of the 250 simulation runs. On the one hand, we see pretty narrow bands which accurately capture even those functions with flat plateaus in the intermediate range (category 3) and with bumps (category 2). Such functions are often hard to estimate in practise. We also see the limits of the method as for example boundary effects. Our smoother can estimate without any bias the linear function. In Table 3 we give the estimated parameter means, together with the standard deviations. The exactness of our simulation is demonstrated by the small total MSE of only $6.83 \cdot 10^{-6}$.

Table 2: Price parameters used in the simulation

	1	2	3	4	5
1	−0.150	−0.100	0.150	0.100	0.280
2		0.250	0.100	−0.250	0.170
3			0.320	−0.220	−0.190
4				−0.200	0.150
5					−0.180

Table 3: Estimated price parameters and standard deviations (in brackets)

	1	2	3	4	5
1	−0.1515	−0.1006	0.1494	0.0997	0.2799
	(0.0140)	(0.0106)	(0.0103)	(0.0090)	(0.0087)
2		0.2494	0.0999	−0.2497	0.1699
		(0.0176)	(0.0129)	(0.0102)	(0.0102)
3			0.3200	−0.2208	−0.1891
			(0.0170)	(0.0096)	(0.0100)
4				−0.1992	0.1490
				(0.0117)	(0.0079)
5					−0.1803
					(0.0117)

To verify the functioning of our new bootstrap procedure we constructed 200 bootstrap samples (with $g = h$ as N is relatively small) along Section 3.3 for each of 100 simulation runs. We calculated 90% bootstrap confidence intervals around the estimator for each simulation. The mean of these upper and lower bounds of these intervals are given in Figure 1 (dashed lines). The fact that the 90% coverage probability intervals and the means of the 90% bootstrap intervals almost coincide indicates that our bootstrap procedure is acceptably accurate.

4.2. Analysing Household Expenditures in Canada

In our empirical study we use the same Canadian data as in Lewbel and Pendakur (2009) and Pendakur and Sperlich (2010) which come from public sources, see also Pendakur (2002). The price and expenditure data are available for 12 years in 5 regions: Atlantic, Quebec, Ontario, Prairies and British Columbia. This yields 60 distinct price vectors, where prices are normalised in a way that all prices of the categories from Ontario in 1986 are one, i.e. $\tilde{\mathbf{p}}_{O,86} = (0, \ldots, 0)$, so these 189 observations define the base price vector and we use them to get the starting values. Note further, to achieve homogeneity we subtracted p^M, the price of the left–out expenditure category, from all other prices and total expenditure.

We use 6952 observations of rental–tenure unattached individuals aged between 25 and 64 with no dependants to minimise demographic variation in preferences. Our analysis includes annual total–expenditure in nine categories: food–in, food–out, rent, clothing, household operations, household furnishing and equipment, private transportation, public transportation and personal care. The left–out category is personal care, so that we get a system of eight expenditure share equations which depend on eight (normalised) log–prices and (normalised) log–total expenditure. These expenditure categories account for about three-quarters of the current consumption of the households in the sample. Summary statistics of the observations are given in Table 4.

We note that this choice of commodities is arbitrary: one could divide these goods into subcategories, or aggregate them up into larger categories. We choose these categories because they offer the finest gradation consistent with largest possible time span for the price data (finer gradations of price data are available, but for shorter periods of time). Another advantage of this choice of commodities is that they are directly comparable with Pendakur and Sperlich (2010).

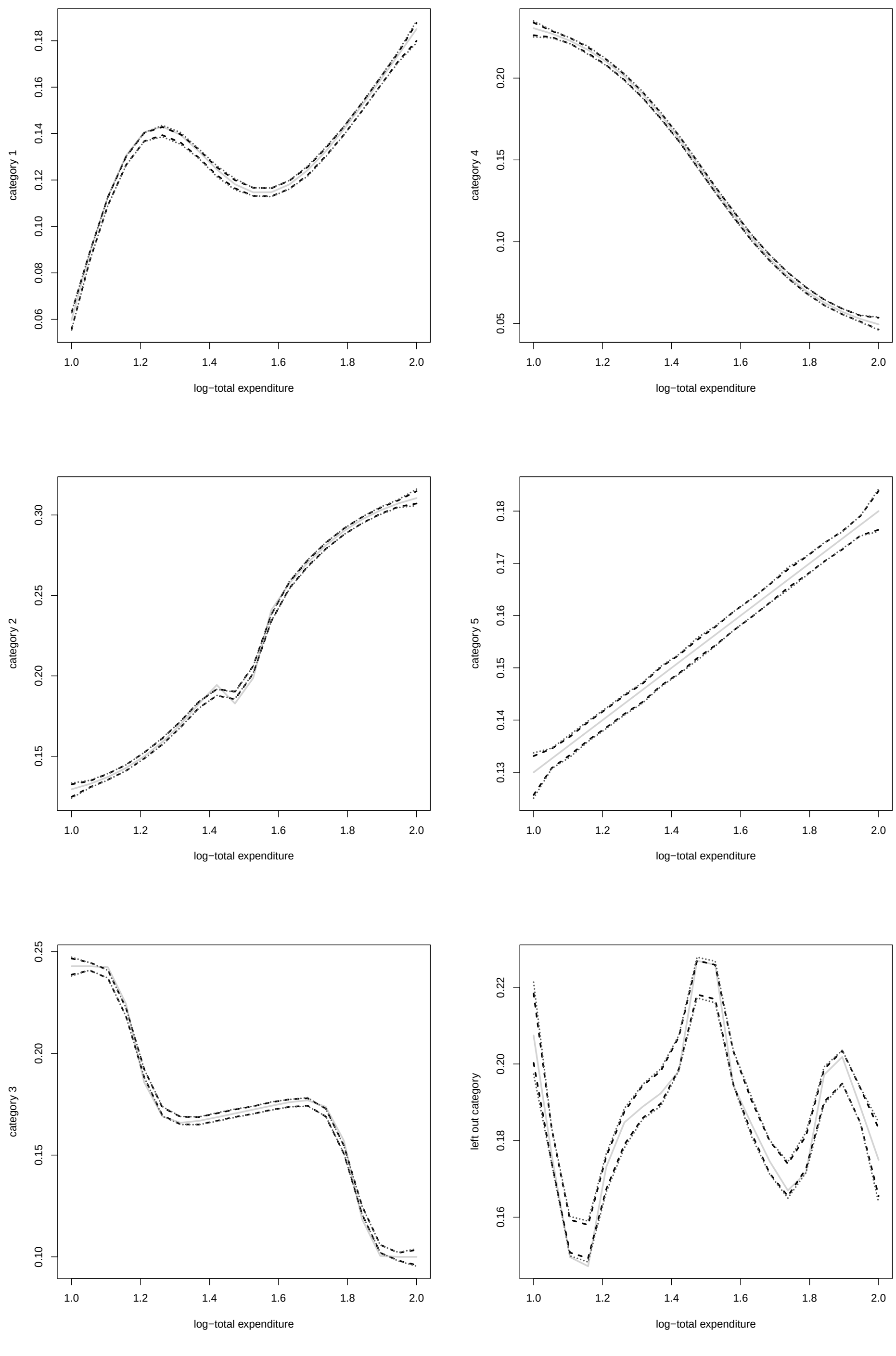

Figure 1: Simulation of 6 different budget share functions (solid line) with 90% coverage probability (dotted lines) and asymptotic 90% bootstrap confidence intervals (dashed lines)

Table 4: The Data

		Min	Max	Mean	Std.
expenditure shares	food–in	0.00	0.63	0.17	0.09
	food–out	0.00	0.64	0.08	0.08
	rent	0.01	0.95	0.40	0.13
	operations	0.00	0.64	0.08	0.04
	furnishing	0.00	0.65	0.04	0.06
	clothing	0.00	0.53	0.09	0.06
	private trans	0.00	0.59	0.08	0.09
	public trans	0.00	0.34	0.04	0.04
log–prices	food–in	−0.39	0.07	−0.03	0.09
	food–out	−0.42	0.25	0.05	0.12
	rent	−0.35	0.14	−0.12	0.15
	operations	−0.28	0.10	−0.04	0.08
	furnishing	−0.16	0.21	−0.03	0.09
	clothing	−0.07	0.44	0.10	0.11
	private trans	−0.51	0.30	−0.09	0.18
	public trans	−0.59	0.40	0.01	0.25
log–total expenditure		3.03	6.26	4.61	0.45

As noted above, when $\tilde{\mathbf{p}} = (0, \ldots, 0)$ (as it does for observations in Ontario 1986), the price effects in expenditure shares amount to zero, yielding

$$\mathbf{w}(\mathbf{p}, x) = \mathbf{w}(\tilde{\mathbf{p}}, \tilde{x}) = \mathbf{f}(\tilde{x}) = \mathbf{f}(x),$$

which we will refer to as the vector of Engel curves. The estimated Engel curves of all expenditure categories can be found in Figure 2 and 3 as solid lines, where the horizontal axes refer to $\tilde{x}$, i.e. the log total expenditures minus p^M.

We include pointwise 90% confidence intervals which we calculated as described in Section 3.3 with heteroscedastic error terms and 500 bootstrap iterations using our new conditionally asymmetric wild bootstrap procedure. To generate the bootstrap samples we used $g = h$ with $h = 0.17$ being somewhat larger than $h_0(N_0/N)^{1/5}$ which would give wiggly estimates $\widehat{\mathbf{f}}$. This was estimated using the Gaussian kernel, and it converged in our setting after about 15 iterations. In all figures, the resulting Engel curves are compared to the ones of Banks et al. (1997) and the ones of Pendakur and Sperlich (2010) when assuming a partial linear cost function with Slutsky symmetry. In terms of local regularity, the estimated values of budget-shares lie entirely within $[0, 1]$. Although we do not assess the global regularity of our estimates or estimator, it is comforting that estimated budget shares satisfy this condition locally.

Food-at-home and food-out are strong necessities and luxuries, respectively, with nearly linear Engel curves in both cases. The near-linearity of these Engel curves has been observed in a large number of empirical investigations, including Banks et al. (1997). Some curvature is observed in the rent and clothing equations, especially near the bottom of the distribution. This curvature is noted in semiparametric work, such as Pendakur and Sperlich (2010) and Lewbel and Pendakur (2009). The most curvature is noted in smaller budget shares like household operation, private transportation and public transportation. The curvature in household operation seems quite strong, and that in private transportation seems decidedly non-quadratic. In the figures one can see that most of the expenditure-share equations are very similar

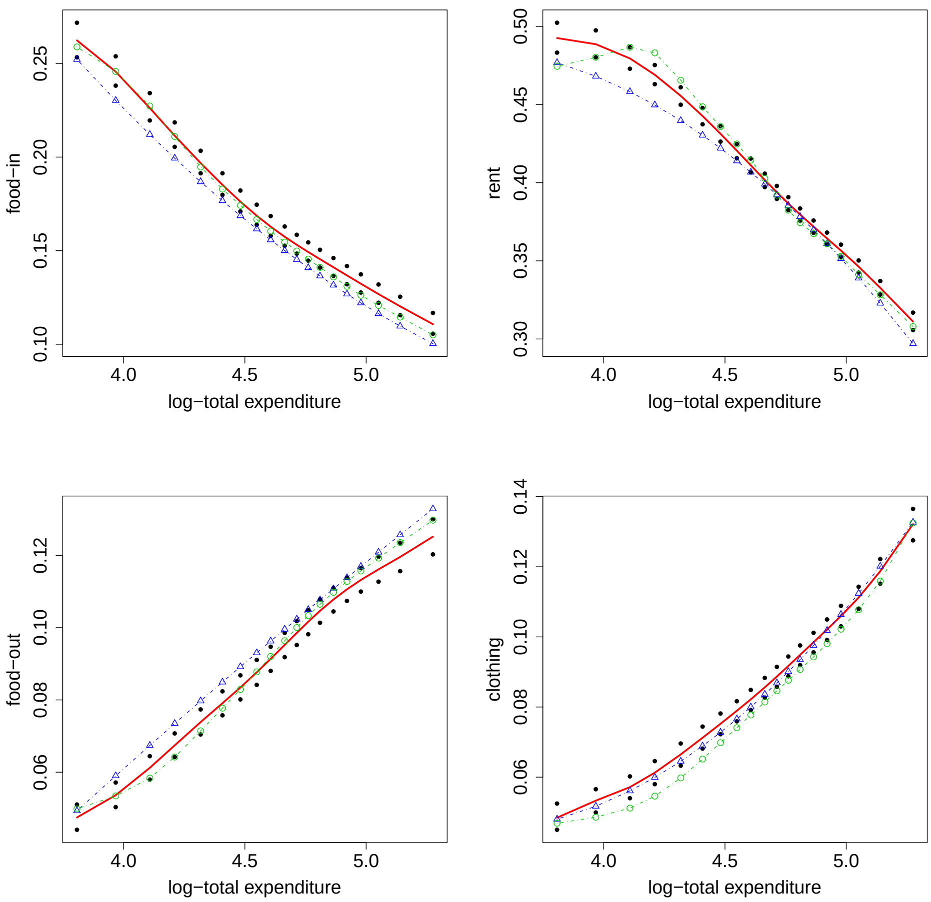

Figure 2: Estimates of food-in, food-out, rent, and clothing (solid line) with 90% pointwise confidence bands (dotted), together with estimates using Banks et al. (1997) (triangles) and Pendakur and Sperlich (2010) (circles).

between the present approach and the partially linear cost function approach of Pendakur and Sperlich (2010). However, two exceptions are the household-operation and public-transportation equations. These are estimated about 0.5 percentage points higher in the present approach. We note that this variation between estimated models is not overwhelmingly large. For example, in some expenditure shares equations, Pendakur and Sperlich (2010) report a difference of more than 0.5 percentage points between symmetry-restricted and symmetry-unrestricted estimates of their partially linear cost function model. This highlights the fact that although the two models are similar in spirit, they are not identical in practise.

Table 5 gives the estimated symmetric price parameters and in brackets the bootstrapped standard deviations. These estimated price effects are in the plausible range, and are similar to those found in Pendakur and Sperlich (2010).

Thus, the estimated Engel curves are plausible and have some evidence of complexity beyond the quadratic form of Banks et al. (1997). Compared to Pendakur and Sperlich (2010), the present approach has an important computational advantage: it is based entirely on observed regressors, and so does not require any numerical inversions to generate a latent regressor. In comparison with Lewbel and Pen-

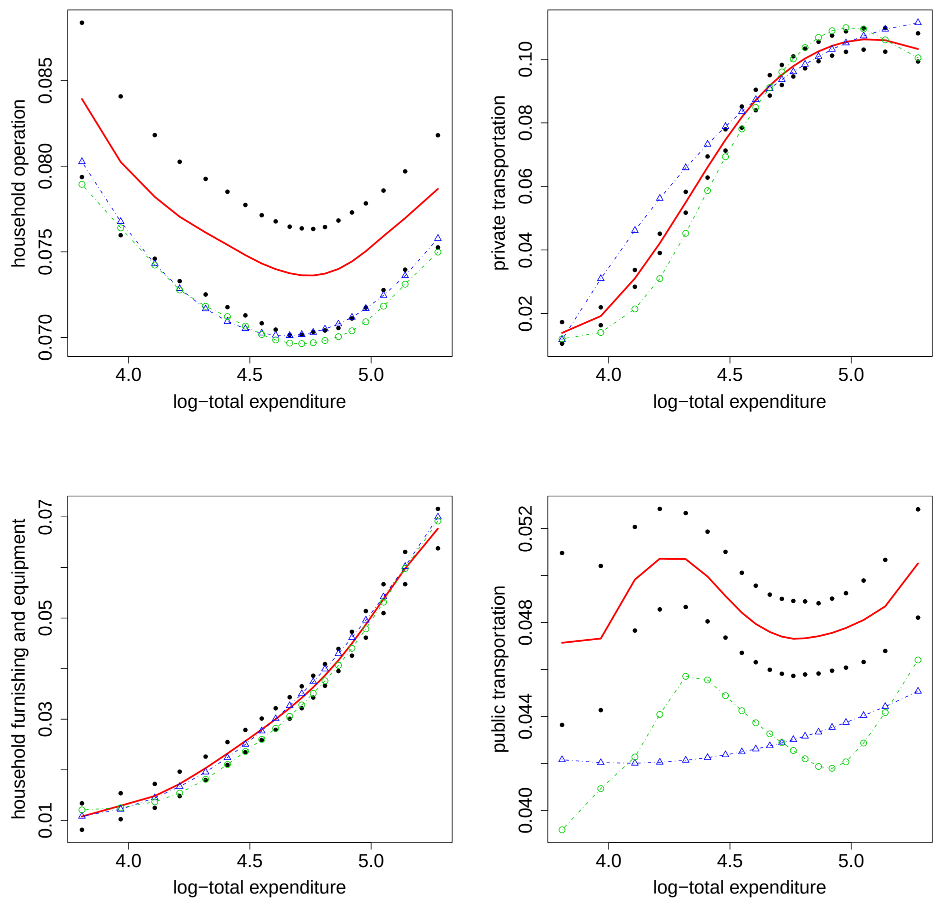

Figure 3: Estimates of household operations, furnishing and equipment, private and public transportation (solid line) with 90% pointwise confidence bands (dashed), together with estimates using Banks et al. (1997) (triangles) and Pendakur and Sperlich (2010) (circles).

dakur (2009), the present approach has an important interpretational difference: whereas Lewbel and Pendakur (2009) must interpret model error terms as unobserved preference heterogeneity parameters, the present approach is based on the more standard view of error terms as measurement or other non-behavioural error.

The varying-coefficients extension is similarly easy to implement. The estimated Engel curves are almost identical to those found in the fixed-coefficients case, with some deviations in the tails. Depending on the bandwidth, the estimated price parameters evaluated at median log-expenditure are statistically indistinguishable from those of the fixed-coefficients model, but their estimated variance is much greater. In particular, we got approximately twice the standard errors for estimated parameters evaluated at median expenditures relative to their fixed-coefficients counterparts. We found that the varying coefficients estimates of $\mathbf{f}$ were very similar to the fixed-coefficient estimates, and so we do not present them here.

Table 5: Estimated symmetric price effects a^{jk} (with bootstrap standard deviations in brackets)

	food–in	food–out	rent	oper	furn	clothing	priv tr	pub tr
food–in	−0.026	0.013	−0.006	−0.008	0.009	0.006	0.037	−0.058
	(0.036)	(0.018)	(0.012)	(0.019)	(0.014)	(0.015)	(0.007)	(0.006)
food–out		−0.035	0.047	0.012	−0.002	−0.069	0.001	−0.045
		(0.014)	(0.007)	(0.012)	(0.010)	(0.009)	(0.005)	(0.005)
rent			0.186	0.023	−0.026	−0.021	−0.036	0.087
			(0.017)	(0.007)	(0.005)	(0.008)	(0.006)	(0.005)
oper				0.040	0.010	−0.016	−0.029	0.023
				(0.017)	(0.011)	(0.011)	(0.004)	(0.005)
furn					−0.038	0.026	−0.017	−0.024
					(0.016)	(0.009)	(0.004)	(0.004)
clothing						0.005	−0.002	−0.014
						(0.010)	(0.005)	(0.004)
priv tr							0.002	0.006
							(0.006)	(0.003)
pub tr								−0.011
								(0.003)

5. Conclusions

We propose a model indirect utility which is nonparametric in the expenditure direction and parametric (with fixed- or varying-coefficients) in the price directions. This utility function implies a consumer demand system that has parametric log–price effects and nonparametric log–total expenditure effects. We avoid the curse of dimensionality typically associated in fully nonparametric estimation of consumer demand since the nonparametric part of the model is only one dimensional. The model is easily restricted to satisfy the rationality conditions of homogeneity and Slutsky symmetry.

We provide a new wild bootstrap procedure that allows for conditional asymmetries and guarantees positive shares. We show the finite sample performance of our estimators in a simulation study, and finally apply our method to Canadian expenditure data.

The application of this model to Canadian price and expenditure data shows not only the potential of the model but also suggests that some expenditure shares more complex than the linear ones in popular parametric demand models. The simulation study reveals further that it is also possible to capture shapes which are difficult to estimate (cf. Hastie and Tibshirani (1984)), such as those with flat plateaus in the intermediate range or with bumps.

Acknowledgements

The first author acknowledges the financial support of the Social Sciences and Humanities Research Council of Canada through its Standard Research Grants program.

The third author acknowledges support by the DFG project FOR916 and the Spanish MTM2008-03010.

References

Banks, J., Blundell, R., Lewbel, A., 1997. Quadratic Engel curves and consumer demand. Review of Economics and Statistics 79, 527-539.

Blundell, R., Chen, X., Kristensen, D., 2003. Semi–nonparametric IV estimation of shape–invariant Engel curves. (Centre for Microdata Methods and Practice, working paper CWP15/03).

Blundell, R., Duncan, A., Pendakur, K., 1998. Semiparametric estimation of consumer demand. Journal of Applied Econometrics 13, 435-461.

Deaton, A., Muellbauer, J., 1980. An almost ideal demand system. American Economic Review 70, 312-326.

Deschamps, P.J., 1988. A note on the maximum likehood estimation of allocation systems. Computational Statistics & Data Analysis 6, 109-112.

Fan, J., Zhang, W., 1999. Statistical estimation in varying coefficient models. Annals of Statistics 27, 1491-1518.

Haag, B.R., Hoderlein, S., Pendakur, K., 2009. Testing and imposing Slutsky symmetry in nonparametric demand systems. Journal of Econometrics 153, 33-50.

Härdle, W., Marron, J.S., 1991. Bootstrap simultaneous error bars for nonparametric regression. The Annals of Statistics 19, 778-796.

Hastie, T.J., Tibshirani, R.J., 1984. Generalized aditive models. (Department of Statistics, Stanford University, Technical Report 98).

Jorgensen, D., Lau, L., Stoker, T., 1980. Welfare comparison and exact aggregation. American Economic Review 70, 268-272.

Lewbel, A., 2000. Consumer demand systems and household equivalence scales, Handbook of Applied Econometrics, Volume II: Microeconomics. Blackwell Publishers Ltd, Oxford.

Lewbel, A., Pendakur, K., 2009. Tricks with Hicks: The EASI demand system. American Economic Review 99, 827-863.

Mas-Colell, A., Whinston, M., Green, J., 1995. Microeconomic Theory. Oxford University Press, New York.

Mazzocchi, M., 2006. Time patterns in UK demand for alcohol and tobacco: an application of the EM algorithm. Computational Statistics & Data Analysis 50, 2191-2205.

Moral-Arce, I., Rodríguez-Póo, J.M., Sperlich, S., 2007. Feasible semiparametric estimation of censored expenditure equations. (Preprint, University of Göttingen, ZfS-2007-13).

Millimet, D., Rusty T., 2008. Estimating demand systems and measuring consumer preferences, Journal of Econometrics 147, 384-395.

Orbe, S., Ferreira, E., Rodriguez-Poo, J., 2003. An algorithm to estimate time-varying parameter SURE models under different types of restriction. Computational Statistics & Data Analysis 42, 363-383.

Pendakur, K., 2002. Taking prices seriously in the measurement of inequality. Journal of Public Economics 86, 47-69.

Pendakur, K., Sperlich, S., 2010. Semiparametric estimation of consumer demand systems in real expenditure. Journal of Applied Econometrics 25, 420-457.

Pollack, R., Wales, T., 1991. Demand system specification and estimation. Oxford University Press, New York.

Roy, R., 1947. La Distribution du Revenu Entre Les Divers Biens, Econometrica 15, 205-225.

Sarmiento, C., 2005. A varying coefficient approach to global flexibility in demand analysis: a semiparametric approximation. American Journal of Agricultural Economics 87, 38-47.

Sperlich, S., 2009. A note on nonparametric estimation with predicted variables. The Econometrics Journal 12, 382-395.

Varian, H.R., 1978. Microeconomic Analysis. Norton, New York.