Semiparametric Indirect Utility and Consumer Demand

Krishna Pendakur*, Michael Scholz, and Stefan Sperlich†

October 15, 2008

Abstract

In this paper, we specify the indirect utility function as a partially linear model, where utility is nonparametric in expenditure and parametric (with fixed- or varying-coefficients) in prices. Since we start with a model of indirect utility, rationality restrictions like homogeneity and Slutsky symmetry are easily imposed. The resulting model for expenditure shares (as function 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. Our basic insight is that given a local polynomial model for the numerator, the denominator is is given by a lower-order local polynomial. The model is thus easily estimated using modified versions of standard local polynomial modeling techniques. We provide Monte Carlo evidence that the proposed techniques work, and implement the model on Canadian consumer expenditure and price micro-data.

Keywords: Indirect Utility, Consumer Demand, Engel Curves, Semiparametric Econometrics, Computational Statistics.

*Department of Economics, Simon Fraser University, Burnaby, BC, CANADA, V5A 1S6, e-mail: pendakur@sfu.ca

†Institut für Statistik und Ökonometrie, Georg-August Universität Göttingen, Platz der Göttinger Sieben 5, 37073 Göttingen, Germany. e-mail: stefan.sperlich@wiwi.uni-goettingen.de.
The last author acknowledges support by the DFG project FOR916.

1 Introduction

The specification and estimation of consumer demand systems, defined as the relationship between demands, prices and expenditures, presents many long-standing problems in econometric theory. Recent work has focused on the inclusion of highly nonlinear relationships between demand and total expenditure into empirical models of consumer demand. Since typical consumer demand microdata typically have a large amount of variation in total expenditure across consumers, complex relationships between demand and expenditure might potentially be identified. Unfortunately, because consumer demand models must satisfy a set of (nonlinear cross-equation) rationality restrictions known as the Slutsky symmetry restrictions, such complex relationships have been hard to incorporate. In this paper, we present a semiparametric approach to the consumer demand problem which allows for the simple 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 arbitrarily flexible Engel curves. However, in the price directions – where the date are less rich – we propose a parametric structure.

Nonparametric approaches to consumer demand started by considering the Engel curve, defined as the relationship of expenditure-shares commanded by each good to the total expenditure of the consumer, at fixed vectors of prices. That is, they considered only 1 nonparametric direction and held the others fixed. Work by [3] and [2] revealed considerable complexity in the shapes of Engel curves. A fully nonparametric approach which consider both price and expenditure directions together and which allows for the imposition of rationality restrictions, has recently been developed by [5]. Here, the shape of the demand curves is not restricted, but curse of dimensionality rears its head: in a world with M price directions and 1 expenditure direction, the researcher faces an $M+1$ dimensional problem[1].

On the other hand, parametric approaches like the popular Almost Ideal [4], Translog [9] and Quadratic Almost Ideal [1] demand models typically impose strict limits on the functional complexity of Engel curves. In these cases, Engel curves must be linear, nearly linear, or quadratic, respectively. This lack of complexity is driven by the need for these parametric models to satisfy the Slutsky symmetry restrictions.

In between the fully nonparametric and the fully parametric approaches, we have the realm of semiparametric econometrics. Two recent papers have explored this area, accomodating the need for the structural model to satisfy Slutsky symmetry. [11]

[1]If homogeneity, another rationality condition, is imposed, this drops to an M dimensional problem and thus not a remedy for the curse of dimensionality.

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 – i.e., nonparametric – 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. [12] 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 (via varying coefficients). While potentially appealing, implementation of their model requires numerical inversions of unknown functions. Both of these semiparametric approaches address the curse of dimensionality: they each have just 1 nonparametric dimension rather than $M + 1$ nonparametric dimensions.

The present paper is a semiparametric contribution, similar in spirit to [12], but does not require the researcher to undertake numerical inversions of unknown functions. [12] write down a model in which expenditure-shares are nonparametric in utility, an unobserved regressor, and 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 write down a model in which utility is nonparametric in log-expenditure and parametric in log-prices. This results in a model of expenditure-shares with only 1 nonparametic dimension which is locally nonlinear but has no unobserved or generated regressors.

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 gives expenditure shares as the ratio of derivatives of indirect utility, expenditure shares in our model are given by a ratio. This ratio has model parameters in both the numerator and denominator. In particular, the ratio which characterises expenditure shares has nonparametric functions in the numerator and their derivatives in the denominator. Because local estimators have to be evaluated locally at a large number of points, a locally nonlinear model, which may take a long time to evaluate at each point, is typically not very useful. 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. Furthermore, large data sets can be handled in acceptable time, and the results are readily interpreted.

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 accomodate the parametric part of the model, we use a profiled estimator. In the applied part of the paper we show the power of our method in a short simulation study, and implement the model with Canadian price and expenditure data.

The paper is organised as follows. In section 2 we introduce the model specification. In section 3 we discuss the basic estimation idea and give the associated algorithm. The empirical part of the paper can be found in section 4, where we present the estimation results of the Canadian microdata and the simulation study. Here, we find that some expenditure–share equations show quite a lot of nonlinearity, i.e. these are S-shaped or even more complex. Section 5 concludes the present work and discusses some possible extensions.

2 A Semiparametric Model for Indirect Utility

Define the indirect utility function $V(p,x)$ to give the maximum utility attained by a consumer when faced with a vector of log–prices $p = (p^1, \ldots, p^M)$ 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 $w = (w^1, \ldots, w^{M-1})$ 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.

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

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

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 $\mathbf{A} = \{a^{kl}\}_{k,l=1}^{M}$ symmetric parameters, i.e. $a^{kl} = a^{lk}$ to fulfil Slutsky symmetry. Second, we consider the varying-coefficient extension of this model:

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

or, in matrix notation,

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

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 to not have nearly as much variation in the price direction, which suggests that partially linear modelling might describe these effects sufficiently well. If additionally, the researcher feels that more may be identified on the strength of observed variation, the varying-coefficients model allows prices 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 dimensionality is still kept to 1 throughout.

2.1.1 Rationality Restrictions: Homogeneity

In both cases, utility is obviously a non–increasing function in prices and for certain f^k non–decreasing in total expenditure. Rationality is comprised of three conditions: homogeneity, symmetry and concavity. As is common in this literature, we will deal only with homogeneity and symmetry. First consider homogeneity, which 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, which implies $V(p,x) = V(p+\lambda, x+\lambda)$ for all λ. 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:

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

in model (1) and analogously in model (3). The sums go only to $M-1$ because the M-th element of the sums is zero. Using $\tilde{x} = x - p^M$, $\tilde{p}^j = p^j - p^M$ and $\tilde{p} = (\tilde{p}^1, \ldots, \tilde{p}^{M-1})$ we may write

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

Given an indirect utility function, Marshallian (uncompensated) expenditure share equations may be recovered via the logarithmic version of Roy's Identity: $w^j(p,x) = [\partial V(p,x)/\partial p^j] \,/\, [\partial V(p,x)/\partial V x]$. In the following we will only think in vectors of length $(M-1)$, i.e. skipping the last elements as this will typically be fixed to

$$w^M(\tilde{p}, \tilde{x}) = 1 - \sum_{i=1}^{M-1} w^i(\tilde{p}, \tilde{x}),$$

guaranteeing that $\sum_{j=1}^{M} w^j = 1$ for all sets of $(\tilde{p}, \tilde{x})$. Then, applying Roy's identity to our indirect utility function, we get the uncompensated 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 model (1), and

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

respectively, where $\nabla_{\tilde{x}} \mathbf{f}(\tilde{x})$ is the $(M-1)$ vector of the derivatives of $\mathbf{f}(\tilde{x})$, and $\nabla_{\tilde{x}} \mathbf{A}(\tilde{x})$ is the $(M-1) \times (M-1)$ matrix function equal to the derivatives of $\mathbf{A}$.

These expressions for budget shares have a nice feature in comparison to [12]. Whereas their model for expenditure shares uses a nonparametric function of a generated regression, the expression above uses only observed regressors. However, in comparison to [12], which is a partially linear model, the above expression is partially linear only in the numerator. The presence of the denominator makes it seem difficult to implement. However, as we show below, use of local polynomial estimators makes this problem fairly easy to solve.

2.1.2 Rationality Restrictions: Slutsky Symmetry

The imposition of Slutsky symmetry requires that $a^{lk} = a^{kl}$ $(a^{lk}(x) = a^{kl}(x))$ for all k, l, or equivalently, that $\mathbf{A} = \mathbf{A}'$ $(\mathbf{A}(x) = \mathbf{A}(x)')$. To see this, start with the definition of the Slutsky matrix, whose elements are given by

$$s_{ij} = \frac{\partial h^i}{\partial b^j} = \frac{\partial g^i}{\partial y} \cdot q^j + \frac{\partial g^i}{\partial b^j}, \tag{8}$$

where h^i denotes the Hicksian demand, g^i the Marshallian demand, y total expenditure, b^j the price, and q^j the quantity of good/category j. With $q^i = g^i = w^i \cdot y/b^i$, where g^i and w^i are functions of $y, b^1, \ldots, b^M$, we get

$$\frac{\partial g^i}{\partial y} = \frac{\partial w^i}{\partial y} \cdot \frac{y}{b^i} + \frac{w^i}{b^i} \quad \text{and} \quad \frac{\partial g^i}{\partial b^j} = \frac{\partial w^i}{\partial b^j} \cdot \frac{y}{b^i}. \tag{9}$$

Using the equations in (8) and (9) we can write the difference of $s_{ij} - s_{ji}$, for $i, j = 1, \ldots, M-1$, as

$$s_{ij} - s_{ji} = \left(\frac{\partial w^i}{\partial y} \cdot w^j - \frac{\partial w^j}{\partial y} \cdot w^i \right) \frac{y^2}{b^i b^j} + \left(\frac{\partial w^i}{\partial b^j} \cdot \frac{y}{b^i} - \frac{\partial w^j}{\partial b^i} \cdot \frac{y}{b^j} \right). \tag{10}$$

With the abbreviations $T^i = \sum a^{ik} \tilde{p}^k$, $S = 1 - \sum \partial f^k / \partial \tilde{x} \cdot \tilde{p}^k$ and $f^i = f^i(\tilde{x})$, we can rewrite (6) in the following way

$$w^i = \frac{f^i + T^i}{S}, \tag{11}$$

but note that w^i depends on $\tilde{x} = \log y - \log b^M$, log-total expenditure, and $\tilde{p}^j = \log b^j - \log b^M$ the log-prices for $j = 1, \ldots, M-1$. Now we can differentiate (11) w.r.t. total expenditure and the j-th price, and obtain with $U = \sum \partial^2 f^k / \partial \tilde{x}^2 \cdot \tilde{p}^k$

$$\frac{\partial w^i}{\partial y} = \frac{\frac{\partial f^i}{\partial \tilde{x}} \cdot S + (f^i + T^i) \cdot U}{y S^2} \quad \text{and} \quad \frac{\partial w^i}{\partial b^j} = \frac{a^{ij} S + (f^i + T^i) \cdot \frac{\partial f^j}{\partial \tilde{x}}}{b^j S^2}.$$

Plugging-in these results and equation (11) in (10), we get immediately that $s_{ij} - s_{ji} = 0$ if $a_{ij} = a_{ji}$.

3 Estimation of the Models

The main advantage of the above models is the enormous dimension reduction for the nonparametric part of the estimation procedure. Instead of $M+1$ or M dimensions for a fully nonparametric model, we face only functions which are one dimensional in total–expenditure. Thus, we circumvent the curse of dimensionality.

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 local linear kernel estimators for the nonparametric part, and in case of model (1) restricted (due to the symmetry restriction) least squares for the parametric coefficients. Clearly, in this case, the method is based on profiled estimators.

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

$$\begin{aligned} \mathbf{f}(t) &\approx \mathbf{f}(\tilde{x}) + \nabla_{\tilde{x}}\mathbf{f}(\tilde{x})(t-\tilde{x}) && (12) \\ &\approx \alpha(\tilde{x}) + \beta(\tilde{x})(t-\tilde{x}), && (13) \end{aligned}$$

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

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

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

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

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

Here, the imposition of homogeneity is via the use of normalised prices and expenditures (tilda'd quantities). 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 $\boldsymbol{\Gamma}(\tilde{x}) = \boldsymbol{\Gamma}(\tilde{x})'$ and $\boldsymbol{\Delta}(\tilde{x}) = \boldsymbol{\Delta}(\tilde{x})'$.

We could also use higher order approximations (and thus higher-order local polynomials), but for this we would need stronger assumptions on data and model (e.g. higher order smoothness of functions and densities). The local linear approach however nests the parametric linear model for any smoothing bandwidth without a bias. Therefore we confine ourselves to the local linear approximation. We will now be more specific but we will see that the steps are easily generalised.

3.2 The Estimation Algorithm

Denote $\Delta_i = \tilde{X}_i - \tilde{x}$, $K_i = K((\tilde{X}_i - \tilde{x})/h)/h$ and $d = M-1$, where K is some symmetric kernel function with the usual properties and h a bandwidth that controls the smoothness of the estimate.

Le us start with the profiled 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 $\tilde{x}$:

$$\sum_{j=1}^{d}\sum_{i=1}^{N}\left(W_i^j-\frac{\alpha^j+\Delta_i\beta^j+\sum\limits_{k=1}^{d}a^{jk}\tilde{P}_i^k}{1-\sum\limits_{k=1}^{d}\beta^k\tilde{P}_i^k}\right)^2 K_i \longrightarrow \underset{\alpha^j,\beta^j}{\text{Min!}} \tag{14}$$

In order to minimize, we set the first derivative equal to zero. Taking the derivative of (14) with respect to α^j, and using the notations $S_i = 1-\sum_{k=1}^{d}\beta^k\tilde{P}_i^k$ and $T_i^j = \sum_{k=1}^{d}a^{jk}\tilde{P}_i^k$, we solve

$$0=\sum_{i=1}^{N}\left(W_i^j-\frac{\alpha^j+\Delta_i\beta^j+T_i^j}{S_i}\right)\frac{K_i}{S_i}. \tag{15}$$

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

$$\alpha^j=\frac{\sum\limits_{i=1}^{N}W_i^jK_i/S_i-\beta^j\sum\limits_{i=1}^{N}K_i\Delta_i/S_i^2-\sum\limits_{i=1}^{N}K_iT_i^j/S_i^2}{\sum\limits_{i=1}^{N}K_i/S_i^2}. \tag{16}$$

On the other hand, by differentiating (14) with respect to β^j (again for $j=1,\ldots,d$), we get the equations

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

This is equivalent to

$$\begin{aligned} 0 \;=\; & \sum_{k=1}^{d}\sum_{i=1}^{N}\left(W_i^k-\frac{\alpha^k+\Delta_i\beta^k+T_i^k}{S_i}\right)K_i\cdot\frac{(\alpha^k+\Delta_i\beta^k+T_i^k)\tilde{P}_i^j}{S_i^2}+\\ & \sum_{i=1}^{N}\left(W_i^j-\frac{\alpha^j+\Delta_i\beta^j+T_i^j}{S_i}\right)K_i\frac{\Delta_i}{S_i}. \end{aligned} \tag{17}$$

Unfortunately, we can not solve equation (17) analytically for β^j. But, for our iterative purpose it is enough to consider the following implicit representation:

$$\begin{aligned}\beta^j &= \Big[\sum_{k=1}^{d}\sum_{i=1}^{N}\left(W_i^k - \frac{\alpha^k+\Delta_i\beta^k+T_i^k}{S_i}\right)K_i\cdot\frac{(\alpha^k+\Delta_i\beta^k+T_i^k)\tilde{P}_i^j}{S_i^2} + \\ &\quad \sum_{i=1}^{N}\left(W_i^j - \frac{\alpha^j+T_i^j}{S_i}\right)K_i\frac{\Delta_i}{S_i}\Big]\Big/\sum_{i=1}^{N}\frac{K_i\Delta_i^2}{S_i^2}. \qquad (18)\end{aligned}$$

We use the implicit representation (18) to calculate new values for β^j. With them we get new S_i so that we can find new α^j:

$$\begin{aligned}\beta^j_{new} &= \Big[\sum_{k=1}^{d}\sum_{i=1}^{N}\left(W_i^k - \frac{\alpha^k_{old}+\Delta_i\beta^k_{old}+T^k_{i,old}}{S_{i,old}}\right)K_i\frac{(\alpha^k_{old}+\Delta_i\beta^k_{old}+T^k_{i,old})\tilde{P}_i^j}{S^2_{i,old}} + \\ &\quad \sum_{i=1}^{N}\left(W_i^j - \frac{\alpha^j_{old}+T^j_{i,old}}{S_{i,old}}\right)K_i\frac{\Delta_i}{S_{i,old}}\Big]\Big/\sum_{i=1}^{N}\frac{K_i\Delta_i^2}{S^2_{i,old}}, \qquad (19)\end{aligned}$$

$$S_{i,new} = 1-\sum_{k=1}^{d}\beta^k_{new}\tilde{P}_i^k, \qquad (20)$$

$$\alpha^j_{new} = \frac{\sum_{i=1}^{N}W_i^jK_i/S_{i,new} - \beta^j_{new}\sum_{i=1}^{N}K_i\Delta_i/S^2_{i,new}\sum_{i=1}^{N}K_iT^j_{i,old}/S^2_{i,new}}{\sum_{i=1}^{N}K_i/S^2_{i,new}}.$$

We repeat these steps to convergence. As noted, this is a profiled estimator with fixed a^{kl}. So, 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

$$W_i^j\cdot(1-\sum_{k=1}^{d}\beta_i^k\tilde{P}_i^k)-\alpha_i^j = \sum_{k=1}^{d}a^{jk}\tilde{P}_i^k. \qquad (21)$$

Details are given in the Appendix. Here, we see why we need $\mathbf{A}$ to be symmetric for identification: we can only identify the sum of the symmetric effects.

The modification of the algorithm to take the varying coefficients $\mathbf{A}(\tilde{x})$ into account is quite straightforward, though tedious. With the same local linear approximation arguments as above, we get the local problem in the neighbourhood of $\tilde{x}$ as

$$\sum_{j=1}^{d}\sum_{i=1}^{N}\left(W_i^j - \frac{\alpha^j+\Delta_i\beta^j+\sum_{k=1}^{d}(\gamma^{jk}+\Delta_i\delta^{jk})\tilde{P}_i^k}{1-\sum_{k=1}^{d}\beta^k\tilde{P}_i^k-\frac{1}{2}\sum_{k=1}^{d}\sum_{l=1}^{d}\delta^{kl}\tilde{P}_i^k\tilde{P}_i^l}\right)^2 K_i \longrightarrow \min_{\theta} \qquad (22)$$

with θ all $\alpha^j, \beta^j, \gamma^{jk}$ 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 (22) 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 $C_i^s = \alpha^s + \Delta_i \beta^s + \sum \Delta_i \delta^{sk} \tilde{P}_i^k$ and

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

with $T_i^{s,-t} = T_i^s - \Delta_i \delta^{st} \tilde{P}_i^t$.

3.3 Practical Considerations

One important 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 $\tilde{P}_i = (0, \ldots, 0)$ for some group of consumers, equation (14) reduces to the well-known local linear case. That is, since $S_i = 0$, we get the objective function

$$\sum_{j=1}^d \sum_{i=1}^N \left(W_i^j - \alpha^j + \Delta_i \beta^j\right)^2 K_i \longrightarrow \underset{\alpha^j, \beta^j}{\text{Min!}}$$

Solving this problem on the sample of consumers where $\tilde{P}_i = (0, \ldots, 0)$ gives us consistent estimates that we can use as starting values for α^j and β^j. As a natural choice for the starting values of γ^{jk} we use the results of the algorithm in Section 3.2 and zero for all δ^{jk}.

For the bandwidth choice, we use the same bandwidth h for all expenditure categories because the functions refer to the same expenditure data in all equations. As bandwidth choice criterion one may use the usual cross–validation or plug-in rules like Silverman's (though this was constructed for density estimates rather than regression).

Note finally for the profiled estimator in the fixed-coefficients model 1), we recommend running the whole estimation algorithm twice: first with an undersmoothing bandwidth in the nonparametric part to keep the possible bias small. The resulting estimate for the coefficient matrix **A** is kept, and used in the second run which uses a larger bandwidth for the nonparametric part to get reasonably smooth function estimates. This is of course unneccessary in the varying coefficients model (3) where we face only nonparametric functions.

4 Empirical Analysis

To get an idea about the finite sample performance of the method, before we analyse household expenditures in Canada, we start with a brief simulation study. Afterwards, we also introduce a wild bootstrap procedure for further inference like the construction of confidence intervals.

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 [10]. 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), the 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.

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 the nonparametric part of the estimation procedure we used the gaussian kernel and a data-adaptive bandwidth of $h \approx 0.034$.

In Figure 1 we have plotted the true functions (solid lines) together with intervals of 90% coverage probabilities for the estimates (dashed 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. However, we also see the limits of the method: for example, boundary effects seem important.

In Table 3 are given the estimated parameter means, together with the standard deviations. The exactness of our simulation results in a very 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

4.2 Bootstrap Inference

The "wild bootstrap" draws new artificial responses from the estimated model (1), so we must first assume that we have semiparametric estimates as described in the

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)

previous sections, namely the estimators $\hat{\alpha}^j$, $\hat{\beta}^j$ and $\hat{a}^{jk}$, $k, j = 1, \ldots, d$. Denote an oversmoothed bandwidth g with $g > h$, and let h is the bandwidth giving us the desired smoothness in the original sample. The basic idea is now (cf. [6]) to use the estimated residuals from a estimates with bandwidth g,

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

to construct a bootstrap sample $\{Y_i^*, X_i, \boldsymbol{P}_i\}_{i=1}^N$ from which we get oversmoothed residuals $\hat{\varepsilon}_i^j$.

Let u be a standard normal random variable and $\hat{\sigma}_\varepsilon$ the estimated standard deviation of the error terms $\hat{\varepsilon}_i^j$. Given homoscedasticity, the wild bootstrap errors are defined by $\varepsilon_i^{j*} = u \cdot \hat{\sigma}_\varepsilon^j$. When allowing for heteroscedasticity, we construct bootstrap errors of type $\varepsilon_i^{j*} = u \cdot \hat{\varepsilon}_i^j$. The generated bootstrap sample is then (for $i = 1, \ldots, N$) given by

$$W_i^{j*} = \frac{\hat{\alpha}^j(\tilde{X}_i) + \sum_{k=1}^{d} \hat{a}^{jk} \tilde{P}_i^k}{1 - \sum_{k=1}^{d} \hat{\beta}^k(\tilde{X}_i)} + \varepsilon_i^{j*}. \tag{23}$$

Given this bootstrap sample, we may estimate the model, and take the variance of estimated model parameters as equal to the variance of model parameters across the bootstrap samples. When we estimate the nonparametric functions, from the

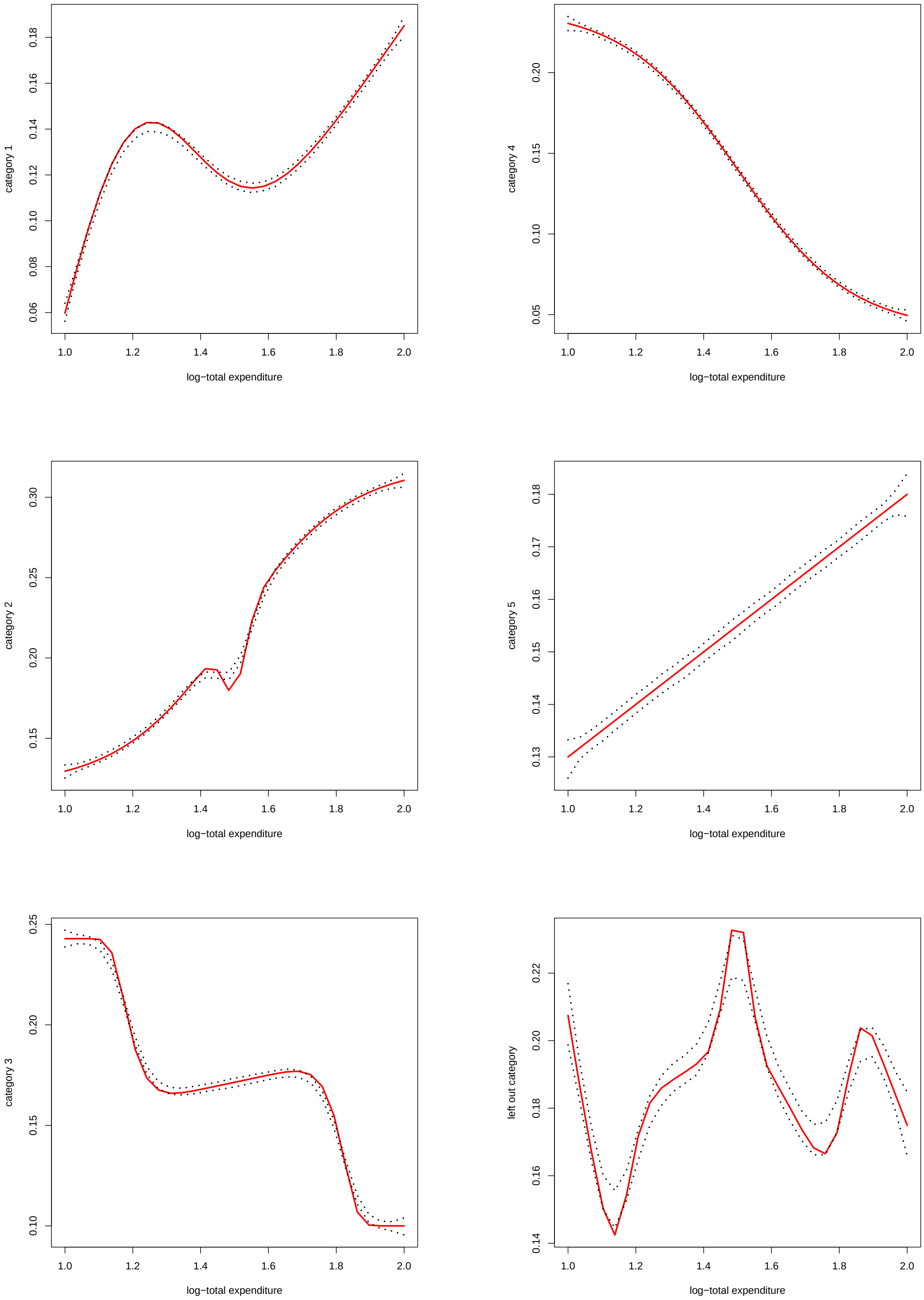

Figure 1: Simulation of 6 different budget share functions (solid line) with 90% coverage probability (dashed lines)

bootstrap or the original sample, we have to use the bandwidth h. Note that we should choose the bandwidth g used in the construction of the bootstrap errors in such a way that it tends to zero at a slower rate than the optimal bandwidth h (cf. [6]).

A disadvantage of this method is that it may happen that boundaries of the confidence intervals are not in $[0, 1]$, especially when the estimate $\hat{\alpha}^j$ at some point x is close to 0 or 1. Note further that the generated bootstrap samples do not necessarily fulfill the required constraints, i.e. $W_i^j \in [0, 1]$ and $\sum_j W_i^j = 1$. In practice, however, this does not matter very much because neither in the first nor in the second step of the estimation procedure 3.2 do we control for these conditions. This is also the reason why we should not force the values W_i^{j*} to satisfy the constraints. One possible solution for this problem would be the usage of tail truncated error terms. This is quite plausible because to get shares between zero and one we cannot allow all possible error values. In the following simulation study we reconsider this argument. The other possibility is to respect the constraints in the estimation algorithm, which may pose an interesting question for further research.

4.3 Analysing Household Expenditures in Canada

In our empirical example we use the same Canadian data as in [11] or [12] which come from public sources. 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. $\tilde{p}_{O,86} = (0, \ldots, 0)$, 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 dependence 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 in the given example 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 quarter of the current consumption of the households in the sample. Summary statistics of the observations are given in Table 4.

As noted above, when $\tilde{p} = (0, \ldots, 0)$ (as it does for observations in Ontario 1986),

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

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. We included into the graphics also the pointwise 90% confidence intervals which we calculated as described in Section 4.2 with heteroscedastic error terms and 500 bootstrap iterations. To generate the bootstrap samples we used the bandwidth $g = 0.9$ and in the estimation of the functions α^j the bandwidth $h = 0.75$. The nonparametric estimation method, where we estimated over a grid of 30 equispaced points using the gaussian kernel, converged in our setting on average after 15 iterations. A nonlinear functional relationship can be observed for example between private transportation and log–total expenditure in Figure 3. As noted in previous work [12] we find again that expenditure shares are near by zero for the bottom quintile of the population, rise through the middle of the population, and fall at the top quintile.

Table 5 gives the estimated symmetric price parameters and in brackets the bootstraped standard deviations.

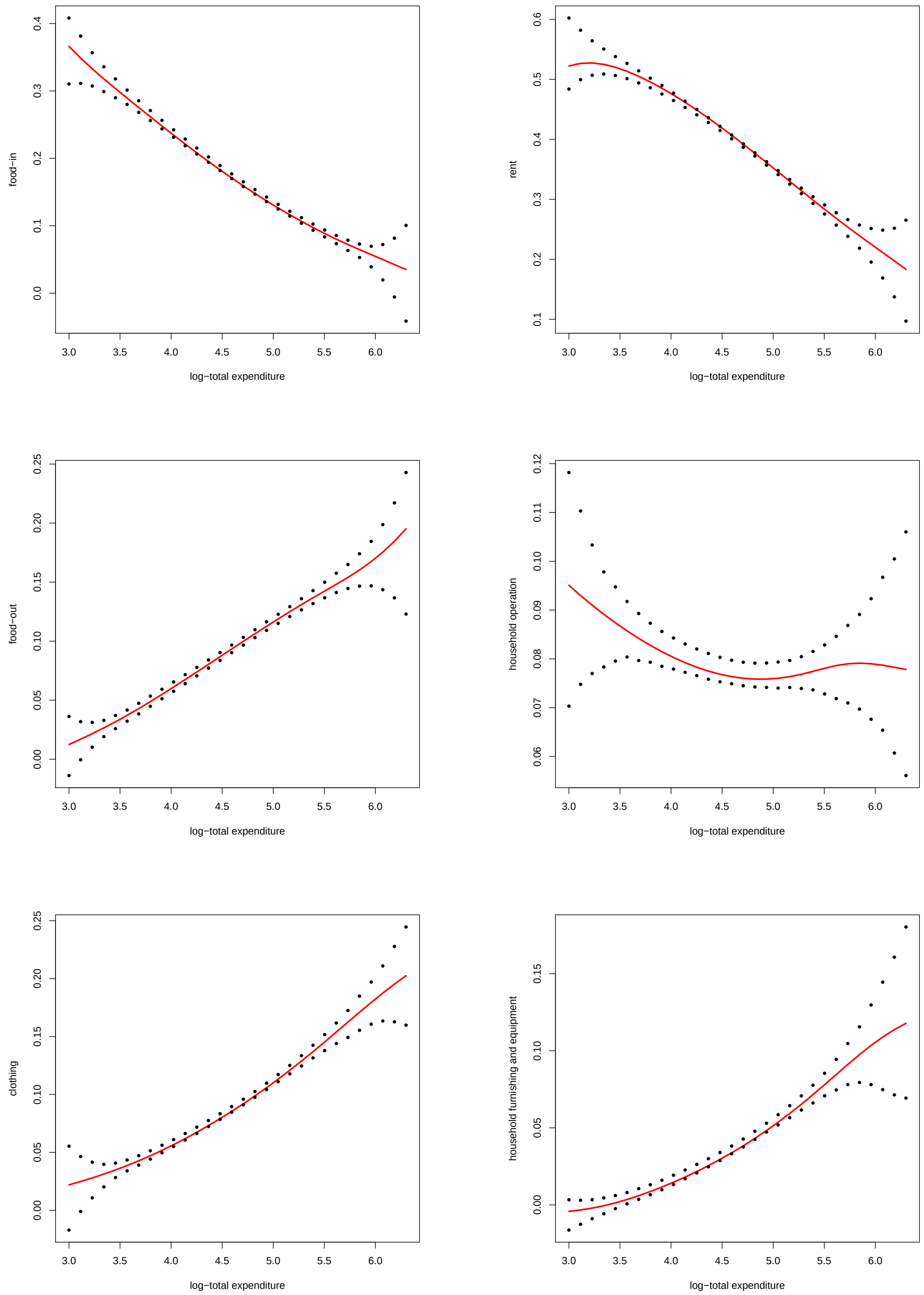

Figure 2: Estimates of food-in, food-out, clothing, rent, household operations, furnishing and equipment (solid line) with 90% pointwise confidence intervals

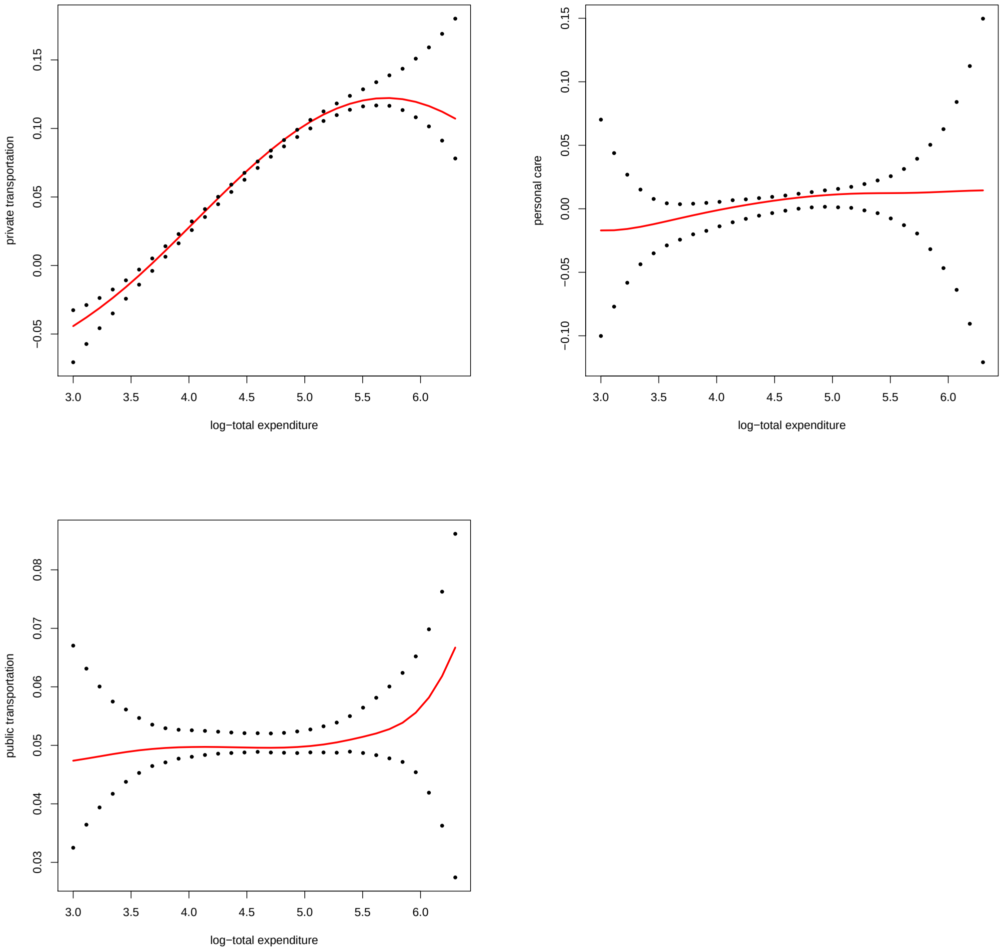

Figure 3: Estimates of personal care, private and public transportation (solid line) with 90% pointwise confidence intervals

These estimated price effects are in the plausible range, and are similar to those found in [12]. However, whereas [12] faced a complex numerical inversion problem for there generated regressor, our model is evaluated entirely on observables.

The varying-coefficients extension is similarly easy to implement. We use the same bandwidths (which are driven in the main by the fit for $\mathbf{f}(\tilde{x})$) as in the fixed-coefficients case. The estimated Engel curves are almost identical to those found in the fixed-coefficients case, with some small deviations in the tails. For the sake of brevity, we do not present those results here, but they are available on request from the authors. 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 see approximately twice the standard errors for estimated parameters evaluated at median expenditures relative

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

	food–in	food–out	rent	oper	furn	clothing	priv tr	pub tr
food–in	−0.026	0.022	−0.003	−0.002	0.017	0.003	0.034	−0.057
	(0.021)	(0.016)	(0.007)	(0.016)	(0.014)	(0.012)	(0.006)	(0.006)
food–out		−0.050	0.055	0.008	−0.006	−0.078	0.002	−0.045
		(0.016)	(0.007)	(0.012)	(0.012)	(0.010)	(0.006)	(0.006)
rent			0.171	0.024	−0.028	−0.015	−0.044	0.093
			(0.010)	(0.005)	(0.005)	(0.005)	(0.006)	(0.004)
oper				0.039	0.007	−0.017	−0.029	0.024
				(0.017)	(0.012)	(0.009)	(0.005)	(0.005)
furn					−0.024	0.017	−0.018	−0.021
					(0.019)	(0.010)	(0.005)	(0.006)
clothing						0.002	0.001	−0.016
						(0.010)	(0.005)	(0.004)
priv tr							−0.001	0.006
							(0.007)	(0.004)
pub tr								−0.011
								(0.004)

to their fixed-coefficients counterparts. Again, we do not present those estimates here for the sake of brevity, but they are available on request.

5 Conclusions

We propose a model which starts with the indirect utility function and implies a consumer demand system that has parametric log–price effects and nonparametric log–total expenditure effects. Furthermore, 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 in log–total expenditure. The model is easily restricted to satisfy the rationality conditions of homogeneity and Slutsky symmetry.

A detailed explanation of the estimation procedure shows the working of the method, and a simulation shows finite sample performance. We provide a bootstrap procedure for possible further inference, 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 the popular AID [4] and Translog [9] demand models. The simulation study reveals further that it is also possible to estimate functions which are difficult to estimate [8], such as those with flat plateaus in the intermediate range or with bumps.

6 Appendix: Restricted Least Squares for a Symmetric Matrix A

Recall that to estimate the symmetric parameters a^{jk}, $j,k = 1,\ldots,d$, we use equation (6) and get, with some algebra, for a single individual i

$$W_i^j \cdot (1 - \sum_{k=1}^{d} \beta_i^k \tilde{P}_i^k) - \alpha_i^j = \sum_{k=1}^{d} a^{jk} \tilde{P}_i^k. \tag{24}$$

Here, the parameters $\alpha_i^j = \alpha^j(\tilde{X}_i)$ are related to the functions f^j at the point $\tilde{X}_i$ and the parameters $\beta_i^j = \beta^j(\tilde{X}_i)$ to its first derivatives. Defining $(\boldsymbol{W})_{ij} := W_i^j \cdot (1 - \sum_{k=1}^{d} \beta_i^k \tilde{P}_i^k) - \alpha_i^j$, $(\boldsymbol{P})_{ik} := \tilde{P}_i^k$ and $(\boldsymbol{A})_{kj} := a^{kj}$ we can formulate equation (24) using matrix notation:

$$\boldsymbol{W} = \boldsymbol{P} \cdot \boldsymbol{A}, \tag{25}$$

where $\boldsymbol{W}$, $\boldsymbol{P}$ are $N \times d$ matrices and $\boldsymbol{A}$ a $d \times d$ symmetric matrix. Note that it is not necessary to start in the model description (1) with symmetric parameters a^{jk}. However, when we start with arbitrary parameters we will end nevertheless in (25) with a symmetric parameter matrix. More specific, we get for (24):

$$W_i^j \cdot (1 - \sum_{k=1}^{d} \beta_i^k \tilde{P}_i^k) - \alpha_i^j = 1/2 \sum_{k=1}^{d} (a^{jk} + a^{kj}) \tilde{P}_i^k$$

and in matrix notation

$$\boldsymbol{W} = \frac{1}{2} \boldsymbol{P}(\boldsymbol{A} + \boldsymbol{A}^t).$$

Obviously, $\boldsymbol{A} + \boldsymbol{A}^t$ is symmetric and we would have an identification problem for nonsymmetric $\boldsymbol{A}$. In other words, even if one does not require symmetry from the beginning, only symmetry of $\boldsymbol{A}$ makes the estimation problem identifiable.

Next, to calculate the unknown matrix $\boldsymbol{A}$ in equation (25) we should not use the standard least square method but want directly make use of the symmetry of $\boldsymbol{A}$. Denote $w^{ij} = (\boldsymbol{W})_{ij}$ and w^j the j–th column of $\boldsymbol{W}$, $p^{ij} = (\boldsymbol{P})_{ij}$, p_i the i–th row and p^j the j–th column of the price–matrix $\boldsymbol{P}$ and a^j also the j–th row of $\boldsymbol{A}$. Note that the symmetric matrix $\boldsymbol{A}$ obtains only $(d^2+d)/2$ different parameters which are found for example in the lower triangular part, including the diagonal elements

$$\boldsymbol{A} = \begin{pmatrix} a_1 & a_2 & \dots & a_d \\ a_2 & a_{d+1} & \dots & a_{2d-1} \\ \dots & \dots & \dots & \dots \\ a_d & a_{2d-1} & \dots & a_{(d^2+d)/2} \end{pmatrix} = \begin{pmatrix} a^{11} & a^{12} & \dots & a^{1d} \\ a^{21} & a^{22} & \dots & a^{2d} \\ \dots & \dots & \dots & \dots \\ a^{d1} & a^{d2} & \dots & a^{dd} \end{pmatrix}.$$

Let $\boldsymbol{a}_p$ be the one–dimensional array formed by these parameters,

$$\boldsymbol{a}_p = (a_1, \dots, a_{(d^2+d)/2}),$$

then we have to find the vector $\boldsymbol{a}_p$ that minimises

$$S := \sum_{j=1}^{d} \sum_{i=1}^{N} (w_{ij} - \langle\, p_i, a^j \,\rangle)^2 \longrightarrow \underset{\boldsymbol{a}_p}{\mathrm{Min!}} \tag{26}$$

We obtain by differentiation of (26) with respect to all elements of $\boldsymbol{a}_p$ the linear equation system $\boldsymbol{B}\boldsymbol{a}_p = \boldsymbol{c}$ which can be solved by standard methods. In detail, we construct the coefficient matrix $\boldsymbol{B}$ and the constant vector $\boldsymbol{c}$ in the following way. For the diagonal elements of $\boldsymbol{A}$ we get

$$\frac{\partial S}{\partial a^{ll}} = -2 \sum_{i=1}^{N} (w^{il} - \langle\, p_i, a^l \,\rangle) p^{il} \stackrel{!}{=} 0$$

for $l = 1, \dots, d$. This is equivalent to

$$\begin{aligned} \sum_{i=1}^{N} w^{il} p^{il} \qquad &= \sum_{i=1}^{N} \langle\, p_i, a^l \,\rangle\, p^{il} \\ &= \sum_{i=1}^{N} \sum_{j=1}^{d} p^{ij} a^{jl} p^{il} = \sum_{j=1}^{d} \sum_{i=1}^{N} p^{ij} p^{il} a^{jl}, \end{aligned}$$

what gives

$$\langle\, w^l, p^l \,\rangle = \sum_{j=1}^{d} \langle\, p^j, p^l \,\rangle\, a^{jl}. \tag{27}$$

For the off–diagonal elements we obtain

$$\frac{\partial S}{\partial a^{kl}} = -2 \sum_{i=1}^{N} (w^{il} - \langle\, p_i, a^l \,\rangle) p^{ik} - 2 \sum_{i=1}^{N} (w^{ik} - \langle\, p_i, a^k \,\rangle) p^{il} \stackrel{!}{=} 0$$

for $k, l = 1, \dots, d$ and $k > l$. This is equivalent to

$$\langle\, w^l, p^k \,\rangle + \langle\, w^k, p^l \,\rangle = \sum_{j=1}^{d} \langle\, p^j, p^k \,\rangle\, a^{jl} + \sum_{j=1}^{d} \langle\, p^j, p^l \,\rangle\, a^{jk}. \tag{28}$$

The t–th entry in $\boldsymbol{c}$ we get now by the left hand side of (27), when $\boldsymbol{a}_p(t)$ corresponds to a diagonal element of $\boldsymbol{A}$, or by the left hand side of (28) otherwise. Note that $t = k + (2d - l)(l - 1)/2$ for $k \geq l$ and $t = 1, \ldots, (d^2 + d)/2$. The t–th row of $\boldsymbol{B}$ we obtain by the right hand side of (27) or (28), where obviously the factors of the a^{jk} are our searched coefficients. So an explicit solution is available such that no iteration is necessary for estimating $\mathbf{A}$ in the partial linear model case. This is exactly the reason for both, the much smaller variance of the resulting estimates in practice and the much higher speed of the algorithm for estimating model (1) compared to the one for estimating the varying coefficients model (3).

Example For the simple case $d = 3$ we get with $\boldsymbol{a}_p = (a^{11}, a^{21}, a^{31}, a^{22}, a^{32}, a^{33})$

$$\begin{pmatrix} \langle p^1,p^1\rangle & \langle p^2,p^1\rangle & \langle p^3,p^1\rangle & 0 & 0 & 0 \\ \langle p^1,p^2\rangle & \langle p^2,p^2\rangle+\langle p^1,p^1\rangle & \langle p^3,p^2\rangle & \langle p^2,p^1\rangle & \langle p^3,p^1\rangle & 0 \\ \langle p^1,p^3\rangle & \langle p^2,p^3\rangle & \langle p^3,p^3\rangle+\langle p^1,p^1\rangle & 0 & \langle p^2,p^1\rangle & \langle p^3,p^1\rangle \\ 0 & \langle p^1,p^2\rangle & 0 & \langle p^2,p^2\rangle & \langle p^3,p^2\rangle & 0 \\ 0 & \langle p^1,p^3\rangle & \langle p^1,p^2\rangle & \langle p^2,p^3\rangle & \langle p^3,p^3\rangle+\langle p^2,p^2\rangle & \langle p^3,p^2\rangle \\ 0 & 0 & \langle p^1,p^3\rangle & 0 & \langle p^2,p^3\rangle & \langle p^3,p^3\rangle \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{pmatrix}$$

$$= \begin{pmatrix} \langle w^1,p^1\rangle \\ \langle w^1,p^2\rangle+\langle w^2,p^1\rangle \\ \langle w^1,p^3\rangle+\langle w^3,p^1\rangle \\ \langle w^2,p^2\rangle \\ \langle w^2,p^3\rangle+\langle w^3,p^2\rangle \\ \langle w^3,p^3\rangle \end{pmatrix}.$$

References

[1] J. Banks, R. Blundell, A. Lewbel (1997) Quadratic Engel Curves and Consumer Demand. *Review of Economics and Statistics,* 79, 527–539.

[2] R. Blundell, X. Chen, D. Kristensen (2003) Semi–Nonparametric IV Estimation of Shape–Invariant Engel Curves. *Centre for Microdata Methods and Practice, working paper CWP15/03.*

[3] R. Blundell, A. Duncan, K. Pendakur (1998) Semiparametric Estimation of Consumer Demand. *Journal of Applied Econometrics,* 13, 435–461.

[4] A. Deaton, J. Muellbauer (1980) An almost ideal demand system. *American Economic Review,* 70, 312–326.

[5] B.R. Haag, S. Hoderlein, K. Pendakur (2006) Testing and Imposing Slutsky Symmetry in Nonparametric Demand Systems. *Working Paper Simon Fraser University.*

[6] W. Härdle, J.S. Marron (1991) Bootstrap Simultaneous Error Bars For Nonparametric Regression. *The Annals of Statistics,* 19, 778-796.

[7] W. Härdle, M. Müller, S. Sperlich, A. Werwatz (2004) *Nonparametric and Semiparametric Models,* Springer Verlag, Berlin.

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

[9] D. Jorgensen, L. Lau, T. Stoker (1980) Welfare comparison and exact aggregation. *American Economic Review,* 70, 268–272.

[10] A. Lewbel (2000) Consumer Demand Systems and Household Equivalence Scales, Handbook of Applied Econometrics, Volume II: Microeconomics. Blackwell Publishers Ltd, Oxford.

[11] A. Lewbel, K. Pendakur (2008) Tricks with Hicks: The EASI Demand System, *American Economic Review*, forthcoming.

[12] K. Pendakur, S. Sperlich (2008) Semiparametric Estimation of Consumer Demand Systems in Real Expenditure. *Journal of Applied Econometrics*, forthcoming