RECENT APPROACHES TO ESTIMATING INCOME DISTRIBUTIONS, ENGEL CURVES AND RELATED FUNCTIONS

Joachim Engel
Wirtschaftstheorie II
Lennestr. 37
53225 Bonn
Germany

Alois Kneip
Institut des Statistique
34 Voie du Roman Pays
1348 Louvain–La–Neuve
Belgium

Key words and phrases: income distributions; cross–sectional Engel curves; non-parametric curve estimation; demand analysis

Subject Classification: Primary C14, C21, Secondary D12, D31

ABSTRACT

The paper surveys recent methods of nonparametric curve estimation in their relevance to some problems in econometrics: the estimation of income distributions, cross–sectional Engel curves and other functions which depend on the demand in large economies. Instead of a comprehensive review of the literature, only aspects of nonparametric estimation are considered as they relate to the applications we have in mind. In density estimation the concept of estimating skewed and heavy–tailed distributions is discussed. Estimating demand curves necessitates in particular to address nonparametric regression under a random design when the data are heteroscedastic.

1 Introduction

Graphical representation of curves is a very efficient way of summarizing large data sets, for drawing inferences about a population from a sample or for detecting structure in data. Assuming a functional relationship describing some structural feature of the data, the objective is to obtain an estimate for the underlying function from noisy observations. Two important examples are estimation of a probability density and of a regression function. If the data consist of random observations $X_i \in \mathbb{R}^d, i = 1, \ldots, n$ for some integer d, density estimation aims at making inferences about the underlying density f that determines the distribution of the X_i, i.e. $P(X_i \in [a,b]) = \int_a^b f(t)\mathrm{d}t, a < b$ for $d = 1$. In contrast, regression estimation requires paired data $(X_i, Y_i), X_i \in \mathbb{R}^d, Y_i \in \mathbb{R}$ and assumes a functional dependence of the form $Y = m(X)$ between the dependent variable Y and the regressor variable X. Then the objective is to estimate the function m which expresses the dependence between X and Y.

An important application of density estimation in economics is estimating income distributions. The estimated income density contains important information on the dispersion of financial resources. The shape of the income distribution (kurtosis, skewness, number of modes) reveals important economic information. While income itself being a highly relevant quantity for many economic studies, economists also take a strong interest in how households spend their earned incomes. Considering expenditures for a certain good as a variable, say Y, depending on household income X, exploring the relationship between X and Y has a long tradition in economic theory (Ernst Engel, 1895; Working, 1943). The function expressing a household's expenditures for a given good as a function of household income (or total expenditure) is known as cross-sectional Engel curve. While this paper is tailored particularly towards these two examples (income density and Engel curve estimation), the concepts presented here are much more general and apply to estimating virtually any probability density and regression function. For a general introduction into the field of nonparametric curve estimation we refer to the recent monographs of Silverman (1985) and Scott (1992) on density estimation and of Eubank (1988), Müller (1988) and Härdle (1990) on regression estimation. The main ideas can also be applied to estimating spectral densities of time series, hazard rate functions, intensity functions of counting processes or any functional depending on one of the above functions. For a further application of nonparametric curve estimation to demand theory, see Härdle, Hildenbrand and Jerison (1991), Hildenbrand and Kneip (1993) and Hildenbrand (1994) who give empirical evidence for the Law of Demand by using nonparametric curve estimation techniques developed by Härdle and Stoker (1989).

The classical approach to curve estimation is based on a parametric model, i.e. $f(x) = f(x, \theta)$ with a known function f and unknown parameter θ. The functional shape of the estimate is already given through assumptions. The quality of the resulting estimator relies heavily on the correctness of this specification.

If the parametric model is correct, good estimators of the parameters and of the curve can usually be obtained. However, if the parametric model is misspecified, the resulting curve estimator is not even consistent. In most applications there is no reliable method providing the statistician with a correct parametric model. Non–parametric estimators are much more flexible because they allow the data themselves to determine the shape of the estimate. The assumption of a particular parametric functional form is dropped and replaced by a much more general *smoothness* assumption postulating differentiability of a certain degree for the function to be estimated. As an example comparing parametric with non–parametric estimators consider the problem of estimating income densities and their evolution over time. Using income data from the British Family Expenditure Survey (FES) between 1968 and 1984 the sample size is about 7000 households per year. Details of the survey are described in the Family Expenditure Survey Handbook by Kemsley, Redpath and Holmes (1980). The traditional parametric model for estimating income densities is based on a lognormal class of densities. The resulting estimate is graphed in the left frame of figure 1 (net income has been divided by mean net income to study relative effects only). Among the parametric estimates one can hardly notice any change at all over the years. This contrasts starkly with the nonparametric estimates in the center of figure 1. There is considerable variation in the relative income densities. Another striking feature of the nonparametric estimates is that they are bimodal – a property ruled out by assumptions in the parametric estimates.

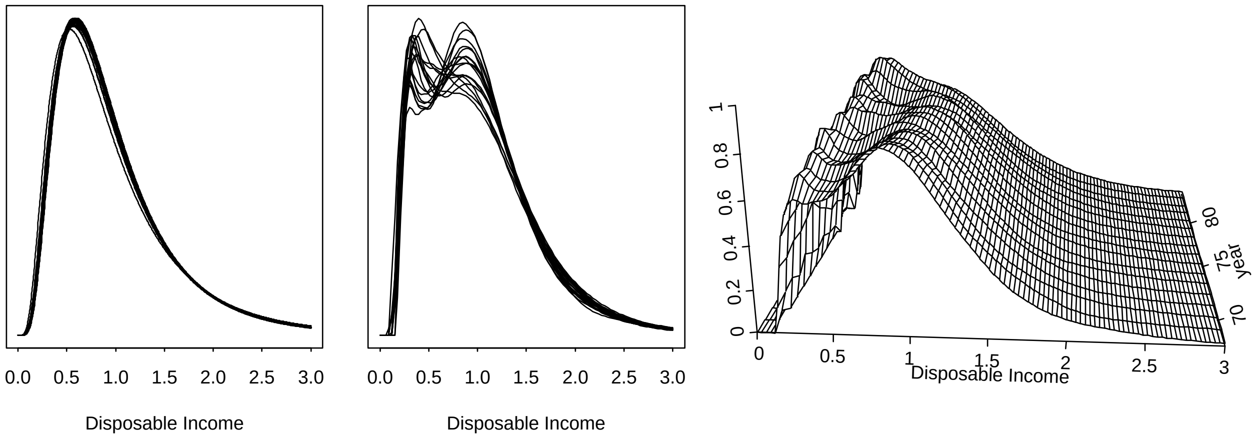

Figure 1: Parametric Lognormal fit (left) versus nonparametric kernel estimate (center and right) of net income density from FES, 1968–1984.

Even more informative is the three–dimensional plot in the right frame of figure 1 which shows nonparametric income density estimates with time as third axis. Here we notice a slow change over the years putting more and more mass from the central mode to a mode located at about .3 times mean income. All this structurally interesting information is masked in the parametric approach.

2 Income Densities

Histogram

Given independent observations $X_1, \ldots, X_n \in \mathbb{R}$ stemming from an unknown distribution, the objective of density estimation is to obtain estimates $\hat{f}$ for the underlying density f. For simplicity of presentation we restrict ourselves to the case of $d = 1$ dimension. The oldest density estimator is the histogram. Given an origin x_0 and a bin width h the bins of a histogram are defined as intervals $[x_0 + mh, x_0 + (m+1)h)$ for positive and negative integers m. Then the equal bin width histogram is defined as the function assigning to a point x the relative number of observations falling into the bin x belongs to:

$$\hat{f}(x) = \frac{1}{nh} \#\{i \mid X_i \text{ is in the same bin as } x \}. \tag{2.1}$$

Hence, to construct a histogram requires specification of an origin x_0 and a bin size h. The bin size h determines the smoothness of the estimate. As illustration consider an estimate of the net income density of British households for the year 1973, based on a sample of size $n = 7203$. Figure 2 shows four different histograms with the same origin but different bin sizes. While for a small bin width the estimate has many modes, we have about two modes for a moderate bin size. Finally, with increasing bin width the estimate becomes unimodal.

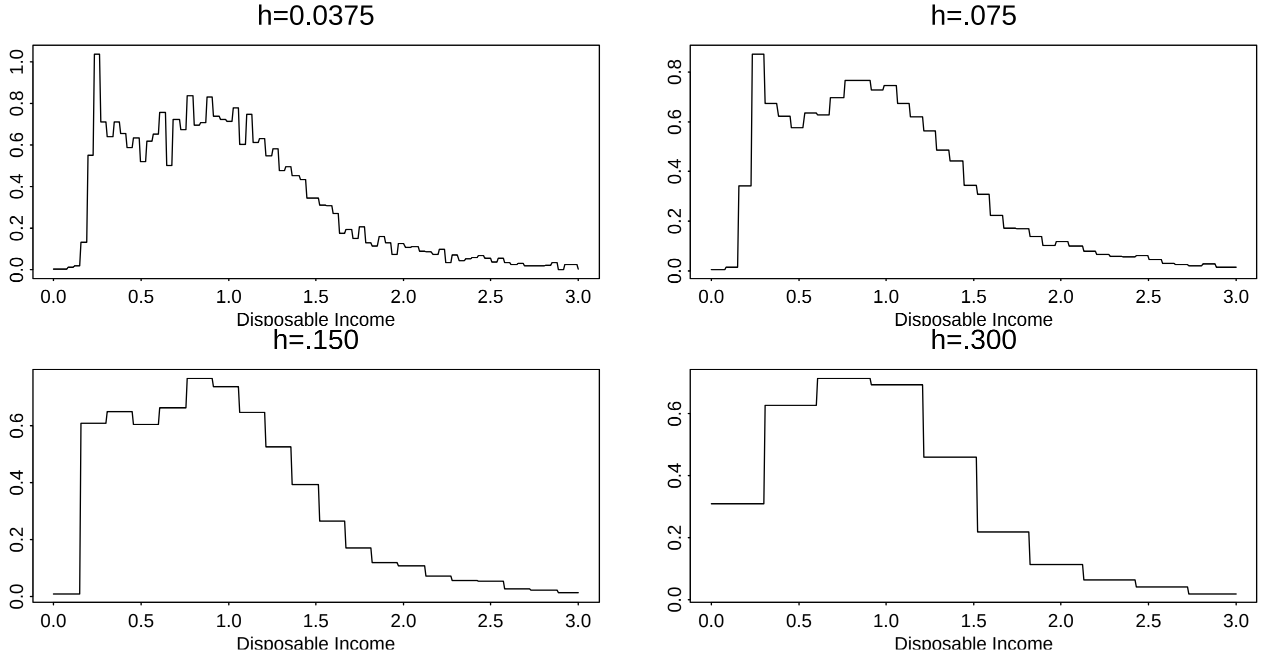

Figure 2: Four different histograms for 1973 FES net income data with respective bin size $h = .0375, h = .075, h = .15, h = .3$

Despite its popularity, the histogram is quite unsatisfactory because of its discontinuities of its dependence on the origin. Figure 3 shows five different histograms as estimates for the 1973 British household income distribution with the same bin size but with different origins. Each successive estimate has an origin shifted by 1/5 of a bin. Eventhough the general message is the same, a closer look at, say, the width of the first mode or the symmetry of the second mode might lead an inexperienced user to different conclusions, depending on how the origin was chosen.

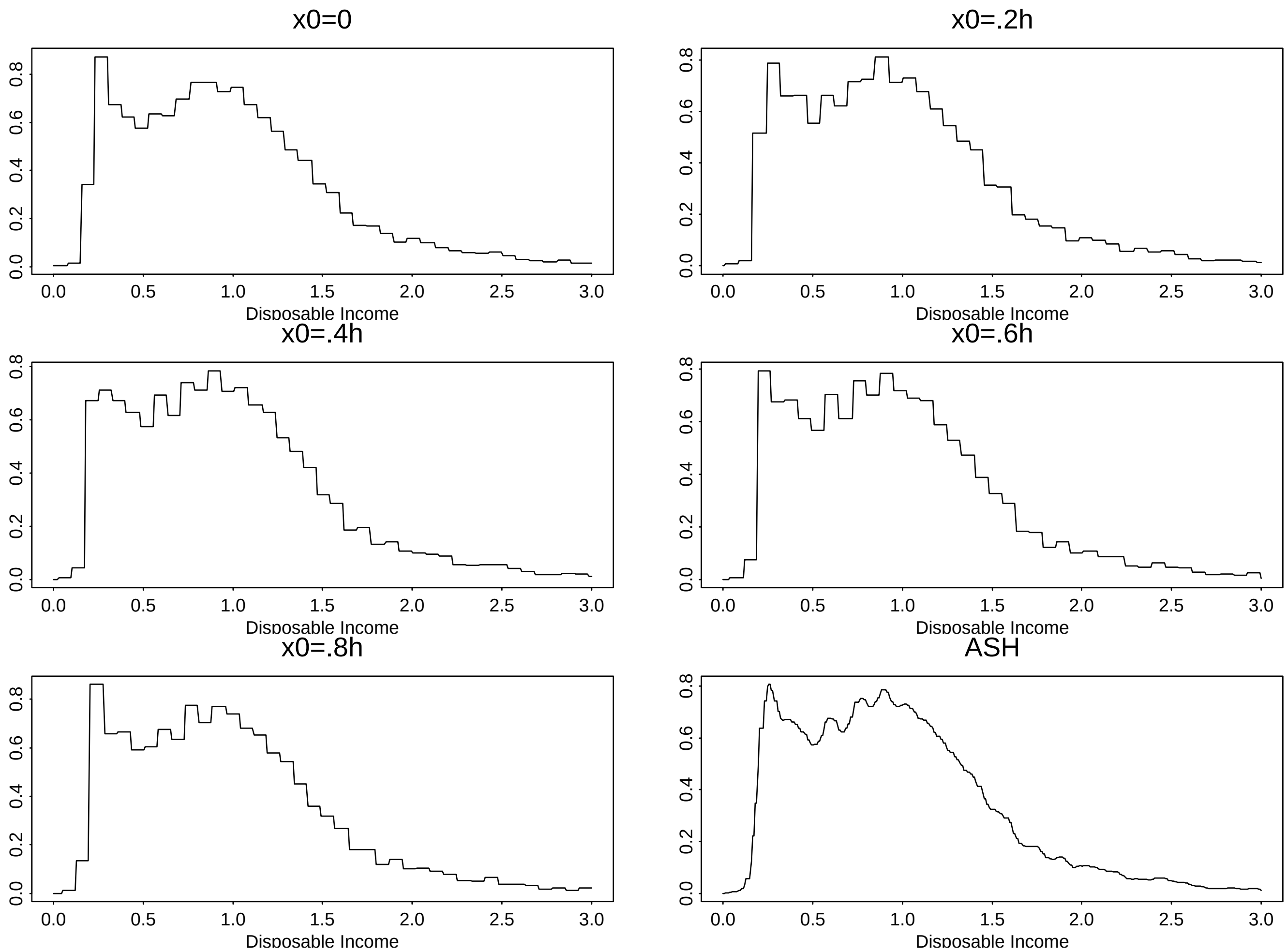

Figure 3: Five different histograms for 1973 FES net income data with bin size $h = .075$ and origin at $x_0 = .0, .2h, .4h, .6h, .8h$. On the bottom left, the average shifted histogram of the five preceding histograms.

Under some technical assumptions regarding the density f bias and variance of the histogram are calculated as follows (see Scott, 1992, pg 51 ff):

$$\text{Bias}\left(\hat{f}(x)\right) = \left(\frac{h}{2} - t(x)\right) f'(x) + O(h^2), \text{Var}\left(\hat{f}(x)\right) = \frac{f(x)}{nh} + O(n^{-1})$$

provided $h \to 0$ and $nh \to \infty$. Here $t(x)$ denotes the distance between x and the left boundary of the bin containing x.

A convenient measure of discrepancy in curve estimation expressing the distance between the true density f and its estimate $\hat{f}$ is the mean integrated squared error (MISE), defined by MISE $= \mathrm{E}\int_{-\infty}^{\infty}\left(\hat{f}(x)-f(x)\right)^2 \mathrm{d}x$. Notice that $\hat{f}$ depending on the random observations $X_1,\ldots,X_n$ is a random quantity itself. Hence MISE, being an expected value of some expression involving $\hat{f}$ represents an average over all possible realizations of the n random observations. Finally, integration is involved in MISE as a global measure, evaluating the quality of estimation over the whole real line instead of estimation at a single point.

MISE can be decomposed as sum of the integrated squared bias and the integrated variance. For the histogram MISE allows representation as

$$\mathrm{MISE}(h)=\frac{1}{nh}+\frac{1}{12}h^2\int_{-\infty}^{\infty} f'(x)^2\mathrm{d}x+O(n^{-1}+h^3).$$

An asymptotically optimal bin width h_{asy} then is calculated as $h_{\text{asy}}=\left(6/n\int_{-\infty}^{\infty} f'(x)^2\mathrm{d}x\right)^{1/3}$ resulting in an asymptotically optimal MISE of the order $\mathrm{MISE}(h_{\text{asy}})=O(n^{-2/3})$.

An idea to free the histogram from its dependence on the origin is the average shifted histogram (ASH) due to Scott (1985). The ASH is defined as average over several, say L, histograms with equal bin width but different origins, shifted by a multiple of $1/L$. If L is sufficiently large, an improvement in the rate of convergence is obtained up to the order $O(n^{-4/5})$. An ASH for the 1973 British income data based on the average of five shifted histograms is shown in the lower right frame of figure 3.

Kernel Density Estimator

Another approach to avoid the problem of choosing the origin in the histogram is the 'moving histogram' as introduced by Rosenblatt(1956). To estimate a density at a given point x the histogram starts with fixed bins and counts the relative number of observations falling into the same bin as point x. In contrast, the moving histogram considers a *moving* bin centered around x:

$$\begin{aligned}\hat{f}(x) &= \frac{1}{2nh}\#\{i \mid x-h<X_i<x+h\}=\frac{1}{2nh}\sum_{i=1}^{n} I_{(x-h,x+h)}(X_i) \qquad (2.2)\\ &\approx \frac{1}{2h}P(x-h<X_i<x+h)\to f(x) \text{as } n\to\infty\ .\end{aligned}$$

While the naive estimator does not dependent on an origin anymore, this estimator is unsatisfactory on other grounds: $\hat{f}$ has $2n$ discontinuities, located at $X_i \pm h$ and zero derivatives everywhere else. A straight generalization avoids these difficulties and leads to the kernel estimator (Parzen, 1962):

$$\hat{f}(x)=\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right). \qquad (2.3)$$

For the particular choice of the rectangular kernel function $K(u) = I_{[-1/2,1/2]}(u)$ we obtain the naive estimator. Just as the naive estimator (2.2) can be considered as a sum of boxes the kernel estimator (2.3) is a sum of 'bumps' centered at the observations. The choice of the kernel function $K(u)$ determines the shape of these bumps, and the scaling factor h, called the *bandwidth*, their width. A small bandwidth leads to a wiggly estimate with several modes, while a moderate bandwidth results in a bimodal estimate. Finally, a large bandwidth smears the mass across the sample space yielding a rather flat unimodal estimate.This is illustrated by Figure 4 which shows kernel estimates based on the same 6 data points, but with differing bandwidths.

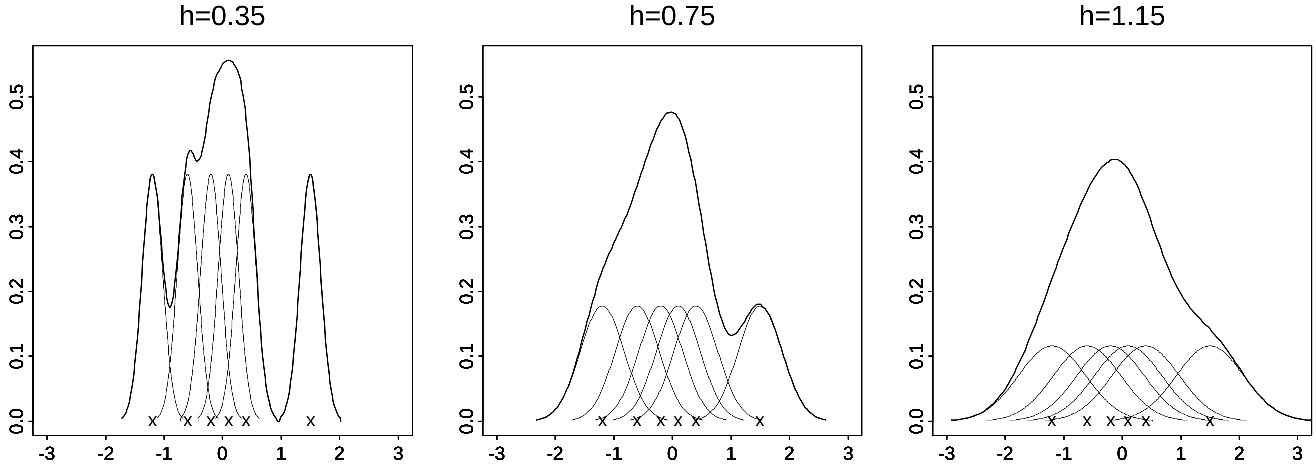

Figure 4: Kernel density estimate based on 6 data points, with bandwidth $h = .35$(left), $h = .75$(center) and $h = 1.15$ (right) and Gaussian kernel; data points are marked by **x**.

The kernel density estimate $\hat{f}$ inherits all analytic properties from the kernel K, i.e. $\hat{f}$ is exactly as many times differentiable as the kernel function. If the kernel function is a probability density, i.e. nonnegative and integrates to 1, then the same holds for $\hat{f}$. Usually one postulates the estimator of a density to be a density itself. When the objective is to estimate a derivative of a density instead, the use of differentiable kernels is advisable: assuming that K is ν–times differentiable, we estimate the ν–th derivative of a density by the ν–th derivative of the estimate of f, i.e. $\hat{f}^{(\nu)}(x) = 1/(nh^{\nu+1})\sum_{i=1}^{n} K^{(\nu)}((x - X_i)/h)$. Therefore, K is usually chosen as a smooth symmetric probability density. However, in some situations, in particular for very large sample sizes, it may be advantageous to apply so–called higher order kernels which assume also negative values (see Gasser, Müller and Mammitzsch, 1985). Figure 5 shows kernel density estimates with three different bandwidths for the British 1973 Household Income data and the shifted power–transformed kernel estimator to be introduced below.

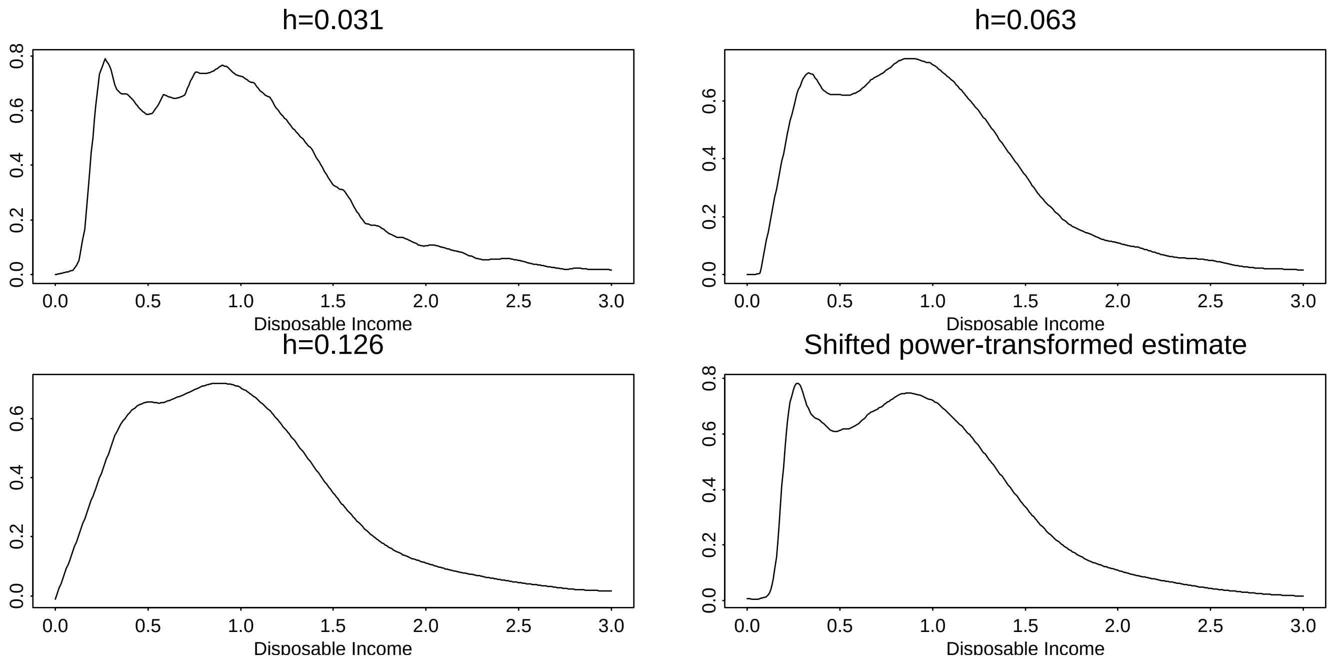

Figure 5: Three kernel density estimates of the 1973 FES net income data with respective bandwidth $h = 0.031, h = 0.063$ and $h = 0.126$. On the lower right the shifted power–transformed kernel density estimate.

Selection of a kernel

Under the assumption that f is twice differentiable and that the kernel K has a vanishing first moment and a second moment unequal zero ($\int xK(x) = 0, \int x^2 K(x) \neq 0$), bias and variance of the kernel density estimator (2.3) at x are given by (Parzen, 1962)

$$\mathrm{Bias}(\hat{f}(x)) = \frac{1}{2}h^2\sigma_K^2 f''(x) + o(h^2), \mathrm{Var}(\hat{f}(x)) = \frac{1}{nh}f(x)R(K) + o([nh]^{-1}) \quad (2.4)$$

as $h \to 0$ and $nh \to \infty$. Here $\sigma_K^2 = \int K(u)u^2 du$ and $R(K) = \int K(u)^2 du$. For the mean integrated squared error we obtain

$$\mathrm{MISE}(h, K) = \left[h^4/4\sigma_K^2 \int_{-\infty}^{\infty} f''(x)^2 \mathrm{d}x + \frac{R(K)}{nh}\right](1 + o(1)). \quad (2.5)$$

Relation (2.4) allows quantification of the effect of varying the bandwidth. A small bandwidth h leads to a small bias but a large variance. A large bandwidth results in small variance but a large bias. Thus there is a trade–off between bias and variance, and a reasonable bandwidth h has to be some sort of compromise. Minimizing (2.5) for h results in a formula for the asymptotically optimal bandwidth:

$$h_{\mathrm{asy}} = \left(\frac{R(K)}{\sigma_K^2 R(f'')}\right)^{1/5} n^{-1/5}, \quad (2.6)$$

where $R(f'') = \int_{-\infty}^{\infty} f''(x)^2 \mathrm{d}x$. Plugging the asymptotically optimal bandwidth h_{asy} back into formula (2.5) leads to

$$\begin{aligned} \mathrm{MISE}(h_{\mathrm{asy}}, K) &= 5/4 C(K)^{2/5} \left(R(f'')\right)^{1/5} n^{-4/5}(1 + o(1)) \\ \text{where} \quad C(K) &= \sigma_K^2 R(K)^2. \end{aligned} \tag{2.7}$$

Hence, MISE is asymptotically minimal for a kernel K minimizing $C(K)$. Subject to the constraint that K is nonnegative and integrates to 1, this leads to a minimization problem which is solved by the *Epanechnikov kernel*

$$K_{\mathrm{e}}(t) = \begin{cases} 3/5(1 - t^2) & \mid t \mid \leq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.8}$$

However, the Epanechnikov kernel is not differentiable at ± 1 and hence not suitable for estimating derivatives. As it will turn out (see below), estimating the second derivative f'' will be of particular importance for data–driven selection of the bandwidth h. Additionally requiring twice differentiability of K in the above minimization problem leads to the *quartic kernel* defined by

$$K_q(t) = \begin{cases} 15/16(1 - t^2)^2 & \text{if } \mid t \mid \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Defining the efficiency of an arbitrary kernel K by $\mathrm{eff}(K) = C(K_{\mathrm{e}})/C(K)$ allows evaluation of any kernel. The Gaussian kernel (K equals normal density) has an efficiency of 95%, the quartic kernel of 99.39% and even the unappealing rectangular kernel that leads to the naive estimator has an efficiency of 92.95 %. For details see Silverman (1985). Hence the asymptotic MISE is quite insensitive to the selection of a kernel function. Despite their slight suboptimality the quartic and the Gaussian kernel are the preferred choice for many researchers because they result in a twice respectively infinitely differentiable estimator $\hat{f}$.

Data–driven bandwidth selection

The ease of finding a solution for appropriate kernels contrasts with the problem of suitable bandwidth selection. While the optimal kernel is determined independently of density f and sample size n, the opposite is true for the optimal bandwidth, as inspection of formula (2.6) shows. Furthermore, given a fixed sample size, if the bandwidth approaches 0, the estimator $\hat{f}$ converges to a sum of delta functions, i.e. $\hat{f}$ equals zero almost everywhere except for sharp peaks located at the observations. On the other hand, with the bandwidth growing to infinity, the estimator converges to the uniform distribution. An efficient data–driven bandwith selector is crucial to all curve estimation problems. A comprehensive survey of the current state of the art is given by Jones, Marron and Park (1992). Two different approaches have been pursued. Methods summarized under the name *cross–validation* try to obtain an estimate of $\mathrm{ISE}(h) = \int (\hat{f} - f)^2$

and then search for a bandwidth h_{CV} minimizing this estimate of the integrated squared error. While cross–validation methods attracted a lot of attention at first, they suffer from a high sample variability, as has been confirmed by mathematical analysis of their asymptotic behavior as well as by empirical finite sample studies.

An alternative to cross–validation is the so–called *plug–in* approach. The starting point is formula (2.6) which states that the asymptotically optimal bandwidth depends on $R(f'')$, a functional of the unknown density f. *Plug–in* methods replace $R(f'')$ in (2.6) by an estimate. Among several variations of this idea the version proposed by Sheather and Jones (1991) is considered by most experts as top–ranking in finite sample performance. An iterative version of this method by Engel, Herrmann and Gasser (1994) has proven to be equivalent, in analytic analysis as well as in simulation studies. We present here the iterative method because of its closeness to a bandwidth selector by Gasser, Kneip and Köhler (1991) in the related field of regression estimation (see section 3). Recognizing that estimating the functional $R(f'') = \int f''(x)\mathrm{d}x$ is a different smoothing problem than estimating the density f itself, we allow when estimating this functional for a different bandwidth α and a different four times differentiable kernel L . An asymptotically optimal value for α is calculated as (Sheather and Jones, 1991)

$$\alpha_{\mathrm{asy}} = \left(\frac{2L^{iv}(0)}{\sigma_L^2 R(f''')n}\right)^{1/7}. \tag{2.9}$$

As described below, the iterative plug–in bandwidth rule h_{PI} is obtained by starting with an initial value for α which is obtained through a scale model referring to the normal distribution and calculating a bandwidth h following formula (2.6). Then an improved value for α is obtained by multiplying h with the factor $cn^{2/35}$ which is derived from asymptotic considerations before iterating three times.

Step 1:

$$h^{(0)} = \left(\frac{R(K)}{\sigma_K^2 \hat{R}(\alpha^{(0)})}\right)^{1/5} n^{-1/5}$$

where $\alpha^{(0)} = 1.06\hat{\sigma}n^{-1/5}$, $\hat{\sigma}^2 = 1/(n-1)\sum_{i=1}^n (x_i - \bar{x})^2$, $\bar{x} = 1/n\sum_{i=1}^n x_i$ and $\hat{R}(\alpha)$ denotes an estimate for $R(f'')$ defined by

$$\hat{R}(\alpha) = \int f''_\alpha(t)^2 dt = \int \left(\frac{1}{n\alpha^3}\sum_{i=1}^n L''\left(\frac{t - X_i}{\alpha}\right)\right)^2 \mathrm{d}t$$

Step 2: For $i = 1, 2, 3$ set

$$\begin{aligned} \alpha^{(i)} &= ch^{(i)} n^{2/35} \\ h^{(i)} &= \left(\frac{R(K)}{\sigma_K^2 \hat{R}(\alpha^{(i)})}\right)^{1/5} n^{-1/5}. \end{aligned}$$

Here the constant c is obtained from the relation $\alpha_{\text{asy}} = cn^{2/35}h_{\text{asy}}$ after replacing the unknown functionals involved in h_{asy} and α_{asy} by nonparametric estimates which in turn are obtained through a normal scale model.

Step 3: Finally, set $h_{\text{PI}} = h^{(3)}$.

Shifted power–transformed kernel estimator

When applied to the 1973 British household data, the above algorithm yields a bandwidth of $h_{\text{PI}} = 0.0631$ resulting in the estimate of the income distribution in the upper right frame of figure 5. The resulting bimodal smooth estimate is quite satisfactory, inviting discussion about the nature of the two modes. A closer look, however, reveals some wiggles in the upper tail of the estimate that are hard to explain by socioeconomic arguments. Do they represent *real* structure in the data or are they artefacts induced by the kernel method? Inspection of density estimates with increased bandwidth in the left lower frame of figure 5 reveals that under a larger bandwidth these wiggles disappear. They are smoothed away. The price paid, however, is that the bimodal structure also disappears. Wiggles in the tails are an immediate consequence of applying the same amount of smoothing throughout the real line when estimating long–tailed or skewed densities. Instead of using a fixed bandwidth, had we been allowed to apply varying amounts of smoothing, say a variable scaling factor of $h = h_i$ applied at observation X_i, then spurious noise in the tails of the estimate could have been avoided: just apply more smoothing in the tails and less smoothing at the center of the distribution. Mathematical analysis shows that in particular skewed and kurtotic densities call for variable smoothing. Various methods for variable smoothing parameter selection have been proposed for kernel density estimation (see Silverman, 1985). A flexible and practical approach is based on transformations: inspection of formula (2.7) tells us that mean integrated squared error is smallest for a density which minimizes $R(f'') = \int f''(x)^2 \mathrm{d}x$. It can be shown that the normal density approximately minimizes the function $R(f'')$. Therefore, Wand, Marron and Ruppert (1991) propose to first transform the data such that the transformed data are close to normal and hence suitable for a fixed bandwidth kernel density estimator. Then the kernel density estimator of the transformed data is back–transformed to obtain the final estimate. Wand et al. propose the parametric class of shifted power tranformations of the form

$$Z_i := T(X_i) = \begin{cases} (\alpha + X_i)^\beta \text{sign}(\beta) & \text{for } \alpha > 0, \beta < 1 \\ \log(X_i + \alpha) & \text{for } \beta = 0. \end{cases}$$

Finally, the back–transform estimate for the density f becomes

$$\hat{f}(x) = \frac{\beta(\alpha + x)^{\beta - 1}}{nh} \sum_{i=1}^{n} K\left(\frac{(\alpha + x)^\beta - (\alpha + X_i)^\beta)}{h}\right). \tag{2.10}$$

The mean integrated squared error of the density estimate of the transformed data Z_i can be expressed as a function $\mathrm{MISE}(\alpha, \beta)$ of the two transformations parameters α and β. We used the transformation based kernel density estimator for the British income data for the years 1968 to 1984. It turned out, that throughout the years $\mathrm{MISE}(\alpha, \beta)$ assumed its minimum for values around $\alpha = .1$ and $\beta = .25$. Therefore, we chose the transformation $Z_i = (X_i + .1)^{.25}$. When applying the quartic kernel and choosing the bandwidth according to the iterative plug–in method described above, the resulting density estimate for the 1973 British income data is shown in the lower right frame of figure 5. Now the wiggles in the right tail have disappeared while the bimodal structure has been distinctly preserved.

The center and right frame of Figure 1 show nonparametric estimates based on the shifted power–transformed kernel estimator of the income density in Great Britain from 1968 to 1984. The striking bimodal structure asks for further discussion of its socio–economic origin. Here also non–parametric density estimation can provide an important contribution. Figure 6 shows density estimates separately for those household whose head is retired (left frame) and whose head is not retired (right frame). Now the bimodal structure has disappeared in both plots. Furthermore, the percentage of households included in the FES–survey whose head is retired has increased from 26 %(1968) to 39 %(1984). Hence

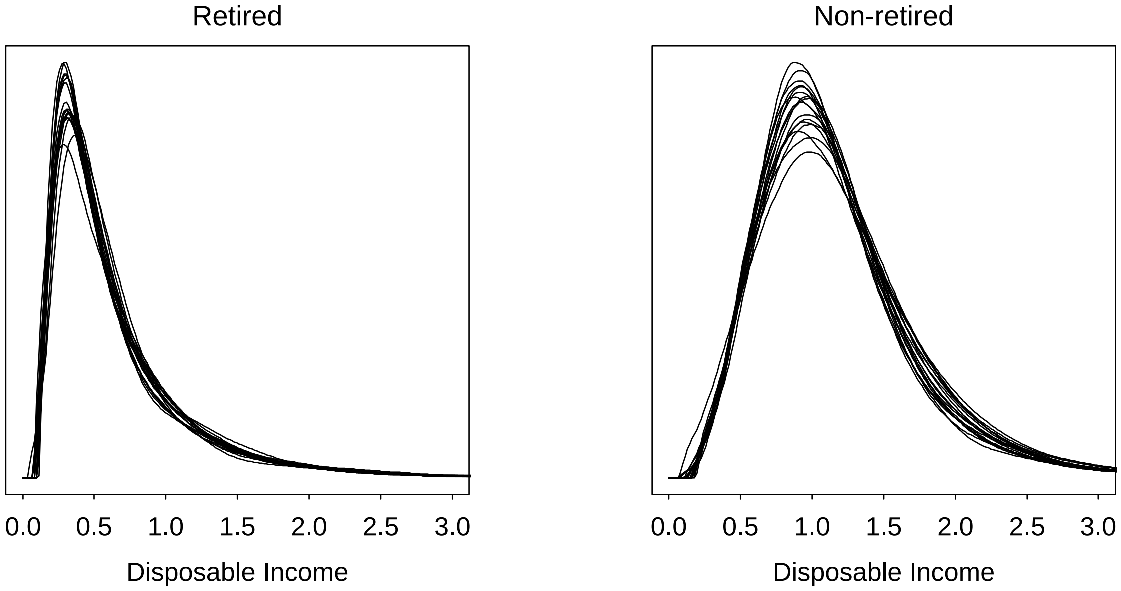

Figure 6: Income density from FES, 1968-1984, households with head retired (left), households with head not retired (right).

the evolution of the sharp peak of the income density at around .3 times mean income may well be caused by demographic (more retired people) and socioeconomic (more retired persons living in their own household) causes.

3 Engel Curves

Cross–sectional Engel curves are functions expressing the demand for a particular good in a large economy as a function of disposable income or total expenditure. They provide important information about cross–sectional expenditure data. For example, they can be used to predict which income classes are most likely to benefit from changes in the tax system. The first published empirical study of demand as a function of the available household budget was done by Ernst Engel (1895) in his 1857 study on consumption behavior of Saxonian families.

The efforts of most modern economies in conducting large household expenditure surveys and the development of new statistical techniques for curve estimation in recent years provided the tools necessary for the new interest of economic theory in empirical approaches to modern demand theory. Härdle and Jerison (1988) apply nonparametric kernel regression to investigate if cross–sectional Engel curves estimated from the British FES data change over time. Bierens and Pott–Buter (1990) apply a multivariate kernel regression estimation to Dutch household data including a subgroup analysis according to number and age of children. While the authors apply one particular type of kernel estimator (*Nadaraya–Watson estimator*) and choose a parameter decisive to all nonparametric methods (*the bandwidth*) by a suboptimal method, the rapid development of mathematical statistics in recent years has produced a multitude of curve estimation methods with refined techniques for bandwidth selection. This section discusses various parametric and nonparametric curve estimation methods and their suitability for estimating Engel curves. Figure 7 shows the data cloud for expenditure on food at home, housing, transportation and leisure from the French EBF 1989 household survey. Details of this survey are described in Enquete Budget de Famille (1979, 1984–85, 1989). Since we are only interested in relative effects, all data were divided by mean disposable income.

Cross–sectional Engel curves are regression functions for the conditional mean demand with disposable income (or total expenditure) as regressor variable. The basic framework for estimating Engel curves is to treat the observations consisting of disposable income and a vector of expenditures for several commodities for each household as a sample $(\{(X_i, Y_i)\}_{i=1,\dots,n}$ of random observations fulfilling

$$Y_i = m(X_i) + \epsilon_i, \quad \epsilon_i \text{ independent residuals, } \mathrm{Var}(\epsilon_i \mid X_i = x) = \sigma^2(x). \quad (3.11)$$

Here the regressor X_i and error terms ϵ_j are independent. The regression model can be expressed mathematically through the conditional expectation:

$$m(x) \quad = \quad \mathrm{E}[Y \mid X = x] = \int y f_{Y|X}(y \mid x) \mathrm{d}y = \frac{\int y f_{Y,X}(x, y) \mathrm{d}y}{f_X(x)} \qquad (3.12)$$

where $f_{X,Y}$, $f_{Y|X}$ and f_X denote respectively the joint density of (X, Y), the conditional density of Y given X and the marginal denisty of the random variable X.

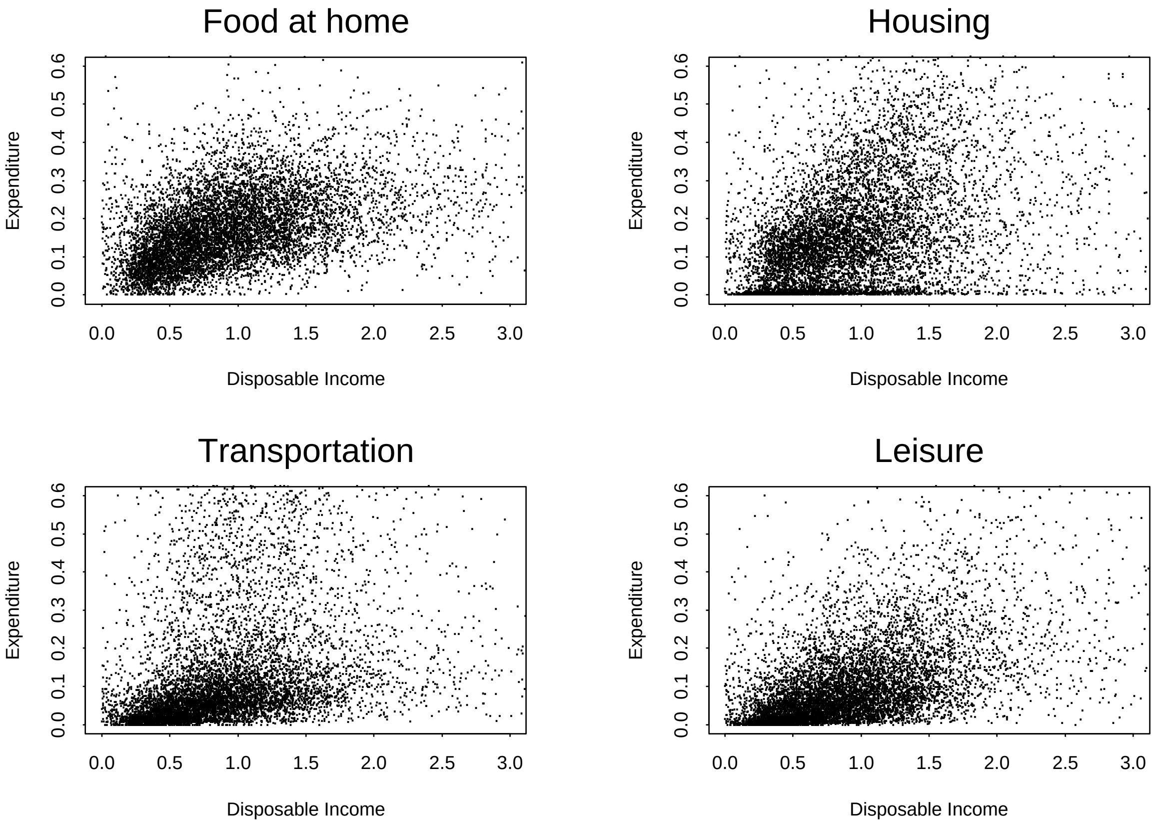

Figure 7: Data cloud of expenditure for food at home, housing, transportation and leisure versus disposable income from 1989 EBF data, $n = 9038$.

Though formally one may introduce Engel curves as functions on the whole positive half line $\mathbb{R}_+$, they are well defined only for $x \in \text{supp}(f_X)$, where $\text{supp}(f_X)$ denotes the support of the income distribution. Furthermore, as we have seen in section 2, the distribution f_X has heavy tails. Therefore, the behavior of Engel curves for extremely large incomes can neither be observed appropriately, nor is it of much interest for practical purposes. Therefore, we restrict the domain of estimation to a subinterval A of $\text{supp}(f_X)$ omitting a small part of the upper and lower tails. The established criterion in nonparametric curve estimation to measure the goodness of an estimator is the mean integrated squared error:

$$\text{MISE} = \text{E} \int_A (\hat{m}(x) - m(x))^2.$$

MISE, depending on the unknown quantity m, is not available in real data applications. However, mathematical theory allows in many situations derivation of an expression for asymptotic MISE.

3.1 Parametric Estimates

The classical approach to curve estimation is based on a parametric model. For estimating Engel curves Working (1943) and Leser (1963) suggested the model

$$m(x,\theta) = \theta_1 x + \theta_2 x \log(x) \tag{3.13}$$

where the parameter $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$ depends on the price system and the time point which both are considered fixed in the context of our static analysis. The authors applied the model (3.13) to food expenditure data and obtained fairly satisfactory fits. The model has been widely used in practice. However, in many other situations the model failed to be accurate and other model specifications were considered. Pollak and Wales (1978) and Howe, Pollak and Wales (1979) introduced the following polynomial model

$$m(x,\theta) = \theta_1 + \theta_2 x + \theta_3 x^2 \quad , \quad (\theta_1, \theta_2, \theta_3) \in \mathbb{R}^3. \tag{3.14}$$

Another interesting approach is due to Deaton (1981), who analyzed Sri Lankan micro household data under a model of the form

$$m(x,\theta) = \theta_1 x + \theta_2 x \log(x) + \theta_3 x (\log(x))^2 \quad , \quad (\theta_1, \theta_2, \theta_3) \in \mathbb{R}^3. \tag{3.15}$$

Kneip (1993) developed a test for the dimension of the space of Engel curves. He further derived a method for estimating the *basic functions* $g_i(x)$ spanning the space of Engel curves, i.e.

$$m(x,\theta) = \sum_{i=1}^{L} \theta_i g_i(x) \quad , \quad \theta = (\theta_1, \ldots, \theta_L) \in \mathbb{R}^L.$$

When applied to the French EBF data Kneip concludes that $L = 4$. Further, in a test developed by Härdle and Kneip(1992) the data conform with the assumption of logarithmic basic functions, i.e.

$$g_i(x) = x(\log(x))^{i-1} \quad , \quad i = 1, \ldots, 4.$$

Hence, following Kneip (1993), a parametric model for the EBF data would have to be at least of polynomial order 3 in $\log(x)$, i.e.

$$m(x,\theta) = \theta_1 x + \theta_2 x \log(x) + \theta_3 x(\log(x))^2 + \theta_4 x(\log(x))^3, (\theta_1, \theta_2, \theta_3, \theta_4) \in \mathbb{R}^4 \tag{3.16}$$

Figure 8 shows Engel curves estimated with the above parametric models for the commodities food at home, housing, transportation and leisure for the 1989 EBF data. The dots represent *binned data* (see below).

Parametric models achieve a rate of convergence for MISE of the order MISE $= O(n^{-1})$ provided the underlying model is correct. If the model is misspecified, the estimators are not even consistent.

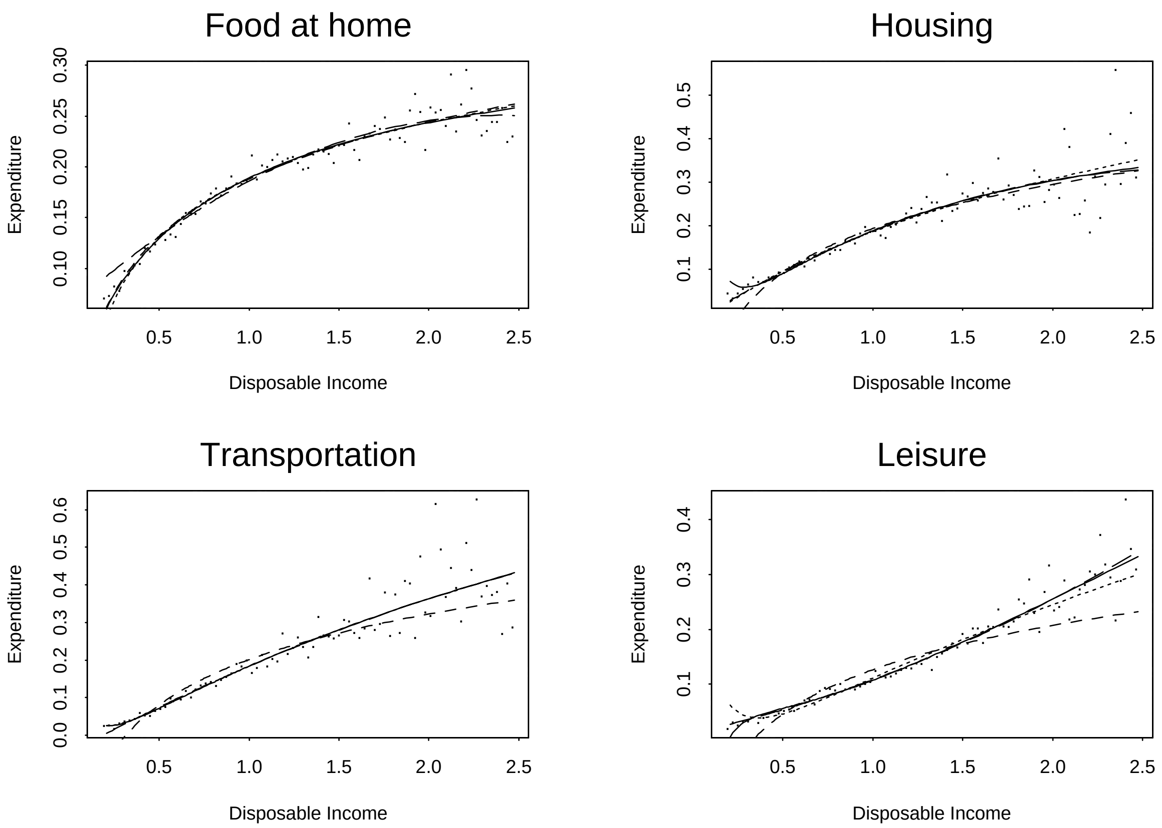

Figure 8: Parametric estimate of Engel Curves: logarithmic polynomial of order 3 (solid line), Deaton model (dotted line), Working model (small dashes) and Pollak & Walsh model (large dashes).

3.2 Regressogram

The problem with parametric regression models is that their derivation may be guided by intuition and experience from the particular field of application, but they rarely have an objective justificaton. The parametric model just may be misspecified. This draw–back calls for nonparametric methods which allow more flexibility. Ernst Engel based his 1857 analysis of Saxonian household data on a simple non–parametric method: the regressogram. It is defined in analogy to the histogram (see section 2): the sample space is divided into fixed bins $B(k) = [b_{k-1}, b_k)$, where $b_k = a + kh, k = 1, 2, \ldots$. Here a denotes the origin and h the bin size. The regressogram is defined by local averages within the bins:

$$\hat{m}(x) = \frac{1}{n_k} \sum_{i=1}^{n} Y_i I_{B(k)}(X_i) \quad \text{for } x \in B(k) \tag{3.17}$$

where $n_k = 1/n \sum_{i=1}^{n} I_{B(k)}(X_i)$ is the number of design points belonging to bin $B(k)$.

Figure 9 shows regressograms for the 1989 EBF Engel curves. The regressogram is intuitively simple. Its disadvantages are the same as for the histogram in density estimation: dependence on the bin location, the bin size, many discontinuities and an asymptotic behavior of MISE $= O(n^{-2/3})$ inferior to other methods.

3.3 Kernel Regression Estimator

Nadaraya–Watson Estimator

An intuitive approach for estimating the regression function $m(x)$ is obtained when replacing the unknown joint and marginal density in (3.12) by nonparametric kernel density estimates. This leads immediately to the regression estimator proposed independently by Nadaraya (1964) and Watson (1964):

$$\begin{aligned} \hat{m}_{\mathrm{NW}}(x) &= \frac{\int y \hat{f}_{X,Y}(x,y)dy}{\hat{f}_X(x)} \\ &= \frac{\int y/(nh^2)\sum_{i=1}^n K((x-X_i)/h)K((y-Y_i)/h)dy}{1/(nh)\sum_{i=1}^n K((x-X_i)/h)} \\ &= \frac{\sum_i Y_i K((x-X_i)/h)}{\sum_i K((x-X_i)/h)}. \end{aligned} \tag{3.18}$$

Here for estimating the bivariate density $f_{X,Y}(x,y)$ we used a so–called product kernel, i.e. a bivariate kernel function $\mathbf{K}$ that allows factorization into two univariate kernel functions: $\mathbf{K}(u,v) = K(u)K(v)$.

Gasser–Müller Estimator

An alternative to the Nadaraya–Watson estimator, due to Gasser and Müller (1979), can be represented as a convolution of the data with the kernel function

$$\hat{m}_{\mathrm{GM}}(x) = \frac{1}{h}\sum_{i=1}^n \int_{s_{i-1}}^{s_i} K\left(\frac{x-z}{h}\right) \mathrm{d}z Y_i. \tag{3.19}$$

Here the data (X_i, Y_i) are ordered according to X_i and $s_i = (X_i + X_{i+1})/2, i = 1, \ldots, n-1, s_0 = 0, s_n = 1$.

Comparing the two kernel estimators the Nadaraya–Watson estimator has a smaller asymptotic variance which is calculated as $\mathrm{Var}(\hat{m}_{\mathrm{NW}}(x)) = \sigma^2/(nhf_X(x))R(K)$ while the Gasser–Müller variance is larger by a factor of 1.5, i.e. $\mathrm{Var}(\hat{m}_{\mathrm{GM}}(x) = 1.5\mathrm{Var}(\hat{m}_{\mathrm{NW}}(x))$ (for details see, e.g. Chu and Marron; 1991). However, considering asymptotic bias the Gasser–Müller estimator is preferable: The asymptotic bias of G–M estimator at an interior point, i.e. a point at least one bandwidth away from the boundary, is easy to interpret and is calculated as follows: $\mathrm{Bias}(\hat{m}_{\mathrm{GM}}(x)) = h^2\sigma_K^2 m''(x)$ where $\sigma_K^2 = \int K(u)u^2\mathrm{d}u$. In contrast, the Nadaraya–Watson estimator has an asymptotic bias in the interior that is much harder to interpret and that may become arbitrarily large when the density of X takes a small value: $\text{Bias}(\hat{m}_{\text{NW}}(x)) = h^2\sigma_K^2\{m''(x) + 2m'(x)f'_X(x)/f_X(x)\}$. Estimation at points near the boundary necessitates the use of *boundary* kernels for both, the Nadaraya–Watson and the Gasser–Müller method in order to guarantee the order of the bias at the boundary to remain as low as in the interior (Gasser et al., 1985). In terms of asymptotic MISE both kernel estimators achieve the same rate of $\text{MISE} = O(n^{-4/5})$.

The Gasser–Müller estimator has some technical advantages: it can be generalized straightforwardly to the problem of estimating derivatives of regression functions, and allows calculation in a fast $O(n)$ algorithm (Seifert, Brockmann, Engel and Gasser; 1994). Nevertheless, under a random design, there is no clear preference between the two kernel regression methods: the Nadaraya–Watson method has smaller variance while the Gasser–Müller estimator has preferable bias behavior.

Binned Convolution Estimator

The Binned Convolution estimator combines the advantages of both kernel methods: it has the asymptotic variance of the Nadaraya–Watson estimator and an asymptotic bias identical with the Gasser–Müller estimator. Kneip and Engel (1994) show that the Binned Convolution estimator is not only easy and fast to compute but also outperforms its competitors in finite sample performance. The binned convolution estimator starts with discretizing the data into bins by calculating bin average values and then applies the convolution estimator to the average bin values. For simplicity of representation we assume the unit interval $[0, 1]$ as domain of estimation for X:

Step 1: Discretize the data into L bins defined by

$$B(k) = [(k-1)/L, k/L), \quad k = 1, \ldots, L \tag{3.20}$$

Step 2: Define "new"data by calculating bin averages:

$$\tilde{Y}_k = \frac{1}{n_k}\sum_{i=1}^{n} Y_i I(X_i \in B(k)), \quad k = 1, \ldots, L \tag{3.21}$$

where $n_k = \#\{X_i \mid X_i \in B(k)\}$

Step 3: Finally, smooth the binned data

$$\hat{m}_{\text{BC}}(x) = \frac{1}{h}\sum_{k=1}^{L}\int_{(k-1)/L}^{k/L} K\left(\frac{x-z}{h}\right) dz\, \tilde{Y}_k. \tag{3.22}$$

An asymptotically optimal choice of the number of bins is calculated as $L = (2.4n/h^2)^{1/3}$ (see Kneip and Engel, 1994).

Bandwidth selection, choice of kernels

All of the above introduced regression methods require specification of a kernel function K and a smoothing factor h. Regarding the choice of a kernel the discussion in the previous section on density estimation applies. The same optimality criteria hold also in the regression case leading to identical optimal kernels. Data–driven bandwidth selection remains a delicate problem. Cross–validation methods had been the preferred choice of the earlier research until their proven high sample variability led to the development of *plug–in* bandwidth selectors.

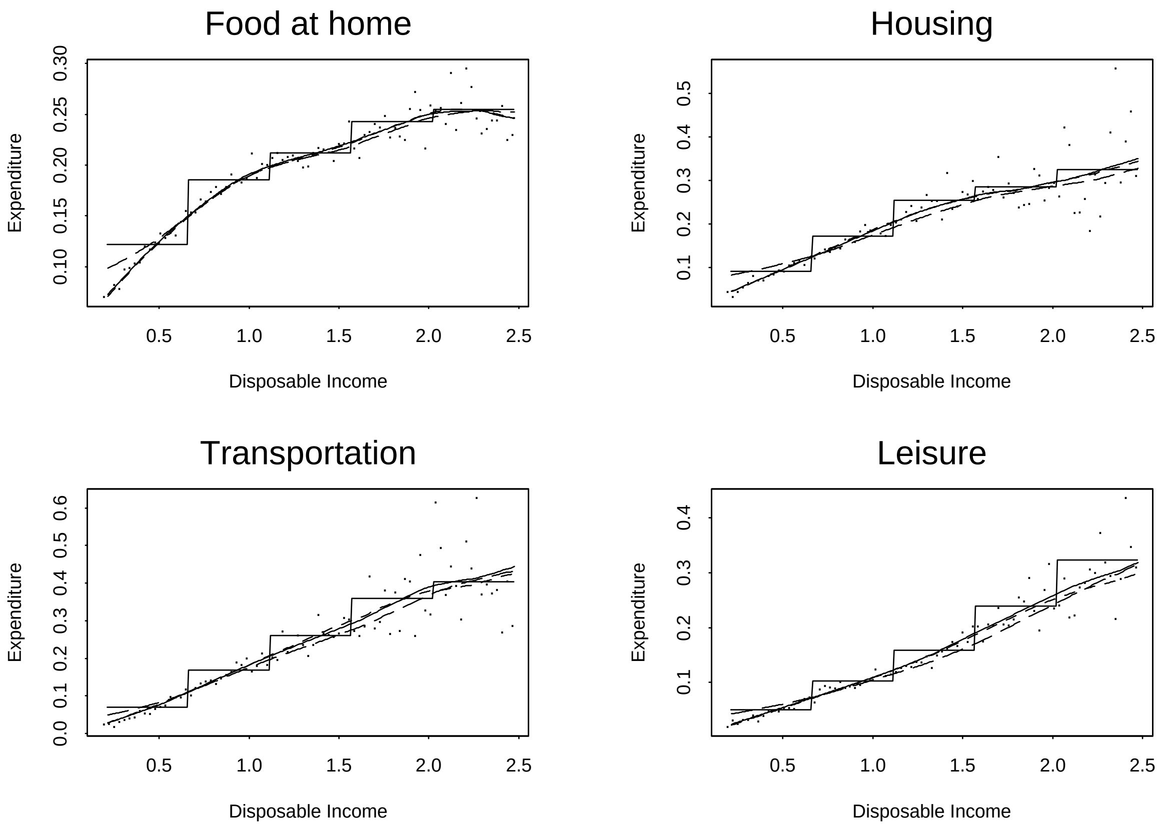

Figure 9: Estimates of Engel Curves: Binned–Convolution estimator (solid line), Nadaraya–Watson estimator (large dashes), Gasser–Müller estimator (small dashes), and Regressogram.

A flexible and fast *plug–in* method is due to Gasser, Kneip and Köhler (1991) and proceeds as follows (again, for simplicity, assume the unit interval $[0, 1]$ as domain of estimation)

Step 1:

$$h_0 = n^{-1/5} \tag{3.23}$$

Step 2: For $k = 1, \ldots, 11$

$$h_k = \left\{ \frac{\hat{\sigma}^2 R(K)}{4n\sigma_K^2 \int_0^1 \hat{m}_2(x; h_{k-1}n^{1/10})^2 \mathrm{d}x} \right\}^{1/5}. \tag{3.24}$$

Here $\hat{m}_2(x; h_{k-1}n^{1/10})$ is an estimator of m'' with bandwidth $h_{k-1}n^{1/10}$, and $\hat{\sigma}^2$ is an estimate for the variance σ^2 obtained, for example, by the method of Gasser, Sroka and Jennen–Steinmetz (1986).

Step 3: $h_{\mathrm{PI}} = h_r$.

The bandwidth inflation by $n^{1/10}$ when estimating the functional involving the second derivative is derived from asymptotic theory (see Gasser et al., 1991).

Figure 9 shows estimates of 1989 EBF Engel curves based on the three different kernel estimators. The dots correspond to the binned values $(\tilde{X}_i, \tilde{Y}_i)$ with $\tilde{X}_i$ denoting the mid–bin point and $\tilde{Y}_i$ calculated according to (3.21).

3.4 Local Polynomial Regression

Another approach that combines the asymptotic advantages of the Nadaraya–Watson and the Gasser–Müller kernel estimator is the method of local polynomial approximation. This method has gained much praise in the mathematical statistics literature of the recent years (e.g., Hastie & Loader, 1993). To estimate the regression function at a point x, a weighted least squares problem centered locally around x has to be solved. The resulting method of local polynomial regression estimation leads to an estimator with an asymptotic behavior identical with the Binned Convolution estimator: its bias coincides with the convolution kernel estimator while its variance is the same as for the Nadaraya–Watson estimator. In addition, local polynomials automatically adjust to the boundary. We illustrate this method for local linear approximation. To estimate $m(x)$ at point x we have to solve the following weighted least squares problem:

$$\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)(Y_i - a - b(x - X_i))^2 \to \min_{a,b} \tag{3.25}$$

Then set $\hat{m}(x) = a$. The smoothing factor h determines the size of the neighborhood for local approximation and the function K determines the weight given to points in that neighborhood. Fan (1993) shows that this method is asymtoptically minimax efficient among all linear smoothers. Local linear estimates of Engel curves are shown in Figure 10.

3.5 Spline Methods

Splines are an attractive alternative to kernel estimators in nonparametric regression. Two different approaches are *smoothing splines* and *regression splines*.

Smoothing splines are defined as twice differentiable functions minimizing

$$S_h(m) = \sum_{i=1}^{n}(Y_i - m(X_i))^2 + h\int m''(x)\mathrm{d}x. \tag{3.26}$$

Here the parameter h determines the smoothness of the estimate. Is $h = 0$, then any curve interpolating the data minimizes the functional S_h, while in the limit of $h \to \infty$ the resulting minimizer of S_h is the linear regression line. The solution of (3.26) is some sort of compromise between the aim of yielding a good fit of the data and the objective to obtain a curve without too much local variation. The integral term in (3.26) is a penalty for roughness. The resulting solution $\hat{m}$ is a cubic spline function with the design points X_i as knots (Reinsch, 1967).

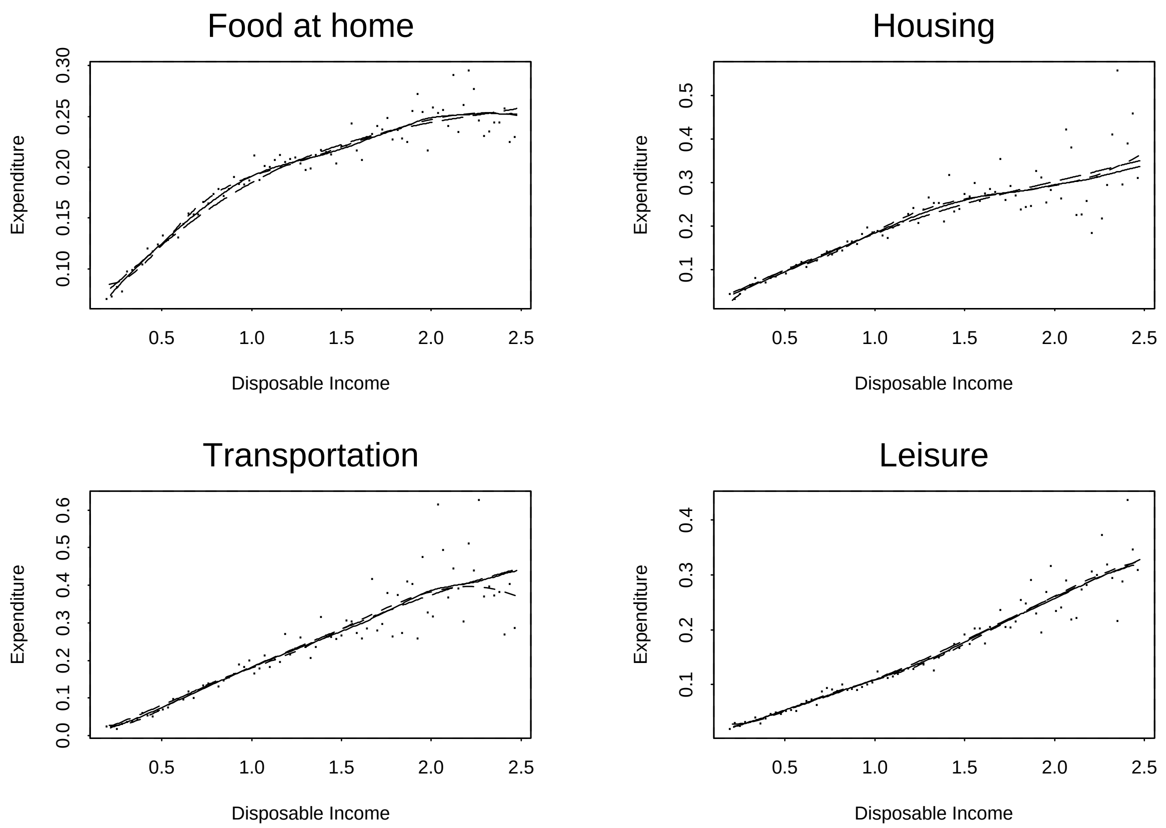

Figure 10: Estimates of Engel Curves, based on 1989 EBF data: Local Linear Estimator (solid), Regression Splines (large dashes), Smoothing Splines (small dashes).

Regression splines are obtained by placing fewer knots at some pre–chosen points. They are defined as minimizer $\hat{m}$ of $S(m) = \sum_{i=1}^{n}(Y_i - m(X_i))^2$ with the following properties: $\hat{m}$ is a cubic polynomial between knots and is twice differentiable at the knot points. Unlike for smoothing splines no explicit roughness penalty is involved. The smoothness of the estimate is determined by the number and the placement of the knots (Friedman and Silverman, 1989).Because of its implicit definition, the derivation of asymptotic properties becomes rather cumbersome. Considerable insight into the concept of smoothing splines is obtained through a result by Silverman (1984), who shows the equivalence of the smoothing spline to a kernel regression estimator that is based on the kernel $K^*(u) = n/2 \exp(-|u|/\sqrt{2}) \sin(|u|/\sqrt{2} + \pi/4)$ and the local, design–adaptive bandwidth $h^*(X_i) = (h/nf_X(X_i))^{1/4}$. Figure 10 shows estimates of Engel curves based on regression splines and on smoothing splines. There is hardly any discernable difference between them and the local linear estimator in this example.

3.6 Variance Estimation, Refinements

All curve estimation methods considered so far assume homoscedastic data, i.e. data of equal variance. But family expenditure data are not likely to fulfill this assumption: households with higher disposable income have much more freedom how to spend their financial resources than poorer household who more or less may be forced to use their income on essential goods for their survival. With minor modifications the method of binned convolution estimation can also be applied to estimate variance curves: Again, the sample space is split into a number of bins. Then instead of calculating a bin mean value in (3.21) one has to calculate the sample variance within each bin. Finally, the binned sample variances are smoothed with the convolution method (3.19).

Step 1: Discretize the data into L bins $B(k) = [(k-1)/L, k/L), k = 1, \ldots, L$

Step 2: Calculate within–bin variances:

$$V_k = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y}_k)^2 I(X_i \in B(k)).$$

Here $\bar{Y}_k$ is the bin mean $\bar{Y}_k = 1/n_k \sum_{i=1}^{n} Y_i I(X_i \in B(k)), n_k := \#\{i \mid X_i \in B(k)\}$.

Step 3: Finally, smooth the within–bin variances

$$\hat{v}(x) = \frac{1}{h} \sum_{k=1}^{L} \int_{(k-1)/L}^{k/L} K\left(\frac{x-z}{h}\right) \mathrm{d}z\, V_k$$

Figure 11 shows estimates based on the binned convolution method for the variance curves of the cross–sectional demand.

Eye inspection of Figure 9 gives a clear message: the assumption of homoscedasticity for the family expenditure data can no longer be sustained. A first approach respecting the heteroscedasticity of the data is based on transformations: in a very rough approximation the variances curves appear linear in the

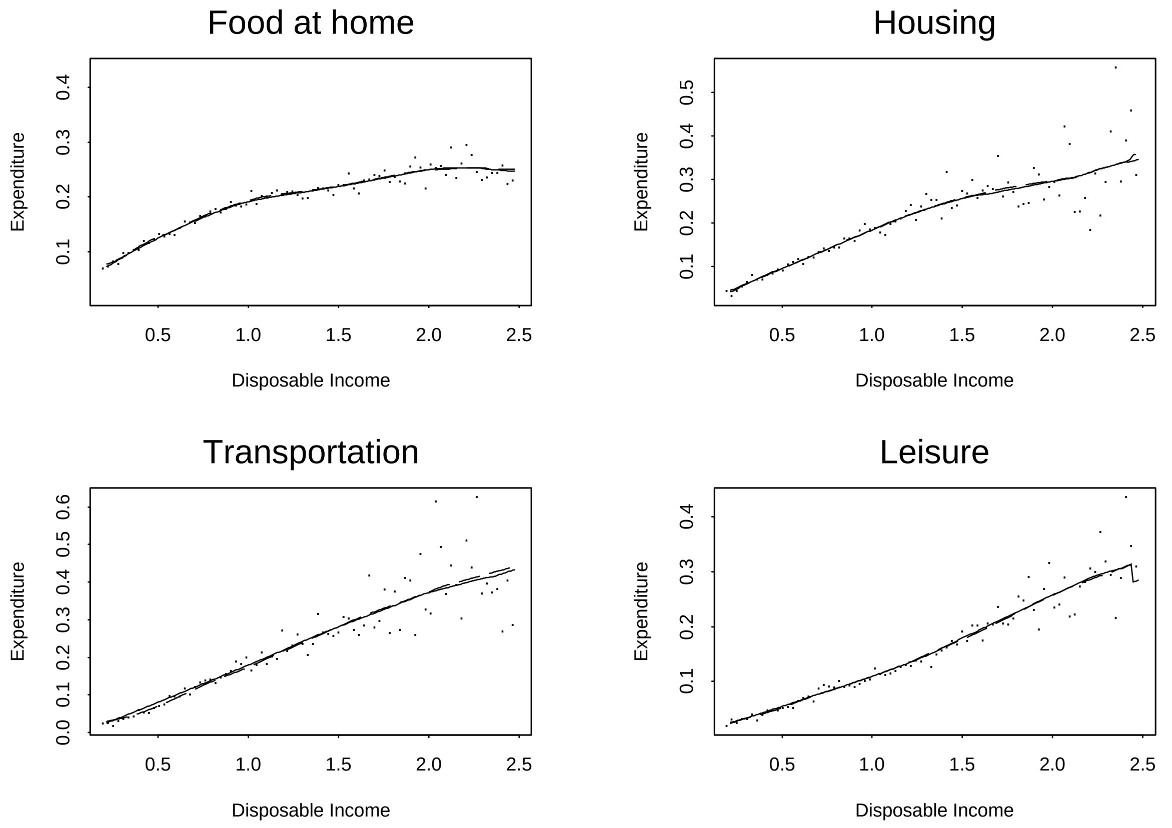

Figure 11: Estimated variance curves of the cross–sectional demand for food at home, housing, transport and leisure.

variable disposable income. Hence, considering budget shares, i.e. expenditures for the particular good under consideration divided by disposable income, may lead to data with approximately constant variance. This leads to *budget share curves* $y = bs(x) = m(x)/x$ given as the relative amount or the portion of the budget spend on a particular commodity. All methods for estimating Engel curves as discussed above can also be applied for estimating Budget Share curves. The only difference is that the raw data, household expenditures for various commodities, are divided by disposable income.

Multiplying the estimated budget share curve with disposable income, i.e. $bs(x) \times x$ results in another estimate of the cross–sectional Engel curve. Notice the conceptual difference: instead of estimating Engel curves from the data directly, we apply nonparametric regression methods to the relative expenditure data and multiply the resulting estimate afterwards with disposable income.

A second approach responding to the heteroscedasticity of the data is based on allowing variable bandwith size across the sample space. Estimating curves from unequal variance data requires more smoothing in areas of the sample space where the data are more variable and less smoothing where the data are less variable. Also, a variable bandwidth estimator adapts to the local curvature of the data. A slight modification of above algorithm for bandwidth selection (3.23), (3.24) leads to a variable bandwidth selector. The method is due to Brockmann, Gasser and Herrmann (1993) and allows adaptation to the local smoothness of the data. The method proceeds as follows:

1. Step 1 and step 2 for $k = 1, \ldots, 8$ as in (3.23) and (3.24)
2. For $k = 9, 10$

$$h_k(x) = \left\{ \frac{\hat{\sigma}^2(x) R(K)}{4n\sigma_K^2 \hat{m}_2(x; h_{k-1}(x) n^{1/10})^2} \right\}^{1/5} \tag{3.27}$$

3. $h_{\mathrm{PI}}(x) = h_{10}(x)$.

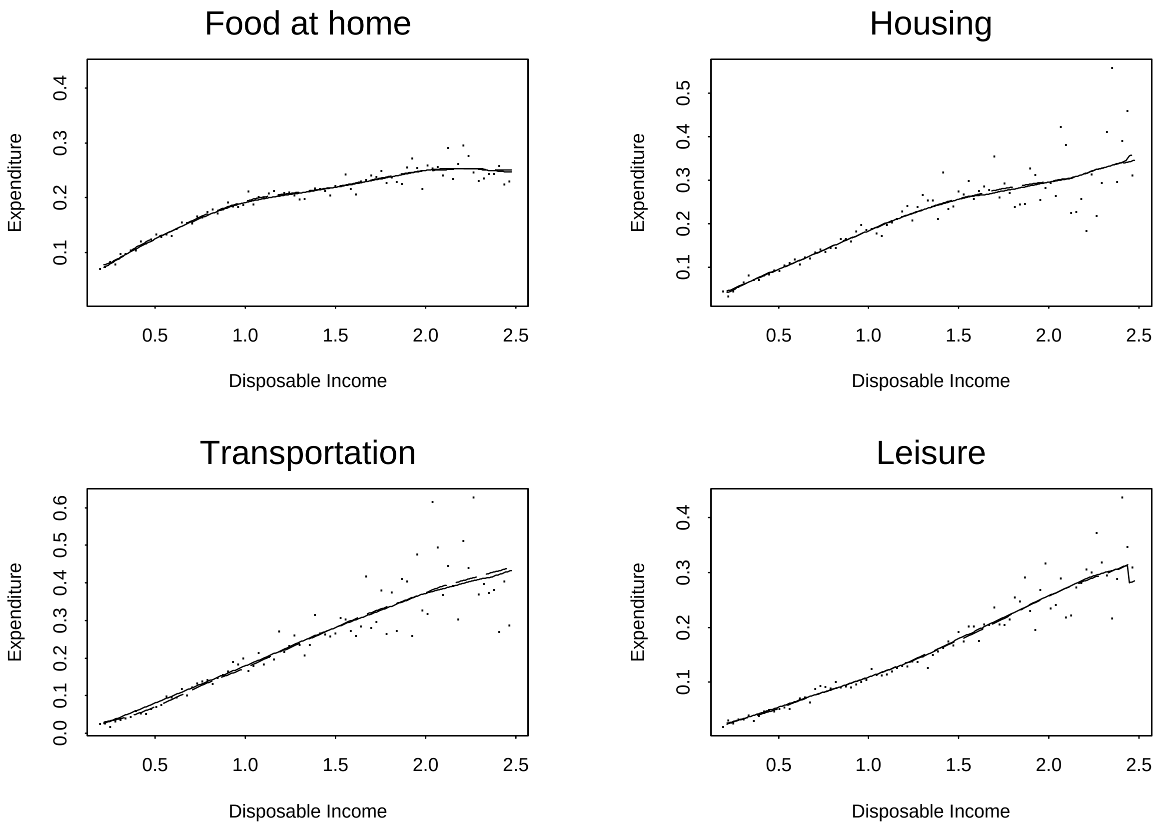

Figure 12: Estimates of Engel Curves based on 1989 EBF data: Binned Convolution estimator with locally adaptive bandwith (solid line) and the Budget Share Curve multiplied with disposable income (dashed line).

Notice that here we select a whole bandwidth function $h(x)$ in order to estimate another function, the cross–sectional Engel curve. Obviously, applying the theoretically best bandwidth function $h(x)$ results in an estimator which is superior to an estimator based on a constant bandwidth parameter, regardless the local smoothness and variability of the data. However, the optimal bandwidth function is not available but has to be estimated instead. It is much more complicated to estimate a bandwidth function than just a fixed bandwidth parameter. The complexity in estimating a bandwidth function introduces additional noise. Therefore, it is the context of estimation given by the local smoothness and variability of the data, that determines if it is advantageous to allow for a variable bandwidth function or to proceed with the simpler concept of a fixed global bandwidth. Figure 12 shows estimates of cross–sectional Engel curves based on the binned convolution estimator under a locally adaptive bandwidth together with curve estimates based on the Budget Share curve (Binned Convolution estimate with fixed bandwidth) multiplied with disposable income. By eye a difference is hardly noticeable.

3.7 Summary

We considered the problem of estimating cross–sectional Engel curves based on data from household expenditure surveys. The data contain information on individual households demand for a variety of (aggregated) goods and disposable income. In statistical terms, they can be characterized as heteroscedastic random design data. The variances increase considerably with increasing income. The sample size is quite large ($n \approx 9000$), with about 75% of the data in an income range between .3 to 1.5 times mean income. For very high income levels (where the variance is high) only few observations are available. We investigated a variety of different methods for estimating Engel curves: parametric models, regressogram, kernel estimators, local polynomial approximation, splines and estimates that first estimate the Budget Share curves and then multiply the result with disposable income.

To evaluate the performance of various Engel curve estimators we calculated the weighted mean average squared error (MASE), preferred here over MISE mainly due to its ease of estimation. The weights are the inverse of the variance and don't necessarily add up to 1. In fact, in our applications their sum is considerably larger than 1. Conceptually, weigthed MASE is a reasonable measure because it requires the estimate to be more accurate at locations where the variance is low. An intuitive estimator for MASE can be derived when representing all nonparametric methods considered here as *linear smoothers*: introducing vector notation, let $Y = (Y_1, \ldots, Y_n)^T, m = (m(X_1), \ldots, m(X_n))^T, \hat{m} = (\hat{m}(X_1), \ldots, \hat{m}(X_n))^T, \epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$, then model (3.11) can be written as

$$Y = m + \epsilon$$

with

$$\mathrm{Var}(\epsilon) = \Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \dots \\ \dots & \dots & \dots \\ 0 & \dots & \sigma_n^2 \end{pmatrix}.$$

Then

$$\hat{m} = S_h Y$$

where the $n \times n$ smoother matrix S_h depends on the particular estimation method and the smoothing parameter h. For example, for the Gasser–Müller estimator we have $S_h^{(ij)} = 1/h \int_{s_{i-1}}^{s_i} K((X_j - u)/h)\mathrm{d}u$. From straightforward calculation it follows that

$$\frac{1}{n}\mathrm{E}(Y - S_h Y)^T \Sigma^{-1} (Y - S_h Y) = \mathrm{MASE} - \frac{2}{n}\mathrm{trace}(S_h) + 1.$$

Hence, MASE can be estimated by

$$\widehat{\mathrm{MASE}} = \frac{1}{n}(Y - S_h Y)^T \hat{\Sigma}^{-1} (Y - S_h Y) + \frac{2}{n}\mathrm{trace}(S_h) - 1. \tag{3.28}$$

Table 1 lists values for estimated MASE averaged over the three years (1979, 1984 and 1989) and 14 (aggregated) commodities.

	Estimator	Weighted MASE
	Working Model	3.3223
Parametric	Deaton Model	.6398
	Pollak & Walsh	.3796
Methods	Logarithmic polynomial of order 3	.3322
	Regressogram	26.5902
Non–	BCE, adaptive bandwidth, via Budget Share	.3128
	Smoothing Spline, via Budget Share	.3113
Parametric	Regression Spline	.2704
	BCE, global bandwidth, via Budget Share	.1711
Estimators	Smoothing Spline	.0829
	BCE, adaptive bandwidth	.0760
	BCE, global bandwidth	.0287

Table 1 Weighted MASE for various Engel Curve estimators.

Low dimensional parametric models suffer from too much inflexibility and seem inappropriate, but already the 3 dimensional model by Pollak and Walsh and the 4 dimenional logarithmic model suggested by Kneip (1993) performed almost as good as nonparametric methods that are based on Budget Share estimates. Despite its supposed variance stabilization there is no benefit in estimating via

Budget Shares curves. This may be explained by the relatively poor performance of the Budget Share estimator for low disposable income where the variance of the budget shares are much higher then the variances of absolute expenditures. Among the various nonparametric methods the two versions of Binned Convolution estimator (fixed and adaptive bandwidth) are the most promising contenders, again confirmed by their practical performance and by theoretical considerations. Engel curves estimates obtained by smoothing splines also lead to results that are very close to the Binned Convolution estimator.

ACKNOWLEDGEMENT

We are grateful to Michael Jerison for his helpful comments. Research supported by Sonderforschungsbereich 303 (project A 3) with funds from the Deutsche Forschungsgemeinschaft

References:

Bierens, H.J. and H. A. Pott-Buter (1990). Specification of household Engel curves by nonparametric regression. *Economteric Reviews* **9**(2), 123–184

Brockmann, M., Gasser, T. and E. Herrmann (1993). Locally adaptive bandwidth choice for kernel regression estimators. *J. American Statist. Assoc.* **88**, 1302–1309

Chu, C.K. and J.S. Marron (1991). Choosing a kernel regression estimator. *Statistical Science* **6**, 404–433

Deaton(1981). Three Essays on a Sri Lankan Household Survey. *Living Standards Measurement Study.* W.D. No.11, Washington: The World Bank

Engel, Ernst (1857). Die Produktions– und Consumptionsverhältnisse des Königreichs Sachsen. *Reprinted 1895 in: Bulletin de Institut International de Statistique* **9**, 1–54

Engel, J., E. Herrmann and T. Gasser (1994). An iterative bandwidth selector for kernel estimation of densities and their derivatives. *J. Nonparametric Statistics*, to appear

Enquete Budget de Famille (1979, 1984–85, 1989): Institut National de la Statistique et des Etudes Economiques. Division *Condition de Vie des Menages*, Paris

Eubank, R. (1988). *Spline Smoothing and Nonparametric Regression.* Dekker, New York.

Fan, J. (1993). Local linear regression smoothers and their mini–max efficiencies. *Ann Statist* **21**, 196–216

Friedman, J.H. and B.W. Silverman (1989). Flexible parsimonious smoothing and additive modelling. *Technometrics* **3**, 3–21

Gasser, T. and H.G. Müller (1979). Kernel estimation of regression functions. In: Th. Gasser and M. Rosenblatt, eds. *Smoothing Techniques for Curve Estimation.* Lecture Notes in Math., Vol. 757, Springer, New York, 23–58

Gasser, T., H.G. Müller and V. Mammitzsch (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc Ser. B* **47**, 238–252

Gasser,T., A. Kneip and W. Köhler (1991). A flexible and fast method for automatic smoothing. *J. American Statist. Assoc.* **86**, 643–652

Gasser, T., L. Sroka and C. Jennen–Steinmetz (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73**, 625–633

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Cambridge

Härdle, W., W. Hildenbrand and M. Jerison (1991). Empirical evidence on the law of demand, *Econometrica* **59**, 1525–1549

Härdle, W. and M. Jerison (1991). Cross–section Engel curves over time. *Recherches Economiques de Louvain* **57**, 1525–1549

Härdle, W. and A. Kneip (1992). Testing a regression model when we have a smooth alternative in mind. *Discussion Paper No. A–178, Sonderforschungsbereich 303, University of Bonn*

Härdle, W. and T. Stoker (1989). Investigating smooth multiple regression by the method of average derivatives. *J American Statist. Assoc.* **84**, 968–995

Hastie, T. and C. Loader (1993). Local regression: automatic kernel carpentry. *Statistical Science* **8**, 120–143

Hildenbrand, W. (1994). *Market Demand: Theory and Empirical Evidence.* Princeton University Press.

Hildenbrand, W. and A. Kneip (1993). Family Expenditure Data, Heteroscedasticity and the Law of Demand. *Ricerche Economiche* **47**, 137–165

Howe, H., Pollak, R.A. and T.J. Wales (1979). Theory and time series estimation of the quadratic expenditure system. *Econometrica* **47**, 1231–1247

Jones, M.C., J.S. Marron and B. Park (1992). Progress in data–based bandwidth selection for kernel density estimation. *Working Paper 92–014, Australian Graduate School of Management*

Leser, C.E.V.(1983). Form of Engel functions. *Econometrica* **31**, 694–703

Kemsley, W.F., Redpath, R.D. and M. Holmes (1980). *Family Expenditure Survey Handbook.* London: Her Majesty's Stationary Office

Kneip A. and J. Engel (1994). A remedy for effective regression estimation under random design. *Discussion paper No A–431, Sonderforschungsbereich 303, University of Bonn*

Müller, H.G. (1988). *Nonparametric Analysis of Longitudinal Data.* Lecture Notes in Statistics, Vol. 46, Springer, Berlin

Nadaraya, E.A. (1964). On estimating regression. *Theory Probab. Appl.* **9**, 141–142

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, 1065–1076

Pollak, R.A. and T.J. Wales (1978). Estimation of complete demand systems from household budget data. *American Economic Review* **68**, 348–359

Reinsch, C.H. (1967). Smoothing by spline functions. *Numer. Math.* **16**, 177–183

Rosenblatt, M.(1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27**, 832–837

Scott,D.W. (1985). Average shifted histograms: Effective nonparametric density estimation in several dimensions. *Ann. Statist.* **13**, 1024–1040

Scott,D.W. (1992). *Multivariate Density Estimation.* Wiley, New York

Seifert,B., M.Brockmann, J. Engel and T. Gasser (1994). Fast algorithms for nonparametric curve estimation. *J. Comp. and Graph. Statist.*, **3** 192–213

Sheather,S.J. and M.C. Jones (1991). A reliable data–based bandwidth selection method for kernel density estimation. *J. R. Statist. Soc. B* **53**, 683–690

Silverman, B. (1984). Spline smoothing: the equivalent variable kernel method. *Ann. Statist.* **12**, 898–916

Silverman, B. (1985). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London

Wand, M.P., J.S. Marron and D. Ruppert (1991). Transformations in density estimation (with discussion), *J. American Statist. Assoc.* **86**, 343–361

Watson, G.S. (1964). Smooth regression analysis. *Sankhyā Ser A* **26**, 359–372

Working, H. (1943). Statistical Laws of Family Expenditure. *J. American Statist. Assoc.* **38**, 43–56