Comparing Nonparametric Regression Quantiles*

Cristian Huse†
LSE

March 24, 2004
Preliminary version. Comments Welcome.

Abstract

This paper investigates how conditional quantiles of a given distribution relate to each other. Given two conditional quantiles estimated nonparametrically, we investigate their relation by linking them through a parametric transformation. Asymptotic normality of the associated parameter vector is established, and the method is illustrated with data from the Family Expenditure Survey (FES) of UK households. The FES records expenditures of households on six broad categories of goods (alcohol, clothing, food, fuel, transport, and "other goods"), and the methodology is applied by estimating and comparing the conditional quantiles of the Engel relation. The only category for which expenditure can explain the shift in the quantile curves is for "other goods" relationship, indicating an increase in heterogeneity for better off households, suggesting a "taste for variety" effect as the expenditure level increases. For the remaining categories one cannot reject the null of a parallel shift of the quantile curves.

Keywords: Quantile Regression, Semiparametric Estimation, Specification Testing, Engel Curve, Household Expenditure, Budget Shares.

JEL Classification: C12 (Hypothesis Testing), C13 (Estimation), C14 (Semi- and Nonparametric Methods), D12 (Consumer Economics: Empirical Analysis).

*I am very grateful to Oliver Linton for guidance throughout this project.
†Email address: C.Huse@lse.ac.uk. Correspondence address: Financial Markets Group, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom.

1 Introduction

When compared to the standard regression model, quantile regression (Koenker and Basset, 1978) enables one to go beyond the study of the conditional mean $E(y|X = x)$, obtained by applying least squares to a regression equation $y = X\beta + u$, and investigating the whole conditional distribution $F(y|X = x)$ by estimating its quantiles, thus obtaining a more general picture of the behaviour of the data at hand. In effect, quantile regression techniques have been used extensively in areas such as microeconometrics with great success [see, for instance, Powell (1994) for a thorough review of the methods].

The standard approach in quantile regression is, as it is in classical linear regression, to specify a linear model $y = X\beta + u$ and, when interested in estimating a conditional quantile τ, assuming that the $\tau-$th quantile of the error distribution is zero [see Buchinsky (1998) and Koenker and Hallock (2001) for nice introductions]. Extensions to nonlinear models parallel those in least squares methods. Although in early days computing the quantile regression estimator was a considerable burden, nowadays this is far from being an issue [see Koenker and D'Orey (1987), and Portnoy and Koenker (1997) for developments in computational methods]. Misspecification issues were addressed recently by Kim and White (2002) $-$ using the fact that any misspecification in the conditional quantile is a form of conditional heteroskedasticity in the error term, thus implying violation of the information matrix equality, their approach parallels White (1980), but now in the L_1-world.

There are situations, however, where the reasearcher would like to allow data to "speak for themselves" instead of imposing a parametric specification on the regression function. A classical example is in Engel curve estimation [Bierens and Pott-Butter (1990), Blundell, Duncan, and Pendakur (1998), Blundell Browning, and Crawford (2003)]. In this particular application, the use of nonparametric techniques results in shapes hard to mimic using parametric methods due to the flexibility offered by the former.

Given a pair of quantile regression curves, it might be of interest to compare them in order to get a more accurate idea of the behaviour of the conditional distribution $F(y|X = x)$. In this paper we compare nonparametric quantile curves by relating them through a parametric function ϕ which links them. In the simplest case possible, ϕ is just a parallel shift, but general nonlinear parameterizations are also allowed, subject to regularity conditions. Relaxing the assumption of a parametric regression function specified a priori yields a more general method $-$ less prone to specification error and more suited to deal with situations where the researcher has no idea about how the regression equation should be specified.

The paper is divided as follows. In Section 2 I briefly discuss quantile regression techniques and some potential problems of the standard model, such as the inconsistency result coming from heteroskedasticity. Section 3 presents the model and discusses identification, estimation, and testing. Section 4 illustrates the method by estimating quantile Engel curves, whereas Section 5 concludes. Proofs and derivations are deferred to the Appendix.

2 Comparing Quantiles

Consider an independent and identically distributed random vector (X, Y), where both X and Y are real-valued, and whose realizations are denoted by (x, y). A parametric linear quantile regression model reads

$$y_i = x_i'\beta + u_i$$

by assuming that the $q-$th quantile of u_i is zero $(0 < q < 1)$, one obtains the $q - th$ quantile of $F(y|X = x)$. One key assumption of the model above is the linearity of the conditional quantile function, not always supported by the data. An immediate consequence of this specification is the homoskedasticity of the conditional distribution $-$ besides being linear, the quantiles are also parallel to each other $-$ a rather stringent assumption in many situations. Violating the null of homoskedasticity is much worse in this case than in the least squares world since, instead of raising only efficiency issues, heteroskedasticity in quantile regression models implies inconsistency of the parameter estimates. This is explored by Kim and White (2002) who, based on this fact, propose specification tests for quantile regression. Given the possible pitfalls involved in misspecifying the regression model, one may well advocate the use of general nonlinear specifications of the regression function, but this might involve more knowledge about the phenomenon under study that what researchers typically have.

In this paper we attempt to avoid misspecification issues in quantile regression by estimating nonparametric quantile curves. Nonparametric techniques have been applied to a variety of problems and offer a great deal of flexibility, as opposed to traditional parametric methods. The price to be paid is a slower rate of convergence, thus the need of larger samples when compared to parametric techniques but, depending on the type of application, the benefits might well outweigh the costs.

Given a set of quantile curves, one has a broad picture of how the conditional distribution $F(Y|X = x)$ behaves. It might also be of interest to investigate how the different quantiles relate eg. how does heterogeneity behave for different values of X, in particular, whether the quantile curves are parallel. Alternatively, one might be interested in the behaviour of the tails of the distribution, giving special attention to either lower or upper quantiles, as in finance and hidrology, respectively. In all these examples, the fundamental question summarizing the different interests is whether, and how, the different quantiles relate, and what drives the relationship among them.

Recently, Koenker and Xiao (2002) investigated the relation between the same quantile, say q, of two distributions. In the context of the treatment-efffects problem, they compare the same quantile of the distribution of the treated and non-treated, assuming the regression models are linear. In what follows, however, we investigate the features of a conditional distribution by studying how its (nonparametrically estimated) quantiles relate. To fix ideas, consider our application, which uses data from a sample of households. The data consists of a cross section of UK households reporting their expenditure patterns. The Engel relation compares the expenditure level of a household to the expenditure shares of different (categories of) goods purchased. However, the widely used Engel curve is a conditional mean of budget shares given expenditure levels, thus neither giving a broad picture of the whole distribution, nor accounting properly for the huge asymmetry in the data, besides being subject to be driven by outlying observations.

Nonparametric Engel curves have been studied extensively in the literature [see Bierens and Pott-Butter (1990) for a seminal contribution, and Blundell, Duncan, and Pendakur (1998), and Blundell Browning, and Crawford (2003) for recent studies] and shown to have desirable properties when compared to their parametric counterparts. Estimating nonparametric quantile curves is then expected to bring together the flexibility of nonparametric methods and the broader picture enabled by using quantile methods. Once the quantile Engel curves are estimated, they are linked by using suitable parameterizations in order to study whether $-$ and how $-$ they relate. For instance, comparing two quantile curves of the distribution of household budget shares conditioned on expenditure levels would provide insights on how consumer heterogeneity evolves for each different good under study, and whether heterogeneity is related to the expenditure level.

3 The Model

The model we consider can be summarized in a two-stage procedure. The first stage consists on the estimation of nonparametric quantile curves. Then, given a parameterization linking the quantile curves, the second step consists on estimating the associated parameter vector, θ.

The first stage consists on estimating nonparametric quantile curves. To this end, consider the regression model

$$w_{ij} = m_{q_j}(x_i) + u_{ij} \ , \ i = 1, ..., N, \ j = 1, 2$$

where (X, W) is an independent and identically distributed random vector, where both X and W are real-valued, and whose realizations are denoted by (x, w) and with the q_j-th quantile of u_j being zero $(0 < q_j < 1)$.

The $q_j - th$ quantile (local linear) regression estimator at the point x_0 is defined as

$$\widehat{m}_{q_j}(x_0) \equiv \widehat{a}_0 = \arg\min_{a_0, a_1} \frac{1}{N} \sum_{i=1}^{N} \rho_{q_j}(t).K\left(\frac{x_i - x_0}{b}\right)$$

where $\rho_{q_j}(.)$ is the check function, with $t = |w_{ij} - a_0 - a_1.(x_i - x_0)|. + (2q_j - 1).(w_{ij} - a_0 - a_1.(x_i - x_0))$.

Early approaches to this estimation problem were proposed by Härdle and Gasser (1984), and Tsybakov (1986) for the local constant fitting method ie. $a_1 = 0$. More recently, Chaudhuri (1991) and Welsh (1996) consider the local polynomial case, which exhibits better properties at the boundary of x.

In the second stage we investigate whether the quantile curves can be related by means of a parametric transformation ϕ. Put formally, we consider the model

$$\widehat{m}_{q_1}(x_i) = \phi(\theta, \widehat{m}_{q_2}(x_i), y_i) + e_i, \ i = 1, ..., N \tag{1}$$

where $\widehat{m}_{q_1}(x_i)$ and $\widehat{m}_{q_2}(x_i)$ are estimates of the quantile curves at the point x_i, ϕ is a parameterization linking the quantile curves, θ is the parameter vector of interest, y_i is a set of explanatory variables, and e_i is an error term.

In a related paper, Härdle and Marron (1990) investigated the relation between two nonparametric (mean) regression functions, establishing the asymptotic normality of the parameter vector of the function linking the curves. In their study, however, they do not aim at explaining what actually drives the relation between the curves, not considering explanatory variables y_i, besides restricting themselves to an affine ϕ ie. a pure location-scale shift model.

Example (Pure location shift). The simplest case is the pure (location) shift viz.

$$\widehat{m}_{q_1}(x_i) = \alpha_0 + \widehat{m}_{q_2}(x_i) + e_i, \ i = 1, ..., N$$

where $\alpha_0 \equiv \alpha_0(q_1, q_2)$ is the shift parameter depending on quantiles (q_1, q_2). ■

The pure location shift model reflects a situation where the quantile curves − regardless of their shapes − are parallel shifts of one another.

Even such a simple model as the pure location shift sheds light on interesting questions. For instance, given J quantile curves and two curves j and j^*, one might be interested in testing wether $\alpha_0 \equiv \alpha_0(q_j, q_{j*}) = \alpha_0(|q_j - q_{j^*}|)$, which can be phrased as whether the parallel shifts are of the same magnitude regardless of the quantiles considered.

Example (Location shift). A slightly more general case is given by

$$\widehat{m}_{q_1}(x_i) = \alpha_0 + y_i'\alpha_1 + \widehat{m}_{q_2}(x_i) + e_i,\ i = 1, ..., N$$

where y_i is a $k-$dimensional column vector of variables one believes should account for the parallel shift $-$ in particular, $y_i = x_i$. ■

Location shifts can well be thought of as too simplistic to explain the interaction among quantile curves, so that we consider other cases of interest.

Example (Location-scale shift). Introducing a scale component in the parameterization allows for the quantile curves to drift apart from each other as the covariate values change. This allows for an heteroskedastic behaviour of the distribution, conditional on the value taken by the covariate. As before, it might be of interest to investigate what makes the curves drift apart, so that we write the location-scale shift model as

$$\widehat{m}_{q_1}(x_i) = \alpha_0 + y_i'\alpha_1 + (\beta_0 + y_i'\beta_1).\widehat{m}_{q_2}(x_i) + e_i,\ i = 1, ..., N$$

■

In what concerns estimation of the parameter vector θ, a standard way of tackling the problem would be to assume the error term e_i in (1) to be mean zero and then consider its (weighted) integrated squared distance. There are, however, two reasons to consider alternatives to this approach. First, in the case of highly asymmetric distributions, one might be willing to consider a broader view of the picture than the conditional mean. Second, in the case of outlying observations, squared distances are known not to be robust, being driven by these outlying observations. For instance, in our application, it is known *a priori* that the data under study are highly asymmetric, possibly with outlying observations in the lower and, especially, higher expenditure levels. This is further documented and discussed in the empirical section.

It then follows that we consider a quantile regression problem also in the second stage of our procedure. The corresponding moment condition[1] is

[1] A particular case of which is given by

$$G_N^{\tau}(\theta, \widehat{m}) \equiv \frac{1}{N}\sum_{i=1}^{N}\left(\tau - \frac{1}{2} + \frac{1}{2} sign\left[\widehat{m}_{q_1}(x_i) - \phi(\theta, \widehat{m}_{q_2}(x_i), y_i)\right]\right).D_{\theta}\phi(\theta, \widehat{m}_{q_2}(x_i), y_i).\widetilde{w}(x_i),$$

using the empirical measure.

$$G_N^\tau(\theta,\widehat{m}) \equiv \int \left(\tau - \frac{1}{2} + \frac{1}{2} sign\left[\widehat{m}_{q_1}(x_i) - \phi(\theta, \widehat{m}_{q_2}(x_i), y_i)\right]\right).D_\theta\phi(\theta,\widehat{m}_{q_2}(x_i),y_i).\widetilde{w}(x_i).dx$$

where $w(x_i)$ is a weighting function which can be used for trimming observations in regions of data scarcity, where kernel estimates are expected to be rather unstable.

The estimator of θ is then defined as

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} G_N^\tau(\theta,\widehat{m})'.\Omega^{-1}.G_N^\tau(\theta,\widehat{m})$$

where Ω^{-1} is a weighting matrix.

Asymptotic properties of this general class of estimators have been studied in Andrews (1994), Newey (1994) and, more recently, by Chen, Linton, and van Keilegom (2003, CLK hereafter).

Example (Location-scale shift). The moment condition in this case reads

$$G_N^\tau(\theta,\widehat{m}) \equiv \int \left(\tau - \frac{1}{2} + \frac{1}{2} sign\left[\widehat{m}_{q_1}(x_i) - \alpha_0 - y_i'\alpha_1 - (\beta_0 + y_i'\beta_1).\widehat{m}_{q_2}(x_i)\right]\right).z_i.\widetilde{w}(x).dx$$

with $z_i = \begin{bmatrix} 1 & y_i & \widehat{m}_{q_2}(x) & y_i.\widehat{m}_{q_2}(x) \end{bmatrix}'$. ■

In what follows I discuss identification, estimation, and testing issues.

3.1 Identification

Identification is verified by checking whether the population moment condition $G^\tau(\theta,m)$ is zero at $(\theta,m)=(\theta_0,m_0)$, and that its derivative $H_1^\tau(\theta,m)$ is non-zero at $(\theta_0,m_0)=(\theta,m)$ on a non-negligible set.

As shown in the Appendix,

$$G^\tau(\theta,m) = E\left(\left[\tau - \frac{1}{2} + \frac{1}{2} sign(e_i)\right].D_\theta\phi(\theta,\widehat{m}_{q_2}(x_i),y_i).\widetilde{w}(x_i)\right)$$

now setting $(\theta,m)=(\theta_0,m_0)$ one obtains that it is zero given $E\left[D_\theta\phi(\theta,\widehat{m}_{q_2}(x_i),y_i)\right]$ bounded, continuous and of full (column) rank, and $\widetilde{w}(.)$ bounded . Moreover,

$$H_1^\tau(\theta_0,m_0) = -\frac{1}{2}E\left[f_e(0|z).D_\theta\phi(\theta_0,\widehat{m}_{0q_2}(x_i),y_{0i}).D_\theta\phi(\theta_0,\widehat{m}_{0q_2}(x_i),y_{0i})'.\widetilde{w}(x_i)\right]$$

which is negative definite given $E\left[D_\theta \phi(\theta_0, \widehat{m}_{0q_2}(x_i), y_{0i})\right]$ bounded, continuous, and of full (column) rank in a neighbourhood of the true parameter value, a bounded weighting function, and an error density bounded away from zero.

In what follows, these conditions will be assumed for the general case, although I carefully discuss and justify them for the specifications we adopt in the empirical section, showing that they imply mild restrictions easy to be verified in practice.

3.2 Estimation

The estimation problem can be understood as a two-stage GMM problem, with the first stage consisting on the estimation of nonparametric quantile regression functions, and the second, on the minimization of a chosen distance between these two curves, given a pre-specified parameterization linking them.

The moment condition of the second stage is given by

$$G_N^\tau(\theta, \widehat{m}) \equiv \int \left(\tau - \frac{1}{2} + \frac{1}{2} sign\left[\widehat{m}_{q_1}(x_i) - \phi(\theta, \widehat{m}_{q_2}(x_i), y_i)\right]\right) .D_\theta \phi(\theta, \widehat{m}_{q_2}(x_i), y_i).\widetilde{w}(x).dx$$

The estimator of θ is then defined as

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} G_N^\tau(\theta, \widehat{m})'.\Omega^{-1}.G_N^\tau(\theta, \widehat{m})$$

where Ω^{-1} is a weighting matrix.

The following result establishes the asymptotic normality of $\widehat{\theta}$. It relies on methods developed in CLK and holds under mild primitive conditions which are discussed below. As is usual in quantile regression methods, boundedness conditions on the error densities are imposed, besides smoothness of the nonparametric quantile regression functions. The conditions on the kernel and the bandwidth are standard, and the smoothness condition is imposed on the population instead of the sample moment condition.

Proposition. Assume that:
1. The compactly supported variable x has absolutely continuous density $f(.)$ bounded away from zero and infinity.
2. The nonparametric quantile regression $\widehat{m}_{q_j}(.)$ is k times differentiable with respect to x on its support in a neighbourhood of the true parameter value θ_0.
3. The bandwidth sequence $\{b_n\}$ is such that $\frac{Nb_N^3}{\log N} \to \infty$ and $Nb_N^4 \to 0$.
4. The kernel function $K(.)$ of order k is a symmetric pdf with support $[-1, 1]$.

5. Each of the error terms $u_j, j = 1, 2$ and ε are independent, independent of each other and of finite variance. The error densities $f_{u_j}(.)$, $j = 1$, 2 and $f_\varepsilon(.)$ are absolutely continuous, and bounded away from zero and infinity.

6. The moment function $G^\tau(\theta, m)$ is twice continuously differentiable in (θ, m), and the derivatives $E\left[\frac{\partial G^\tau(\theta,m)}{\partial\theta}\right]$ and $E\left[\frac{\partial^2 G^\tau(\theta,m)}{\partial\theta^2}\right]$ are bounded, continuous and of full (column) rank in a neighbourhood of the true parameter value.

7. The weighting function $\widetilde{w}(x)$ is bounded, non-negative, and non-identically zero.

Then the asymptotic distribution of $\widehat{\theta}$ is

$$\sqrt{N}\left(\widehat{\theta} - \theta_0\right) \to^d \mathcal{N}(0, \Phi), \text{ with } \Phi = \left(H_1^\tau\right)^{-1}.V_1^\tau.\left(H_1^\tau\right)^{-1}$$

where H_1^τ is the derivative of the moment function with respect to the parameter θ, and V_1^τ is the variance of the moment conditions plus the variance of their functional derivatives evaluated at the direction $[\widehat{m} - m_0]$.

Proof. See Appendix.

Example (Location-scale shift continued). The moment condition for the location-scale shift model reads

$$G_N^\tau(\theta, \widehat{m}) \equiv \int \left(\tau - \frac{1}{2} + \frac{1}{2} sign\left[\widehat{m}_{q_1}(x_i) - \alpha_0 - y_i'\alpha_1 - (\beta_0 + y_i'\beta_1).\widehat{m}_{q_2}(x_i)\right]\right).z_i.w(x).dx$$

with $z_i = \begin{bmatrix} 1 & y_i & \widehat{m}_{q_2}(x) & y_i.\widehat{m}_{q_2}(x) \end{bmatrix}'$. As shown in the Appendix, the covariance matrix takes the sandwich form, $\Phi = \left(H_1^\tau\right)^{-1}.V_1^\tau.\left(H_1^\tau\right)^{-1}$, where

$$H_1^\tau(\theta_0, m_0) = -E\left[f_\varepsilon(0|z).z_i z_i'.\widetilde{w}(x)\right]$$

and

$$V_1^\tau = E[G_N^\tau(\theta_0, m_0).G_N^{\tau\prime}(\theta_0, m_0) + H_{2N}^\tau(\theta_0, m_0).H_{2N}^{\tau\prime}(\theta_0, m_0).[\widehat{m} - m_0]]$$

The outer terms have the same form as in standard quantile regression, whereas the inner term is somewhat different due to the H_{2N}^τ terms, which account for the preliminary nonparametric estimation step. This can be understood as the "price to be paid" for having used nonparametric estimation beforehand. The components of V_1^τ are

$$G_N^\tau(\theta_0, m_0) \equiv \int \left[\tau - \frac{1}{2} + \frac{1}{2} sign(\widehat{v}_{i0} - z_{i0}'\theta)\right].z_{i0}.\widetilde{w}(x_i).dx$$

with $v = \begin{bmatrix} m_{q_1}(x_1) & \dots & m_{q_1}(x_N) \end{bmatrix}'$, and

$$H_{2N}^\tau(\theta, m)[m - m_0] = \begin{bmatrix} \int \delta(\varepsilon_i).\{\zeta_1 - (\beta_0 + x_i'\beta_1).\zeta_2\}.\widetilde{w}(x_i).dx \\ \int x_i'.\delta(\varepsilon_i).\{\zeta_1 - (\beta_0 + x_i'\beta_1).\zeta_2\}.\widetilde{w}(x_i).dx \\ \int \left(\zeta_2.sign(\varepsilon_i) + m_{q_2}(x_i).\delta(\varepsilon_i).\{\zeta_1 - (\beta_0 + x_i'\beta_1).\zeta_2\}\right).\widetilde{w}(x_i).dx \\ \int \left(x_i'.\zeta_2.sign(\varepsilon_i) + x_i'.m_{q_2}(x_i).\delta(\varepsilon_i).\{\zeta_1 - (\beta_0 + x_i'\beta_1).\zeta_2\}\right).\widetilde{w}(x_i).dx \end{bmatrix}$$

where $\delta(.)$ is the delta function, obtained by differentiating the sign function, and

$$\widehat{\zeta}_j = \widehat{m}_{q_j}(x_j) - m_{q_j}(x_j) = \frac{1}{f_{u_j}(0)} \cdot \frac{1}{N_j} \sum_{i=1}^{N_j} [q_j - 1\{u_j \leq 0\}], j = 1, 2$$

■

Discussion of the Conditions.

I now discuss the conditions with special attention to the application. Condition 1 amounts to assuming that the logarithm of total expenditure has compact support and density bounded away from zero. But it is certainly bounded below from zero, so that it remains assuming it has an upper bound, a mild restriction. Conditions 2, 4, and 6 are standard regularity conditions, requiring smoothness of the preliminary nonparametric step. Note that smoothness is imposed on the population rather than the sample moment condition, a much weaker requirement. Condition 3 is needed in order to obtain uniform consistency in the first stage estimation. Condition 5 is a standard regularity condition in quantile regression, and is needed in order to obtain a well-defined covariance matrix. Condition 6 is immediately satisfied, since $0 < \widehat{m}_{q_j}(.) < 1$ for every x. Condition 7 is satisfied by a trimming function such as the indicator function or by a function of the density of x, which may be of use to eliminate unstable estimates.

3.3 Hypothesis Testing

Given estimates $\widehat{\theta}$ of θ, it may be of interest to test whether the model can be reduced. Assuming the existence of a $\widetilde{\theta} \in \Theta$ such that $\widehat{m}_{q_1}(x) = \widehat{m}_{q_2}(x)$, consider the general nonlinear null hypothesis

$$H_0 : \varphi(\theta) = 0$$

against the two-sided alternative

$$H_1 : \varphi(\theta) \neq 0$$

The rationale to obtain the asymptotic distribution of the test stems from the classical Wald testing principle. Once established that $\widehat{\theta} \rightarrow^d \mathcal{N}(\theta_0, \Phi)$, where $\rightarrow^d$ stands for convergence in distribution, under the null one obtains $\varphi(\widehat{\theta}) \rightarrow^d \mathcal{N}\left(0, D_\theta\varphi(\widehat{\theta}).\Phi.D_\theta\varphi(\widehat{\theta})'\right)$ using the delta method, so that

$$W = \varphi(\widehat{\theta}).\left(0, D_\theta\varphi(\widehat{\theta}).\Phi.D_\theta\varphi(\widehat{\theta})'\right)^{-1}.\varphi(\widehat{\theta})' \rightarrow^d \chi_r^2$$

In practice, when calculating the test statistic, Φ is replaced with a consistent estimate $\widehat{\Phi}$. This generates a decision rule according to which the null is rejected whenever W exceeds the $(1-\gamma)-$th percentile of the χ_r^2 distribution, given a significance level γ.

Example. As in the application, we consider the linear case. More specifically, given two estimated quantile curves of the Engel relation, one might be interested in testing whether the covariate x has any explanatory power in what regards the joint behaviour of $m_{q_1}(.)$ and $m_{q_2}(.)$ ie. $\alpha_1 = \beta_1 = 0$. Further, one might be interested in testing whether one quantile curve is just a shifted version of the other ie. $\alpha_1 = \beta_1 = 0$ and $\beta_0 = 1$. These tests can be written as

$$H_0 : R\theta - c = 0$$

against

$$H_1 : R\theta - c \neq 0$$

where R, θ, and c are of dimensions $(r \times p)$, $(p \times 1)$, and $(r \times 1)$, respectively, r being the number of restrictions involved. The test statistic then reduces to

$$W = (R\widehat{\theta} - c).(R\Phi R')^{-1}.(R\widehat{\theta} - c)' \to^d \chi_r^2$$

4 Application

In this section we apply the above methodology to compare conditional quantiles of the Engel relation, which expresses the budget share spent on a particular good (or group of goods) as a function of the expenditure level. Engel curves can be used to classify goods into luxuries, necessities, and inferior goods [see Deaton and Muellbauer (1980) for a thorough discussion]. Luxuries are goods that take up a larger budget share for better-off households, the opposite holding for necessities, whereas inferior goods are those the purchase of which decreases absolutely, and not only in relative terms, as the expenditure level increases.

To account for the fact that nonparametric tests might perform poorly even for samples of moderate size, the test statistics will be computed not only from the original sample, but also from its bootstrapped version. By doing so one should expect these problems to be mitigated, given the connection of the bootstrap and higher-order asymptotic expansions.

In what regards the Engel relation, one widely used functional form is the Working-Leser specification

$$w_{ij} = m_j(ln\ x_i) + u_{ij}$$

where w_{ij} is the budget share of the $j-th$ good for individual i, $ln\ x_i$ is the natural logarithm of total expenditure, and u_{ij} is an error term usually assumed to satisfy $E(u_{ij}|x_i) = 0$.

To illustrate the method the 1980-82 British Family Expenditure Survey (FES) dataset[2] is used, as in Blundell, Duncan and Pendakur (1998, BDP hereafter). For this study only the subset of households having only one child is considered, resulting in 594 observations[3]. The FES considers six broad categories of goods - food, domestic fuel, clothing, alcohol, transport and other goods. Total expenditure and income are measured in British pounds (£) per week. Table 1 displays some sample statistics.

TABLE 1 - Sample Statistics of the FES data

Variable	Mean	Std Deviation	JB p-value
food share	0.343	0.109	2×10^{-3}
fuel share	0.093	0.053	$< 2 \times 10^{-16}$
clothing share	0.106	0.098	$< 2 \times 10^{-16}$
alcohol share	0.067	0.069	$< 2 \times 10^{-16}$
transport share	0.138	0.109	$< 2 \times 10^{-16}$
other goods share	0.253	0.104	$< 2 \times 10^{-16}$
total expenditure	94.74	45.84	$< 2 \times 10^{-16}$
total net income	134.22	70.45	$< 2 \times 10^{-16}$
log total expenditure	4.46	0.41	6×10^{-7}
log net income	4.81	0.40	$< 2 \times 10^{-16}$

Note: The last column displays the p-values for the Jarque-Bera test statistic.

Following the tradition in the literature, BDP assumed $E(u_{ij}|x_i) = 0$ and estimated $m_j(.)$ nonparametrically. Although nonparametric estimation of Engel curves yields desirable estimates, this only enables one to address the study of the conditional mean of the Engel relation. Although theory is usually based on the behaviour of the conditional mean of the budget shares given the expenditure level, it would be desirable to study its entire distribution, in order to have a broader picture of the Engel relation. Nonparametric quantile regression allows one to investigate the behaviour of the distribution function of the budget shares conditional on total expenditure, which is likely to give insights on issues such as the behaviour of consumer heterogeneity conditioned on the expenditure level. To address this issue, in what follows we provide estimates of quantile Engel curves (QEC) comparing them in order to assess their behaviour

[2] Both the dataset and its documentation can be freely downloaded from the Journal of Applied Econometrics webpage.

[3] Most of the computations were done writing programs using the *quantreg library* by Roger Koenker, and the *sm library* by Adrian Bowman and Adelchi Azzalini, both for the freely available software R.

4.1 Estimating Mean Engel Curves

In this section we estimate Engel curves with the Working-Leser specification $w_{ij} = m_j(ln\ x_i) + u_{ij}$ under $E(u_{ij}|x_i) = 0$ for all the six categories of goods. Figures 1-6 display nonparametric Engel Curves estimated using the local linear estimator with the Gaussian kernel.

In this particular application, bandwidth choice is an especially critical issue. We started by using the cross-validation criterion available in the *sm library*. When using the full sample, the cross-validation criterion computed over a grid of points on the interval $[0, 30]$ was usually remarkably flat, with the location of (local) minima at very small values $-$ resulting in very wiggly estimates far from how one expects an Engel curve to behave $-$ and global minima at very large values. Given theses preliminary findings, we performed bandwidth choice after trimming the data. After defining μ and σ^2 as, respectively, the mean and variance of the logarithm of total expenditure of a household, we defined the first trimming function as $1\{\mu - 2\sigma <$ log(Total Expenditure)$< \mu + 2\sigma\}$, where $1\{.\}$ denotes the indicator function, and the second one as $1\{\mu - 1.5\sigma <$ log(Total Expenditure)$< \mu + 1.5\sigma\}$. Table 2 displays the results.

TABLE 2 - Bandwidths obtained using cross-validation

	Food	**Fuel**	**Clothing**	**Alcohol**	**Transport**	**Other Goods**
$h_{\text{full sample}}$	$-$	0.7251	$-$	0.4738	$-$	$-$
$h_{\mu\pm2\sigma}$	2.037	$-$	$-$	0.1703	0.4264	0.2071
$h_{\mu\pm1.5\sigma}$	0.1172	0.5204	0.1226	0.1723	0.2459	0.1507

Note: "$-$" denotes non-convergence of the cross-validation criterion to a global minimum in the interval $[0, 30]$.

Once obtained the bandwidth values, we experimented the values by estimating each Engel curve. In what regards the food Engel curve, the second and third schemes yielded, respectively, over- and undersmoothed Engel curves $-$ a bandwith value of one still showed a high degree of oversmoothing, whereas values between 0.5 and 0.75 showed a more satisfactory behaviour. For the fuel Engel curve, the values obtained produced fairly similar (both of them satisfactory) behaviour. The bandwidth obtained for the clothing Engel curve was expectedly too low, resulting in a very wiggly estimate, the same happening for the estimates of the "other goods" Engel curve. For the both the alcohol and the transport Engel curves, the largest values generated reasonable estimates.

Given the above findings I estimate the Engel curves using bandwidths 0.5 and 0.75 for all categories of goods. As one can see from Figures 1-6, in some cases one cannot even discern between the thin (bandwidth = 0.5) and the thick lines (bandwidth = 0.75). Actually, their behaviour tends to differ in a noticeable way for extreme observations only.

FIGURES 1-6 ABOUT HERE

As shown in Figures 1-6, the shape of the mean Engel curves for food, fuel, and clothing are downward-sloping, although the slope of the former is much more remarkable. This reflects the fact that they are considered necessities. Visual inspection of the alcohol mean Engel curve suggests a non-monotonic relation – slightly upward-sloping up to a threshold expenditure level, and slightly downward-sloping thereafter. On the other hand, the mean Engel curves for transport and "other goods" reflect the fact that they are luxury goods. The upward-sloping "other goods" Engel curve can be interpreted as "taste for variety/diversity", whereas the one for transport can be seen as a consequence of the fact that better-off households would be willing to use private means of transportation instead of public transport.

It is worth noticing, however, that due to scarcity of observations, estimates for the higher expenditure levels seem to be heavily influenced by outliers, resulting in somewhat abrupt changes in shape in this region. This provides justification for the introduction of trimming or weighting observations when estimating θ, so as to exclude those extreme observations and obtain estimates both more stable and in consonance of what one would expect from economic theory.

4.2 Estimating Quantile Engel Curves

Having investigated the behaviour of the conditional mean of the Engel relation we now investigate their quantiles. In what follows we consider a measure of dispersion of the distribution of budget shares conditional on expenditure levels by estimating its 75th and 25th quantiles[4]. Intuitively, this can be seen as a way to study heterogeneity of expenditure patterns among households. Following our notation, we set $q_1 = 0.75$ and $q_2 = 0.25$ and compare the quantile curves testing whether the location-scale shift can be explained by the expenditure level.

Bandwidth choice seems to play a much bigger role in quantile estimation than when estimating mean Engel curves. Bandwidth values between 0.5 and 0.75 usually resulted in extremely wiggly estimates, especially for the upper quantile curve $m_{q_1}(.)$. After experimenting with several values we decided to use bandwidth values 1 and 1.5. This amount of oversmoothing is certainly

[4]More extreme quantiles were also experimented, but were not as well behaved as the ones chosen. It seems much harder to find nonparametric estimates with desirable properties as one estimates more extreme quantiles. In particular, the amount of oversmoothing needed is remarkable, especially due to the scarcity of data in the upper quantiles. The figures with the quantile Engel curves illustrate this issue, whereas Welsh (1996) discuss similar findings in different applications.

responsible for delivering estimates much closer to linearity than one would expect but, to our knowledge, theory is silent about how to implement sensible choices in such a setting. Figures 7-12 display estimates for the 25th and 75th quantiles (together with the 50th quantile, for the sake of completeness), which were computed using the Epanechnikov kernel. Thin lines are estimates obtained using the smaller bandwidth (bandwidth =1), whereas thick lines are estimates obtained using larger bandwidth values (bandwidth=1.5). Vertical lines show the threshold values obtained by applying the trimming function $1\{\mu - 2\sigma <$ log(Total Expenditure)$< \mu + 2\sigma\}$ to the data, which eliminates 2 and 20 observations in the lowest and highest income levels, repectively. Interestingly, as opposed to parametric quantile regression models, the nonparametric quantile curves do not cross, for a given value of the bandwidth parameter.

FIGURES 7-12 ABOUT HERE

By estimating QEC's one is able to address heterogeneity issues instead of being restricted to measures of central tendency such as the conditional mean. In the case of the food share, for instance, one can observe that heterogenenity tends to slightly decrease with the expenditure level, instead of restricting oneself to the fact that the mean (or median) Engel curve is downward-sloping. For the fuel share, the mean Engel curve is also downward-sloping, but heterogeneity tends to decrease at a much faster rate.

In what regards the alcohol budget shares, inspection of the QEC's suggests that there is a non-monotone relation between the 25th and 75th quantile curves, whereas the transport quantile curves seem to be parallel. For the "other goods" relation, the quantile curves tend to drift apart as the expenditure level increases, suggesting an increase in heterogeneity. The more interesting behaviour of how considering only the mean Engel curve can be misleading seems to come for the clothing relation $-$ although the mean Engel curve is downward-sloping, the quantile curves are upward-sloping, and visual inspection of their shapes suggests that heterogeneity tends to increase with the expenditure level. Moreover, the slight departures from linearity in most of the cases can be attibuted to the great deal of oversmoothing imposed.

4.3 Comparing Quantile Engel Curves

Once estimated the pair of quantile curves, the next step is to investigate the relation between them. Our strategy is as follows. First, we consider $e_i = \widehat{m}_{q_1}(x_i) - \alpha_0 - x_i'\alpha_1 - (\beta_0 + x_i'\beta_1).\widehat{m}_{q_2}(x_i)$ ie. the location-scale shift model (or the "full model"). We then focus on whether this specification can be reduced to a more parsimonious one. Specifically, we are interested in testing whether

(i) $\alpha_1 = \beta_1 = 0$ ie. whether the expenditure level can explain the changes in heterogeneity; and (ii) $\alpha_1 = \beta_1 = 0$ and $\beta_0 = 1$ ie. whether the quantile curves can be seen as parallel shifts of one another.

The parameter vector $\theta \equiv (\alpha_0, \alpha_1, \beta_0, \beta_1)'$ is then estimated for each pair of quantile curves. For all of the following computations, the trimming function $1\{\mu - 2\sigma < \log(\text{Total Expenditure}) < \mu + 2\sigma\}$ (illustrated in Figures 7-12) is used. Given the need of nonparametric estimation in the preliminary step in order to obtain estimates of θ, one might well expect the asymptotic covariance matrix to be poorly estimated for small samples. In order to mitigate this inconveniences, we compare the asymptotic values with bootstrapped ones [see Horowitz (2001) for a thorough discussion and Buchinsky (1994, 1998) for its application to quantile regression] using the Design Matrix Bootstrap Estimator[5]. Table 3 displays the results for the full model.

The results in Table 3 show a number of interesting results:

LS1. for a relatively small sample such as ours (less than 600 observations), the bootstrap delivers much sharper standard errors than the asymptotic ones;
LS2. when taken individually, the coefficients $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ are not found to be statistically significant for the food, fuel, clothing, alcohol, and transport relationships. If this is confirmed when testing the null $H_0 : \boldsymbol{\alpha}_1 = \boldsymbol{\beta}_1 = 0$, this suggests that the expenditure level cannot explain neither location not scale shifts of the QEC's;
LS3. the coefficients $\boldsymbol{\alpha}_1, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1$ of the "other goods" relation are all found to be statistically significant, reflecting changes in heterogeneity as the expenditure level increases;
LS4. when taken individually, the estimates for $\boldsymbol{\beta}_0$ are not found to be statistically different from unity for *any* of the relationships. As a result, assuming the joint null $H_0 : \boldsymbol{\alpha}_1 = 0, \boldsymbol{\beta}_1 = 0$ is not rejected, one cannot reject that the QEC's (for clothing, alcohol, and transport) are parallel shifted versions of one another.

[5] As opposed to the Error Bootstrap Estimator, the Design Matrix Bootstrap Estimator does not rely on the independence assumption to yield a consistent estimator of the asymptotic covariance matrix [see Buchinsky (1998) for a thorough discussion]. The basic idea is to randomly sample (y_i^*, x_i^*) with replacement from the original sample in order to form B (artificial samples) of size N ($B = 1000$ in this application). After obtaining $\widehat{m}^*_{q_j,b}(.)$, $j = 1, 2$ and $\widehat{\theta}^*_b$, $b = 1, ..., B$, the bootstrap estimator of the covariance matrix Φ is obtained as $\widehat{\Phi}^B = \frac{N}{B} \sum_{b=1}^{B} (\widehat{\theta}^*_b - \overline{\theta}^*)(\widehat{\theta}^*_b - \overline{\theta}^*)'$. Although one could use $\overline{\theta}^* = \frac{1}{B} \sum_{b=1}^{B} \widehat{\theta}^*_b$ as the pivotal value, Buchinsky shows that plugging $\widehat{\theta}$ yields better small-sample properties, and this is followed in our application.

TABLE 3 - Parameter Estimates - Full Model

Variable	$\boldsymbol{\alpha}_0$	$\boldsymbol{\alpha}_1$	$\boldsymbol{\beta}_0$	$\boldsymbol{\beta}_1$
food share	0.0429	0.0000	1.2693	-0.0002
	(19.68)	(3.73)	(66.09)	(12.48)
	[0.144]	[0.046]	[0.631]**	[0.174]
fuel share	0.0001	-0.0029	1.9863	0.0388
	(298.75)	(67.49)	(5110.78)	(1156.41)
	[0.043]	[0.002]	[0.680]***	[0.030]
clothing share	0.0697	-0.0007	2.4797	0.0117
	(30.79)	(7.03)	(880.70)	(199.26)
	[0.006]***	[0.018]	[1.028]**	[0.663]
alcohol share	0.06776	-0.0000	2.4842	-0.0018
	(100.81)	(22.93)	(8037.51)	(1819.10)
	[0.029]**	[0.010]	[1.960]	[0.842]
transport share	0.1461	-0.0006	0.5930	0.0177
	(237.28)	(47.07)	(4258.30)	(848.47)
	[0.031]***	[0.005]	[0.743]	[0.110]
other goods share	-0.1900	-0.0175	2.7060	0.0971
	(595.81)	(134.94)	(3226.86)	(730.36)
	[0.171]	[0.004]***	[0.911]***	[0.019]***

Note: (i) Standard errors of the estimates appear in parentheses, whereas their bootstrapped versions ($B = 1000$) appear inside square brackets. (ii) "***","**", and "*" denote individual rejection of $\alpha_0 = 0$, $\beta_1 = 0$, $\alpha_1 = 0$, and $\beta_0 = 0$, at the 1, 5, and 10% significance levels, respectively.

In what follows we explore the plausibility of LS2-4 by testing whether the location-scale shift model can be reduced.

4.3.1 Can Expenditure Explain Shifts ?

Given the results for the full model, the first natural question is whether expenditure can explain the changes in heterogeneity. This can be done by considering the null $H_0 : \boldsymbol{\alpha}_1 = 0, \boldsymbol{\beta}_1 = 0$ of "pure" location-scale shifts against a two-sided alternative. Table 4 displays the results.

TABLE 4 - Testing for Pure Location-Scale Shifts

Variable	$\mathbf{H}_0$: $\boldsymbol{\alpha}_1 = \mathbf{0}, \boldsymbol{\beta}_1 = \mathbf{0}$
food share	$1.5 \text{ x } 10^{-5}$
	[1.75]
fuel share	$3.3 \text{ x } 10^{-8}$
	[0.002]
clothing share	$3.4 \text{ x } 10^{-9}$
	[1.70]
alcohol share	$9.4 \text{ x } 10^{-13}$
	[1.75]
transport share	$4.3 \text{ x } 10^{-10}$
	[1.34]
other goods share	$1.8 \text{ x } 10^{-8}$
	$[13.47]^{***}$

Note: (i) Bootstrapped versions ($B = 1000$) of the test statistics appear inside square brackets. (ii) Rejection of the null, should it occur, is denoted by *, **, and ***, which denote rejection at the 10, 5 and 1% significance level, respectively.

From the results of Table 4, the only category of goods for which the null of the pure location-scale shift model is rejected is the "other goods" relation, for which $-$ as already shown in Table 3 $-$ the expenditure level plays a key role in explaining the location-scale shift. For the remaining relations, the main lesson from Table 4 is that changes in heterogeneity $-$ should they occur $-$ cannot be explained by the expenditure level.

4.3.2 Are There Any Parallel Quantile Curves ?

The second natural question one might ask is whether the quantile curves can be seen as parallel shifts of one another ie. the null $H_0 : \boldsymbol{\alpha}_1 = 0, \boldsymbol{\beta}_0 = 1, \boldsymbol{\beta}_1 = 0$ against a two-sided alternative. We investigate the null of parallel quantile curves for the remaining categories other than the "other goods" one. Table 5 displays the results.

TABLE 5 - Testing for Parallel Quantile Engel Curves

Variable	$\mathbf{H}_0$: $\boldsymbol{\alpha}_1 = \mathbf{0}, \boldsymbol{\beta}_0 = \mathbf{1}, \boldsymbol{\beta}_1 = \mathbf{0}$
food share	1.5 x 10^{-3}
	[2.95]
fuel share	2.5 x 10^{-5}
	[0.016]
clothing share	2.5 x 10^{-6}
	[5.44]
alcohol share	3.1 x 10^{-8}
	[3.84]
transport share	8.6 x 10^{-9}
	[0.61]

Note: (i) Bootstrapped versions ($B = 1000$) of the test statistics appear inside square brackets. (ii) Rejection of the null, should it occur, is denoted by *, **, and ***, which denote rejection at the 10, 5 and 1% significance level, respectively.

From the results in Table 5 one cannot reject the null according to which the quantile curves are parallel (location) shifts of one another, considering the standard significance levels. In other terms, one cannot reject the null of a "pure" location-scale shift model. We then estimate this specification for all categories of goods for which the null of parallel curves is not rejected. For each pair of quantile curves we consider $e_i = \widehat{m}_{q_1}(x_i) - \alpha_0 - \beta_0.\widehat{m}_{q_2}(x_i)$ and the corresponding estimation problem. Table 6 displays the results.

The estimates in Table 6 once again allow us to confront asymptotic and bootstrapped standard errors $-$ which are connected to higher-order asymptotics, thus reflecting more accurately the properties of the estimators. In general, asymptotic values would lead to rejection of the null of parallel quantile curves, except for the transport relation, whereas bootstrapped estimates would not reject the null, suggesting no changes in heterogeneity as expenditure increases.

TABLE 6 - Parameter Estimates - Pure Location-Scale Shift Model

Variable	α_0	β_0
food share	0.0430	1.2685
	(0.0644)	(0.0039)***
	[0.0168]**	[0.9795]
fuel share	-0.0126	2.1562
	(0.2089)	(0.2270)***
	[0.0518]	[1.5919]
clothing share	0.0668	2.5290
	(0.1125)	(0.2381)***
	[0.0159]***	[1.8469]
alcohol share	0.0677	2.4762
	(0.1746)	(0.2771)***
	[0.0221]***	[1.8461]
transport share	0.1433	0.6744
	(0.0784)*	(0.2055)***
	[0.0797]*	[0.3246]***

Note: (i) Standard errors of the estimates appear in parentheses, whereas their bootstrapped versions ($B = 1000$) appear inside square brackets. (ii) "***","**", and "*" denote individual rejection of $\alpha_0 = 0$ and $\beta_0 = 0$, at the 1, 5, and 10% significance levels, respectively.

5 Conclusion

In this paper we compare nonparametrically estimated conditional quantiles by linking them using a parametric transformation and testing whether they could be related as "location-scale shifts" of one another. Being nonparametric, the quantile curves are less prone to misspecification than standard parametric methods, besides offering more flexibility.

The method is illustrated with a microeconometric application. Using data from a sample of UK households from the FES (Family Expenditure Survey), we estimate and compare quantile Engel curves. That is, instead of restricting ourselves to the (mean) Engel curve, we consider the Engel relation in a much broader sense by analysing features of the distribution function of budget shares conditioned on (the logarithm of) the expenditure level. This enables one to test the relation between different quantiles in order to study how household heterogeneity evolves as the expenditure level increases, as well as whether changes in heterogeneity can be explained by the expenditure level.

When applying the methodology to the FES dataset, the main findings are

(i) for food, fuel, clothing, alcohol, and transport, the expenditure level cannot explain the shifts of the quantile budget shares;

(ii) for the same five categories of goods in (i) for which the expenditure cannot explain the shift in the quantile curves, the null of parallel quantile curves cannot be rejected, suggesting that heterogeneity does not change significantly as the expenditure level increases;

(iii) for the "other goods" category, expenditure can explain the both the loication and scale shifts of the quantile curves, thus reflecting the increase in heterogeneity for better-off households.

The methods here presented combine results from quantile regression and smoothing techniques, and can be extended to consider other problems of interest in the quantile literature, such as quantile treatment effects, as in the application of Koenker and Xiao (2002). This is left for further research.

6 Appendix 1: The CLK Theorems

In what follows we denote the data by $\{Z_i\}, i = 1, ..., n$, $Z_i = (x_i, y_i)$, with support $Z \subset \mathbb{R}^d$, the finite-dimensional compact parameter set by $\Theta \subset \mathbb{R}^k$, the infinite-dimensional parameter set by $\mathcal{M}$, assuming also that the latter is a vector space of functions endowed with a pseudo-metric $||.||_{\mathcal{M}}$ (meaning that $||a(x) - b(x)||_{\mathcal{M}} = 0$ does not necessarily imply $a = b$). The true finite- and infinite-dimensional parameters are denoted respectively by $\theta_0 \in \Theta$ and $m_0 \in \mathcal{M}$ and, given a matrix A, its norm is defined as $||A|| = (tr(A'WA))^{1/2}$, with W being a symmetric positive definite matrix.

We consider a non-random measurable vector-valued function $M : \mathbb{R}^k \times \mathcal{M} \rightarrow \mathbb{R}^p$, $p \geq k$, with $G(\theta, m) = E[g(Z_i, \theta, m)]$ such that $||G(\theta, m_0)||$ attains its minimum at $\theta = \theta_0$, as well as a random vector-valued function $G_N : \mathbb{R}^k \times \mathcal{M} \rightarrow \mathbb{R}^p$ with $G_N(\theta, m) = N^{-1} \sum\limits_{i=1}^{N} g(Z_i, \theta, m)$, and $G_N(\theta, m_0)$ close to $G(\theta, m_0)$. Note that smoothness assumptions are imposed only on $G(\theta, m)$ at (θ_0, m_0).

The estimation problem under consideration can be stated as follows: The final aim is to obtain a two-stage estimator $\widehat{\theta}$ of a finite-dimensional parameter $\theta_0 \in \Theta \subset \mathbb{R}^p$ depending on some preliminary nonparametric estimates. In the preliminary stage, infinite-dimensional parameters m are estimated, say, nonparametric density or regression estimators. In the final stage, one obtains $\widehat{\theta}$ of θ_0 that depends on the previously estimated $\widehat{m}$ and approximately solves the problem $\min_{\theta \in \Theta} ||G_N(\theta, \widehat{m})||$. To develop the asymptotic theory, one will need to differentiate $G(\theta, m)$ with respect to its arguments. Differentiating with respect to θ is straightforward, but a different notion of differentiability is needed for the infinite-dimensional part. For any $\theta \in \Theta$, $G(\theta, m)$ is said to be pathwise differentiable at at m in the direction $[\overline{m} - m]$ if $\{m + \tau(\overline{m} - m) : \tau \in [0, 1]\} \subset \mathcal{M}$, and $\lim_{\tau \rightarrow 0}[G(\theta, m + \tau(\overline{m} - m)) - G(\theta, m)]/\tau$ exists, with the limit being the pathwise derivative.

The conditions needed require the estimator $\widehat{\theta}$ to be an approximate minimizer of the sample criterion function, to be unique in a neighbourhood of the true parameter value θ_0, the uniform consistency of the preliminary nonparametric estimator $\widehat{m}$ (under the pseudo-metric $||.||_{\mathcal{M}}$), as well as a sort of stochastic equicontinuity condition.

CLK's Theorem 1. Suppose that $\theta_0 \in \Theta$ satisfies $G(\theta_0, m_0) = 0$, and that

1.1 - $||G_N(\widehat{\theta}, \widehat{m})|| \leq \inf_{\theta \in \Theta} ||G_N(\theta, \widehat{m})|| + o_p(1)$.

1.2 - For all $\delta > 0$, there exists $\epsilon(\delta) > 0$ such that $\inf_{||\theta - \theta_0|| > \delta_n} ||G_N(\theta, m_0)|| \geq \epsilon(\delta) > 0$

1.3 - Uniformly for all $\theta \in \Theta$, $G_N(\theta, m)$ is continuous [with respect to the metric $||.||_{\mathcal{M}}$] in m at $m = m_0$

1.4 - $||\widehat{m} - m_0||_{\mathcal{M}} = o_p(1)$

1.5 - For all sequences of positive numbers $\{\delta_n\}$ with $\delta_n = o(1)$,

$$\sup \frac{||G_N(\theta, m) - G(\theta, m)||}{1 + ||G_N(\theta, m)|| + ||G(\theta, m)||} = o_p(1) \qquad (2)$$

where the sup is taken over $||\theta - \theta_0|| < \delta_n$, $||m(x) - m_0(x)||_{\mathcal{M}} < \delta_n$
Then,

$$\widehat{\theta} - \theta_0 = o_p(1)$$

It is worth noticing that Condition (1.5) is implied by an alternative replacing its fraction with $||G_N(\theta, m) - G(\theta, m)||$

For asymptotic normality, further requirement, besides the consistency of the parameter vector, include regularity conditions on the derivatives of $G(\theta, m_0)$.

CLK's Theorem 2. Suppose that $\theta_0 \in int(\Theta)$ satisfies $G(\theta_0, m_0) = 0$, $\widehat{\theta} - \theta_0 = o_p(1)$, and that

2.1 - $||G_N(\widehat{\theta}, \widehat{m})|| = \inf_{||\theta - \theta_0|| < \delta_n} ||G_N(\theta, \widehat{m})|| + o_p(N^{-1/2})$ for some positive sequence $\delta_n = o(1)$.

2.2 - The ordinary partial derivative in θ of $G(\theta, m_0)$, denoted $H_1(\theta, m_0)$, exists in a neighbourhood of θ_0, is continuous at $\theta = \theta_0$, and is of full (column) rank

2.3 - The pathwise derivative $H_2(\theta_0, m_0)$ of $G(\theta, m_0)$ exists in all directions $[m - m_0]$ and satisfies

$$(i)\ ||G(\theta, m) - G(\theta, m_0) - H_2(\theta, m_0)[m - m_0]|| \leq c.||m - m_0||^2_{\mathcal{M}}$$

$$(ii)\ ||H_2(\theta, m_0)[m - m_0] - H_2(\theta_0, m_0)[m - m_0]|| \leq c'||\theta - \theta_0||.o(1)$$

for all θ with $||\theta - \theta_0|| = o(1)$, all m with $||m - m_0||_{\mathcal{M}} = o(1)$, and constants $c, c' \in [0, \infty)$.

2.4 - With probability approaching one (wpa1), $\widehat{m} \in \mathcal{M}$, and $c.||\widehat{m} - m_0||^2_{\mathcal{M}} = o_p(N^{-1/2})$ uniformly over θ with $||\theta - \theta_0|| = o(1)$.

2.5 - For all sequences of positive numbers $\{\delta_n\}$ with $\delta_n = o(1)$,

$$\sup \frac{\sqrt{N}||G_N(\theta, m) - G(\theta, m) - G_N(\theta_0, m_0)||}{1 + \sqrt{N}\left[||G_N(\theta, m)|| + ||G(\theta, m)||\right]} = o_p(1) \qquad (3)$$

where the sup is taken over $||\theta - \theta_0|| < \delta_n$, $||m(x) - m_0(x)||_{\mathcal{M}} < \delta_n$

2.6 - For some finite matrix V_1,

$$\sqrt{N}\left(G_N(\theta_0, m_0) + H_2(\theta_0, m_0)[\widehat{m} - m_0]\right) \to^d \mathcal{N}(0, V_1)$$

Then,

$$\sqrt{N}\left(\widehat{\theta} - \theta_0\right) \to^d \mathcal{N}(0, \Omega)\text{, where } \Omega = (H_1'WH_1)^{-1}.H_1'WV_1WH_1.(H_1'WH_1)^{-1}$$

Remark. Condition 2.5 is implied by the following condition

$$\sup \sqrt{N}||G_N(\theta, m) - G(\theta, m) - G_N(\theta_0, m_0)|| = o_p(1)$$

with the sup taken over the same region as before. To verify Condition 2.5, CLK establish a result relying on more primitive assumptions.

CLK's Theorem 3. Suppose that each component g_j of $g = (g_1, ..., g_l)'$ takes the form $g_j(z, \theta, m) = g_{cj}(z, \theta, m) + g_{lcj}(z, \theta, m)$, and satisfies

3.1 - $g_{cj}(z, \theta, m)$ is Hölder-continuous with respect to (θ, m) in the sense

$$|g_{cj}(z, \theta_1, m_1) - g_{cj}(z, \theta_2, m_2)| \leq b_j(z). \left(||\theta_1 - \theta_2||^{s_{1j}} + ||m_1 - m_2||_{\mathcal{M}}^{s_j}\right)$$

for $s_{1j}, s_j \in (0, 1]$, a measurable function $b_j(.)$ with $E[b_j(.)]^r < \infty$, $r \geq 2$

3.2 - $g_{lcj}(., \theta, m)$ is locally uniformly $L_r(P)-$continuous ($r \geq 2$) with respect to (θ, m) in the following sense

$$\left(E\left[\sup_{(\theta^*, m^*): ||\theta^* - \theta|| < \delta, ||m^* - m||_{\mathcal{M}} < \delta} |g_{lcj}(Z, \theta^*, m^*) - g_{lcj}(Z, \theta, m)|\right]\right)^{1/r} \leq K_j.\delta^{s_j}$$

for all $(\theta, m) \in \Theta \times \mathcal{M}$, all small positive value $\delta = o(1)$, and for some constants $s_j \in (0, 1]$, $K_j > 0$.

3.3 - Θ is a compact subset of $\mathbb{R}^k$, and $\int_0^\infty \sqrt{\log N(\epsilon^{1/s_j}, \mathcal{M}, ||.||_{\mathcal{M}})} d\epsilon < \infty$ for $j = 1, ..., l$.

Then, for all positive δ_n such that $\delta_n = o(1)$,

$$\sup \sqrt{N}||G_N(\theta, m) - G(\theta, m) - G_N(\theta_0, m_0)|| = o_p(1)$$

where the sup is taken over $||\theta - \theta_0|| < \delta_n$, $||m(x) - m_0(x)||_{\mathcal{M}} < \delta_n$.

Remark. It is worth mentioning that, as discussed in CLK, Condition 3.2 allows for discontinuous moment functions such as sign and indicator functions of (θ, m) (CLK, pp. 1598). Moreover, $s_j = 1$ correspond to the regular case, when Condition 3.3 is readily verified. If $s_j < 1$, then a higher degree of smoothness of m is required, so that the class of admissible functions $\mathcal{M}$ has to be restricted (CLK, pp. 1598).

CLK also show that the bootstrap can provide consistent estimates of the asymptotic variance. In what follows, let the superscript * denote a probability or moment computed under the bootstrap distribution conditional on the original data set $\{Z_i\}_{i=1}^N$.

CLK's Theorem B. Assuming that the conditions of CLK's Theorem 2 hold with "in probability" replaced by "almost surely"; that conditions 2.2 and 2.3 hold with m_0 replaced by m with $||m - m_0|| = o(1)$; that $H_1(\theta, m)$

is continuous [with respect to the metric $||.||_{\mathcal{M}}$] in m at (θ_0, m_0). Assuming further that the following bootstrap conditions hold

4B. With P^*-probability tending to one, $\widehat{m}^* \in \mathcal{M}$, and $||\widehat{m}^* - \widehat{m}||_{\mathcal{M}} = o_{P^*}(N^{-1/4})$

6B. $\sqrt{N}.\left(G_N^*(\widehat{\theta}, \widehat{m}) - G_N(\widehat{\theta}, \widehat{m}) + H_2(\widehat{\theta}, \widehat{m})[\widehat{m}^* - m_0]\right) = \mathcal{N}(0, V_1) + o_{P^*}(1)$

Then, $\sqrt{N}(\widehat{\theta}^* - \theta_0)$ converges in distribution to a $\mathcal{N}(0, \Omega)$ distribution in P^*-probability.

7 Appendix 2: Verifying Conditions

In what follows we derive the asymptotic covariance matrix of $\widehat{\theta}$ and provide sufficient conditions for asymptotic normality to hold. To estimate the parameter vector θ we minimize the absolute distance between $m_{q_1}(.)$ and $m_{q2}(.)$ allowing for unequal weights or trimming through a suitably chosen weight function.

In what follows we compute the covariance matrix of $\widehat{\theta}$. First note that

$$G_N^{\tau}(\theta,\widehat{m}) = \int \left(\tau - \frac{1}{2} + \frac{1}{2} sign\left[\widehat{m}_{q_1}(x_i) - \phi(\theta, \widehat{m}_{q_2}(x_i), y_i)\right]\right).D_\theta\phi(\theta,\widehat{m}_{q_2}(x_i),y_i).\widetilde{w}(x_i).dx$$

The estimator of θ is then defined as

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} G_N^{\tau}(\theta,\widehat{m})'.\Omega^{-1}.G_N^{\tau}(\theta,\widehat{m})$$

where Ω^{-1} is a weighting matrix.

Note that

$$G_N^{\tau}(\theta_0,m_0) = \int \left(\tau - \frac{1}{2} + \frac{1}{2} sign\left[m_{0q_1}(x_i) - \phi(\theta_0, m_{0q_2}(x_i), y_{0i})\right]\right).D_\theta\phi(\theta_0,\widehat{m}_{0q_2}(x_i),y_{0i}).\widetilde{w}(x_i).dx$$

converging to

$$G^{\tau}(\theta_0,m_0) = E\left(\left[\tau - \frac{1}{2} + \frac{1}{2} sign(e_{0i})\right].D_\theta\phi(\theta_0,\widehat{m}_{0q_2}(x_i),y_{0i}).\widetilde{w}(x_i)\right)$$

We start by verifying consistency.

Condition 1.2 is verified by showing that $G^{\tau}(\theta,m)$ is zero at $(\theta,m) = (\theta_0,m_0)$, and that $H_1 < 0$ at $\theta=\theta_0$ on a non-negligible set. But

$$\begin{aligned} H_{1N}^{\tau}(\theta_0,m_0) &= -\frac{1}{2}\int \delta(m_{0q_1}(x_i) - \phi(\theta_0, m_{0q_2}(x_i), y_{0i})). \\ &\quad .D_\theta\phi(\theta_0,\widehat{m}_{0q_2}(x_i),y_{0i}).D_\theta\phi(\theta_0,\widehat{m}_{0q_2}(x_i),y_{0i})'.\widetilde{w}(x_i).dx \\ &\quad + \int\left[\tau - \frac{1}{2} + \frac{1}{2} sign(m_{0q_1}(x_i) - \phi(\theta_0, m_{0q_2}(x_i), y_{0i}))\right]. \\ &\quad .D_{\theta\theta}\phi(\theta_0,\widehat{m}_{0q_2}(x_i),y_{0i}).\widetilde{w}(x_i).dx \end{aligned}$$

which converges to

$$H_1^{\tau}(\theta_0,m_0) = -\frac{1}{2}E\left[f_e(0|z).D_\theta\phi(\theta_0,\widehat{m}_{0q_2}(x_i),y_{0i}).D_\theta\phi(\theta_0,\widehat{m}_{0q_2}(x_i),y_{0i})'.\widetilde{w}(x)\right]$$

which is negative definite given $E\left[D_\theta\phi(\theta_0, \widehat{m}_{0q_2}(x_i), y_{0i})\right]$ bounded, continuous, and of full (column) rank in a neighbourhood of the true parameter value (Condition 6), a bounded weighting function (Condition 7), and an error density bounded away from zero (Condition 5)

Condition 1.3 follows from smoothness of m (Condition 2).

Condition 2.4 of CLK is satisfied by Assumptions 3 and 4. In particular, by $K(.)$ is a symmetric pdf with support $[-1, 1]$ - for instance, the Epanechnikov kernel.

Condition 1.5 will be dealt with when discussing Condition 2.5.

To verify CLK's Theorem 2, note that one needs H_1, and V_1, the asymptotic covariance matrix described in condition 2.6. The former was obtained as

$$H_1^\tau(\theta_0, m_0) = -E\left[f_\varepsilon(0|z).D_\theta\phi(\theta_0, \widehat{m}_{0q_2}(x_i), y_{0i}).D_\theta\phi(\theta_0, \widehat{m}_{0q_2}(x_i), y_{0i})'.\widetilde{w}(x)\right]$$

Sufficient conditions for Condition 2.2 to be satisfied are $E\left[D_\theta\phi(\theta, \widehat{m}_{q_2}(x_i), y_i).D_\theta\phi(\theta, \widehat{m}_{q_2}(x_i), y_i)'\right]$ being of full rank in a neighbourhood of the true parameter value, and the error density being bounded away from zero (Condition 6).

Condition 2.3 will be dealt with when considering Condition 2.6.

Condition 2.4 of CLK is satisfied by Assumptions 3 and 4. In particular, by $K(.)$ is a symmetric pdf with support $[-1, 1]$ - for instance, the Epanechnikov kernel.

To verify Condition 2.5 we use CLK's Theorem 3. To check Conditions 3.1 and 3.2 of this result. First, define $G_N^\tau(\theta, m) = \frac{1}{N}\sum_{i=1}^{N} g_i(\theta, m)$, Then note that

$$\begin{aligned}
|g_j(\theta^*, m^*) - g_j(\theta, m)| &= \left| \begin{array}{c} \left[\tau - \frac{1}{2} + \frac{1}{2}sign(e_j^*)\right].D_\theta\phi(\theta^*, m_{q_2}^*(x_i), y_j^*).\widetilde{w}(x_j) \\ -\left[\tau - \frac{1}{2} + \frac{1}{2}sign(e_j)\right].D_\theta\phi(\theta, m_{q_2}(x_i), y_j).\widetilde{w}(x_j) \end{array} \right| \\
&= \left|\left[\tau - \frac{1}{2} + \frac{1}{2}sign(e_j^*)\right].D_\theta\phi(\theta^*, m_{q_2}^*(x_i), y_j^*).\widetilde{w}(x_j)\right| \\
&\quad + \left|\left[\tau - \frac{1}{2} + \frac{1}{2}sign(e_j)\right].D_\theta\phi(\theta, m_{q_2}(x_i), y_j).\widetilde{w}(x_j)\right| \\
&= \left|\left[\tau - \frac{1}{2} + \frac{1}{2}sign(e_j^*)\right]\right| + \left|D_\theta\phi(\theta^*, m_{q_2}^*(x_i), y_j^*)\right| + |\widetilde{w}(x_j)| \\
&\quad + \left|\left[\tau - \frac{1}{2} + \frac{1}{2}sign(e_j)\right]\right| + |D_\theta\phi(\theta, m_{q_2}(x_i), y_j)| + |\widetilde{w}(x_j)|
\end{aligned}$$

The first and fourth terms are bounded by definition, whereas the third and sixth term are bounded by Condition 7. The second and fifth terms are bounded by Condition 6. It then follows that Conditions 3.1 and 3.2 of CLK are satisfied for $s_j = 1$, $K_j < \infty$. In what concerns the remaining condition we refer to the remark after CLK's Theorem 3 (or to CLK, pp. 1598) and argue that since $s_j = 1$, Condition 3.3 is satisfied. As a result, Condition 2.5 is satisfied.

To verify Condition 2.6 one needs to obtain V_1. First consider

$$G_N^\tau(\theta, m) = \int \left[\tau - \frac{1}{2} + \frac{1}{2} sign(e_i)\right].D_\theta\phi(\theta, m_{q_2}(x_i), y_i).\widetilde{w}(x_i).dx$$

Now recall that $sign(e_{i0}) = sign(m_{0q_1}(x_i) - \phi(\theta_0, m_{0q_2}(x_i), y_{0i}))$ is a binomial random variable with zero mean and unit variance. Thus, given e_i independent, of finite variance, and independent of x, applying a Central Limit Theorem results in asymptotic normality of this term. The second term concerning 2.6 is $H_2^\tau(\theta_0, m_0)[\widehat{m} - m_0]$. Note that

$$H_{2N}^\tau(\theta, m)[m - m_0] = \frac{\partial G_N^\tau(\theta, m + \gamma\zeta)}{\partial\gamma}|_{\gamma=0}$$

where $\zeta_j = m_{q_j} - m_{0q_j}, j = 1, 2$. But

$$\begin{aligned} G_N^\tau(\theta, m + \gamma\zeta) \quad = \quad & \int \left[\tau - \frac{1}{2} + \frac{1}{2}.sign(m_{q_1}(x_i) + \gamma\zeta_1 - \phi(\theta, m_{q_2}(x_i) + \gamma\zeta_2, y_i))\right] \\ & .D_\theta\phi(\theta, m_{q_2}(x_i) + \gamma\zeta_2, y_i).\widetilde{w}(x_i).dx \end{aligned}$$

Differentiating with respect to γ yields

$$\begin{aligned} & \frac{\partial G_N^\tau(\theta, m + \gamma\zeta)}{\partial\gamma} \\ = \; & \int \tfrac{1}{2}.\delta(m_{q_1}(x_i) + \gamma\zeta_1 - \phi(\theta, m_{q_2}(x_i) + \gamma\zeta_2, y_i)).\{\zeta_1 - D_\gamma\phi(\theta, m_{q_2}(x_i) + \gamma\zeta_2, y_i).\zeta_2\} \\ & .D_\theta\phi(\theta, m_{q_2}(x_i) + \gamma\zeta_2, y_i).\widetilde{w}(x_i).dx \\ & + \int \left[\tau - \tfrac{1}{2} + \tfrac{1}{2}.sign(m_{q_1}(x_i) + \gamma\zeta_1 - \phi(\theta, m_{q_2}(x_i) + \gamma\zeta_2, y_i))\right] \\ & .D_{\gamma\theta}\phi(\theta, m_{q_2}(x_i) + \gamma\zeta_2, y_i).\zeta_2.\widetilde{w}(x_i).dx \end{aligned}$$

setting $\gamma = 0$ results in

$$\begin{aligned} H_{2N}^\tau(\theta, m)[m - m_0] = \; & \int \tfrac{1}{2}.\delta(m_{q_1}(x_i) - \phi(\theta, m_{q_2}(x_i), z_i)).\{\zeta_1 - D_\gamma\phi(\theta, m_{q_2}(x_i), z_i).\zeta_2\} \\ & .D_\theta\phi(\theta, m_{q_2}(x_i), z_i).\widetilde{w}(x_i).dx \\ & + \int \left[\tau - \tfrac{1}{2} + \tfrac{1}{2}.sign(m_{q_1}(x_i) - \phi(\theta, m_{q_2}(x_i), z_i))\right] \\ & .D_{\gamma\theta}\phi(\theta, m_{q_2}(x_i), z_i).\zeta_2.\widetilde{w}(x_i).dx \end{aligned}$$

evaluating at $(\theta, m) = (\theta_0, m_0)$ in the direction $\widehat{m} - m_0$ yields

$$\begin{aligned} H_{2N}^\tau(\theta_0, m_0)[m - m_0] = \; & \int \tfrac{1}{2}.\delta(e_{i0}).\{\zeta_1 - D_\gamma\phi(\theta, m_{q_2}(x_i), y_i).\zeta_2\}.D_\theta\phi(\theta, m_{q_2}(x_i), y_i).\widetilde{w}(x_i).dx \\ & + \int \left[\tau - \tfrac{1}{2} + \tfrac{1}{2}.sign(e_{i0})\right].D_{\gamma\theta}\phi(\theta, m_{q_2}(x_i), y_i).\zeta_2.\widetilde{w}(x_i).dx \end{aligned}$$

but the second term is zero, so that

$$H_{2N}^{\tau}(\theta_0, m_0)[\widehat{m}-m_0] = \int \frac{1}{2}.\delta(e_{i0}).\{\zeta_1 - D_\gamma\phi(\theta, m_{q_2}(x_i), y_i).\zeta_2\}.D_\theta\phi(\theta, m_{q_2}(x_i), y_i).\widetilde{w}(x_i).dx$$

It then follows that

$$G_N^{\tau}(\theta_0, m_0) = \int \left(\tau - \frac{1}{2} + \frac{1}{2} sign\left[m_{0q_1}(x_i) - \phi(\theta_0, m_{0q_2}(x_i), y_{0i})\right]\right).D_\theta\phi(\theta_0, \widehat{m}_{0q_2}(x_i), y_{0i}).\widetilde{w}(x_i).dx$$

$$H_{2N}^{\tau}(\theta_0, m_0)[\widehat{m}-m_0] = \int \frac{1}{2}.\delta(e_{i0}).\{\widehat{\zeta}_1 - D_\gamma\phi(\theta, m_{q_2}(x_i), y_i).\widehat{\zeta}_2\}.D_\theta\phi(\theta, m_{q_2}(x_i), y_i).\widetilde{w}(x_i).dx$$

where

$$\widehat{\zeta}_j = \widehat{m}_{q_j}(x_j) - m_{q_j}(x_j) = \frac{1}{f_{u_j}(0)}.\frac{1}{N_j}\sum_{i=1}^{N_j}\left[q_j - 1\{u_j \leq 0\}\right], j = 1, 2$$

the latter being the local Bahadur representation of Chaudhuri (1991). The term $G_N^{\tau}(\theta_0, m_0) + H_{2N}^{\tau}(\theta_0, m_0)[\widehat{m} - m_0]$ is asymptotically normal with zero mean and finite variance under regularity conditions such as $f_{u_j}(.)$ and $f_\varepsilon(.)$ being bounded away from both zero and infinity and *i.i.d.* observations. The corresponding covariance matrix V_1^{τ} is then obtained as

$$V_1 = E\left(G_N^{\tau}(\theta_0, m_0).G_N^{\tau\prime}(\theta_0, m_0) + H_{2N}^{\tau}(\theta_0, m_0).H_{2N}^{\tau\prime}(\theta_0, m_0).[\widehat{m} - m_0]\right)$$

Proposition 1 summarizes the results.
■

In what concerns verifying the validity of CLK's Theorem B, the reasoning is analogous. In particular, Condition 4B can be verified in the same fashion as Condition 4 for a variety of kernel density and regression estimators. In this case, it is also happens that $\sqrt{N}.H_2^{\tau}(\widehat{\theta}, \widehat{m})[\widehat{m}^* - m_0]$ is approximately a sum of mean zero and independent random variables (under P^*) which can be expected to satisfy a central limit theorem (CLK, pp. 1596). ■

Example (Location-scale shift). Let

$$\widehat{m}_{q_1}(x_i) = \alpha_0 + x_i'\alpha_1 + (\beta_0 + x_i'\beta_1).\widehat{m}_{q_2}(x_i) + e_i$$

Define

$$z = \begin{bmatrix} 1 & ... & 1 \\ x_1 & ... & x_N \\ m_{q_2}(x_1) & ... & m_{q_2}(x_N) \\ x_1.m_{q_2}(x_1) & ... & x_N.m_{q_2}(x_N) \end{bmatrix}'$$

$v = \left[\begin{array}{ccc} m_{q_1}(x_1) & ... & m_{q_1}(x_N) \end{array} \right]'$, $\varepsilon_i = v_i - z_i'\theta$, and $\widetilde{w}(x) = \left[\begin{array}{ccc} \frac{w(x_1)}{f(x_1)} & ... & \frac{w(x_N)}{f(x_N)} \end{array} \right]'$, with $f(.)$ being the density of x. We also use hats to denote estimated quantities, and z_i', v_i, and $\widetilde{w}(x_i)$ to represent the $i-th$ row of the respective matrices. In what follows we compute the covariance matrix of $\widehat{\theta}$. First note that

$$G_N^\tau(\theta, \widehat{m}) \equiv \int \left[\tau - \frac{1}{2} + \frac{1}{2} sign(\widehat{v}_i - \widehat{z}_i'\theta) \right] .z_i.\widetilde{w}(x_i).dx$$

so that

$$G_N^\tau(\theta_0, m_0) \equiv \int \left[\tau - \frac{1}{2} + \frac{1}{2} sign(\widehat{v}_{i0} - z_{i0}'\theta) \right] .z_{i0}.\widetilde{w}(x_i).dx$$

converging to

$$\begin{aligned} G^\tau(\theta_0, m_0) &= E\left[z_0.sign(\varepsilon_0).\widetilde{w}(x)\right] \\ & \quad E\left(\left[\tau - \frac{1}{2} + \frac{1}{2} sign(\varepsilon_0) \right] .z_{i0}.\widetilde{w}(x_i) \right) \end{aligned}$$

To verify CLK's Theorem 2, note that one needs H_1, and V_1, the asymptotic covariance matrix described in condition 2.6. The former is given by

$$H_1^\tau(\theta_0, m_0) = -E\left[f_\varepsilon(0|z).z_i z_i'.\widetilde{w}(x)\right]$$

Sufficient conditions for Condition 2.2 to be satisfied are $E(z_i z_i')$ being of full rank and the error density being bounded away from zero - for instance, neither of x_i, $m_{q_2}(x_i)$, or $x_i.m_{q_2}(x_i)$ can be constant across i. Once this occurs, $E(z'z)$ is positive-definite, continuous, and full rank.

To verify Condition 2.5 we use CLK's Theorem 3. To check Conditions 3.1 and 3.2 of this result, note that

$$|g_j(\theta^*, m^*) - g_j(\theta, m)| = \left| \left[\tau - \frac{1}{2} + \frac{1}{2} sign(\varepsilon_j^*) \right] .z_j^*.\widetilde{w}(x_j) - \left[\tau - \frac{1}{2} + \frac{1}{2} sign(\varepsilon_j) \right] .z_j.\widetilde{w}(x_j) \right|$$

where $\varepsilon^* = \widehat{m}_{q_1}^*(x) - \alpha_0^* - x_i'\alpha_1^* - (\beta_0^* + x_i'\beta_1^*).\widehat{m}_{q_2}^*(x)$, and $\varepsilon = \widehat{m}_{q_1}(x) - \alpha_0 - x_i'\alpha_1 - (\beta_0 + x_i'\beta_1).\widehat{m}_{q_2}(x)$.

The rows of $|g_j(\theta^*, m^*) - g_j(\theta, m)|$ are majorated by, respectively

[row 1]: $\frac{1}{2}|sign(\varepsilon_j^*) - sign(\varepsilon_j)|.|\widetilde{w}(x_j)| \leq |\widetilde{w}(x_j)|$

[row 2]: $\frac{1}{2}|sign(\varepsilon_j^*) - sign(\varepsilon_j)|.|x_j|.|\widetilde{w}(x_j)| \leq |x_j|.|\widetilde{w}(x_j)|$

[row 3]: $\frac{1}{2}|sign(\varepsilon_j^*).\widehat{m}_{q_2}^*(x_j) - sign(\varepsilon_j).\widehat{m}_{q_2}(x_j)|.|\widetilde{w}(x_j)| \leq |\widehat{m}_{q_2}^*(x_j) - \widehat{m}_{q_2}(x_j)|.|\widetilde{w}(x_j)| \leq |\widehat{m}_{q_2}^*(x_j)|.|\widetilde{w}(x_j)| + |\widehat{m}_{q_2}(x_j)|.|\widetilde{w}(x_j)|$

[row 4]: $\frac{1}{2}|sign(\varepsilon_j^*).\widehat{m}_{q_2}^*(x_j) - sign(\varepsilon_j).\widehat{m}_{q_2}(x_j)|.|x_j|.|\widetilde{w}(x_j)| \leq |\widehat{m}_{q_2}^*(x_j) - \widehat{m}_{q_2}(x_j)|.|x_j|.|\widetilde{w}(x_j)| \leq |\widehat{m}_{q_2}^*(x_j)|.|\widetilde{w}(x_j)| + |\widehat{m}_{q_2}(x_j)|.|x_j|.|\widetilde{w}(x_j)|$

Sufficient conditions are given by a bounded weighting function $\widetilde{w}(x_j)$ (such as an indicator function), a bounded x_j (equivalent to requiring that there is an upper bound to the expenditure level), and bounded $\widehat{m}_{q_2}(.)$ (which it is by assumption, since budget shares lie between 0 and 1). It then follows that Conditions 3.1 and 3.2 of CLK are satisfied for $s_j = 1$, $K_j < \infty$. In what concerns the remaining condition we refer to the remark after CLK's Theorem 3 (or to CLK, pp. 1598) and argue that since $s_j = 1$, Condition 3.3 is satisfied. As a result, Condition 2.5 is satisfied.

To verify Condition 2.6 one needs to obtain V_1. First consider

$$G_N^\tau(\theta, m) = \int \left[\tau - \frac{1}{2} + \frac{1}{2} sign(\varepsilon_i)\right] .z_i.\widetilde{w}(x_i).dx$$

Now recall that $sign(\varepsilon_{i0}) = sign(v_{i0} - z'_{i0}\theta_0)$ is a binomial random variable with zero mean and unit variance. Thus, given ε independent, of finite variance, and independent of x, applying a Central Limit Theorem results in asymptotic normality of this term. The second term concerning 2.6 is $H_2^\tau(\theta_0, m_0)[\widehat{m} - m_0]$. Note that

$$H_{2N}^\tau(\theta, m)[m - m_0] = \frac{\partial G_N^\tau(\theta, m + \gamma\zeta)}{\partial \gamma}|_{\gamma=0}$$

where $\zeta = m - m_0$. But

$$G_N^\tau(\theta, m + \gamma\zeta) = \int \left[\tau - \frac{1}{2} + \frac{1}{2} sign(v_{i,\gamma\zeta} - z'_{i,\gamma\zeta}\theta)\right] .z_{i,\gamma\zeta}.\widetilde{w}(x_i).dx$$

where $z_{i,\gamma\zeta} = \left[\begin{array}{cccc} 1 & x_i & m_{q_2}(x_i) + \gamma\zeta_2 & x_i.\left(m_{q_2}(x_i) + \gamma\zeta_2\right) \end{array}\right]'$, $v_{i,\gamma\zeta} = m_{q_1}(x_1) + \gamma\zeta$, and $\varepsilon_{i,\gamma\zeta} = v_{i,\gamma\zeta} - z'_{i,\gamma\zeta}\theta$.

Differentiating with respect to γ yields

$$\frac{\partial G_N^\tau(\theta, m + \gamma\zeta)}{\partial \gamma}$$
$$= \left[\begin{array}{c} \int \delta(\varepsilon_{i,\gamma\zeta}).\{\zeta_1 - (\beta_0 + x'_i\beta_1).\zeta_2\}.\widetilde{w}(x_i).dx \\ \int x'_i.\delta(\varepsilon_{i,\gamma\zeta}).\{\zeta_1 - (\beta_0 + x'_i\beta_1).\zeta_2\}.\widetilde{w}(x_i).dx \\ \int \left(\zeta_2.sign(\varepsilon_{i,\gamma\zeta}) + \{m_{q_2}(x_i) + \gamma\zeta_2\}.\delta(\varepsilon_{i,\gamma\zeta}).\{\zeta_1 - (\beta_0 + x'_i\beta_1).\zeta_2\}\right).\widetilde{w}(x_i).dx \\ \int \left(x'_i.\zeta_2.sign(\varepsilon_{i,\gamma\zeta}) + x'_i.\{m_{q_2}(x_i) + \gamma\zeta_2\}.\delta(\varepsilon_{i,\gamma\zeta}).\{\zeta_1 - (\beta_0 + x'_i\beta_1).\zeta_2\}\right).\widetilde{w}(x_i).dx \end{array}\right]$$

setting $\gamma = 0$ results in

$$H_{2N}^\tau(\theta, m)[m - m_0] = \left[\begin{array}{c} \int \delta(\varepsilon_i).\{\zeta_1 - (\beta_0 + x'_i\beta_1).\zeta_2\}.\widetilde{w}(x_i).dx \\ \int x'_i.\delta(\varepsilon_i).\{\zeta_1 - (\beta_0 + x'_i\beta_1).\zeta_2\}.\widetilde{w}(x_i).dx \\ \int \left(\zeta_2.sign(\varepsilon_i) + m_{q_2}(x_i).\delta(\varepsilon_i).\{\zeta_1 - (\beta_0 + x'_i\beta_1).\zeta_2\}\right).\widetilde{w}(x_i).dx \\ \int \left(x'_i.\zeta_2.sign(\varepsilon_i) + x'_i.m_{q_2}(x_i).\delta(\varepsilon_i).\{\zeta_1 - (\beta_0 + x'_i\beta_1).\zeta_2\}\right).\widetilde{w}(x_i).dx \end{array}\right]$$

It then follows that

$$
\begin{aligned}
& G_N^{\tau}(\theta_0, m_0) + H_{2N}^{\tau}(\theta_0, m_0)[\widehat{m} - m_0] \\
= & \left[\begin{array}{c}
\int sign(\varepsilon_i(\theta_0, \widehat{m})).\widetilde{w}(x_i).dx + \int \delta(\varepsilon_i(\theta_0, \widehat{m})).\widehat{\xi}_i.\widetilde{w}(x_i).dx \\
\int x_i'.sign(\varepsilon_i(\theta_0, \widehat{m})).\widetilde{w}(x_i).dx + \int x_i'.\delta(\varepsilon_i(\theta_0, \widehat{m})).\widehat{\xi}_i.\widetilde{w}(x_i).dx \\
\int \widehat{m}_{q_2}(x_i).sign(\varepsilon_i(\theta_0, \widehat{m})).\widetilde{w}(x_i).dx + \\
+ \int \left(\widehat{\zeta}_2.sign(\varepsilon_i(\theta_0, \widehat{m})) + \widehat{m}_{q_2}(x_i).\delta(\varepsilon_i(\theta_0, \widehat{m})).\widehat{\xi}_i \right).\widetilde{w}(x_i).dx \\
\int x_i'.\widehat{m}_{q_2}(x_i).sign(\varepsilon_i(\theta_0, \widehat{m})).\widetilde{w}(x_i).dx + \\
+ \int \left(x_i'.\widehat{\zeta}_2.sign(\varepsilon_i(\theta_0, \widehat{m})) + x_i'.\widehat{m}_{q_2}(x_i).\delta(\varepsilon_i(\theta_0, \widehat{m})).\widehat{\xi}_i \right).\widetilde{w}(x_i).dx
\end{array} \right]
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{\xi}_i &= \{\widehat{\zeta}_1 - (\beta_0 + x_i'\beta_1).\widehat{\zeta}_2\} \\
\widehat{\zeta}_j &= \widehat{m}_{q_j}(x_j) - m_{q_j}(x_j) = \frac{1}{f_{u_j}(0)}.\frac{1}{N_j}\sum_{i=1}^{N_j}\left[q_j - 1\{u_j \leq 0\}\right], j = 1, 2
\end{aligned}
$$

the latter being the local Bahadur representation of Chaudhuri (1991). The term $G_N^{\tau}(\theta_0, m_0) + H_{2N}^{\tau}(\theta_0, m_0)[\widehat{m} - m_0]$ is asymptotically normal with zero mean and finite variance under regularity conditions such as $f_{u_j}(.)$ and $f_{\varepsilon}(.)$ being bounded away from both zero and infinity and *i.i.d.* observations. The corresponding covariance matrix V_1^{τ} is then obtained as

$$
V_1^{\tau} = E[G_N^{\tau}(\theta_0, m_0).G_N^{\tau\prime}(\theta_0, m_0) + H_{2N}^{\tau}(\theta_0, m_0).H_{2N}^{\tau\prime}(\theta_0, m_0).[\widehat{m} - m_0]]
$$

■

References

[1] Andrews, D. W. K. (1994), Asymptotics for Semiparametric Econometric Models Via Stochastic Equicontinuity, Econometrica, 62 (1), 43-72.

[2] Bierens, H. J. and H. Pott-Butter (1990), Specification of Household Engel Curves by Nonparametric Regression (with discussion), Econometric Reviews 9, 123-184.

[3] Blundell, R., M. Browning, and I. Crawford (2003), Nonparametric Engel Curves and Revealed Preference, Econometrica 71, 205-240.

[4] Blundell, R., A. Duncan, and K. Pendakur (1998), Semiparametric Estimation and Consumer Demand, Journal of Applied Econometrics, 13 (5), 1-30.

[5] Buchinsky, M. (1994), Changes in the US Wage Structure 1963-1987: Application of Quantile Regression, Econometrica 62 (2), 405-458.

[6] Buchinsky, M. (1998), Recent Advances in Quantile Regression Models: A Practical Guide for Empirical Research, Journal of Human Resources, 33, 88-126.

[7] Chaudhuri, P. (1991), Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation, Annals of Statistics, 19, 760-777.

[8] Chen, X., O. Linton, and I. van Keilegom (2003), Estimation of Semiparametric Models when the Criterion Function is not Smooth, Econometrica forthcoming.

[9] Deaton, A., and J. Muellbauer (1980), Economics and Consumer Behavior. Cambridge: Cambridge University Press.

[10] Härdle, W., and Th. Gasser (1984), Robust Non-Parametric Function Fitting, Journal of the Royal Statistical Society. Series B, 46 (1), 42-51.

[11] Härdle, W., and J. S. Marron (1990), Semiparametric Comparison of Regression Curves with Similar Shape, Annals of Statistics, 18(1), 63-89.

[12] Horowitz, J. (2001), The Bootstrap, Handbook of Econometrics, Vol. V, J.J. Heckman and E.E. Leamer, eds., Elsevier Science B.V., Ch. 52, pp. 3159-3228.

[13] Kim, T-H., and H. White (2002), Estimation, Inference and specification Testing for Possible Misspecified Quantile Regression, Advances in Econometrics forthcoming.

[14] Koenker, R. (2004), Quanterg: A Quantile Regression Package for R, http://www.R-project.org

[15] Koenker, R. and G. Bassett (1978), Regression Quantiles, Econometrica, 46, 33-50.

[16] Koenker, R., and V. D'Orey (1987), Computing Regression Quantiles, Applied Statistics 36, 383-393.

[17] Koenker, R. and K. Hallock (2001), Quantile Regression, Journal of Economic Perspectives, 15, 143-156.

[18] Koenker, R., and Z. Xiao (2002), Inference on the Quantile Regression Process, Econometrica, 70 (4), 1583-1612.

[19] Newey, W. K. (1994), The Asymptotic Variance of Semiparametric Estimators, Econometrica, 62 (6), 1349-1382.

[20] Portnoy, S. and R. Koenker (1997), The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-error vs. Absolute-error Estimators, Statistical Science, 12, 279-300.

[21] Powell, J. L. (1994), Estimation in Semiparametric Models, in The Handbook of Econometrics, vol. IV, eds. R. F. Engle and D. L. McFadden. North Holland.

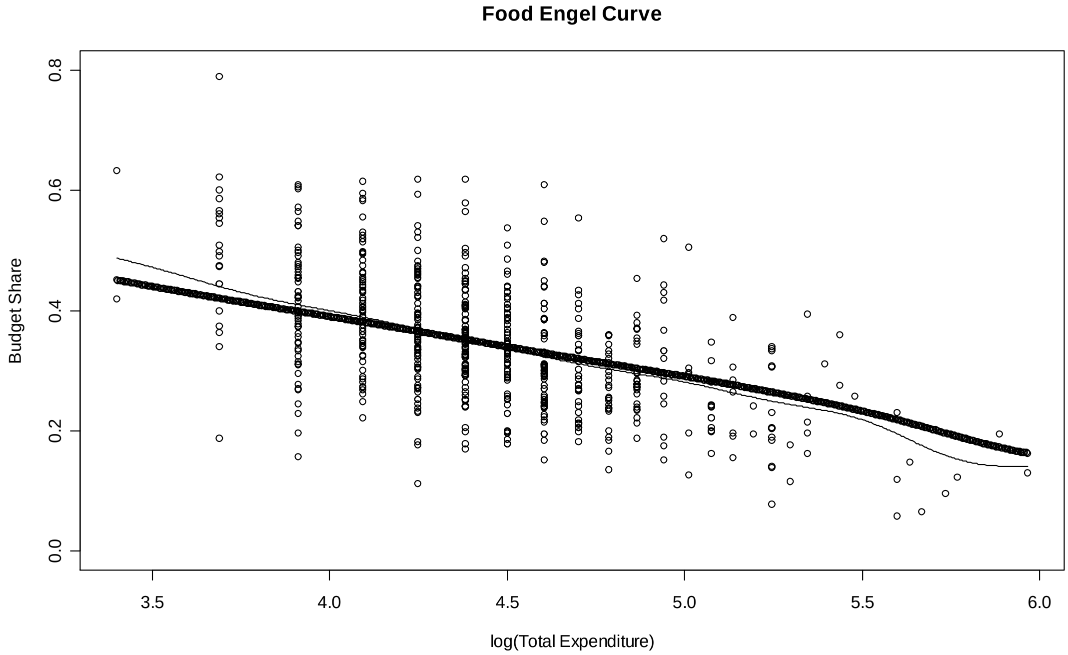
[22] Tsybakov, A. B. (1986), Robust Reconstruction of Functions by Local-Approximation Method, Problems Inform. Transmision 22, 133-146.

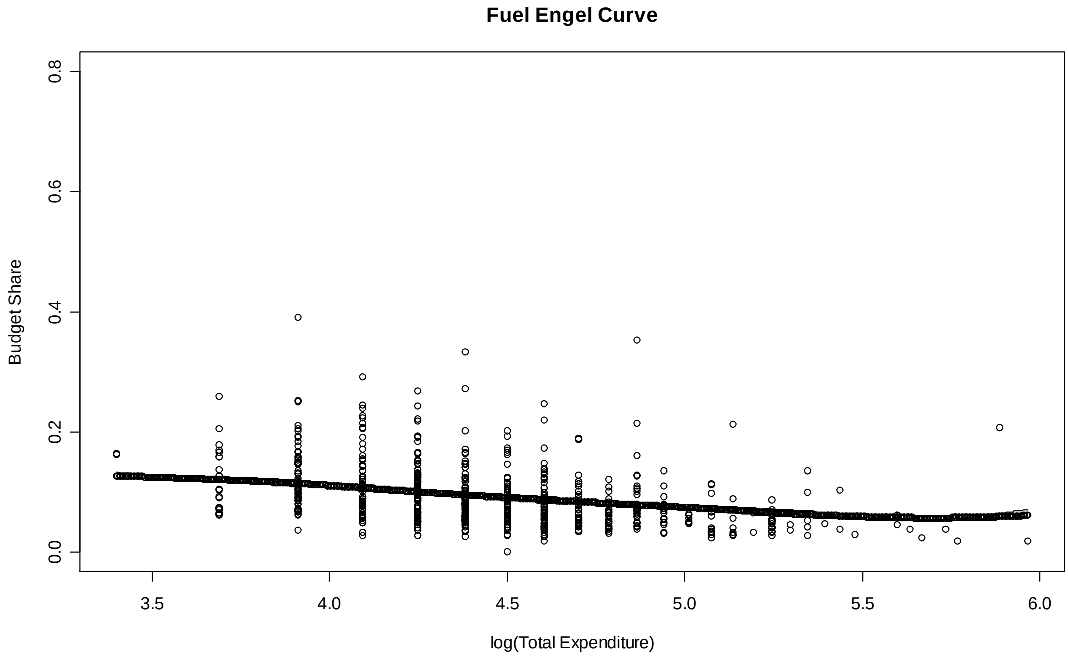
[23] Welsh, A. H. (1996), Robust Estimation of Smooth Regression and Spread Functions and Their Derivatives, Statistica Sinica 6, 347-366.

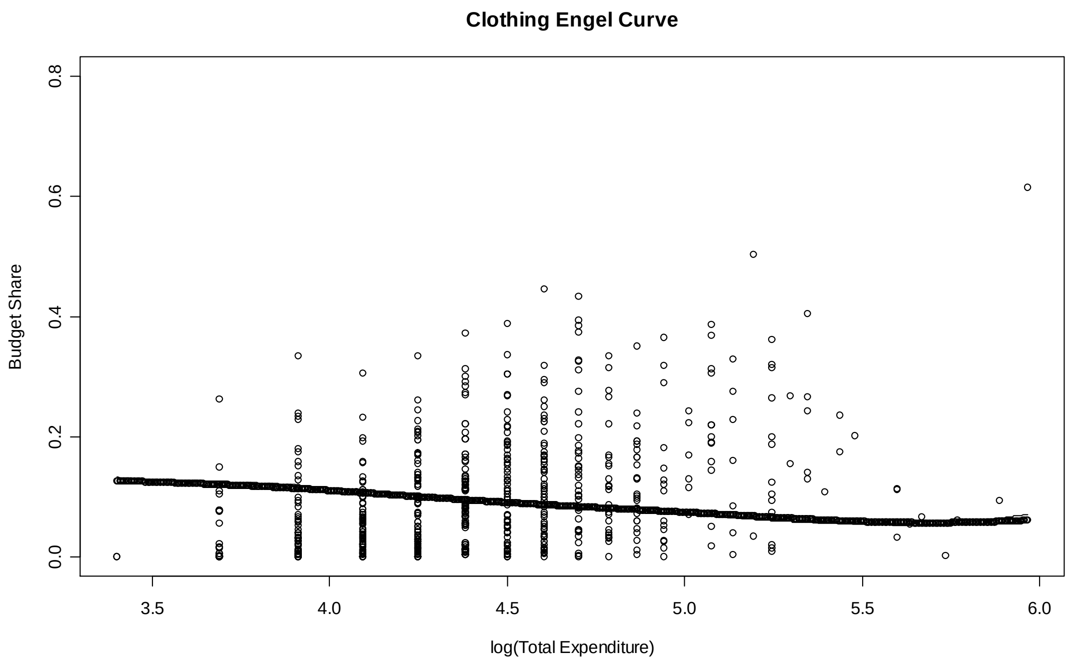
[24] White, H. (1980), A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, Econometrica, 48, 817-838.

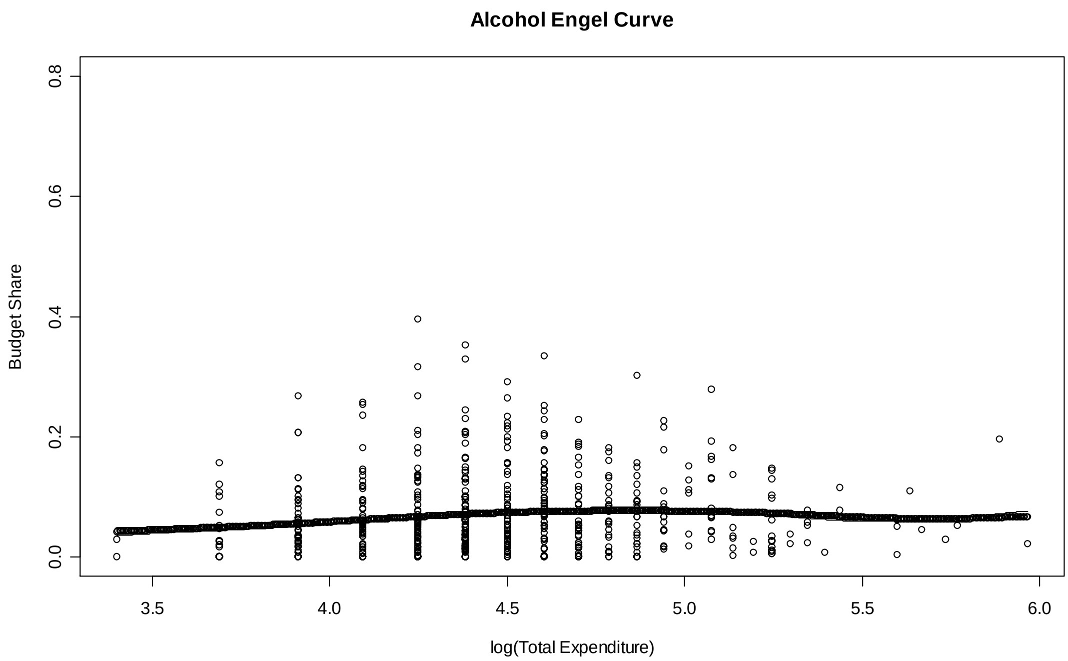
Figures 1-6 – Nonparametric Engel Curves

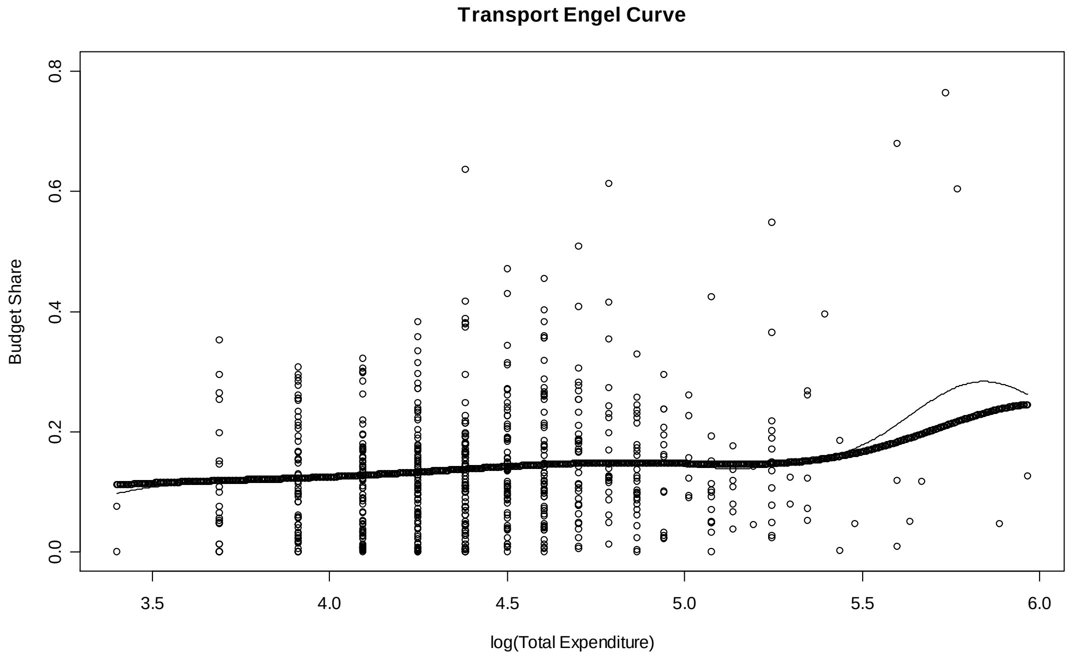
Thin lines: bandwidth = 0.5
Thick lines: bandwidth = 0.75

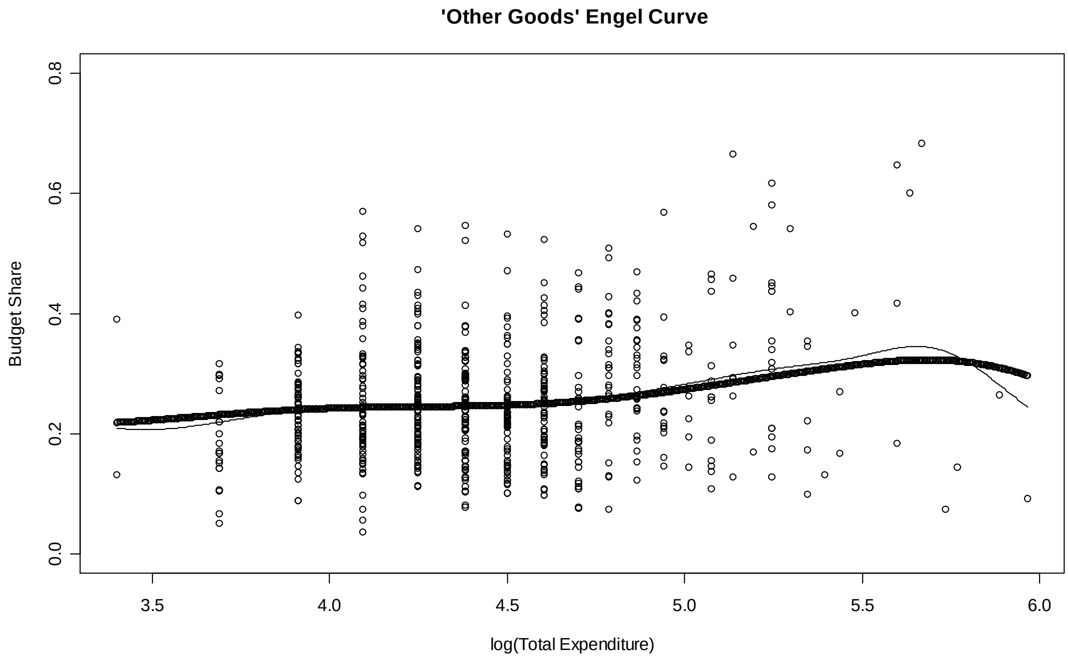





Clothing Engel Curve
Budget Share
0.0
0.2
0.4
0.6
0.8
3.5
4.0
4.5
5.0
5.5
6.0
log(Total Expenditure)

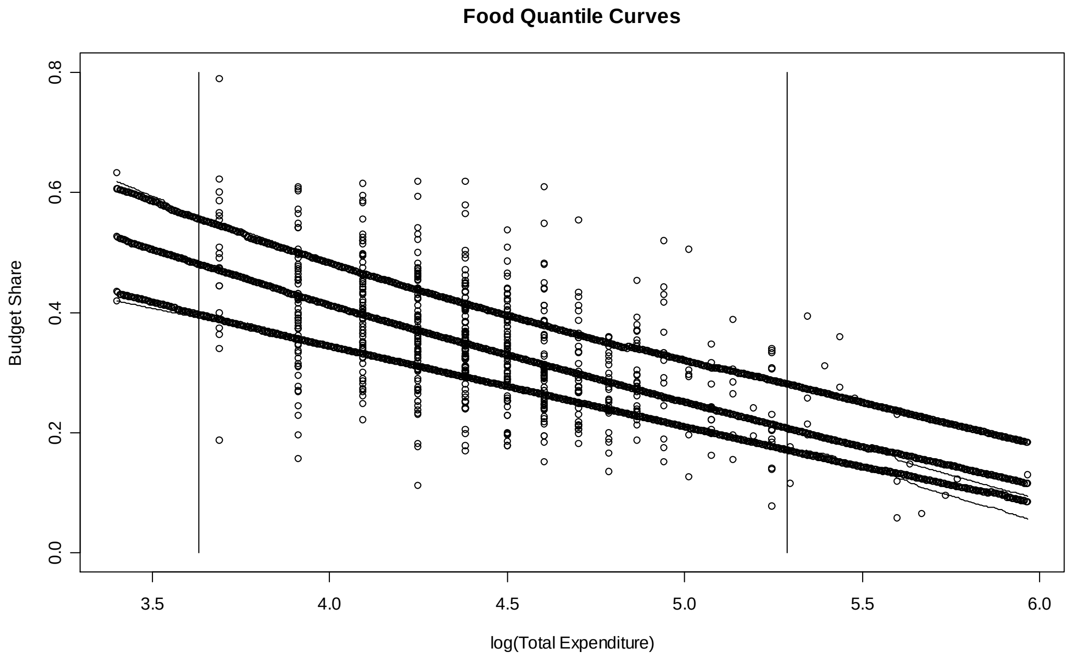

Alcohol Engel Curve
Budget Share
0.0
0.2
0.4
0.6
0.8
3.5
4.0
4.5
5.0
5.5
6.0
log(Total Expenditure)

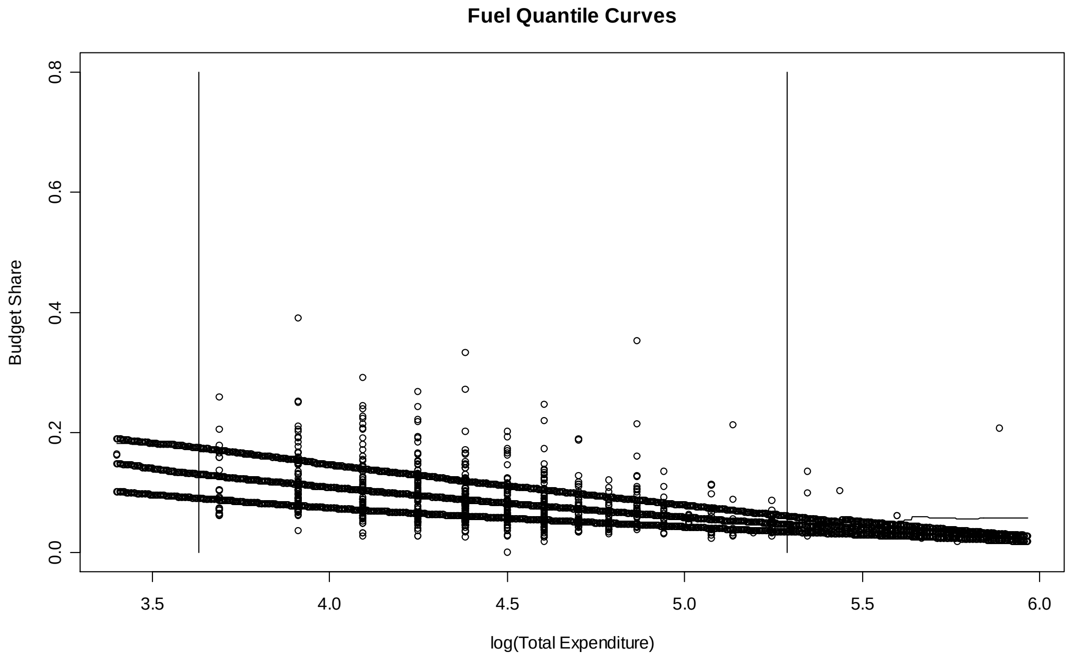

Transport Engel Curve
Budget Share
0.0
0.2
0.4
0.6
0.8
3.5
4.0
4.5
5.0
5.5
6.0
log(Total Expenditure)

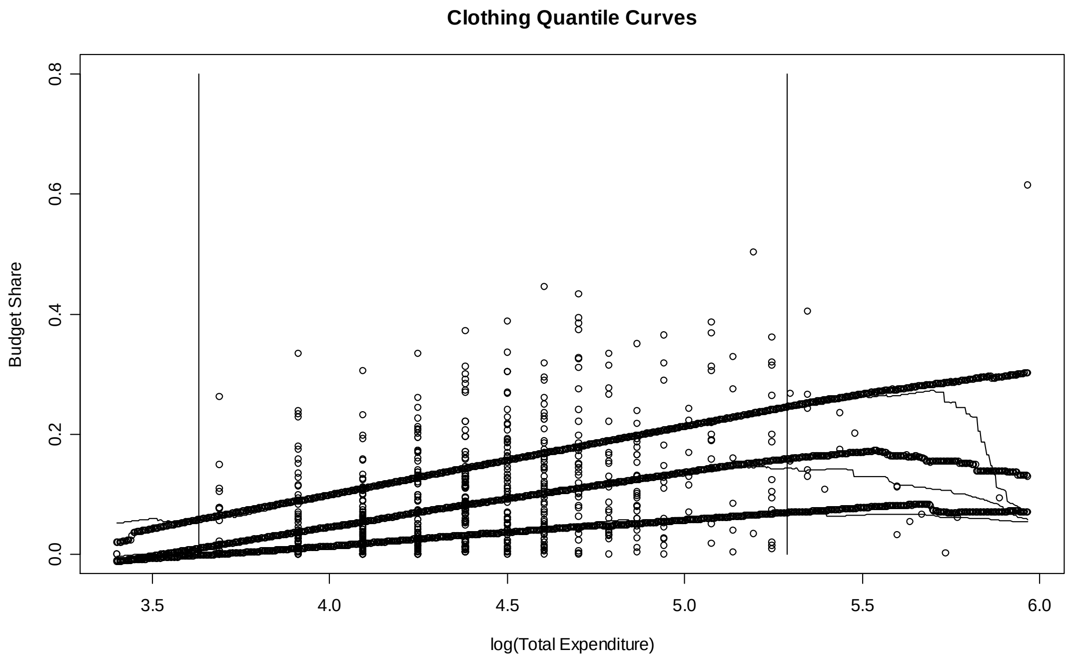

'Other Goods' Engel Curve
Budget Share
0.0
0.2
0.4
0.6
0.8
3.5
4.0
4.5
5.0
5.5
6.0
log(Total Expenditure)

Figures 7-12 – Quantile Engel Curves

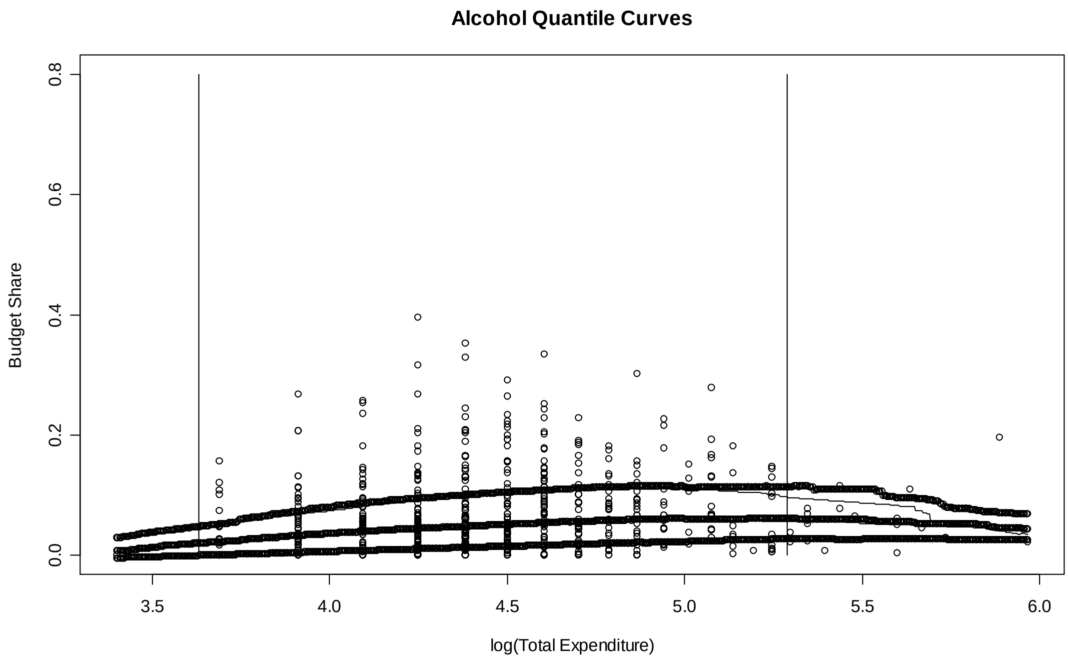
Thin lines: bandwidth = 1.0
Thick lines: bandwidth = 1.5



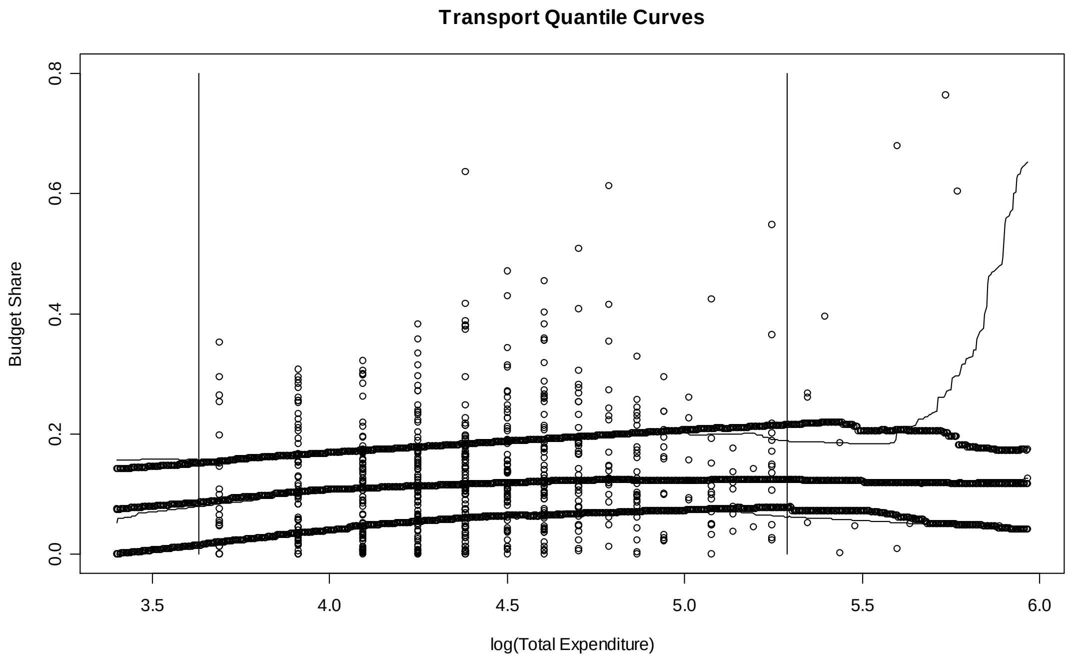


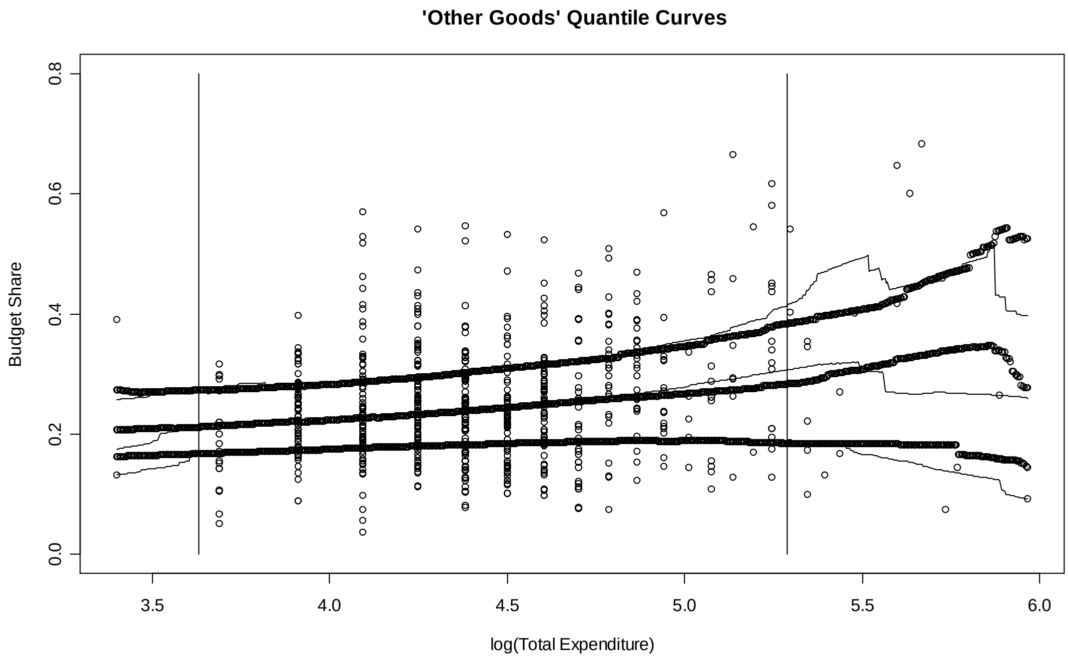


Clothing Quantile Curves
Budget Share
log(Total Expenditure)
0.0
0.2
0.4
0.6
0.8
3.5
4.0
4.5
5.0
5.5
6.0



Alcohol Quantile Curves
Budget Share
log(Total Expenditure)
0.0
0.2
0.4
0.6
0.8
3.5
4.0
4.5
5.0
5.5
6.0


Transport Quantile Curves
Budget Share
0.0
0.2
0.4
0.6
0.8
3.5
4.0
4.5
5.0
5.5
6.0
log(Total Expenditure)


'Other Goods' Quantile Curves
Budget Share
0.0
0.2
0.4
0.6
0.8
3.5
4.0
4.5
5.0
5.5
6.0
log(Total Expenditure)