TESTING NON-NESTED SEMIPARAMETRIC MODELS: AN APPLICATION TO ENGEL CURVES SPECIFICATION

MIGUEL A. DELGADO[a,*] AND JUAN MORA[b]

[a]*Departamento de Estadistica y Econometria, Universidad Carlos III de Madrid, Madrid, Spain*
[b]*Departamento de Fundamentos del Análisis Económico, Universidad de Alicante, Alicante, Spain*

SUMMARY

This paper proposes a test statistic for discriminating between two partly non-linear regression models whose parametric components are non-nested. The statistic has the form of a *J*-test based on a parameter which artificially nests the null and alternative hypotheses. We study in detail the realistic case where all regressors in the non-linear part are discrete and then no smoothing is required on estimating the non-parametric components. We also consider the general case where continuous and discrete regressors are present. The performance of the test in finite samples is discussed in the context of some Monte Carlo experiments. The test is well motivated for specification testing of Engel curves. We provide an application using data from the 1980 Spanish Expenditure Survey.

1. INTRODUCTION

The semiparametric partly linear regression (SPLR) model has recently attracted considerable attention. We are interested in the estimation of the parameters entering in the linear part of a regression model which is partly non-linear in certain explanatory variables. The functional form of the non-linear part of the model is not parametrically specified. Estimation methods for the linear part have been proposed by Chen (1988), Speckman (1988) and Robinson (1988), among others.

In this paper, we propose a specification test of non-tested SPLR models belonging to the class of *J*-tests suggested by Davidson and MacKinnon (1981). We motivate the test in the context of functional specification of Engel curves with cross-sectional data. The focus of interest is the specification of the relation between expenditure and income given other explanatory variables about household characteristics.

The rest of the paper is organized as follows. In Section 2 we present the test statistic and in Section 3 we derive its asymptotic properties. Section 4 provides some Monte Carlo simulations in order to illustrate the performance of the test in small and moderate samples. In Section 5 the test is applied to specification testing of Engel curves using data from the Spanish Expenditure Survey.

2. THE TEST STATISTIC

Data consists of independent observations $\{(Y_i, X_i, Z_i), i=1, \ldots, n\}$ identically distributed as the $\mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q$-valued random variable (Y, X, Z), where $X=(X_1', X_2', X_3')$ and X_1, X_2, X_3 take values in $\mathbb{R}^{p_1}$, $\mathbb{R}^{p_2}$ and $\mathbb{R}^{p_3}$, respectively ($p_1+p_2+p_3=p, p_1>0, p_2>0$). We face the competing hypotheses:

$$\mathrm{H}_0\colon \Pr\{E[Y \mid X, Z] = X_1'\beta_1 + X_3'\beta_3 + g_1(Z)\} = 1 \tag{1}$$

$$\mathrm{H}_1\colon \Pr\{E[Y \mid X, Z] = X_2'\beta_2 + X_3'\beta_3 + g_2(Z)\} = 1 \tag{2}$$

where $g_1, g_2\colon \mathbb{R}^q \to \mathbb{R}$ are unknown functions, which may include an intercept term, and β_1, β_2 and β_3 are unknown parameters. The variables X_1 and X_2 are, possibly, transformations of some given set of variables. In the application to Engel curves of Section 5, X_1 and X_2 are non-linear transformations of income and Z includes variables like 'size of household' or 'age of reference person (RP)'.

In order to avoid the unknown functions $g_1(\cdot)$ and $g_2(\cdot)$, following the approach in Speckman (1988) and Robinson (1988) we represent the regression models in H_0 and H_1 as

$$\mathrm{H}_0\colon \Pr\{E[\varepsilon_Y \mid X, Z] = \varepsilon_1'\beta_1 + \varepsilon_3'\beta_3\} = 1 \tag{3}$$

$$\mathrm{H}_1\colon \Pr\{E[\varepsilon_Y \mid X, Z] = \varepsilon_2'\beta_2 + \varepsilon_3'\beta_3\} = 1 \tag{4}$$

where $\varepsilon_Y \equiv Y - m_Y(Z)$, $\varepsilon_j \equiv X_j - m_j(Z)$, $j=1, 2, 3$, $m_Y(\mathcal{Z}) \equiv E[Y \mid Z = \mathcal{Z}]$ and $m_j(\mathcal{Z}) \equiv E[X_j \mid Z = \mathcal{Z}], j=1, 2, 3$.

The hypotheses can be represented in terms of a parameter δ which artificially links the regression models in (3) and (4) by means of the composite hypothesis

$$\mathrm{H}_\mathrm{C}\colon \Pr\{E[\varepsilon_Y \mid X, Z] = (1-\delta)\varepsilon_1'\beta_1 + \delta\varepsilon_2'\beta_2 + \varepsilon_3'\beta_3\} = 1 \tag{5}$$

Hypotheses (1) and (2) become, in terms of δ,

$$\mathrm{H}_0\colon \delta = 0 \text{ versus } \mathrm{H}_1\colon \delta = 1 \tag{6}$$

If ε_{Yi} and ε_{ji}, $j=1, 2, 3$, were known, a J-test statistic could be based on an estimate of δ in the regression model:

$$\varepsilon_{Yi} = \varepsilon_{1i}'\gamma + q_i\delta + \varepsilon_{3i}'\beta_3 + \text{Error}, \quad 1 \leqslant i \leqslant n \tag{7}$$

where $\varepsilon_{Yi} \equiv Y_i - m_Y(Z_i)$, $\varepsilon_{ji} \equiv X_{ji} - m_j(Z_i), j=1, 2, 3$, $q_i \equiv \varepsilon_{2i}'\bar{\beta}_2$, and $\bar{\beta}_2$ is a consistent estimate of β_2 under H_1. Notice that model (7) results from reparameterizing model (5) with $\gamma = \beta_1(1-\delta)$ and then replacing β_2 by $\bar{\beta}_2$ to solve the identification problem. In our context, we will first replace ε_{Yi}, ε_{1i}, ε_{3i} and q_i in model (7) by suitable non-parametric estimates and then estimate δ jointly with γ and β_3 by ordinary least squares (OLS).

First, we estimate $m_Y(\mathcal{Z})$ and $m_k(\mathcal{Z})$ by non-parametric estimates $\hat{m}_Y(\mathcal{Z})$ and $\hat{m}_j(\mathcal{Z})$ ($j=1, 2, 3$), respectively, and, hence, ε_{Yi} and ε_{ji} are estimated by $\hat{\varepsilon}_{Yi} \equiv [Y_i - \hat{m}_Y(Z_i)]I_i$ and $\hat{\varepsilon}_{ji} \equiv [X_{ji} - \hat{m}_j(Z_i)]I_i$ ($j=1, 2, 3$), respectively. The indicator function I_i trims out those observations where the denominator of the non-parametric regression estimate is too small. Second, q_i is estimated by $\hat{q}_i \equiv \hat{\varepsilon}_{2i}'\bar{\beta}_2 I_i$, where $\bar{\beta}_2$ is the OLS estimate of β_2 in the regression

$\hat{\varepsilon}_{Yi} = \hat{\varepsilon}_{2i}\tilde{\beta}_2 + \hat{\varepsilon}'_{3i}\tilde{\beta}_3+$ Residual. Finally, $(\gamma, \beta'_3, \delta)'$ are jointly estimated by $(\tilde{\gamma}, \tilde{\beta}'_3, \tilde{\delta})'$ using OLS in the regression

$$\hat{\varepsilon}_{Yi} = \hat{\varepsilon}_{1i}\tilde{\gamma} + \hat{\varepsilon}'_{3i}\tilde{\beta}_3 + \hat{q}_i\tilde{\delta} + \text{Residual} \tag{8}$$

The test-statistic is the t-ratio of $\tilde{\delta}$ in regression (8). Its asymptotic properties are studied in the next section.

3. ASYMPTOTIC PROPERTIES OF THE TEST

First we describe the non-parametric estimates which are used in the test statistic and then present the theoretical results of the paper.

3.1. Non-parametric Estimates

We suppose first that all variables in Z are discrete — this is the case in many real situations where variables are dummies, qualitative variables, counts or continuous variables recorded at intervals (see, for example, Section 5). Thus, we assume that

$$\exists \mathcal{D} \subset \mathbb{R}^q, \mathcal{D} \text{ countable set, with } P(Z \subset \mathcal{D}) = 1 \text{ and } Z_i \in \mathcal{D} \Rightarrow P(Z = Z_i) > 0 \tag{9}$$

In this case, given observations $\{\xi_1, \ldots, \xi_n\}$ of a random vector ξ, non-parametric estimates of $m_{\xi i} \equiv E[\xi_i \mid Z_i]$ can be expressed as weighted averages of the form

$$\hat{m}^{(f)}_{\xi i} \equiv \sum_{j \neq i} \xi_j W^{(f)}_{nj}(Z_i),$$

where the superscript f is introduced in order to distinguish among different types of weights and, unless otherwise stated all sumations run from 1 to n. Observe that this is a 'leave-one-out' estimate because ξ_i is not used to estimate $m_{\xi i}$. As regressors are discrete, the simplest non-parametric weights we can use are the non-smoothing weights, defined as

$$W^{(1)}_{nj}(Z_i) \equiv I(Z_j = Z_i) \Big/ \sum_{m \neq i} I(Z_m = Z_i), \tag{10}$$

where $I(A)$ is the indicator function of event A and, hereafter, 0/0 is defined to be 0. In some situations the non-smoothing weights can perform poorly. As usual, the amount of smoothing must decrease as the sample size increases and, therefore, in most cases any smooth estimate will eventually coincide with the non-smoothing estimate. A natural way of smoothing in this context is k-nearest-neighbours (k-NN) where one chooses the k observations closest to Z_i according to a given metric. The precise definition of k-NN weights is as follows. Given a sequence of positive integers k_n and constants c_{jn} satisfying that $\sum_{j=1}^{n} c_{jn} = 1$, $c_{jn} \geqslant \cdots \geqslant c_{nn} \geqslant 0$ and $j > k_n \Rightarrow c_{jn} = 0$, the k-NN weights are

$$W^{(2)}_{nj}(\mathcal{Z}) = \left(\sum_{m=1}^{e(j,n,\mathcal{Z})} c_{d(j,n,\mathcal{Z})+m} \right) \Big/ e(j, n, \mathcal{Z}) \tag{11}$$

where $e(j, n, \mathcal{Z}) \equiv \#\{m : 1 \leqslant m \leqslant n, \rho_n(Z_m, \mathcal{Z}) = \rho_n(Z_j, \mathcal{Z})\}$, $d(j, n, \mathcal{Z}) \equiv \#\{m : 1 \leqslant m \leqslant n, \rho_n(Z_m, \mathcal{Z}) < \rho_n(Z_j, \mathcal{Z})\}$ and $\rho_n(u, v) = (\Sigma_l((u^{(l)} - v^{(l)})/s_{nl})^2)^{1/2}$. The sum in ρ_n extends over all l, $1 \leqslant l \leqslant q$, such that $s_{nl} > 0$ and s_{nl} denotes the sample standard deviation of $Z_1^{(l)}, \ldots, Z_n^{(l)}$ ($Z^{(l)}$ is the lth coordinate of Z, $1 \leqslant l \leqslant q$). There are different possible k-NN estimates, according to various choices of the sequence c_{jn}. The uniform k-NN estimate ($c_{jn} = I(1 \leqslant j \leqslant k_n)/k_n$) is the most popular one. Another alternative c_{jn} are defined in Stone (1977). The value k_n is usually referred to as smoothing value. All regression estimates can be viewed as local averages around the point at which regression is estimated. The k-NN weights are intuitively appealing because one decides how many points are used in these local averages.

In Delgado and Mora (1995a, b) it is shown that, when regressors are discrete, the non-smoothing estimate is globally consistent in the sense of Stone (1977) and estimates based on different smoothing techniques, including kernels and the regressogram, are asymptotically equivalent, in a very strong way, to the non-smoothing estimate.

Let us assume now that Z satisfies that

$$\left.\begin{array}{r} Z' = (Z^{(1)'}, Z^{(2)'}, \text{ where } Z^{(1)} \subset \mathbb{R}^r \text{ is discrete and} \\ Z^{(2)} \subset \mathbb{R}^s \text{ is absolutely continuous; } r + s = q,\ s \geqslant 1 \end{array}\right\} \tag{12}$$

For any dependent variable ξ, we estimate $E[\xi_i \mid Z_i]$ using Nadaraya–Watson kernel weights for the continuous regressors and non-smoothing weights for the discrete regressors. Hence, the weights in this case are

$$W_{nj}^{(3)}(Z_i) \equiv \Psi_{ij}(h_n) I(Z_i^{(1)} = Z_j^{(1)}) / \Sigma_m \Psi_{im}(h_n) I(Z_i^{(1)} = Z_m^{(1)}) \tag{13}$$

where $\Psi_{ij}(h_n) \equiv \Psi((Z_i^{(2)} - Z_j^{(2)})/h_n)$, Ψ: $\mathbb{R}^s \to \mathbb{R}$ is defined as $\Psi(c) = \Pi_{i=1}^s \psi(c_i)$ for any $c \in \mathbb{R}^s$ (ψ: $\mathbb{R} \to \mathbb{R}$ is a univariate kernel function) and $h_n > 0$ is a sequence of smoothing values. In this case, the non-parametric estimate we use is $\hat{m}_{\zeta i}^{(3)} \equiv \Sigma_j \zeta_j W_{nj}^{(3)}(Z_i)$. Note that this is not a 'leave-one-out' estimate. We do not need to use 'leave-one-out' estimates here because universal consistency results are not required. However, stronger technical conditions will be required when working with them.

3.2. Theoretical Results

As mentioned above, our statistic is based on the t-ratio of $\tilde{\delta}$ in equation (8), irrespective of the non-parametric method we choose. That is, the test statistic is

$$\hat{J}_n^{(f)} \equiv \tilde{\delta}/\mathrm{SE}(\tilde{\delta}) \quad f = 1, 2, 3 \tag{14}$$

where the superscript f indicates the non-parametric estimation procedure employed and $\mathrm{SE}(\tilde{\delta})$ is the standard error provided by any statistical package, which may be robust to heteroscedasticity of unknown form if the researcher suspects that this problem might arise. Observe that the computation of $\hat{J}_n^{(f)}$ from model (8) also requires the use of a trimming function I_i which is defined in the Appendix.

The following theorem justifies an asymptotic test based on $\hat{J}_n^{(f)}$, $f = 1, 2$, when all regressors are discrete.

Theorem 1. Under assumption (9) and other regularity conditions stated in the Appendix,

(a) $\lim_{n\to\infty} Pr(\hat{J}_n^{(f)} \geqslant z_\alpha) = \alpha, f = 1, 2,$ under H_0

where z_α is such that $Pr(Z \geqslant z_\alpha) = \alpha$ for the standard normal distribution Z.

(b) $\lim_{n\to\infty} Pr(\hat{J}_n^{(f)} \geqslant c) = 1, \forall c > 0, f = 1, 2,$ under H_1 □

Thus, Theorem 1 proves that critical values can be approximated by a standard normal distribution and the test is consistent.

When Z contains any continuous variables we have the following result:

Theorem 2. Under assumption (12) and other regularity conditions stated in the Appendix, the test statistic $\hat{J}_n^{(3)}$ is also asymptotically normal under the null hypothesis and consistent under the alternative one, as in Theorem 1. □

Theorems 1 and 2 generalize earlier results obtained in a completely parametric environment by Davidson and MacKinnon (1981) and others. Theorem 1 is based on results in Delgado and Mora (1995a), whereas Theorem 2 also uses results in Robinson (1988).

Loosely speaking, when Z is discrete and non-smoothing weights are used, only moment conditions on X and the error term are required for the asymptotic results to follow, and heteroscedasticity can be easily handled. If k-NN weights are used, conditions on the rate of convergence of k_n and moment conditions on Z are also required. When Z contains continuous variables, assumptions are much stronger and include conditions on the rate of convergence of h_n and the order of the kernel function $\psi(\cdot)$, which are related to smoothness conditions on $g(\cdot)$, $E[X|Z = z]$ and the conditional density functions of Z; independence between regressors and errors is also required. This latter assumption is also required by Robinson (1988) and seems difficult to relax when Nadaraya–Watson regression estimates are used. Fan *et al.* (1995) have shown that it is possible to relax this assumption, allowing for heteroscedasticity of unknown form, using density-weighted least squares. The procedure which Fan *et al.* (1995) propose can also be applied in our context, but note that if the true model is homoscedastic then this procedure yields inefficient estimates. Observe also that optimal bandwidth selection in this context is not discussed (see Linton, 1995).

4. MONTE CARLO SIMULATION

We illustrate the performance of our test using Monte Carlo experiments. First we study the size of the test. We generate n i.i.d. observations from the model

$$Y_i = X_{i1} + F(Z_i) + U_{i1} \tag{15}$$

for $i = 1, \ldots, n$, and $X_{i1} = F(Z_i) + U_{i2}$, $X_{i2} = F(Z_i) + \varphi_1 U_{i2} + \varphi_2 U_{i3}$, where $U_i = (U_{i1}, U_{i2}, U_{i3})' \sim N(0, I_3)$, Z_i follows a Poisson distribution with parameter $\lambda = 2$ independent of U_i; φ_1, φ_2 and function $F(\cdot)$ will be specified below. The hypotheses to be tested are:

$$\begin{aligned} &H_0\colon \Pr\{E[Y \mid X_1, X_2, Z] = \beta_1 X_1 + g_1(Z)\} = 1; \\ &H_1\colon \Pr\{E[Y \mid X_1, X_2, Z] = \beta_2 X_2 + g_2(Z)\} = 1 \end{aligned} \tag{16}$$

Observe that parameters φ_1 and φ_2 determine the correlation between regressors in the null and alternative hypotheses; in particular $\rho \equiv \mathrm{Corr}(\varepsilon_1, \varepsilon_2) = \varphi_1/(\varphi_1^2 + \varphi_2^2)^{1/2}$, where $\varepsilon_j \equiv X_j - E[X_j|Z]$, $j = 1, 2$. We have performed various experiments in order to analyse the influence of $F(\cdot)$, φ_1 and φ_2 on the size of the test. Specifically, we consider five different models:

Model 1:	$\varphi_1 = 2/3$,	$\varphi_2 = 5^{1/2}/3$,	$(\rho = 2/3)$,	$F(Z) = Z$
Model 2:	$\varphi_1 = 2/3$,	$\varphi_2 = 5^{1/2}/3$,	$(\rho = 2/3)$,	$F(Z) = Z^2/5$
Model 3:	$\varphi_1 = 0{\cdot}99$,	$\varphi_2 = (1-0{\cdot}99^2)^{1/2}$,	$(\rho = 0{\cdot}99)$,	$F(Z) = Z$
Model 4:	$\varphi_1 = 1/2$,	$\varphi_2 = 3^{1/2}/2$,	$(\rho = 1/2)$,	$F(Z) = Z$
Model 5:	$\varphi_1 = 0$,	$\varphi_2 = 1$,	$(\rho = 0)$,	$F(Z) = Z$

In all experiments our results are based on $n = 100$ observations and $r = 1000$ replications. We compute the semiparametric test statistic using uniform k-NN weights and $k_n = 2, 5, 8$. These values have been selected by previous graphical inspection of various non-parametric estimates computed from some Monte Carlo samples. These smoothing values have been kept unchanged in all experiments as our theoretical results do not allow for data-driven k_n. Note that, as Z is discrete, $k_n = 2$ does not mean that the non-parametric estimate is computed as a weighted average with two observations (see equation (11)). As a benchmark, we also compute a purely parametric J-test which assumes $g_j(Z) = Z$ $(j = 1, 2)$ in equation (16). All experiments were performed with FORTRAN77 programs and run on HP Apollo Work Stations at Universidad Carlos III. (All programs and data used in this paper are available at the *Journal of Applied Econometrics*' Data Archive). P-value plots (that is, empirical distribution function of P-values; see, for example, Davidson and MacKinnon, 1994) for Models 1 and 2 are shown in Figures 1 and 2, respectively. Reported results correspond to three semiparametric test statistics, computed with different smoothing values, and the parametric test-statistic. P-value plots for Models 3, 4 and 5 are shown in Figure 3. Reported results correspond to semiparametric test statistics with $k_n = 5$.

In Figure 1 we observe that, as expected, in Model 1 the parametric test performs better than all semiparametric ones, but the latter also behave reasonably well, especially if k_n is small. In all cases, the test has a tendency to over-reject but this problem is not very serious in this Monte

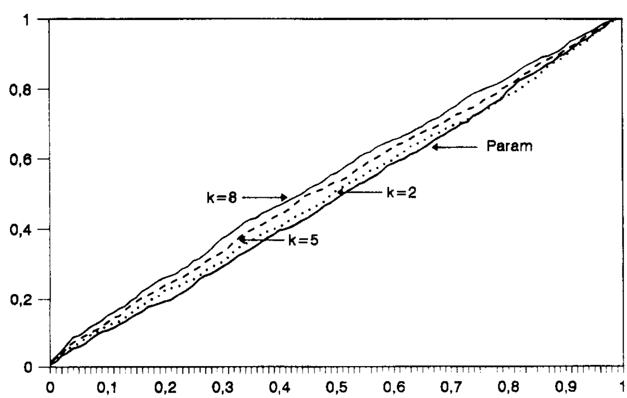

Figure 1. SIZE: P-value plot. Parametric test and semiparametric test with different smoothing values in Model 1

Carlo experiment (for example, when significance level is 0·05, the empirical level is 0·054 for the parametric test and 0·061 for the semiparametric test with $k_n = 2$). Observe that in this experiment with only one non-overlapping variable the exact null distribution, conditional on the observed regressors, is Student's t (see, for instance, Davidson and MacKinnon, 1993, p. 384). The over-rejection of the parametric test is explained by the fact that we are implementing the asymptotic test, based on the standard normal distribution. The problem of over-rejection may be more serious if the sample size is small or the two rival models have more than one non-overlapping variable; in these situations certain small-sample adjustments would probably be desirable (see Pesaran, 1974 or Godfrey and Pesaran, 1983). In Figure 2 we observe that, when $F(\cdot)$ is not linear, the semiparametric test is quite sensitive to the choice of k_n, but it still works reasonably well if k_n is adequately chosen; on interpreting the curve labelled 'Param' on Figure 2, it must be taken into account that both the null and alternative hypotheses faced by the researcher in that case are false. In Figure 3 we observe that, as expected, the performance of the

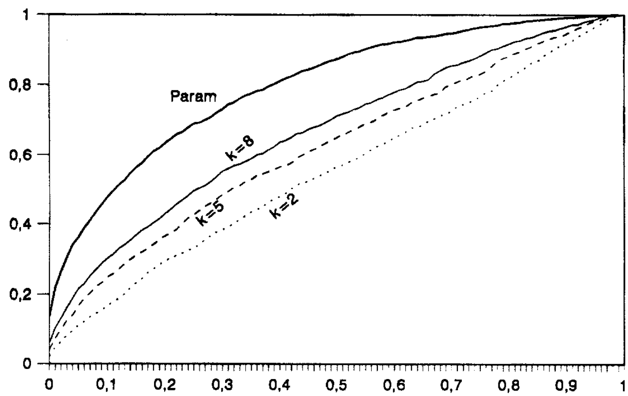

Figure 2. SIZE: P-value plot. Parametric test and semiparametric test with different smoothing values in Model 2

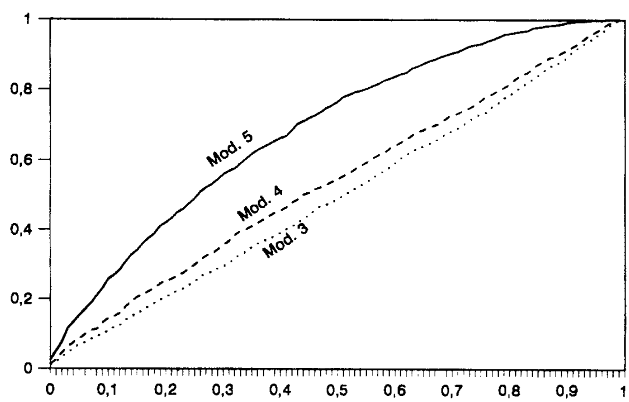

Figure 3. SIZE: P-value plot. Semiparametric test with smoothing value $k = 5$ in Models 3, 4, and 5

test statistic worsens as the correlation between regressors approaches 0 (note that Theorem 1 does not apply for Model 5 because condition $\lambda_2 \neq 0$ does not hold; see Appendix).

To study the power of the test we generate n i.i.d. observations from

$$Y_i = \beta_2 X_{i2} + F(Z_i) + U_{i1} \qquad (17)$$

for $i = 1, \ldots, n$ and X_{i1}, X_{i2}, Z_i, U_i are as defined before (below equation (15)), β_2 varies between $-0{\cdot}6$ and $0{\cdot}6$ and φ_1, φ_2 and $F(\cdot)$ are defined by:

Model 6:	$\varphi_1 = 2/3$,	$\varphi_2 = 5^{1/2}/3$,	$(\rho = 2/3)$,	$F(Z) = Z$
Model 7:	$\varphi_1 = 2/3$,	$\varphi_2 = 5^{1/2}/3$,	$(\rho = 2/3)$,	$F(Z) = Z^2/5$

Observe that if $\beta_2 = 0$ then both the null and alternative hypotheses are true, but Theorem 1 does not apply (conditions $\lambda_2 \neq 0$ and $\beta_2 \neq 0$ do not hold; see Appendix). The hypotheses to be tested are in equation (16). The results we report are based on $n = 100$ observations and $r = 1000$ replications and the experiments were carried out with the same characteristics as before. The percentage of rejections of the null hypothesis when $\alpha = 0{\cdot}05$ is shown in Figures 4 and 5, for Models 6 and 7, respectively.

In Figure 4 we observe that, when $F(\cdot)$ is linear (Model 6), the test is extremely powerful and behaves almost as well as the parametric one. In Figure 5 we observe that, when $F(\cdot)$ is not linear (Model 7), the semiparametric test continues to yield acceptable results whereas, obviously, the parametric test based on wrong specifications of the null and alternative hypotheses may produce misleading results.

We have also generated various models as before, but with continuous Z (taken from a normal distribution). The results we obtained, using higher order kernels, are entirely similar to those reported for discrete Z.

5. TESTING FUNCTIONAL FORMS OF ENGEL CURVES

The first attempt of estimating the regression curve relating expenditure with income is due to Engel (1895), who proposed the regressogram, the first non-parametric regression estimate. Since

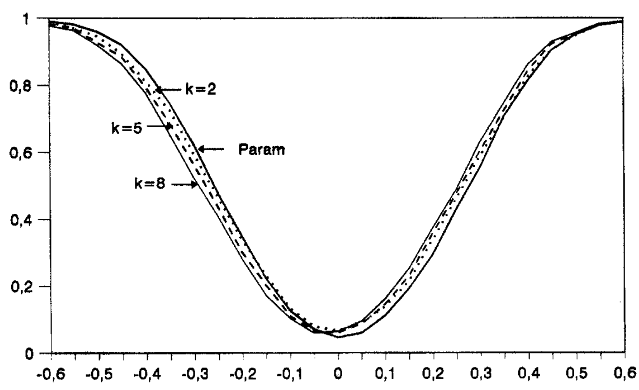

Figure 4. Empirical power function (significance level = 0·05). Parametric test and semiparametric test with different smoothing values in Model 6

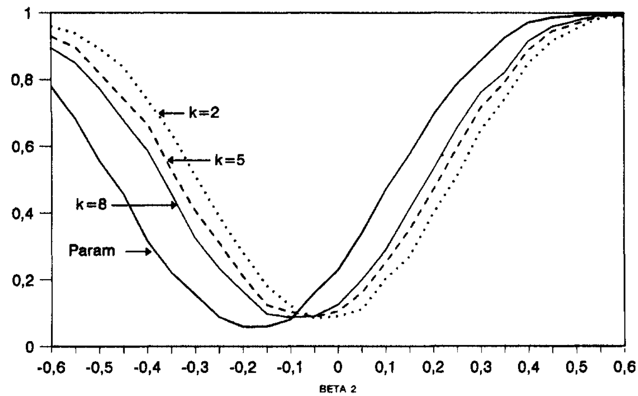

Figure 5. Empirical power function (significance level = 0·05). Parametric test and semiparametric test with different smoothing values in Model 7

then, there have been several alternative formulations for the functional form of Engel curves. The most popular one is due to Working (1943) and Leser (1963), who proposed a log-linear relationship. The validity of the Working–Leser formulation in certain goods has been questioned by other authors; see Deaton and Muellbauer (1980), Lewbel (1991), Banks *et al.* (1994), Deaton (1981) or Pollak and Wales (1978, 1980), to mention only a few. There is also some empirical evidence using non-parametric estimates of the Engel curve; see, for instance, Banks *et al.* (1994), Härdle and Mammen (1993) and Gozalo (1992). In this section we test the Working–Leser specification of Engel curves and the Engel curve form derived from the Quadratic Expenditure System (QES) of Pollak and Wales (1978, 1980), taking into account a vector of other possibly relevant variables Z containing characteristics of households (Z is usually referred to as 'demographic variables'). The information of these variables is usually incorporated into the modelling of Engel curves by means of 'demographic translating' and 'demographic scaling' (see Pollak and Wales, 1980). These procedures are simple to implement, but recent empirical research does not support this specification (see Gozalo, 1992). We propose a different way of introducing the information regarding demographic characteristics. We introduce all these variables in an unknown function $g(\cdot)$ in the semiparametric way discussed in previous sections.

We consider Food Engel curves in its share form, specifying the relationship between total expenditure of a household and expenditure spent on food. The dependent variable is $Y = pq/X$, where p is price, q is quantity demanded and X is total expenditure. We use data from the 1980 Spanish Family Expenditure Survey (FES) described in Alonso *et al.* (1994). This survey contains 23,972 observations with detailed information on household characteristics, total income and expenditure on several categories. The sample is designed to be representative of the Spanish population.

We want to test the following relationships:

$$\mathrm{H}_0\colon E[Y \mid X, Z] = \alpha_1 \log(X) + \alpha_2 \log(X)^2 + g_1(Z) \quad \text{a.s.} \tag{18}$$

$$\mathrm{H}_1\colon E[Y \mid X, Z] = \gamma_1 X + \gamma_2 X^{-1} + g_2(Z) \quad \text{a.s.} \tag{19}$$

where α's and γ's are parameters. Both specifications are semiparametric. In H_0 the relation between expenditure and income, given Z, is a Generalized Working–Leser (GWL) Form, whereas in H_1 the proposed Engel Curve is the one derived from the Quadratic Expenditure System (QES). Observe that in these equations no parametric form is specified for $g_1(\cdot)$ and $g_2(\cdot)$, but no interaction effects are allowed between X and Z, that is, both specifications are additive in X, Z and, hence, demographic variables are only allowed to produce changes in the intercept term. First we have estimated equations (18) and (19) separately using as vector of demographic variables $Z = (Z_1, Z_2, Z_3, Z_4)$, where $Z_1 \equiv$ Age of 'reference person' in the household (i.e. the member of the household with greatest income), $Z_2 \equiv$ Size of household, $Z_3 \equiv$ Size of the town where the household is placed (categorized into five groups according to the number of inhabitants in thousands NI: $Z_3(NI) = i$ if $NI \in I_i$, where $I_1 = (0, 2)$, $I_2 = [2, 10)$, $I_3[10, 50)$, $I_4 = [50, 200)$, $I_5 = [200, \infty)$) and $Z_4 \equiv$ Sex of reference person (1 if the reference person is a woman). For illustrative purposes, in Table I we report the semiparametric estimates which have been obtained for equations (18) and (19). These estimates were computed using the semiparametric estimation procedure described in Section 2. We used uniform k-NN weights with various values of k. We have also tested specification (18) with (19) as an alternative using the semiparametric test described in Section 2 (Test 1), and then repeated the test reversing the null and alternative hypotheses (Test 2). In Table II we report the results we obtained. As in Section 4, all computations were made with FORTRAN77 programs.

In Table I we observe that in both equations both parameters are significantly different from 0 (hereafter, all conclusions will be drawn taking $\alpha = 0{\cdot}05$ as the significance level). In Table II we observe that in both cases we reject the null hypothesis. However, the results for Test 1 are somewhat different from those obtained for Test 2 because the estimate of δ is much closer to 1 in Test 2 than in Test 1 (though in neither case would the null hypothesis $\delta = 1$ be accepted). That is, both

Table I. Semiparametric estimates (whole sample)

		GWL form (18)	
Bandwidth	$k = 160$	$k = 350$	$k = 750$
$\hat{\alpha}_1$	0·296 (0·002)	0·391 (0·036)	0·456 (0·044)
$\hat{\alpha}_2$	−0·017 (0·001)	−0·020 (0·001)	−0·023 (0·002)
		QES form (19)	
Bandwidth	$k = 160$	$k = 350$	$k = 750$
$\hat{\alpha}_1$	−1·2E − 7 (4E − 9)	−1·2E − 7 (4E − 9)	−1·2E − 7 (4E − 9)
$\hat{\alpha}_2$	1·0E + 4 (2E + 3)	9·9E + 3 (2E + 3)	9·6E + 3 (2E + 3)

Heteroscedasticity-consistent SE in parentheses.

Table II. Semiparametric test (whole sample)

	Test 1: (18) versus (19)			Test 2: (19) versus (18)		
Bandwidth	$k = 160$	$k = 350$	$k = 750$	$k = 160$	$k = 350$	$k = 750$
$\hat{\delta}$	−0·30	−0·40	−0·48	1·35	1·38	1·42
$\lvert t\rvert$	9·31	11·49	12·94	43·81	43·22	42·85
$\lvert t^*\rvert$	7·87	9·60	10·70	39·56	39·63	39·26

t is the semiparametric standard t-ratio; t^* is the semiparametric heteroscedasticity-consistent t-ratio.

models are grossly incompatible with the data. The QES form is rejected even more decisively than the GWL form, though both forms are decisively rejected.

We have searched for explanations for these negative results. First, we have analysed the performance of the test statistic in this specific situation using a Monte Carlo experiment. We have generated observations from two models which mimic the behaviour of the observations. Specifically, i.i.d. observations $\{(Y_i, X_i, Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4})\}$ were generated as follows:

Model 8: X_i is lognormal with $\log(X_i) \sim N(13{\cdot}4, 0{\cdot}52)$; Z_{i1} is $N(50{\cdot}5, 15{\cdot}1^2)$; $Z_{i2} = Z^*_{i2} + 1$ where Z^*_{i2} is Poisson with mean 2·5; Z_{i3} is discrete uniform with support $\{1, 2, 3, 4, 5\}$; Z_{i4} is a Bernoulli variable with mean 0·2; and $Y_i = \beta'_1 \times (\log(X_i), \log(X_i)^2)' + F(Z_i) + U_i$, where U_i is $N(0, 0{\cdot}01^2)$, $\beta'_1 = (0{\cdot}296, -0{\cdot}017)$, $F(Z_i) = \{I(Z_{i2} > 3) - I(Z_{i2} < 2)\}/10$. All variables are generated independently.
Model 9: All variables as before, except Y_i, which is now generated as $Y_i = (X_i, X_i^{-1}) \times \beta_2 + F(Z_i) + U_i$, where $\beta'_2 = (-1{\cdot}2E-7, 1{\cdot}0E+4)$.

Generating data in this way, the mean and variance of X_i and Z_{i1} coincide approximately with the sample mean and variance of variables 'total expenditure' and 'age of reference person'. Parameters β_1, β_2 and the variance of the error term are similar to the semiparametric estimates previously obtained. Thus, each artificial data set is, to a certain extent, similar to the observations we use. In Model 8 the GWL specification is correct and in Model 9 the QES form is the true one. Our results are based on $n = 2000$ observations and $r = 100$ replications. The test statistic was computed using uniform k-NN weights with two different k_n selected by previous inspection of some Monte Carlo samples. If nominal size is $\alpha = 0{\cdot}05$ and observations are generated from Model 8, the percentage of rejections of the null hypothesis is 0·061 ($k = 150$) or 0·052 ($k = 250$) when GWL is the null hypothesis and virtually 1·00 when QES is the null hypothesis. If $\alpha = 0{\cdot}05$ and observations are generated from Model 9, the percentage of rejections of the null hypothesis is almost 1·00 when GWL is the null hypothesis and 0·067 ($k = 150$) or 0·052 ($k = 250$) when QES is the null hypothesis. These results show that the test statistic performs adequately here. So, it seems unreasonable to think that the negative results obtained in Table II are a consequence of the poor performance of the test statistic.

It might happen that the rejection of both GWL and QES forms is a consequence of considering too heterogeneous a sample. In order to analyse this conjecture, we performed Tests 1 and 2 again but now considering only those observations corresponding to 'standard' households, i.e. those households consisting of one woman and one man between 18 and 64 years of age and one, two or three children below 18. The sample size then decreased from 23,972 to 6710, but the results were again similar to those we had previously obtained and we do not report them here. We then reduced the data set in a different way: we performed again both tests considering only those households whose total expenditure was within the (0·1, 0·8) quantiles. Table III reports the results obtained when performing the test with this data set.

We observe that the GWL form is not rejected as null hypothesis when $\alpha = 0{\cdot}05$ (if we use $k = 160$ or 350), whereas, when the QES form is the null hypothesis, δ is significantly different from 0 but not significantly different from 1. Thus, the GWL form adequately explains all results obtained with this subsample. But this is by no means a surprise. It is well known that the GWL form adequately explains the relationship between total expenditure X and share food Y, except for those observations contained in the upper tail of X (see, for example, Banks *et al.*, 1994), and those observations were not taken into account here.

Table III. Semiparametric test ($n = 16{,}779$)

Bandwidth	Test 1: (18) versus (19) $k = 160$	$k = 350$	$k = 750$	Test 2: (19) versus (18) $k = 160$	$k = 350$	$k = 750$
$\hat{\delta}$	0·53	0·50	0·95	1·13	1·34	1·85
$\|t\|$	1·33	1·12	2·13	4·35	3·28	3·49
$\|t^*\|$	1·18	1·18	2·25	4·28	3·14	3·29

t is the semiparametric standard t-ratio; t^* is the semiparametric heteroscedasticity-consistent t-ratio.

As mentioned above, in equations (18) and (19) we do not allow for interaction effects between X and Z. In order to examine to what extent interaction effects may affect our results we have also estimated equations (18) and (19) and performed the semiparametric test splitting the sample into different groups according to household characteristics. Specifically, we consider the following groups (the number of observations in each group is also reported):

Group 1: Households whose only member is a man, $n = 467$
Group 2: Urban households whose only member is a woman, $n = 1035$
Group 3: Rural households whose only member is a woman, $n = 443$
Group 4: Urban households with 2 to 5 members and a man as RP, $n = 11{,}930$
Group 5: Urban households with 2 to 5 members and a woman as RP, $n = 1328$
Group 6: Rural households with 2 to 5 members and a man as RP, $n = 4938$
Group 7: Rural households with 2 to 5 members and a woman as RP, $n = 400$
Group 8: Urban households with 6 to 12 members, $n = 2518$
Group 9: Rural households with 6 to 12 members, $n = 898$

Groups have been defined in such a way that the number of observations in each group is greater than or equal to 400 and all households in each group have similar demographic characteristics. Now we consider equations (18) and (19) for each group, but taking $Z = Z_1$ (all other demographic variables have already been taken into account on constructing groups). First, we estimated both equations for each group semiparametrically. We used as smoothing values $k = 75$ in Group 7, $k = 100$ in Groups 1 and 3, $k = 150$ in Group 9, $k = 200$ in Groups 2 and 5, $k = 250$ in Group 8, $k = 350$ in Group 6 and $k = 500$ in Group 4 (the smoothing value varies because the number of observations is different in each group). Instead of reporting all estimates we obtained, we prefer to examine graphically a selection of our results. We depict in Figures 6–8 three different estimates of the Food Engel curve for those groups with the highest number of observations (Groups 4, 6 and 8), taking $Z_1 = 50$. The three estimates we depict are: a kernel non-parametric estimate of $E[Y \mid X = x,\ Z_1 = 50]$; an estimate of the GWL form (18) (obtained from $Y = \hat{\alpha}_1 \log(X) + \hat{\alpha}_2 \log(X)^2 + \hat{g}$, where $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are the semiparametric estimates obtained, $\hat{g} \equiv \hat{m}_Y(50) - \hat{m}_1(50)\hat{\alpha}_1 - \hat{m}_2(50)\hat{\alpha}_2$ and $\hat{m}_Y(\mathcal{Z})$, $\hat{m}_1(\mathcal{Z})$, $\hat{m}_2(\mathcal{Z})$ denote k-NN non-parametric estimates of $E[Y \mid Z = \mathcal{Z}]$, $E[\log(X) \mid Z = \mathcal{Z}]$ and $E[\log(X)^2 \mid Z = \mathcal{Z}]$, respectively); and an estimate of the QES form (19) (obtained similarly). Vertical bounds are included in these figures specifying the range of X which falls within the (0·1, 0·8) quantiles.

In most cases both parameters are significantly different from 0: the only exceptions to this are α_2 in Group 5 (GWL Form) and γ_2 in Groups 3 and 7 (QES Form). If we examine Figures 6–8 we observe that in Groups 4, 6 and 8 the GWL estimate is usually closer to the non-parametric estimate than the QES estimate; this is also true in Groups 1, 7 and 9; in Groups 2 and 3 neither of the semiparametric estimates seems to be close to the non-parametric one and in Group 5 the

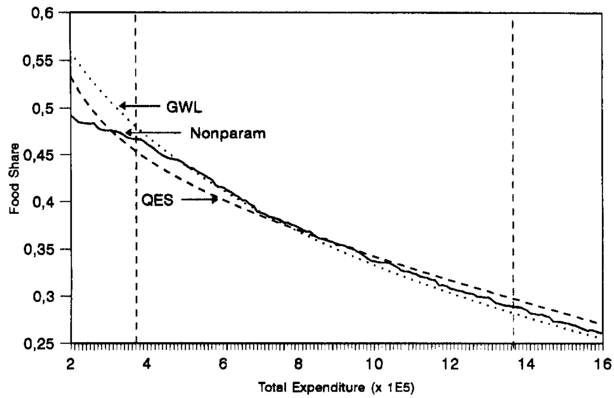

Figure 6. Food Engel curve, Group 4. Semiparametric estimates for GWL and QES forms and nonparametric estimate

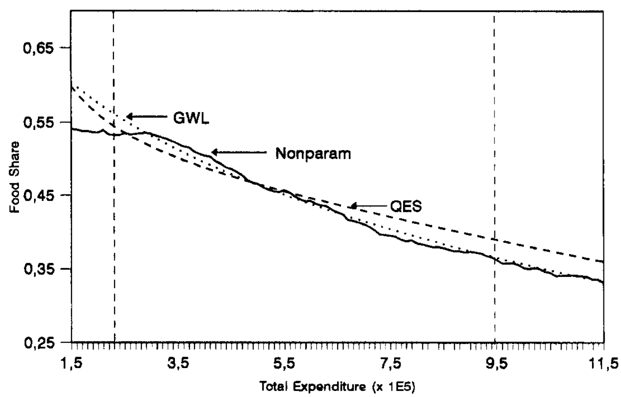

Figure 7. Food Engel curve, Group 6. Semiparametric estimates for GWL and QES forms and nonparametric estimate

QES estimate seems to perform better. As expected, when X is within the (0·1, 0·8) quantiles, in all groups all estimates are closer to each other than when X is on the tails.

Finally, we have also performed the semiparametric test separately in each group. As before, first we used the whole data in each group and then reduced the data set considering only those observations for which X was within the (0·1, 0·8) quantiles. We report our results in Tables IV and V.

Some interesting conclusions may be drawn from these tables. In Table IV we observe that in the majority of cases (six of nine) the test manages to discriminate between the two models, in favour of GWL, even when extreme values of X are kept in the sample. However, in Table V we observe that results are less conclusive when observations corresponding to the tails of X are

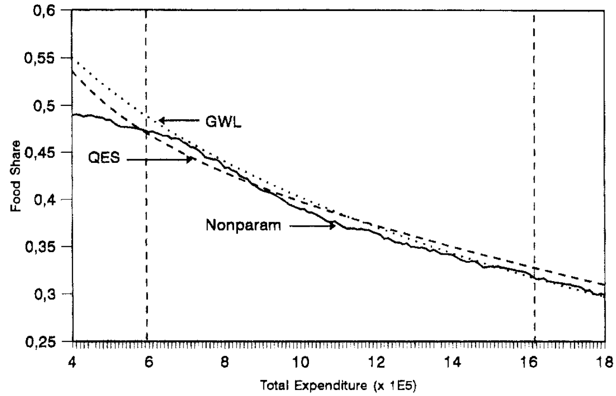

Figure 8. Food Engel curve, Group 8. Semiparametric estimates for GWL and QES forms and nonparametric estimate

Table IV. Semiparametric test (sample into groups, all observ.)

	G.1	G.2	G.3	G.4	G.5	G.6	G.7	G.8	G.9
Test 1: (18) versus (19)									
k	30	50	30	150	50	100	30	75	50
$\hat{\delta}$	0·14	−0·07	0·64	−0·47	0·17	−0·33	0·32	−0·43	−0·11
$\|t\|$	0·52	0·41	2·28	8·46	1·00	3·52	1·87	3·25	0·57
$\|t^*\|$	0·53	0·36	2·50	7·98	0·71	3·28	0·95	3·62	0·57
Test 2: (19) versus (18)									
$\hat{\delta}$	1·28	1·92	1·05	1·43	1·03	1·43	1·49	1·48	1·94
$\|t\|$	4·54	9·62	3·55	27·29	6·01	15·48	7·59	11·37	8·65
$\|t^*\|$	5·27	10·19	3·52	25·87	5·09	13·54	7·41	13·75	6·91

t is the semiparametric standard t-ratio; t^* is the semiparametric heteroscedasticity-consistent t-ratio.

Table V. Semiparametric test (sample into groups, obs. within (0·1, 0·8)-quant.)

	G.1	G.2	G.3	G.4	G.5	G.6	G.7	G.8	G.9
Test 1: (18) versus (19)									
k	25	40	25	120	40	80	25	60	40
$\hat{\delta}$	0·45	0·11	0·15	0·57	0·45	2·76	4·24	0·00	2·78
$\|t\|$	0·43	0·12	0·21	0·52	0·52	1·77	1·18	0·01	2·62
$\|t^*\|$	0·48	0·12	0·19	0·48	0·47	1·86	1·13	0·01	2·69
Test 2: (19) versus (18)									
$\hat{\delta}$	0·93	1·06	0·94	0·85	0·91	0·80	1·28	1·01	1·67
$\|t\|$	1·16	2·01	2·05	1·73	2·04	1·70	2·96	2·11	3·41
$\|t^*\|$	1·21	1·94	1·83	1·46	1·92	1·70	2·96	1·98	3·53

t is the semiparametric standard t-ratio; t^* is the semiparametric heteroscedasticity-consistent t-ratio.

excluded: the GWL form continues to perform better, but both models seem compatible with data in six cases (using heteroscedasticity-consistent t-ratios), both models are rejected in one case and the GWL form is favoured only in the other two cases.

To summarize, if demographic variables are incorporated additively to the formulation of Engel curves, then our results show that neither the GWL or QES specifications can be accepted as suitable when the whole data set (observations from the 1980 Spanish FES) is used. However, the GWL form explains adequately the results obtained when observations with extreme values of X are removed. On the other hand, if the formulation of Engel curves assumes that demographic variables may also affect the shape of the curve, then the GWL form adequately explains the results obtained in many cases when the whole data set is used and in almost all cases when observations falling in the tails of X are removed. However, the QES is incompatible with data in all cases when the whole data set is used but in some cases it may explain the results obtained when observations with extreme values of X are removed. It is worth noting, finally, that the poor performance of certain specifications when the whole data set is used may be also explained by the possible endogeneity of income and certain households characteristics (like number of members in the household) as some authors have argued (see, for instance, Deaton, 1986 or Pudney, 1989).

APPENDIX

Theorem 1. Let $\{(Y_i, X_{1i}, X_{2i}, X_{3i}, Z_i), 1 \leqslant i \leqslant n\}$ be i.i.d. observations from an $\mathbb{R} \times \mathbb{R}^{p_1} \times \mathbb{R}^{p_2} \times \mathbb{R}^{p_3}\mathbb{R}^{q}$-valued observable random variable (Y, X_1, X_2, X_3, Z) $(p_1 > 0,\ p_2 > 0,\ p_1 + p_2 + p_3 = p)$. Assume that $E|Y| < \infty$, condition (9) holds, $E\|X\|^4 < \infty$, $E[U^4] < \infty$ (where $U \equiv Y - E[Y | X, Z])$, and $\Phi \equiv E\{(X - E[X | Z])(X - E[X | Z])'\}$ is positive definite.

Suppose we use as a trimming function $I_i = I(\Sigma_{c \neq i} I(Z_c = Z_i) > 0)$ and denote

$$\Sigma_{23} \equiv \begin{pmatrix} \Phi_{22} & \Phi_{23} \\ \Phi'_{23} & \Phi_{33} \end{pmatrix} \quad \Xi \equiv \begin{pmatrix} \Phi'_{12} & \Phi_{23} \\ \Phi'_{13} & \Phi_{33} \end{pmatrix} \quad \begin{pmatrix} \lambda_2 \\ \lambda_3 \end{pmatrix} \equiv \Sigma_{23}^{-1} \times \Xi \times \begin{pmatrix} \beta_1 \\ \beta_3 \end{pmatrix}$$

where $\Phi_{rs} \equiv E\{(X_r - E[X_r | Z])(X_s - E[X_s | Z])'\}$. Then we will show that:

(1) If $\mathrm{Var}(Y | X, Z) = \sigma^2 \in (0, \infty)$ and we use non-smoothing estimates $\hat{m}^{(1)}_{\zeta i}$ for $\zeta = X_1, X_2, X_3, Y$, then under H_0 (i.e. if equation (1) holds), if $\lambda_2 \neq 0$, $\hat{J}^{(1)}_n \xrightarrow{d} N(0, 1)$ and under H_1 (if equation (2) holds) if $\beta_2 \neq 0$ $\forall \rho > 0$ $\lim_{n \to \infty} Pr(|\hat{J}^{(1)}_n| > \rho) = 1$.

(2) If $\mathrm{Var}(Y | X, Z) = \sigma^2 \in (0, \infty)$, $E[g(Z)^2] < \infty$, we use k-NN estimates $\hat{m}^{(2)}_{\zeta i}$ with uniform, quadratic or triangular weights and $k_n/n + 1/k_n = o(1)$ then, under H_0, if $\lambda_2 \neq 0$, $\hat{J}^{(2)}_n \xrightarrow{d} n(0, 1)$ and, under H_1, if $\beta_2 \neq 0$ $\forall \rho > 0$ $\lim_{n \to \infty} Pr|\hat{J}^{(2)}_n| > \rho) = 1$.

(3) If $\mathrm{Var}(Y | X, Z) = \sigma^2(X, Z) \in (0, \infty)$ and we use non-smoothing estimates $\hat{m}^{(1)}_{\zeta i}$ then, under H_0, if $\hat{\lambda}_2 \neq 0$, $\hat{J}^{(1)^*}_n \xrightarrow{d} n(0, 1)$ and under H_1, if $\beta_2 \neq 0, \forall \rho > 0$ $\lim_{n \to \infty} Pr(|\hat{J}^{(1)^*}_n| > \rho) = 1$, where $J_n^{(1)^*}$ is the heteroscedasticity-consistent t-ratio, constructed as specified below.

Proof. (i) First suppose that H_0 is true. Denote $\hat{\varepsilon}_{Xi} \equiv (\hat{\varepsilon}'_{1i}, \hat{\varepsilon}'_{2i}, \hat{\varepsilon}'_{3i})'$, $\hat{\Phi} \equiv n^{-1}\Sigma_i \hat{\varepsilon}_{Xi}\hat{\varepsilon}'_{Xi} I_i$, $\hat{\omega}_i \equiv (\hat{\varepsilon}'_{2i}, \hat{\varepsilon}'_{3i})'$. Then $(\bar{\beta}'_2, \bar{\beta}'_3)' = \{n^{-1}\Sigma_i \hat{\omega}_i \hat{\omega}'_i I_i\}^{-1} n^{-1} \Sigma_i \hat{\omega}_i \hat{\varepsilon}_{Yi} I_i$. Using equation (A.4) in Delgado and Mora (1995a) (hereafter DM), $\hat{\Phi} - \Phi = o_P(1)$ and thus $n^{-1}\Sigma_i \hat{\omega}_i \hat{\omega}'_i I_i - \Sigma_{23} = o_P(1)$. Moreover, if $I_i = 1$, then $\Sigma_{j \neq i} W_{nj}(Z_i) g(Z_j) = g(Z_i)$ and, therefore, $\hat{\varepsilon}_{Yi} I_i = \hat{\varepsilon}'_{1i}\beta_1 I_i + \hat{\varepsilon}'_{3i}\beta_3 I_i + \hat{\varepsilon}_{Ui} I_i$. Hence, $n^{-1}\Sigma_i \hat{\omega}_i \hat{\varepsilon}_{Yi} I_i = n^{-1} \Sigma_i \hat{\omega}_i (\hat{\varepsilon}'_{1i}, \hat{\varepsilon}'_{3i}) I_i \times (\beta'_1, \beta'_3)' + n^{-1}\Sigma_i \hat{\omega}_i \hat{\varepsilon}_{Ui} I_i$. Now, the second term here is $o_P(1)$ (using equation (A.3) in DM) and the first term converges to $\Xi \times (\beta'_1, \beta'_3)'$. Thus

$(\bar{\beta}_2' - \lambda_2', \bar{\beta}_3' - \lambda_3')' = o_P(1)$. Assume now that $\lambda_2 \neq 0$, and denote $\Gamma \equiv H(\lambda_2)'\Phi H(\lambda_2)$ where, for $u \in \mathbb{R}^{p_2}$, $H(u)$ denotes the $p \times (p_1 + 1 + p_3)$ block-diagonal matrix with the first block, I_{p_1}, the second block u and the third block I_{p_3}. Observe that Γ is non-singular because $\lambda_2 \neq 0$. As H_0 is true, with the same reasoning as before, if $\hat{\Gamma} \equiv H(\bar{\beta}_2)'\hat{\Phi}H(\bar{\beta}_2)$, then $n^{1/2}(\tilde{\gamma}_1' - \beta_1', \tilde{\delta}, \tilde{\gamma}_3' - \beta_3')' = \hat{\Gamma}^{-1}H(\bar{\beta}_2)'n^{-1/2}\Sigma_i\hat{\varepsilon}_{Xi}\hat{\varepsilon}_{Ui}I_i$. Now, $n^{-1/2}\Sigma_i\hat{\varepsilon}_{Xi}\hat{\varepsilon}_{Ui}I_i$ converges in distribution to $N(0, \sigma^2\Phi)$, by equation (A.3) in DM; moreover, as $\bar{\beta}_2 - \lambda_2 = o_P(1)$ and equation (A.4) in DM holds, then $H(\bar{\beta}_2) - H(\lambda_2) = o_P(1)$ and $\hat{\Gamma}^{-1} - \Gamma^{-1} = o_P(1)$. Thus, $n^{1/2}(\tilde{\gamma}_1' - \beta_1', \tilde{\delta}, \tilde{\gamma}_3' - \beta_3')' \xrightarrow{d} N(0, \sigma^2\Gamma^{-1})$. On the other hand, $\tilde{\sigma}^2 - \sigma^2 = o_P(1)$, where $\tilde{\sigma}^2$ is the OLS estimate of the error variance in equation (8), because $\tilde{\sigma}^2 = n^{-1}\Sigma_i\{\hat{\varepsilon}_{1i}'(\beta_1 - \hat{\gamma}_1)I_i + \hat{\varepsilon}_{3i}'(\beta_3 - \tilde{\gamma}_3)I_i + \hat{\varepsilon}_{Ui}I_i - \hat{\varepsilon}_{2i}'\bar{\beta}_2\tilde{\delta}I_i\}^2$ and all terms here are $o_P(1)$ except $n^{-1}\Sigma_i\hat{\varepsilon}_{Ui}^2I_i$, which converges to σ^2 as DM prove in their equation (A.4). The asymptotic result for $\hat{J}_n^{(1)}$ under H_0 follows now because $\hat{J}_n^{(1)} = n^{1/2}\tilde{\delta}/(\tilde{\sigma}^2\hat{a})^{1/2}$, where $\hat{a}$ denotes the $(p_1 + 1)$th diagonal element in $\hat{\Gamma}^{-1}$.

Now suppose that H_1 is true. Then, as a consequence from Theorem 2 in DM, $n^{1/2}(\bar{\beta}_2' - \beta_2', \bar{\beta}_3' - \beta_3')' \xrightarrow{d} N(0, \sigma^2\Sigma_{23}^{-1})$. Moreover, $\hat{\varepsilon}_{Yi}I_i = (\hat{\varepsilon}_{1i}', \hat{\varepsilon}_{2i}'\beta_2, \hat{\varepsilon}_{3i}') \times (0, 1, \beta_3')'I_i + (\beta_2 - \bar{\beta}_2)'\ \hat{\varepsilon}_{2i}I_i + \hat{\varepsilon}_{Ui}I_i$. Hence, $(\tilde{\gamma}_1', \tilde{\delta}, \tilde{\gamma}_3')' = (0', 1, \beta_3')' + \hat{\Gamma}^{-1}H(\bar{\beta}_2)' \times \{n^{-1}\Sigma_i\hat{\varepsilon}_{Xi}\hat{\varepsilon}_{2i}'I_i\}(\beta_2 - \bar{\beta}_2) + \hat{\Gamma}^{-1}H(\bar{\beta}_2)' \times \{n^{-1}\Sigma_i\hat{\varepsilon}_{Xi}\hat{\varepsilon}_{Ui}I_i\}$. As $\beta_2 \neq 0$, $H(\beta_2)'\Phi H(\beta_2)$ is non-singular and $\hat{\Gamma}^{-1} - \{H(\beta_2)'\Phi H(\beta_2)\}^{-1} = o_P(1)$. Thus, $(\tilde{\gamma}_1', \tilde{\delta} - 1, \tilde{\gamma}_3' - \beta_3')' = o_P(1)$. On the other hand, $\tilde{\sigma}^2 - \sigma^2 = o_P(1)$ as before. The asymptotic result for $\hat{J}_n^{(1)}$ under H_1 follows from these results.

(ii) Follows similarly as part (i) using the Corollary in DM.

(iii) In the heteroscedastic model, conditions (A.3) and (A.4) in DM no longer hold. Instead we have that $n^{-1/2}\Sigma_i\hat{\varepsilon}_{Xi}\hat{\varepsilon}_{Ui}I_i \xrightarrow{d} N(0, \Phi^{-1}\Omega\Phi^{-1})$, $\hat{\Phi} - \Phi = o_P(1)$, $\hat{\Omega} - \Omega = o_P(1)$ where $\Omega \equiv E\{\sigma^2(X, Z)(X - E[X \mid Z])(X - E[X \mid Z])'\}$ and $\hat{\Omega} \equiv n^{-1}\Sigma_i\hat{\varepsilon}_{Xi}\hat{\varepsilon}_{Xi}'\hat{\varepsilon}_{Ui}^2I_i$. These conditions follow in a similar way to equation (A.3) and (A.4) in DM. The heteroscedasticity-consistent t-ratio is now $J_n^{(1)*} \equiv n^{1/2}\tilde{\delta}/\hat{b}^{1/2}$, where $\hat{b}$ denotes the $(p_1 + 1)$th diagonal element in $\hat{\Gamma}^{-1}H(\bar{\beta}_2)'\hat{\Omega}H(\bar{\beta}_2)\hat{\Gamma}^{-1}$. As in part (i), it may be proved that $\bar{\beta}_2 - \lambda_2 = o_P(1)$ and if $\lambda_2 \neq 0$, $n^{1/2}(\tilde{\gamma}_1' - \beta_1', \tilde{\delta}, \tilde{\gamma}_3' - \beta_3')' \xrightarrow{d} N(0, \Gamma^{-1}H(\lambda_2)'\Omega H(\lambda_2)\Gamma^{-1})$ and hence $J_n^{(1)*} \xrightarrow{d} N(0, 1)$. Similarly under H_1 $n^{1/2}(\bar{\beta}_2' - \beta_2', \bar{\beta}_3' - \beta_3')' \xrightarrow{d} N(0, \Sigma_{23}^{-1}E[\sigma^2(X, Z)(\varepsilon_2', \varepsilon_3')'(\varepsilon_2', \varepsilon_3')]\Sigma_{23}^{-1})$ and if $\beta_2 \neq 0$ then $(\hat{\gamma}_1', \tilde{\delta} - 1, \hat{\gamma}_3' - \beta_3')' = o_P(1)$ and $\forall \rho > 0$ $Pr(| \hat{J}_n^{(1)*} | > \rho) \to 1$. □

Theorem 2. All results stated in Theorem 1(i) hold when assumption (9) is replaced by (12), non-parametric estimates $m_{\zeta i}^{(1)}$ are replaced by $m_{\zeta i}^{(3)}$, the trimming function I_i is replaced by $I_i^* \equiv I[(nh_n^s)^{-1}\Sigma_c\Psi_{ic}(h_n)I(Z_i^{(1)} = Z_c^{(1)}) > b_n]$ (where $b_n > 0$ is a sequence of trimming values) and it is also assumed that:

$$U \equiv Y - E[Y \mid X, Z] \text{ and } (X, Z) \text{ are independent} \tag{A1}$$

$$\exists \nu \in \mathbb{N} : f_{\mathcal{Z}} \in \mathcal{G}_\nu^\infty, \theta_{\mathcal{Z}} \in \mathcal{G}_\nu^4 \text{ and } \xi_{\mathcal{Z}} \in \mathcal{G}_\nu^2 \text{ uniformly in } \mathcal{D} \tag{A2}$$

$$b_n \to 0,\ nb_n^{-4}h_n^{4\nu} \to 0,\ nb_n^4h_n^{2s} \to \infty \text{ (as } n \to \infty) \text{ and} \tag{A3}$$

$$\text{The kernel function } \psi \text{ is in class } \mathcal{K}_{2\nu-1} \tag{A4}$$

where, given $\mathcal{Z} \in \mathcal{D} \subseteq \mathbb{R}^r$, $f_{\mathcal{Z}} : \mathbb{R}^s \to \mathbb{R}$ denotes the density function of $Z^{(2)} \mid Z^{(1)} = \mathcal{Z}$; $\theta_{\mathcal{Z}}$: $\mathbb{R}^s \to \mathbb{R}$ is defined as $\theta_{\mathcal{Z}}(a) = g(\mathcal{Z}, a)$ for $a \in \mathbb{R}^s$ ($g(.,.)$ as in equation (1) for $(\mathcal{Z}, a) \in \mathbb{R}^q$); and $\xi_{\mathcal{Z}}$: $\mathbb{R}^s \to \mathbb{R}$ is defined as $\xi_{\mathcal{Z}}(a) = E[X \mid Z^{(1)} = \mathcal{Z}, Z^{(2)} = a]$ for $a \in \mathbb{R}^s$.

Proof: First we comment briefly on assumptions in Theorem 2: classes $\mathcal{G}_\mu^\lambda$, $\mathcal{G}_\mu^\infty$ and $\mathcal{K}_f$ (for $\lambda > 0$, $\mu > 0$ and $f \in \mathbb{N}$) are as defined in Robinson (1988); 'uniformly in $\mathcal{D}$' means that the constants which appear in the definition do not depend on $\mathcal{Z}$; equation (A2) specifies the degree of smoothness in f_Z, θ_Z and ξ_Z which is required; equation (A3) gives conditions on the rate of convergence of h_n and b_n; equation (A4) specifies the relation between the degree of smoothness and the order of kernel $\psi(\cdot)$; note that equation (A3) implies $2v > s$ (so, ψ is at least of order s).

Theorem 2 follows in the same way as Theorem 1 replacing references to DM by references to generalizations of Propositions 1–14 in Robinson (1988). The latter contains results only for a partly linear regression model with absolutely continuous Z, but, adding uniformness conditions, they may be easily extended to the case when Z contains discrete and absolutely continuous random variables (Delgado and Mora, 1995b). Observe that if we rewrite conditions $vi - x$ in Robinson (1988) with $\lambda = \mu = \nu \equiv v$, we obtain equations (A2)–(A4). In fact it would have been possible to give Theorem 2 with weaker smoothness conditions (allowing for different degree of smoothness in f_Z, θ_Z and ξ_Z as in Robinson, 1988), but we have preferred this version for simplicity. □

ACKNOWLEDGEMENTS

We are grateful to three anonymous referees for their valuable comments on a previous version of the paper. This article is based on research funded by Spanish Direccion General de Enseñanza Superior (DGES), reference number PB95-0292. Financial support from IVIE is also gratefully acknowledged.

REFERENCES

Alonso, M. D., A. Lara, R. Arevalo and J. Ruiz-Castillo (1994), 'La encuesta de presupuestos familiares de 1980–81', DT 94-12, Universidad Carlos III de Madrid.

Banks, J., R. Blundell and A. Lewbel (1994), 'Quadratic Engel curves, welfare measurement and consumer demand', Discussion Paper, University College, London.

→ Chen, H. (1988), 'Convergence rates for parametric components in a partly linear model', *Annals of Statistics*, **16**, 136–46.

→ Davidson, R. and J. MacKinnon (1981), 'Several tests for model specification in the presence of alternative hypotheses', *Econometrica*, **49**, 781–93.

Davidson, R. and J. MacKinnon (1993), *Estimation and inference in Econometrics*, Oxford University Press, New York.

Davidson, R. and J. MacKinnon (1994), 'Graphical methods for investigating the size and power of hypothesis tests', Queen's University, Institute for Economic Research, Discussion Paper No. 903.

Deaton, A. (1981), 'Three essays on a Sri Lanka household survey', *Living Standard Measurement Study*, Working Paper No. 11, Washington: World Bank.

Deaton, A. (1986), 'Demand analysis', in Z. Griliches and M. Intrilligator (eds), *Handbook of Econometrics*, Elsevier, New York.

Deaton, A. and J. Muellbauer (1980), 'An almost ideal demand system', *American Economic Review*, **70**, 312–26.

Delgado, M. A. and J. Mora (1995a), 'Nonparametric and semiparametric estimation with discrete regressors', *Econometrica*, **63**, 1477–85.

Delgado, M. A. and J. Mora (1995b), 'On asymptotic inferences in nonparametric and semiparametric models with discrete and mixed regressors', *Investigaciones Economicas*, **19**, 435–67.

Engel, E. (1895), 'Die lebenskosten belgischer arbeiterfamilien früher und jetzt', *International Statistical Institute Bulletin*, **9**, 1–74.

Fan, Y., Q. Li and T. Stengos (1995), 'Root-*n*-consistent semiparametric regression with conditional heteroskedastic disturbances', *Journal of Quantitative Economics*, **11**, 229–40.

Godfrey, L. and M. H. Pesaran (1983), 'Tests of non-nested regression models: small sample adjustments and Monte Carlo evidence', *Journal of Econometrics*, **21**, 133–54.

Gozalo, P. L. (1992), 'Nonparametric analysis of Engel curves: Estimation and Testing of Demographic Effects', Working Paper No. 92-15, Brown University.
→ Härdle, W. and E. Mammen (1993), 'Comparing nonparametric versus parametric regression fits', *Annals of Statistics*, **21**, 1926–47.
→ Leser, C. E. V. (1963), 'Forms of Engel curve functions', *Econometrica*, **31**, 694–703.
Lewbel, A. (1991), 'The rank of demand systems: theory and nonparametric estimation', *Econometrica*, **59**, 377–91.
→ Linton, O. (1995), 'Second order approximation in the partially linear regression model', *Econometrica*, **63**, 1079–112.
→ Pesaran, M. H. (1974), 'On the general problem of model selection', *Review of Economic Studies*, **41**, 153–71.
Pollak, R. A. and T. J. Wales (1978), 'Estimation of complete demand systems from household budget data: the linear and quadratic expenditure systems', *American Economic Review*, **68**, 348–59.
→ Pollak, R. A. and T. J. Wales (1980), 'Comparison of the quadratic expenditure system and translog demand systems with alternative specifications of demographic effects', *Econometrica*, **48**, 595–612.
Pudney, S. (1989), *Modelling Individual Choice*, Blackwell, Oxford.
→ Robinson, P. M. (1988), 'Root-*n*-consistent semiparametric regression', *Econometrica*, **56**, 931–54.
→ Speckman, P. (1988), 'Kernel smoothing in partially linear models', *Journal of the Royal Statistical Society B*, **50**, 413–46.
Stone, C. J. (1977), 'Consistent nonparametric regression', *Annals of Statistics*, **4**, 595–645.
→ Working, H. (1943), 'Statistical laws of family expenditure', *Journal of the American Statistical Association*, **38**, 43–56.