ON THE PERFORMANCE OF NONPARAMETRIC SPECIFICATION TESTS IN REGRESSION MODELS[*]

Daniel Miles and Juan Mora[**]

WP-AD 2002-13

Correspondence to: Juan Mora. Universidad de Alicante, Departamento de Fundamentos del Análisis Económico, Apartado de Correos, 99, E-03080 Alicante (Spain). Tel.: +34 965903614 / Fax: +34 965903898 / E-mail: juan@merlin.fae.ua.es.

[**] Daniel Miles: University of Vigo; Juan Mora: University of Alicante.

ON THE PERFORMANCE OF NONPARAMETRIC SPECIFICATION TESTS IN REGRESSION MODELS

Daniel Miles and Juan Mora

A B S T R A C T

Some recently developed nonparametric specification tests for regression models are described in a unified way. The common characteristic of these tests is that they are consistent against any alternative hypothesis. The performance of the test statistics is compared by means of Monte Carlo simulations, analysing how heteroskedasticity, number of regressors and bandwidth selection influence the results. The statistics which do not use a bandwidth perform slightly better if the regression model has only one regressor; otherwise, some of the statistics which use a bandwidth behave better if the bandwidth is chosen adequately. These statistics are applied to test the specification of three commonly used Mincer-type wage equations with Uruguayan and Spanish data; all of them are rejected.

KEYWORDS: Nonparametric Specification Tests; Wage equations.

1 INTRODUCTION

The usual approach for examining a relation among economic variables is to specify a functional form depending on unknown parameters that are then estimated and tested using a data set. This approach, however, is problematic since economic theory often does not suggest such functional forms. At best it results in a list of potential explanatory variables. Further, if the chosen functional form does not properly capture the nature of the data, there is a danger of arriving at incorrect conclusions. Moreover, most empirical studies do not test the assumed parametric specification properly through the use of a general alternative hypothesis.

The first concern of this paper is to present newly developed, appropriate nonparametric specification tests and to compare their performance with moderate sample sizes by means of simulations. Specifically, we are interested in analysing how bandwidth selection may affect the behaviour of the statistics, how sensitive they are to the homoskedasticity assumption, and to what extent the performance of the statistics improves when bootstrap critical values are used instead of asymptotic ones.

The second concern of this paper is to apply these procedures to study the appropriateness of certain wage equations. Characterizing wage profiles is a long-standing issue, given its key role in understanding poverty, internal migrations, wage inequality, the performance of earnings-based pension plans, the consequences of implementing universal educational programmes and so on (see Manski 2000). The Mincer-type wage equation, which specifies the logarithm of wage as a linear combination of years of schooling, experience, squared experience and other individual characteristics, is the functional form that has been most commonly used to estimate expected earnings profiles. As Willis (1986) states, as an empirical tool "the Mincer earnings function has been one of the great success stories of modern labour economics". In this paper we analyse the validity of various Mincer-type wage equations using Uruguayan and Spanish data.

The remainder of the paper is structured as follows. Section 2 describes the test-statistics considered here. Section 3 presents the results of three Monte Carlo experiments. Section 4 describes Mincer-type wage equations, the data sets and the empirical results. Finally, in Section 5, the conclusions are presented.

2 NONPARAMETRIC TESTS FOR SPECIFICATION OF RE-GRESSION MODELS

In this section, we present the basic ideas behind the non-parametric statistics recently developed to test the specification of a regression model. Let (X, Y) denote a $\mathbb{R}^d \times \mathbb{R}$ random vector and assume that Y is integrable so that the regression function $m(x) = E(Y \mid X = x)$, $x \in \mathbb{R}^d$, is well defined. Parametric modelling assumes that $m(\cdot)$ belongs to a given family $\mathcal{M} = \{m(\cdot, \theta) : \theta \in \Theta \subset \mathbb{R}^p\}$, i.e. $m(x) = m(x, \theta_0)$ for some "true" parameter value θ_0. Therefore, statistical inference based on $\mathcal{M}$ should be accompanied by a test for $H_0 : m(\cdot) \in \mathcal{M}$ versus the non-parametric alternative $H_1 : m(\cdot) \notin \mathcal{M}$.

The procedures proposed in the literature for performing tests of this type can be classified into two groups: those that use a smoothing value ("smoothing tests") and those that do not ("non-smoothing tests"). All of these procedures require the use of a parametric estimator of $m(x)$ under H_0; the natural one is $m(x, \widehat{\theta})$, where $\widehat{\theta}$ is any root-n-consistent estimator of θ_0, e.g. the nonlinear least squares estimator under certain assumptions. The smoothing tests also require the use of a nonparametric estimator of $m(x)$; the most popular one is the Nadaraya-Watson estimator: $\widehat{m}_h(x) = \sum_{i=1}^{n} K[(x - X_i)/h]Y_i / \sum_{i=1}^{n} K[(x - X_i)/h]$, where $\{(X_i, Y_i)\}_{i=1}^{n}$ are the observations, h is a smoothing value (or bandwidth) and $K(\cdot) : \mathbb{R}^d \to \mathbb{R}$ is a symmetric kernel function (hereafter, $0/0$ is arbitrarily defined to be 0). The main advantage of the smoothing tests over the non-smoothing ones is that their null asymptotic distribution is known; their main disadvantage is that their performance (size and power) depends, crucially, on the choice of the smoothing value. We now briefly describe the test-statistics that will be considered here. In each case, we discuss how critical values can be obtained, assuming that certain technical conditions, which we do not specify, are satisfied. The wild bootstrap procedure which is sometimes required to obtain critical values is described at the end of this section. Under H_1, all of these statistics diverge to $+\infty$, hence, the critical region is always one-sided.

The statistic proposed by Gozalo (1993) is based on the squared difference between the parametric estimator and the Bierens-Nadaraya-Watson nonparametric estimator at L points $\{x_l\}_{l=1}^{L}$ in the support of X :

$$T_n^{(G)} = nh^d \sum_{l=1}^{L} \frac{[m(x_l, \widehat{\theta}) - \widehat{m}_{h,\delta}(x_l)]^2}{\widehat{v}_h(x_l)},$$

where $\widehat{m}_{h,\delta}(x) = [\widehat{m}_h(x) - (\frac{h}{s})^r \widehat{m}_s(x)]/[1 - (\frac{h}{s})^r]$, $\widehat{m}_s(x)$ is defined as $\widehat{m}_h(x)$ replacing h by $s = hn^{(1-\delta)/(2r+d)}$, δ is any value in $(0, 1)$, r is the order of the kernel function, $\widehat{v}_h(x) = c_K \widehat{V}(x)/\{\frac{1}{nh^d} \sum_{i=1}^{n} K[(x - X_i)/h]\}$, $c_K = \int K(u)^2 du$ and $\widehat{V}(x)$ is a consistent estimator of $Var(Y \mid X = x)$. If this conditional variance

3

is assumed to be constant, it is possible to use $\widehat{V}(x) = n^{-1}\sum_{i=1}^{n}[Y_i - \widehat{m}_h(X_i)]^2$; but if no parametric form is assumed for it, a sensible choice would be $\widehat{V}(x) = \sum_{i=1}^{n} K[(x - X_i)/h][Y_i - \widehat{m}_h(X_i)]^2/\sum_{i=1}^{n} K[(x - X_i)/h]$. The asymptotic null distribution of $T_n^{(G)}$ is χ_L^2. This test-statistic is only consistent if the regression function specified under H_0 and the true regression function are different at x_l for some $l \in \{1, ..., L\}$.

Härdle and Mammen (1993) base their statistic on the weighted squared difference between the Nadaraya-Watson estimator and a kernel-smoothed parametric estimator:

$$T_n^{(HM)} = nh^{d/2}\int[\widehat{m}_h(x) - \widehat{m}_{h,\widehat{\theta}}(x)]^2\pi(x)dx,$$

where $\widehat{m}_{h,\widehat{\theta}}(x) = \sum_{i=1}^{n} K[(x-X_i)/h]m(X_i,\widehat{\theta})/\sum_{i=1}^{n} K[(x-X_i)/h]$, $\pi(\cdot) : \mathbb{R}^d \to \mathbb{R}$ is a weight function, the range of integration is the support of X and the integral can be numerically approximated. The asymptotic null distribution of $T_n^{(HM)}$ is normal, but the authors recommend computing critical values using a boostrap procedure described below. This test-statistic is consistent against any alternative hypothesis.

The statistic proposed by Zheng (1996) can be viewed as a conditional moment test-statistic based on the following moment condition, which holds under the null hypothesis: $E[UE(U \mid X)p(X)] = 0$, where $U = Y - m(X, \theta_0)$ and $p(\cdot)$ is the density function of X. This moment condition suggests basing the statistic on $n^{-1}\sum_{i=1}^{n} \hat{U}_i\hat{E}(U_i \mid X_i)$, where $\hat{U}_i$ and $\hat{E}(U_i \mid X_i)$ are suitable estimates of U_i and $E(U_i \mid X_i)$. Specifically, using kernel smoothers and parametric residuals $e_i = Y_i - m(X_i, \widehat{\theta})$, the statistic they derive is:

$$T_n^{(Z)} = \frac{\sum_{i=1}^{n} e_i \sum_{j=1,j\neq i}^{n} K[(X_i - X_j)/h]e_j}{\left(2\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\{K[(X_i - X_j)/h]e_ie_j\}^2\right)^{1/2}}.$$

The asymptotic null distribution of $T_n^{(Z)}$ is standard normal, and it is also consistent against any alternative. Li and Wang (1998) have proven that the performance of $T_n^{(Z)}$ could be improved if critical values are obtained using a bootstrap procedure.

The idea behind the statistic proposed by Ellison and Ellison (2000) is the same as the one in Zheng (1996), but the sample analogue is constructed using normalized weights whose sum is one for each observation and, additionally, a finite-sample correction term is included. When kernel weights are used, the statistic they propose becomes:

$$T_n^{(EE)} = \frac{\sum_{i=1}^{n} e_i \sum_{j=1,j\neq i}^{n} K_{h,ij}e_j}{\left(2\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} K_{h,ij}^2 e_i^2 e_j^2\right)^{1/2}} + \frac{1+d}{\left(2\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} K_{h,ij}^2\right)^{1/2}},$$

where $K_{h,ij} = \frac{1}{2}K[(X_i-X_j)/h](\{\sum_{l=1,l\neq i}^{n} K[(X_i-X_l)/h]\}^{-1} + \{\sum_{l=1,l\neq j}^{n} K[(X_j - X_l)/h]\}^{-1})$. The asymptotic null distribution of $T_n^{(EE)}$ is standard normal; it is also consistent against any alternative hypothesis.

Horowitz and Härdle (1994) consider a slightly different context from ours, but their statistic can also be used to test our null and alternative hypotheses; here, we adapt their motivation and statistic to our context. Their statistic can be viewed as a conditional-moment test-statistic, based on the moment condition $E\{UE[U \mid m(X,\theta_0)]w[m(X,\theta_0)]\} = 0$, which holds under H_0 for any non-negative weight function $w(\cdot) : \mathbb{R} \to \mathbb{R}$. This moment condition suggests basing the statistic on $n^{-1}\sum_{i=1}^{n} \hat{U}_i[\hat{Y}_i - m(X_i,\hat{\theta})]$, where $\hat{U}_i$ and $\hat{Y}_i$ are suitable estimates of U_i and $E[Y_i \mid m(X_i,\theta_0)]$. Using kernels, and after certain normalizations, the statistic derived is:

$$T_n^{(HH)} = \frac{\sum_{j=1}^{n} e_i[\widehat{Y}_{h,\delta,j} - m(X_j,\widehat{\theta})]w[m(X_j,\widehat{\theta})]}{\{\frac{2c_K}{nh}\sum_{j=1}^{n} \widehat{V}_j^2 w[m(X_j,\widehat{\theta})]^2/\widehat{p}_{h,j}\}^{1/2}},$$

where $\widehat{Y}_{h,\delta,j} = [\widehat{Y}_{h,j} - (\frac{h}{s})^r\widehat{Y}_{s,j}]/[1-(\frac{h}{s})^r]$, $\widehat{Y}_{h,j} = (nh\widehat{p}_{h,j})^{-1}\sum_{i=1,i\neq j}^{n} K\{[m(X_j,\widehat{\theta}) - m(X_i,\widehat{\theta})]/h\}Y_i$, $\widehat{p}_{h,j} = (nh)^{-1}\sum_{i=1,i\neq j}^{n} K\{[m(X_j,\widehat{\theta}) - m(X_i,\widehat{\theta})]/h\}$, $\widehat{Y}_{s,j}$, $\widehat{p}_{s,j}$ are defined as $\widehat{Y}_{h,j}$, $\widehat{p}_{h,j}$ replacing h by $s = hn^{(1-\delta)/(2r+d)}$, and $\widehat{V}_j$ is a consistent estimator of $Var[Y_j \mid m(X_j,\theta)]$. If this conditional variance is assumed to be constant, it is possible to use $\widehat{V}_j = n^{-1}\sum_{i=1}^{n}(Y_i - \widehat{Y}_{h,\delta,i})^2$; but if no parametric form is assumed for it, a sensible choice would be $\widehat{V}_j = (nh\widehat{p}_{h,j})^{-1}\sum_{i=1,i\neq j}^{n} K\{[m(X_j,\widehat{\theta}) - m(X_i,\widehat{\theta})]/h\}(Y_i - \widehat{Y}_{h,\delta,i})^2$. Under H_0, $T_n^{(HH)}$ converges to a standard normal distribution. The main advantage of $T_n^{(HH)}$ over the other smoothing tests is that the nonparametric estimation that is made has only one explanatory variable; hence, the "curse of dimensionality" of nonparametric estimators is avoided and the statistic is expected to perform well, even with moderate sample sizes. The disadvantage of $T_n^{(HH)}$ is that, unlike all the other statistics described here, it is not consistent against all alternatives, but only against those for which $E(\{E[Y \lfloor m(X,\theta_1)] - m(X,\theta_1)\}^2 w[m(X,\theta_1)]) > 0$, where θ_1 is the probability limit of $\widehat{\theta}$.

The non-smoothing tests base their statistics on empirical processes defined in such a way that any deviation from the null hypothesis is detected. Bierens (1990) and Bierens and Ploberger (1997) suggest considering the empirical process $Z_n(t) = n^{-1/2}\sum_{i=1}^{n} e_i\omega(t, X_i)$, for $t \in \Xi \subseteq \mathbb{R}^d$ and a smooth function $\omega(\cdot) : \Xi \times \mathbb{R}^d \to \mathbb{R}$, and then using a functional of this process as the test-statistic. If $\omega(\cdot)$ is chosen adequately, statistics constructed in this way can be viewed as conditional-moment test-statistics based on an infinite set of moment conditions; hence, they are consistent against all possible alternatives. Here, we consider the Bierens statistic $\int Z_n(t)^2 dF(t)$, where $F(\cdot)$ is the product of d standard normal distribution functions, $\omega(t, X) = \exp[t'\Phi(X)]$ for a suitable function $\Phi(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$, and the range of integration is $\mathbb{R}^d$. The

statistic thus obtained is:

$$T_n^{(B)} = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} e_i e_j \exp[\frac{1}{2} \sum_{l=1}^{d} (Z_{il} + Z_{jl})^2],$$

where Z_{il} is the lth component of $Z_i = \Phi(X_i)$. Note that our choice of $F(\cdot)$ does not satisfy the requirement of compact support contained in Bierens and Ploberger (1997), but we follow the suggestion of Fan and Li (2000), who study this version of the Bierens statistic and show that the asymptotic null distribution of $T_n^{(B)}$ can be approximated by a bootstrap procedure.

Stute (1997) observes that H_0 holds if and only if $E[UI(X \le t)] = 0$ for all $t \in \mathbb{R}^d$, where $I(\cdot)$ is the indicator function. Hence, he proposes considering the empirical process $R_n(t) = n^{-1/2} \sum_{i=1}^{n} e_i I(X_i \le t)$, and using a functional of this process as the test-statistic. Here, we consider the Cramér-von Mises statistic

$$T_n^{(S)} = n^{-1} \sum_{j=1}^{n} R_n(X_j)^2 = n^{-2} \sum_{j=1}^{n} [\sum_{i=1}^{n} e_i I(X_i \le X_j)]^2,$$

whose asymptotic distribution, studied in detail in Stute (1997), can also be approximated by bootstrap.

The bootstrap procedure which allows us to obtain critical values when using $T_n^{(B)}$ and $T_n^{(S)}$ is usually referred to as "wild bootstrap" and works as follows: i) generate independent $\{e_i^*\}_{i=1}^{n}$, where the distribution of e_i^* is discrete with $\Pr\{e_i^* = \frac{1+\sqrt{5}}{2} e_i\} = \frac{5-\sqrt{5}}{10}$ and $\Pr\{e_i^* = \frac{1-\sqrt{5}}{2} e_i\} = \frac{5+\sqrt{5}}{10}$; ii) compute the bootstrap data $\{(X_i^*, Y_i^*)\}_{i=1}^{n}$, where $X_i^* = X_i$ and $Y_i^* = m(X_i, \widehat{\theta}) + e_i^*$ and, with these data, the corresponding bootstrap statistic T_n^*; iii) repeat the process B times to obtain $\{T_{n,j}^*\}_{j=1}^{B}$, and reject H_0 if $T_n > T_{n,(1-\alpha)}^*$, where $T_{n,(1-\alpha)}^*$ is the $(1-\alpha)$ quantile of $\{T_{n,j}^*\}_{j=1}^{B}$ and α is the significance level. This "wild bootstrap" procedure also allows us to obtain critical values when using the statistic proposed by Härdle and Mammen (1993), but in this case, in the first step e_i must be replaced by $e_{h,i} = Y_i - \widehat{m}_h(X_i)$. The validity of the wild bootstrap procedure in this context has been studied in Härdle and Mammen (1993), Li and Wang (1998), Stute et al. (1998) and Fan and Li (2000), among others.

3 SIZE AND POWER OF SPECIFICATION TESTS: SIMULATION RESULTS

We examine the behaviour of the specification tests discussed above with a moderate sample size by means of some Monte Carlo experiments. We first generate $n = 100$ independent observations as follows: $X_i \sim N(0, 1)$, $U_i \sim N(0, 1)$, X_i and U_i independent, and $Y_i = m(X_i) + U_i$, where

$$m(x) = x + c(x^2 - 1)$$

and the value of c varies; henceforth, we will refer to this experiment as Model 1. To analyse if the statistics are sensitive to the assumption of homoscedasticity, we also perform another experiment with the same characteristics as Model 1, but defining $Y_i = m(X_i) + (1 + X_i^2/2)^{1/2} U_i$; this experiment will be referred to as Model 2. In both models, the null hypothesis we test is that the regression funcion is linear, i.e. $E(Y \mid X = x) = \theta_{01} + \theta_{02} x$ for some $\theta_0 = (\theta_{01}, \theta_{02})$ in $\mathbb{R}^2$. Hence H_0 is true if and only if $c = 0$. To examine the performance of the tests when the number of regressors is greater than one, we also perform an experiment with $n = 100$ independent observations as follows: $X_i = (X_{1i}, X_{2i}) \sim$ bivariate normal distribution with mean $(0, 0)$, $var(X_{1i}) = var(X_{2i}) = 1$, $cov(X_{1i}, X_{2i}) = 0$, $U_i \sim N(0, 1)$, X_i and U_i independent, and $Y_i = m(X_{1i}, X_{2i}) + U_i$, where

$$m(x_1, x_2) = x_1 + x_2 + c(x_1^2 - 1)(x_2^2 - 1)$$

and the value of c varies; this experiment will be referred to as Model 3. As before, we also test the null hypothesis that the regression function is linear, i.e. $E(Y \mid X_1 = x_1, X_2 = x_2) = \theta_{01} + \theta_{02} x_1 + \theta_{03} x_2$ for some $\theta_0 = (\theta_{01}, \theta_{02}, \theta_{03})$ in $\mathbb{R}^3$, and again H_0 is true if and only if $c = 0$.

Parameter θ_0 is always estimated by least squares. In all univariate kernel estimations we use the quartic kernel $K(u) = \frac{15}{16}(1 - u^2)^2 I(|u| \leq 1)$, so $c_K = \frac{5}{7}$ and $r = 2$. In the multivariate estimations we use the product of quartic kernels. As regards the bandwidth, in the univariate estimations we use $h = \lambda S_X n^{-1/5}$, where S_X is the sample standard deviation of the regressor. In the multivariate estimations, each regressor is previously divided by its sample standard deviation and we then use $h = \lambda n^{-1/(d+4)}$. Suitable choices for λ were selected from plots obtained with some samples, but in all of the cases we report the results for various λ in order to see how this choice affects the performance of the statistics.

The statistic $T_n^{(G)}$ is computed in Models 1 and 2 using $L = 3$ points: -1, 0, 1; in Model 3, $L = 5$ points are considered: $(-1, -1)$, $(-1, 1)$, $(0, 0)$, $(1, -1)$, $(1, 1)$; in all cases, $T_n^{(G)}$ is computed with $\delta = 0.5$ and without assuming any parametric form for $Var(Y \mid X = x)$. The statistic $T_n^{(HM)}$ is computed with

7

$\pi(x) = I(x \in [-1.96, 1.96])$ in Models 1 and 2, and $\pi(x) = I(x \in [-1.8, 1.8] \times [-1.8, 1.8])$ in Model 3; in all cases, the integral in $T_n^{(HM)}$ is approximated numerically. The statistic $T_n^{(HH)}$ is computed with $\delta = 0.5$, $w(\cdot) = \pi(\cdot)$ and without assuming any parametric form for $Var[Y_i \mid m(X_i, \theta)]$. As suggested in Bierens (1990), $T_n^{(B)}$ is computed with $\Phi(\cdot) = (\Phi(\cdot)_1, ..., \Phi(\cdot)_d)$, where for $j = 1, ..., d$, we denote $\Phi(u)_j \equiv \arctan[(u - \overline{X}_j)/S_{Xj}]$, and $\overline{X}_j$ and S_{X_j} are the sample mean and standard deviation of $\{X_{ji}\}_{i=1}^n$; in this case, additional experiments suggest that the choice of $\omega(\cdot)$ and $F(.)$ does not play a crucial role in the results. Finally, whenever a bootstrap procedure is required we use $B = 500$ bootstrap replications. All the results we report are based on 2000 replications of the data-generating process, and have been carried out using GAUSS programmes, which are available from the authors on request.

We first discuss the results obtained when there is only one regressor. In Tables 1 and 2, we report the proportion of rejections of H_0 in Models 1 and 2, respectively, when nominal significance level is $\alpha = 0.05$.

TABLE 1: Proportion of Rejections of H_0 in Model 1 ($\alpha = 0.05$)

c	Gozalo $\lambda=3.0$	$\lambda=3.5$	$\lambda=4.0$	Härdle and Mammen $\lambda=3.0$	$\lambda=3.5$	$\lambda=4.0$	Zheng $\lambda=.20$	$\lambda=.40$	$\lambda=.60$
-0.5	0.909	0.930	0.935	0.991	0.991	0.993	0.522	0.707	0.797
-0.4	0.723	0.754	0.768	0.955	0.960	0.956	0.313	0.447	0.548
-0.3	0.455	0.485	0.525	0.799	0.812	0.809	0.176	0.261	0.312
-0.2	0.202	0.228	0.290	0.480	0.488	0.486	0.091	0.105	0.105
-0.1	0.082	0.093	0.144	0.178	0.171	0.159	0.053	0.045	0.044
0	0.052	0.059	0.099	0.064	0.057	0.050	0.034	0.037	0.027
0.1	0.094	0.099	0.149	0.173	0.164	0.165	0.052	0.050	0.048
0.2	0.213	0.231	0.286	0.475	0.483	0.481	0.094	0.113	0.127
0.3	0.464	0.494	0.536	0.814	0.818	0.821	0.190	0.256	0.303
0.4	0.746	0.782	0.797	0.937	0.946	0.947	0.319	0.451	0.543
0.5	0.902	0.918	0.924	0.991	0.993	0.996	0.516	0.697	0.795

c	Ellison and Ellison $\lambda=3.0$	$\lambda=3.5$	$\lambda=4.0$	Horowitz and Härdle $\lambda=3.0$	$\lambda=3.5$	$\lambda=4.0$	Bierens	Stute
-0.5	0.995	0.993	0.993	0.977	0.974	0.969	0.997	0.987
-0.4	0.972	0.969	0.964	0.919	0.918	0.910	0.988	0.934
-0.3	0.835	0.827	0.816	0.726	0.726	0.714	0.902	0.763
-0.2	0.494	0.472	0.458	0.379	0.370	0.350	0.624	0.444
-0.1	0.164	0.132	0.109	0.118	0.089	0.076	0.216	0.150
0	0.082	0.052	0.031	0.077	0.048	0.027	0.048	0.050
0.1	0.175	0.144	0.119	0.134	0.110	0.076	0.215	0.160
0.2	0.502	0.474	0.446	0.372	0.354	0.340	0.626	0.441
0.3	0.838	0.830	0.810	0.740	0.735	0.706	0.901	0.777
0.4	0.976	0.972	0.968	0.927	0.925	0.919	0.988	0.941
0.5	0.994	0.994	0.993	0.972	0.967	0.966	0.999	0.988

TABLE 2: Proportion of Rejections of H_0 in Model 2 ($\alpha = 0.05$)

	Gozalo			Härdle and Mammen			Zheng		
c	λ=3.0	λ=3.5	λ=4.0	λ=3.0	λ=3.5	λ=4.0	λ=.20	λ=.40	λ=.60
-0.5	0.862	0.873	0.874	0.939	0.941	0.942	0.402	0.566	0.652
-0.4	0.671	0.689	0.703	0.821	0.821	0.816	0.242	0.350	0.418
-0.3	0.450	0.457	0.478	0.608	0.611	0.606	0.146	0.217	0.263
-0.2	0.241	0.250	0.283	0.356	0.349	0.340	0.084	0.088	0.094
-0.1	0.117	0.127	0.170	0.175	0.165	0.153	0.060	0.049	0.053
0	0.090	0.097	0.124	0.091	0.086	0.079	0.041	0.039	0.034
0.1	0.133	0.142	0.173	0.166	0.165	0.148	0.060	0.059	0.053
0.2	0.247	0.254	0.294	0.351	0.347	0.330	0.090	0.107	0.120
0.3	0.440	0.468	0.492	0.620	0.629	0.618	0.161	0.212	0.248
0.4	0.704	0.726	0.724	0.807	0.808	0.804	0.253	0.348	0.419
0.5	0.856	0.863	0.860	0.930	0.939	0.936	0.393	0.557	0.646

	Ellison and Ellison			Horowitz and Härdle			Bierens	Stute
c	λ=3.0	λ=3.5	λ=4.0	λ=3.0	λ=3.5	λ=4.0		
-0.5	0.937	0.932	0.932	0.884	0.874	0.861	0.950	0.931
-0.4	0.826	0.821	0.810	0.737	0.733	0.716	0.848	0.805
-0.3	0.611	0.604	0.589	0.505	0.484	0.463	0.622	0.576
-0.2	0.343	0.324	0.300	0.269	0.246	0.222	0.317	0.291
-0.1	0.168	0.136	0.115	0.137	0.103	0.079	0.107	0.121
0	0.122	0.091	0.066	0.102	0.069	0.044	0.041	0.055
0.1	0.184	0.148	0.123	0.150	0.116	0.095	0.115	0.120
0.2	0.340	0.319	0.292	0.260	0.227	0.209	0.314	0.296
0.3	0.607	0.586	0.569	0.507	0.483	0.465	0.613	0.567
0.4	0.832	0.828	0.821	0.758	0.744	0.726	0.856	0.816
0.5	0.927	0.923	0.922	0.866	0.865	0.854	0.946	0.920

The first thing we observe in Table 1 is that the empirical size of the tests is correct in most cases: only $T_n^{(Z)}$ has an empirical size that is slightly below its nominal size. Regarding the smoothing tests, the first conclusion to be drawn is that bandwidth does not seem to play a crucial role, as long as it is selected within a reasonable range. This is probably not so surprising considering that a particular version of the Bierens test corresponds to a kernel-regression based test with a fixed bandwidth, as Fan and Li (2000) have shown. In Model 1, assuming that the bandwidth has been correctly chosen, we observe that $T_n^{(HM)}$ and $T_n^{(EE)}$ are the smoothing tests that perform better and that they both perform quite similarly. It is also possible to rank the other smoothing statistics (from better to worse) as follows: $T_n^{(HH)}$, $T_n^{(G)}$, $T_n^{(Z)}$. Regarding the non-smoothing tests, according to their power functions, $T_n^{(B)}$ performs better than $T_n^{(S)}$; in fact, $T_n^{(B)}$ performs even better than any other smoothing test, whereas $T_n^{(S)}$ performs similarly to $T_n^{(HM)}$ and $T_n^{(EE)}$. In the results of Model 2 we observe that heteroskedasticity causes more sensitivity to bandwidth choice in smoothing tests, some distortions in the size of the statistics (especially $T_n^{(G)}$) and, in all of the cases, a decrease in power. The statistics, however, continue to perform well, in the sense that any departure from the null is detected and the empirical size is approximately correct. The conclusions drawn about the comparative performance of the statistics from the results of Model 1 still hold.

In Table 3, we report the proportion of rejections of H_0 in Model 3, which has two regressors, when the nominal significance level is $\alpha = 0.05$.

TABLE 3: Proportion of Rejections of H_0 in Model 3 ($\alpha = 0.05$)

c	Gozalo $\lambda=2.5$	$\lambda=3.0$	$\lambda=3.5$	Härdle and Mammen $\lambda=3.0$	$\lambda=3.5$	$\lambda=4.0$	Zheng $\lambda=.25$	$\lambda=.50$	$\lambda=.75$
-0.5	0.661	0.692	0.709	0.786	0.644	0.476	0.095	0.221	0.340
-0.4	0.553	0.559	0.600	0.616	0.475	0.363	0.064	0.120	0.186
-0.3	0.389	0.384	0.444	0.412	0.286	0.197	0.057	0.093	0.116
-0.2	0.288	0.295	0.359	0.251	0.182	0.134	0.046	0.058	0.069
-0.1	0.221	0.204	0.274	0.131	0.091	0.058	0.035	0.039	0.035
0	0.180	0.176	0.245	0.098	0.059	0.034	0.038	0.049	0.043
0.1	0.213	0.207	0.273	0.116	0.088	0.051	0.043	0.049	0.044
0.2	0.277	0.273	0.334	0.247	0.194	0.133	0.042	0.060	0.066
0.3	0.389	0.401	0.468	0.417	0.301	0.218	0.064	0.091	0.116
0.4	0.539	0.555	0.587	0.608	0.460	0.354	0.076	0.146	0.193
0.5	0.663	0.692	0.708	0.790	0.640	0.478	0.095	0.206	0.334

c	Ellison and Ellison $\lambda=3.5$	$\lambda=4.0$	$\lambda=4.5$	Horowitz and Härdle $\lambda=4.0$	$\lambda=4.5$	$\lambda=5.0$	Bierens	Stute
-0.5	0.759	0.705	0.617	0.455	0.426	0.389	0.094	0.158
-0.4	0.658	0.591	0.503	0.371	0.341	0.310	0.080	0.113
-0.3	0.472	0.418	0.354	0.260	0.230	0.201	0.061	0.085
-0.2	0.268	0.217	0.180	0.188	0.156	0.130	0.042	0.080
-0.1	0.128	0.094	0.067	0.111	0.081	0.064	0.041	0.051
0	0.079	0.044	0.033	0.083	0.053	0.038	0.043	0.043
0.1	0.116	0.081	0.057	0.114	0.089	0.063	0.034	0.066
0.2	0.278	0.237	0.196	0.177	0.147	0.127	0.050	0.070
0.3	0.468	0.391	0.325	0.258	0.225	0.201	0.058	0.100
0.4	0.652	0.592	0.518	0.396	0.374	0.333	0.072	0.118
0.5	0.766	0.709	0.625	0.467	0.439	0.406	0.097	0.149

12

In this table, we first observe that the performance of all the tests is worse here than when there is only one regressor. The empirical size of the test is correct in all cases, except for Gozalo's statistic. In this case, we carried out additional experiments to verify that the sample size required to obtain the correct size is much greater. With regard to the other four smoothing statistics, the one that seems less sensitive to bandwidth choice is Zheng's statistic, which, however, is the one with the worst power properties. In fact, among smoothing statistics with the correct size, the ranking that is obtained according to the power functions is similar as in Model 1, although $T_n^{(EE)}$ now performs somewhat better than $T_n^{(HM)}$. The performance of the non-smoothing tests in Model 3 is dramatically worse than in Models 1 and 2; they both have approximately correct size, but their power is clearly below that of the smoothing-tests. With additional experiments, we verified that the value of c required to obtain an empirical power above 0.5 with $T_n^{(B)}$ and $T_n^{(S)}$ in Model 3 is approximately ± 1.

To sum up, in the models with one regressor, the statistic $T_n^{(B)}$ proposed by Bierens (1990) is preferable, although the statistics $T_n^{(S)}$ proposed by Stute (1997), $T_n^{(EE)}$ proposed by Ellison and Ellison (2000) and $T_n^{(HM)}$ proposed by Härdle and Mammen (1993) have similar properties; $T_n^{(B)}$ and $T_n^{(S)}$ have the advantage of not requiring the use of any smoothing value, although a bootstrap procedure is required to compute critical values. In the model with two regressors, however, some of the smoothing statistics perform better than the non-smoothing ones; specifically, $T_n^{(EE)}$ is the one that yields the best results, and it has the additional advantage of being able to be implemented with critical values from the standard normal distribution.

4 TESTING MINCER-TYPE WAGE EQUATIONS

In this section, we apply the statistics we analysed in the previous sections to test three Mincer-type wage equation specifications, using Uruguayan and Spanish data. The first specification we test is the traditional Mincer-type model, in which the logarithm of real hourly wages w_i is regressed on a constant, years of schooling ed_i, years of potential work experience ex_i and its square:

$$w_i = \beta_1 + \beta_2 ed_i + \beta_3 ex_i + \beta_4 ex_i^2 + u_i.$$

This model is used in many papers on applied labour, including Buchinsky (1994) or Di Nardo et al. (1996). The second specification adds cubic and quartic terms on years of potential work experience:

$$w_i = \beta_1 + \beta_2 ed_i + \beta_3 ex_i + \beta_4 ex_i^2 + \beta_5 ex_i^3 + \beta_6 ex_i^4 + u_i.$$

Murphy and Welch (1992) indicate that this specification provides a better fit than the traditional model, which only includes ex_i and its square. Finally, we examine a third specification which is quite similar to the first one but which incorporates an interaction between education and potential work experience:

$$w_i = \beta_1 + \beta_2 ed_i + \beta_3 ex_i + \beta_4 ex_i^2 + \beta_5 ed_i ex_i + u_i.$$

The data on Uruguay was obtained from the Uruguayan Household Survey, from 1986 through 1997 (Encuesta de Hogares, Instituto Nacional de Estadística, Uruguay). The frame of the survey is the Urugayan civilian population that lives in housing units, and is sub-divided into two separate surveys: one for the metropolitan area of Montevideo and another for the rest of the country. We use the first of these surveys exclusively, since almost half of the Uruguayan population lives in Montevideo and two-thirds of its economic activity takes place there. The sample is composed of males older than 13, which this is the legal working age in Uruguay. We only included individuals who had a positive salary in the month preceding the interview, and who had worked during the week before the interview, in either the private or the public sector. However, self-employed individuals, workers without salaries, entrepreneurs and those who had never worked were excluded (see Bucheli et al., 2000, for a detailed description of the survey). The Spanish data was obtained from the Encuesta de Estructura Salarial (1995). This survey gives detailed information on individuals, their earnings, their characteristics (age, education, etc) as well as their employers' characteristics (firm size, economic activity, type of wage-bargaining, etc). For a description of the sample, see Cantó et al. (2001). Here, we use the subsample of full-time male workers.

The variable of interest is the real hourly wage. For the Uruguayan data,

it was defined as the salary in the month before the interview, divided by four times the hours worked during the week before the interview. Thus, we assume that the number of hours worked for each individual week during the previous month is equal to the number of hours worked during the week before the interview. For the Spanish data, this was defined as the monthly salary divided by total hours worked in the month (see Di Nardo et al. 1996 for a justification of the use of real hourly wages).

Before testing the wage specifications, we perform another Monte Carlo experiment to see how the test statistics behave with data that is similar to what will be used in the application. We generate $n = 500$ independent observations, as follows: $X_i = (X_{1i}, X_{2i}) \sim$ bivariate normal distribution with mean $(8.44, 25.52)$, $var(X_{1i}) = 14.44$, $var(X_{2i}) = 149.8$, $cov(X_{1i}, X_{2i}) = 11.17$, $U_i \sim N(0, 0.197)$, X_i and U_i independent, and $Y_i = m(X_{1i}, X_{2i}) + U_i$, where

$$m(x_1, x_2) = 5.66 + 0.814x_1 + 0.043x_2 - 0.000465x_2^2 + c(x_1^2 - 1)(x_2^2 - 1)$$

and the value of c varies. This experiment will be referred to as Model 4. All of the parameters in this model have been obtained from the estimation of the first specification for the Mincer-type model with Spanish data. We test the null hypothesis $E(Y \mid X_1 = x_1, X_2 = x_2) = \theta_{01} + \theta_{02}x_1 + \theta_{03}x_2 + \theta_{04}x_2^2$ for some $\theta_0 = (\theta_{01}, \theta_{02}, \theta_{03}, \theta_{04})$ in $\mathbb{R}^4$. All estimations are performed with the same characteristics as in Model 3, except that in the Gozalo test the $L = 5$ points we consider are $(8.44, 25.52)$ and $(8.44 \pm 14.14^{1/2}, 25.52 \pm 149.8^{1/2})$. In Table 4, we report the proportion of rejections of H_0 in this model when the nominal significance level is $\alpha = 0.05$. These results are based on 2000 replications of the data. When a bootstrap procedure is required $B = 200$ bootstrap replications are used. In this table we observe that all of the statistics have approximately the correct size, except for $T_n^{(G)}$, which will therefore be excluded from the application. All of the statistics detect departures from the null, although the power function obtained differs greatly among statistics, what might have been expected considering the results obtained for Model 3.

TABLE 4: Proportion of Rejections of H_0 in Model 4 ($\alpha = 0.05$)

c	Gozalo			Härdle and Mammen			Zheng		
	$\lambda=1.0$	$\lambda=1.5$	$\lambda=2.0$	$\lambda=1.6$	$\lambda=1.8$	$\lambda=2.0$	$\lambda=.20$	$\lambda=.40$	$\lambda=.60$
$-.005$	0.752	0.831	0.913	0.886	0.945	0.989	0.131	0.312	0.486
$-.004$	0.691	0.751	0.822	0.852	0.916	0.957	0.103	0.188	0.289
$-.003$	0.615	0.649	0.738	0.760	0.803	0.806	0.067	0.096	0.130
$-.002$	0.569	0.567	0.641	0.523	0.470	0.411	0.056	0.054	0.067
$-.001$	0.547	0.508	0.598	0.258	0.154	0.096	0.054	0.052	0.053
0	0.545	0.520	0.604	0.143	0.068	0.025	0.042	0.046	0.036
$.001$	0.548	0.556	0.641	0.222	0.140	0.081	0.049	0.051	0.051
$.002$	0.610	0.608	0.735	0.513	0.471	0.409	0.064	0.068	0.076
$.003$	0.654	0.702	0.797	0.761	0.807	0.799	0.064	0.103	0.143
$.004$	0.717	0.778	0.852	0.833	0.921	0.967	0.103	0.177	0.284
$.005$	0.775	0.850	0.923	0.851	0.941	0.988	0.134	0.291	0.510

c	Ellison and Ellison			Horowitz and Härdle			Bierens	Stute
	$\lambda=4.0$	$\lambda=4.5$	$\lambda=5.0$	$\lambda=20$	$\lambda=25$	$\lambda=30$		
$-.005$	1	1	1	0.188	0.129	0.084	0.731	1
$-.004$	0.999	1	1	0.181	0.110	0.067	0.610	1
$-.003$	0.983	0.985	0.986	0.167	0.102	0.061	0.382	0.983
$-.002$	0.743	0.757	0.766	0.129	0.079	0.047	0.206	0.754
$-.001$	0.239	0.220	0.224	0.093	0.058	0.032	0.078	0.214
0	0.088	0.058	0.047	0.069	0.047	0.029	0.029	0.033
$.001$	0.231	0.216	0.217	0.100	0.059	0.032	0.063	0.227
$.002$	0.729	0.749	0.764	0.180	0.102	0.049	0.154	0.798
$.003$	0.983	0.986	0.990	0.328	0.193	0.098	0.298	0.990
$.004$	1	1	1	0.547	0.358	0.195	0.416	1
$.005$	1	1	1	0.732	0.517	0.321	0.765	1

Finally, in Table 5 we report the p-value we obtain when we test each of the three wage specifications. We use Uruguayan data for 1986, 1989, 1993 and 1997, and Spanish data for 1995. The number of Uruguayan observations in each year is $n = 4509$ for 1986, $n = 5064$ for 1989, $n = 4920$ for 1993 and $n = 4847$ for 1997. For the estimation with the Spanish data, we randomly selected a subsample of $n = 5982$ observations (5% of the total sample). The estimations are performed with the same characteristics as those in Table 4, although, in this case, we use $B = 1000$ bootstrap replications whenever required.

TABLE 5: Testing Mincer-Type Wage Equations: P-Values

	$T_n^{(HM)}$		$T_n^{(Z)}$		$T_n^{(EE)}$		$T_n^{(HH)}$		$T_n^{(B)}$	$T_n^{(S)}$
Specif. 1	h	P-v	h	P-v	h	P-v	h	P-v	P-v.	P-v.
1986-URU	.648	.000	.615	.000	.738	.000	.558	.011	.023	.008
1989-URU	.639	.000	.603	.000	.724	.000	.545	.000	.005	.010
1993-URU	.643	.000	.606	.000	.727	.000	1.10	.034	.004	.006
1997-URU	.686	.000	.608	.000	1.46	.000	.550	.000	.004	.000
1995-SPA	.512	.000	.473	.000	.709	.000	.531	.000	.000	.000
Specif. 2	h	P-v	h	P-v	h	P-v	h	P-v	P-v.	P-v.
1986-URU	.648	.000	.615	.000	.738	.000	.558	.001	.021	.003
1989-URU	.639	.000	.603	.000	.724	.000	.545	.010	.011	.000
1993-URU	.643	.000	.606	.000	.727	.001	1.10	.097	.008	.000
1997-URU	.686	.000	.608	.001	1.46	.000	.550	.001	.002	.000
1995-SPA	.512	.000	.473	.000	.709	.000	.531	.000	.009	.000
Specif. 3	h	P-v	h	P-v	h	P-v	h	P-v	P-v.	P-v.
1986-URU	.648	.000	.615	.000	.738	.000	.558	.031	.020	.018
1989-URU	.639	.000	.603	.000	.724	.000	.545	.000	.016	.016
1993-URU	.643	.000	.606	.000	.727	.000	1.10	.001	.012	.006
1997-URU	.686	.000	.608	.000	1.46	.000	.550	.000	.004	.006
1995-SPA	.512	.000	.473	.000	.709	.000	.531	.000	.003	.002

We observe in Table 5 that almost all of the specifications are rejected for all of the data sets with the usual significance levels. Note that this result is relatively robust, as we verified that the statistics have the correct size in this context even with much smaller sample sizes. Additionally, the heteroskedasticity problem that could arise from the heterogeneity of the sample would cause a loss of power in the test, as seen in the previous section, and therefore, does not affect our conclusions. These results are of particular interest for applied work, as they imply that any conclusion drawn about the earning profiles based on the estimation of Mincer-type wage equations, like the ones analyzed here, could be misleading. In the words of Manski (2000), "empirical findings are only as credible as the identifying assumptions imposed".

5   CONCLUDING REMARKS

The main objective of this study was to present, in a unified way, the various non-parametric specification tests for regression models that have recently appeared in the literature, and to compare their performances in a common framework. Our results show that when there is only one regressor, the non-smoothing tests perform slightly better than the smoothing ones, especially the implementation we consider for the Bierens statistic. Moreover, they have the obvious advantage that no bandwidth selection is required, though their implementation requires the use of a bootstrap procedure. When the number of regressors is greater than one, some of the smoothing tests we consider here perform better. Specifically, the statistic proposed in Ellison and Ellison (2000) exhibits good properties in all of the models we have simulated, and it has the additional advantage of being able to be implemented with critical values from the standard normal distribution.

The second objective of this study was to emphasize the importance of testing parametric assumptions in applied work. Specifically, we have shown that the well-known Mincer-type wage equation is clearly rejected with Uruguayan and Spanish data, even considering several generalizations that have appeared in the literature. This negative result casts doubts on the conclusions that have been derived from this type of wage equation in recent applied work.

References

[1] Bierens, H., 1990. A consistent conditional moment test of functional form. Econometrica 58, 1443-1458.

[2] Bierens, H., Ploberger, W., 1997. Asymptotic theory of integrated conditional moment tests. Econometrica 65, 1129-1152.

[3] Bucheli, M., Miles, D., Vigorito, A., 2000. Un análisis dinámico de la toma de decisiones de los hogares en América Latina: el caso Uruguayo. Revista de Economía del Banco Central del Uruguay 7, 5-56.

[4] Buchinsky, M., 1994. Changes in the U.S. wage structure 1963-1987: application of quantile regression. Econometrica 62, 405-458.

[5] Cantó, O., Cardoso, A.R., Jimeno, J.F., 2001. Earnings inequality in Spain and Portugal: contrasts and similarities. In: Cohen, D., Piketty, T., Saint-Paul, G. (Ed.), The New Economics of Rising Inequalities, Oxford University Press.

[6] Di Nardo, J., Fortin, N., Lemieux, T., 1996. Labor market institutions and the distribution of wages: a semiparametric approach. Econometrica 64, 1001-1044.

[7] Ellison, G., Ellison, S.F., 2000. A simple framework for non-parametric specification testing. Journal of Econometrics 96, 1-23.

[8] Fan Y., Li, Q., 2000. Consistent model specification tests: kernel-based tests versus Bierens' ICM Tests. Econometric Theory 16, 1016-1041.

[9] Gozalo, P., 1993. A consistent model specification test for non-parametric estimation of regression function models. Econometric Theory 9, 451-477.

[10] Härdle, W., Mammen, E., 1993. Comparing non-parametric versus parametric regression fits. Annals of Statistics 21, 1926-1947.

[11] Horowitz, J., Härdle, W., 1994. Testing a parametric model against a semiparametric alternative. Econometric Theory 10, 821-848.

[12] Li, Q., Wang, S., 1998. A simple consistent bootstrap test for a parametric regression function. Journal of Econometrics 87, 145-165.

[13] Manski, C., 2000. Economic analysis of social interactions. Journal of Economic Perspectives 14, 115-36.

[14] Murphy, W., Welch, G., 1992. The structure of wages. Quarterly Journal of Economics 107, 285-326.

[15] Stute, W., 1997. Nonparametric model checks for regression. Annals of Statistics 25, 613-641.

[16] Stute, W., González-Manteiga, W., Presedo-Quindimil, M., 1998. Bootstrap approximations in model checks for regression. Journal of the American Statistical Association 93, 141-149.

[17] Willis, R., 1986. Wage determinants: a survey and reinterpretation of human capital earnings functions. In: Ashenfelter, O., Layard, R. (Ed.), Handbook of Labor Economics, vol. 1, North Holland, Amsterdam, 525-602.

[18] Zheng, J., 1996. A consistent test of functional form via non-parametric estimation techniques. Journal of Econometrics 75, 263-289.