An equality test across nonparametric regressions[*]

Pascal Lavergne[†]

INRA-ESR, Toulouse

March 1997, revised May 2000

Abstract

We propose a procedure for testing the equality of several nonparametric multivariate regressions. We allow the regressors' designs and the number of observations to differ across subsamples. The division into subsamples is defined through a variable C which can be either fixed or random. For a random C, our procedure is a general test of significance for qualitative variables in a nonparametric regression. For a fixed C, our procedure provides a "nonparametric analysis of covariance," which is valid for cross-section or panel data. In both cases, the test is a one-sided normal test and is consistent against all alternatives.

Keywords: Hypothesis testing, Nonparametric methods, Qualitative variables, Covariance analysis.

JEL classification: Primary C52; Secondary C14.

[†]Corresponding address: INRA-ESR, BP 27, 31326 CASTANET-TOLOSAN Cedex FRANCE. E-mail address: lavergne@toulouse.inra.fr

1 Introduction

A classic problem in econometrics is determining whether the form of a regression function remains the same for two or more separate subsamples. Beginning with Chow's (1960) work, a lot of attention has been devoted in the econometric literature to testing equality of regression functions. The related procedures have been used for various economic problems, such as testing for gender or race discrimination in earnings functions, testing for stability over time of economic relationships, testing of disequilibrium models or testing for switching firms' strategies in microeconometric models derived from game theory. Classical procedures assume a parametric form, usually a linear one, for the regression functions under test. As is well-known, specifying an incorrect parametric form can lead to serious errors in inference. Indeed, rejection of the equality hypothesis can be solely due to misspecification of the model. Reversely, overacceptance of the null hypothesis can appear as a consequence of misspecification. Therefore, it is advisable to use a testing procedure free of any parametric assumption.

The issue of comparing regression curves in a nonparametric context has been first addressed in the particular setup of two subsamples with a one-dimensional regressor.[1] When the regressor's designs are identical, it is possible to use the differences in the dependent variable between the two subsamples to build a test statistic. Hall and Hart (1990) propose a Cramer-von-Mises type statistic while Delgado (1993) studies a Kolmogorov-Smirnov type statistic. Alternatively, one can directly use the mean squared difference between nonparametric regression estimates. This idea has been worked out in the fixed design case by King, Hart and Wehrly (1991), who propose a test under the assumption of normal residuals, and by Härdle and Marron (1990), who consider testing whether two curves are equal up to a parametric transformation. Recent work focuses on specific extensions of the initial setup. A first extension deals with more than two subsamples. Young and Bowman's (1995) test compares several nonparametric regressions depending on a one-dimensional random variable with normal residuals. Fan and Lin (1998) propose testing procedures based on the adaptive Neyman and the wavelet thresholding approaches for comparing multiple curves indexed by time. A second extension deals with non-identical designs. Kulasekera (1995) extends Hall and Hart's procedure to this case. Pinkse and Robinson (1995) generalizes some of Härdle and Marron's (1990) results to random designs and give some guidelines for testing whether two regression functions are similar up to a linear transforma-

---

[1]Testing equality of nonparametric density functions is a closely related problem, see e.g. Mammen (1992).

tion. Hall, Huber and Speckman (1997) build covariate-matching based tests that are taylored against the specific alternative that one of the regression function is always greater than the other. Koul and Schick (1997) propose analogous procedures that incorporate smoothing techniques. Matching is actually used in an earlier illuminating paper by Quade (1982), who deals simultaneously with non-identical designs and multiple subsamples. Other authors considering a general setup are Baltagi, Hidalgo and Li (1996) on testing for poolability of panel data, Munk and Dette (1998) and Yatchew (1999), whose test is based on the comparison of residual variance estimators.

In applied econometrics, we often consider more than one explanatory variable and deal simultaneously with more than two subsamples. More crucially, it is scarcely the case that we have control on the design of explanatory variables. Furthermore, the division into subsamples is also frequently governed by a random variable. In view of practical use in econometrics, we propose a general asymptotic joint test of equality across nonparametric regressions that is consistent against any alternative. It extends previous work in many directions. First, we allow for different unknown data distributions across subsamples. Second, we also allow for any dimension of the explanatory variables. Third, we deal with any number of subsamples. Fourth, we consider random explanatory variables and allow the regressors' designs and the number of observations to differ across subsamples. Fifth, in contrast to previous work that exclusively focuses on the case where the division into subsamples is defined through a fixed variable C, we consider as a leading case the situation where a random qualitative variable C defines the split into different subsamples, as frequently arises in economic applications. We subsequently extend our procedure to the usual setup.

When C is fixed, our procedure provides a "nonparametric analysis of covariance" that has numerous potential applications in and outside the field of econometrics. When C is random, our procedure is a general significance test for qualitative variables in a nonparametric regression. This supplements previous work on testing for omitted continuous variables in nonparametric regression, see Aït-Sahalia, Bickel and Stoker (1994), Fan and Li (1996), Gozalo (1995), Lavergne and Vuong (2000) and Li (1999).[2] However, our analysis has a distinctive feature with respect to previous work on smooth tests for omitted variables. Indeed, our null and alternative models

---

[2]More generally, our test is in line with smooth specification tests of parametric models previously proposed in the literature, see Delgado, Dominguez and Lavergne (1998), Härdle and Mammen (1993), Zheng (1996) and some of the previous references.

have the same effective dimension due to the discrete nature of C, whereas previous work on nonparametric significance testing compare models with different dimensions.[3] Hence in our setup, nonparametric estimators of the null and the alternative models have similar rates of convergence and both affect the asymptotic distribution of the test statistic, in contrast to what happens when models have different dimensions.

The paper is organised as follows. In Section 2, we consider the leading case where the splitting variable is random. We set up our testing framework and we derive the basic statistic for testing equality of nonparametric regression functions. We characterize its asymptotic distribution not only under the null hypothesis but also under a sequence of local alternatives. We then derive a consistent testing procedure and discuss its implementation. In Section 3, we treat the case of a fixed splitting variable and relate it to a nonparametric analysis of covariance. We show how the assumptions of Section 2 can be weakened to deal with cross-section and panel data. Section 4 studies the small sample behavior of our test through some simulation experiments. The Conclusion summarizes our main findings. All the proofs are relegated to Appendices.

2   Case of a random C

2.1   The testing framework

Let C be a discrete variable on $\boldsymbol{C} = \{1, \ldots, \bar{C}\}$, with corresponding strictly positive probabilities $p_1, \ldots, p_{\bar{C}}$. Let $\{(C_i, X_i, Y_i),\ i = 1, \ldots, n\}$ be a sample of i.i.d. observations from (C, X, Y) taking values on $\boldsymbol{C} \times I\!\!R^p \times I\!\!R$. Consider the general regression model

$$Y_i = R(X_i, C_i) + U_i, \qquad E\left[U_i | X_i, C_i\right] = 0 \qquad i = 1, \ldots, n, \tag{2.1}$$

where $R(\cdot, \cdot)$ denotes the regression function of Y on X and C. For any c, a nonparametric kernel estimator of $f_c(\cdot)$, the conditional density of X given $C = c$, is

$$f_{n,c}(x) = (n_c h_c^p)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h_c}\right) I\!\![C_i = c], \qquad \forall x \in I\!\!R^p,$$

---

[3]This is also true for smooth specification tests of parametric regression models.

where $n_c = \sum_{i=1}^{n} 1\!\!1 [C_i = c]$, $K(\cdot)$ is a kernel on $I\!\!R^p$ and h_c is a bandwidth parameter converging to zero.[4] A nonparametric kernel estimator of $R(\cdot, c)$ is obtained as

$$R_n(x, c) = \frac{(n_c h_c^p)^{-1} \sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h_c}\right) 1\!\!1[C_i = c]}{f_{n,c}(x)} \qquad \forall x \in I\!\!R^p.$$

In the formulas, we use non-smoothing weights for the qualitative variable C. If there exists a natural ranking of the modalities of C relevant for the regression model, non-smoothing weights can be replaced by smooth ones without affecting the estimators' asymptotic properties, see Bierens (1987) and Delgado and Mora (1995), or our test's validity.

If we overlook the information concerning the splitting as given by the C_i's, we would consider instead the pooling regression model

$$Y_i = r(X_i) + u_i, \qquad E[u_i | X_i] = 0, \qquad i = 1, \ldots, n. \tag{2.2}$$

Thus we will estimate the function $r(\cdot)$ by a kernel estimate on the whole sample given by

$$r_n(x) = \frac{(n h^p)^{-1} \sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h}\right)}{f_n(x)}, \quad \text{where } f_n(x) = (n h^p)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \qquad \forall x \in I\!\!R^p,$$

and h is a bandwidth parameter converging to zero. These estimators respectively converge to $r(\cdot)$ and $f(\cdot)$, the marginal density of X.

The hypothesis of interest is the constancy of the regression function $R(\cdot, C = c)$ for different values of c. This means that we are not loosing any information by disregarding the C_i's and estimating the simpler regression Model (2.2) instead of (2.1). Thus the null hypothesis reads

$$H_0 : R(X, C) = r(X) \text{ a.s.}$$

This intuitive formulation enables us to deal with the testing problem as a comparison of two nested models, whatever the number of subsamples is. As H_0 corresponds to the non-significance of the discrete variable C, we can built our test statistic in a way similar to Lavergne and Vuong (2000), who test for omitted continuous variables. Let $u \equiv Y - r(X)$. We consider an estimate of $E\left[E^2(u | X, C) \Psi(X, C)\right] = E\left[(R(X, C) - r(X))^2 \Psi(X, C)\right]$, which is zero under H_0 and strictly positive under any alternative to H_0, for any function $\Psi(\cdot, \cdot)$ such that $\Psi(\cdot, c)$ is strictly positive on the support of $f_c(\cdot)$ for any c. Because of the form of the kernel estimate, it is convenient

---

[4]For exposition's ease, we do not make explicit the dependence of the bandwidths on the (sub)sample size.

4

to use $f^2(X)f_C(X)$ as a weighting function.[5] This device is analogous to the one used in other semiparametric estimation and testing problems, see e.g. Powell, Stock and Stoker (1989), Fan and Li (1996), Lavergne and Vuong (2000) and Zheng (1996). If the quantities $u_i f(X_i)$ were observed, a sample analog of $E\left[E^2(uf(X)|X,C)f_C(X)\right] = E\left[uf(X)E(uf(X)|X,C)f_C(X)\right]$ would be

$$V_{0n} = \frac{1}{n^{(2)}} \sum_a u_i u_j f(X_i) f(X_j) \frac{1}{h_{C_i}^p} \tilde{K}_{nij} w_{nij}, \text{ where } \tilde{K}_{nij} \equiv (1/h_{C_i}^p) K\left[\frac{X_i - X_j}{h_{C_i}}\right],$$

$w_{nij} = \frac{n-1}{n_{C_i}-1} I\!\!I[C_i = C_j]$, $\sum_a$ denotes summation over the arrangements of m distinct elements $\{i_1, \ldots, i_m\}$ from $\{1, \ldots, n\}$, and $n^{(m)} = n!/(n-m)!$ is the number of these arrangements. Now, because we do not know the u_i's and $f(X_i)$'s, we replace them by their kernel estimates, and we drop equal indices in the final sum.[6] We then obtain the statistic

$$V_n = \frac{1}{n^{(4)}} \sum_a (Y_i - Y_k)(Y_j - Y_l) K_{nik} K_{njl} \tilde{K}_{nij} w_{nij}, \text{ where } K_{nij} \equiv (1/h^p) K\left[\frac{X_i - X_j}{h}\right]. \quad (2.3)$$

2.2 Asymptotic behavior of V_n

Theorem 1 gives the behavior of V_n under the set of hypotheses

$$H_{1n}: \qquad R(X,C) = r(X) + \delta_n d(X,C), \ \delta_n \in [0,1]$$

A fixed alternative corresponds to $\delta_n = 1 \ \forall n$, while the null hypothesis corresponds to $\delta_n = 0 \ \forall n$.[7] This general formulation allows to deal with some local alternatives whose rates of convergence to H_0 are given by the rate of decrease of δ_n to zero.

We call $\mathcal{U}^p$ the class of integrable uniformly continuous functions from $I\!\!R^p$ to $I\!\!R$, and $\mathcal{D}_{m,q}^p$ the class of m-times differentiable functions from $I\!\!R^p$ to $I\!\!R$ with derivatives of order m that are uniformly Lipschitz continuous of order q, $q \in [0,1)$. For $m \geq 2$, we define $\mathcal{K}_m^p$ as the class of integrable functions $K(\cdot)$ from $I\!\!R^p$ to $I\!\!R$ with compact support, satisfying $\int K(s)\, ds = 1$ and

$$\int s_1^{\alpha_1} \ldots s_p^{\alpha_p} K(s)\, ds = 0 \quad \text{for } \alpha_i \in \{0, 1, \ldots, m-1\}, \ \forall i = 1, \ldots, p, \quad 0 < \sum_{i=1}^{p} \alpha_i \leq m - 1.[8]$$

---

[5]We could also choose $f^2(X)f_C(X)v(X)$, $v(\cdot)$ being a known function, used for instance to restrict attention to a particular interval. Our results can be easily extended to this formulation.

[6]For significance testing of continuous variables, Lavergne and Vuong (2000) show that dropping equal indices does not change the asymptotic distribution of their statistic but reduces its small-sample bias. In our case, dropping equal indices is essential to get an asymptotically unbiased statistic under H_0.

[7]We let $d(X,C) \equiv 0$ if $\delta_n = 0$.

[8]It suffices that $\int K(s)\, ds > 0$ to obtain a consistent test, but the unity integral assumption is not restrictive.

We let $\sigma_C^2(X) \equiv E\left[u^2 | X, C\right] = E\left[(Y - r(X))^2 | X, C\right]$ and we label it the "conditional variance" (with respect to both X and C) from Model (2.2). We also let $w_{CC'} \equiv \frac{1}{p_C} 1\!\!1\left[C = C'\right]$ and $K_a(\cdot) \equiv (1/a)K\left(\cdot/a\right)$. We define $*$ as the convolution operator, i.e. for kernels $K(\cdot)$ and $L(\cdot)$,

$$(K * L)(s) = \int_{I\!\!R^p} K(v)L(s - v)\, dv.$$

Assumption 2.1 $\{(C_i, X_i, Y_i),\ i = 1, \ldots, n\}$ *is an i.i.d. sample from a random variable* (C, X, Y) *on* $\boldsymbol{C} \times I\!\!R^p \times I\!\!R$, *where* C *is a discrete variable on* $\boldsymbol{C} = \{1, \ldots, \bar{C}\}$, *with corresponding strictly positive probabilities* $p_1, \ldots, p_{\bar{C}}$, *and* $E\left(Y^8\right) < \infty$.

Assumption 2.2 *(i) For each* $c = 1, \ldots, \bar{C}$, $f_c(\cdot)$ *and* $R(\cdot, c)f_c(\cdot)$ *belong to* $\mathcal{U}^p \cap \mathcal{D}_{m,q}^p$, $m \geq 2$, *and also* $\sigma_c^2(\cdot)f_c(\cdot)$ *belongs to* $\mathcal{U}^p$. *(ii)* $K \in \mathcal{K}_m^p$, $m \geq 2$.

Our statistic involves one bandwidth h for the pooling null model as well as a set of $\bar{C}$ bandwidths for the alternative general model. We restrict our analysis to the following case.

Assumption 2.3 *For* $c = 1, \ldots, \bar{C}$, $h_c/h \to a_c$ *with* $0 < a_c < \infty$ *and* $nh^p \to +\infty$.

Theorem 1 *Under Assumptions 2.1, 2.2 and 2.3, if* $nh^{(p/2)+2(m+q)} \to 0$, *then as* $n \to +\infty$,

$$
\begin{aligned}
(i) \quad & nh^{p/2}V_n \ \xrightarrow{d} \ N(A\mu, \omega^2) && \text{if } \delta_n^2 nh^{p/2} \to A < \infty, \\
(ii) \quad & nh^{p/2}V_n \ \xrightarrow{p} \ +\infty && \text{if } \delta_n^2 nh^{p/2} \to +\infty,
\end{aligned}
$$

where $\mu = E\left[d^2(X, C)f^2(X)f_C(X)\right]$, $\omega^2 = 2E\left[\sigma_C^2(X)\sigma_{C'}^2(X)f^4(X)E_{CC'}(X)\right]$,

$$E_{CC'}(X) = \int \left[K_{a_C}(s)w_{CC'} - 2(K * K_{a_C})(s)\frac{f_C(X)}{f(X)} + \frac{g^2(X, s)}{f^2(X)}\right]^2 ds$$

and $g^2(X, s) = \sum_{c=1}^{\bar{C}} p_c f_c^2(X)(K * K * K_{a_c})(s)$.

We postpone discussion of our assumptions to the next subsection. As shown in the proofs, the behavior of V_n depends upon whether H_0 holds or does not hold. Under the alternative, V_n converges to a normal distribution with the usual $\sqrt{n}$-rate of convergence. But under the null, the $\sqrt{n}$-asymptotic distribution of V_n is degenerate. This leads us to consider higher-order terms in the asymptotic expansion of V_n. Similar situations also arise in parametric specification testing using functional estimation or significance testing of continuous covariates in nonparametric regression. In such procedures, one also compares two nested models with statistics similar to

6

V_{0n}, where the unknown elements of the null model are replaced by parametric or nonparametric estimators. Because in the latter cases, the null model has a lower effective dimension than the alternative model, estimators in the former have faster pointwise rates of convergence than estimators in the latter. Thus plugging-in estimators in V_{0n} does not affect its asymptotic behavior. In contrast, both the general Model (2.1) and the restricted Model (2.2) have the same effective dimension, so that estimators have similar rates of convergence in both models. Consequently, the asymptotic behavior of V_n differs from the one of V_{0n}. Nevertheless, our results show that plugging-in estimators influences the asymptotic variance under H_0, but affects neither the asymptotic expectation nor the rate of convergence.

In writing the asymptotic variance ω^2, we have adopted the following conventions:

$$E\left[\sigma_C^2(X)\sigma_{C'}^2(X)\Psi(X)\right] \equiv \sum_{c,c'} p_c p_{c'} \int \sigma_c^2(x)\sigma_{c'}^2(x)\Psi(x)f_c(x)f_{c'}(x)\,dx,$$

$$E\left[\sigma_C^2(X)\sigma_{C'}^2(X)\Psi(X)\,\mathit{1\!I}\left[C=C'\right]\right] \equiv \sum_{c,c'} p_c p_{c'} \int \sigma_c^2(x)\sigma_{c'}^2(x)\Psi(x)f_c(x)f_{c'}(x)\,\mathit{1\!I}\left[c=c'\right]\,dx. \quad(2.4)$$

The asymptotic variance of V_n under H_0 has a quite complicated form.[9] First, it depends on the cross-products between $\sigma_c^2(\cdot)$ and $\sigma_{c'}^2(\cdot)$ for different c and c', see Härdle and Marron (1990) for a similar feature. Second, it explicitly depends on the difference in designs between subsamples, through the ratios $f_c(\cdot)/f(\cdot)$ and $g^2(\cdot,\cdot)/f^2(\cdot)$. The first quantity is the ratio of the conditional density of X given $C=c$ to the "average" marginal density $f(\cdot) = \sum_{c=1}^{\bar{C}} p_c f_c(\cdot)$. To interpret the second quantity, assume that $a_c = a$ for any c. Then $g^2(\cdot,s)/f^2(\cdot) = G^2(\cdot)/f^2(\cdot)(K*K*K_a)(s)$, where $G^2(\cdot) = \left(\sum_{c=1}^{\bar{C}} p_c f_c^2(\cdot)\right)$. Thus the second ratio can be given the interpretation of a "normalized variance" of $f_c(\cdot)$. When X is independent of C, both $f_c(\cdot)/f(\cdot)$ and $G^2(\cdot)/f^2(\cdot)$ equal one for any x and c. As we do not require independence, the designs can markedly differ across subsamples and, in general, the latter ratios introduce different weights across the subsamples and the values of explanatory variables. Therefore, it seems impossible to find a kernel that would minimize the variance irrespective of the regressors' designs.

---

[9]It is written in a way that suggests an estimator similar to V_n. See the Appendix for an extensive form of ω^2.

2.3 Discussion on the assumptions and bandwidths' choice

Assumption 2.1 allows for any dependence between (X, Y) and C.[10] Hence, the residuals' distribution is not restricted to be identical for different values of C. The residuals can also be conditionally heteroskedastic. Assumption 2.2 requires smoothness conditions on the underlying functions and kernels that are standard in nonparametric estimation. The compactness of the support of $K(\cdot)$ could be relaxed, but this would lead to very tedious proofs. Our assumptions on the bandwidth h include the usual ones, and specifically imply that $h \to 0$ as $n \to \infty$, while its rate of decrease is restricted by $nh^p \to +\infty$. The condition $nh^{(p/2)+2(m+q)} \to 0$ relates the rate of convergence of the statistic and its bias rate. When comparing two univariate nonparametric regression curves, Härdle and Marron (1990) obtain a statistic with a bias of order $(1/nh)$, which needs to be estimated to construct a test.[11] In our case, the leading term that determines the asymptotic distribution of V_n is unbiased under the null hypothesis. A bias of order $h^{2(m+q)}$ appears in the remaining terms because of our ignorance of the true function $r(\cdot)$. This bias is controlled through the aforementioned condition. With respect to the optimal rate for estimation of the regression function, i.e. $h \propto n^{-1/[p+2(m+q)]}$, this condition implies undersmoothing as is usual in semiparametric estimation, see Robinson (1988) and Powell, Stock and Stoker (1989) among others.[12] Finally, note that our conditions implicitely require $m > p/4$.

Assumption 2.3 ensures on the one hand that each bandwidth h_c converges to zero at a similar rate and on the other hand that the bandwidth h for the pooling model decreases at the same rate. This allows for dependence of the bandwidths on the (sub)sample sizes. The first condition is imposed in all work on the topic using smoothing techniques and is even strengthened to equal bandwidths across subsamples in most papers, see the Introduction for references. The second condition is also imposed in Young and Bowman (1995), where a pooled nonparametric regression is compared to nonparametric regressions based upon subsamples. These authors further argue for equal bandwidths when the design densities are identical across subsamples, so as to equate estimation biases in all models under H_0. By contrast, Baltagi, Hidalgo and Li (1996) assume oversmoothing of the pooling model by imposing $h_c/h^2 = o(1)$.

---

[10] That is, it allows for $\Pr[(X, Y) \le (x, y), C = c] \ne \Pr[(X, Y) \le (x, y), C = c']$ for $c \ne c'$. In particular, the distribution of the regressors can vary across subsamples. See the end of Section 3.1 for a discussion on this point.

[11] A similar feature would appear if we would have not dropped equal indices in V_n.

[12] However, the optimal bandwidth for testing is likely to be different from the optimal bandwidth for estimation, see Guerre and Lavergne (1999) for the case of nonparametric specification testing.

Why is it sensible to use similar amounts of smoothing for all estimators? From an estimation viewpoint, there is no reason why we should employ different smoothing rates in each model. Indeed, a discrete variable does not affect the rate of convergence of nonparametric estimators and does not create any bias in estimation, see e.g. Bierens (1987) and Delgado and Mora (1995). Moreover, each subsample size increases at the same rate as the whole sample size. Therefore, if one wants to select the parameters with respect to some optimality measure, the resulting bandwidths, while depending on possibly different unknown constants, should asymptotically follow the same rate of decrease to zero.[13]

From a testing viewpoint, using bandwidths that fulfil Assumption 2.3 seems the easiest way to put models on equal footing in the testing procedure, because the regressors' dimension is the same between models. For a more theoretical justification, let us consider what happens if we use different amounts of smoothing in the competing models. In this aim, we introduce a new bandwidth a for the pooling model and consider the statistic

$$V_n = \frac{1}{n^{(4)}} \sum_a (Y_i - Y_k)(Y_j - Y_l) \frac{1}{a^{2p}} K\left(\frac{X_i - X_k}{a}\right) K\left(\frac{X_j - X_l}{a}\right) \tilde{K}_{nij} w_{nij}.$$

We maintain Assumption 2.3, which now only imposes similar rates of convergence for the h_c's. If we assume oversmoothing in the null model with respect to the general one by imposing $h/a = o(1)$, V_n is asymptotically equivalent to V_{0n} and has a $nh^{p/2}$ rate of convergence.[14] Conversely, if we assume that $a/h = o(1)$, we obtain a rate of convergence of $na^{p/2}$ for V_n.[15] These results, together with Theorem 1 under the assumption $a = h$, gives us a rate of convergence for V_n of $\min\left(nh^{p/2}, na^{p/2}\right)$. To get the most powerful test, we should clearly maximize this rate by equating h and a, that is, by choosing similar amount of smoothing in the competing models.[16] Hence Assumption 2.3 allows our statistic to be as efficient as possible and in this sense is not restrictive. From a practical viewpoint, our formal analysis under Assumption 2.3 leads to an approximation that is likely to be more useful than a result obtained under one of the alternative assumptions $a/h = o(1)$ or $h/a = o(1)$. Each alternative assumption is akin to ignore some interaction terms in the variance of the statistic, which are however present because these

---

[13]We are also using the same kernel for all estimators because, from their definitions, both functions $r(\cdot)$ and $R(\cdot, \cdot)$ have similar smoothness properties. Hence the order of the kernel should be the same for both models. However, this restriction is not crucial, and a more general result could account for different kernels.

[14]This easily follows by adapting our proof. See Lavergne and Vuong (2000) for a related result.

[15]This also follows by adapting our proof.

[16]The supplementary restrictions on the bandwidths are similar across cases.

ratios are never zero in practice. By contrast, our approach explicitly takes these interaction terms into account.

2.4 Testing procedure and extensions

From a reasoning analogous to the one leading to (2.3), the variance ω^2 can be estimated as

$$\omega_n^2 = \frac{2}{n^{(6)}} \sum_a (Y_i - Y_k)(Y_i - Y_{k'})(Y_j - Y_l)(Y_j - Y_{l'}) K_{nik} K_{nik'} K_{njl} K_{njl'} K_{nij} E_{nij},$$

where $E_{nij} = \mathbb{1}[f_n(X_i) \geq b_n] \int \left[K(s) w_{nij} - 2(K * K)(s) \frac{f_{n,C_i}(X_i)}{f_n(X_i)} + \frac{g_n^2(X_i, s)}{f_n^2(X_i)} \right]^2 ds,$

$g_n^2(x, s) = \sum_{c=1}^{\bar{C}} (n_c/n) f_{n,c}^2(x)(K * K * K_{a_c})(s)$, $\forall x \in I\!\!R^p$, and b_n is a trimming parameter such that $b_n = o(1)$. This trimming is necessary to get rid of edge effects, that is, to avoid undue influence from observation points where the density $f(\cdot)$ is small.[17] An alternative estimator, which is computationally less demanding but more biased in small samples, is

$$\omega_n^2 = \frac{2}{n^{(2)}} \sum_a u_{ni}^2 f_{ni}^2 u_{nj}^2 f_{ni}^2 \tilde{K}_{nij} E_{nij}, \quad \text{where} \quad u_{ni} f_{ni} \equiv [Y_i - r_n(X_i)] f_n(X_i). \tag{2.5}$$

Theorem 2 *Under Assumptions 2.1, 2.2 and 2.3, if $(b_n \sqrt{n} h_n^p)^{-1} \to 0$, $b_n^{-1} h^{m+q} \to 0$ and $n h^{(p/2)+2(m+q)} \to 0$, then as $n \to +\infty$,*

$$(i) \quad n h^{p/2} V_n / \omega_n \xrightarrow{d} N(A\mu/\omega, 1) \qquad \text{if } \delta_n^2 n h^{p/2} \to A < \infty,$$

$$(ii) \quad n h^{p/2} V_n / \omega_n \xrightarrow{p} +\infty \qquad \text{if } \delta_n^2 n h^{p/2} \to +\infty.$$

Theorem 2 provides an equality test across nonparametric regressions. By letting $\delta_n = 0$, the test statistic is asymptotically $N(0, 1)$ under H_0 and, by letting $\delta_n = 1$, it diverges to $+\infty$ under any fixed alternative to H_0. The resulting test is therefore a *one-sided* normal test and is consistent against any fixed alternative. In addition, the test has power to detect local alternatives of the type H_{1n} provided that $\delta_n^2 n h^{p/2} \to +\infty$. In practice, V_n can be more easily computed through

$$
\begin{aligned}
n^{(4)} V_n &= \sum_a u_{ni} f_{ni} u_{nj} f_{nj} \tilde{K}_{nij} w_{nij} + 2 \sum_a u_{ni} f_{ni} (Y_i - Y_j) K_{nij} \tilde{K}_{nij} w_{nij} \\
&\quad - \sum_a (Y_i - Y_j)^2 K_{nij}^2 \tilde{K}_{nij} w_{nij} - \sum_a (Y_i - Y_k)(Y_j - Y_k) K_{nik} K_{njk} \tilde{K}_{nij} w_{nij}.
\end{aligned}
$$

Different extensions of the procedure can be proposed. First, the procedure can be applied to test the significance of any set of qualitative variables in a nonparametric regression. The

---

[17]Unreported simulation results indicate that trimming, though necessary in theory, is not crucial in practice.

variable C is used to recover any combination of the values of the initial discrete variables. Second, one can easily introduce discrete variables in the regressors that are not under test. That is, we can consider (X, D) instead of X, where D is a set of discrete covariates. In that case, one should introduce D in the different functions considered, so that $r(X)$ becomes $r(X, D)$, . . . The asymptotic properties of V_n are as in Theorem 1. As noted before, we can equivalently use either smooth or non-smoothing weights for the discrete variables in D without affecting our procedure's properties. Third, though the theory is developped for generic bandwidths, our result can be extended to vanishing individual bandwidths for each univariate regressor. Fourth, as suggested by a referee, we can allow for data-driven bandwidths.

Theorem 3 *If $K(\cdot)$ is differentiable with bounded partial derivatives, Theorem 2 extends to the case of random data-driven $\widehat{h}$ and $\widehat{h}_c$, $c = 1, \ldots, \bar{C}$, if there exist deterministic sequences h and h_c, $c = 1, \ldots, \bar{C}$, that fulfil the assumptions of Theorem 2 such that $\left(\widehat{h} - h\right)/h$ and $\left(\widehat{h}_c - h_c\right)/h_c$, $c = 1, \ldots, \bar{C}$, are all $o_p(1)$.*

Fifth, as detailled in the next section, both the independence assumption and the identical distribution assumption can be relaxed to some extent.

3   Case of a fixed C

3.1   Cross-section data

There exist situations where the variable defining the division of subsamples is not random, for instance when testing for poolability of cross-section data. This is also true for experiments in which one can control for treatments. The results from the previous section can be adapted to this context. Specifically, let C be a variable taking integer values in $\{1, \ldots, \bar{C}\}$. For each c, assume that we have at hand an i.i.d. sample of size n_c from a random variable (X_c, Y_c) on $\mathbb{R}^p \times \mathbb{R}$, such that X_c has marginal density $f_c(\cdot)$. We employ similar notations as in Section 2, so that, for each c, the sample from (X_c, Y_c) is denoted $\left\{(X_i, Y_i), i = 1 + \sum_{c' < c} n_{c'}, \ldots, \sum_{c' \le c} n_{c'}\right\}$ and $\sum_{c=1}^{\bar{C}} n_c = n$. We then consider the general regression model

$$Y_i = R(X_i, c) + U_i, \qquad E[U_i|X_i] = 0 \qquad i = 1, \ldots, n, \tag{3.1}$$

so that now $R(\cdot, c)$ denotes the regression function of Y_c on X_c. For each c, nonparametric kernel estimators of $R(\cdot, c)$ and $f_c(\cdot)$ are defined as in the previous section. Overlooking the

information given by C and assuming falsely that the observations constitute an i.i.d. sample leads to consider the pooling regression model (2.2). Nonparametric kernel estimators of $f_n(\cdot)$ and $r_n(\cdot)$ are defined as in the previous section, but their interpretation radically changes. Here $f_n(\cdot)$ estimates $f(\cdot) = \sum_{c=1}^{\bar{C}} n_c f_c(\cdot)/n$, which is not the marginal density of an observed random variable X, but the density of a hypothetical variable constructed from the different X_c's. Similarly, $r_n(\cdot)$ estimates $r(\cdot) = \sum_{c=1}^{\bar{C}} n_c R(\cdot, c) f_c(\cdot)/n f(\cdot)$, which is no more a conditional expectation function, but a weighted average of the regression functions $R(\cdot, c)$, $c = 1, \ldots, \bar{C}$.

While the two models are now interpreted differently, the framework is similar. The null hypothesis of interest is still the constancy of the regression function $R(\cdot, c)$ for different values of c, so that it is possible to estimate it from the pooling model, even though the densities $f_c(\cdot)$ differ for different values of c. Thus, the null hypothesis of interest reads

$$H_0 : R(X, c) = r(X) \text{ a.s.} \qquad \forall c = 1, \ldots, \bar{C}$$

and the statistic V_n is constructed as in (2.3). We let $\sigma_c^2(X) \equiv E\left[(Y_c - r(X_c))^2 \,|\, X_c = X\right]$. The following assumption summarizes the considered setup for a fixed C.

Assumption 3.1 *Let C be a fixed variable taking integer values in $\{1, \ldots, \bar{C}\}$. For $c = 1, \ldots, \bar{C}$, each subsample $\left\{(X_i, Y_i), i = 1 + \sum_{c' < c} n_{c'}, \ldots, \sum_{c' \le c} n_{c'}\right\}$ is an i.i.d. sample of size n_c from a random variable (X_c, Y_c) on $\mathbb{R}^p \times \mathbb{R}$, such that X_c has marginal density $f_c(\cdot)$ and $E\left(Y_c^8\right) < \infty$. The subsamples are independent. For any c, $n_c/n \to p_c$ as $n = \sum_{c=1}^{\bar{C}} n_c \to +\infty$.*

Theorem 4 *Under Assumptions 3.1, 2.2 and 2.3, if $nh^{p/2 + 2(m+q)} \to 0$, then as $n \to +\infty$,*

$$
\begin{aligned}
(i) \quad & nh^{p/2} V_n \xrightarrow{d} N(A\mu, \omega^2) && \text{if } \delta_n^2 nh^{p/2} \to A < \infty, \\
(ii) \quad & nh^{p/2} V_n \xrightarrow{p} +\infty && \text{if } \delta_n^2 nh^{p/2} \to +\infty,
\end{aligned}
$$

where $\mu = \sum_{c=1}^{\bar{C}} p_c E\left[d^2(X, c) f^2(X) f_c(X)\right]$, $\omega^2 = 2 \sum_{c,c'} p_c p_{c'} E\left[\sigma_c^2(X) \sigma_{c'}^2(X) f^4(X) E_{cc'}(X)\right]$,

$$E_{cc'}(X) = \int \left[K_{a_c}(s) w_{cc'} - 2(K * K_{a_c})(s)\frac{f_c(X)}{f(X)} + \frac{g^2(X, s)}{f^2(X)}\right]^2 ds,$$

$w_{cc'} = (1/p_c)\, \mathbb{1}[c = c']$ and $g^2(X, s) = \sum_{c=1}^{\bar{C}} p_c f_c^2(X)(K * K * K_{a_c})(s)$.[18]

Compared to the previous section, we have relaxed the assumption of identically distributed data across subsamples, but still assume independence across subsamples, which is typically

---

[18]In writing ω^2, we use similar conventions as in (2.4).

the case for cross-section data. The proof of Theorem 4 mainly follows the one of Theorem 1, see Appendix A for some brief explanations. Theorems 2 and 3 can be similarly extended. A testing procedure can then be based on $nh^{p/2}V_n/\omega_n$, where ω_n is an estimator of the asymptotic variance ω^2 analogous to the ones in Section 2.4. As before, the test is a one-sided normal test, is consistent against any alternative and detects local alternatives of the type H_{1n} approaching the null at a rate slower than $(nh^{p/2})^{-1/2}$.[19]

There are interesting connections between our procedure and analysis of covariance. The simple analysis of variance model reads

$$Y_i = \beta_{C_i} + U_i, \qquad E[U_i] = 0.$$

For testing the hypothesis $\beta_c = \beta$, for all c, the usual testing procedure is built upon

$$S = (1/n) \sum_c n_c \left[\bar{Y}_c - \bar{Y}\right]^2,$$

where $\bar{Y}_c = (1/n_c) \sum_{i=1}^n Y_i \math4[C_i = c]$ and $\bar{Y} = (1/n) \sum_{i=1}^n Y_i$. Now S equivalently reads

$$S = (1/n^4) \sum_{i,j,k,l} (Y_i - Y_k)(Y_j - Y_l) \frac{n}{n_{C_i}} \math4[C_i = C_j].$$

Our statistic V_n is analogous to S (with the slight difference that it excludes equal indices in the sum), but weights the first differences in the dependent variable by quantities that depend upon explanatory variables. Thus our testing procedure provides a nonparametric analysis of covariance, which allows to detect any differential effect of the regressors on the dependent variable across the considered subsamples without imposing any parametric assumption.

Three main remarks follow from this interpretation. The first remark is related to the choice of weights in the null hypothesis

$$H_0 : R(X, c) = r(X) \equiv \frac{\sum_{c=1}^{\bar{C}} p_c R(X, c) f_c(X)}{\sum_{c=1}^{\bar{C}} p_c f_c(X)} \text{ a.s.} \quad \forall c = 1, \ldots, \bar{C}.$$

When the observations constitute a unique random sample, it is meaningful to use a weighting scheme proportional to $p_c f_c(.)$. When C is fixed, the use of "frequencies" as weights in H_0 is no longer readily interpretable. Obviously, there exist other equivalent formulations of the

---

[19] By contrast, Yatchew's (1999) test can detect local alternatives that converges to the null at a rate slower than $n^{-1/4}$, which is in turn slower than $(nh^{p/2})^{-1/2}$.

hypothesis that $R(\cdot, c)$ is constant in c that use different weighting schemes. To each formulation corresponds a test statistic, whose relative merits generally depend on the data at hand.

The second remark concerns problems in application and interpretation of the procedure. As noted by Scheffé (1959, p. 198), "it is sometimes said that the analysis of covariance is valid only if the treatments do not affect the values of the concomitant variables. The dictum that the analysis of covariance can be used only in this case would thus confine it to a very restricted situation. The analysis of covariance can be applied to get tests of hypotheses that have correct significance level, but the sense of using these tests must be considered separately in each application." This statement remains true for the nonparametric analysis of covariance proposed here. Specifically, we allow the regressors' designs to vary across subsamples, so that the "treatments" — i.e. the discrete variable C — may affect explanatory variables. The procedure is widely applicable, but may give a right answer to a wrong question. If some of the regressors are "part of the treatment", e.g. if the regressors have different supports depending on the values of C, then the null hypothesis H_0 is no longer meaningful.[20] This second remark is equally valid in the case of a random C.

Third, our procedure only applies for testing the strict equality of regression functions. If one wants to test equality up to some parametric transformations, one should build a specific test statistic that accounts for this at the outset, as done by Härdle and Marron (1990). Even in the simple case of testing for parallelism of the regression curves, which is easily entertained within the linear parametric analysis of covariance framework, adapting our test is not completely straightforward. This extension is left for future work.

3.2 Panel data

A useful extension of our test is testing for poolability of panel data. Consider the model

$$Y_{it} = R_t(X_{it}) + U_{it}, \qquad i = 1, \ldots, n_0, \qquad t = 1, \ldots, T. \tag{3.2}$$

At each period t, $\{X_{it}, i = 1, \ldots, n_0\}$ is an i.i.d. sample from X_t with density $f_t(\cdot)$. The number T of time periods is fixed while the sample size for each period n_0 goes to infinity. This is the standard case, which is of most practical relevance, since one usually has a large number of

---

[20]In Quade (1982), the first proposed method is valid only under the assumption that the distribution of X does not vary conditionally to C. The second one does not require such an assumption and is close in spirit to our analysis, with the major difference that the bandwidth is held fixed.

cross-section units but data covering only a few time-periods.[21] The null hypothesis of interest is the constancy of the regression function $R_t(\cdot)$ over time, that is

$$H_0 : R_t(X) = r(X) \text{ a.s.}$$

where $r(\cdot) \equiv (1/T) \sum_{t=1}^{T} f_t(\cdot) R_t(\cdot)/f(\cdot)$ and $f(\cdot) \equiv (1/T) \sum_{t=1}^{T} f_t(\cdot)$. The statistic V_n is

$$V_n = \frac{T}{n^{(4)}} \sum_a (Y_{it} - Y_{kr})(Y_{jt} - Y_{ls}) \frac{1}{h^{2p}} K\left(\frac{X_{it}-X_{kr}}{h}\right) \frac{1}{h_t^p} K\left(\frac{X_{jt}-X_{ls}}{h}\right) K\left(\frac{X_{it}-X_{jt}}{h_t}\right). \qquad (3.3)$$

Here $n = n_0 T$, $\sum_a$ denotes summation over arrangements of distinct indices $\{it, jt, kr, ls\}$, h and h_t, $t = 1, \ldots, T$, are bandwidth parameters such that $h_t/h \to a_t$, $0 < a_t < +\infty$, for all t. The results of the Section 3.1, where we imposed independence across subsamples, do not readily apply in this context. Nevertheless, as we argue below, this assumption can be weakened and the asymptotic behavior of V_n as $n_0 \to \infty$ is given by Theorem 4.[22] Its asymptotic variance is

$$\omega^2 = (2/T^2) \sum_{t=1}^{T} \sum_{t'=1}^{T} E\left[\sigma_t^2(X)\sigma_{t'}^2(X)f^4(X)E_{tt'}(X)\right],$$

$$\text{where } E_{tt'} = \int \left[T\, K_{a_t}(s)\, I\!I\left[t = t'\right] - 2(K * K_{a_t})(s)\frac{f_t(X)}{f(X)} + \frac{g^2(X,s)}{f^2(X)}\right]^2 ds$$

and $g(X,s) = (1/T) \sum_{t=1}^{T} f_t^2(\cdot)(K * K * K_{a_t})(s)$.

The usual way of considering panel data models in econometrics is to see $R_t(\cdot)$ as the conditional expectation of Y_t given all past explanatory variables $\{X_1, \ldots, X_t\}$ and a time-independent latent variable l. This formulation is quite general, and in particular allows for some lagged dependent variable in the regressors, so that further restrictions are usually imposed on the model. Chamberlain (1984) distinguishes two main restrictions: lack of residual serial correlation and no structural lagged dependent variables.

There is residual serial correlation conditional on a latent variable l if Y_t is not independent of $\{Y_1, \ldots, Y_{t-1}\}$ *conditional on* $\{X_1, \ldots, X_t, l\}$.

The relationship of X to Y is static conditional on a latent variable l if X is strictly exogeneous conditional on l and if Y_t is independent of $\{X_1, \ldots, X_{t-1}\}$ *conditional on X_t and l. If the*

---

[21]The increasing availability of micropanels in which T is not negligible relative to n_0 motivates rising interest for further analysis. See e.g. Alvarez and Arellano (1998) for recent results and an account of issues that still need to be addressed in a parametric context.

[22]A brief account of the necessary adaptations of the proof is given in Appendix A.

relationship of X to Y is static conditional on a latent variable l, then there are no structural lagged dependent variables.

Our analysis imposes the two restrictions of no serial residual correlation and of a static relationship of X to Y (both conditional on a latent variable l). First, though Assumption 3.1 imposes independence between subsamples, inspection of the proofs reveals that we can alleviate the independence requirement and replace it by the assumption of no serial residual correlation. This assumption allows for fixed individual effects correlated with the regressors. In a nonparametric context, such effects are not separately identifiable.[23] Second, Model (3.2) also assumes that the regression function does not depend on $\{X_1, \ldots, X_{t-1}\}$. This is true when the relationship of X to Y is static conditional on a latent variable l. But, as shown by Chamberlain (1984), there is no restriction to assume a static conditional relationship in a fully nonparametric context. It is restrictive only when combined with a specific functional form of the distribution. Hence, the restrictions imposed in our analysis are not as stringent as they may appear at first.

Baltagi, Hidalgo and Li (1996) consider a statistic close to ours, with the crucial difference that they use a smoothing parameter a for the pooling model and the same bandwidth h at each period for Model (3.2). They subsequently impose $h/a^2 = o(1)$. As a consequence, the asymptotic variance of the test statistic becomes

$$\omega_0^2 = 2 \int K^2(s)\, ds \sum_{t=1}^{T} E\left[\sigma_t^4(X) f^4(X) f_t(X) \right].$$

Contrary to ω^2, the variance ω_0^2 depends neither on cross-products of conditional variances between periods nor explicitly on the designs' differences between periods.[24] This occurs because using different amounts of smoothing in the null and the alternative model results in pulling out any cross-effect between periods. This is why the previous authors do not need the assumption of no serial residual correlation. As argued in Section 2.3, the restriction $h/a^2 = o(1)$ does not allow to attain the maximum rate of convergence of the statistic. The only justification advocated by the authors is that otherwise one should estimate the expectation of V_n to recenter the test statistic, and that this would introduce finite-sample bias. This argument is invalid from both a theoretical viewpoint, as our analysis demonstrates, and a practical viewpoint, as shown by the following simulation results.

---

[23]On this topic, see the discussion of Baltagi, Hidalgo and Li (1996).

[24]Our formula gives the general form of the asymptotic variance ω_0^2. The one given in Baltagi, Hidalgo and Li (1996) assumes that the density of the regressors remains the same across time and has $f^5(\cdot)$ in place of $f^4(\cdot) f_t(\cdot)$.

4 Small sample behavior

We now investigate the behavior of our test through a Monte-Carlo experiment in the case of a random C taking two values 0 or 1 with equal probabilities. We generate the data through

$$Y = \beta X + \gamma X^3 + I\!I[C = 0] \ d(X) + U, \tag{4.1}$$

where conditional on C, X is generated as $N(C, 1)$ and U is independently distributed as $N(0, 1)$. We impose the restriction that $E[d(X)|C = 0] = 0$ and we set parameters β and γ to -4 and 1 respectively, so that $E(Y|C = 0) = E(Y|C = 1) = 0$.

The null hypothesis corresponds to $d(X) \equiv 0$. The alternatives are chosen to investigate the power of our test when their magnitude or frequency vary. For the magnitude, we consider three linear alternatives of the form

$$d(X) = \alpha X,$$

with $\alpha = 0.5, 1$ and 2 corresponding respectively to DGP_1, DGP_2 and DGP_3. This allows to compare the performances of our procedure to the standard Chow test based on the true Model (4.1). Alternatives corresponding to varying frequencies are defined through

$$d(X) = \sin(\alpha \pi X),$$

with $\alpha = 2, 1, 2/3$ and $1/2$ corresponding respectively to DGP_4, DGP_5, DGP_6 and DGP_7. These departures from the null are of special interest, for it is known that smooth tests of parametric specification and nonparametric significance tests for continuous regressors are sensitive to the frequency of the alternative, see Hart (1997) and Lavergne and Vuong (2000).

We consider small ($n = 100$) and moderate ($n = 200$) sample sizes and run 2000 replications. Unreported simulation results show that the trimming parameter has very little influence on the results, so that it is arbitrarily set to 0 in all experiments. The bandwidth parameter h is chosen as $h = \hat{s}_X \ n^{-\lambda}$, where $\hat{s}_X$ is the estimated standard deviation for all observations of X. We let $h_0 = h_1$, so that $a_0 = a_1 = a$. This choice allows to save much computations and has also the advantage that it mimics the choice of an uninformed practitioner who only knows that $p_0 = p_1 = 1/2$. For ease of computation, we choose the uniform kernel with support $[-1/2, 1/2]$. Appendix C gives the values with respect to a of the integrals that enter ω^2.[25]

---

[25]The Mathematica notebook used to compute the integrals can be adapted for other kernels, and is available from the author upon request.

We first focus on how bandwidths' choice for the alternative model, i.e. $h_0 = h_1$, influence the test's outcome, while keeping fixed the restricted model's bandwidth h, as done by Lavergne and Vuong (2000) in another context. We then let a vary in the grid $(0.5, 1, 1.5, 2)$, while keeping $h = \hat{s}_X n^{-1/5}$.[26] Table 1 reports our results for the null hypothesis (DGP_0) and the linear alternatives. For each case, the first row gives the mean of our test statistic with its standard deviation in parentheses. The second row gives empirical rejection rates of the test, the first figure corresponds to a 5% nominal level and the second one to a 10% nominal level. The third row gives corresponding size-corrected empirical rejection rates for the alternative, except for DGP_0 where it gives the empirical critical values at 5% and 10% levels. For each sample size, the last row reports rejection rates of the Chow test for the same nominal levels.

The first column of Table 1 relates to the null hypothesis. For any choice of a, the test statistic is very close to be unbiased. The standard deviation of our test statistic is close to, but stays smaller than one. This is due partly to the fact that, to save computations, we use the simplest estimator of the variance (2.5), which is positively biased in small samples. A similar feature appears in the simulations performed by Lavergne and Vuong (2000) on their nonparametric significance test for continuous regressors.[27] While the mean seems to have an inversed U-shape with respect to a, no clear relationship appears to link the standard deviation and a. Under the null hypothesis, empirical sizes are quite close to nominal levels. There is no evidence of systematic under or overrejection at a 5% nominal level. At a 10% level, there is slight underrejection for all our choices of a. Finally the empirical critical values are very close to the ones derived from the normal approximation, with for instance a value of 1.645 at a 5% level when $a = 1$ and $n = 200$.

Regarding the linear alternatives, we find as expected that power is increasing with the sample size and the magnitude of the departure from the null, as measured by α. The choice $a = 0.5$ leads to small power, especially for alternatives of little amplitude. Though, our test can reasonably detect moderate linear alternatives for a bandwidth $h_0 = h_1$ at least as great as h. Furthermore, for an alternative of amplitude $\alpha = 2$, the power performance of our test can equal that of the Chow test, although the design is ideal for the latter. The highest power

---

[26] This choice corresponds to a practical rule-of-thumb and is not supposed to be optimal for the testing problem at hand. In our setup, we have $m = 2$ and $q = 1$, so that h fulfils Theorem 1's conditions.

[27] In the latter study, it has also been observed that better estimators of the variance are obtained by using $K^2(\cdot)$ instead of $K(\cdot)$ times the integral of $K^2(\cdot)$. A similar feature is expected to hold in our case.

is attained for the largest considered bandwidth, which could be expected since the alternative is linear and the kernel smoother is a straight line for large bandwidths. Though, too large a bandwidth is likely to give low power.

Table 2 has the same structure as Table 1 and reports results relative to the sinus alternatives. The empirical power of the test exhibits an inverse U-shape as a function of the bandwidth $h_0 = h_1$ for the high frequency alternative DGP_4. This suggests that the bandwidth should be adapted to the frequency of the alternative. This is not surprising, as a similar feature appears in other smooth tests, see e.g. the simulations results of Delgado, Dominguez and Lavergne (1998) and Lavergne and Vuong (2000). For comparative purposes, we also provide the empirical rejection rates of the Chow test assuming a linear specification in X. The lowest frequency alternative DGP_7 is close to a linear specification in the range $[-1, 1]$. Given that X is $N(0, 1)$ when $C = 0$, the Chow test therefore performs quite well, while our test has high power for large a. For higher frequency alternatives DGP_4 and DGP_5, the Chow test has either trivial or low power irrespective of the sample size, while the empirical power of our test can exceed 85% for a moderate sample size of 200.

We now focus on how the choice of h influences our test, while keeping fixed the ratio of bandwidths $h_0/h = h_1/h = 1$. We then let λ vary in the grid $(1/3, 1/4, 1/5, 1/6)$. Table 3 (respectively Table 4) has the same structure as Table 1 (Table 2) and report our results for the null hypothesis and the linear alternatives (respectively the sinus alternatives). For a choice of λ of at least $1/5$, the statistic is close to be unbiased under H_0, while its expectation increases when λ decreases to $1/6$ as expected. The variance is decreasing in λ. On the whole, decreasing λ leads to more rejections. For linear alternatives, empirical power is decreasing in λ, whether it is size-adjusted or not. For sinus alternatives, empirical power can be low when λ is large, especially for a small sample size, while it highly increases when λ decreases. For DGP_4 and $n = 100$, size-adjusted power exhibits an inverse U-shape as a function of h. In the other cases, it highly increases when h increases.

5 Conclusion

In this paper we propose a general test of equality across nonparametric regressions. The procedure is based on the comparison of the regression functions for each subsample with the general one that pools all observations. It applies in a variety of situations, whether the division into

subsamples is defined in a random way or not. In our presentation, we have first considered the leading situation where a random qualitative variable defines the split into different subsamples. Then, by considering the case where the split depends on a fixed qualitative variable, we have shown how our basic assumptions can be weakened so that our test applies to cross-section and panel data. In summary, our testing procedure is applicable in any case where the observations are i.i.d. within each subsample and where the residuals are uncorrelated across subsamples.

The characteristic feature of our testing problem is that both models (under the null and under the alternative) have the same effective dimension. Hence our procedure uses similar amount of smoothing for the pooling and the alternative model. We have justified this assumption and investigated thoroughly its implications. Though, bandwidths' choice is clearly a key issue for application of the test. Bootstrap methods could be a way to bypass this problem, as they usually provide better approximations to the null distribution than asymptotics do and can be much less sensitive to bandwidths' choice, see e.g. Delgado, Dominguez and Lavergne (1998). This possibility should be investigated from both a theoretical and a practical viewpoint.

Appendix A: Proofs

In what follows, $u_i \equiv Y_i - r(X_i)$, $U_i \equiv Y_i - R(X_i, C_i)$, $f_i \equiv f(X_i)$, $r_i \equiv r_i(X_i)$, $d_i \equiv d(X_i, C_i)$, $\sigma_i^2 \equiv \sigma_{C_i}^2(X_i)$ and $Z_i \equiv (C_i, X_i, Y_i)$, $i = 0, 1, \ldots, n$. Also $K_{nik} \equiv h^{-p} K\left[(X_i - X_k)/h\right]$, $h_i \equiv h_{n_{C_i}}$, $\boldsymbol{K}(\cdot) \equiv |K(\cdot)|$ and $i, j, k, l, i', j', k', l'$ refer to indices that are pairwise different unless stated otherwise. We let $\widehat{f}_i \equiv (n-1)^{-1} \sum_{k \neq i} K_{nik}$, and more generally for any index set I not containing i with cardinality $|I|$, $\widehat{f}_i^I \equiv (n - 1 - |I|)^{-1} \sum_{k \neq i, k \notin I} K_{nik}$. To ease exposition, we assume $h_c/h = a_c$, for all c.

A.1   Proof of Theorem 2.1

As $Y_i - Y_k = (u_i - u_k) + (r_i - r_k)$, and as $K(\cdot)$ is even by Assumption 2.2, we have from (2.3)

$$
\begin{aligned}
V_n &= \frac{1}{n^{(4)}} \sum_a (u_i - u_k)(u_j - u_l) K_{nik} K_{njl} \tilde{K}_{nij} w_{nij} + \frac{2}{n^{(4)}} \sum_a (u_i - u_k)(r_j - r_l) K_{nik} K_{njl} \tilde{K}_{nij} w_{nij} \\
&\quad + \frac{1}{n^{(4)}} \sum_a (r_i - r_k)(r_j - r_l) K_{nik} K_{njl} \tilde{K}_{nij} w_{nij} = I_1 + 2I_2 + I_3,
\end{aligned}
$$

where $I_1 = \dfrac{n-2}{n-3} \dfrac{1}{n^{(2)}} \sum_a u_i u_j f_i f_j \tilde{K}_{nij} w_{nij} + \dfrac{2(n-2)}{n-3} \dfrac{1}{n^{(2)}} \sum_a u_i (\widehat{f}_i^j - f_i) u_j f_j \tilde{K}_{nij} w_{nij}$

$\qquad\qquad + \dfrac{n-2}{n-3} \dfrac{1}{n^{(2)}} \sum_a u_i (\widehat{f}_i^j - f_i) u_j (\widehat{f}_j^i - f_j) \tilde{K}_{nij} w_{nij} - \dfrac{2}{n^{(3)}} \sum_a u_i f_i u_l K_{njl} \tilde{K}_{nij} w_{nij}$

$$- \frac{2}{n^{(3)}} \sum_a u_i(\widehat{f}_i^{j,l} - f_i)u_l K_{njl}\tilde{K}_{nij}w_{nij} + \frac{1}{n^{(4)}} \sum_a u_k u_l K_{nik} K_{njl}\tilde{K}_{nij}w_{nij}$$

$$- \frac{1}{n^{(4)}} \sum_a u_i u_j K_{nik} K_{njk}\tilde{K}_{nij}w_{nij} = \frac{n-2}{n-3}[V_{0n} + 2I_{1,1} + I_{1,2}] - 2I_{1,3} - 2I_{1,4} + I_{1,5} - I_{1,6},$$

$$I_2 = \frac{1}{n^{(3)}} \sum_a u_i f_i(r_j - r_l) K_{njl}\tilde{K}_{nij}w_{nij} + \frac{1}{n^{(3)}} \sum_a u_i(\widehat{f}_i^{j,l} - f_i)(r_j - r_l) K_{njl}\tilde{K}_{nij}w_{nij}$$

$$- \frac{1}{n^{(4)}} \sum_a u_k(r_j - r_l) K_{nik} K_{njl}\tilde{K}_{nij}w_{nij} = I_{2,1} + I_{2,2} - I_{2,3}.$$

Propositions 1 to 12 study each of the above terms. Collecting results, it follows that

$nh^{p/2}[V_{0n} - 2I_{1,3} + I_{1,5}] = L_n + \delta_n^2 nh^{p/2}\mu_n + \delta_n\sqrt{n}h^{p/2}O_p(1) + o_p(1)$, where $L_n \xrightarrow{d} N(0, \omega^2)$ and $\mu_n \to \mu$

and $nh^{p/2}[V_n - (V_{0n} - 2I_{1,3} + I_{1,5})] = \delta_n^2 nh^{p/2}o_p(1) + \delta_n\sqrt{n}h^{p/2}O_p(1) + \delta_n nh^{p/2}h^{(m+q)}O_p(1) + o_p(1)$.

If $\delta_n^2 nh^{p/2} \to A$, $\delta_n^2\sqrt{n}h^{p/2}$ and $\delta_n nh^{p/2}h^{(m+q)}$ are both $o(1)$, so that $nh^{p/2}V_n \xrightarrow{d} N(A\mu, \omega^2)$.

If $\delta_n^2 nh^{p/2} \to +\infty$, $\delta_n^2\sqrt{n}h^{p/2}$ and $\delta_n nh^{p/2}h^{(m+q)}$ are both $o(\delta_n^2 nh^{p/2})$, so that $nh^{p/2}V_n \xrightarrow{p} +\infty$. Q.E.D.

In the following proofs, we will rely on the following lemma.

Lemma 1 (Fan and Li, 1996) *Let U_n be a U-statistic of order m with symmetrical kernel H_n such that $E[H_n(Z_1, \ldots, Z_m)|Z_1] = 0$ a.s. and $E[H_n^2(Z_1, \ldots, Z_m)] < \infty$ for each n.*
Let $H_{n,s} = E[H_n(Z_1, \ldots, Z_m)|(Z_1, \ldots, Z_s)]$, $\xi_s = \text{Var}[H_{n,s}(Z_1, \ldots, Z_s)]$ for $s = 2, \ldots, m$ and $G_n(Z_1, Z_2) = E[H_{n,2}(Z_1, Z_0), H_{n,2}(Z_2, Z_0)|(Z_1, Z_2)]$. If $\xi_s/\xi_2 = o(n^{s-2})$ for $s = 1, \ldots, m$, and

$$\frac{E[G_n^2(Z_1, Z_2)] + n^{-1}E[H_{n,2}^4(Z_1, Z_2)]}{E^2[H_{n,2}^2(Z_1, Z_2)]} \to 0 \quad \text{as } n \to \infty,$$

then nU_n is asymptotically normal with zero mean and variance $(1/2)(m^{(2)})^2 \xi_2$.

As we consider U-statistics with non-symmetric kernel in our proofs, we briefly explain how Lemma 1 extends to this case. One first needs to replace $H_n(Z_1, \ldots, Z_m)$ by the symmetric kernel

$$\tilde{H}_n(Z_1, \ldots, Z_m) = \frac{1}{m!} \sum_p H_n(Z_{i_1}, \ldots, Z_{i_m}),$$

where $\sum_p$ denotes summation over the $m!$ permutations of $(1, \ldots, m)$. If $E\left[\tilde{H}_n(Z_1, \ldots, Z_m)|Z_1\right] = 0$ a.s., the U-statistic is degenerate and under the assumptions of Lemma 1 converges in distribution to a centered normal distribution. Its asymptotic variance is given by

$$(1/2)\left(m^{(2)}\right)^2 \xi_2 = (1/2)\left(m^{(2)}\right)^2 (1/m!)^2 \sum_p \sum_{p'} E[H_n(Z_{i_1}, \ldots, Z_{i_m})H_n(Z_{i'_1}, \ldots, Z_{i'_m})|Z_1, Z_2].$$

Hence one needs to determine all the terms in the double summation. Similar expressions are derived for $\xi_s, s = 3, \ldots, m$. An analogous result also holds for two U-statistics of order m_1 and m_2 with respective kernels $H_{1,n}$ and $H_{2,n}$. Their asymptotic covariance is then given by

$$(1/2)m_1^{(2)}m_2^{(2)}(1/m_1!)(1/m_2!) \sum_p \sum_{p'} E\left[H_{1,n}(Z_{i_1}, \ldots, Z_{i_{m_1}})H_{2,n}(Z_{i'_1}, \ldots, Z_{i'_{m_2}})|Z_1, Z_2\right].$$

Proposition 1 $nh^{p/2}V_{0n} = nh^{p/2}U_{0n} + \delta_n^2 nh^{p/2}\mu_n + \delta_n\sqrt{n}h^{p/2}B_n$, where $\mu_n \longrightarrow \mu$, $nh^{p/2}U_{0n} \overset{d}{\longrightarrow} N(0, \omega_0^2)$ and $B_n \overset{d}{\longrightarrow} 2N(0, \zeta - \delta^2\mu^2)$, with $\delta = \lim_{n\to\infty}\delta_n$ and μ, ω_0^2 and ζ subsequently defined.

PROOF: $V_{0n} = U_{0n} + W_{0n} - \theta_n$, where $H_n(Z_i, Z_j) = u_i u_j f_i f_j \tilde{K}_{nij} w_{nij}$, $\theta_n = E[H_n(Z_1, Z_0)]$, $W_{0n} = (2/n)\sum_i E[H_n(Z_i, Z_0)|Z_i]$ and

$$U_{0n} = \binom{n}{2}^{-1}\sum_{i<j}\tilde{H}_n(Z_i, Z_j) = \binom{n}{2}^{-1}\sum_{i<j}\{H_n(Z_i, Z_j) - E[H_n(Z_i, Z_0)|Z_i] - E[H_n(Z_0, Z_j)|Z_j] + \theta_n\}.$$

(i)
$$\begin{aligned}
\theta_n &= E\left[u_i f_i u_j f_j \tilde{K}_{nij} w_{nij}\right] = E\left[(U_i + \delta_n d_i)f_i(U_j + \delta_n d_j)f_j \tilde{K}_{nij} w_{nij}\right] \\
&= \delta_n^2 E\left[d_i f_i d_j f_j \tilde{K}_{nij} w_{nij}\right] = \delta_n^2 E\left[d_i f_i w_{nij} E\left(d_j f_j \tilde{K}_{nij}|X_i, C_i, C_j\right)\right].
\end{aligned}$$

Now $E\left[\delta_n d_j f_j \tilde{K}_{nij}|X_i, C_i, C_j\right] = \delta_n d(X_i, C_j)f(X_i)f_{C_j}(X_i) + o(\delta_n)$ uniformly in X_i by Lemma 2, as $\delta_n d(X, C)f(X)f_C(X) \in \mathcal{U}^p$, $\forall C$. Therefore $\theta_n = \delta_n^2 E\left[d_i^2 f_i^2 E(w_{nij}|C_i)f_{C_i}(X_i)\right] + o(\delta_n^2) = \delta_n^2 \mu_n$ with $\mu_n \longrightarrow \mu = E\left[d^2(X, C)f^2(X)f_C(X)\right]$, as $E(w_{nij}|C_i) = 1$.

(ii) Distribution of W_{0n}: $E\left[E^2(H_n(Z_i, Z_0)|Z_i)\right] = E\left[u_i^2 f_i^2 E^2(u_0 f_0 \tilde{K}_{ni0} w_{ni0}|Z_i)\right] = \delta_n^2 \zeta_n$ with $\zeta_n \longrightarrow \zeta = E\left[\sigma_C^2(X)d^2(X, C)f^4(X)f_C^2(X)\right]$, as $\delta_n d(X, C)f(X)f_C(X) \in \mathcal{U}^p$, $\forall C$.

Now $E \mid E[H_n(Z_i, Z_j)|Z_i] \mid^\nu = E \mid u_i^\nu f_i^\nu E^\nu\left[u_0 f_0 \tilde{K}_{ni0} w_{ni0}|Z_i\right] \mid = O(1) = o(n^{\nu/2-1})$ for some $2 < \nu \le 4$, as $E \mid Y^8 \mid < \infty$. Thus, by Theorem 7.1 of Hoeffding (1948), $\sqrt{n}[W_{0n} - 2\theta_n] \to 2\delta N\left(0, \zeta - \delta^2\mu^2\right)$.

(iii) Distribution of U_{0n}: As $E\left[\tilde{H}_n(Z_i, Z_j)|Z_i\right] = 0$, by Lemma 1,

$$nh^{p/2}U_{0n} \overset{d}{\longrightarrow} N(0, \tau^2) \qquad \text{if} \qquad \frac{E[\tilde{G}_n^2] + n^{-1}E[\tilde{H}_n^4]}{E^2[\tilde{H}_n^2]} = o(1),$$

where $\tilde{G}_n(Z_i, Z_j) = E\left[\tilde{H}_n(Z_i, Z_0)\tilde{H}_n(Z_j, Z_0)|Z_i, Z_j\right]$ and $\tau^2 = 2\lim_{n\to\infty}h^p E(\tilde{H}_n^2)$. By definition of $\tilde{H}_n(Z_i, Z_j)$, the above is equivalent to

$$\frac{E[G_n^2] + n^{-1}E[H_n^4]}{E^2[H_n^2]} = o(1), \tag{A.1}$$

where $G_n(Z_i, Z_j) = E[H_n(Z_i, Z_0)H_n(Z_j, Z_0)|Z_i, Z_j]$, and $\tau^2 = 2\lim_{n\to\infty}h^p E(H_n^2)$.

As $\sigma_C^2(X)f^2(X)f_C^2(X) \in \mathcal{U}^p$, $\forall C$, $E\left[H_n^2(Z_i, Z_j)\right] = E\left[\sigma_{C_i}^2(X_i)\sigma_{C_j}^2(X_j)f_i^2 f_j^2 \tilde{K}_{nij}^2 w_{nij}^2\right] = h^{-p}\omega_{0,n}^2/2$, where $\omega_{0,n}^2 \longrightarrow \omega_0^2 = 2E\left[\sigma_C^2(X)\sigma_{C'}^2(X)w_{CC'}^2 f^4(X)\int K_{a_C}^2(t)\,dt\right]$, using (2.4).

Moreover, as $E(u^4|X, C)f^4(X)f_C(X) \in \mathcal{U}^p$, $\forall C$,

$$E\left[H_n^4\right] = E\left[u_i^4 u_j^4 f_i^4 f_j^4 \tilde{K}_{nij}^4 w_{nij}^4\right] = E\left[E(u_i^4|X_i, C_i)f_i^4 E(u_j^4|X_j, C_j)f_j^4 \tilde{K}_{nij}^4 w_{nij}^4\right] = O(h^{-3p}).$$

As $G_n(Z_i, Z_j) = u_i f_i u_j f_j E\left[\sigma_{C_0}^2(X_0)f^2(X_0)\tilde{K}_{ni0}\tilde{K}_{nj0}w_{ni0}w_{nj0}|Z_i, Z_j\right]$, we have

$$E[G_n^2] = \int \sigma_i^2 f_i^2 \sigma_j^2 f_j^2 E\left[\int \sigma_{C_0}^2 \tilde{K}_{ni0}\tilde{K}_{nj0}f^2(X_0)f_{C_0}(X_0)\,dX_0\, w_{ni0}w_{nj0}|Z_i, Z_j\right]^2 f_i f_j\, dX_i dX_j\, d\nu(C_i, C_j)$$

22

$$= h^{-p} \int \sigma_i^2 f_i^2 \sigma_{C_j}^2 (X_i + ht) f^2(X_i + ht)$$

$$E \left[\int \sigma_{C_0}^2 (X_i - hs) K_{a_{C_0}}(s) K_{a_{C_0}}(s + t) f^2(X_i - hs) f_{C_0}(X_i - hs) \, ds \, w_{ni0} w_{nj0} | Z_i, Z_j \right]^2$$

$$f_{C_i}(X_i) f_{C_j}(X_i + ht) \, dX_i dt \, d\nu \, (C_i, C_j)$$

$$= h^{-p} E \left[\sigma_C^8(X) f^8(X) f_C^3(X) \int (K_{a_C} * K_{a_C})^2 (t) \, dt \right] + o(h^{-p}) = O(h^{-p}),$$

where $\nu(C_i, C_j)$ denotes the distribution of (C_i, C_j). Thus (A.1) holds as $h \to 0$ and $nh^p \to \infty$. Q.E.D.

Proposition 2 $nh^{p/2}I_{1,3} = nh^{p/2}U_{1n} + \delta_n \sqrt{n} h^{p/2} O_p(1) + o_p(1)$, *where* $nh^{p/2}U_{1n}$ *is asymptotically centered normal with variance* $2E \left[\sigma_C^2(X) \sigma_{C'}^2(X) f^2(X) f_C^2(X) \int (K * K_{a_C})^2 (t) \, dt \right]$.

PROOF: $I_{1,3}$ is a U-statistic with kernel $H_n(Z_i, Z_j, Z_l) = u_i f_i u_l K_{njl} \tilde{K}_{nij} w_{nij}$. We now compute the corresponding ξ_s, $s = 1, 2, 3$.

(i) $\xi_2 = h^{-p} 2E \left[\sigma_C^2(X) \sigma_{C'}^2(X) f^2(X) f_C^2(X) \int (K * K_{a_C})^2 (t) \, dt \right] + o(h^{-p})$. Indeed we have $E(H_n | Z_i, Z_j) = 0$, $E(H_n | Z_i, Z_l) = u_i f_i u_l E \left(K_{njl} \tilde{K}_{nij} w_{nij} | Z_i, Z_l \right)$ and $E(H_n | Z_j, Z_l) = \delta_n u_l K_{njl} E \left(d_i f_i \tilde{K}_{nij} w_{nij} | Z_j \right)$.

Now $E \left[E^2(H_n | Z_i, Z_l) \right] = E \left[\sigma_i^2 f_i^2 \sigma_l^2 E^2 \left(K_{njl} \tilde{K}_{nij} w_{nij} | Z_i, Z_l \right) \right]$

$$= h^{-p} E \left[\sigma_C^2(X) \sigma_{C'}^2(X) f^2(X) f_C^2(X) \int (K * K_{a_C})^2 (t) \, dt \right] + o(h^{-p}),$$

$$E \left[E^2(H_n | Z_j, Z_l) \right] = \delta_n^2 E \left[\sigma_l^2 K_{njl}^2 d_j^2 f_j^2 f_{C_j}^2 (X_j) \right] = \delta_n^2 O(h^{-p}) E \left[u_l^2 \boldsymbol{K}_{njl} d_j^2 f_j^2 f_{C_j}^2 (X_j) \right] = O(\delta_n^2 h^{-p}),$$

$$E \left[E(H_n | Z_i, Z_l) E(H_n | Z_j, Z_l) \right]$$

$$= \delta_n^2 E \left[d_i f_i \sigma_l^2 K_{njl} E \left(K_{njl} \tilde{K}_{nij} w_{nij} | Z_i, Z_l \right) E \left(d_i f_i \tilde{K}_{nij} w_{nij} | Z_j \right) \right]$$

$$= \delta_n^2 O(h^{-p}) E \left[|d_i| f_i \sigma_l^2 \boldsymbol{K}_{njl} E \left(\tilde{\boldsymbol{K}}_{nij} w_{nij} | Z_i, Z_l \right) | E \left(d_i f_i \tilde{K}_{nij} w_{nij} | Z_j \right) | \right] = O(\delta_n^2 h^{-p}).$$

(ii) $\xi_1 = O(\delta_n^2)$. Indeed $E(H_n | Z_i) = E(H_n | Z_j) = 0$ and $E(H_n | Z_l) = \delta_n u_l E \left(d_i f_i K_{njl} \tilde{K}_{nij} w_{nij} | Z_l \right)$. Then $E \left[E^2(H_n | Z_l) \right] = \delta_n^2 E \left[u_l^2 E^2 \left(K_{njl} E(d_i f_i \tilde{K}_{nij} w_{nij} | Z_j, Z_l) | Z_l \right) \right] = O(\delta_n^2)$.

(iii) $E[H_n] = 0$.

(iv) $\xi_3 = O(h^{-2p})$, as $E \left[H_n^2 \right] = O(h^{-2p}) E \left[u_i^2 u_l^2 f_i^2 \boldsymbol{K}_{njl} \tilde{\boldsymbol{K}}_{nij} w_{nij}^2 \right] = O(h^{-2p})$.

Thus $E \left[nh^{p/2} (I_{1,3} - U_{1n}) \right]^2 = \delta_n^2 nh^p O(1)$, where $U_{1n} = (1/n^{(3)}) \sum_a [H_n(Z_i, Z_j, Z_l) - E(H_n | Z_l)]$. Similarly, $E \left[H_{n,2}^4 \right] = O(h^{-3p})$ and $E \left[G_n^2 \right] = O(h^{-p})$. The result then follows from Lemma 1. Q.E.D.

Proposition 3 $nh^{p/2}I_{1,5}$ *is asymptotically centered normal with variance* $2E \left[\sigma_C^2(X) \sigma_{C'}^2(X) \int g^4(X, t) \, dt \right]$, *where* $g^2(x, t) = \sum_c p_c f_c^2(x)(K * K * K_{a_c})^2(t)$.

PROOF: $I_{1,5}$ is a U-statistic with kernel $H_n(Z_i, Z_j, Z_k, Z_l) = u_k u_l K_{nik} K_{njl} \tilde{K}_{nij} w_{nij}$.

23

(i) $\xi_3 = O(h^{-2p})$. Indeed we have $E(H_n|Z_i, Z_j, Z_k) = E(H_n|Z_i, Z_j, Z_l) = 0$, and

$$
\begin{aligned}
E(H_n|Z_i, Z_k, Z_l) &= u_k u_l K_{nik} E\left(K_{njl} \tilde{K}_{nij} w_{nij}|Z_i, Z_l\right), \\
E(H_n|Z_j, Z_k, Z_l) &= u_k u_l K_{njl} E\left(K_{nik} \tilde{K}_{nij} w_{nij}|Z_j, Z_k\right).
\end{aligned}
$$

Then
$$
\begin{aligned}
E\left[E^2(H_n|Z_i, Z_k, Z_l)\right] &= E\left[u_k^2 u_l^2 K_{nik}^2 E\left(K_{njl}\tilde{K}_{nij}w_{nij}|Z_i, Z_l\right) E\left(K_{nj'l}\tilde{K}_{nij'}w_{nij'}|Z_i, Z_l\right)\right] \\
&= O(h^{-2p}) E\left[u_k^2 u_l^2 \boldsymbol{K}_{nik} E\left(\boldsymbol{K}_{njl}\tilde{\boldsymbol{K}}_{nij}w_{nij}|Z_i, Z_l\right) E\left(\tilde{\boldsymbol{K}}_{nij'}w_{nij'}|Z_i, Z_l\right)\right] \\
&= O(h^{-2p}) E\left[u_k^2 u_l^2 \boldsymbol{K}_{nik} \boldsymbol{K}_{njl}\tilde{\boldsymbol{K}}_{nij}w_{nij}f_{C_i}^2(X_i)\right] = O(h^{-2p}),
\end{aligned}
$$

Similarly $E\left[E^2(H_n|Z_j, Z_k, Z_l)\right]$ and $E\left[E(H_n|Z_i, Z_k, Z_l)E(H_n|Z_j, Z_k, Z_l)\right]$ are $O(h^{-2p})$.

(ii) $\xi_2 = h^{-p}2E\left[\sigma_C^2(X)\sigma_{C'}^2 \int g^4(X,t)\,dt\right] + o_p(h^{-p})$. Indeed we have $E(H_n|Z_i, Z_j) = E(H_n|Z_i, Z_k) = E(H_n|Z_i, Z_l) = E(H_n|Z_j, Z_k) = E(H_n|Z_j, Z_l) = 0$ and $E(H_n|Z_k, Z_l) = u_k u_l E\left(K_{nik}K_{njl}\tilde{K}_{nij}w_{nij}|Z_k, Z_l\right)$, so that $E\left[E^2(H_n|Z_k, Z_l)\right] = E\left[\sigma_k^2\sigma_l^2 E^2\left(K_{nik}K_{njl}\tilde{K}_{nij}w_{nij}|Z_k, Z_l\right)\right]$. Now

$$
\begin{aligned}
E\left(K_{nik}K_{njl}\tilde{K}_{nij}w_{nij}|Z_k, Z_l\right) &= E\left[w_{nij}\int K_{nik}K_{njl}\tilde{K}_{nij}f_{C_i}(X_i)f_{C_j}(X_j)\,dX_i dX_j|Z_k, Z_l\right] \\
&= E\left[w_{nij}f_{C_i}(X_k)f_{C_j}(X_l)\int K(u)K(v)h^{-p}K_{a_{C_i}}(u-v+h^{-1}(X_k-X_l))\,dudv|Z_k, Z_l\right]
\end{aligned}
$$

uniformly in (X_i, X_j) as $f_C(\cdot) \in \mathcal{U}^p$, $\forall C$. Therefore

$$
\begin{aligned}
&E\left[E^2(H_n|Z_k, Z_l)\right] \\
&= \int \sigma_k^2\sigma_l^2 \left[\int w_{nij}f_{C_i}(X_k)f_{C_j}(X_l)\int K(u)K(v)h^{-p}K_{a_{C_i}}(u-v+h^{-1}(X_k-X_l))\,dudv\,d\nu(C_i,C_j)\right]^2 \\
&\qquad\qquad\qquad\qquad\qquad f_{C_k}(X_k)f_{C_l}(X_l)\,dX_k dX_l\,d\nu(C_k,C_l) \\
&= h^{-p}\int \sigma_k^2\sigma_{C_l}^2(X_k-ht)\left[\int w_{nij}f_{C_i}(X_k)f_{C_i}(X_k-ht)\int K(u)K(v)K_{a_{C_i}}(u-v+t)\,dudv\,d\nu(C_i,C_j)\right]^2 \\
&\qquad\qquad\qquad\qquad\qquad f_{C_k}(X_k)f_{C_l}(X_k-ht)\,dX_k dt\,d\nu(C_k,C_l) \\
&= h^{-p}E\left[\sigma_C^2(X)\sigma_{C'}^2(X)\int g^4(X,s)\,ds\right] + o(h^{-p}),
\end{aligned}
$$

where $g^2(x,s) = \sum_c p_c f_c^2(x)(K*K*K_{a_c})^2(s)$.

(iii) $\xi_1 = 0$, as $E(H_n|Z_i) = E(H_n|Z_j) = E(H_n|Z_k) = E(H_n|Z_l) = 0$.

(iv) $E[H_n] = 0$.

(v) $\xi_4 = O(h^{-3p})$, as $E[H_n^2] = E\left[u_k^2 u_l^2 K_{nik}^2 K_{njl}^2 \tilde{K}_{nij}^2 w_{nij}^2\right] = O(h^{-3p})E\left[u_k^2 u_l^2 \boldsymbol{K}_{nik}\boldsymbol{K}_{njl}\tilde{\boldsymbol{K}}_{nij}w_{nij}^2\right] = O(h^{-3p})$.

Similarly $E\left[H_{n,2}^4\right] = O(h^{-3p})$ and $E\left[G_n^2\right] = O(h^{-p})$. The result follows from Lemma 1. Q.E.D.

Proposition 4 $nh^{p/2}(U_{0n} - 2U_{1n} + I_{1,5}) \overset{d}{\longrightarrow} N(0, \omega^2)$.

PROOF: To show this, we apply the Cramer-Wold device. We compute the covariances between U_{0n}, U_{1n} and $I_{1,5}$.

COVARIANCE BETWEEN U_{0n} AND U_{1n}: $\xi_{2,2}$ is determined by

$$E\left[u_i f_i u_l f_l \tilde{K}_{nil} w_{nil} E\left(u_i f_i u_l K_{njl} \tilde{K}_{nij} w_{nij}|Z_i, Z_l\right)\right] = E\left[\sigma_i^2 f_i^2 \sigma_l^2 f_l \tilde{K}_{nil} w_{nil} E\left(K_{njl}\tilde{K}_{nij}w_{nij}|Z_i, Z_l\right)\right]$$

$$= h^{-p}E\left[\sigma_C^2(X)\sigma_{C'}^2(X)w_{CC'}f^3(X)f_C(X)\int K_{a_C}(t)(K*K_{a_C})(t)\,dt\right] + o(h^{-p}).$$

COVARIANCE BETWEEN U_{0n} AND $I_{1,5}$: $\xi_{2,2}$ is determined by

$$E\left[u_k f_k u_l f_l \tilde{K}_{nkl} w_{nkl} E\left(u_k u_l K_{nik} K_{njl} \tilde{K}_{nij} w_{nij}|Z_k, Z_l\right)\right]$$

$$= E\left[\sigma_k^2 f_k \sigma_l^2 f_l \tilde{K}_{nkl} w_{nkl} E\left(K_{nik} K_{njl} \tilde{K}_{nij} w_{nij}|Z_k, Z_l\right)\right]$$

$$= h^{-p}E\left[\sigma_C^2(X)\sigma_{C'}^2(X)w_{CC'}f^2(X)f_{C''}^2(X)\int K_{a_C}(t)(K*K*K_{a_{C''}})(t)\,dt\right] + o(h^{-p}).$$

COVARIANCE BETWEEN U_{1n} AND $I_{1,5}$: $\xi_{3,3} = O(h^{-p})$. and $\xi_{2,2}$ is determined by

$$E\left[E\left(u_k f_k u_l K_{njl} \tilde{K}_{nkj} w_{nkj}|Z_k, Z_l\right) E\left(u_k u_l K_{nik} K_{nj'l} \tilde{K}_{nij'} w_{nij'}|Z_k, Z_l\right)\right]$$

$$= E\left[\sigma_k^2 f_k \sigma_l^2 E\left(K_{njl}\tilde{K}_{nkj}w_{nkj}|Z_k, Z_l\right) E\left(K_{nik}K_{nj'l}\tilde{K}_{nij'}w_{nij'}|Z_k, Z_l\right)\right]$$

$$= h^{-p}E\left[\sigma_C^2(X)\sigma_{C'}^2(X)f(X)f_C(X)f_{C''}^2(X)\int(K*K_{a_C})(t)(K*K*K_{a_C''})(t)\,dt\right] + o(h^{-p}).$$

Hence $nh^{p/2}\left[U_{0n} - 2U_{1n} + I_{1,5}\right] \xrightarrow{d} N(0, \omega^2)$, where

$$\omega^2 = 2E\left[\sigma_C^2(X)\sigma_{C'}^2(X)w_{CC'}^2 f^4(X)\int K_{a_C}^2(t)\,dt\right] + 2E\left[\sigma_C^2(X)\sigma_{C'}^2(X)\int g^4(X,t)\,dt\right]$$

$$+ 8E\left[\sigma_C^2(X)\sigma_{C'}^2(X)f^2(X)f_C^2(X)\int (K*K_{a_C})^2(t)\,dt\right]$$

$$- 8E\left[\sigma_C^2(X)\sigma_{C'}^2(X)w_{CC'}f^3(X)f_C(X)\int K_{a_C}(t)(K*K_{a_C})(t)\,dt\right]$$

$$+ 4E\left[\sigma_C^2(X)\sigma_{C'}^2(X)w_{CC'}f^2(X)f_{C''}^2(X)\int K_{a_C}(t)(K*K*K_{a_{C''}})(t)\,dt\right]$$

$$- 8E\left[\sigma_C^2(X)\sigma_{C'}^2(X)f(X)f_C(X)f_{C''}^2(X)\int (K*K_{a_C})(t)(K*K*K_{a_C''})(t)\,dt\right]$$

$$= 2E\left[\sigma_C^2(X)\sigma_{C'}^2 f^4(X)E_{CC'}(X)\right],$$

with $E_{CC'}(X) = \int\left[K_{a_C}(t)w_{CC'} - 2(K*K_{a_C})(t)\dfrac{f_C(X)}{f(X)} + \dfrac{g^2(X,t)}{f^2(X)}\right]^2 dt.$ \hfill Q.E.D.

Proposition 5 $nh^{p/2}I_{1,6} = \delta_n^2 nh^{p/2}o_p(1) + o_p(1).$

Proposition 6 $nh^{p/2}I_{2,1} = \delta_n\sqrt{n}h^{p/2}o_p(1) + \delta_n nh^{p/2}h^{(m+q)}O_p(1) + o_p(1).$

Proposition 7 $nh^{p/2}I_{2,3} = o_p(1).$

25

Proposition 8 $nh^{p/2}I_3 = nh^{p/2}O_p(h^{2(m+q)}) + o_p(1)$.

PROOFS OF PROPOSITIONS 5–8: They follow the same lines as the ones of Proposition 4 to 7 by Lavergne and Vuong (2000), provided that K_{nij} is replaced by $\tilde{K}_{nij}$, $L(\cdot)$ by $K(\cdot)$ and g by h.

Proposition 9 $nh^{p/2}I_{1,1} = \delta_n^2 nh^{p/2}o_p(1) + \delta_n\sqrt{n}h^{p/2}o_p(1) + o_p(1)$.

PROOF: Let $\Delta f_i^j \equiv (\widehat{f}_i^j - f_i)$. We have $I_{1,1} = (1/n^{(2)}) \sum_a u_i \Delta f_i^j u_j f_j \tilde{K}_{nij} w_{nij}$ so that

$$E(I_{1,1}^2) = \left(\frac{1}{n^{(2)}}\right)^2 \left[\sum_a u_i \Delta f_i^j u_j f_j \tilde{K}_{nij} w_{nij}\right]\left[\sum_a u_{i'} \Delta f_{i'}^{j'} u_{j'} f_{j'} K_{ni'j'} w_{ni'j'}\right],$$

where the sums are taken over all arrangements of different indices In what follows, $\tilde{K}_{nij} \equiv \tilde{K}_{nij} w_{nij}$.
(i) All indices are different: $n^{(4)}$ terms.

$$E\left[u_i \Delta f_i^j u_j f_j \tilde{K}_{nij} u_{i'} \Delta f_{i'}^{j'} u_{j'} f_{j'} \tilde{K}_{ni'j'}\right] = \delta_n^4 E\left[\Delta f_i^j f_j \Delta f_{i'}^{j'} f_{j'} d_i d_j d_{i'} d_{j'} \tilde{K}_{nij} \tilde{K}_{ni'j'}\right]$$

$$= \delta_n^4 E\left[f_j f_{j'} d_i d_j d_{i'} d_{j'} \tilde{K}_{nij} \tilde{K}_{ni'j'} E\left(\Delta f_i^j \Delta f_{i'}^{j'} | Z_i, Z_j, Z_{i'}, Z_{j'}\right)\right]$$

$$= \delta_n^4 \lambda_n E\left|f_j f_{j'} d_i d_j d_{i'} d_{j'} \tilde{K}_{nij} \tilde{K}_{ni'j'}\right| = O(\delta_n^4 \lambda_n),$$

where $\lambda_n = E\left[\Delta^2 f_i^j | Z_i, Z_j, Z_{i'}, Z_{j'}\right]$.
(ii) One index is common to $\{i, j\}$ and $\{i', j'\}$: $4n^{(3)}$ terms.

$(i' = i)$ $\quad E\left[u_i^2 \Delta f_i^j u_j f_j \tilde{K}_{nij} \Delta f_i^{j'} u_{j'} f_{j'} \tilde{K}_{nij'}\right] = \delta_n^2 \lambda_n E\left|f_j f_{j'} u_i^2 d_j d_{j'} \tilde{K}_{nij} \tilde{K}_{nij'}\right| = O(\delta_n^2 \lambda_n)$,

$(j' = j)$ $\quad E\left[u_i \Delta f_i^j u_j^2 f_j^2 \tilde{K}_{nij} u_{i'} \Delta f_{i'}^j \tilde{K}_{ni'j}\right] = \delta_n^2 \lambda_n E\left|f_j^2 d_i u_j^2 d_{i'} \tilde{K}_{nij} \tilde{K}_{ni'j}\right| = O(\delta_n^2 \lambda_n)$,

$(i' = j)$ $\quad E\left[u_i \Delta f_i^j u_j^2 f_j \tilde{K}_{nij} \Delta f_j^{j'} u_{j'} f_{j'} \tilde{K}_{njj'}\right] = \delta_n^2 \lambda_n E\left|f_j f_{j'} d_i u_j^2 d_{j'} \tilde{K}_{nij} \tilde{K}_{njj'}\right| = O(\delta_n^2 \lambda_n)$.

The case $j' = i$ is similar to $i' = j$.
(iii) Two indices in common to $\{i, j\}$ and $\{i', j'\}$: $2n^{(2)}$ terms. We have $E\left[u_i^2 u_j^2 \left(\Delta f_i^j\right)^2 f_j^2 \tilde{K}_{nij}^2\right] = O(\lambda_n/h^p)$ and $E\left[u_i^2 u_j^2 \Delta f_i^j \Delta f_j^i f_i f_j \tilde{K}_{nij}^2\right] = O(\lambda_n/h^p)$.
Therefore, $E\left(nh^{p/2}I_{1,1}\right)^2 = \delta_n^4 n^2 h^p O(\lambda_n) + \delta_n^2 nh^p O(\lambda_n) + O(\lambda_n)$. The proposition then follows from $\lambda_n = o(1)$ uniformly, see Lemma 3. Q.E.D.

Proposition 10 $nh^{p/2}I_{1,2} = \delta_n^2 nh^{p/2}o_p(1) + \delta_n\sqrt{n}h^{p/2}o_p(1) + o_p(1)$.

Proposition 11 $nh^{p/2}I_{1,4} = \delta_n^2 nh^{p/2}o_p(1) + \delta_n\sqrt{n}h^{p/2}o_p(1) + o_p(1)$.

Proposition 12 $nh^{p/2}I_{2,2} = \delta_n^2 nh^{p/2}o_p(1) + \delta_n\sqrt{n}h^{p/2}o_p(1) + o_p(1)$.

PROOFS OF PROPOSITIONS 10—12: They are very similar to Proposition 11's proof and are not reported. The two last proofs use the second part of Lemma 3.

A.2 Proof of Theorem 2

The consistency of both forms of ω_n^2 can be proven using similar arguments as in the proof of Theorem 1 and as in Part (i) of Theorem 1 of Lavergne and Vuong (1996) for the treatment of the trimming parameter. In particular, $(b_n \sqrt{n} h_n^p)^{-1} = o(1)$ and $b_n^{-1} h^{m+q} = o(1)$ ensures that $b_n^{-1} \sup_{x \in I\!\!R^p} |f_n(x) - f(x)| = o_p(1)$. Hence, $f_n(\cdot)$ can be replaced by $f(\cdot)$ in E_{nij} without changing the probability limit of ω_n^2.

A.3 Proof of Theorem 3

We use similar arguments to that in Mammen (1992), who deals with testing equality of density functions. However, the technical analysis is long and tedious, so that we only provide a brief sketch of the proof in the special case where $a_c = 1$, $c = 1, \ldots, \bar{C}$.

Let us explicit the dependence of V_n on the bandwidth. Under the null hypothesis, we shall show that $n\hat{h}^{p/2}V_n(\hat{h}) - nh^{p/2}V_n(h) = o_p(1)$. For this equality to hold, we need to show tightness of the process $n(\nu h)^{p/2}V_n(\nu h)$ for $\nu \in [B_1, B_2]$, with $0 < B_1 < 1 < B_2 < \infty$. Now $n(\nu h)^{p/2}V_n(\nu h)$ converges in distribution to the same limit for any ν. Hence it is sufficient to show that for $\nu_1, \nu_2 \in [B_1, B_2]$,

$$E\left[n(\nu_1 h)^{p/2}V_n(\nu_1 h) - n(\nu_2 h)^{p/2}V_n(\nu_2 h)\right]^2 = O\left[(\nu_1 - \nu_2)^2\right],$$

see Billingsley (1968). Now because $\left[\sum_{i=1}^{11} \lambda_i\right]^2 \leq 11 \sum_{i=1}^{n} \lambda_i^2$, it is sufficient to show an analogous result for each of the eleven terms of Propositions 1 to 3 and 5 to 12. For instance, we can show that

$$E\left[n(\nu_1 h)^{p/2}V_{0n}(\nu_1 h) - n(\nu_2 h)^{p/2}V_{0n}(\nu_2 h)\right]^2 = O(1) \int \left[\nu_1^{-p/2}K\left(x/\nu_1\right) - \nu_2^{-p/2}K\left(x/\nu_2\right)\right] dx$$

is $O\left[(\nu_1 - \nu_2)^2\right]$ by a Taylor expansion of $\nu_2^{-p/2}K\left[x/\nu_2\right]$ around ν_1.

Under the alternative hypothesis, it is sufficient to show that for $\nu_1, \nu_2 \in [B_1, B_2]$,

$$E\left[V_n(\nu_1 h) - V_n(\nu_2 h)\right]^2 = O\left[(\nu_1 - \nu_2)^2\right].$$

An analogous result for ω_n^2 then implies the desired result.

A.4 Proof of Theorem 4

The proof of Theorem 4 is analogous to the proof of Theorem 1. To deal with V_{0n}, $I_{1,3}$ and $I_{1,5}$, we use a straightforward generalization of Lemma 1, which accounts for the fact that observations may not be identically distributed across subsamples, although they are independent. This result is not formally stated, but one can easily check that it holds by examining the proofs of Lemma B.4 of Fan and Li (1996) and of Theorem 1 of Hall (1984): the latter proof relies on a martingale central limit theorem that still applies, see Hall and Heyde (1980, Chapter 3).

27

A.5 Panel data

Under our assumptions, $u_{it} \equiv Y_{it} - r(X_{it})$ is independent of $\mathcal{Z}_{it-1} = \{Y_{i1}, \ldots, Y_{it-1}, X_{i1}, \ldots, X_{it-1}\}$ conditionally on $\{X_{it}, l\}$ and $E[u_{it}|\mathcal{Z}_{it-1}, X_{it}, l] = R_t(X_{it}) - r(X_{it})$ is zero under H_0. Also by assumption $E[u_{it}u_{jt'}|\mathcal{Z}_{it-1}, X_{it}, \mathcal{Z}_{jt'-1}, X_{jt'}, l] = 0$, $\forall i, j, t \neq t'$. Then $U_{0,n}$, $U_{1,n}$ and $I_{1,5}$ are degenerate U-statistics under H_0 and a generalization of Lemma 1 can be applied. The remaining terms are dealt with as in the proof of Theorem 1, using Lemma 4 for $I_{1,1}$, $I_{1,2}$, $I_{1,4}$ and $I_{2,2}$.

Appendix B: Technical lemmas

Lemma 2 *For $l(\cdot) \in \mathcal{U}^p$, $\sup_{x \in I\!\!R^p} \left| \int l(X) \frac{1}{h^p} K\left(\frac{x-X}{h} \right) dX - l(x) \right| \to 0$ when $h \to 0$.*

PROOF: This result comes from the well-known Bochner lemma. Q.E.D.

Lemma 3 $E\left[\left(\widehat{f}_i^j - f_i \right)^2 | Z_i, Z_j, Z_{i'}, Z_{j'} \right] = o(1)$ *and* $E\left[\left(\widehat{f}_i^{j,l} - f_i \right)^2 | Z_i, Z_j, Z_l, Z_{i'}, Z_{j'}, Z_{l'} \right] = o(1)$ *uniformly in their arguments, if $f_c(X) \in \mathcal{U}^p$ $\forall c$ and $nh^p \to \infty$.*

PROOF: The first expectation is

$$E\left[\left(\widehat{f}_i^j - E(\widehat{f}_i^j | Z_i, Z_j, Z_{i'}, Z_{j'}) \right)^2 | Z_i, Z_j, Z_{i'}, Z_{j'} \right] + \left[E\left(\widehat{f}_i^j | Z_i, Z_j, Z_{i'}, Z_{j'} \right) - f_i \right]^2 .$$

Because $\widehat{f}_i^j - E(\widehat{f}_i^j | Z_i, Z_j, Z_{i'}, Z_{j'}) = (n-2)^{-1} \sum_{k \notin \{i,j,i',j'\}} \left(K_{nik} - E(K_{nik}|Z_i) \right)$, whose summands are independent with zero mean conditional on Z_i,

$$\begin{aligned}
E\left[\left(\widehat{f}_i^j - E(\widehat{f}_i^j | Z_i, Z_j, Z_{i'}, Z_{j'}) \right)^2 | Z_i, Z_j, Z_{i'}, Z_{j'} \right] &\leq (n-2)^{-2} \sum_{k \notin \{i,j,i',j'\}} E\left[(K_{nik} - E(K_{nik}|Z_i))^2 | Z_i \right] \\
&\leq (n-2)^{-2} \sum_{k \notin \{i,j,i',j'\}} E\left[K_{nik}^2 | Z_i \right] = O(nh^p)^{-1}.
\end{aligned}$$

$$\begin{aligned}
\text{Now} \qquad \left[E\left(\widehat{f}_i^j | Z_i, Z_j, Z_{i'}, Z_{j'} \right) - f_i \right]^2 &= \left[\frac{1}{n-2} (K_{nii'} + K_{nij'} - f_i) + \frac{n-4}{n-2} E(K_{nik} - f_i|Z_i) \right]^2 \\
&\leq \left[O(n^{-1}h^{-p}) + O(n^{-1}) + o(1) \right]^2 = o(1),
\end{aligned}$$

where we use $f_i = \sum_c p_c f_c(X_i)$. The proof for the second part is similar and is not reported. Q.E.D.

Lemma 4 *The result of Lemma 3 holds for panel data.*

PROOF: The proof follows Lemma 3's proof, with the difference that

$$E\left[\left(\widehat{f}_i^j - E(\widehat{f}_i^j | Z_i, Z_j, Z_{i'}, Z_{j'}) \right)^2 | Z_i, Z_j, Z_{i'}, Z_{j'} \right]$$

$$\leq (n-2)^{-2} \sum_{k,k' \notin \{i,j,i',j'\}} E\left[(K_{nik} - E(K_{nik}|Z_i)) (K_{nik'} - E(K_{nik'}|Z_i)) | Z_i \right].$$

By conditioning upon (Z_i, Z_k), all terms such that k and k' correspond to different individuals vanishes. At most nT terms correspond to same individuals and all are $O(h^p)^{-1}$ by the Cauchy-Schwartz inequality. Thus the right-hand side is $O(nh^p)^{-1}$ and Lemma 4 follows. Q.E.D.

Appendix C

We thereafter give the integrals that enter ω^2 when $K(\cdot)$ is uniform on $[-1/2, 1/2]$.

$$\int K_a^2(s)\,ds \quad = \quad \frac{1}{a}$$

$$\int K_a(s)(K * K_a)(s)\,ds \quad = \quad \begin{cases} 1 & \text{if } a < 1/2 \\ \frac{-1+4a}{4a^2} & \text{if } a \geq 1/2 \end{cases}$$

$$\int (K * K_a)^2(s)\,ds \quad = \quad \int K_a(s)(K * K * K_a)(s)\,ds$$

$$= \quad \begin{cases} 1 - \frac{a}{3} & \text{if } a < 1 \\ \frac{-1+3a}{3a^2} & \text{if } a \geq 1 \end{cases}$$

$$\int (K * K_a)(s)(K * K * K_a)(s)\,ds \quad = \quad \begin{cases} \frac{3}{4} - \frac{a^2}{6} & \text{if } a < 1/2 \\ \frac{3-24a+216a^2-96a^3+16a^4}{192a^2} & \text{if } 1/2 \leq a \leq 3/2 \\ \frac{-13+32a^2}{32a^2} & \text{if } a > 3/2 \end{cases}$$

$$\int (K * K * K_a)^2(s)\,ds \quad = \quad \begin{cases} \frac{40-10a^2+3a^3}{60} & \text{if } a < 1 \\ \frac{4-20a+80a^2-40a^3+10a^4-a^5}{60a^2} & \text{if } 1 \leq a \leq 2 \\ \frac{-7+15a^2}{15a^2} & \text{if } a > 2. \end{cases}$$

29

REFERENCES

Aït-Sahalia, Y., P. Bickel and T. M. Stoker, 1994, Goodness-of-fit tests for regression using kernel methods, University of Chicago.

Alvarez, J. and M. Arellano, 1998, The time series and cross-section asymptotics of dynamic panel data estimators, CEMFI.

Baltagi, B.H., J. Hidalgo and Q. Li, 1996, A nonparametric test for poolability using panel data Journal of Econometrics , 75, 345–367.

Bierens, H.J., 1987, Kernel estimators of regression functions, in: T. Bewley, ed., Advances in Econometrics (Cambridge University Press, Cambridge) 99–144.

Billingsley, P., 1968, Convergence of Probability Measures (John Wiley & Sons, New-York).

Chamberlain, G., 1984, Panel data, in: Z. Griliches and M.D. Intriligator, eds., Handbook of Econometrics, Vol. 2 (North-Holland: Amsterdam) Chap. 5, 285–330.

Chow, G.C., 1960, Tests of equality between sets of coefficients in two linear regressions, Econometrica 28(3), 591–605.

Delgado, M.A., 1993, Testing the equality of nonparametric regression curves, Statistics and Probability Letters 17, 199–204.

Delgado, M.A., M.A. Dominguez and P. Lavergne, 1998, Asymptotic and bootstrap specification tests of nonlinear in variables econometric models, Universidad Carlos III de Madrid.

Delgado, M.A. and J. Mora, 1995, Nonparametric and semiparametric estimation with discrete regressors, Econometrica 63(6), 1477–1484.

Fan, J. and S.-K. Lin, 1998, Test of significance when data are curves, Journal of the American Statistical Association 93 (443), 1007–1021.

Fan, Y. and Q. Li, 1996, Consistent model specification tests: omitted variables and semiparametric functional forms, Econometrica 64 (4), 865–890.

Gozalo, P.L., 1995, Nonparametric specification testing with $\sqrt{n}$- local power and bootstrap critical values, Brown University.

Guerre, E. and P. Lavergne, 1999, Minimax rates for nonparametric specification testing in regression models, LSTA, Université Paris 6.

Hall, P., 1984, Central limit theorem for integrated square error of multivariate nonparametric density estimators, Journal of Multivariate Analysis 14, 1–16.

Hall, P. and J.D. Hart, 1990, Bootstrap test for difference bewteen means in nonparametric regression, Journal of the American Statistical Association 85, 1039–1049.

Hall, P. and C.C. Heyde, 1980, Martingale Limit Theory and its Application (Academic Press: New-York).

Hall, P., C. Huber and P.L. Speckman, 1997, Covariance-matched one-sided tests for the difference between functional means, Journal of the American Statistical Association 92(439), 1074–1083.

Härdle, W. and E. Mammen, 1993, Comparing nonparametric versus parametric regression fits, Annals of Statistics 21 (4), 1926–1947.

Härdle, W. and J.S. Marron, 1990, Semiparametric comparison of regression curves, Annals of Statistics 18 (1), 63–89.

Hart, J.D., 1997, Nonparametric Smoothing and Lack-of-Fit Tests (Springer Verlag: New-York).

Hoeffding, W., 1948, A class of statistics with asymptotically normal distribution, Annals of Mathematical Statistics 15, 293–325.

King, E.C., J.D. Hart and T.E. Wehrly, 1991, Testing the equality of two regression curves using linear smoothers, Statistics and Probability Letters 12, 239–247.

Koul, H.L. and A. Schick, 1997, Testing for the equality of two nonparametric regression curves, Journal of Statistical Planning and Inference 65, 293–314.

Kulasekera, K.B., 1995, Comparison of regression curves using quasi-residuals, Journal of the American Statistical Association 90(431), 1085–1093.

Lavergne, P. and Q.H. Vuong, 1996, Nonparametric selection of regressors: the nonnested case, Econometrica 64(1), 207–219.

Lavergne, P. and Q.H. Vuong, 2000, Nonparametric significance testing, Econometric Theory , to appear.

Li, Q., 1999, Consistent model specification tests for time series econometrics models, Journal of Econometrics 92(1), 101–147.

Mammen, E., 1992, When Does Bootstrap Work? Asymptotic Results and Simulations, Lecture Notes in Statistics, 77 (Springer-Verlag: New-York).

Munk, A. and H. Dette, 1998, Nonparametric comparison of several regression functions: exact and asymptotic theory, Annals of Statistics 26(6), 2339–2368.

Powell, J.L., H. Stock and T.M. Stoker, 1989, Semiparametric estimation of index coefficients, Econometrica 57, 1403–1430.

Quade, D., 1982, Nonparametric analysis of covariance, Biometrics 38, 597–611.

Pinkse, C.A.P and P.M. Robinson, 1995, Pooling nonparametric estimates of regression funtions with a similar shape, in: G.S. Maddala, P.C.B. Phillips and T.N. Srinivasan, eds., Advances in Econometrics and Quantitative Economics: Essays in Honor of Professor C.R. Rao (B. Blackwell: Oxford) 172–197.

Robinson, P., 1988, Root-N consistent semiparametric regression, Econometrica 56, 931–954.

Scheffé, H., 1959, The Analysis of Variance (John Wiley & Sons: New-York).

Yatchew, A., 1999, An elementary nonparametric differencing test of equality of regression functions, Economic Letters 62, 271–278.

Young, S. G. and A. W. Bowman, 1995, Non-parametric analysis of covariance, Biometrics 51, 920–931.

Zheng, J.X., 1996, A consistent test of functional form via nonparametric estimation techniques, Journal of Econometrics 75 (2), 263–289.

Table 1: Null and Linear Alternatives — $h \propto n^{-1/5}$

n	a	DGP_0		DGP_1		DGP_2		DGP_3	
100	0.5	0.001	(0.854)	0.271	(0.935)	0.966	(1.116)	2.725	(1.334)
		4.3%	7.6%	8.8%	14.3%	25.3%	36.4%	78.6%	86.2%
		1.546	1.111	9.8%	17.9%	28.1%	42.2%	80.6%	88.2%
	1.0	0.061	(0.846)	0.490	(1.048)	1.559	(1.404)	4.129	(1.792)
		4.4%	8.3%	13.1%	19.7%	43.1%	53.6%	91.3%	94.4%
		1.592	1.170	14.0%	22.5%	44.3%	57.0%	92.1%	95.2%
	1.5	0.036	(0.876)	0.582	(1.158)	1.949	(1.641)	5.203	(2.159)
		5.6%	8.9%	16.8%	23.7%	51.6%	61.7%	95.6%	97.5%
		1.696	1.176	15.9%	26.0%	50.3%	64.0%	95.3%	97.8%
	2.0	0.034	(0.875)	0.675	(1.230)	2.280	(1.826)	6.093	(2.469)
		5.0%	8.6%	20.3%	26.0%	58.6%	66.6%	97.4%	98.4%
		1.642	1.171	20.3%	28.8%	58.6%	68.8%	97.4%	98.7%
	Chow test	4.4%	9.6%	51.2%	64.6%	98.7%	99.4%	100.0%	100.0%
200	0.5	0.031	(0.921)	0.595	(1.086)	2.063	(1.427)	5.771	(1.816)
		5.1%	9.5%	16.3%	23.2%	59.8%	69.7%	98.7%	99.5%
		1.655	1.241	16.2%	24.3%	59.7%	71.0%	98.6%	99.5%
	1.0	0.042	(0.899)	0.932	(1.265)	3.200	(1.870)	8.667	(2.504)
		5.0%	8.8%	24.5%	34.1%	79.8%	85.4%	99.9%	100.0%
		1.645	1.224	24.5%	36.0%	79.8%	86.0%	99.9%	100.0%
	1.5	0.053	(0.910)	1.175	(1.415)	4.031	(2.213)	10.838	(3.035)
		5.9%	9.7%	32.5%	41.3%	86.4%	90.3%	100.0%	100.0%
		1.725	1.262	30.1%	41.9%	85.2%	90.5%	100.0%	100.0%
	2.0	0.042	(0.900)	1.354	(1.529)	4.694	(2.493)	12.624	(3.486)
		5.6%	9.3%	36.0%	44.9%	89.8%	92.2%	100.0%	100.0%
		1.718	1.237	34.3%	46.0%	89.3%	92.7%	100.0%	100.0%
	Chow test	5.3%	9.8%	92.8%	96.6%	100.0%	100.0%	100.0%	100.0%

Each cell contains the empirical mean of the test statistic with its standard deviation in parentheses on the first line, empirical rejection percentages at 5% and 10% nominal levels on the second line, and size-corrected empirical rejection percentages at 5% and 10% levels on the third line. For the null hypothesis, i.e. DGP_0, empirical critical values for 5% and 10% levels are on the third line.

Table 2: Sinus Alternatives — $h \propto n^{-1/5}$

n	a	DGP_4		DGP_5		DGP_6		DGP_7	
100	0.5	0.981	(1.069)	0.877	(1.136)	0.956	(1.161)	0.934	(1.142)
		25.6%	36.2%	23.9%	32.8 %	27.2%	36.2%	25.8%	35.0%
		28.3%	42.1%	26.4%	38.2 %	29.2%	40.6%	28.2%	40.7%
	1.0	1.241	(1.171)	1.353	(1.375)	1.529	(1.467)	1.510	(1.444)
		32.9%	45.2%	36.2%	46.8%	40.7%	51.0%	41.9%	50.4%
		34.2%	48.8%	37.6%	49.7%	42.0%	54.2%	42.9%	53.9%
	1.5	0.943	(1.109)	1.535	(1.549)	1.853	(1.678)	1.861	(1.665)
		23.1%	32.5%	42.0%	50.3 %	48.3%	57.5%	48.5%	57.7%
		22.0%	36.2%	40.6%	52.9 %	47.2%	59.7%	47.5%	60.2%
	2.0	0.346	(1.018)	1.575	(1.588)	2.085	(1.827)	2.142	(1.835)
		10.1%	15.2%	41.3%	51.2 %	52.9%	61.6%	54.1%	63.0%
		10.1%	17.1%	41.4%	54.3 %	53.0%	64.5%	54.3%	65.7%
	Chow test	4.8%	9.5%	14.8%	21.2%	74.9%	82.3%	83.8%	90.9%
200	0.5	2.037	(1.389)	1.933	(1.427)	2.078	(1.452)	2.014	(1.429)
		59.5%	69.8%	55.2%	64.5%	58.9%	68.5%	57.2%	68.2%
		59.3%	70.7%	55.0%	65.8%	58.7%	69.5%	56.9%	68.9%
	1.0	2.626	(1.577)	2.870	(1.825)	3.182	(1.917)	3.112	(1.882)
		71.2%	79.4%	72.4%	79.0%	77.5%	84.0%	77.4%	84.2%
		71.1%	80.4%	72.4%	80.3%	77.5%	84.9%	77.4%	84.9%
	1.5	2.347	(1.498)	3.395	(2.082)	3.934	(2.242)	3.892	(2.216)
		64.7%	74.4%	79.0%	84.3%	85.3%	89.6%	84.2%	88.8%
		62.6%	74.7%	77.5%	84.5%	84.1%	89.8%	83.4%	89.2%
	2.0	1.452	(1.299)	3.588	(2.192)	4.440	(2.491)	4.470	(2.484)
		38.8%	48.4%	79.5%	85.8%	87.3%	91.4%	87.9%	91.5%
		36.4%	50.2%	78.5%	86.6%	86.3%	92.0%	86.8%	91.9%
	Chow test	5.5%	10.1%	13.4%	21.0%	93.3%	95.9%	99.6%	100.0%

Each cell contains the empirical mean of the test statistic with its standard deviation in parentheses on the first line, empirical rejection percentages at 5% and 10% nominal levels on the second line, and size-corrected empirical rejection percentages at 5% and 10% levels on the third line.

Table 3: Null and Linear Alternatives — $h \propto n^{-\lambda}$, $a = 1$

n	λ	DGP_0		DGP_1		DGP_2		DGP_3	
100	1/3	0.013	(0.762)	0.252	(0.852)	0.874	(1.037)	2.398	(1.274)
		3.2%	6.5%	6.7%	11.6%	21.1%	32.0%	72.1%	80.3%
		1.420	1.027	9.9%	17.6%	27.8%	40.6%	77.5%	85.9%
	1/4	0.019	(0.808)	0.365	(0.953)	1.248	(1.239)	3.394	(1.567)
		3.8%	7.5%	10.7%	16.0%	33.8%	44.9%	86.0%	90.5%
		1.517	1.072	12.2%	20.2%	37.6%	51.3%	88.1%	92.8%
	1/5	0.061	(0.846)	0.490	(1.048)	1.559	(1.404)	4.129	(1.792)
		4.4%	8.3%	13.1%	19.7%	43.1%	53.6%	91.3%	94.4%
		1.592	1.170	14.0%	22.5%	44.3%	57.0%	92.1%	95.2%
	1/6	0.094	(0.880)	0.578	(1.112)	1.777	(1.512)	4.653	(1.937)
		5.7%	9.2%	16.2%	22.9%	48.4%	59.0%	94.1%	96.2%
		1.744	1.221	14.2%	24.4%	46.1%	60.5%	93.4%	96.5%
200	1/3	0.029	(0.829)	0.546	(1.009)	1.881	(1.353)	5.130	(1.741)
		4.1%	7.4%	13.2%	21.2%	54.1%	64.4%	97.9%	98.9%
		1.486	1.096	16.8%	26.5%	58.8%	70.7%	98.5%	99.3%
	1/4	0.014	(0.899)	0.745	(1.189)	2.630	(1.684)	7.194	(2.197)
		4.7%	8.2%	19.7%	28.1%	69.6%	78.2%	99.6%	99.8%
		1.622	1.138	20.2%	31.8%	70.1%	80.7%	99.6%	99.9%
	1/5	0.042	(0.899)	0.932	(1.265)	3.200	(1.870)	8.667	(2.504)
		5.0%	8.8%	24.5%	34.1%	79.8%	85.4%	99.9%	100.0%
		1.645	1.224	24.5%	36.0%	79.8%	86.0%	99.9%	100.0%
	1/6	0.104	(0.913)	1.112	(1.332)	3.654	(2.022)	9.772	(2.730)
		6.0%	9.6%	30.5%	40.4%	84.2%	89.1%	99.9%	100.0%
		1.777	1.258	28.1%	40.7%	82.0%	89.2%	99.9%	100.0%

Each cell contains the empirical mean of the test statistic with its standard deviation in parentheses on the first line, empirical rejection percentages at 5% and 10% nominal levels on the second line, and size-corrected empirical rejection percentages at 5% and 10% levels on the third line. For the null hypothesis, i.e. DGP_0, empirical critical values for 5% and 10% levels are on the third line.

Table 4: Sinus Alternatives — $h \propto n^{-\lambda}$, $a = 1$

n	λ	DGP_4		DGP_5		DGP_6		DGP_7	
100	1/3	0.822	(1.006)	0.826	(1.080)	0.886	(1.098)	0.860	(1.075)
		19.6%	29.4%	21.5%	30.6 %	23.8%	33.1%	22.2%	31.3%
		25.3%	37.7%	26.6%	38.1 %	29.1%	40.2%	28.3%	39.0%
	1/4	1.079	(1.122)	1.127	(1.261)	1.245	(1.313)	1.217	(1.283)
		27.9%	38.1%	29.8%	40.2%	33.3%	43.3%	33.0%	42.7%
		31.4%	45.5%	32.9%	46.6%	37.3%	48.9%	36.9%	48.4%
	1/5	1.241	(1.171)	1.353	(1.375)	1.529	(1.467)	1.510	(1.444)
		32.9%	45.2%	36.2%	46.8 %	40.7%	51.0%	41.9%	50.4%
		34.2%	48.9%	37.6%	49.8 %	42.0%	54.2%	42.9%	54.0%
	1/6	1.327	(1.179)	1.492	(1.464)	1.717	(1.576)	1.711	(1.553)
		35.8%	46.6%	41.7%	50.5 %	45.8%	56.0%	45.8%	56.4%
		32.7%	48.9%	39.0%	52.0 %	42.8%	58.0%	43.6%	57.9%
200	1/3	1.759	(1.324)	1.800	(1.355)	1.925	(1.378)	1.855	(1.358)
		49.6%	60.5%	51.2%	61.8 %	54.7%	65.1%	52.0%	63.8%
		53.9%	65.6%	55.8%	67.1 %	59.7%	70.1%	57.1%	69.3%
	1/4	2.314	(1.535)	2.449	(1.677)	2.658	(1.722)	2.577	(1.690)
		63.8%	72.7%	64.1%	73.8%	70.3%	78.4%	68.2%	77.2%
		64.2%	76.0%	64.7%	77.3%	70.7%	80.5%	68.9%	80.5%
	1/5	2.626	(1.577)	2.870	(1.825)	3.182	(1.917)	3.112	(1.882)
		71.2%	79.4%	72.4%	79.0 %	77.5%	84.0%	77.4%	84.2%
		71.2%	80.4%	72.4%	80.3 %	77.5%	84.9%	77.4%	84.9%
	1/6	2.835	(1.595)	3.162	(1.942)	3.584	(2.058)	3.535	(2.026)
		75.3%	83.5%	76.7%	83.7 %	82.4%	87.5%	81.8%	86.6%
		72.5%	84.0%	73.9%	84.0 %	79.4%	88.0%	79.9%	87.0%

Each cell contains the empirical mean of the test statistic with its standard deviation in parentheses on the first line, empirical rejection percentages at 5% and 10% nominal levels on the second line, and size-corrected empirical rejection percentages at 5% and 10% levels on the third line.