An equality test across nonparametric regressions

Pascal Lavergne*
INRA-ESR

March 97, revised August 98

Part of this work was done while visiting Universidad Carlos III de Madrid. Financial support from INRA and from the European Commision through research training grant ERBFMBICT961595 is gratefully acknowledged. The paper was presented at Universidad Carlos III and CEMFI (Madrid), DELTA (Paris), Camp Econometrics 98 (Catalina Island) and ESEM 98 (Berlin). I thank the participants and especially Manuel Arellano, Richard Blundell, Jean-Marc Robin and Quang Vuong for their comments.

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

Abstract

A procedure for testing equality across nonparametric regressions is proposed. The procedure allows for any dimension of the explanatory variables and for any number of subsamples. We consider the case of random explanatory variables and allow the designs of the regressors 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. In the case of a random C, our procedure is a general test of significance for qualitative variables in a nonparametric regression. In the case of a fixed C, our procedure provides a "nonparametric analysis of covariance." In both case, the test is a one-sided normal test and is consistent against all alternatives. We study its small sample behavior through Monte-Carlo simulations.

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

JEL classification: Primary C52; Secondary C14.

Résumé

Une procédure pour tester l'égalité d'une régression non-paramétrique entre différents groupes est proposée. La procédure autorise des régresseurs multidimensionnels et un nombre quelconque de groupes. Nous considérons des variables explicatives aléatoires et envisageons le cas où les valeurs de ces variables et le nombre d'observations diffèrent suivant les groupes. La division entre groupes est défini à partir d'une variable C qui peut être fixe ou aléatoire. Lorsque C est aléatoire, la procédure est un test de significativité de variables qualitatives dans une régression non-paramétrique. Lorsque C est fixe, la procédure est analogue à une analyse de covariance non-paramétrique. Dans les deux cas, nous obtenons un test normal unilatéral consistent contre toute alternative. Nous étudions son comportement en petits échantillons par des simulations.

Mots-Clés: Test d'hypothèse, Régression non-paramétrique, Variables qualitatives, Analyse de covariance.

An equality test across nonparametric regressions

Pascal Lavergne

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 tests have been used in various economic problems. Some instances are testing for gender or race discrimination in earnings functions, testing for stability over time of economic relationships, and in particular testing for poolability of panel data, testing of disequilibrium models, testing for switching of firms' strategies in microeconometric models derived from game-theory, ... The classical testing procedures assume a parametric form, usually a linear one, for the regression functions under test. But as is well-known, specifying incorrect parametric forms can lead to serious errors in inference. Indeed, rejection of the equality hypothesis can be due solely 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 problem of comparing regression curves in a nonparametric context has been mostly studied in the particular setup of two subsamples with a one-dimensional regressor. In this case, 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. The related procedures require identical regressor's designs. Kulasekera (1995) extend Hall and Hart's procedure to the case of two curves with different designs of explanatory variables using quasi-residuals, built from use of a nonparametric regression estimated on the first subsample and applied to the observations of the second subsample. Alternatively, one can directly use the mean squared differences between nonparametric regression estimates. This idea has been worked out in the fixed design case, when the two curves are assumed to be equal up to a known parametric transformation by Härdle and Marron (1990), and under the assumption of normality of the residuals by King, Hart and Wehrly (1991). More recently, Young and Bowman (1995) have proposed a test that compares several regressions depending on a one-dimensional random variable with normal residuals.

However, 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. In view of practical use in econometrics, this paper proposes a general asymptotic joint test of equality across nonparametric regressions that is consistent against any alternative to the null hypothesis. It extends previous work in many directions. First, our assumptions does not require normality or homoscedasticity of the regression errors and residuals are allowed to have different distributions across subsamples. Second, it allows for any dimension of the explanatory variables. Third, we deal with any number of subsamples. Fourth, we consider the case of random explanatory variables and allow the designs of the regressors and the number of observations to differ across subsamples. Fifth, the division into subsamples is defined through a variable which can be either fixed or random. As a leading case, we consider the situation where a random qualitative variable defines the split into different subsamples, as frequently arises in economic applications. Our procedure is then a general test of significance for qualitative variables in a nonparametric regression. It 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) and Lavergne and Vuong (1995). We subsequently extend our procedure to the setup where the split depends on a fixed qualitative variable. Our procedure here provides a "nonparametric analysis of covariance" that has numerous potential applications in and outside the field of econometrics.

For designing a general procedure, we formalize the problem as one of comparison of only two nested models, irrespective to the number of subsamples considered. Thus we can build a test statistic that compares nonparametric estimators under the null model and under the alternative. Such a comparison is analogous to the one performed in many consistent testing procedures for parametric specification of regression functions or significance of continuous covariates. However, our work has a distinctive feature with respect to previous work on testing against a nonparametric alternative. Indeed, tests of a parametric specification using nonparametric estimation use the fact that the parametric estimator in the null model has a faster rate of convergence than the nonparametric estimator in the alternative model. Similarly, tests for significance of continuous variables in nonparametric regression crucially rely on the difference in pointwise rates of convergence of the estimators in the competing models, which is related to the different dimension of the regressors sets. In contrast, we argue that in the particular testing issue that we address, there is no justification for such discrepancy in rates of convergence. Moreover there is no need to require it for deriving a consistent testing procedure. Therefore, we consider equal rates of convergence for estimators in each model and we investigate thoroughly the implications of this peculiarity.

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 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 simulations. The Conclusion summarizes our main findings. All the proofs are relegated to the last Section 6.

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.s 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. Let $K(\cdot)$ be a kernel on $I\!R^p$ and h_n a bandwidth. For any c, a nonparametric kernel estimator of $f_c(\cdot)$, the conditional density of X given $C = c$, is

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

where $n_c = \sum_{i=1}^n I\!I[C_i = c]$. A nonparametric kernel estimator of $R(\cdot, c)$ is obtained as

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

In the formulas, we use non-smoothing weights for the qualitative variable C. If there exists a natural ranking of the modalities of C that is likely to be relevant in the regression model, non-smoothing weights can be replaced by smooth ones without changing the estimators' properties, see Delgado and Mora (1995).

If we overlook the information concerning the splitting as given by the C_i's, we would consider instead the 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 its kernel estimate on the whole sample

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

where

$$f_n(x) = (n h_n^p)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \qquad \forall x \in I\!R^p.$$

These estimators converge respectively to $r(\cdot) = \sum_{c=1}^{\bar{C}} p_c R(\cdot, c) f_c(\cdot)/f(\cdot)$, the conditional expectation of Y given X, and $f(\cdot) = \sum_{c=1}^{\bar{C}} p_c f_c(\cdot)$, 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, i.e. across the subsamples defined by the variable C. Equivalently, it means that we are not loosing any information by disregrading the C_i's and estimating the simpler regression Model (2.2) instead of (2.1). Thus the null hypothesis can write

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

This intuitive formulation enables us to deal with the testing problem as a comparison of two nested models, whatever the number of subsamples is.[1] Because we aim to compare the unknown regression functions $R(X, C)$ and $r(X)$, we will rely on their respective estimators $R_n(X, C)$ and $r_n(X)$. We are using the same amount of smoothing, as well as the same kernel, for both

[1]Quade (1982), Young and Bowman (1995) and Koul and Schick (1996) also use a pooling model to built tests of equality of regression functions. The latters consider only the two samples case and design testing procedures that are consistent against one-sided alternatives only.

estimators. There are many reasons for this choice. First, from an estimation viewpoint, there is no reason why we should employ different parameters in each model. The sample size is the same in both models. Moreover, it is known that a discrete variable does not affect the rate of convergence of nonparametric estimators and does not create any bias in estimation, see Bierens (1987) and Delgado and Mora (1995). Similarly, from their definitions, both functions $r(\cdot)$ and $R(\cdot,\cdot)$ have similar smoothness properties, so that the order of the kernel should be the same for both models. 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.[2] Second, from a testing viewpoint, using different amount of smoothing for each of the two models may lead to incorrect inferences. Indeed, it is likely to attenuate the discrepancies between the regression functions if the alternative were to hold. Conversely, it may introduce spurious differences between the two models when they are in fact equivalent.[3]

The last point is illustrated by Figure 1. From 200 observations generated as in Section 4 under the null hypothesis, we estimate separate regression functions for the two subsamples and compare them first (cf. Figure 1a) to the pooled estimated curve with the same bandwidth and second (cf. Figure 1b) to an oversmoothed pooled estimated curve (pooled estimates are represented as discontinuous lines). In Figure 1a, the pooled estimated curve always appears to lie in between the two separate regression functions. This is because the nonparametric estimators fulfill $r_n(\cdot) = \sum_{c=1}^{\bar{C}} (n_c/n) R_n(\cdot, c) f_{n,c}(\cdot)/f_n(\cdot)$, which is the empirical counterpart of the equality $r(\cdot) = \sum_{c=1}^{\bar{C}} p_c R(\cdot, c) f_c(\cdot)/f(\cdot)$. In contrast, when oversmoothing the pooling model, the pooled estimated curve lie within some intervals either below or above both separate curves. Furthermore, in the tails where only observations from one group are available, the pooled estimated curve from Figure 1b can markedly differ from the estimator on this subsample, while in Figure 1a the two are identical. Therefore, using the same bandwidth and kernel parameters seems to be the easiest way to put both model on equal footing in the testing procedure. This also constitutes a practical advantage for implementation, because the behavior of estimators under the null and alternative model are driven by only one free smoothing parameter. By contrast, other

[2]We could also allow for different bandwidths in our results such that their ratio tends to a non-null constant. However, the determination of this constant itself would be a difficult issue.

[3]Young and Bowman (1995) give supplementary justifications for using similar amount of smoothing in the two models.

testing procedures using nonparametric estimation, i.e. parametric specification tests against a nonparametric alternative or significance testing of continuous variables in nonparametric regression, heavily rely on the fact that the estimator in the null model is independently determined from the competing estimator under the alternative.

As the null hypothesis of interest corresponds to the non-significance of the discrete variable C, we can built our test statistic in a way similar to Lavergne and Vuong (1995), who deal with significance testing of a continuous variable. Let u denote the difference between Y and $r(X)$. For testing H_0 and obtaining a procedure consistent against any alternative, 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 $\Psi(X,C)$ that is strictly positive and non zero on the support of (X,C). Because of the form of the kernel estimate, it is convenient to use $f^2(X)f_C(X)$ as a weighting function. 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) and Lavergne and Vuong (1995). 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

$$V_{0n} = \frac{1}{n^{(2)}}\sum_a u_i u_j f(X_i) f(X_j) \frac{1}{h_n^p} K\left(\frac{X_i - X_j}{h_n}\right) w_{nij},$$

where $w_{nij} = \frac{n-1}{n_{C_i}-1}\mathbb{I}[C_i = C_j]$, $\sum_a$ 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. Dropping suitable terms as in Lavergne and Vuong (1995), we obtain the statistic

$$V_n = \frac{1}{n^{(4)}}\sum_a (Y_i - Y_k)(Y_j - Y_l) K_{nik} K_{njl} K_{nij} w_{nij}, \qquad (2.3)$$

where $K_{nij} \equiv (1/h_n^p) K\left[(X_i - X_j)/h_n\right]$.[4]

2.2 Asymptotic behavior of V_n

Theorem 1 gives the behavior of V_n under the hypotheses

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

[4]Lavergne and Vuong (1995) show that dropping similar indices in the sum does not change the asymptotic distribution of their statistic but reduces its small-sample bias. In our case, dropping similar indices is essential to obtain the asymptotic distribution of our statistic.

where $\{\delta_n, n = 1, \ldots\}$ is a sequence of reals from $[0,1]$. The fixed alternative corresponds to $\delta_n = 1\ \forall n$, while the null corresponds to $\delta_n = 0\ \forall n$.[5] Moreover, 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 0.

For stating and commenting our results, we need some definitions and notations. 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 let $w_{CC'} \equiv \frac{1}{p_C} I\!I[C = C']$ and define $*$ as the convolution operator, i.e.

$$(K * K)(u) = \int_{I\!R^p} K(t)K(u - t)\, dt.$$

We call $\mathcal{U}^p$ the class of integrable uniformly continuous functions from $I\!R^p$ to $I\!R$, and $\mathcal{D}^p_{m,q}$ 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, $q \in (0,1)$. Moreover, we define $\mathcal{K}^p_m$, $m \geq 2$, as the class of integrable functions K from $I\!R^p$ to $I\!R$ with compact support, satisfying $\int K(s)\, ds = 1$ and

$$\int s_1^{\alpha_1} \ldots s_p^{\alpha_p} K(s)\, ds = 0 \quad \text{for } 0 < \sum_{i=1}^{p} \alpha_i \leq m - 1.$$[6]

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 where Y has finite eight moment.*

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

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

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

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

$$E_{CC'}(X) = \int \left[K(t)w_{CC'} - 2(K * K)(t)\frac{f_C(X)}{f(X)} + (K * K * K)(t)\frac{g^2(X)}{f^2(X)}\right]^2 dt$$

and $g^2(X) = \sum_{c=1}^{\bar{C}} p_c f_c^2(X)$.

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

[6] The unity integral assumption is actually not necessary, but we impose it as it is not restrictive.

We first discuss our assumptions. Assumption 2.1 allows for dependence between (X, Y) and C. In particular, the distribution of the regressors can vary across subsamples.[7] Similarly, the residuals distributions are not restricted to be identical for different values of C. The residuals can also be heteroscedastic with respect to X. 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 more tedious proofs. Our assumptions on the bandwidth include the usual ones, and specifically imply that h_n goes to zero as the sample size grows, while its rate of decrease is restricted by $nh_n^p \to +\infty$. The last condition relates the rate of convergence of the statistic and its bias rate. When comparing two nonparametric regression curves, Härdle and Marron (1990) obtain a statistic with a bias of order $(1/nh_n^p)$. In our context, the bias is of order $h_n^{2(m+q)}$ and is controlled through the condition $nh_n^{p/2}h_n^{2(m+q)} \to 0$. With respect to the optimal rate for estimating the regression function, i.e. $h_n \propto n^{-1/[p+2(m+q)]}$, this implies undersmoothing as is usual in semiparametric estimation, see Robinson (1988) and Powell, Stock and Stoker (1989) among others.

As shown in the proofs, the behavior of V_n depends on whether the null hypothesis holds or not. Under the alternative, V_n asymptotically converges to a normal distribution with the usual $\sqrt{n}$-rate of convergence. But under the null, the asymptotic distribution of V_n has both a null expectation and a zero asymptotic variance. This degeneracy leads us to consider higher-order terms in the asymptotic expansion of V_n. For this we use a central limit theorem for degenerate U-statistics, see Fan and Li (1996). Similar situations also arise in other studies of testing problems, as 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 V_{0n}, where the elements u_i of the null model are replaced by parametric or nonparametric estimators. But because in the latter cases estimators of the null regression model have a pointwise faster rate of convergence than estimators in the alternative general model, plugging-in estimators in V_{0n} does not affect its asymptotic behavior. In contrast, in our case, the estimators in the general Model (2.1) and the restricted Model (2.2) have similar pointwise rates of convergence. Consequently, the asymptotic behavior of V_n differs from the one of V_{0n}. Nevertheless, our results show that in our setup plugging-in estimators of u_i influences the asymptotic variance under the null hypothesis, but affects neither the asymptotic

[7]See the end of Section 3.1 for a discussion on this point.

expectation nor the rate of convergence under H_0.

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

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

The asymptotic variance of V_n under the null hypothesis has a quite complicated form. First, it depends on the cross-products between $\sigma_c^2(\cdot)$ and $\sigma_{c'}^2(\cdot)$ for different c and c', that is on the cross-products of "conditional variances" from Model (2.2) between different subsamples. Second, it explicitely depends on the difference in the designs between subsamples, through the ratios $f_c(\cdot)/f(\cdot)$ and $g^2(\cdot)/f^2(\cdot)$. The first quantity is the ratio of the conditional density of X given $C = c$ to the "average" marginal density $f(\cdot) = \sum_{c=1}^{\bar{C}} p_c f_c(\cdot)$. The second equals $\left(\sum_{c=1}^{\bar{C}} p_c f_c^2(x)\right)/f^2(\cdot)$ and can be given the interpretation of a "normalized variance" of $f_c(\cdot)$. In the case where X is independent of C, both ratios equal one for any x and c. But, as we do not require such an independence assumption, the designs may differ markedly across subsamples. Hence, in general, these ratios introduce very different weightings across the subsamples and the values of the explanatory variables. Therefore, even in the simple case with two subsamples with identical sizes, it seems impossible to find a kernel that would minimize the variance irrespective of the designs of the explanatory variables.

Had we used different amounts of smoothing in the two models, and specifically imposed oversmoothing in Model (2.2) with respect to Model (2.1), the results of Theorem 1 would still hold. But the asymptotic variance of V_n would then reduce to the one of V_{0n}, i.e.

$$\omega_0^2 = 2E\left[\sigma_C^2(X)\sigma_{C'}^2(X) f^4(X) w_{CC'}^2\right] \int K^2(t)\, dt$$
$$= 2\int K^2(t)\, dt \sum_{c=1}^{\bar{C}} \int \sigma_c^4(X) f^4(X) f_c^2(X)\, dX.$$

It is noticeable that the variance ω_0^2 has none of the features of the variance ω^2. It does not depend at all on the cross-products of the conditional variances. It does not explicitely depend on the differences in the designs (though obviously the different $f_c(\cdot)$, $c = 1, \ldots, \bar{C}$, play a role in integration). Moreover, it is also independent of the probabilities p_c, $c = 1, \ldots, \bar{C}$, so that each value of C plays the same role in the variance whatever its probability of occurence is. These findings appear as supplementary justifications for not using different amounts of smoothing in each model.

More generally, one could derive the asymptotic variance when using a specific bandwidth for each model, with their ratio converging to a finite constant. Varying this constant gives more or less weight to the different terms in the asymptotic variance. In general, we cannot say which choice of bandwidths would minimize this variance. Oversmoothing of the pooling model comes to the specific choice of a bandwidths' ratio converging to zero. This leads in particular to ignore some interaction terms in the asymptotic variance, which are however present in finite samples. Our approach explicitely takes these interaction terms into account and aims to control for them by imposing identical bandwidths.

2.3 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

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

$g_n^2(x) = \sum_{c=1}^{\bar{C}} (n_c/n) f_{n,c}^2(x)$, $\forall x \in \mathbb{R}^p$, and b_n is a trimming parameter such that $b_n = o(1)$. An alternative estimator, which is computationally less demanding but more biased in small samples, is

$$\omega_n^2 = \frac{2}{n^{(2)}} \sum_a u_{ni}^2 f_n^2(X_i) u_{nj}^2 f_n^2(X_j) K_{nij} E_{nij}, \tag{2.5}$$

where $u_{ni} \equiv Y_i - r_n(X_i)$. 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, an assumption on b_n that ensures consistency of ω_n^2 is that $b_n^{-1} \sup_{x \in \mathbb{R}^p} |f_{n,c}(x) - f_c(x)| = o_p(1)$, for all c. In view of our Assumption 2, sufficient conditions are $(b_n \sqrt{n} h_n^p)^{-1} = o(1)$ and $b_n^{-1} h^{m+q} = o(1)$, see Lavergne and Vuong (1996) for details.[8]

[8] Our simulation results indicate that the trimming parameter, though necessary in theory, is not crucial in practice.

Therefore, we can propose $nh^{p/2}V_n/\omega_n$ as a test statistic for testing equality across nonparametric regressions. From Theorem 1, by letting $\delta_n = 0$, this test statistic is asymptotically $N(0,1)$ under the null hypothesis, and by letting $\delta_n = 1$, it diverges to $+\infty$ under any fixed alternative to H_0. Thus, for implementing the testing procedure, one chooses a critical value from the standard normal distribution for some significance level. If the value of the test statistic is larger than this critical value, then one rejects the null hypothesis of equality of the regression functions. If the value of the test statistic is smaller than the critical value, then one accepts the null hypothesis, i.e. one concludes to the non-significance of the qualitative variable C in the regression function of Model (2.1). The test is therefore a *one-sided* normal test and is consistent against any fixed alternative. In addition, by Theorem 1, the test has power to detect local alternatives of the type H_{1n} approaching the null at a rate slower than $(nh^{p/2})^{-1/2}$.

Different extensions of the procedure can be proposed. First, as C is a qualitative random variable with any fixed number of possible values, the procedure can be applied to test the significance of any set of qualitative variables in a nonparametric regression. The variable C is then 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, so that $R(X,C)$ becomes $R(X,D,C)$, $r(X)$ becomes $r(X,D)$, ... The rate of convergence of V_n will be unaffected as the discrete variables has no influence on the rate of convergence of nonparametric estimators. The asymptotic null distribution will be similar to the one of Theorem 1, with the arguments D added in the expression of ω^2. As noted before, we can equivalently use either smooth or non-smoothing weights for the discrete variables in D, as well as for those in C, without affecting the asymptotic properties of our procedure. Third, as detailled in the next section, both assumptions of independent observations and identically distributed observations 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, such as those concerning different industries and/or different countries. This is also true for experiments in which one can control for some factors. The general results from the previous section can be adapted to be used in this context. Specifically, let C be a variable taking integer values in $\{1,\ldots,\bar{C}\}$. For each c, we assume that we have at hand a n_c i.i.d. sample from a random variable (X_c, Y_c) on $I\!R^p \times I\!R$, 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 $\left\{(X_i, Y_i), i = 1 + \sum_{c'<c} n_{c'}, \ldots, \sum_{c'\leq c} n_{c'}\right\}$ and $\sum_{c=1}^{\bar{C}} n_c = n$. We then consider the general regression model

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

so that now $R(X_i, c)$ 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 a i.i.d. sample leads to consider the regression model

$$Y_i = r(X_i) + u_i, \qquad E\left[u_i|X_i\right] = 0, \qquad i = 1,\ldots,n. \tag{3.2}$$

Nonparametric kernel estimators of $f_n(\cdot)$ and $r_n(\cdot)$ are defined as in the previous section, but their interpretation changes radically. Here $f_n(\cdot)$ estimates $f(\cdot) = \sum_{c=1}^{\bar{C}} n_c f_c(\cdot)/n$, which is no more 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 of Y_c on X_c, $c = 1,\ldots,\bar{C}$. However, $f(\cdot)$ and $r(\cdot)$ play here exactly the same role as before, so that we may call them “marginal density” and “restricted regression” by abuse of language.

While the two models are now interpreted differently, the framework is really similar. The null hypothesis of interest is still the constancy of the regression function $R(\cdot, c)$ for different values of c. Equivalently, it means that it is possible to pool the data and to estimate the function $R(\cdot, c)$ in Model (3.1) through the simpler Model (3.2), even though the densities $f_c(\cdot)$ for different values of c differ. Thus, the null hypothesis of interest can write $H_0 : R(X, c) = r(X)$ a.s. $\forall c = 1,\ldots,\bar{C}$ and the statistic V_n is constructed as in Section 2.1.

Theorem 2 gives our general result for a fixed C. We let $\sigma_c^2(X) \equiv E\left[(Y_c - r(X_c))^2 | X_c = X\right]$ and $w_{cc'} \equiv \frac{1}{p_c} I\!I[c = c']$, with $p_c = (n_c/n)$ (which is asumed to be fixed as C is fixed).

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 $\left\{(X_i,Y_i), i = 1+\sum_{c'<c} n_{c'},\ldots,\sum_{c'\leq c} n_{c'}\right\}$ is an i.i.d. sample from a random variable (X_c,Y_c) on $I\!R^p \times I\!R$, such that X_c has marginal density $f_c(\cdot)$ and Y_c has finite eight moment. Moreover, the subsamples are independent.*

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

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

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

$$E_{cc'}(X) = \int \left[K(t)w_{cc'} - 2(K*K)(t)\frac{f_c(X)}{f(X)} + (K*K*K)(t)\frac{g^2(X)}{f^2(X)}\right]^2 dt$$

and $g^2(X) = \sum_{c=1}^{\bar{C}} p_c f_c^2(X)$.[9]

Compared to the previous section, we have relaxed the assumption of identically distributed data across subsamples, but still we assume independent observations across subsamples, which is typically the case for cross-section data. The proof of Theorem 2 mainly follows the one of Theorem 1, see Section 6.2 for some brief explanations. From our theorem, it is straightforward to deduce a testing procedure based on $nh^{p/2}V_n/\omega_n$, where ω_n is an estimator of the asymptotic variance ω^2 similar to the ones given in the previous section. The test is as before a one-sided normal test, consistent against any alternative and detects local alternatives of the type H_{1n} provided that $\delta_n^2 nh_n^{p/2} \to +\infty$.

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

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

For testing the hypothesis $\beta_c = \beta, \forall c = 1,\ldots,\bar{C}$, the usual testing procedure is built upon

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

[9] In writing ω^2, we use the convention in (2.4).

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

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

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 on explanatory variables, namely by $K_{nik}K_{njl}K_{nij}$. Thus our testing procedure provides a "non-parametric analysis of covariance." Indeed, it allows to test if there exists any differential effect of the regressors on the dependent variable across the considered subsamples without imposing any parametric assumption at the outset.

Three main remarks follow from the interpretation of our procedure as a nonparametric analysis of covariance. The first remark is related to the choice of weights in the null hypothesis considered, which writes

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

In the case where the observations constitute a unique random sample, it is meaningful to use a weighting scheme proportional to $p_c f_c(.)$. However, in the present setup where C is fixed, the use of "frequencies" as weights in writing the null hypothesis is no longer readily interpretable. Obviously, there exist other equivalent formulations of the null hypothesis of interest, i.e. the constancy of $R(\cdot, c)$ with respect to c, that use different weighting schemes. To each formulation corresponds a test statistic. The relative merits of the different procedures will generally depend of the particular 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, our analysis allows the density of explanatory variables to vary across subsamples, so that the "treatments" (i.e. the discrete variable C) may affect the explanatory variables.[10] Therefore, 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. This second remark obviously extends to the case of a random C treated in the previous section.

Third, our procedure only applies for testing the strict equality of the whole 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 in 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 procedure is not completely straightforward. This and other extensions will be the topic of further work.

3.2 Panel data

One potential useful econometric application of our test is testing for poolability of panel data. We consider this problem separately for two main reasons. First, we need to detail the assumptions under which our test is applicable. Second, we want to compare our theoretical results with the ones in Baltagi, Hidalgo and Li (1996), which is to our knowledge the only work to date that proposes a nonparametric test of poolability for panel data.

Let us consider the panel data model

$$Y_{it} = R_t(X_{it}) + U_{it}, \qquad i = 1, \ldots, n_0, \qquad t = 1, \ldots, T. \tag{3.3}$$

At each period t, $\{X_{it}, i = 1, \ldots, n_0\}$ is a i.i.d. sample from X_t with density $f_t(\cdot)$. The null hypothesis of interest is the constancy of the regression function $R_t(\cdot)$ over time, that is

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

[10] After the first version of this paper was written, we have discovered an early paper by Quade (1982), who proposes nonparametric analysis of covariance methods. A first analysis, labelled analysis of covariance by matching, is valid only under the assumption that the distribution of X does not vary conditionally to C. The second one, named analysis of matched difference, does not require this assumption and is very close in spirit to our analysis, with the major difference that the bandwidth is considered as fixed.

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 here writes

$$V_n = \frac{T}{n^{(4)}} \sum_a (Y_{it} - Y_{kr})(Y_{jt} - Y_{ls}) \frac{1}{h_n^{3p}} K\left(\frac{X_{it} - X_{kr}}{h}\right) K\left(\frac{X_{jt} - X_{ls}}{h}\right) K\left(\frac{X_{it} - X_{jt}}{h}\right), \tag{3.4}$$

with $n = n_0 T$ and $\sum_a$ denotes summation over the arrangements of distinct indices $\{it, jt, kr, ls\}$. The results of the previous subsection, where we imposed independence across subsamples, do not readily apply in this context. Nevertheless, as we argue below, this assumption can be weakened without changing the results.[11] Thus, the asymptotic behavior of V_n as n_0 goes to infinity is given by Theorem 2. Its asymptotic variance is

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

where

$$E_{tt'} = \int \left[T\ K(u) I\!I[t = t'] - 2(K * K)(u) \frac{f_t(X)}{f(X)} + (K * K * K)(u) \frac{g^2(X)}{f^2(X)}\right]^2 du$$

and $g(\cdot) = (1/T) \sum_{t=1}^{T} f_t^2(\cdot)$.

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 allow 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. We here recall the fundamental definitions.

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 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

[11] A brief account of the necessary adaptations of the proof is given in Section 6.3.

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. Indeed, in a nonparametric context, such effects are included in the regression function, i.e. they are not separately identifiable.[12] Second, the formulation (3.3) 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 of our analysis are not as stringent as they may appear at first.

Baltagi, Hidalgo and Li (1996) consider a statistic which is basically built as ours, with the important difference that they introduce two different smoothing parameters h_n and a_n, using h_n for the general Model (3.3) and a_n for the model that pools the data. Subsequently, they require oversmoothing of the null regression model, i.e. the pooling one, relative to the general alternative one by imposing $h_n/a_n^2 = o(1)$. As a consequence, the asymptotic variance of the transformed statistic is

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

Contrary to ω^2, the variance ω_0^2 does not depend neither on the cross-products of conditional variances between periods nor on the differences in the designs between periods.[13] This occurs because using different amounts of smoothing in the null and the alternative model results in pulling out any cross-effect between periods in the test statistic. This is also the reason why the previous authors do not need the assumption of no serial residual correlation. However, as fully argued in the previous section, such oversmoothing of the pooling model does not seem justified for testing constancy of the regression functions across the different periods.

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

[13] The formula of ω_0^2 corrects a mistake in the formula of the asymptotic variance in Baltagi, Hidalgo and Li (1996), by replacing $f^5(\cdot)$ by $f^4(\cdot)f_t(\cdot)$. This mistake comes from their implicit assumption that the density of the regressors remains the same across time.

4 Small sample behavior

In this section, we investigate the behavior of our test in the case of a random qualitative variable C taking two values 0 or 1. We generate the data through

$$Y = aX + bX^3 + \mathbb{I}[C = 0]\ d(X) + U \tag{4.1}$$

where conditional on C, X is generated as $N(C, 1)$ and U is independently distributed as $N(0, \sigma_C^2)$. The null hypothesis corresponds to $d(X) \equiv 0$, and we consider different forms of alternatives as specified by $d(\cdot)$. We impose the restriction that $E[d(X)|C = 0] = 0$ and we set parameters a and b to -4 and 1 respectively, so that the conditional expectation of Y given C is independent of C.

We consider small ($n = 100$) and moderate ($n = 250$) sample sizes and run 2000 replications. For ease of computations, we choose the uniform kernel with support $[-1/2, 1/2]$. The bandwidth parameter is chosen as $h_n = a\ \hat{s}_X^2\ n^{-1/5}$, where $\hat{s}_X^2$ is the estimated standard deviation for all observations of X. The choice of $a = 1$ corresponds to the usual rule-of-thumb in kernel estimation and we let a vary so as to investigate the sensitivity of our testing procedure to the choice of the bandwidth. Unreported simulations show that the trimming parameter has very little influence on the results, so that it is arbitrarily set to 0 in all experiments.

The design of the alternatives has been chosen to investigate the power of our test with respect to the magnitude and the frequency of $d(\cdot)$. 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_5 and DGP_7. These departures from the null are of special interest, as 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 (1995).

We first consider the case of equal probabilities of $C = 0$ and $C = 1$, choosing identical residual variances $\sigma_0^2 = \sigma_1^2 = 1$. Table 1 reports our results for the null hypothesis (DGP_0) and the linear alternatives as we let a vary in the grid $(0.2, 0.5, 1, 1.5, 2)$. For each case, the first row gives the mean with standard deviation in parentheses of our test. The second row gives empirical levels of rejections for our test, the first figure corresponds to a 5% nominal level, while the second one corresponds to a 10% nominal level. For each sample size, the last row reports empirical rejection rates of the Chow test for the same nominal levels.

The first column relates to the null hypothesis. The mean of our test statistic is close to zero for small and moderate bandwidths, then increases as the bandwidth constant goes from 1 to 2. The test is closest to be unbiased with slight undersmoothing with respect to the rule-of-thumb. The standard deviation of our test statistic grows with the smoothing parameter, 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 (1995) on their nonparametric significance test for continuous regressors.[14] Under the null hypothesis, empirical sizes are much higher than the nominal ones for $a = 2$ because of the bias of the statistic, and much smaller than desired for a less than 1, because of the variance estimation bias. It is quite difficult to draw conclusions about a best choice for the bandwidth in terms of empirical size, as the variance estimation problem leads to systematic underrejections in our procedure. Because the same holds true under any alternative, the small sample power performances of our test are also understated.

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. Rough undersmoothing leads to small power, especially for alternatives of little amplitude. Though, our test can reasonably detect quite small linear alternatives such as DGP_1 for bandwidths that are greater than the rule-of-thumb. Furthermore, for alternatives of moderate amplitude, the power performance of our test can equal that of the Chow test, although the design is ideal for the latter. Our results also indicate that the highest power is attained for the largest tried bandwidth, though using an infinite bandwidth should ultimately lead to a trivial power.

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

Table 2 has the same structure as Table 1 and reports results relative to the sinus alternatives. For $n = 100$, our test has relatively low power against sinus alternatives when the bandwidth is smaller than the rule-of-thumb. When increasing the bandwidth, its performances improve except against the high-frequency alternative DGP_4, in which case its empirical power exhibits an inverse U-shape as a function of h_n. A different and striking feature appears from our results for $n = 250$. The empirical power of our test is only slightly affected by the frequency of the departure from the null. For all four alternatives it is close to the one observed against the linear alternative DGP_2. This is in sharp contrast with smooth tests for parametric specification or for significance of continuous regressors, which are very sensitive to the frequency of the alternative. For instance, in testing omitted continuous regressors, Lavergne and Vuong (1995) find that the bandwidth in the general model has to be adapted to the frequency of the alternative, namely, the higher the frequency, the smaller the bandwidth should be. This occurs because in the latter test, the behavior of the estimator under the null is idenpendently driven by another bandwidth parameter. On the contrary, our procedure uses the same smoothing parameter in the general and the pooling model. Then the bandwidth affects both estimators under the null hypothesis and under the alternative. As a consequence, our testing procedure appears to be very robust to the frequency of the considered alternatives for a moderate sample size. Our results show that h_n needs not be adjusted to detect departures from the null of varying frequencies, and in all considered cases, the maximum power is achieved for the largest tried bandwidth.[15]

For comparative purpose, 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 power higher than the latter for bandwidth constants greater than 1. 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 90% for a moderate sample size of 250.

To investigate the properties of the test under varying circumstances, we consider two different variations of the initial setup. We first study a case where there is a large discrepancy in the population with respect to values of C by letting $p_0 = 0.2$. In the second variation, we investigate the influence of residual variances by letting $\sigma_1^2 = 2$. (In both cases, the other

[15] Unreported results show that the bandwidth needs be very large to observe a decrease in empirical power.

characteristics of the data generating process are unchanged with respect to the initial setup.) Relying on what we have learned from the previous simulations, we focus on the null and linear alternatives for $n = 250$ and choose values of a among $(1, 1.5, 2)$. Table 3 reports results relative to these two situations. In the first variation and under the null hypothesis, the test statistic has roughly similar mean as in the initial setup, but has less variation, so that the test has smaller empirical size. The empirical power of the test is adversely affected with respect to the initial results, as could be expected given that only one fifth of the sample, on average, has a different behavior of the rest of the sample. Still our test enjoys reasonnable power properties against DGP_2 and very good ones against the largest alternative DGP_3, for which the percentage of rejections is always above 80%. In the second variation, the mean and the standard deviation of the test statistic have decreased with respect to the first set of simulations. Indeed, it could be expected that the bias in variance estimation is larger when residual variance is greater. The test is also less powerful against departures from the null of small amplitude, but is roughly unchanged against DGP_3, with rejection percentages greater than 96% for $n = 250$.

5 Conclusion

In this paper we propose a general test of equality across nonparametric regressions. It is based on the comparison of the regression function for each subsample with the general one that pools all the observations. It applies in a variety of situations, and in particular whether or not the division into subsamples is defined in a random way. In our presentation, we have first considered the leading situation where a random qualitative variable defines the split into different subsamples and where all observations are independent and identically distributed. 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.s within each subsample and under the assumption that the residuals are uncorrelated across subsamples.

The characteristic feature of our procedure is that it uses a common smoothing parameter for the pooling estimator and the estimators based on the subsamples. We have justified this choice and investigated thoroughly its implications. Besides the practical advantage that the practitioner needs only choose one smoothing parameter, another one is that our test is much less sensitive to the frequency of the alternative as shown in our simulations. Though, the bandwidth choice is clearly a key issue for application of the test. Bootstrap methods could be a way to bypass this problem, as bootstrap tests usually provide better approximations to the asymptotic null distribution than asymptotics do and can be much less sensitive to bandwidth choice, see e.g. Delgado, Dominguez and Lavergne (1998). This possibility should be investigated both from a theoretical and a practical viewpoint.

6 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 stands for (C_i, X_i, Y_i), $i = 0, 1, \ldots, n$. Also $h \equiv h_n$, $K_{nij} \equiv h_n^{-p} K\left[(X_i - X_j)/h_n\right]$, $\boldsymbol{K} \equiv |K|$ and $i, j, k, l, i', j', k', l'$ refer to indices that are pairwise different unless stated otherwise. We let $\widehat{f}_i = (n-1)^{-1} \sum_{k \neq i} K_{nik}$, and more generally for any index set I not containing i with cardinality $|I|$, $\widehat{f}_i^I = (n - 1 - |I|)^{-1} \sum_{k \neq i, k \notin I} K_{nik}$.

6.1 Proof of Theorem 2.1

6.1.1 Outline of the proof

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

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

where

$$
\begin{aligned}
I_1 &= \frac{n-2}{n-3} \frac{1}{n^{(2)}} \sum_a u_i u_j f_i f_j K_{nij} w_{nij} + \frac{2(n-2)}{n-3} \frac{1}{n^{(2)}} \sum_a u_i (\widehat{f}_i^j - f_i) u_j f_j K_{nij} w_{nij} \\
&\quad + \frac{n-2}{n-3} \frac{1}{n^{(2)}} \sum_a u_i (\widehat{f}_i^j - f_i) u_j (\widehat{f}_j^i - f_j) K_{nij} w_{nij} - \frac{2}{n^{(3)}} \sum_a u_i f_i u_l K_{njl} K_{nij} w_{nij} \\
&\quad - \frac{2}{n^{(3)}} \sum_a u_i (\widehat{f}_i^{j,l} - f_i) u_l K_{njl} K_{nij} w_{nij} + \frac{1}{n^{(4)}} \sum_a u_k u_l K_{nik} K_{njl} K_{nij} w_{nij}
\end{aligned}
$$

$$
\begin{aligned}
&\quad - \frac{1}{n^{(4)}} \sum_a u_i u_j K_{nik} K_{njk} K_{nij} w_{nij} \\
&= \frac{n-2}{n-3} [V_{0n} + 2I_{1,1} + I_{1,2}] - 2I_{1,3} - 2I_{1,4} + I_{1,5} - I_{1,6}, \\
I_2 &= \frac{1}{n^{(3)}} \sum_a u_i f_i (r_j - r_l) K_{njl} K_{nij} w_{nij} + \frac{1}{n^{(3)}} \sum_a u_i (\widehat{f}_i^{j,l} - f_i)(r_j - r_l) K_{njl} K_{nij} w_{nij} \\
&\quad - \frac{1}{n^{(4)}} \sum_a u_k (r_j - r_l) K_{nik} K_{njl} K_{nij} w_{nij} = I_{2,1} + I_{2,2} - I_{2,3}.
\end{aligned}
$$

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

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

where $L_n \xrightarrow{d} N(0, \omega^2)$,

$$nh^{p/2} [I_1 - (V_{0n} - 2I_{1,3} + I_{1,5})] = \delta_n^2 nh^{p/2} o_p(1) + \delta_n \sqrt{n} h^{p/2} O_p(1) + o_p(1),$$

$$nh^{p/2} I_2 = \delta_n^2 nh^{p/2} o_p(1) + \delta_n \sqrt{n} h^{p/2} o_p(1) + \delta_n nh^{p/2} h^{(m+q)} O_p(1) + o_p(1),$$

and $nh^{p/2} I_3 = o_p(1)$. Therefore

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

In case (i),

$$
\begin{aligned}
\delta_n \sqrt{n} h^{p/2} &= (\delta_n^2 nh^{p/2})^{1/2} h^{p/4} = o(1) \\
\delta_n nh^{p/2} h^{(m+q)} &= (\delta_n^2 nh^{p/2})^{1/2} (nh^{p/2} h^{2(m+q)})^{1/2} = o(1).
\end{aligned}
$$

Thus $nh^{p/2} [V_n - (V_{0n} - 2I_{1,3} + I_{1,5})] = o_p(1)$ and $nh^{p/2} [V_{0n} - 2I_{1,3} + I_{1,5}] \xrightarrow{d} N(C\mu, \omega^2)$.
In case (ii),

$$
\begin{aligned}
\delta_n \sqrt{n} h^{p/2} &= (\delta_n^2 nh^{p/2}) \frac{h^{p/4}}{(\delta_n^2 nh^{p/2})^{1/2}} = o(\delta_n^2 nh^{p/2}) \\
\delta_n nh^{p/2} h^{(m+q)} &= (\delta_n^2 nh^{p/2}) \frac{(nh^{p/2} h^{2(m+q)})^{1/2}}{(\delta_n^2 nh^{p/2})^{1/2}} = o(\delta_n^2 nh^{p/2}).
\end{aligned}
$$

Thus $nh^{p/2} [V_n - (V_{0n} - 2I_{1,3} + I_{1,5})] = o_p(\delta_n^2 nh^{p/2})$ and $nh^{p/2} [V_{0n} - 2I_{1,3} + I_{1,5}] = \delta_n^2 nh^{p/2} \mu + o_p(\delta_n^2 nh^{p/2})$. Theorem 1 follows. Q.E.D.

6.1.2 U-Statistics

Let $U_{k,n} = (1/n^{(m_k)}) \sum_a H_{k,n}(Z_{i_1}, \ldots, Z_{i_{m_k}})$, $k = 1, 2$, be arbitrary U-statistics, where the Z_i's are identically distributed but the $H_{k,n}$ are not necessarily symmetric. Then,

$$\begin{aligned} E(U_{1,n}U_{2,n}) &= \frac{1}{n^{(m_1)}} \frac{1}{n^{(m_2)}} \sum_{s=0}^{\min(m_1,m_2)} \frac{n^{(m_1+m_2-s)}}{s!} \sum_{|\Delta_1|=s=|\Delta_2|}^{(s)} J(\Delta_1, \Delta_2) \\ &= \sum_{s=0}^{\min(m_1,m_2)} O(n^{-s})\xi_{s,s} \end{aligned} \tag{6.1}$$

where $\xi_{s,s} = (1/s!) \sum^{(s)} J(\Delta_1, \Delta_2)$, and

$$E(U_{1n}^2) = \left(\frac{1}{n^{(m_1)}}\right)^2 \sum_{s=0}^{m_1} \frac{n^{(2m_1-s)}}{s!} \sum_{|\Delta_1|=s=|\Delta_2|}^{(s)} I(\Delta_1, \Delta_2) = \sum_{s=0}^{m_1} O(n^{-s})\xi_s \tag{6.2}$$

where $\xi_s = (1/s!) \sum^{(s)} I(\Delta_1, \Delta_2)$ and $\sum^{(s)}$ denotes summation over sets Δ_1 and Δ_2 of ordered positions of length s,

$$J(\Delta_1, \Delta_2) = E\left[H_{1,n}(Z_{i_1}, \ldots, Z_{i_{m_1}})H_{2,n}(Z_{j_1}, \ldots, Z_{j_{m_2}})\right],$$

$$I(\Delta_1, \Delta_2) = E\left[H_{1,n}(Z_{i_1}, \ldots, Z_{i_{m_1}})H_{1,n}(Z_{j_1}, \ldots, Z_{j_{m_1}})\right],$$

and the i's in position Δ_1 coincide with the j's in position Δ_2 and are pairwise distinct otherwise. In what follows, we intensively use (6.1) and (6.2) to bound $E(U_{1,n}U_{2,n})$ and $E(U_{1,n}^2)$. Indeed, if $\bar{Z}_s$ denotes the vector of common Z_i's, we have by conditioning on $\bar{Z}_s$ that

$$\begin{aligned} J^2(\Delta_1, \Delta_2) &= E^2\Big[E[H_{1,n}(Z_{i_1}, \ldots, Z_{i_{m_1}})|\bar{Z}_s]E[H_{2,n}(Z_{j_1}, \ldots, Z_{j_{m_2}})|\bar{Z}_s]\Big] \\ &\leq E\left[E^2[H_{1,n}(Z_{i_1}, \ldots, Z_{i_{m_1}})|\bar{Z}_s]\right] E\left[E^2[H_{2,n}(Z_{j_1}, \ldots, Z_{j_{m_2}})|\bar{Z}_s]\right] \end{aligned}$$

by Cauchy-Schwartz inequality, and we have also the similar inequality for $I(\Delta_1, \Delta_1)$.
We will also rely on the following lemma from Fan and Li (1996).

Lemma 1 *: Let U_n be a U-statistic of order m with symmetric 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 $s = 1, \ldots, m$ 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 $s = 3, \ldots, m$ and

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

as $n \to \infty$*, then* nU_n *is asymptotically normal with zero mean and variance* $(1/2)\left(m^{(2)}\right)^2 \xi_2$.

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

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

Hence one needs to determine all the terms in the double summation. Similar expressions are derived for ξ_s, $s = 3, \ldots, m$.
It can also be easily shown that an anologous result holds for two U-statistics of order m_1 and m_2 with repective kernels $H_{1,n}$ and $H_{2,n}$. Their asymptotic covariance is then given by

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

6.1.3 Behavior of V_{0n}

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 $\mu_n \longrightarrow \mu$, $nh^{p/2}U_{0n} \xrightarrow{d} N(0, \omega_0^2)$ *and* $B_n \xrightarrow{d} 2N(0, \xi - \delta^2\mu^2)$, *with* $\delta = \lim_{n\to\infty}\delta_n$ *and* $\xi = E\left[\sigma_C^2(X) d^2(X, C) f^4(X) f_C^2(X)\right]$.

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

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

(i) Limit of θ_n:

$$\theta_n = E\left[u_i f_i u_j f_j K_{nij} w_{nij}\right]$$

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

Now

$$\begin{aligned} E\left[\delta_n d_j f_j K_{nij} | X_i, C_i, C_j\right] &= \int \delta_n d(X_j, C_j) f\left(X_j\right) K_{nij} f_{C_j}\left(X_j\right) dX_j \\ &= \delta_n d(X_i, C_j) f\left(X_i\right) f_{C_j}\left(X_i\right) + o(\delta_n) \end{aligned}$$

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

$$\begin{aligned} \theta_n &= \delta_n^2 E\left[d_i f_i w_{nij} d(X_i, C_j) f(X_i) f_{C_j}(X_i)\right] + o(\delta_n^2) \\ &= \delta_n^2 E\left[d_i f_i w_{nij} d(X_i, C_i) f(X_i) f_{C_i}(X_i)\right] + o(\delta_n^2) \\ &= \delta_n^2 E\left[d_i^2 f_i^2 E\left(w_{nij} | C_i\right) f_{C_i}(X_i)\right] + o(\delta_n^2) = \delta_n^2 \mu_n \end{aligned}$$

with $\mu_n \longrightarrow \mu = E\left[d^2(X, C) f^2(X) f_C(X)\right]$.

(ii) Distribution of W_{0n}:

$$\begin{aligned} E\left[E^2\left(H_n(Z_i, Z_0) | Z_i\right)\right] &= E\left[u_i^2 f_i^2 E^2(u_0 f_0 K_{ni0} w_{ni0} | Z_i)\right] \\ &= \delta_n^2 E\left[u_i^2 f_i^2 E^2(d_0 f_0 K_{ni0} w_{ni0} | Z_i)\right] = \delta_n^2 \xi_n \end{aligned}$$

with $\xi_n \longrightarrow \xi = E\left[\sigma_C^2(X) d^2(X, C) f^4(X) f_C^2(X)\right]$, as $\delta_n d(X, C) f(X) f_C(X) \in \mathcal{U}^p, \ \forall C$.

Now $E \mid E\left[H_n(Z_i, Z_j) | Z_i\right] \mid^\nu = E \mid u_i^\nu f_i^\nu E^\nu\left[u_0 f_0 K_{ni0} w_{ni0} | Z_i\right] \mid = O(1) = o(n^{\nu/2-1})$ for $2 < \nu \leq 4$, as $E \mid Y^{2\nu} \mid < \infty$. Thus, by Theorem 7.1 of Hoeffding (1948),

$$\sqrt{n}\left[W_{0n} - 2\theta_n\right] \rightarrow 2\delta N\left(0, \xi - \delta^2 \mu^2\right).$$

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

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

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

$$\frac{E[G_n^2] + n^{-1} E[H_n^4]}{E^2[H_n^2]} = o(1), \tag{6.3}$$

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 $\sigma_C^2(X)f^2(X)f_C^2(X) \in \mathcal{U}^p,\ \forall C$,

$$E\left[H_n^2(Z_i, Z_j)\right] = E\left[\sigma_{C_i}^2(X_i)\sigma_{C_j}^2(X_j)f_i^2 f_j^2 K_{nij}^2 w_{nij}^2\right] = h^{-p}\omega_{0,n}^2/2,$$

where $\omega_{0,n}^2 \longrightarrow \omega_0^2 = 2E\left[\sigma_C^2(X)\sigma_{C'}^2(X)w_{CC'}^2 f^4(X)\right]\int K^2(t)\,dt$, using the definition (2.4).
Moreover, as $E(u^4|X, C)f^4(X)f_C(X) \in \mathcal{U}^p,\ \forall C$,

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

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

$$\begin{aligned}
&E[G_n^2]\\
&= \int \sigma_i^2 f_i^2 \sigma_j^2 f_j^2 E\left[\int \sigma_{C_0}^2 K_{ni0}K_{nj0} f^2(X_0) f_{C_0}(X_0)\,dX_0\, w_{ni0}w_{nj0}|Z_i, Z_j\right]^2 f_i f_j\, dX_i dX_j\, d\nu\,(C_i, C_j)\\
&= h^{-p}\int \sigma_i^2 f_i^2 \sigma_{C_j}^2(X_i + ht) f^2(X_i + ht)\\
&\qquad E\left[\int \sigma_{C_0}^2(X_i - hs)K(s)K(s+t)\, f^2(X_i - hs) f_{C_0}(X_i - hs)\,ds\, w_{ni0}w_{nj0}|Z_i, Z_j\right]^2\\
&\qquad\qquad f_{C_i}(X_i) f_{C_j}(X_i + ht)\,dX_i dt\,d\nu\,(C_i, C_j)\\
&= h^{-p}E\left[\sigma_C^8(X)f^8(X)f_C^3(X)\right]\int (K * K)^2(t)\,dt + o(h^{-p})\\
&= O(h^{-p}),
\end{aligned}$$

where $\nu\,(C_i, C_j)$ denotes the distribution of (C_i, C_j). Thus condition (6.3) holds as $h \to 0$ and $nh^p \to \infty$. Collecting results, Proposition 1 follows. Q.E.D.

6.1.4 Behavior of $I_{1,3}$

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 normal with mean 0 and variance $2E\left[\sigma_C^2(X)\sigma_{C'}^2(X)w_{CC'}f^2(X)f_C^2(X)\right]\int (K*K)^2(t)\,dt$.*

Proof: We have $I_{1,3} = (1/n^{(3)})\sum_a u_i f_i u_l K_{njl}K_{nij}w_{nij}$, which is a U-statistic with kernel

$$H_n(Z_i, Z_j, Z_l) = u_i f_i u_l K_{njl} K_{nij} w_{nij}.$$

We now compute the corresponding $\xi_s,\ s = 0, 1, 2, 3$.

(i) $\xi_2 = h^{-p} 2E\left[\sigma_C^2(X)\sigma_{C'}^2(X)f^2(X)f_C^2(X)\right]\int (K*K)^2(t)\,dt + o(h^{-p})$. Indeed we have

$$\begin{aligned}
E(H_n|Z_i,Z_j) &= u_i f_i K_{nij} w_{nij} E\left(u_l K_{njl}|Z_j\right) = 0,\\
E(H_n|Z_i,Z_l) &= u_i f_i u_l E\left(K_{njl}K_{nij}w_{nij}|Z_i,Z_l\right),\\
E(H_n|Z_j,Z_l) &= u_l K_{njl} E\left(u_i f_i K_{nij} w_{nij}|Z_j\right) = \delta_n u_l K_{njl} E\left(d_i f_i K_{nij} w_{nij}|Z_j\right).
\end{aligned}$$

Then,

$$\begin{aligned}
E\left[E^2(H_n|Z_i,Z_l)\right] &= E\left[u_i^2 f_i^2 u_l^2 E^2\left(K_{njl}K_{nij}w_{nij}|Z_i,Z_l\right)\right]\\
&= E\left[\sigma_i^2 f_i^2 \sigma_l^2 E^2\left(K_{njl}K_{nij}w_{nij}|Z_i,Z_l\right)\right]\\
&= h^{-p}E\left[\sigma_C^2(X)\sigma_{C'}^2(X)f^2(X)f_C^2(X)\right]\int (K*K)^2(t)\,dt + o(h^{-p}),\\
E\left[E^2(H_n|Z_j,Z_l)\right] &= \delta_n^2 E\left[u_l^2 K_{njl}^2 E^2\left(d_i f_i K_{nij} w_{nij}|Z_j\right)\right]\\
&= \delta_n^2 E\left[\sigma_l^2 K_{njl}^2 d_j^2 f_j^2 f_{C_j}^2(X_j)\right]\\
&= \delta_n^2 O(h^{-p}) E\left[u_l^2 \boldsymbol{K}_{njl} d_j^2 f_j^2 f_{C_j}^2(X_j)\right] = O(\delta_n^2 h^{-p})
\end{aligned}$$

and

$$\begin{aligned}
&E\left[E(H_n|Z_i,Z_l)E(H_n|Z_j,Z_l)\right]\\
&= E\left[u_i f_i u_l^2 K_{njl} E\left(K_{njl}K_{nij}w_{nij}|Z_i,Z_l\right) E\left(u_i f_i K_{nij} w_{nij}|Z_j\right).\right]\\
&= \delta_n^2 E\left[d_i f_i \sigma_l^2 K_{njl} E\left(K_{njl}K_{nij}w_{nij}|Z_i,Z_l\right) E\left(d_i f_i K_{nij} w_{nij}|Z_j\right)\right]\\
&= \delta_n^2 O(h^{-p}) E\left[d_i f_i \sigma_l^2 K_{njl} E\left(\boldsymbol{K}_{nij} w_{nij}|Z_i,Z_l\right) E\left(d_i f_i K_{nij} w_{nij}|Z_j\right)\right]\\
&= O(\delta_n^2 h^{-p}).
\end{aligned}$$

(ii) $\xi_1 = O(\delta_n^2)$. Indeed we have $E\left(H_n|Z_i\right) = E\left(H_n|Z_j\right) = 0$ and

$$E(H_n|Z_l) = u_l E\left(u_i f_i K_{njl}K_{nij}w_{nij}|Z_l\right) = \delta_n u_l E\left(d_i f_i K_{njl} K_{nij} w_{nij}|Z_l\right).\text{Then}$$

$$\begin{aligned}
E\left[E^2(H_n|Z_l)\right] &= \delta_n^2 E\left[u_l^2 E^2\left(d_i f_i K_{njl}K_{nij}w_{nij}|Z_l\right)\right]\\
&= \delta_n^2 E\left[u_l^2 E^2\left(K_{njl}E(d_i f_i K_{nij} w_{nij}|Z_j,Z_l)|Z_l\right)\right]\\
&= O(\delta_n^2).
\end{aligned}$$

(iii) $E[H_n] = 0$. Thus $\xi_0 = 0$.

(iv) $\xi_3 = O(h^{-2p})$, as

$$
\begin{aligned}
E\left[H_n^2\right] &= E\left[u_i^2 u_l^2 f_i^2 K_{njl}^2 K_{nij}^2 w_{nij}^2\right] \\
&= O(h^{-2p}) E\left[u_i^2 u_l^2 f_i^2 \boldsymbol{K}_{njl} \boldsymbol{K}_{nij} w_{nij}^2\right] = O(h^{-2p}).
\end{aligned}
$$

Similarly to Proposition 1, it is easy to show that $E\left[H_{n,2}^4\right] = O(h^{-3p})$ and $E\left[G_n^2\right] = O(h^{-p})$. Thus Lemma 1 shows that

$$
nh^{p/2} U_{1n} = nh^{p/2} (1/n^{(3)}) \sum_a \left[H_n(Z_i, Z_j, Z_l) - E(H_n|Z_l)\right],
$$

is asymptotically normal with variance $2E\left[\sigma_C^2(X)\sigma_{C'}^2(X) f^2(X) f_C^2(X)\right] \int (K * K)^2(t)\, dt$ and zero mean. As $E\left[nh^{p/2}\left(I_{1,3} - U_{1n}\right)\right]^2 = \delta_n^2 nh^p O(1)$, Proposition 2 follows. Q.E.D.

6.1.5 Behavior of $I_{1,5}$

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

Proof: We have $I_{1,5} = (1/n^{(4)}) \sum_a u_k u_l K_{nik} K_{njl} K_{nij} w_{nij}$ which is a U-statistic with kernel

$$
H_n(Z_i, Z_j, Z_k, Z_l) = u_k u_l K_{nik} K_{njl} K_{nij} w_{nij}.
$$

We now compute the corresponding ξ_s, $s = 0, 1, 2, 3, 4$.

(i) $\xi_3 = O(h^{-2p})$. Indeed we have

$$
\begin{aligned}
E(H_n|Z_i, Z_j, Z_k) &= u_k K_{nik} K_{nij} w_{nij} E\left(u_l K_{njl}|Z_j\right) = 0, \\
E(H_n|Z_i, Z_j, Z_l) &= u_l K_{njl} K_{nij} w_{nij} E\left(u_k K_{nik}|Z_i\right) = 0, \\
E(H_n|Z_i, Z_k, Z_l) &= u_k u_l K_{nik} E\left(K_{njl} K_{nij} w_{nij}|Z_i, Z_l\right), \\
E(H_n|Z_j, Z_k, Z_l) &= u_k u_l K_{njl} E\left(K_{nik} K_{nij} w_{nij}|Z_j, Z_k\right).
\end{aligned}
$$

Then,

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

(ii) $\xi_2 = h^{-p}2E\left[\sigma^2_C(X)\sigma^2_{C'}g^4(X)\right]\int(K*K*K)^2(t)\,dt + o_p(h^{-p})$. 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

$$E(H_n|Z_k,Z_l) = u_k u_l E\left(K_{nik}K_{njl}K_{nij}w_{nij}|Z_k,Z_l\right),$$

so that

$$\begin{aligned} E\left[E^2(H_n|Z_k,Z_l)\right] &= E\left[u_k^2u_l^2E^2\left(K_{nik}K_{njl}K_{nij}w_{nij}|Z_k,Z_l\right)\right] \\ &= E\left[\sigma_k^2\sigma_l^2E^2\left(K_{nik}K_{njl}K_{nij}w_{nij}|Z_k,Z_l\right)\right]. \end{aligned}$$

Now

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

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

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

where $g^2(x) = \sum_c p_cf_c^2(x)$.

(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) $E[H_n] = 0$. Thus $\xi_0 = 0$.

(v) $\xi_4 = O(h^{-3p})$, as

$$\begin{aligned} E\left[H_n^2\right] &= E\left[u_k^2u_l^2K_{nik}^2K_{njl}^2K_{nij}^2w_{nij}^2\right] \\ &= O(h^{-3p})E\left[u_k^2u_l^2\boldsymbol{K}_{nik}\boldsymbol{K}_{njl}\boldsymbol{K}_{nij}w_{nij}^2\right] = O(h^{-3p}). \end{aligned}$$

Similarly to Proposition 1, it is easy to show that $E\left[H_{n,2}^4\right] = O(h^{-3p})$ and $E\left[G_n^2\right] = O(h^{-p})$. Thus Proposition 3 follows. Q.E.D.

6.1.6 Behavior of $V_{0n} - 2I_{1,3} + I_{1,5}$

Proposition 4 *:* $nh^{p/2}\left(V_{0n} - 2I_{1,3} + I_{1,5}\right) \xrightarrow{d} N(0, \omega^2)$.

PROOF: To determine the asymptotic distribution of $U_{0n} - 2U_{1n} + I_{1,5}$, we apply the Cramer-Wold device. We compute the covariances between U_{0n}, U_{1n} and $I_{1,5}$ by using (6.1) and the formula given at the end of Section 6.1.2.

Covariance between U_{0n} and U_{1n}: In this case $\xi_{2,2}$ is determined by

$$
\begin{aligned}
& E\left[u_i f_i u_l f_l K_{nil} w_{nil} E\left(u_i f_i u_l K_{njl} K_{nij} w_{nij} | Z_i, Z_l\right)\right] \\
&\quad = E\left[u_i^2 f_i^2 u_l^2 f_l K_{nil} w_{nil} E\left(K_{njl} K_{nij} w_{nij} | Z_i, Z_l\right)\right] \\
&\quad = E\left[\sigma_i^2 f_i^2 \sigma_l^2 f_l K_{nil} w_{nil} E\left(K_{njl} K_{nij} w_{nij} | Z_i, Z_l\right)\right] \\
&\quad = h^{-p} E\left[\sigma_C^2(X)\sigma_{C'}^2(X) w_{CC'} f^3(X) f_C(X)\right] \int K(t)(K*K)(t)\, dt + o(h^{-p}).
\end{aligned}
$$

Thus $\text{Cov}(nh^{p/2}U_{0n}, nh^{p/2}U_{1n}) \to 2E\left[\sigma_C^2(X)\sigma_{C'}^2(X) w_{CC'} f^3(X) f_C(X)\right] \int K(t)(K*K)(t)\, dt$.

Covariance between U_{0n} and $I_{1,5}$: In this case $\xi_{2,2}$ is determined by

$$
\begin{aligned}
& E\left[u_k f_k u_l f_l K_{nkl} w_{nkl} E\left(u_k u_l K_{nik} K_{njl} K_{nij} w_{nij} | Z_k, Z_l\right)\right] \\
&\quad = E\left[u_k^2 f_k u_l^2 f_l K_{nkl} w_{nkl} E\left(K_{nik} K_{njl} K_{nij} w_{nij} | Z_k, Z_l\right)\right] \\
&\quad = E\left[\sigma_k^2 f_k \sigma_l^2 f_l K_{nkl} w_{nkl} E\left(K_{nik} K_{njl} K_{nij} w_{nij} | Z_k, Z_l\right)\right] \\
&\quad = h^{-p} E\left[\sigma_C^2(X)\sigma_{C'}^2(X) w_{CC'} f^2(X) g^2(X)\right] \int K(t)(K*K*K)(t)\, dt + o(h^{-p}).
\end{aligned}
$$

Thus $\text{Cov}(nh^{p/2}U_{0n}, nh^{p/2}I_{1,5}) \to 2E\left[\sigma_C^2(X)\sigma_{C'}^2(X) w_{CC'} f^2(X) g^2(X)\right] \int K(t)(K*K*K)(t)\, dt$.

Covariance between U_{1n} and $I_{1,5}$: In this case $\xi_{2,2}$ is determined by

$$
\begin{aligned}
& E\left[E\left(u_k f_k u_l K_{njl} K_{nkj} w_{nkj} | Z_k, Z_l\right) E\left(u_k u_l K_{nik} K_{nj'l} K_{nij'} w_{nij'} | Z_k, Z_l\right)\right] \\
&\quad = E\left[\sigma_k^2 f_k \sigma_l^2 E\left(K_{njl} K_{nkj} w_{nkj} | Z_k, Z_l\right) E\left(K_{nik} K_{nj'l} K_{nij'} w_{nij'} | Z_k, Z_l\right)\right] \\
&\quad = h^{-p} E\left[\sigma_C^2(X)\sigma_{C'}^2(X) f(X) f_C(X) g^2(X)\right] \int (K*K)(t)(K*K*K)(t)\, dt + o(h^{-p}).
\end{aligned}
$$

Similarly $\xi_{3,3} = O(h^{-p})$.
Thus $\mathrm{Cov}(nh^{p/2}U_{1n}, nh^{p/2}I_{1,5}) \to 2E\left[\sigma_C^2(X)\sigma_{C'}^2(X)f(X)f_C(X)g^2(X)\right]\int (K*K)(t)(K*K*K)(t)\,dt.$

Conclusion:

$$nh^{p/2}\left[V_{0n} - 2I_{1,3} + I_{1,5}\right] = L_n + \delta_n^2 nh^{p/2}\mu + \delta_n\sqrt{n}h^{p/2}O_p(1) + \delta_n^2 nh^{p/2}o_p(1),$$

where L_n is asymptotically $N(0,\omega^2)$. Moreover,

$$\begin{aligned}
\omega^2 &= \mathrm{Var}\left[nh^{p/2}(U_{0n} - 2U_{1n} + I_{1,5})\right] \\
&= 2E\left[\sigma_C^2(X)\sigma_{C'}^2(X)w_{CC'}^2 f^4(X)\right]\int K^2(t)\,dt + 2E\left[\sigma_C^2(X)\sigma_{C'}^2(X)g^4(X)\right]\int (K*K*K)^2(t)\,dt \\
&\quad + 8E\left[\sigma_C^2(X)\sigma_{C'}^2(X)f^2(X)f_C^2(X)\right]\int (K*K)^2(t)\,dt \\
&\quad - 8E\left[\sigma_C^2(X)\sigma_{C'}^2(X)w_{CC'}f^3(X)f_C(X)\right]\int K(t)(K*K)(t)\,dt \\
&\quad + 4E\left[\sigma_C^2(X)\sigma_{C'}^2(X)w_{CC'}f^2(X)g^2(X)\right]\int K(t)(K*K*K)(t)\,dt \\
&\quad - 8E\left[\sigma_C^2(X)\sigma_{C'}^2(X)f(X)f_C(X)g^2(X)\right]\int (K*K)(t)(K*K*K)(t)\,dt \\
&= 2E\left[\sigma_C^2(X)\sigma_{C'}^2 f^4(X)E_{CC'}(X)\right],
\end{aligned}$$

$$\text{with } E_{CC'}(X) = \int\left[K(t)w_{CC'} - 2(K*K)(t)\frac{f_c(X)}{f(X)} + (K*K*K)(t)\frac{g^2(X)}{f^2(X)}\right]^2 dt.$$

6.1.7 The remaining terms

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

PROOF: We have $(n-3)I_{1,6} = (1/n^{(3)})\sum_a u_i u_j K_{nik}K_{njk}K_{nij}w_{nij}$ which is a U-statistic with kernel

$$H_n(Z_i,Z_j,Z_k) = u_i u_j K_{nik}K_{njk}K_{nij}w_{nij}.$$

In order to use (6.2), we need to compute the corresponding ξ_s, $s = 0,1,2,3$.
(i) $\xi_2 = O(h^{-3p})$. Indeed we have

$$\begin{aligned}
E(H_n|Z_i,Z_j) &= u_i u_j K_{nij}w_{nij}E\left(K_{nik}K_{njk}|Z_i,Z_j\right), \\
E(H_n|Z_i,Z_k) &= u_i K_{nik}E\left(u_j K_{njk}K_{nij}w_{nij}|Z_i,Z_k\right) \\
&= \delta_n u_i K_{nik}E\left(d_j K_{njk}K_{nij}w_{nij}|Z_i,Z_k\right),
\end{aligned}$$

$$\begin{aligned} E(H_n|Z_j,Z_k) &= u_jK_{njk}E\left(u_iK_{nik}K_{nij}w_{nij}|Z_j,Z_k\right) \\ &= \delta_n u_jK_{njk}E\left(d_iK_{nik}K_{nij}w_{nij}|Z_j,Z_k\right). \end{aligned}$$

Then

$$\begin{aligned} E\left[E^2(H_n|Z_i,Z_j)\right] &= E\left[u_i^2u_j^2K_{nij}^2w_{nij}^2E^2\left(K_{nik}K_{njk}|Z_i,Z_j\right)\right] \\ &= O(h^{-3p})E\left[u_i^2u_j^2\boldsymbol{K}_{nij}w_{nij}^2E^2\left(\boldsymbol{K}_{nik}|Z_i,Z_j\right)\right] = O(h^{-3p}), \\ E\left[E^2(H_n|Z_i,Z_k)\right] &= \delta_n^2E\left[u_i^2K_{nik}^2E^2\left(d_jK_{njk}K_{nij}w_{nij}|Z_i,Z_k\right)\right] \\ &= O(\delta_n^2h^{-3p})E\left[u_i^2\boldsymbol{K}_{nik}E^2\left(|d_j|\boldsymbol{K}_{njk}w_{nij}|Z_i,Z_k\right)\right] = O(h^{-3p}), \\ E\left[E^2(H_n|Z_j,Z_k)\right] &= \delta_n^2E\left[u_j^2K_{njk}^2E^2\left(d_iK_{nik}K_{nij}w_{nij}|Z_j,Z_k\right)\right] = O(h^{-3p}). \end{aligned}$$

(ii) $\xi_1 = O(h^{-2p})$. Indeed we have

$$\begin{aligned} E(H_n|Z_i) &= u_iE\left(u_jK_{nik}K_{njk}K_{nij}w_{nij}|Z_i\right) = \delta_nu_iE\left(d_jK_{nik}K_{njk}K_{nij}w_{nij}|Z_i\right), \\ E(H_n|Z_j) &= u_jE\left(u_iK_{nik}K_{njk}K_{nij}w_{nij}|Z_j\right) = \delta_nu_jE\left(d_iK_{nik}K_{njk}K_{nij}w_{nij}|Z_j\right), \\ E(H_n|Z_k) &= E\left(u_iu_jK_{nik}K_{njk}K_{nij}w_{nij}|Z_k\right) = \delta_n^2E\left(d_id_jK_{nik}K_{njk}K_{nij}w_{nij}|Z_k\right). \end{aligned}$$

Then by successive application of Lemma 2

$$\begin{aligned} E\left[E^2(H_n|Z_i)\right] &= \delta_n^2E\left[u_i^2E^2\left(d_jK_{nik}K_{njk}K_{nij}w_{nij}|Z_i\right)\right] \\ &= O(\delta_n^2h^{-2p})E\left[u_i^2E^2\left(|d_j|\boldsymbol{K}_{nik}\boldsymbol{K}_{njk}w_{nij}|Z_i\right)\right] = O(h^{-2p}), \\ E\left[E^2(H_n|Z_j)\right] &= \delta_n^2E\left[u_j^2E^2\left(d_iK_{nik}K_{njk}K_{nij}w_{nij}|Z_j\right)\right] = O(h^{-2p}), \\ E\left[E^2(H_n|Z_i)\right] &= \delta_n^4E\left[E^2\left(d_id_jK_{nik}K_{njk}K_{nij}w_{nij}|Z_k\right)\right] = O(h^{-2p}). \end{aligned}$$

(iii) $E\left[H_n\right] = E\left[u_iu_jK_{nik}K_{njk}K_{nij}w_{nij}\right] = \delta_n^2E\left[d_id_jK_{nik}K_{njk}K_{nij}w_{nij}\right] = O(\delta_n^2h^{-p})$.
(iv) $\xi_3 = O(h^{-4p})$, as

$$\begin{aligned} E\left[H_n^2\right] &= E\left[u_i^2u_j^2K_{nik}^2K_{njk}^2K_{nij}^2w_{nij}^2\right] \\ &= O(h^{-4p})E\left[u_i^2u_j^2\boldsymbol{K}_{nik}\boldsymbol{K}_{njk}w_{nij}^2\right] = O(h^{-4p}). \end{aligned}$$

Collecting results, $E(nh^{p/2}I_{1,6})^2 = O(\delta_n^2nh^{p/2})^2(nh^p)^{-2} + O(nh^p)^{-1} + O(nh^p)^{-2} + O(nh^p)^{-3}$. Q.E.D.

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

PROOF: We have $I_{2,1} = (1/n^{(3)}) \sum_a u_i f_i (r_j - r_l) K_{njl} K_{nij} w_{nij}$, which is a U-statistic with kernel

$$H_n(Z_i, Z_j, Z_l) = u_i f_i (r_j - r_l) K_{njl} K_{nij} w_{nij}.$$

In order to use (6.2), we need to compute the corresponding ξ_s, $s = 0, 1, 2, 3$.

(i) $\xi_2 = o(h^{-p})$. Indeed we have

$$\begin{aligned} E(H_n|Z_i, Z_j) &= u_i f_i K_{nij} w_{nij} E\left((r_j - r_l) K_{njl}|Z_j\right), \\ E(H_n|Z_i, Z_l) &= u_i f_i E\left((r_j - r_l) K_{njl} K_{nij} w_{nij}|Z_i, Z_l\right), \\ E(H_n|Z_j, Z_l) &= (r_j - r_l) K_{njl} E\left(u_i f_i K_{nij} w_{nij}|Z_j\right) \\ &= \delta_n (r_j - r_l) K_{njl} E\left(d_i f_i K_{nij} w_{nij}|Z_j\right). \end{aligned}$$

Then using the fact that $E\left[(r_j - r_l) K_{njl}|Z_j\right] = O(h^{(m+q)}) = o(1)$ uniformly in Z_j,

$$\begin{aligned} E\left[E^2(H_n|Z_i, Z_j)\right] &= E\left[u_i^2 f_i^2 K_{nij}^2 E^2\left((r_j - r_l) K_{njl} w_{nij}|Z_j\right)\right] \\ &= O(h^{-p}) E\left[u_i^2 f_i^2 \boldsymbol{K}_{nij} E^2\left((r_j - r_l) K_{njl} w_{nij}|Z_j\right)\right] = o(h^{-p}). \end{aligned}$$

Also we have

$$\begin{aligned} E\left[E^2(H_n|Z_i, Z_l)\right] &= E\left[u_i^2 f_i^2 E^2\Big((r_j - r_l) K_{njl} K_{nij} w_{nij}|Z_i, Z_l\Big)\right] \\ &= E\left[u_i^2 f_i^2 E\Big((r_j - r_l) K_{njl} K_{nij} w_{nij}|Z_i, Z_l\Big) E\left((r_{j'} - r_l) K_{nj'l} K_{nij'} w_{nij'}|Z_i, Z_l\right)\right] \\ &= O(h^{-p}) E\left[u_i^2 f_i^2 E\Big(|r_j - r_l| \boldsymbol{K}_{njl} \boldsymbol{K}_{nij} w_{nij}|Z_i, Z_l\Big) E\left(|r_{j'} - r_l| \boldsymbol{K}_{nij'} w_{nij'}|Z_i, Z_l\right)\right] \\ &= O(h^{-p}) E\left[u_i^2 f_i^2 |r_j - r_l| \boldsymbol{K}_{njl} \boldsymbol{K}_{nij} w_{nij} |r_i - r_l| f_{C_i}(X_i)\right] = o(h^{-p}), \\ E\left[E^2(H_n|Z_j, Z_l)\right] &= \delta_n^2 E\left[(r_j - r_l)^2 K_{njl}^2 E^2\left(d_i f_i K_{nij} w_{nij}|Z_j\right)\right] \\ &= \delta_n^2 E\left[(r_j - r_l)^2 K_{njl}^2 d_j^2 f_j^2 f_{C_j}^2(X_j)\right] \\ &= O(\delta_n^2 h^{-p}) E\left[(r_j - r_l)^2 \boldsymbol{K}_{njl} d_j^2 f_j^2 f_{C_j}^2(X_j)\right] = o(h^{-p}). \end{aligned}$$

(ii) $\xi_1 = O(h^{2(m+q)}) + o(\delta_n^2)$. Indeed

$$\begin{aligned} E\left[E^2(H_n|Z_i)\right] &= E\left[u_i^2 f_i^2 E^2\Big((r_j - r_l) K_{njl} K_{nij} w_{nij}|Z_i\Big)\right] \\ &= E\left[u_i^2 f_i^2 E^2\Big(K_{nij} w_{nij} E\Big((r_j - r_l) K_{njl}|Z_j\Big)|Z_i\Big)\right] \\ &= O(h^{2(m+q)}) E\left[u_i^2 f_i^2 E^2(\boldsymbol{K}_{nij} w_{nij})\right] = O(h^{2(m+q)}), \\ E\left[E^2(H_n|Z_j)\right] &= E\left[E^2\left(u_i f_i (r_j - r_l) K_{njl} K_{nij} w_{nij}|Z_j\right)\right] = O(h^{(m+q)}), \\ E\left[E^2(H_n|Z_l)\right] &= E\left[E^2\left(u_i f_i (r_j - r_l) K_{njl} K_{nij} w_{nij}|Z_l\right)\right] \end{aligned}$$

$$
\begin{aligned}
&= E\left[E^2\Big((r_j - r_l)K_{njl}E\left(u_i f_i K_{nij} w_{nij}|Z_j\right)|Z_l\Big)\right] \\
&= \delta_n^2 E\left[E^2\Big((r_j - r_l)K_{njl} d_j f_j f_{C_j}(X_j)|Z_l\Big)\right] = o(\delta_n^2).
\end{aligned}
$$

(iii)

$$
\begin{aligned}
E(H_n) &= E\left[u_i f_i (r_j - r_l)K_{njl}K_{nij}w_{nij}\right] \\
&= E\left[(r_j - r_l)K_{njl}E(u_i f_i K_{nij} w_{nij}|Z_j)\right] \\
&= \delta_n E\left[(r_j - r_l)K_{njl} d_j f_j f_{C_j}(X_j)\right] \\
&= \delta_n E\left[d_j f_j f_{C_j}(X_j) E\Big((r_j - r_l)K_{njl}|Z_j\Big)\right] \\
&= O(\delta_n h^{(m+q)}) E\left[d_j f_j f_{C_j}(X_j)\right] = O(\delta_n h^{(m+q)}).
\end{aligned}
$$

(iv) $\xi_3 = o(h^{-2p})$, as $E\left[H_n^2\right] = E\left[u_i^2 f_i^2 (r_j - r_l)^2 K_{njl}^2 K_{nij}^2 w_{nij}^2\right] = o(h^{-2p})$.

Collecting results, $E(nh^{p/2}I_{2,1})^2 = \delta_n^2 n^2 h^p O(h^{2(m+q)}) + nh^p O(h^{2(m+q)}) + o(\delta_n^2 nh^p) + o(1) + o(nh^p)^{-1}$. Q.E.D.

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

PROOF: We have $I_{2,3} = (1/n^{(4)}) \sum_a u_k (r_j - r_l) K_{nik} K_{njl} K_{nij} w_{nij}$, which is a U-statistic with kernel

$$
H_n(Z_i, Z_j, Z_k, Z_l) = u_k (r_j - r_l) K_{nik} K_{njl} K_{nij} w_{nij}.
$$

In order to use (6.2), we need to compute the corresponding ξ_s, $s = 0, 1, 2, 3, 4$.

(i) $\xi_3 = o(h^{-2p})$. Indeed we have

$$
\begin{aligned}
E(H_n|Z_i, Z_j, Z_k) &= u_k K_{nik} K_{nij} w_{nij} E\left((r_j - r_l)K_{njl}|Z_j\right), \\
E(H_n|Z_i, Z_j, Z_l) &= (r_j - r_l) K_{njl} K_{nij} w_{nij} E\left(u_k K_{nik}|Z_i\right) = 0, \\
E(H_n|Z_i, Z_k, Z_l) &= u_k K_{nik} E\left((r_j - r_l) K_{njl} K_{nij} w_{nij}|Z_i, Z_l\right), \\
E(H_n|Z_j, Z_k, Z_l) &= u_k (r_j - r_l) K_{njl} E\left(K_{nik} K_{nij} w_{nij}|Z_j, Z_k\right).
\end{aligned}
$$

Then

$$
E\left[E^2(H_n|Z_i, Z_j, Z_k)\right] = E\left[u_k^2 K_{nik}^2 K_{nij}^2 w_{nij}^2 E^2\Big((r_j - r_l)K_{njl}|Z_j\Big)\right] = o(h^{-2p}),
$$

$$
\begin{aligned}
&E\left[E^2(H_n|Z_i, Z_k, Z_l)\right] \\
&= E\left[u_k^2 K_{nik}^2 E^2\Big((r_j - r_l) K_{njl} K_{nij} w_{nij}|Z_i, Z_l\Big)\right] \\
&= E\left[u_k^2 K_{nik}^2 E\Big((r_j - r_l) K_{njl} K_{nij} w_{nij}|Z_i, Z_l\Big) E\Big((r_{j'} - r_l) K_{nj'l} K_{nij'} w_{nij'}|Z_i, Z_l\Big)\right]
\end{aligned}
$$

$$=O(g^{-p_1})E\left[u_k^2K_{nik}^2E\Big(|r_j-r_l|\boldsymbol{K}_{njl}\boldsymbol{K}_{nij}w_{nij}|Z_i,Z_l\Big)E\Big(|r_{j'}-r_l|\boldsymbol{K}_{nij'}w_{nij'}|Z_i,Z_l\Big)\right]$$
$$=O(h^{-2p})E\left[u_k^2\boldsymbol{K}_{nik}|r_j-r_l|\boldsymbol{K}_{njl}\boldsymbol{K}_{nij}|r_i-r_l|f_{C_i}(X_i)\right]=o(h^{-2p}),$$

$$E\left[E^2(H_n|Z_j,Z_k,Z_l)\right]=E\left[u_k^2(r_j-r_l)^2K_{njl}^2E^2\left(K_{nik}K_{nij}w_{nij}|Z_j,Z_k\right)\right]=o(h^{-2p}).$$

(ii) $\xi_2=o(h^{-p})$. Indeed we have $E(H_n|Z_i,Z_j)=E(H_n|Z_i,Z_l)=E(H_n|Z_j,Z_l)=0$,

$$\begin{aligned}E(H_n|Z_i,Z_k)&=u_kK_{nik}E\left((r_j-r_l)K_{njl}K_{nij}w_{nij}|Z_i\right),\\ E(H_n|Z_j,Z_k)&=u_kE\left(K_{nik}(r_j-r_l)K_{njl}K_{nij}w_{nij}|Z_j,Z_k\right),\\ E(H_n|Z_k,Z_l)&=u_kE\left(K_{nik}(r_j-r_l)K_{njl}K_{nij}w_{nij}|Z_k,Z_l\right).\end{aligned}$$

Then

$$\begin{aligned}E\left[E^2(H_n|Z_i,Z_k)\right]&=E\left[u_k^2K_{nik}^2E^2\Big((r_j-r_l)K_{njl}K_{nij}w_{nij}|Z_i\Big)\right]\\ &=E\left[u_k^2K_{nik}^2E^2\Big(K_{nij}w_{nij}E\Big((r_j-r_l)K_{njl}|Z_j\Big)|Z_i\Big)\right]\\ &=o(1)E\left[u_k^2K_{nik}^2E^2\Big(\boldsymbol{K}_{nij}w_{nij}|Z_i\Big)\right]\\ &=o(h^{-p})E\left[u_k^2\boldsymbol{K}_{nik}f_{C_i}^2(X_i)\right]=o(h^{-p}),\\ E\left[E^2(H_n|Z_j,Z_k)\right]&=E\left[u_k^2E^2\Big(K_{nik}(r_j-r_l)K_{njl}K_{nij}w_{nij}|Z_j,Z_k\Big)\right]\\ &=E\left[u_k^2E^2\Big(K_{nik}K_{nij}w_{nij}E\Big((r_j-r_l)K_{njl}|Z_j\Big)|Z_j,Z_k\Big)\right]\\ &=o(1)E\left[u_k^2E\Big(\boldsymbol{K}_{nik}\boldsymbol{K}_{nij}w_{nij}|Z_j,Z_k\Big)E\Big(\boldsymbol{K}_{ni'k}\boldsymbol{K}_{ni'j}w_{ni'j}|Z_j,Z_k\Big)\right]\\ &=o(h^{-p})E\left[u_k^2E\Big(\boldsymbol{K}_{nik}\boldsymbol{K}_{nij}w_{nij}|Z_j,Z_k\Big)E\Big(\boldsymbol{K}_{ni'j}w_{ni'j}|Z_j,Z_k\Big)\right]\\ &=o(h^{-p})E\left[u_k^2\boldsymbol{K}_{nik}\boldsymbol{K}_{nij}f_{C_j}(X_j)\right]=o(h^{-p}),\end{aligned}$$

$$\begin{aligned}&E\left[E^2(H_n|Z_k,Z_l)\right]\\ &=E\left[u_k^2E^2\Big(K_{nik}(r_j-r_l)K_{njl}K_{nij}w_{nij}|Z_k,Z_l\Big)\right]\\ &=E\left[u_k^2E\Big(K_{nik}(r_j-r_l)K_{njl}K_{nij}w_{nij}|Z_k,Z_l\Big)E\Big(K_{ni'k}(r_{j'}-r_l)K_{nj'l}K_{ni'j'}w_{ni'j'}|Z_k,Z_l\Big)\right]\\ &=O(h^{-p})E\left[u_k^2E\Big(\boldsymbol{K}_{nik}|r_j-r_l|\boldsymbol{K}_{njl}\boldsymbol{K}_{nij}w_{nij}|Z_k,Z_l\Big)E\Big(|r_{j'}-r_l|\boldsymbol{K}_{nj'l}\boldsymbol{K}_{ni'j'}w_{ni'j'}|Z_k,Z_l\Big)\right]\\ &=o(h^{-p})E\left[u_k^2\boldsymbol{K}_{nik}|r_j-r_l|\boldsymbol{K}_{njl}\boldsymbol{K}_{nij}w_{nij}\right]=o(h^{-p}).\end{aligned}$$

(iii) $\xi_1=O(h^{2(m+q)})$. Indeed we have $E\left(H_n|Z_i\right)=E\left(H_n|Z_j\right)=E(H_n|Z_l)=0$ and

$$\begin{aligned}E\left[E^2(H_n|Z_k)\right]&=E\left[u_k^2E^2\Big(K_{nik}(r_j-r_l)K_{njl}K_{nij}w_{nij}|Z_k\Big)\right]\\ &=E\left[u_k^2E^2\Big(K_{nik}K_{nij}w_{nij}E\Big(r_j-r_l)K_{njl}|Z_j\Big)|Z_k\Big)\right]=O(h^{2(m+q)}).\end{aligned}$$

(iv) $E[H_n]=0$. Thus $\xi_0=0$.
(v) $\xi_4=o(h^{-3p})$, as $E[H_n^2]=E\left[u_k^2K_{nik}^2(r_j-r_l)^2K_{njl}^2K_{nij}^2w_{nij}^2\right]=o(h^{-3p})$.
Collecting results, $E(nh^{p/2}I_{2,3})^2=nh^pO(h^{2(m+q)})+o(1)+o(nh^p)^{-1}+o(nh^p)^{-2}$. Q.E.D.

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

Proof: We have $I_3=(1/n^{(4)})\sum_a(r_i-r_k)(r_j-r_l)K_{nik}K_{njl}K_{nij}w_{nij}$, which is a U-statistic with kernel

$$H_n(Z_i,Z_j,Z_k,Z_l)=(r_i-r_k)(r_j-r_l)K_{nik}K_{njl}K_{nij}w_{nij}.$$

In order to use (6.2), we need to compute the corresponding ξ_c, $c=0,1,2,3,4$. Similarly to the proof of Proposition 7 for $I_{2,3}$, we can show that $\xi_4=o(h^{-3p})$, $\xi_3=o(h^{-2p})$, $\xi_2=o(h^{-p})$, $\xi_1=o(h^{2(m+q)})$. On the other hand,

$$E[H_n]=E[(r_i-r_k)(r_j-r_l)K_{nik}K_{njl}K_{nij}w_{nij}]=O(h^{2(m+q)}),$$

so that $E\left(nh^{p/2}I_3\right)^2=n^2h^pO(h^{4(m+q)})+o(1)$. Q.E.D.

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

Proof: We denote $(\widehat{f}_i^j-f_i)$ by Δf_i^j. We have $I_{1,1}=(1/n^{(2)})\sum_a u_i\Delta f_i^j u_jf_jK_{nij}w_{nij}$ so that

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

where the first (respectively the second) sum is taken over all arrangements of different indices i and j (respectively different indices i' and j'). In what follows, $\tilde{K}_{nij}=K_{nij}w_{nij}$. We consider three situations.
(i) All indices are different: $n^{(4)}$ terms.

$$\begin{aligned}E\left[u_i\Delta f_i^ju_jf_j\tilde{K}_{nij}u_{i'}\Delta f_{i'}^{j'}u_{j'}f_{j'}\tilde{K}_{ni'j'}\right]&=E\left[\Delta f_i^jf_j\Delta f_{i'}^{j'}f_{j'}E\left(u_iu_ju_{i'}u_{j'}\tilde{K}_{nij}\tilde{K}_{ni'j'}|\bar{X}\right)\right]\\&=\delta_n^4E\left[\Delta f_i^jf_j\Delta f_{i'}^{j'}f_{j'}E\left(d_id_jd_{i'}d_{j'}\tilde{K}_{nij}\tilde{K}_{ni'j'}|\bar{X}\right)\right]\\&=\delta_n^4E\left[\Delta f_i^jf_j\Delta f_{i'}^{j'}f_{j'}d_id_jd_{i'}d_{j'}\tilde{K}_{nij}\tilde{K}_{ni'j'}\right]\\&=\delta_n^4E\left[f_jf_{j'}d_id_jd_{i'}d_{j'}\tilde{K}_{nij}\tilde{K}_{ni'j'}E\left(\Delta f_i^j\Delta f_{i'}^{j'}|Z_i,Z_j,Z_{i'},Z_{j'}\right)\right]\\&=\delta_n^4\lambda_nE\left|f_jf_{j'}d_id_jd_{i'}d_{j'}\tilde{K}_{nij}\tilde{K}_{ni'j'}\right|\\&=O(\delta_n^4\lambda_n),\end{aligned}$$

where $\bar{X}$ is the σ-algebra generated by all the (X_i, C_i) and $\lambda_n = E\left[\Delta^2 f_i^j | Z_i, Z_j, Z_{i'}, Z_{j'}\right] = o(1)$ uniformly by Lemma 3.

(ii) One index is common to $\{i,j\}$ and $\{i',j'\}$: $4n^{(3)}$ terms.

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

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

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

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

$$E\left[u_i^2 u_j^2 \left(\Delta f_i^j\right)^2 f_j^2 \tilde{K}_{nij}^2\right] = O(\lambda_n / h^p) \quad \text{and} \quad E\left[u_i^2 u_j^2 \Delta f_i^j \Delta f_j^i f_i f_j \tilde{K}_{nij}^2\right] = O(\lambda_n / h^p).$$

Therefore,

$$E\left(nh^{p/2} I_{1,1}\right)^2 = \delta_n^4 n^2 h^p O(\lambda_n) + \delta_n^2 n h^p O(\lambda_n) + O(\lambda_n).$$

The proposition then follows from $\lambda_n = o(1)$, see Lemma 3. Q.E.D.

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

Proof: The proof is very similar to the proof of Proposition 9 for $I_{1,1}$ and is not reported.

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

Proof: We denote $(\widehat{f}_i^{j,l} - f_i)$ by $\Delta f_i^{j,l}$. We have $I_{1,4} = (1/n^{(3)}) \sum_a u_i \Delta f_i^{j,l} u_l K_{njl} \tilde{K}_{nij}$, so that

$$E(I_{1,4}^2) = \left(\frac{1}{n^{(3)}}\right)^2 \left[\sum_a u_i \Delta f_i^{j,l} u_l K_{njl} \tilde{K}_{nij}\right] \left[\sum_a u_{i'} \Delta f_{i'}^{j',l'} u_{l'} K_{nj'l'} \tilde{K}_{ni'j'}\right],$$

where the first (respectively the second) sum is taken over all arrangements of pairwise different indices i, j and l (respectively pairwise different indices i', j' and l'). We consider four situations.
(i) All indices are different: $n^{(6)}$ terms.

$$\begin{aligned}
&E\left[u_i\Delta f_i^{j,l}u_lK_{njl}\tilde{K}_{nij}u_{i'}\Delta f_{i'}^{j',l'}u_{l'}K_{nj'l'}\tilde{K}_{ni'j'}\right]\\
&= E\left[\Delta f_i^{j,l}K_{njl}\Delta f_{i'}^{j',l'}K_{nj'l'}E(u_iu_lu_{i'}u_{l'}\tilde{K}_{nij}\tilde{K}_{ni'j'}|\bar{X})\right]\\
&= \delta_n^4E\left[\Delta f_i^{j,l}K_{njl}\Delta f_{i'}^{j',l'}K_{nj'l'}d_id_ld_{i'}d_{l'}\tilde{K}_{nij}\tilde{K}_{ni'j'}\right]\\
&= \delta_n^4E\left[K_{njl}K_{nj'l'}d_id_ld_{i'}d_{l'}\tilde{K}_{nij}\tilde{K}_{ni'j'}E\left(\Delta f_i^{j,l}\Delta f_{i'}^{j',l'}|Z_i,Z_j,Z_l,Z_{i'},Z_{j'},Z_{l'}\right)\right]\\
&= O(\delta_n^4\lambda_n),
\end{aligned}$$

where $\lambda_n = E\left[\Delta^2 f_i^{j,l}|Z_i,Z_j,Z_l,Z_{i'},Z_{j'},Z_{l'}\right] = o(1)$ uniformly by Lemma 3.
(ii) One index is common to $\{i,j,l\}$ and $\{i',j',l'\}$: $9n^{(5)}$ terms.

$$\begin{aligned}
(i'=i)\quad &E\left[u_i^2\Delta f_i^{j,l}u_lK_{njl}\tilde{K}_{nij}\Delta f_i^{j',l'}u_{l'}K_{nj'l'}\tilde{K}_{nij'}\right]\\
&= E\left[\Delta f_i^{j,l}K_{njl}\Delta f_i^{j',l'}K_{nj'l'}E(u_i^2u_lu_{l'}\tilde{K}_{nij}\tilde{K}_{nij'}|\bar{X})\right]\\
&= \delta_n^2E\left[\Delta f_i^{j,l}K_{njl}\Delta f_i^{j',l'}K_{nj'l'}u_i^2d_ld_{l'}\tilde{K}_{nij}\tilde{K}_{nij'}\right]\\
&= O(\delta_n^2\lambda_n),
\end{aligned}$$

Similar computations can be made for the cases $j=j'$, $l=l'$, $i=j'$ (or $j=i'$), $i=l'$ (or $l=i'$), $j=l'$ (or $l=j'$). The corresponding expectations are all $O(\delta_n^2\lambda_n)$.
(iii) Two indices are common to $\{i,j,l\}$ and $\{i',j',l'\}$: $18n^{(4)}$ terms.

$$\begin{aligned}
(i=i' \text{ and } j=j')\quad &E\left[u_i^2\Delta f_i^{j,l}u_lK_{njl}\Delta f_i^{j,l'}u_{l'}K_{njl'}\tilde{K}_{nij}^2\right]\\
&= E\left[\Delta f_i^{j,l}K_{njl}\Delta f_i^{j,l'}K_{njl'}E(u_i^2u_lu_{l'}\tilde{K}_{nij}^2|\bar{X})\right]\\
&= \delta_n^2E\left[\Delta f_i^{j,l}K_{njl}\Delta f_i^{j,l'}K_{njl'}u_i^2d_ld_{l'}\tilde{K}_{nij}^2\right]\\
&= O(\delta_n^2\lambda_n/h^p) = O(\lambda_n/h^p),
\end{aligned}$$

The other cases are: ($i=i'$ and $l=l'$), ($j=j'$ and $l=l'$), ($i=j'$ and $j=i'$), ($i=j'$ and $j=l'$) (or $i=l'$ and $l=j'$) (or $j=i'$ and $l=j'$) (or $j=l'$ and $l=i'$), ($i=j'$ and $l=i'$) (or $i=l'$ and $j=i'$), ($i=l'$ and $j=j'$) (or $l=i'$ and $j=j'$), ($i=i'$ and $j=l'$) (or $i=i'$ and $l=j'$), ($i=l'$ and $l=i'$), ($i=j'$ and $l=l'$) (or $j=i'$ and $l=l'$), ($j=l'$ and $l=j'$). It can be similarly checked that the corresponding expectations are all $O(\lambda_n/h^p)$.
(iv) Three indices are common to $\{i,j,l\}$ and $\{i',j',l'\}$: $6n^{(3)}$ terms.
For instance, if ($i=i'$, $j=j'$ and $l=l'$), $E\left[u_i^2\left(\Delta f_i^{j,l}\right)^2u_l^2K_{njl}^2\tilde{K}_{nij}^2\right] = O(\lambda_n/h^{2p})$. The remaining cases are: ($i=i'$ and $j=l'$ and $l=j'$), ($i=j'$ and $j=i'$ and $l=l'$), ($i=j'$ and $j=l'$ and $l=i'$), ($i=l'$ and $j=i'$ and $l=j'$), ($i=l'$ and $j=j'$ and $l=i'$). The corresponding expectations are all $O(\lambda_n/h^{2p})$.

Therefore,

$$E\left[nh^{p/2}I_{1,4}\right]^2 = \delta_n^4 n^2 h^p O(\lambda_n) + \delta_n^2 nh^p O(\lambda_n) + O(\lambda_n)(1 + (nh^p)^{-1}).$$

The proposition then follows from $\lambda_n = o(1)$, see Lemma 3. Q.E.D.

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

Proof: The proof is very similar to the proof of Proposition 11 for $I_{1,4}$ and is not reported.

6.2 Proof of Theorem 2

The proof of Theorem 2 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 and shown but one easily check that it holds by looking at the proofs of Lemma B.4 of Fan and Li (1996) and Theorem 1 of Hall (1984) (the latter proof relies on a martingale central limt theorem that still applies in this case). To deal with the remaining terms, one uses analogs of (6.1) and (6.2) for independent but not necessarily identically distributed random variables.

6.3 Panel data

The difficulty to adapt the proof of Theorem 2 comes from the fact that the observations may not be independent across subsamples. But under our assumptions, $u_{it} \equiv Y_{it} - r(X_{it})$ is independent of $\mathcal{Z}_{t-1} = \{Y_{i1}, \ldots, Y_{i,t-1}, X_{i1}, \ldots, X_{i,t-1}\}$ conditionally on $\{X_{it}, l\}$ and $E[u_{it}|\mathcal{Z}_{t-1}, X_{it}, l] = R_t(X_{it}) - r(X_{it})$, which is zero under H_0. 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 the terms $I_{1,1}$, $I_{1,2}$, $I_{1,4}$ and $I_{2,2}$.

6.4 Technical lemmas

Lemma 2 *For any function* $l \in \mathcal{U}^p$*,*

$$\sup_{x \in \mathbb{R}^p} \left| \int l(X) \frac{1}{h^p} K\left(\frac{x - X}{h}\right) dX - l(x) \right| \to 0.$$

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

Lemma 3 *: If the density $f_c(X)$ belongs to $\mathcal{U}^p \quad \forall c$, and $nh^p \to \infty$, $E\left[\Delta^2 f_i^j | Z_i, Z_j, Z_{i'}, Z_{j'}\right] =$* $o(1)$ and $E\left[\Delta^2 f_i^{j,l} | Z_i, Z_j, Z_l, Z_{i'}, Z_{j'}, Z_{l'}\right] = o(1)$, *where $\Delta f_i^j = \widehat{f}_i^j - f_i$ and $\Delta f_i^{j,l} = \widehat{f}_i^{j,l} - f_i$.*

PROOF: From the definition of Δf_i^j,

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

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, conditional on Z_i, independent with zero mean,

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

As $E\left(\widehat{f}_i^j | Z_i, Z_j, Z_{i'}, Z_{j'}\right) = (n-2)^{-1}\left[K_{nii'} + K_{nij'} + (n-4)E(K_{nik}|Z_i)\right]$,

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

where we use $f_i = \sum_c p_c f_c(X_i)$.

The proof for the second part is similar and is therefore 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

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

By conditioning upon (Z_i, Z_k), one can see that all the terms such that k and k' correspond to different individuals are zero. There are at most nT terms corresponding to same individuals. These terms are all $O(h^p)^{-1}$ by applying Cauchy-Schwartz inequality. Then the right-hand side is an $O(nh^p)^{-1}$ and Lemma 4 follows. Q.E.D.

REFERENCES

AÏT-SAHALIA, Y., P. BICKEL and T. M. STOKER (1994): "Goodness-of-fit tests for regression using kernel methods," University of Chicago.

BALTAGI, B.H., J. HIDALGO and Q. LI (1996): "A nonparametric test for poolability using panel data," *Journal of Econometrics*, 75, pp. 345–367.

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

CHAMBERLAIN, G. (1984): "Panel data," in *Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M.D. Intriligator. Amsterdam: North-Holland, Chap. 5, pp. 285–330.

CHOW, G.C. (1960): "Tests of equality between sets of coefficients in two linear regressions," *Econometrica*, 28(3), pp. 591–605.

DELGADO, M.A. (1993): "Testing the equality of nonparametric regresion curves," *Statistics and Probability Letters*, 17, pp. 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), pp. 1477–1484.

FAN, Y. and Q. LI (1996): "Consistent model specification tests: omitted variables and semiparametric functional forms," *Econometrica*, 64 (4), pp. 865–890.

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

HALL, P. (1984) : "Central limit theorem for integrated square error of multivariate nonparametric density estimators," *Journal of Multivariate Analysis*, 14, pp. 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, pp. 1039–1049.

HÄRDLE, W. and J.S. MARRON (1990): "Semiparametric comparison of regression curves," *Annals of Statistics*, 18 (1), pp. 63–89.

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

HOEFFDING, W. (1948): "A class of statistics with asymptotically normal distribution," *Annals of Mathematical Statistics*, 15, pp. 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, pp. 239–247.

KOUL, H.L. and A. SCHICK (1996): "Testing for the equality of two nonparametric regression curves," Michigan State University.

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

LAVERGNE, P. and Q.H. VUONG (1995): "Nonparametric significance testing," INRA-ESR Toulouse.

LAVERGNE, P. and Q.H. VUONG (1996): "Nonparametric selection of regressors: the nonnested case," *Econometrica*, 64(1), pp. 207–219.

POWELL, J.L., H. STOCK and T.M. STOKER (1989): "Semiparametric estimation of index coefficients," *Econometrica*, 57, pp. 1403–1430.

QUADE, D. (1982): "Nonparametric analysis of covariance," *Biometrics*, 38, pp. 597–611.

ROBINSON, P. (1988): "Root-N consistent semiparametric regression," *Econometrica*, 56, pp. 931–954.

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

YOUNG, S. G. and A. W. BOWMAN (1995): "Non-parametric analysis of covariance," *Biometrics*, 51, pp. 920–931.

Table 1: Null and Linear Alternatives

n	a	DGP_0		DGP_1		DGP_2		DGP_3	
100	0.2	0.042	(0.536)	0.112	(0.566)	0.311	(0.642)	0.864	(0.782)
		1.2%	2.6%	1.6%	4.2%	3.9%	8.8%	16.0%	29.5%
	0.5	0.027	(0.698)	0.237	(0.788)	0.796	(0.968)	2.189	(1.178)
		2.9%	5.9%	5.7%	10.1%	19.0%	28.7%	67.3%	77.9%
	1.0	0.059	(0.767)	0.452	(0.959)	1.438	(1.295)	3.781	(1.631)
		3.4%	7.0%	11.1%	17.5%	39.8%	50.4%	90.2%	93.9%
	1.5	0.184	(0.833)	0.708	(1.091)	1.998	(1.529)	5.039	(1.966)
		5.2%	9.9%	17.6%	25.4%	54.9%	63.4%	96.3%	97.8%
	2.0	0.607	(0.909)	1.225	(1.230)	2.715	(1.727)	6.219	(2.206)
		12.4%	19.3%	30.8%	41.0%	70.5%	77.1%	98.7%	99.4%
	Chow test	4.4%	9.6%	51.2%	64.6%	98.7%	99.4%	100.0%	100.0%
250	0.2	0.006	(0.647)	0.280	(0.729)	1.023	(0.916)	2.928	(1.096)
		1.6%	3.5%	4.9%	9.5%	24.1%	35.6%	87.2%	93.7%
	0.5	-0.013	(0.756)	0.593	(0.952)	2.185	(1.326)	6.043	(1.694)
		2.9%	5.8%	13.9%	22.1%	62.5%	73.2%	99.8%	99.9%
	1.0	0.038	(0.823)	1.035	(1.208)	3.588	(1.829)	9.635	(2.437)
		4.2%	7.7%	26.0%	36.6%	86.2%	91.1%	100.0%	100.0%
	1.5	0.188	(0.836)	1.491	(1.423)	4.743	(2.239)	12.408	(2.992)
		5.7%	9.9%	38.9%	47.6%	93.8%	96.0%	100.0%	100.0%
	2.0	0.706	(0.890)	2.250	(1.619)	6.022	(2.549)	14.925	(3.374)
		13.3%	22.0%	58.6%	69.3%	98.0%	98.9%	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 mean of the test statistic with its standard deviation in parentheses on the first line, and empirical levels at 5% and 10% nominal levels on the second line.

Table 2: Sinus Alternatives

n	a	DGP_4		DGP_5		DGP_6		DGP_7	
100	0.2	0.327	(0.647)	0.317	(0.641)	0.326	(0.647)	0.312	(0.645)
		3.4%	9.2%	3.8%	8.8%	4.2%	8.6%	4.0%	8.3 %
	0.5	0.744	(0.936)	0.749	(0.994)	0.803	(1.015)	0.779	(0.996)
		16.9%	25.6%	18.1%	27.3 %	20.3%	29.3%	19.4%	28.4%
	1.0	1.134	(1.071)	1.239	(1.254)	1.404	(1.342)	1.387	(1.324)
		28.6%	41.1%	33.3%	43.5%	37.3%	47.9%	38.0%	47.2%
	1.5	1.215	(1.043)	1.536	(1.420)	1.863	(1.553)	1.890	(1.542)
		30.0%	43.5%	40.9%	50.1 %	49.6%	59.2%	51.0%	60.4%
	2.0	0.951	(0.962)	1.897	(1.462)	2.425	(1.709)	2.526	(1.720)
		21.2%	31.1%	51.4%	61.6 %	63.3%	71.6%	66.3%	74.3%
	Chow test	4.8%	9.5%	14.8%	21.2%	74.9%	82.3%	83.8%	90.9%
250	0.2	1.003	(0.932)	1.040	(0.938)	1.068	(0.944)	1.017	(0.926)
		23.7%	36.3%	24.0%	36.7%	25.6%	38.1%	24.3%	35.5%
	0.5	2.034	(1.331)	2.181	(1.399)	2.276	(1.415)	2.171	(1.368)
		59.0%	68.8%	61.7%	71.3%	64.8%	74.4%	61.8%	71.6%
	1.0	3.004	(1.552)	3.386	(1.818)	3.666	(1.910)	3.542	(1.863)
		80.8%	87.0%	82.9%	89.1%	85.8%	90.7%	84.9%	90.2%
	1.5	3.385	(1.490)	4.111	(2.070)	4.695	(2.308)	4.622	(2.271)
		88.8%	93.4%	90.1%	94.3%	92.5%	95.3%	93.0%	95.4%
	2.0	3.210	(1.283)	4.754	(2.218)	5.727	(2.593)	5.771	(2.573)
		90.7%	95.4%	94.7%	97.4%	96.6%	98.1%	97.1%	98.5%
	Chow test	5.5%	10.1%	13.4%	21.0%	93.3%	95.9%	99.6%	100.0%

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

Table 3: Null and Linear Alternatives (n=250)

a	DGP_0		DGP_1		DGP_2		DGP_3	
$p_0 = 0.2$								
1.0	0.065	(0.559)	0.460	(0.694)	1.267	(0.823)	2.410	(0.889)
	0.9%	2.9%	6.0%	12.1%	30.2%	46.0%	81.2%	91.9%
1.5	0.161	(0.582)	0.680	(0.783)	1.697	(0.951)	3.067	(1.050)
	1.8%	4.3%	11.3%	20.2%	49.7%	64.6%	93.3%	97.9%
2.0	0.570	(0.669)	1.209	(0.939)	2.364	(1.119)	3.828	(1.217)
	6.7%	13.8%	28.2%	41.3%	73.1%	84.4%	98.4%	99.5%
$\sigma_1^2 = 2$								
1.0	0.012	(0.770)	0.375	(0.918)	1.399	(1.264)	4.838	(2.041)
	3.5%	6.5%	9.5%	14.7%	37.2%	48.3%	96.3%	97.9%
1.5	0.058	(0.767)	0.532	(0.999)	1.838	(1.471)	6.203	(2.471)
	3.8%	6.8%	12.3%	18.5%	48.8%	59.7%	98.7%	99.2%
2.0	0.251	(0.780)	0.814	(1.088)	2.335	(1.650)	7.406	(2.803)
	5.3%	8.9%	18.7%	27.5%	61.6%	71.2%	99.5%	99.7%

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