INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2007-01

Nonparametric Lack-of-fit Tests for Parametric Mean-Regression Model with Censored Data

O. LOPEZ[1]
V. PATILEA[2]

[1] CREST-ENSAI and IRMAR. Email : lopez@ensai.fr

[2] CREST-ENSAI. Email : patilea@ensai.fr

Address for correspondence : Campus de Ker Lann, rue Blaise Pascal – BP 37203 – 35172 Bruz Cedex, France.

Nonparametric lack-of-fit tests for parametric mean-regression models with censored data

O. Lopez * and V. Patilea†

March, 2007

*CREST-ENSAI and IRMAR. Email: lopez@ensai.fr.

†CREST-ENSAI. Email: patilea@ensai.fr. Address for correspondence: Campus de Ker Lann, Rue Blaise Pascal BP 37203, 35172 Bruz cedex, France

Abstract

We develop two kernel smoothing based tests of a parametric mean-regression model against a nonparametric alternative when the response variable is right-censored. The new test statistics are inspired by the synthetic data and the weighted least squares approaches for estimating the parameters of a (non)linear regression model under censoring. The asymptotic critical values of our tests are given by the quantiles of the standard normal law. The tests are consistent against fixed alternatives, local Pitman alternatives and uniformly over alternatives in Hölder classes of functions of known regularity.

Key words: Hypothesis testing, censored data, Kaplan-Meier integral, local alternative

MSC 2000: 62G10, 62G08, 62N01

Résumé

Nous développons deux procédures de test d'un modèle de régression paramétrique contre une alternative non paramétrique, dans le cas où la variable expliquée est censurée aléatoirement à droite. Ces deux procédures sont basées sur un lissage par noyau. Ces nouvelles statistiques de test sont inspirées des approches "synthetic data" et moindres carrés pondérés utilisées pour l'estimation dans des modèles de régression en présence de censure. Les valeurs critiques asymptotiques de nos tests sont données par les quantiles d'une loi normale centrée réduite. Les tests sont consistants contre des alternatives fixes, des alternatives locales de type Pitman, et uniformément sur des alternatives appartenant à des classes de Hölder de régularité connue.

Mots clés : Test d'hypothèses, données censurées, intégrale Kaplan-Meier, alternative locale

1 Introduction

Parametric mean-regression models, in particular the linear model, are valuable tools for exploring the relationship between a response and a set of explanatory variables (covariates). However, in survival analysis such models are overshadowed by the fashionable proportional hazard models and the accelerated failure time models where one imposes a form for the conditional law of the response given the covariates. Even though mean-regression models involve weaker assumptions on the conditional law of the responses, the popularity of the parametric mean-regressions with censored data greatly suffers from the difficulty to perform statistical inference when not all responses are available.

The existing methods for the estimation of the parameters of the mean-regression in the presence of right censoring can be split into two main categories: i) *weighted least squares (WLS)* based on the uncensored observations but suitably weighted to account for censorship (see Zhou 1992a, Stute 1999); and ii) *synthetic data (SD)* estimators obtained by ordinary least squares with transformed responses, using a transformation that preserves the conditional expectation and that can be estimated from data (e.g., Buckley and James 1979, Koul *et al.* 1981, Leurgans 1987).

This paper's main purpose focuses on a further step in the statistical inference for parametric mean-regression models under right censoring, that is nonparametric lack-of-fit testing. Checking the adequacy of a parametric regression function against a purely nonparametric alternative has received a large amount of attention in the non-censored case and several approaches have been proposed. See, amongst many others, Härdle and Mammen (1993), Zheng (1996), Stute (1997), Dette (1999), Horowitz and Spokoiny (2001), Fan and Huang (2001), Guerre and Lavergne (2005), and the references therein. But for right-censored data, these approaches are not directly applicable. To our knowledge, very few solutions for nonparametric regression checks with right-censored responses have been proposed. Following the approach of Stute (1997), Stute *et al.* (2000) introduced two tests based on an empirical process marked by *weighted* residuals, the role of the weights being to account for censoring. The limit of their marked empirical process is a rather complicated centered Gaussian process and therefore the implementation of the test requires numerical calculations. Sánchez-Sellero *et al.* (2005) reconsidered this type of test and provided a complete proof of its asymptotic level. However, for technical reasons, Sánchez-Sellero *et al.* (2005) drop some observations in the right tail of the response variable and therefore the resulting tests *are no longer omnibus*. Moreover, neither Stute *et al.* (2000) nor Sánchez-Sellero *et al.* (2005) studied the consistency of the tests against a sequence of alternatives approaching the null hypothesis. Pardo-Fernandez *et al.* (2005) proposed another test for parametric models in censored regression that is based on the comparison of two estimators, parametric and nonparametric, of the distribution of the errors. As the latter estimator is based on a nonparametric location-scale model, the test of Pardo-Fernandez *et al.* (2005) is not consistent against *any* alternative.

In this paper we consider two versions adapted for right-censored responses of the kernel-based test statistic studied by Zheng (1996). See also Härdle and Mammen (1993), Horowitz and Spokoiny (2001), Guerre and Lavergne (2005) for closely related test statistics. In the non-censored case, the kernel-based test statistic we consider is a suitably normalized $U-$statistic built from the estimated residuals of the parametric model. Under suitable conditions, the test statistic converges in law to a standard normal when the model is correct. The problem in presence of censoring is that estimated residuals can be computed only for uncensored observations. The two solutions we propose are inspired by the WLS and SD estimation approaches mentioned above. On one hand, we build a *weighted* $U-$statistic using estimated residuals with the weights estimated from data. Once again, the weights account for censoring. On the other hand, we build a $U-$statistic using estimated *synthetic* residuals where the synthetic residuals are the difference between the synthetic responses and the predictions given by the model. Two smoothing-based test statistics are obtained after suitably normalizing each of these $U-$statistics.

The paper is organized as follows. In section 2 we recall the weighted least squares and synthetic data approaches for (non)linear regression models when the response is right-censored. Section 3 shows how to build two kernel based test statistics adapted for censored responses. Section 4 deals with the asymptotic behavior of the two omnibus tests that we derive. The main results in this paper show that the asymptotic study of our tests boils down to the asymptotic study of kernel-based tests without censoring but with suitably transformed observations. As a consequence, the asymptotic critical values of the new tests are given by the quantiles of the standard normal law. Moreover, the asymptotic consistency of our tests is obtained by arguments similar to those used for kernel based tests in the non-censored case. In particular, we study the consistency of the new tests against fixed alternatives, local Pitman type alternatives and the consistency uniformly over Hölder classes of alternatives of known regularity. It is worthwhile to notice that the results of asymptotic equivalence between our two test statistics and two kernel-based test statistics built with transformed (non-censored) observations are obtained uniformly in the bandwidth. This motivates us to propose the construction of a data-driven procedure inspired by the maximum test approach of Horowitz and Spokoiny (2001). However, to keep this paper at reasonable length, the detailed investigation of this data-driven procedure is left for future work. Finally, in section 5 we illustrate the performance of the new tests using a small simulation experiment.

2 Preliminaries

Consider the model $Y = m(X) + \varepsilon$, where $Y \in \mathbb{R}$, $X \in \mathbb{R}^p$, $\mathbb{E}(\varepsilon \mid X) = 0$ almost surely (a.s.), and $m(\cdot)$ is an unknown function. In presence of random right censoring, the response Y is not always available. Instead of (Y, X), one observes a random sample from (T, δ, X) with

$$T = Y \wedge C, \qquad \delta = \mathbf{1}_{\{Y \le C\}},$$

where C is the "censoring" random variable, and $\mathbf{1}_A$ denotes the indicator function of the set A. In our setting, the variable X is not subject to censoring and is fully observed. We want to check whether the regression function $m(\cdot)$ belongs to a parametric family

$$\mathcal{M} = \left\{ f(\theta, \cdot) : \theta \in \Theta \subset \mathbb{R}^d \right\}$$

where f is a known function. Our null hypothesis then writes

$$H_0 \ : \text{for some } \theta_0, \quad \mathbb{E}(Y|X) = f(\theta_0, X) \text{ a.s.}, \tag{2.1}$$

while the alternative is $\mathbb{P}[\mathbb{E}(Y|X) = f(\theta, X)] \le c$ for every $\theta \in \Theta$ and some $c < 1$. For testing H_0, first we need to estimate θ_0.

2.1 Estimating (non)linear regressions with censored data

Since the observed variable T does not have the same conditional expectation as Y, classical techniques for estimating parametric (non)linear regression models like $\mathcal{M}$ must be adapted to account for censorship. Several adapted procedures have been proposed, that we classify in two groups: synthetic data (SD) procedures and weighted least squares (WLS). In the SD approach one replaces the variable T with some transformation of the data Y^*, a transformation which preserves the conditional expectation of Y. Several transformations have been proposed, see for instance Buckley and James (1979), Leurgans (1987), Zheng (1987). In the following, we will restrain ourselves to the transformation first proposed by Koul *et al.* (1981), that is

$$Y^* = \frac{\delta T}{1 - G(T-)}, \tag{2.2}$$

where $G(t) = \mathbb{P}(C \leq t)$. The following assumptions will be used throughout this paper to ensure that $\mathbb{E}(Y^* \mid X) = \mathbb{E}(Y \mid X)$ for Y^* defined in (2.2).

Assumption 1 *Y and C are independent.*

Assumption 2 $\mathbb{P}(Y \leq C \mid X, Y) = \mathbb{P}(Y \leq C \mid Y)$.

These assumptions are quite common in the survival analysis literature when covariates are present. Assumption 1 is an usual identification condition when working with the Kaplan-Meier estimator. Stute (1993), pages 462-3, provides a detailed discussion on Assumption 2. These assumptions may be inappropriate for some data sets. However, they are often satisfied in randomized clinical trials when the failure time Y of each subject is either observed or administratively censored at the end of the follow-up period. Notice that Assumption 2 is flexible enough to allow for a dependence between X and C. Moreover, Assumptions 1 and 2 imply the following general property: for any integrable $\phi(T, X)$,

$$\mathbb{E}\left[\frac{\delta}{1 - G(T-)}\,\phi(T, X) \mid X\right] = \mathbb{E}\left[\phi(Y, X) \mid X\right]. \tag{2.3}$$

Unfortunately, one cannot compute the transformation (2.2) when the function G is unknown. Therefore, given the i.i.d. observations $(T_1, \delta_1, X_1), ..., (T_n, \delta_n, X_n)$, Koul *et al.*

(1981) proposed to replace G with its Kaplan-Meier estimate

$$\hat{G}(t) = 1 - \prod_{\{j:T_j \leq t\}} \left(1 - \frac{1}{R_n(T_j)}\right)^{1-\delta_j}, \qquad \text{with} \qquad R_n(t) = \sum_{k=1}^{n} \mathbf{1}_{\{t \leq T_k\}},$$

and to compute

$$\hat{Y}_i^* = \frac{\delta_i T_i}{1 - \hat{G}(T_i-)}, \quad i = 1, ..., n. \tag{2.4}$$

Next, Koul *et al.* (1981) proposed to estimate θ_0 by $\hat{\theta}^{SD}$ that minimizes

$$M_n^{SD}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left[\hat{Y}_i^* - f(\theta, X_i)\right]^2$$

over Θ. They obtained the consistency of $\hat{\theta}^{SD}$ and the asymptotic normality of $\sqrt{n}(\hat{\theta}^{SD} - \theta_0)$ in the particular case of a linear regression model. Delecroix *et al.* (2006) generalized these results to more general functions $f(\theta, x)$.

The WLS approach consists in applying weighted least squares techniques directly to variables T_i, that is computing $\hat{\theta}^{WLS}$ which minimizes

$$M_n^{WLS}(\theta) = \sum_{i=1}^{n} W_{in} \left[T_i - f(\theta, X_i)\right]^2,$$

with a specific choice of W_{in} that compensates for the fact that Y is censored. More precisely, the weights W_{in} are defined by

$$W_{in} = \frac{\delta_i}{n\left[1 - \hat{G}(T_i-)\right]}. \tag{2.5}$$

Zhou (1992a) studied an estimator like $\hat{\theta}^{WLS}$ in the case of linear regression. Under Assumptions 1 and 2, Stute (1999) generalized this approach to nonlinear regressions. Using the Kaplan-Meier estimator $\hat{F}_{(X,Y)}(x, y)$ of $F_{(X,Y)}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$ introduced by Stute (1993), Stute (1999) interpreted $\hat{\theta}^{WLS}$ as the minimizer of

$$\int \left[y - f(\theta, x)\right]^2 d\hat{F}_{(X,Y)}(x, y) \tag{2.6}$$

with respect to θ. Indeed, on one hand, by definition, at observation i the jump of $\hat{F}_{(X,Y)}$ is equal to the jump of the Kaplan-Meier estimate of $F(t) = \mathbb{P}(Y \leq t)$. On the other hand, it can be easily shown that the jump of $\hat{F}(t)$ at observation i is equal to the weight W_{in} defined in (2.5). Using the properties of Kaplan-Meier integrals, one can deduce consistency and root-n-asymptotic normality for $\hat{\theta}^{WLS}$. See Stute (1999, 1993) or Delecroix *et al.* (2006) for more details. It is worthwhile to notice that a choice of W_{in} as in (2.5) connects $M_n^{WLS}(\theta)$ to $M_n^{SD}(\theta)$ since, in this case, $\hat{Y}_i^* = nW_{in}T_i$. In the following section, we will see how to extend the purpose of the two methodologies (SD and WLS) from estimation to testing.

3 Nonparametric test procedures under censoring

To better explain the new approach, first the case where Y is not censored is reconsidered. Then, testing the adequacy of model $\mathcal{M}$ is equivalent to testing

$$\text{for some } \theta_0, \quad Q(\theta_0) = 0 \quad \text{where} \quad Q(\theta) = \mathbb{E}\left[U(\theta)\,\mathbb{E}\left[U(\theta) \mid X\right] g(X)\right],$$

$U(\theta) = Y - f(\theta, X)$ and g denotes the density of X that is assumed to exist. The choice of g avoids handling denominators close to zero. When the responses are *not censored*, one may estimate $Q(\theta_0)$ by the kernel-based estimator

$$Q_n(\hat{\theta}) = \frac{1}{n(n-1)h^p} \sum_{i \neq j} U_i(\hat{\theta}) U_j(\hat{\theta}) K_h(X_i - X_j) \tag{3.7}$$

where $\hat{\theta}$ is an estimator of θ_0 such that $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$, $U_i(\theta) = Y_i - f(\theta, X_i)$, K is some $p-$dimensional kernel function, h denotes the bandwidth and for $x \in \mathbb{R}^p$, $K_h(x) = K(x/h)$. See Zheng (1996). See also Horowitz and Spokoiny (2001) or Guerre and Lavergne (2005). For the purpose of adapting this smoothing approach to the case where the responses are right-censored, it is worthwhile to notice that $Q_n(\hat{\theta})$ is equal to

$$\begin{aligned} \frac{n}{(n-1)h^p} \int\!\!\int & \left[y_1 - f\left(\hat{\theta}, x_1\right)\right] \left[y_2 - f(\hat{\theta}, x_2)\right] K_h(x_1 - x_2) \\ & \times \mathbf{1}_{\{(x_1,y_1) \neq (x_2,y_2)\}} d\hat{F}_{emp}(x_1, y_1)\, d\hat{F}_{emp}(x_2, y_2), \end{aligned} \tag{3.8}$$

where $\hat{F}_{emp}$ is the empirical distribution function of the sample of X and Y.

Using a consistent estimate $\hat{V}_n^2$ of the asymptotic variance of $nh^{p/2}Q_n(\hat{\theta})$, the smoothing based test statistic with non-censored responses is

$$T_n^{NC} = nh^{p/2} \frac{Q_n(\hat{\theta})}{\hat{V}_n}. \tag{3.9}$$

Under the null hypothesis the statistic behaves asymptotically as a standard normal and therefore the nonparametric test is defined as *"Reject H_0 when $T_n^{NC} \geq z_{1-\alpha}$"*, where $z_{1-\alpha}$ is the $(1-\alpha)$th quantile of the standard normal law. As an estimate $\hat{V}_n^2$, one could use either

$$\hat{V}_n^2 = \frac{2}{n(n-1)h^p} \sum_{i \neq j} U_i^2(\hat{\theta}) U_j^2(\hat{\theta}) K_h^2 \left(X_i - X_j\right)$$

or

$$\hat{V}_n^2 = \frac{2}{n(n-1)h^p} \sum_{i \neq j} \hat{\sigma}^2 \left(X_i\right) \hat{\sigma}^2 \left(X_j\right) K_h^2 \left(X_i - X_j\right), \tag{3.10}$$

with $\hat{\sigma}^2 (x)$ a nonparametric estimator of $\sigma^2 (x) = Var(\varepsilon \mid X = x)$. The former choice for $\hat{V}_n^2$ is simpler but is likely to decrease the power of the test because the squares of the estimated residuals of the parametric model produce an upward biased estimate of $\sigma^2 (x)$ under the alternative hypothesis.

In the presence of censored responses, the test statistic (3.9) cannot be computed since $\hat{F}_{emp}$ is unavailable. However, it is possible to find the analogue of $Q_n(\hat{\theta})$ using the two approaches of section 2.

3.1 Two test statistics with right-censored responses

In the following, the observations are $(T_1, \delta_1, X_1), ..., (T_n, \delta_n, X_n)$, a random sample from (T, δ, X). In the spirit of the SD approach, consider the estimated synthetic responses $\hat{Y}_1^*, ..., \hat{Y}_n^*$ obtained from formula (2.4) and the empirical distribution function

$$\hat{F}^* (x, y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\left\{X_i \leq x, \hat{Y}_i^* \leq y\right\}}.$$

Now, the analogue of $Q_n(\hat{\theta})$ is

$$\begin{aligned} Q_n^{SD}(\hat{\theta}) &= \frac{n}{(n-1)\, h^p} \int \int \left[y_1^* - f(\hat{\theta}, x_1)\right] \left[y_2^* - f(\hat{\theta}, x_2)\right] \\ &\quad \times K_h(x_1 - x_2) \mathbf{1}_{\left\{(x_1, y_1^*) \neq (x_2, y_2^*)\right\}} d\hat{F}^*(x_1, y_1^*)\, d\hat{F}^*(x_2, y_2^*) \\ &= \frac{1}{n(n-1)h^p} \sum_{i \neq j} \hat{U}_i^{SD}(\hat{\theta}) \hat{U}_j^{SD}(\hat{\theta}) K_h(X_i - X_j), \end{aligned} \tag{3.11}$$

where $\hat{\theta} = \hat{\theta}^{SD}$ and

$$\hat{U}_i^{SD}(\theta) = \frac{\delta_i}{1 - \hat{G}(T_i -)} T_i - f(\theta, X_i) = nW_{in}T_i - f(\theta, X_i) \tag{3.12}$$

are the estimated synthetic residuals. The statistic $Q_n^{SD}(\theta)$ estimates

$$Q^{SD}(\theta) = \mathbb{E}\left[U^{SD}(\theta)\,\mathbb{E}\left[U^{SD}(\theta) \mid X\right] g(X)\right]$$

with $U^{SD}(\theta) = \delta T\left[1 - G(T-)\right]^{-1} - f(\theta, X)$. By (2.3), if Assumptions 1 and 2 hold then the null hypothesis is equivalent to $Q^{SD}(\theta_0) = 0$. Therefore $Q_n^{SD}(\hat{\theta})$ can serve to build our first test statistic.

On the other hand, following the WLS approach we can replace $\hat{F}_{emp}(x, y)$ in (3.8) with the Kaplan-Meier estimator

$$\hat{F}_{(X,Y)}(x, y) = \sum_{i=1}^{n} W_{in}\mathbf{1}_{\{X_i \le x,\, T_i \le y\}}$$

to obtain a second $U-$statistic

$$\begin{aligned} Q_n^{WLS}(\hat{\theta}) = & \frac{n}{(n-1)\,h^p}\iint\left[y_1 - f(\hat{\theta}, x_1)\right]\left[y_2 - f(\hat{\theta}, x_2)\right]K_h(x_1 - x_2) \\ & \times\mathbf{1}_{\{(x_1,y_1)\neq(x_2,y_2)\}}d\hat{F}_{(X,Y)}(x_1, y_1)\,d\hat{F}_{(X,Y)}(x_2, y_2) \\ = & \frac{1}{n(n-1)h^p}\sum_{i\neq j}\hat{U}_i^{WLS}(\hat{\theta})\hat{U}_j^{WLS}(\hat{\theta})K_h(X_i - X_j), \end{aligned} \tag{3.13}$$

with $\hat{\theta} = \hat{\theta}^{WLS}$ and

$$\hat{U}_i^{WLS}(\theta) = \frac{\delta_i}{1 - \hat{G}(T_i-)}\left[T_i - f(\theta, X_i)\right] = nW_{in}\left[T_i - f(\theta, X_i)\right]. \tag{3.14}$$

The statistic $Q_n^{WLS}(\theta)$ estimates

$$Q^{WLS}(\theta) = \mathbb{E}\left[U^{WLS}(\theta)\,\mathbb{E}\left[U^{WLS}(\theta) \mid X\right] g(X)\right]$$

with $U^{WLS}(\theta) = \delta\left[1 - G(T-)\right]^{-1}\left[T - f(\theta, X)\right]$. By (2.3), the null hypothesis is equivalent to $Q^{WLS}(\theta_0) = 0$ and therefore $Q_n^{WLS}(\hat{\theta})$ can be used to build our second test statistic, provided that Assumptions 1 and 2 hold true.

Now, given consistent estimates $\left[\hat{V}_n^{SD}\right]^2$ and $\left[\hat{V}_n^{WLS}\right]^2$ of the asymptotic variance of $nh^{p/2}Q_n^{SD}(\hat{\theta})$ and $nh^{p/2}Q_n^{WLS}(\hat{\theta})$, respectively, we propose the following test statistics:

$$T_n^{SD} = T_n^{SD}(\hat{\theta}) = nh^{p/2}\frac{Q_n^{SD}(\hat{\theta})}{\hat{V}_n^{SD}}, \qquad T_n^{WLS} = T_n^{WLS}(\hat{\theta}) = nh^{p/2}\frac{Q_n^{WLS}(\hat{\theta})}{\hat{V}_n^{WLS}}.$$

The corresponding omnibus tests are

$$\text{“Reject } H_0 \text{ when } T_n^{SD} \ge z_{1-\alpha} \quad (\text{resp. } T_n^{WLS} \ge z_{1-\alpha})\text{ ”}. \tag{3.15}$$

To estimate the variance of $nh^{p/2}Q_n^{SD}(\hat{\theta})$ we consider

$$\left[\hat{V}_n^{SD}\right]^2 = \frac{2}{n(n-1)h^p} \sum_{i \neq j} \left[\hat{U}_i^{SD}(\hat{\theta})\right]^2 \left[\hat{U}_j^{SD}(\hat{\theta})\right]^2 K_h^2\left(X_i - X_j\right). \tag{3.16}$$

The variance of $nh^{p/2}Q_n^{WLS}(\hat{\theta})$ is estimated similarly with $\hat{U}_i^{SD}(\hat{\theta})$ replaced by $\hat{U}_i^{WLS}(\hat{\theta})$. Alternative variance estimates are discussed in section 4.

Checking the validity of a parametric conditional model has attracted much attention in survival analysis. Hjort (1990) and Lin and Spiekerman (1996) considered goodness-of-fit statistics based on martingale residuals, while Gray and Pierce (1985) showed how Neyman's smooth tests may be adapted to censored data. See chapter 10 of Lawless (2003) for a review of the methods for testing the lack-of-fit. All these techniques can be used to check whether some parametric form of the *conditional law* of the response variable given the explanatory variables is consistent with observed data. Therefore, these techniques are only of limited use in our framework where we aim to check the adequacy of some parametric form of the *conditional expectation* of the response variable given the covariates. The standard normal limit of the test statistics T_n^{SD} and T_n^{WLS} under the null hypothesis, a property that will be proved in the following, yields the simple one-sided tests (3.15) for checking mean-regressions. By contrast, the only alternative test statistics available in the literature (see Stute *et al.* 2000) have a complicated limit and there is no simple way to construct the critical values of the associated tests.

4 Asymptotic analysis

The most difficult part of the study of our tests is the investigation of the properties of $Q_n^{SD}(\theta)$ and $Q_n^{WLS}(\theta)$. These quadratic forms are difficult to analyze even when H_0 holds true and θ is equal to θ_0, since they do not rely on i.i.d. quantities U_i, as the quadratic form (3.7) does. In fact, due to the presence of $\hat{G}$ in (3.12) and (3.14), each $\hat{U}_i^{SD}(\theta_0)$ and $\hat{U}_i^{WLS}(\theta_0)$ depend on the whole sample. Then, a key point is to show that under H_0, in some sense, $Q_n^{SD}(\hat{\theta})$ and $Q_n^{WLS}(\hat{\theta})$ are asymptotically equivalent to the "ideal" quadratic forms

$$\tilde{Q}_n^{SD}(\theta_0) = \frac{1}{n(n-1)h^p} \sum_{i \neq j} U_i^{SD}(\theta_0)\, U_j^{SD}(\theta_0)\, K_h\left(X_i - X_j\right) \tag{4.17}$$

and

$$\tilde{Q}_n^{WLS}(\theta_0) = \frac{1}{n(n-1)h^p} \sum_{i \neq j} U_i^{WLS}(\theta_0) U_j^{WLS}(\theta_0) K_h(X_i - X_j), \tag{4.18}$$

respectively, where

$$\begin{aligned} U_i^{SD}(\theta) &= \frac{\delta_i}{1 - G(T_i-)} T_i - f(\theta, X_i) = \gamma(T_i) T_i - f(\theta, X_i), \\ U_i^{WLS}(\theta) &= \frac{\delta_i}{1 - G(T_i-)} [T_i - f(\theta, X_i)] = \gamma(T_i)[T_i - f(\theta, X_i)]. \end{aligned}$$

The asymptotic study of $\tilde{Q}_n^{SD}(\theta_0)$ and $\tilde{Q}_n^{WLS}(\theta_0)$ can be done like in the non-censored case. Therefore, the asymptotic level of our tests will be obtained as a consequence of the equivalence result and using techniques for kernel-based tests in the i.i.d. case. See, for instance, Zheng (1996), Horowitz and Spokoiny (2001), Guerre and Lavergne (2005). A similar equivalence result deduced under fixed or moving alternatives will serve for studying the asymptotic consistency of our tests.

4.1 Assumptions

In addition to Assumptions 1 and 2 in section 2, we will use the following assumptions.

Assumption 3 *(i) F and G are continuous.*

(ii) $-\infty < \tau_F \leq \tau_G \leq \infty$, where $\tau_L = \inf\{t \mid L(t) = 1\}$ for any distribution function L.

Assumption 3 (i) is introduced for convenience purposes. Considered together with Assumptions 1, it implies $\mathbb{P}(Y = C) = 0$ and this latter condition simplifies the results on Kaplan-Meier integrals (see Stute 1995, Sánchez *et al.* 2005) and justifies the definition of the Kaplan-Meier estimate $\hat{G}$. When $\tau_F > \tau_G$, there is no way to access information about the law of Y beyond τ_G, so that, in general, there is no way to consistently estimate θ_0. Assumption 3 (ii) allows one to avoid this case.

Assumption 4 *(Data): (i) Let $(\varepsilon_1, C_1, X_1), ..., (\varepsilon_n, C_n, X_n)$ be an independent sample of (ε, C, X) where $\varepsilon, C \in \mathbb{R}$ and $X \in \mathbb{R}^p$, and suppose $\mathbb{E}(\varepsilon \mid X) = 0$ a.s.*

(ii) X is a random vector with bounded support $\mathcal{X}$ and bounded density g.

(iii) There exist some constants c_{inf}, c_{sup} such that for each $x \in \mathcal{X}$

$$0 < c_{inf} \leq \mathbb{E}\left[\varepsilon^2 \mid X = x\right] \leq \mathbb{E}\left[\left\{1 + \varepsilon^2\right\}\{1 - G(Y)\}^{-1} \mid X = x\right] \leq c_{sup} < \infty.$$

(iv) $\mathbb{E}\left[\{1+\varepsilon^4\}\,\delta\{1-G(Y)\}^{-4}\right]=\mathbb{E}\left[\{1+\varepsilon^4\}\,\gamma(T)^4\right]<\infty.$

Assumptions 4 (iii)-(iv) are counterparts of assumptions on the conditional variance and the fourth moment of the residuals that are usually imposed in the non-censored case. See, e.g., Guerre and Lavergne (2005). Now, define $\nabla_\theta f(\theta,x) = \partial f(\theta,x)/\partial\theta$, $\nabla_\theta^2 f(\theta,x)=\partial^2 f(\theta,x)/\partial\theta\partial\theta'$, whenever these derivatives exist. For any matrix A let $\|A\|_2$ denote its 2-norm, that is $\|A\|_2=\sup_{v\neq 0}\|Wv\|/\|v\|$, where $\|v\|$ is the Euclidean norm of the vector v.

Assumption 5 *(Parametric model): The parameter set Θ is a compact subset of $\mathbb{R}^d$, $d\geq 1$, and θ_0 in an interior point of Θ. The parametric regression model $\mathcal{M}=\{f(\theta,\cdot):\theta\in\Theta\}$ satisfies:*

(i) Differentiability in θ: for each $x\in\mathcal{X}$, $f(\theta,x)$ is twice differentiable with respect to θ. There exists a finite constant c_1 such that for each $\theta\in\Theta$ and $x\in\mathcal{X}$, $|f(\theta,x)|+\|\nabla_\theta f(\theta,x)\|+\|\nabla_\theta^2 f(\theta,x)\|_2\leq c_1$. Moreover, there exist finite constants $a, c_2>0$ such that for each θ and x, $|\nabla_\theta^2 f(\theta,x)_{jk}-\nabla_\theta^2 f(\theta_0,x)_{jk}|\leq c_2\|\theta-\theta_0\|^a$, where $\nabla_\theta^2 f(\theta,x)_{jk}$ is the element jk of the matrix $\nabla_\theta^2 f(\theta,x)$.

(ii) Identifiability: there exists a nonnegative bounded function Φ with $\mathbb{E}[\Phi(X)]>0$ such that for each $\theta\in\Theta$ and $x\in\mathcal{X}$, $|f(\theta,x)-f(\theta_0,x)|\geq\Phi(x)\|\theta-\theta_0\|$.

Assumption 6 *(Kernel smoother): (i) If $x=(x_1,...,x_p)$, let $K(x)=\tilde{K}(x_1)...\tilde{K}(x_p)$ where $\tilde{K}$ is a symmetric continuous density of bounded variation on $\mathbb{R}$. The Fourier Transform $\hat{\tilde{K}}$ of $\tilde{K}$ is positive, integrable and non-increasing on $[0,\infty)$.*

(ii) The bandwidth h belongs to an interval $\mathcal{H}_n=[h_{min},h_{max}]$, $n\geq 1$, such that $h_{max}\to 0$ and $nh_{min}^{3p}\to\infty$.

Condition (i) of Assumption 6 holds, for instance, for normal, Laplace or Cauchy densities. The condition non-increasing Fourier Transform for $\hat{\tilde{K}}$ is a convenient assumption that will serve only for deriving our asymptotic equivalence results *uniformly* in the bandwidth. Concerning the range for the bandwidth, in view of equation (A.9) in the Appendix, it is clear that h_{min} may be taken of smaller rate if Assumption 4-(iv) above and Assumption 7 below are made more restrictive. The following assumption is connected to the asymptotic theory of Kaplan-Meier integrals, see condition (1.6) of Stute (1995) and Stute (1996). It will serve us to control the jumps of the Kaplan-Meier estimator. Below, $a \vee b$ denotes the maximum of a and b.

Assumption 7 *Let* $H(t) = \mathbb{P}(T \leq t)$, $t \in \mathbb{R}$, *and*

$$q_\rho(x) = \mathbb{E}\left[\{|Y| + 1\} C(Y)^{1/2+\rho} \mid X = x\right],$$

where

$$C(y) = \int_{-\infty}^{y} \frac{dG(t)}{[1 - H(t)][1 - G(t)]} \vee 1.$$

There exists some $0 < \rho < 1/2$ *such that* $\mathbb{E}[q_\rho^2(X)] < \infty$.

The function $C(\cdot)$ in Assumption 7 appears also in Bose and Sen (2002) who derive an i.i.d. representation for Kaplan-Meier $U-$statistics. Their general result would have been useful for studying our test statistics. Unfortunately, they impose $\rho = 1/2$ for deriving their representation, see Bose and Sen's Theorem 1 and Remark 1. This condition is unrealistic in our framework.

4.2 Behavior of the tests under the null hypothesis

The following theorem gives an asymptotic representation of the test statistics T_n^{SD} and T_n^{WLS} under H_0 stated in (2.1). To simplify notation, in the following we replace the superscripts SD and WLS with 0 and 1, respectively. For instance we write T_n^0 and Q_n^0 (resp. T_n^1 and Q_n^1) instead of T_n^{SD} and Q_n^{SD} (resp. T_n^{WLS} and Q_n^{WLS}). As before, $\hat{\theta}$ stands for $\hat{\theta}^{SD}$ or $\hat{\theta}^{WLS}$, depending on the approach considered.

Theorem 4.1 *Let Assumptions 1 to 7 hold. Under* H_0, *for* $\beta = 0$ *or 1*

$$\sup_{h \in \mathcal{H}_n} \left\{ \left| nh^{p/2} Q_n^\beta(\hat{\theta}) - nh^{p/2} \tilde{Q}_n^\beta(\theta_0) \right| + \left| \frac{\tilde{V}_n^\beta(\theta_0)}{\hat{V}_n^\beta} - 1 \right| \right\} \to 0,$$

in probability, where

$$\left[\tilde{V}_n^\beta(\theta_0)\right]^2 = \frac{2}{n(n-1)h^p} \sum_{i \neq j} \left[U_i^\beta(\theta_0)\right]^2 \left[U_j^\beta(\theta_0)\right]^2 K_h^2\left(X_i - X_j\right).$$

Moreover, under H_0 *and for* $\beta = 0$ *or 1*

$$\sup_{h \in \mathcal{H}_n} \left| T_n^\beta(\hat{\theta}) - \frac{nh^{p/2} \tilde{Q}_n^\beta(\theta_0)}{\tilde{V}_n^\beta(\theta_0)} \right| = o_P(1).$$

Corollary 4.2 *Under Assumptions 1 to 7 the two tests defined in equation (3.15) have asymptotic level* α*.*

Proof of Theorem 4.1. We give here the main steps of the proof. Technical arguments are postponed to the Appendix.

Step 1. First, notice that the assumptions ensure $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$ (see, e.g., Delecroix *et al.* 2006 or Lemma A.12 in the Appendix). Next, by Lemma A.6

$$\sup_{h \in \mathcal{H}_n} h^{p/2} \left| Q_n^\beta(\hat{\theta}) - Q_n^\beta(\theta_0) \right| = o_P(n^{-1}),$$

and thus we reduce the problem to the study of $Q_n^\beta(\theta_0)$.

Step 2. Let us simplify notation: for $\beta = 0$ or 1 and $i = 1, ..., n$, write U_i^β (resp. $\hat{U}_i^\beta$) instead of $U_i^\beta(\theta_0)$ (resp. $\hat{U}_i^\beta(\theta_0)$). Now decompose

$$\begin{aligned} Q_n^\beta(\theta_0) &= \frac{1}{n(n-1)h^p} \sum_{i \neq j} U_i^\beta U_j^\beta K_h(X_i - X_j) \\ &\quad + \frac{2}{n(n-1)h^p} \sum_{i \neq j} \left[\hat{U}_i^\beta - U_i^\beta \right] U_j^\beta K_h(X_i - X_j) \\ &\quad + \frac{1}{n(n-1)h^p} \sum_{i \neq j} \left[\hat{U}_i^\beta - U_i^\beta \right] \left[\hat{U}_j^\beta - U_j^\beta \right] K_h(X_i - X_j) \\ &= \tilde{Q}_n^\beta(\theta_0) + 2Q_{n1}^\beta + Q_{n2}^\beta. \end{aligned} \tag{4.19}$$

Fix $\tau < \tau_H = \inf\{t : H(t) = 1\}$ arbitrarily. To show that Q_{n1}^β is negligible, first we study a truncated version of this quantity, that is

$$Q_{n1}^\beta(\tau) = \frac{1}{n(n-1)h^p} \sum_{i \neq j} \left[\hat{U}_i^\beta - U_i^\beta \right] \mathbf{1}_{\{T_i \leq \tau\}} U_j^\beta K_h(X_i - X_j). \tag{4.20}$$

Since $\hat{U}_i^\beta - U_i^\beta$ can be decomposed into two parts

$$\begin{aligned} \hat{U}_i^\beta - U_i^\beta &= \frac{\hat{G}(T_i -) - G(T_i)}{[1 - G(T_i)]^2} \delta_i \left[T_i - \beta f(\theta_0, X_i) \right] \\ &\quad + \frac{\left[\hat{G}(T_i -) - G(T_i) \right]^2}{[1 - G(T_i)]^2 \left[1 - \hat{G}(T_i -) \right]} \delta_i \left[T_i - \beta f(\theta_0, X_i) \right], \end{aligned}$$

we can separate $Q_{n1}^\beta(\tau)$ into two sums $Q_{n11}^\beta(\tau)$ and $Q_{n12}^\beta(\tau)$, respectively. For $Q_{n12}^\beta(\tau)$, take absolute values and recall that $\sup_{t \leq \tau} \left| \hat{G}(t) - G(t) \right| = O_P(n^{-1/2})$, provided that

$G(\tau) < 1$ (cf. Gill, 1983). This allows us to take $[\hat{G}(T_i-) - G(T_i)]^2$ outside the sum of absolute values and thus to bound $|Q^{\beta}_{n12}(\tau)|$ by a factor $O_P(n^{-1})$ times a $U-$process indexed by $h \in \mathcal{H}_n$. By the rate of uniform convergence of $U-$processes indexed by Euclidean families of functions for a square integrable envelope (see Sherman, 1994) and the condition $nh^{2p}_{min} \to \infty$, we conclude that $\sup_{h\in\mathcal{H}_n} |Q^{\beta}_{n12}(\tau)| = O_P(n^{-1})$. For $Q^{\beta}_{n11}(\tau)$, we use the an i.i.d. representation of $\hat{G}(t-) - G(t)$ with a remainder of order $O_P(n^{-1})$ uniformly in t where $t \leq \tau$. See Theorem 1.1 of Stute (1995) and Theorem 1 of Sánchez-Sellero *et al.* (2005). Replacing $\hat{G}(T_i-) - G(T_i)$ with the sum in the i.i.d. representation plus a remainder, the rate of $Q^{\beta}_{n11}(\tau)$ will be given by the rate of uniform convergence of two $U-$processes indexed by $h \in \mathcal{H}_n$. Using Corollary 4 of Sherman (1994) and the condition $nh^{2p}_{min} \to \infty$, we deduce that for any $\zeta > 0$, $\sup_{h\in\mathcal{H}_n} |h^{\zeta} Q^{\beta}_{n11}(\tau)| = o_P(n^{-1})$. As a consequence

$$\sup_{h\in\mathcal{H}_n} \left| h^{p/2} Q^{\beta}_{n1}(\tau) \right| = o_P(n^{-1}). \tag{4.21}$$

See Lemma A.7 in the Appendix for the details. To derive the rate of Q^{β}_{n2}, take absolute values and use Lemma A.1 in the Appendix to bound $|\hat{U}^{\beta}_i - U^{\beta}_i|$ by $O_P(n^{-1/2})$ times a function of T_i and δ_i which is not square integrable. Consequently, $|Q^{\beta}_{n2}|$ can be bounded by $O_P(n^{-1})$ times a second order $U-$statistic indexed by h. Here, we can no longer apply Sherman (1994)'s results on the rates of uniform convergence for handling this $U-$statistic because the square integrable envelope condition fails. However, the expectation of this $U-$statistic remains bounded when $h \to 0$. This implies $Q^{\beta}_{n2} = O_P(n^{-1})$ when considering a sequence $h \to 0$. To obtain this rate *uniformly* in $h \in \mathcal{H}_n$, that is

$$\sup_{h\in\mathcal{H}_n} \left| Q^{\beta}_{n2} \right| = O_P(n^{-1}), \tag{4.22}$$

we use the monotonicity property of the Fourier Transform of the kernel K and the condition $nh^{2p}_{min} \to \infty$. See Lemma A.8 in the Appendix for the details.

Step 3. Since by definition $Q^{\beta}_{n1}(\tau_H) = Q^{\beta}_{n1}$, it remains to make $\tau \uparrow \tau_H$. By Lemma A.9 in the Appendix

$$\sup_{h\in\mathcal{H}_n} h^{p/2} \left| Q^{\beta}_{n1}(\tau) - Q^{\beta}_{n1} \right| = C_{\tau} \times O_P(n^{-1}),$$

with the $O_P(n^{-1})$ factor independent of τ and C_τ tending to zero when $\tau \uparrow \tau_H$. Use (4.21) and the Cramér-Slutsky argument from Theorem 1.1 of Stute (1995) to deduce that

$$\sup_{h\in\mathcal{H}_n} \left| nh^{p/2} Q_{n1}^{\beta} \right| = o_P(1).$$

From this and (4.22) we obtain

$$\sup_{h\in\mathcal{H}_n} \left| nh^{p/2} Q_n^{\beta}(\theta_0) - nh^{p/2} \tilde{Q}_n^{\beta}(\theta_0) \right| = o_P(1).$$

Step 4. The result for $\hat{V}_n^{\beta}$ is contained in Lemma A.10 in the Appendix. The second part of the theorem follows if we recall that $\tilde{V}_n^{\beta}(\theta_0)$ converges in probability to a strictly positive limit and $nh^{p/2}\tilde{Q}_n^{\beta}(\theta_0)$ is bounded in probability. ■

Remark 1. To estimate the variance $nh^{p/2}Q_n^0(\hat{\theta})$ we considered (3.16). Alternatively, extending the idea behind the equation (3.10) to the right-censoring framework, one may replace in (3.16) the estimated squared residual $\hat{U}_i^0(\hat{\theta})^2$ with a nonparametric estimate of $\sigma^{*\,2}(x) = Var(Y^* \mid X = x)$, the conditional variance of the synthetic responses. It is easy to check that $Var(Y^* \mid X) = \mathbb{E}\left[U^0(\theta_0)^2 \mid X\right]$ under H_0 and, in general, $Var(Y^* \mid X) < \mathbb{E}\left[U^0(\theta_0)^2 \mid X\right]$ if the regression model is not correct. To estimate $\sigma^{*\,2}(\cdot)$, one can use

$$\hat{\sigma}_n^{*\,2}(x) = \frac{\sum_{i=1}^n \hat{Y}_i^{*\,2} L((X_i - x)/b_n)}{\sum_{i=1}^n L((X_i - x)/b_n)} - \left(\frac{\sum_{i=1}^n \hat{Y}_i^{*} L((X_i - x)/b_n)}{\sum_{i=1}^n L((X_i - x)/b_n)} \right)^2, \tag{4.23}$$

$x \in \mathcal{X}$, with L a multivariate kernel and b_n a bandwidth parameter chosen independently of $\mathcal{H}_n$. If

$$\sup_{x\in\mathcal{X}} \left| \hat{\sigma}_n^{*\,2}(x) - \sigma^{*\,2}(x) \right| \to 0 \tag{4.24}$$

in probability, we can redefine

$$\left[\hat{V}_n^0\right]^2 = \frac{2}{n(n-1)h^p} \sum_{i\neq j} \hat{\sigma}_n^{*\,2}(X_i)\hat{\sigma}_n^{*\,2}(X_j) K_h^2(X_i - X_j) \tag{4.25}$$

and the test statistic $T_n^0(\hat{\theta})$ accordingly. Since (4.24) and our assumptions imply $\hat{V}_n^0 - \tilde{V}_n^0 = o_P(1)$ uniformly in $h \in H_n$, where here

$$\left[\tilde{V}_n^0\right]^2 = \frac{2}{n(n-1)h^p} \sum_{i\neq j} \sigma^{*\,2}(X_i)\sigma^{*\,2}(X_j) K_h^2(X_i - X_j), \tag{4.26}$$

the new test statistic $T_n^0(\hat{\theta})$ has the same standard normal asymptotic law under H_0 and potentially leads to a more powerful test. In Lemma A.11 in the Appendix we provide a set of sufficient conditions for obtaining $\sup_{x\in\mathcal{X}} |\hat{\sigma}_n^{*2}(x) - \sigma_n^{*2}(x)| \to 0$, in probability, regardless of whether H_0 is true, where $\sigma_n^{*2}(\cdot)$ is defined like $\hat{\sigma}_n^{*2}(\cdot)$ but with estimated synthetic observations $\hat{Y}_i^*$ replaced with the true (unknown) ones Y_i^*. To obtain (4.24), our result can be completed by the arguments for i.i.d. data like in Horowitz and Spokoiny (2001) or Guerre and Lavergne (2005) allowing to deduce $\sup_{x\in\mathcal{X}} |\sigma_n^{*2}(x) - \sigma^{*2}(x)| \to 0$ in probability. In the WLS approach, the question of how to build an estimate of the variance of $nh^{p/2}Q_n^1(\hat{\theta})$ that (theoretically) performs better than $\hat{V}_n^1$ when H_0 is not true seems harder and therefore is left open. ■

Remark 2. The tests we propose depend on the choice of the smoothing parameter $h \in \mathcal{H}_n$. Following a well-known data-driven method for choosing the smoothing parameter, in the synthetic data approach we can define

$$T_n^{opt} = \max_{h\in\mathcal{H}_{1n}} T_n^0(\hat{\theta}) \tag{4.27}$$

where the maximum is taken over a finite subset $\mathcal{H}_{1n} \subset \mathcal{H}_n$. Typically, $\mathcal{H}_{1n}$ is a geometric grid in $\mathcal{H}_n$ and the number of elements in $\mathcal{H}_{1n}$ increases as $n \to \infty$. See Horowitz and Spokoiny (2001). The resulting test is

"Reject H_0 when $T_n^{opt} \geq t_\alpha^{opt}$ ",

where t_α^{opt} is the $\alpha-$level critical value for T_n^{opt}. Like in the non-censored case, this critical value cannot be evaluated in applications because θ_0 and the law of the errors ε_i are unknown. Horowitz and Spokoiny (2001) proposed a simulation procedure for approximating the critical value t_α^{opt}. Their procedure can be adapted to our synthetic data based test. To build the test statistic $T_n^0(\hat{\theta})$ to be compared to the approximate critical value, we propose the use of the standard deviation estimate $\hat{V}_n^0$ based on the nonparametric estimate $\hat{\sigma}_n^{*2}(X_i)$ introduced in equation (4.25).

1. (*Create synthetic observations*) For each $i = 1, ..., n$, generate $Y_i^{*,b} = f(\hat{\theta}, X_i) + \omega_i^b$, where ω_i^b is sampled randomly from the normal distribution $N\left[0, \hat{\sigma}_n^{*2}(X_i)\right]$ and $\hat{\sigma}_n^{*2}(\cdot)$ is defined in (4.23).

2. (*Build the test statistic with synthetic data*) Use the data $\left\{Y_i^{*,b}, X_i : i = 1, ..., n\right\}$ to estimate θ by ordinary least squares, that is minimizing $\sum_{i=1}^n \left[Y_i^{*,b} - f(\theta, X_i)\right]^2$ with respect to θ, and to estimate $\sigma^{*2}(\cdot)$ nonparametrically by replacing $\hat{Y}_i^*$ with $Y_i^{*,b}$ in (4.23). Denote the resulting estimates by $\hat{\theta}^b$ and $(\sigma_n^{*,b})^2(\cdot)$, respectively. For each $h \in \mathcal{H}_{1n}$, compute the statistic $\tilde{T}_n^0(\hat{\theta}) = nh^{p/2}\tilde{Q}_n^0(\hat{\theta})/\hat{V}_n^0$ that is obtained by replacing Y_i^* and $\hat{\theta}$ with $Y_i^{*,b}$ and $\hat{\theta}^b$ in the definition of $\tilde{Q}_n^0(\hat{\theta})$, and $\hat{\sigma}_n^{*2}(X_i)$ with $(\sigma_n^{*,b})^2(X_i)$ on the right-side hand of (4.25). Take the maximum of $\tilde{T}_n^0(\hat{\theta})$ over $h \in \mathcal{H}_{1n}$ to compute a value of $T_n^{opt,b}$.

3. Estimate t_α^{opt} by the $(1-\alpha)$th quantile of the empirical distribution of $T_n^{opt,b}$ that is obtained by repeating steps 1 and 2 many times.

It is worthwhile to notice that *no Kaplan-Meier estimate* is involved in this simulation procedure. The uniformity with respect to h of the convergence stated in Theorem 4.1 guarantees the asymptotic validity of this simulation procedure for approximating t_α^{opt} as soon as this procedure is asymptotically correct with synthetic (non-censored) responses. See Horowitz and Spokoiny (2001) for a set of technical conditions ensuring the asymptotic validity of the simulation procedure with non-censored responses in a related test. ■

4.3 Behavior of the tests under the alternatives

Consider a sequence of measurable functions $\lambda_n(x)$, $n \geq 1$, and the sequence of alternatives

$$H_{1n} : Y_{in} = f(\theta_0, X_i) + \lambda_n(X_i) + \varepsilon_i, \qquad 1 \leq i \leq n. \tag{4.28}$$

For simplicity, assume that there exists some constant M_λ such that for all $n \geq 1$, $0 \leq |\lambda_n(\cdot)| \leq M_\lambda < \infty$.

Assumption 8 *(i) The censoring times $C_1, ..., C_n$ represent an independent sample from the continuous distribution function G (the same for each n) and are independent of the variables $Y_{1n}, ..., Y_{nn}$ with continuous distribution function $F^{(n)}$.*

(ii) For each n, $\mathbb{P}(Y_{1n} \leq C_1 \mid X_1, Y_{1n}) = \mathbb{P}(Y_{1n} \leq C_1 \mid Y_{1n})$.

Notice that the second part of this assumption is always true if C is independent of ε and X. Now, for each n define $T_{in} = Y_{in} \wedge C_i$ and $\delta_{in} = \mathbf{1}_{\{Y_i \leq C_i\}}$, $i = 1, ..., n$, and let $H^{(n)}$ denote the distribution function of $T_{1n}, ..., T_{nn}$, that is $H^{(n)}(y) = \mathbb{P}(T_{1n} \leq y)$. Let us point out that the two test statistics we propose rely on the Kaplan-Meier estimator that is computed from the observations (T_{in}, δ_{in}), $i = 1, ..., n$. If $\lambda_n(\cdot)$ changes with n, the law of the observations is different for each n. Therefore, in order to control the jumps of the Kaplan-Meier estimator and the conditional variance of the residuals $U_i^\beta(\theta)$ we need the following assumption.

Assumption 9 *(i) There exist some constants c_{inf}, c_{sup} such that for each $x \in \mathcal{X}$*

$$0 < c_{inf} \leq \mathbb{E}\left[\varepsilon^2 \mid X = x\right] \leq \mathbb{E}\left[\left\{1 + \varepsilon^2\right\}\{1 - G(Y_{1n})\}^{-1} \mid X = x\right] \leq c_{sup} < \infty.$$

(ii) There exists some constant M such that for all $n \geq 1$, $\mathbb{E}\left[\{1 + \varepsilon^4\}\,\gamma(Y_{1n})^4\right] \leq M < \infty$ where $\gamma(Y_{1n}) = \delta_{1n}\{1 - G(Y_{1n})\}^{-1}$.

(iii) Let $F_{Y|X=x}^{(n)}(y) = \mathbb{P}(Y_{1n} \leq y \mid X_1 = x)$ and

$$q_\rho^{(n)}(x) = \int \{|y| + 1\} C^{(n)}(y)^{1/2+\rho} dF_{Y|X=x}^{(n)}(y)$$

where

$$C^{(n)}(y) = \int_{-\infty}^{y} \frac{dG(t)}{[1 - H^{(n)}(t)][1 - G(t)]} \vee 1.$$

There exist $0 < \rho < 1/2$ and a function $q_\rho(x)$ with $\mathbb{E}[q_\rho^2(X)] < \infty$ such that for all n, $0 \leq q_\rho^{(n)} \leq q_\rho$.

Let $\hat{V}_n^\beta(\theta)^2$ be the estimator obtained after replacing $\hat{\theta}$ with θ on the right-hand side of (3.16). Once again, our purpose is to transfer the problem of consistency against the alternatives H_{1n} in classical i.i.d. framework. The first step in this transfer is realized in a general setup in the following lemma proved in the Appendix. Next, we will be more specific on the type of alternatives considered in order to derive the asymptotic consistency.

Lemma 4.3 *Let Assumptions 4-(i) and (ii), 5, 6, 8 and 9-(ii) and (iii) hold true. Then, under the alternatives H_{1n}, for $\beta = 0$ or 1*

$$\left|Q_n^\beta(\theta) - \tilde{Q}_n^\beta(\theta)\right| \leq \left[\tilde{Q}_n^\beta(\theta) + R_{n1}\right]^{1/2} R_{n2}^{1/2} - R_{n3} + R_{n2} - R_{n4}$$

with $\sup_{\theta \in \Theta,\, h \in \mathcal{H}_n} \left\{h^p\,|R_{n1}| + |R_{n2}| + h^{p/2}\,|R_{n3}| + |R_{n4}|\right\} = O_P(n^{-1})$.

4.3.1 Consistency against a fixed alternative

We now investigate the consistency of our tests against a fixed alternative

$$H_1 : Y = m(X) + \varepsilon,$$

where $\mathbb{E}(\varepsilon \mid X) = 0$ a.s. and, for simplicity, we assume $0 \le |m(\cdot)| \le M_\lambda < \infty$ for some constant M_λ. The following assumption identifies the limit of $\hat{\theta}$ the SD or WLS estimator and states that the regression model is wrong.

Assumption 10 *There exists $\bar{\theta}$ an interior point of Θ such that*

$$\text{for any } \theta \in \Theta \setminus \{\bar{\theta}\}, \quad 0 < \mathbb{E}\left[\left\{m(X) - f\left(\bar{\theta}, X\right)\right\}^2\right] < \mathbb{E}\left[\{m(X) - f(\theta, X)\}^2\right].$$

Theorem 4.4 *Let Assumption 10, Assumption 9-(i) and the assumptions of Lemma 4.3 hold true. Under H_1, for $\beta = 0$ or 1*

$$\sup_{h\in\mathcal{H}_n} \left| Q_n^\beta(\hat{\theta}) - \mathbb{E}\left[\left\{m(X) - f\left(\bar{\theta}, X\right)\right\}^2 g(X)\right]\right| = o_P(1) \quad \text{and} \quad \sup_{h\in\mathcal{H}_n} |\hat{V}_n^\beta - c| = o_P(1),$$

where $c > 0$ is some positive constant. Consequently, the tests in (3.15) are consistent.

See the Appendix for the proof. It is worthwhile to notice that the limit of $Q_n^\beta(\hat{\theta})$ under the alternative H_1 does not depend on the censoring and is the same for $\beta = 0$ or $\beta = 1$. However, the limits of the standard deviations $\hat{V}_n^\beta$ depend on β and the degree of censoring in the data (see the Appendix for the expressions of these limits). In general our tests lose power if the degree of censoring increases. Looking at the limits of $\hat{V}_n^\beta$, one concludes that none of the two tests we propose is more powerful than the other, that is depending on the laws of Y and C, either the SD or the WLS test will have better finite sample properties.

4.3.2 Consistency against Pitman local alternatives

Let $\lambda(\cdot)$ be a bounded measurable function of X and consider the Pitman alternatives

$$H_{1n} : \ Y_{in} = f(\theta_0, X_i) + r_n\lambda(X_i) + \varepsilon_i, \qquad 1 \le i \le n,$$

with $r_n \to 0$ when $n \to \infty$. For simplicity, we will assume that $\lambda(\cdot)$ is a bounded function and $\mathbb{E}[\lambda(X)\nabla_\theta f(\theta_0, X)] = 0$. The later condition will make the estimator $\hat{\theta}$ to converge to θ_0 at the rate $O_P(n^{-1/2})$. See Lemma A.12 in the Appendix. The following result (see the Appendix for the proof) implies that our tests are consistent against the Pitman local alternatives H_{1n}, provided that r_n decreases to zero slower than $n^{-1/2}h^{-p/4}$.

Theorem 4.5 *Let Assumption 9-(i) and the assumptions of Lemma 4.3 hold. Suppose that $\lambda(\cdot)$ is bounded and $\mathbb{E}\left[\lambda(X)\nabla_\theta f\left(\theta_0, X\right)\right] = 0$. Under H_{1n}, for $\beta = 0$ or 1 the test statistics $T_n^\beta(\hat{\theta})$ converge in law to a normal distribution $N(\mu, 1)$, with $\mu > 0$, provided that $r_n = n^{-1/2}h^{-p/4}$.*

4.3.3 Consistency against a sequence of smooth alternatives

In this section we provide conditions under which our tests are consistent against alternatives H_{1n} like in (4.28) defined by functions $\lambda_n(\cdot)$ in a Hölder smoothness class that vanish as $n \to \infty$. The regularity s of the Hölder class is supposed *known* and the rate to which the functions $\lambda_n(\cdot)$ approach zero can be made arbitrarily close to the optimal rate of testing $n^{-2s/(4s+p)}$, provided that $s > 5p/4$. Note that we have to impose this more restrictive condition on the regularity s (the usual condition is $s \geq p/4$, see, for instance, Horowitz and Spokoiny 2001) because of our conditions on the left endpoint of the bandwidth range $\mathcal{H}_n$. See Assumption 6-(ii) and the subsequent comments. For s and $L > 0$, define the Hölder class $C(L, s)$ as

$$C(L, s) = \left\{ f(\cdot) : |f(x_1) - f(x_2)| \leq L|x_1 - x_2|^s, \quad \forall x_1, x_2 \in \mathcal{X} \right\}, \quad \text{for } s \in (0, 1],$$

while for $s > 1$, $C(L, s)$ is the class of functions having the $[s]$-th partial derivatives in $C(L, s - [s])$, where $[s]$ denotes the integer part of s. As a corollary of the following theorem one may deduce that the optimal rate of testing parametric mean-regressions when s is known is not altered by the censorship, provided that $s > 5p/4$. The proof of the theorem is postponed to the Appendix.

Theorem 4.6 *Let Assumption 9-(i) and the assumptions of Lemma 4.3 hold. Moreover, the density $g(\cdot)$ is bounded from below by a positive constant. Let κ_n, $n \geq 1$ be a sequence of positive real numbers. Consider a sequence of functions $\lambda_n(\cdot)$ such that for all $n \geq 1$, $\lambda_n(\cdot) \in C(L, s)$ for some known $s > 5p/4$ and some $L > 0$. Moreover, $\mathbb{E}\left[\lambda_n^2(X)\right] \to 0$ as*

$n \to \infty$ and for each $n \geq 1$, $\mathbb{E}[\lambda_n (X) \nabla_\theta f (\theta_0, X)] = 0$ and

$$\|\lambda_n\|_n := \left[n^{-1} \sum_{i=1}^{n} \lambda_n^2(X_i) \right]^{1/2} \geq \kappa_n n^{-\frac{2s}{4s+p}}. \tag{4.29}$$

If h is of order $n^{-2/(4s+p)}$, the tests defined in (3.15) are consistent against the alternatives H_{1n} defined by the functions $\lambda_n(\cdot)$ whenever κ_n diverges.

Remark 2 ***(continued).*** In Theorem 4.6 we supposed that the regularity s is known and thus the rate of the bandwidth that allows to detect departures from the null hypothesis like in (4.29) is known. More generally, it would be useful to propose a data-driven selection procedure for h that adapts to the unknown smoothness of the functions $\lambda_n(\cdot)$ and that allows these functions to converge to zero at a rate which is arbitrarily close to the fastest possible rate. In the case of non-censored, if s is unknown but $s \geq p/4$, the optimal rate of testing is $(n^{-1}\sqrt{\log \log n})^{2s/(4s+p)}$, see for instance Horowitz and Spokoiny (2001). The maximum test procedure (4.27) represent a potential solution in the synthetic data testing approach. Consider the test statistic built with the true synthetic observations and the true value of the parameter θ_0, $\tilde{T}_n^0(\theta_0) = nh^{p/2}\tilde{Q}_n^0(\theta_0)/\tilde{V}_n^0$ with $\tilde{V}_n^0$ defined like in (4.26). Suppose that under the alternatives H_{1n} defined by functions $\lambda_n(\cdot)$ like in Theorem 4.6 with some $\kappa_n \uparrow \infty$,

$$\lim_{n \to \infty} \mathbb{P} \left(\max_{h \in \mathcal{H}_{1n}} \tilde{T}_n^0(\theta_0) \geq t_\alpha^{opt} \right) = 1. \tag{4.30}$$

By Lemma 4.3 and the arguments used in the proof of Theorem 4.6 to replace $\hat{\theta}$ by θ_0,

$$\lim_{n \to \infty} \mathbb{P} \left(\max_{h \in \mathcal{H}_{1n}} T_n^0(\hat{\theta}) \geq t_\alpha^{opt} \right) = 1,$$

provided that $\lambda_n(\cdot)$ also satisfy condition (4.29) when κ_n is replaced by $c\,\kappa_n$ with c some constant greater but arbitrarily close to 1. In view of the proof of our Theorem 4.6, it is expected that any sequence κ_n such that $\kappa_n\,[\log \log n]^{-s/(4s+p)} \to \infty$ should suffices to obtain (4.29) provided that $\mathcal{H}_{1n}$ is a geometric grid like in Horowitz and Spokoiny (2001). However, the deeper investigation of the conditions allowing to obtain (4.30) is left for future work. ∎

5 Simulation study

The purpose of the small simulation study presented below was to compare in finite samples the new tests with the tests of Stute *et al.* (2000) based on their statistics D_n and W_n^2. The regression model considered was $Y = \theta_{01} + \theta_{02}X + \varepsilon$ with X uniformly distributed on the interval $[-\sqrt{3}, \sqrt{3}]$ and ε a standard normal residual term. A linear regression function appears, for instance, in the so-called accelerated failure time (AFT) model that has found considerable interest in the survival data literature. Here, we took the parameters $(\theta_{01}, \theta_{02}) = (1, 3)$ and considered a censoring variable C with an exponential distribution of mean μ. The parameter μ served to control the proportion of censored observations that was fixed to 30%, 40% or 50%.

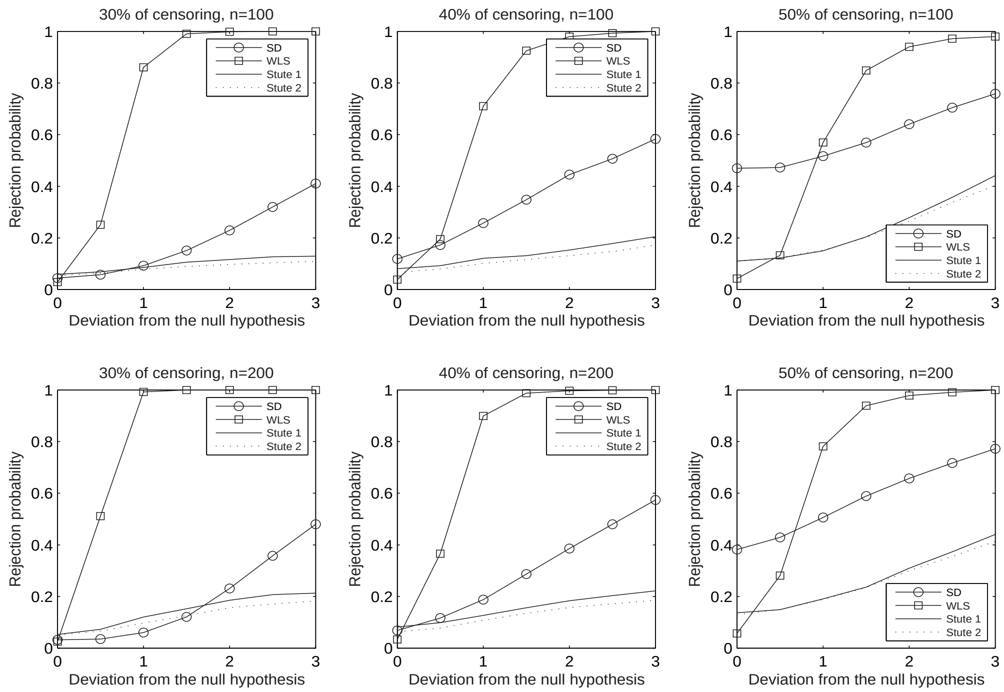

Figure 1: Rejection probabilities for the test based on T_n^{SD}, T_n^{WLS}, D_n (Stute 1) and W_n^2 (Stute 2) test statistics.

The linear regression model was tested against alternatives with the form

$$H_1 : \; Y_i = \theta_{01} + \theta_{02}X + d\cos(2\pi(X_i/\sqrt{3})) + \varepsilon_i, \qquad 1 \le i \le n$$

with $d \in \{0.5, 1, ..., 2.5, 3\}$. The way the alternatives were defined rendered the amount of censoring practically stable on the null and under the alternatives. The levels considered were $\alpha = 0.05$ and $\alpha = 0.1$, the sample sizes were taken $n = 100$ and $n = 200$ and for each sample size we generated 5000 samples. We used the bandwidth $h = 0.1$ for the kernel-based tests. The test statistic T_n^{SD} (resp. T_n^{WLS}) was built using the estimator $\hat{\theta}^{SD}$ (resp. $\hat{\theta}^{WLS}$). The critical values for our tests were those given by the standard normal law while for the test proposed by Stute *et al.* (2000) we followed their bootstrap procedure (with 5000 bootstrap samples) to compute the critical values. The asymptotic distribution of test statistics D_n and W_n^2 used by Stute *et al.* (2000) depend on the asymptotic distribution of the estimator of θ_0. To focus the attention on the performances of the testing approaches, we computed the values of D_n and W_n^2 using the true values of the parameters θ_{01}, θ_{02}. This resulted in improved rejection probabilities under the null and under the alternatives for the corresponding tests. The results of the simulations are presented in Figure 1. To save space, only the case $\alpha = 0.05$ is reported, the results obtained with $\alpha = 0.1$ being very similar.

This preliminary empirical investigation shows that in the setup considered, the test based on T_n^{WLS} outperforms the test built with T_n^{SD} and the tests obtained with the weighted marked empirical process approach of Stute *et al.* (2000). The level of the WLS kernel-based test is satisfactory close to the nominal level for all probabilities of censoring considered. On contrary, the level of the SD-based test drastically deteriorates when the probability of censoring increases. With a few minor exceptions, the rejection probabilities under the alternatives are higher for the kernel-based tests (even much higher for the WLS test) than for the tests based on the marked empirical process approach.

Appendix

First, we prove some technical lemmas that will be used in the proofs of our main results. In the following, $H(t) = \mathbb{P}(T \leq t)$ and $\tau_H = \inf\{t : H(t) = 1\}$. Moreover, we refer to Nolan and Pollard (1987) and Sherman (1994) for the definition of Euclidean classes of functions. Finally, M, c, c_1, ... are constants that may be different from line to line.

A.1 Technical lemmas

The point (ii) of the following lemma is a key ingredient. It provides a bound for the difference between the weights W_{in} built with $\hat{G}$ and the ideal weights one would obtain if G were known. In the following result, for each sample size n, the lifetimes Y are supposed independent with a same law which may depend on n. This generality is needed under a sequence of alternatives approaching the null hypothesis.

Lemma A.1 *Let $Y_{1n}, ..., Y_{nn}$ be an independent sample from a continuous distribution function $F^{(n)}$, $n \geq 1$. Independent of these, let $C_1, ..., C_n$ be an independent sample from a continuous distribution function G (the same for each n). Let $T_{in} = Y_{in} \wedge C_i$ and $\delta_{in} = \mathbf{1}_{\{Y_{in} \leq C_i\}}$, $i = 1, ..., n$, and for each n, let $H^{(n)}$ denote the distribution function of $T_{1n}, ..., T_{nn}$. Denote $\gamma(T_{in}) = \delta_{in} [1 - G(T_{in})]^{-1}$ and let $T_{(n)n} = \max_{1 \leq i \leq n} T_{in}$. Then,*

i)

$$\sup_{1 \leq i \leq n} \frac{1 - \hat{G}(T_{in}-)}{1 - G(T_{in})} = O_P(1) \quad \text{and} \quad \sup_{1 \leq i \leq n} \frac{1 - G(T_{in})}{1 - \hat{G}(T_{in}-)} = O_P(1)\,; \tag{A.1}$$

ii) for all $0 \leq \alpha \leq 1/2$ and $\eta > 0$,

$$|nW_{in} - \gamma(T_{in})| \leq \frac{\delta_{in}}{1 - G(T_{in})} \{C^{(n)}(T_{in})\}^{\alpha + \eta} \times O_P\left(n^{-\alpha}\right),$$

where the $O_P(n^{-\alpha})$ factor does not depend on i.

Proof. Since we only consider the distribution functions $\hat{G}$ and G at the sample points, we can transform data and suppose without loss of generality for this proof that the variables $Y_{1n}, ..., Y_{nn}$ and $C_1, ..., C_n$ are nonnegative.

i) Since by assumption $P(Y_{in} = C_i) = 0$, we can redefine $1 - \delta_{in} = \mathbf{1}_{\{C_i \leq Y_{in}\}}$ and study $\hat{G}$ as the Kaplan-Meier estimator of the lifetimes C_i in presence of the censoring times Y_{in}. The first part of (A.1) follows from Theorem 3.2.4 in Fleming and Harrington (1991) (or Lemma 2.6 of Gill, 1983). The fact the distribution of the i.i.d. variables $Y_{1n}, ..., Y_{nn}$ depends on n is of no consequence for the continuous time martingale arguments applied for each n in the proof of the Theorem 3.2.4 in Fleming and Harrington (1991). The second part of can be obtained for instance as a consequence of Theorem 2.2 in Zhou (1991). Once again, a careful inspection of Zhou's proofs shows that his arguments apply for each n and therefore his Theorem 2.2 apply to our case.

ii) Fix $\eta > 0$ arbitrarily. First, we need to prove that

$$\sup_{y \leq T_{(n)n}} \left[C^{(n)}(y)\right]^{-1/2-\eta} \left|Z^{(n)}(y)\right| = O_P(1), \tag{A.2}$$

where for each $n \geq 1$

$$Z^{(n)}(y) = \sqrt{n}\frac{\hat{G}(y) - G(y)}{1 - G(y)}$$

is the Kaplan-Meier process and

$$C^{(n)}(y) = \int_0^y \frac{dG(t)}{[1 - H^{(n)}(t)]\,[1 - G(t)]} \vee 1.$$

Notice that for any $\eta > 0$, $\left[C^{(n)}(\cdot)\right]^{-1/2-\eta}$ is a continuous, nonnegative, non-increasing and nonrandom function. For each $n \geq 1$, by Lemma 2.9 of Gill (1983),

$$\sup_{y \leq T_{(n)n}} \left[C^{(n)}(y)\right]^{-1/2-\eta} \left|Z^{(n)}(y)\right| \leq 2 \sup_{y \leq T_{(n)n}} \left|\int_0^y \left[C^{(n)}(t)\right]^{-1/2-\eta} dZ^{(n)}(t)\right|. \tag{A.3}$$

Now, for each $n \geq 1$, proceed as in the proof of Theorem 2.1 of Gill (1983). That is, for each τ' such that $H^{(n)}(\tau'-) < 1$, we have for any $\upsilon > 0$ by the inequality of Lenglart (see, e.g., Theorem 3.4.1 in Fleming and Harrington, 1991)

$$\begin{aligned} &\mathbb{P}\left[\sup_{y \leq \tau' \wedge T_{(n)n}} \left|\int_0^y \left[C^{(n)}(t)\right]^{-1/2-\eta} dZ^{(n)}(t)\right| > \epsilon\right] \\ \leq\;& \frac{\upsilon}{\epsilon^2} + \mathbb{P}\left[\left|\int_0^{\tau' \wedge T_{(n)n}} \left[C^{(n)}(t)\right]^{-1-2\eta} \frac{\left[1 - \hat{G}(t-)\right]^2}{[1 - G(t)]^2} \frac{n}{R_n(t)} \frac{dG(t)}{1 - G(t)}\right| > \upsilon\right] \end{aligned}$$

where $R_n(t) = \sum_{i=1}^n \mathbf{1}_{\{T_{in} \geq t\}}$. Next, use Theorem 3.2.4 in Fleming and Harrington (1991) and Lemma 2.7 of Gill (1983) to obtain

$$\begin{aligned} \mathbb{P}\left[\sup_{y \leq \tau' \wedge T_{(n)n}} \left|\int_0^y \left[C^{(n)}(t)\right]^{-1/2-\eta} dZ^{(n)}(t)\right| > \epsilon\right] \;\leq\;& \frac{\upsilon}{\epsilon^2} + \xi + (1/\xi)\exp(1 - 1/\xi) \\ &+ \mathbb{P}\left[\left|\int_0^{\tau'} \frac{\xi^{-3}\left[C^{(n)}(t)\right]^{-1-2\eta}}{1 - H^{(n)}(t)} \frac{dG(t)}{1 - G(t)}\right| > \upsilon\right] \end{aligned}$$

for any $\xi \in (0, 1)$. The fact $H^{(n)}$, the distribution of the observations, depends on n is of no consequence for the continuous time martingale arguments applied for each n in the proof of Lemma 2.7 of Gill (1983). Letting $\tau' \uparrow \tau_{H^{(n)}} = \inf\{t : H^{(n)}(t) = 1\}$ and choosing a finite constant υ independent of n such that

$$\upsilon \geq \xi^{-3} \int_0^{\tau_{H^{(n)}}} \left[C^{(n)}(t)\right]^{-1-2\eta} dC^{(n)}(t)$$

(such a υ exists because $C^{(n)}(\cdot) \geq 1$ and $C^{(n)}(\tau_{H^{(n)}}) = \infty$) we obtain

$$\sup_{y \leq \tau' \wedge T_{(n)n}} \left| \int_0^y \left[C^{(n)}(t) \right]^{-1/2-\eta} dZ^{(n)}(t) \right| = O_P(1).$$

Finally, use (A.3) to derive (A.2). Now, by definition, (A.2) and (A.1)

$$\begin{aligned} nW_{in} - \gamma(T_{in}) &= \frac{\delta_{in}}{1 - G(T_{in})} \frac{\hat{G}(T_{in}-) - G(T_{in})}{1 - G(T_{in})} \frac{1 - G(T_{in})}{1 - \hat{G}(T_{in}-)} \\ &= \frac{\delta_{in} n^{-\alpha}}{1 - G(T_{in})} \{C^{(n)}(T_{in})\}^{\alpha+\eta} \\ &\quad \times \left| \{C^{(n)}(T_{in})\}^{-1/2-\eta/2\alpha} Z^{(n)}(T_{in}-) \right|^{2\alpha} \\ &\quad \times \left[\frac{\hat{G}(T_{in}-) - G(T_{in})}{1 - G(T_{in})} \right]^{1-2\alpha} \frac{1 - G(T_{in})}{1 - \hat{G}(T_{in}-)} \\ &= \frac{\delta_{in}}{1 - G(T_{in})} \{C^{(n)}(T_{in})\}^{\alpha+\eta} \times O_P\left(n^{-\alpha}\right), \end{aligned}$$

with the $O_P(n^{-\alpha})$ factor independent of i. ■

The following lemma will help to state our equivalence results uniformly with respect to h in an interval.

Lemma A.2 *Let $v_1, ..., v_n$ be a sequence of real numbers and $0 < h_m \leq h_M < \infty$. Suppose that Assumption 6-(i) holds true. If*

$$U(h) = \frac{1}{n^2 h^p} \sum_{1 \leq i \neq j \leq n} v_i v_j K_h(X_i - X_j) \quad \text{and} \quad D(h) = \frac{K(0)}{n^2 h^p} \sum_{i=1}^n v_i^2,$$

then for any $h \in [h_m, h_M]$

$$U(h_M) + D(h_M) - D(h_m) \leq U(h) \leq U(h_m) + D(h_m) - D(h_M).$$

Proof. First, consider $p = 1$. Using the Inverse Fourier Transform,

$$U(h) = \int \hat{K}(hu) \left| \frac{1}{n} \sum_{i=1}^n v_i \exp(2i\pi u' X_i) \right|^2 du - D(h) = \tilde{U}(h) - D(h).$$

Now, by the properties of K, $\hat{K}(h_M u) \leq \hat{K}(hu) \leq \hat{K}(h_m u)$ and thus

$$U(h) = \tilde{U}(h) - D(h) \leq \tilde{U}(h_m) - D(h_M) = U(h_m) + D(h_m) - D(h_M).$$

The other inequality follows similarly. For $p \geq 1$, K is a product of univariate kernels and the argument with one regressor applies componentwise. ■

Let A_h be the $n \times n$ symmetric matrix with generic element

$$a_{ij}(h) = [h^p n(n-1)]^{-1} K_h(X_i - X_j)\mathbf{1}_{\{i \neq j\}}. \tag{A.4}$$

Lemma A.3 *Let $v_1, ..., v_n$ and $w_1, ..., w_n$ be sequences of real numbers. Suppose that Assumptions 4 (i)-(ii) and 6 (ii) hold true. If*

$$U(h) = \frac{1}{n^2 h^p} \sum_{1 \leq i \neq j \leq n} v_i w_j K_h(X_i - X_j),$$

then

$$\sup_{h \in \mathcal{H}_n} |U(h)| \leq O_P(1) \left[\frac{1}{n} \sum_{i=1}^n v_i^2 \right]^{1/2} \left[\frac{1}{n} \sum_{i=1}^n w_i^2 \right]^{1/2}$$

Proof. Since for any $n-$dimensional vectors z_1, z_2, $|z_1' A_h z_2| \leq \|A_h\|_2 \|z_1\| \|z_2\|$, it suffices to suitably bound $\|A_h\|_2$ uniformly in h. By simple algebra,

$$\text{for any } z \in \mathbb{R}^n, \quad \|A_h z\|^2 \leq \left[\max_{1 \leq i \leq n} \left(\sum_{j=1, j \neq i}^n a_{ij}(h) \right) \right]^2 \|z\|^2.$$

Hence

$$\|A_h\|_2 \leq \frac{\Delta_n h^{-p}}{n-1} + \frac{1}{(n-1)} \sup_{h>0,\, x \in \mathbb{R}^p} \mathbb{E}\left[h^{-p} K_h(x - X)\right] + \frac{K(0) h^{-p}}{n(n-1)}$$

where

$$\Delta_n = \sup_{h>0,\, x \in \mathbb{R}^p} \left| \frac{1}{n} \sum_{j=1}^n \{K_h(x - X_j) - \mathbb{E}[K_h(x - X_j)]\} \right|.$$

By a change of variables and the boundedness of the density g

$$\mathbb{E}\left[h^{-p} K_h(x - X)\right] = \int_{\mathbb{R}^p} K(x') g(x - hx') dx' \leq c \tag{A.5}$$

for some constant $c > 0$. Hence, for any $h \in H_n$

$$\|A_h\|_2 \leq \frac{c}{n} \left\{ 1 + \frac{\Delta_n}{h^p} \right\} \leq \frac{c}{n} \left\{ 1 + \frac{\Delta_n}{h_{min}^p} \right\}$$

for some $c > 0$ independent of h. By Lemma 22(ii) of Nolan and Pollard (1987) and the rate of an empirical process indexed by an Euclidean family for a constant envelope (e.g., Pakes and Pollard 1989, van der Vaart and Wellner 1996), $\Delta_n = O_P(n^{-1/2})$. The result follows as $nh_{min}^{2p} \to \infty$. ■

Lemma A.4 *Suppose that Assumptions 4-(ii), 6 and 7 hold true. Then:*

$$\sup_{h\in\mathcal{H}_n} \mathbb{E}\left[q_\rho(X_1)q_\rho(X_2)h^{-p}K_h\left(X_1-X_2\right)\right] \leq M$$

where M is some finite constant and

$$\sup_{h\in\mathcal{H}_n} \mathbb{E}\left[q_{\rho,\tau}(X_1)q_{\rho,\tau}(X_2)h^{-p}K_h\left(X_1-X_2\right)\right] \to 0 \qquad as\ \tau\uparrow\tau_H,$$

where $q_{\rho,\tau}(x) = \mathbb{E}\left[\{|Y|+1\}\mathbf{1}_{\{Y>\tau\}}C(Y)^{1/2+\rho} \mid X=x\right]$, $x\in\mathcal{X}$.

Proof. Apply the Inverse Fourier Transform and use the fact that $\hat{K}$ is nonnegative and bounded to write

$$\begin{aligned}\left|\mathbb{E}\left[q_{\rho,\tau}(X_1)q_{\rho,\tau}(X_2)h^{-p}K_h\left(X_1-X_2\right)\right]\right| &= \int \left|\widehat{q_{\rho,\tau}g}(u)\right|^2\hat{K}(hu)du \\ &\leq \int \left|\widehat{q_{\rho,\tau}g}(u)\right|^2 du = \mathbb{E}[q^2_{\rho,\tau}(X)g(X)],\end{aligned}$$

where for the last equality use Parseval's identity with the function $q_{\rho,\tau}(\cdot)g(\cdot)\in L^1(\mathbb{R}^p)\cap L^2(\mathbb{R}^p)$ (see Rudin 1987). Now, for each $x\in\mathcal{X}$, $q_{\rho,\tau}(x)\downarrow 0$ as $\tau\uparrow\tau_H$. Assumption 7 and Lebesgue's Dominated Convergence Theorem yield the second statement. For the first quantity in the statement we can write

$$\begin{aligned}\left|\mathbb{E}\left[q_\rho(X_1)q_\rho(X_2)h^{-p}K_h\left(X_1-X_2\right)\right]\right| &= \int \left|\widehat{q_\rho g}(u)\right|^2 du \\ &\quad + \int \left|\widehat{q_\rho g}(u)\right|^2\left[1-\hat{K}(hu)\right]du.\end{aligned}$$

Since for each u, $0\leq 1-\hat{K}(hu)\leq 1-\hat{K}(h_{min}u)\downarrow 0$, by Parseval's identity and dominated convergence deduce that the expectation in the last display converges to $\mathbb{E}[q^2_\rho(X)g(X)]<\infty$ uniformly in $h\in\mathcal{H}_n$. This implies the first part of the statement. ■

Lemma A.5 *Let $X_1, X_2, ...$ be a sample as in Assumption 4-(i) and (ii) and let Assumption 6 hold true. For each $n\geq 1$, let $u_{1n},...,u_{nn}$ be a sequence of random variables that are independent given $X_1,...,X_n$. For each n and i, the law of u_{in} given $X_1,...,X_n$ depends only on X_i. Assume $\mathbb{E}\left(u_{in}\mid X_i\right)=0$ and $\mathbb{E}\left(u^2_{in}\mid X_i\right)=\sigma^2_n\left(X_i\right)$ and suppose that for each x and n we have $0\leq\sigma^2_n\left(x\right)\leq\overline{\sigma}^2_n<\infty$. Then*

$$\frac{1}{n\left(n-1\right)}\sum_{1\leq i\neq j\leq n} u_{in}u_{jn}\frac{1}{h^p}K_h\left(X_i-X_j\right) = \overline{\sigma}^2_n O_P\left(n^{-1}h^{-p/2}\right). \tag{A.6}$$

Let $\lambda_n(\cdot)$, $n \geq 1$ be a sequence of measurable functions and let

$$U_n = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \lambda_n(X_i) u_{jn} \frac{1}{h^p} K_h(X_i - X_j).$$

If A_h is the matrix defined in (A.4) and $\|\lambda_n\|_n^2$ denotes $n^{-1}\sum_{i=1}^n \lambda_n^2(X_i)$, then

$$\mathbb{E}\left[|U_n| \mid X_1, ..., X_n\right] \leq c\,\overline{\sigma}_n n^{1/2} \|A_h\|_2 \|\lambda_n\|_n \tag{A.7}$$

for some finite constant c independent of n and of the sequence $\lambda_n(\cdot)$, $n \geq 1$.

Proof. By elementary calculus, the variance of the degenerate $U-$statistic in (A.6) is of order $O_P(n^{-2}h^{-p})$ and thus we obtain stated rate from Chebyshev's inequality. Next, following the lines of Guerre and Lavergne (2005), let

$$\overline{\lambda}_n(X_i) = \frac{1}{n(n-1)} \sum_{j=1, i\neq j}^{n} \lambda_n(X_j) \frac{1}{h^p} K_h(X_i - X_j)$$

and use Marcinkiewicz-Zygmund inequality (e.g., Chow and Teicher, 1997, page 386), Jensen's inequality and the properties of the $\|\cdot\|_2$ to write

$$\begin{aligned} \mathbb{E}\left[\left|\sum_{i=1}^n u_{in}\overline{\lambda}_n(X_i)\right| \mid X_1, ..., X_n\right] &\leq c\,\mathbb{E}\left[\left(\sum_{i=1}^n u_{in}^2 \overline{\lambda}_n^2(X_i)\right)^{1/2} \mid X_1, ..., X_n\right] \\ \leq c\left[\sum_{i=1}^n \mathbb{E}\left(u_{in}^2 \mid X_i\right)\overline{\lambda}_n^2(X_i)\right]^{1/2} &\leq c\,\overline{\sigma}_n \left[\sum_{i=1}^n \overline{\lambda}_n^2(X_i)\right]^{1/2} \leq c\,\overline{\sigma}_n n^{1/2}\|A_h\|_2 \|\lambda_n\|_n, \end{aligned}$$

where c is a constant independent of n and of the sequence $\lambda_n(\cdot)$, $n \geq 1$. ■

A.2 Proofs

Lemma A.6 *Let the assumptions of Theorem 4.1 hold and fix $\zeta \in (0, 1/2)$ arbitrarily. Under H_0, for $\beta = 0$ or 1*

$$\sup_{h \in \mathcal{H}_n} h^{\zeta} \left|Q_n^{\beta}(\hat{\theta}) - Q_n^{\beta}(\theta_0)\right| = O_P(n^{-1}).$$

Proof. By definition

$$\hat{U}_i^{\beta}(\hat{\theta}) - \hat{U}_i^{\beta}(\theta_0) = (nW_{in})^{\beta}\left[f(\hat{\theta}, X_i) - f(\theta_0, X_i)\right],$$

where by convention $(nW_{in})^{\beta} = 1$ for $\beta = 0$ and $(nW_{in})^{\beta} = nW_{in}$ for $\beta = 1$. A similar convention applies for $\gamma^{\beta}(T_i)$. Let $a_{ij}(h) = [h^p n(n-1)]^{-1} K_h(X_i - X_j)$ and write

$$\begin{aligned} Q_n^{\beta}(\hat{\theta}) &= Q_n^{\beta}(\theta_0) \\ &\quad +2\sum_{i\neq j} \hat{U}_i^{\beta}(\theta_0)(nW_{jn})^{\beta}[f(\hat{\theta}, X_j) - f(\theta_0, X_j)]a_{ij}(h) \\ &\quad +\sum_{i\neq j}\left(n^2 W_{in}W_{jn}\right)^{\beta}[f(\hat{\theta}, X_i) - f(\theta_0, X_i)][f(\hat{\theta}, X_j) - f(\theta_0, X_j)]a_{ij}(h) \\ &= Q_n^{\beta}(\theta_0) + 2Q_{n1}^{\beta}(\hat{\theta}, \theta_0) + Q_{n2}^{\beta}(\hat{\theta}, \theta_0). \end{aligned}$$

By Assumption 5, there exists some constant c independent of h such that

$$\left|Q_{n2}^{\beta}(\hat{\theta}, \theta_0)\right| \leq c\|\hat{\theta} - \theta_0\|^2 \times \sum_{i\neq j}(nW_{in})^{\beta}(nW_{jn})^{\beta} a_{ij}(h).$$

Using the first part of equation (A.1) we obtain

$$\left|Q_{n2}^{\beta}(\hat{\theta}, \theta_0)\right| \leq O_P(1)\,\|\hat{\theta} - \theta_0\|^2 \sum_{i\neq j}\gamma^{\beta}(T_i)\,\gamma^{\beta}(T_j)\,a_{ij}(h).$$

As $\mathbb{E}[\gamma^2(T)]$ is finite (see Assumption 4-(iv)) and $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$ apply Lemma A.3 to deduce that

$$\sup_{h\in\mathcal{H}_n}\left|Q_{n2}^{\beta}\left(\hat{\theta}, \theta_0\right)\right| = O_P\left(n^{-1}\right).$$

To investigate Q_{n1}^{β}, let

$$\tilde{Q}_{n1}^{\beta}(\hat{\theta}, \theta_0) = \sum_{i\neq j} U_i^{\beta}(\theta_0)\gamma^{\beta}(T_j)\,[f(\hat{\theta}, X_j) - f(\theta_0, X_j)]a_{ij}(h).$$

By Taylor expansion in θ, Assumption 5-(i), Lemma A.3 and the condition $\mathbb{E}[U_i^{\beta}(\theta_0)^2 + \gamma^{\beta}(T)^2] < \infty$,

$$\begin{aligned} \tilde{Q}_{n1}^{\beta}\left(\hat{\theta}, \theta_0\right) &= \frac{(\hat{\theta} - \theta_0)'}{n(n-1)h^p}\sum_{i\neq j}\left\{U_i^{\beta}(\theta_0)\gamma^{\beta}(T_j)\right. \\ &\qquad \left.\times\nabla_{\theta} f(\theta_0, X_j)K_h(X_i - X_j)\right\} + \|\hat{\theta} - \theta_0\|^2 O_P(1) \\ &= h^{-p}(\hat{\theta} - \theta_0)'\tilde{S}_{n1}^{\beta}(h) + \|\hat{\theta} - \theta_0\|^2 O_P(1), \end{aligned}$$

with the $O_P(1)$ factor independent of h. For the zero mean $U-$process $\tilde{S}_{n1}^{\beta}(h)$ apply the Hoeffding decomposition and write it as a sum of degenerate $U-$processes of order 2 and

1, say $\tilde{S}^{\beta}_{n11}(h)$ and $\tilde{S}^{\beta}_{n12}(h)$, indexed by families defined by h that are Euclidean for square integrable envelopes (this property is ensured by the bounded variation of the kernel $\tilde{K}$, Lemma 22-(ii) of Nolan and Pollard 1987, and Lemma 5 of Sherman 1994). By Corollary 4 of Sherman (1994), the rate of the uniform convergence of $\tilde{S}^{\beta}_{n11}(h)$ is $O_P(n^{-1})$. Deduce

$$\sup_{h\in\mathcal{H}_n} h^{-p}\left|\tilde{S}^{\beta}_{n11}(h)\right| = O_P(n^{-1/2}).$$

On the other hand, $h^{-p}\tilde{S}^{\beta}_{n12}(h)$ writes like $n^{-1}\sum_{i=1}^{n} U_i^{\beta}(\theta_0)\phi_i$ with

$$\phi_i = \mathbb{E}[\gamma^{\beta}(T_j)\,\nabla_\theta f(\theta_0, X_j)h^{-p}K_h(X_i - X_j)\mid X_i].$$

Notice that $|\phi_i| \leq M$, for some constant M. Let $h_L \leq h_{min} \leq h_{L-1} < ... < h_1 < h_0 = h_{max}$ a geometric grid of bandwidths such that $h_l = h_{l-1}h_{max}^c$, $l = 1, ..., L$ with $c > 0$ to be chosen below. By definition $\mathcal{H}_n \subset \bigcup_{l=1}^{L} H_l$, where $H_l = [h_l, h_{l-1}]$. Fix arbitrarily $\alpha \in (0, 1)$ such that $1 - \zeta/p < \alpha$. For each $l = 1, ..., L$, by the definition of H_l and Sherman's (1994) Main Corollary

$$\begin{aligned}
\mathbb{E}\left[\sup_{h\in H_l}|n^{1/2}h^{\zeta-p}\tilde{S}^{\beta}_{n12}(h)|\right] &\leq h_l^{\zeta-p}\mathbb{E}\left[\sup_{h\in H_l}|n^{1/2}\tilde{S}^{\beta}_{n12}(h)|\right]\\
&\leq \Lambda_1 h_l^{\zeta-p}\left[\mathbb{E}\sup_{h\in H_l}\{h^{2p}\frac{1}{2n}\sum_{i=1}^{2n}U_i^{\beta}(\theta_0)^2\phi_i^2\}^{\alpha}\right]^{1/2}\\
&\leq \Lambda_2 h_l^{\zeta-(1-\alpha)p}\left(\frac{h_{l-1}}{h_l}\right)^{\alpha p}\left[\frac{1}{2n}\sum_{i=1}^{2n}U_i^{\beta}(\theta_0)^2\right]^{\alpha/2}\\
&= h_{max}^{a_l}\,O_P(1),
\end{aligned}$$

where Λ_1, Λ_2 are constants that depend on α and τ (and p) but not on n and l and $a_l = 1+\{l\,[\zeta - (1-\alpha)\,p] - p\alpha\}\,c$. The Euclidean property for a square integrable envelope required in Sherman's Main Corollary is ensured by the bounded variation of the kernel $\tilde{K}$, Lemma 22-(ii) of Nolan and Pollard (1987) and Lemma 5 of Sherman (1994). Take c such that $1+(\zeta - p)\,c > 0$. Looking at the sum of the geometric series with common ratio $h_{max}^{[\zeta-(1-\alpha)p]c}$ and starting term $h_{max}^{1+(\zeta-p)c}$, deduce that $\mathbb{E}\left[\sup_{h\in\mathcal{H}_n}|n^{1/2}h^{\zeta-p}\tilde{S}^{\beta}_{n12}(h)|\right] \to 0$. This, combined with Chebyshev's inequality, provide the order of $h^{\zeta-p}\tilde{S}^{\beta}_{n12}(h)$ uniformly in $h \in \mathcal{H}_n$. Collecting results and using $\|\hat{\theta} - \theta_0\|h_{min}^{-p} \to 0$, in probability deduce

$$\sup_{h\in\mathcal{H}_n} h^{\zeta}\left|\tilde{Q}^{\beta}_{n1}\left(\hat{\theta}, \theta_0\right)\right| = O_P(n^{-1}).$$

Next, rewrite

$$
\begin{aligned}
Q_{n1}^{\beta}(\hat{\theta},\theta_0) &= \tilde{Q}_{n1}^{\beta}\left(\hat{\theta},\theta_0\right) \\
&\quad +\sum_{i\neq j}[\hat{U}_i^{\beta}(\theta_0)-U_i^{\beta}(\theta_0)]\gamma^{\beta}(T_j)\,[f(\hat{\theta},X_j)-f(\theta_0,X_j)]a_{ij}(h) \\
&\quad +\sum_{i\neq j}U_i^{\beta}(\theta_0)\left[(nW_{jn})^{\beta}-\gamma^{\beta}(T_j)\right][f(\hat{\theta},X_j)-f(\theta_0,X_j)]a_{ij}(h) \\
&\quad +\sum_{i\neq j}[\hat{U}_i^{\beta}(\theta_0)-U_i^{\beta}(\theta_0)]\left[(nW_{jn})^{\beta}-\gamma^{\beta}(T_j)\right][f(\hat{\theta},X_j)-f(\theta_0,X_j)]a_{ij}(h) \\
&= \tilde{Q}_{n1}^{\beta}\left(\hat{\theta},\theta_0\right)+\tilde{Q}_{n11}^{\beta}+\tilde{Q}_{n12}^{\beta}+\tilde{Q}_{n13}^{\beta}.
\end{aligned}
$$

To show the negligibility of $\tilde{Q}_{n11}^{\beta}$ to $\tilde{Q}_{n13}^{\beta}$ we can no longer use the quick argument of Lemma A.3 because the random variables we have to manipulate are no longer square integrable. Indeed, by definition

$$
\hat{U}_i^{\beta}(\theta_0)-U_i^{\beta}(\theta_0)=[nW_{in}-\gamma(T_i)]\,[T_i-\beta f(\theta_0,X_i)]
$$

and the problem comes from the bound of $|nW_{in}-\gamma(T_i)|$ given by Lemma A.1 which contains $C(T_i)^{\alpha+\eta}$ (with $\eta>0$), a quantity that is not square integrable if we need to take $\alpha=1/2$. To show the negligibility of $\tilde{Q}_{n11}^{\beta}$ to $\tilde{Q}_{n13}^{\beta}$, apply Lemma A.1 with $\alpha=1/2$ and η equal to ρ from Assumption 7, and use Taylor expansion to bound $|f(\hat{\theta},X_j)-f(\theta_0,X_j)|$ by a constant times $\|\hat{\theta}-\theta_0\|$. Hence, $\tilde{Q}_{n11}^{\beta}$ to $\tilde{Q}_{n13}^{\beta}$ are bounded by

$$
O_P(n^{-1})\times\sum_{i\neq j}\frac{\gamma(T_i)\,|T_i-\beta f(\theta_0,X_i)|}{[C(T_i)]^{-(1/2+\rho)}}\gamma^{\beta}(T_j)\,a_{ij}(h)=O_P(n^{-1})\times B_{n1},
$$

$$
O_P(n^{-1})\times\sum_{i\neq j}\frac{\gamma(T_i)}{[C(T_i)]^{-(1/2+\rho)}}\gamma^{\beta}(T_j)\,a_{ij}(h)=O_P(n^{-1})\times B_{n2},
$$

and

$$
O_P(n^{-1})\times\sum_{i\neq j}\frac{\gamma(T_i)a_{ij}(h)}{[C(T_i)]^{-(1/2+\rho)}}\left(\frac{\hat{G}(T_j-)-G(T_j)}{1-G(T_j)}\gamma(T_j)\right)^{\beta}=O_P(n^{-1})\times B_{n3},
$$

respectively. To uniformly bound B_{n1}, by the boundedness of the regression

$$
0\leq B_{n1}\leq Ch^{-p}\left\{\Delta_{1n}+\Delta_{2n}\right\}\sum_{i=1}^{n}[C(T_i)]^{1/2+\rho}\gamma(T_i)\left\{|T_i|+1\right\}
$$

where C is some constant and

$$
\Delta_{1n}=\sup_{h\in\mathcal{H}_n,\,x}\left|\frac{1}{n}\sum_{j=1}^{n}\left\{K_h(x-X)\,\gamma^{\beta}(T)-\mathbb{E}\left[K_h(x-X)\,\gamma^{\beta}(T)\right]\right\}\right|
$$

and $\Delta_{2n} = \mathbb{E}\left[K_h(x-X)\gamma^\beta(T)\right] = \mathbb{E}\left[K_h(x-X)\right]$. Like in (A.5), $|h^{-p}\Delta_{2n}| \leq C_2$ for some C_2 independent of h. By the rate of uniform convergence of an empirical process indexed by an Euclidean family for a square integrable envelope and using $h^p\sqrt{n} \to \infty$, deduce that $|h^{-p}\Delta_{1n}| \leq C_1$ with C_1 independent of $h \in \mathcal{H}_n$. Finally,

$$\mathbb{E}\left[\gamma(T)\{|T|+1\}C(T)^{1/2+\rho}\right] = \mathbb{E}\left[\{|Y|+1\}C(Y)^{1/2+\rho}\right] = \mathbb{E}\left[q_\rho(X)\right] < \infty.$$

Deduce that $\sup_{h\in\mathcal{H}_n} B_{n1} = O_P(1)$. Similar arguments apply for B_{n2}. For B_{n3}, the only case that remains to study is $\beta = 1$. Use the first part of equation (A.1) to bring this case to that of B_{n1}. Collecting results, $\sup_{h\in\mathcal{H}_n} h^\gamma |Q_{n1}^\beta(\hat\theta,\theta_0)| = O_P(n^{-1})$. ■

Lemma A.7 *Let the assumptions of Theorem 4.1 hold true. If* $\tau < \tau_H$ *and*

$$Q_{n1}^\beta(\tau) = \frac{1}{n(n-1)h^p}\sum_{i\neq j}\left[\hat U_i^\beta - U_i^\beta\right]\mathbf{1}_{\{T_i\leq\tau\}}U_j^\beta K_h(X_i - X_j), \quad \beta = 0, 1,$$

then for any $\zeta \in (0, 1/2)$, $\sup_{h\in\mathcal{H}_n} h^\zeta \left|Q_{n1}^\beta(\tau)\right| = O_P(n^{-1})$.

Proof. If $w_i^\beta = \delta_i\left[T_i - \beta f(\theta_0, X_i)\right]\left[1 - G(T_i)\right]^{-2}$ we can write $Q_{n1}^\beta(\tau) = Q_{n11}^\beta(\tau) + Q_{n12}^\beta(\tau)$ with

$$Q_{n11}^\beta(\tau) = \frac{1}{n(n-1)h^p}\sum_{i\neq j}\left[\hat G(T_i-) - G(T_i)\right]\mathbf{1}_{\{T_i\leq\tau\}}w_i^\beta U_j^\beta K_h(X_i - X_j)$$

and

$$Q_{n12}^\beta(\tau) = \frac{1}{n(n-1)h^p}\sum_{i\neq j}\frac{\left[\hat G(T_i-) - G(T_i)\right]^2}{1 - \hat G(T_i-)}\mathbf{1}_{\{T_i\leq\tau\}}w_i^\beta U_j^\beta K_h(X_i - X_j).$$

By Theorem 2.1 of Gill (1983), $\sup_{1\leq i\leq n}\left[\hat G(T_i-) - G(T_i)\right]^2\mathbf{1}_{\{T_i\leq\tau\}} = O_P(n^{-1})$. (The fact that the left endpoint of the support of the variables T_i may be $-\infty$ is of no consequence since we only consider $\hat G$ and G at the sample points.) Meanwhile, $\sup_{1\leq i\leq n} G(T_i) \leq G(\tau) < 1$. These facts, Lemma A.3 and Assumption 4-(iv) imply

$$\begin{aligned}\sup_{h\in\mathcal{H}_n}\left|Q_{n12}^\beta(\tau)\right| &\leq O_P(n^{-1})\left(\frac{1}{n}\sum_{i=1}^n\left[w_i^\beta\right]^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^n\left[U_i^\beta\right]^2\right)^{1/2} \\ &= O_P(n^{-1}).\end{aligned}$$

To handle $Q^{\beta}_{n11}(\tau)$, we use the i.i.d. representation

$$\hat{G}(t-) - G(t) = \frac{1}{n}\sum_{k=1}^{n}\psi(T_k, t) + R_n(t)$$

with $\sup_{t\leq\tau}|R_n(t)| = O_P(n^{-1})$ and for each $t \leq \tau$,

$$\mathbb{E}[\psi(T_k, t)] = 0 \tag{A.8}$$

and $|\psi(T_k, t)| \leq M_1$ for some constant M_1 independent of t (but depending on τ). See Stute (1995) or Sánchez-Sellero *et al.* (2005) for the definition of the function $\psi(\cdot,\cdot)$. The representation (A.8) can be derived along the lines of the proof Theorem 1.1 of Stute (1995) in the case where his condition (2.3) holds. This representation is also a consequence of Theorem 1 of Sánchez-Sellero *et al.* (2005) applied for the class of indicator functions of the intervals $(-\infty, t)$ with $t \leq \tau$. Now, we can write

$$\begin{aligned} Q^{\beta}_{n11}(\tau) &= \frac{1}{n^2(n-1)h^p}\sum_{i\neq j\neq k}\psi(T_k, T_i)\,\mathbf{1}_{\{T_i\leq\tau\}}w_i^{\beta}U_j^{\beta}K_h(X_i - X_j) \\ &\quad +\frac{1}{n}\frac{1}{n(n-1)h^p}\sum_{i\neq j}\psi(T_i, T_i)\,\mathbf{1}_{\{T_i\leq\tau\}}w_i^{\beta}U_j^{\beta}K_h(X_i - X_j) \\ &\quad +\frac{1}{n}\frac{1}{n(n-1)h^p}\sum_{i\neq j}\mathbf{1}_{\{T_i\leq\tau\}}w_i^{\beta}\psi(T_j, T_j)\,U_j^{\beta}K_h(X_i - X_j) + \text{ remainder} \\ &= (n-2)\,n^{-1}Q^{\beta}_{n111}(\tau) + n^{-1}Q^{\beta}_{n112}(\tau) + n^{-1}Q^{\beta}_{n113}(\tau) + O_P(n^{-1}). \end{aligned}$$

By Lemma A.3, the fact that $\psi(\cdot,\cdot)$ is bounded and w_i^{β}, U_j^{β} are square integrable

$$\sup_{h\in\mathcal{H}_n}\left\{\left|Q^{\beta}_{n112}(\tau)\right| + \left|Q^{\beta}_{n113}(\tau)\right|\right\} = O_P(1).$$

For the remaining term $Q^{\beta}_{n111}(\tau)$ which is a $U-$process of order 3, apply the Hoeffding decomposition and write it as the sum of two degenerate $U-$processes

$$Q^{\beta}_{n1111}(\tau) = Q^{\beta}_{n111}(\tau) - Q^{\beta}_{n1112}(\tau)$$

and $Q^{\beta}_{n1112}(\tau) = n^{-1}(n-1)^{-1}\sum_{j\neq k}\phi_{jk}U_j^{\beta}$, where

$$\phi_{jk} = \mathbb{E}\left[\psi(T_k, T_i)\,\mathbf{1}_{\{T_i\leq\tau\}}w_i^{\beta}h^{-p}K_h(X_i - X_j) \mid X_j, T_k\right].$$

Notice that $|\phi_{jk}| \leq M_2$ for some constant M_2. The fact that $\mathbb{E}\left[U_j^{\beta} \mid X_j\right] = 0$ a.s. and the property (A.8) make that the other terms in the Hoeffding decomposition of $Q^{\beta}_{n111}(\tau)$ are null. Corollary 4 of Sherman (1994) implies $\sup_{h\in\mathcal{H}_n} h^p \left|Q^\beta_{n1111}(\tau)\right| = O_P(n^{-3/2})$. Thus

$$\sup_{h\in\mathcal{H}_n} \left|Q^\beta_{n1111}(\tau)\right| = o_P(n^{-1}).$$

Next, fix $\zeta \in (0,1/2)$ and $\alpha \in (0,1)$ such that $1-\zeta/p < \alpha$, and consider the intervals H_l like in the proof of our Lemma A.6. For each H_l, by Sherman's (1994) Main Corollary

$$\begin{aligned}
\mathbb{E}\left[\sup_{h\in H_l} |nh^\zeta Q^\beta_{n1112}(\tau)|\right] &\leq h_l^{\zeta-p}\mathbb{E}\left[\sup_{h\in H_l} |nh^p Q^\beta_{n1112}(\tau)|\right] \\
&\leq \Lambda_1 h_l^{\zeta-p}\left[\mathbb{E}\sup_{h\in H_l}\left\{\frac{h^{2p}}{4n^2}\sum_{1\leq j,k\leq 2n}\phi^2_{jk}\left[U_j^\beta\right]^2\right\}^\alpha\right]^{1/2} \\
&\leq \Lambda_2 h_l^{\zeta-(1-\alpha)p}\left(\frac{h_{l-1}}{h_l}\right)^{\alpha p}\left[\frac{1}{2n}\sum_{j=1}^{2n}\left[U_j^\beta\right]^2\right]^{\alpha/2} \\
&= h_{max}^{a_l} O_P(1),
\end{aligned}$$

where Λ_1, Λ_2 are constants and a_l is like in the proof of Lemma A.6. Sum over all l to obtain that $nh^\zeta Q^\beta_{n1112}(\tau) = o_P(1)$ uniformly in $h \in \mathcal{H}_n$. This ends the proof. ■

Lemma A.8 *Let the assumptions of Theorem 4.1 hold true and let*

$$Q^\beta_{n2} = \frac{1}{n(n-1)h^p}\sum_{i\neq j}\left[\hat{U}_i^\beta - U_i^\beta\right]\left[\hat{U}_j^\beta - U_j^\beta\right]K_h(X_i - X_j), \quad \beta = 0,1.$$

Then $\sup_{h\in\mathcal{H}_n}\left|Q^\beta_{n2}\right| = O_P(n^{-1})$.

Proof. We apply Lemma A.1 with $\alpha = 1/2$ to bound $|\hat{U}_i^\beta - U_i^\beta|$ and we obtain

$$\left|Q^\beta_{n2}\right| \leq \frac{O_P(n^{-1})}{n(n-1)}\sum_{i\neq j}\frac{\{|T_i|+1\}\gamma(T_i)}{[C(T_i)]^{-(1/2+\rho)}}h^{-p}K_h(X_i - X_j)\,\frac{\{|T_j|+1\}\gamma(T_j)}{[C(T_j)]^{-(1/2+\rho)}}.$$

By (2.3) and taking conditional expectations, the expectation of the term in the sum is

$$\begin{aligned}
&\mathbb{E}\left[\frac{\{|Y_1|+1\}}{[C(Y_1)]^{-(1/2+\rho)}}h^{-p}K_h(X_1 - X_2)\,\frac{\{|Y_2|+1\}}{[C(Y_2)]^{-(1/2+\rho)}}\right] \\
&= \mathbb{E}\left[q_\rho(X_1)q_\rho(X_2)h^{-p}K_h(X_1 - X_2)\right]
\end{aligned}$$

and thus it is bounded by Lemma A.4. Deduce that $Q^\beta_{n2} = O_P(n^{-1})$. To derive this rate uniformly in $h \in \mathcal{H}_n$, we can use Lemma A.2. To apply Lemma A.2 it remains to prove that deduce

$$\frac{K(0)}{n(n-1)h^p}\sum_{j=1}^n\left[\hat{U}_j^\beta - U_j^\beta\right]^2 = O_P(n^{-1})$$

for $h = h_{min}$ and $h = h_{max}$. For this purpose apply Lemma A.1 with $\alpha = 1/4$ to bound $|\hat{U}_j^\beta - U_j^\beta|$ and recall that $\sqrt{n}h_{min}^p \to \infty$ and $\mathbb{E}[C(Y)^{1/2+\rho}] < \infty$. ■

Lemma A.9 *Let Q_{n1}^β and $Q_{n1}^\beta(\tau)$ be defined as in (4.19) and (4.20), respectively. Under the assumptions of Theorem 4.1, for $\beta = 0$ or 1*

$$\sup_{h\in\mathcal{H}_n} h^{p/2}\left|Q_{n1}^\beta(\tau) - Q_{n1}^\beta\right| = C_\tau \times O_P(n^{-1}),$$

with the $O_P(n^{-1})$ factor independent of τ and C_τ tending to zero when $\tau \uparrow \tau_H$.

Proof. Write

$$\begin{aligned}
&\frac{n-1}{n}h^{p/2}\left[Q_{n1}^\beta(\tau) - Q_{n1}^\beta\right] \\
= &\ \frac{1}{n^2h^{p/2}}\sum_{1\le i,j\le n} U_i^\beta K_h(X_i - X_j)\left(U_j^\beta - \hat{U}_j^\beta\right)\mathbf{1}_{\{T_j>\tau\}} \\
&-\frac{K(0)}{n^2h^{p/2}}\sum_{j=1}^n U_j^\beta\left(U_j^\beta - \hat{U}_j^\beta\right)\ \mathbf{1}_{\{T_j>\tau\}} \\
= &\ S_1 - S_2
\end{aligned}$$

By the Fourier Transform and Cauchy-Schwarz inequality

$$\begin{aligned}
|S_1| \le &\left[\int \hat{K}(hu)\left|\frac{1}{n}\sum_{j=1}^n\left(U_j^\beta - \hat{U}_j^\beta\right)\exp(2i\pi u'X_j)\mathbf{1}_{\{T_j>\tau\}}\right|^2 du\right]^{1/2} \\
&\times\left[h^p\int \hat{K}(hu)\left|\frac{1}{n}\sum_{j=1}^n U_j^\beta\exp(-2i\pi u'X_j)\right|^2 du\right]^{1/2} = [S_{11}]^{1/2}[S_{12}]^{1/2}.
\end{aligned}$$

By the monotonicity of $\hat{\hat{K}}$, to obtain the uniform rate for S_{11} it suffices to take $h = h_{min}$. Now, by the inverse Fourier Transform S_{11} can be rewritten

$$\begin{aligned}
S_{11} = &\ \frac{1}{n^2h_{min}^p}\sum_{i\ne j}(U_i^\beta - \hat{U}_i^\beta)\mathbf{1}_{\{T_i>\tau\}}K_{h_{min}}(X_i - X_j)(U_j^\beta - \hat{U}_j^\beta)\mathbf{1}_{\{T_j>\tau\}} \\
&+\frac{K(0)}{n^2h_{min}^p}\sum_{j=1}^n\left(U_j^\beta - \hat{U}_j^\beta\right)^2\mathbf{1}_{\{T_j>\tau\}} = S_{111} + S_{112}.
\end{aligned}$$

To handle S_{111}, apply Lemma A.1 with $\alpha = 1/2$. Then, $|S_{111}|$ is bounded by

$$\frac{O_P(n^{-1})}{n^2h_{min}^p}\sum_{i\ne j}\frac{\{|T_i|+1\}\mathbf{1}_{\{T_i>\tau\}}\gamma(T_i)}{[C(T_i)]^{-(1/2+\rho)}}K_{h_{min}}(X_i - X_j)\frac{\{|T_j|+1\}\mathbf{1}_{\{T_j>\tau\}}\gamma(T_j)}{[C(T_j)]^{-(1/2+\rho)}},$$

where the $O_P(n^{-1})$ rate does not depend on τ. By (2.3) and taking conditional expectations, the expectation of a term in the last sum is

$$\begin{aligned}
&\mathbb{E}\left[\frac{\{|Y_1|+1\}\mathbf{1}_{\{Y_1>\tau\}}}{[C(Y_1)]^{-(1/2+\rho)}}K_{h_{min}}(X_1-X_2)\frac{\{|Y_2|+1\}\mathbf{1}_{\{Y_2>\tau\}}}{[C(Y_2)]^{-(1/2+\rho)}}\right] \\
&= \mathbb{E}\left[q_{\rho,\tau}(X_1)q_{\rho,\tau}(X_2)K_{h_{min}}(X_1-X_2)\right]
\end{aligned}$$

with $q_{\rho,\tau}$ defined in Lemma A.4. Apply Lemma A.4 and deduce that $|S_{111}|$ is bounded by $C_\tau \times O_P(n^{-1})$ for some constant C_τ independent of n but tending to zero as $\tau \uparrow \tau_H$. To bound S_{112}, apply Lemma A.1 with $\alpha = 1/6$ to obtain

$$\begin{aligned}
|S_{112}| &\leq \frac{1}{n^2 h_{min}^p}\sum_{j=1}^n \left(U_j^\beta - \hat{U}_j^\beta\right)^2 \mathbf{1}_{\{T_j>\tau\}}K(0) \qquad \text{(A.9)}\\
&\leq n^{-1/3}h_{min}^{-p}O_P(n^{-1})\frac{1}{n}\sum_{j=1}^n \frac{\gamma(T_j)^2\{|T_j|+1\}^2}{[C(T_j)]^{-(1/3+2\rho/3)}}.
\end{aligned}$$

By Hölder inequality, the expectation of a term in the last sum is bounded by

$$\mathbb{E}^{1/3}\left[\delta\{|T|+1\}^4[1-G(T)]^{-3}\right]\mathbb{E}^{2/3}\left[\{|T|+1\}C(T)^{1/2+\rho}\right],$$

which is finite under Assumptions 4-(iv) and 7. Finally, recall that $nh_{min}^{3p}\to\infty$. Collecting results, $\sup_{h\in\mathcal{H}_n} S_{11} = C_\tau \times O_P(n^{-1})$. To handle S_{12}, by the inverse Fourier Transform and Corollary 4 of Sherman (1994) we obtain

$$S_{12} = \frac{1}{n^2}\sum_{i\neq j} U_i^\beta U_j^\beta K_h(X_i - X_j) + \frac{K(0)}{n^2}\sum_{j=1}^n \left[U_j^\beta\right]^2 = O_P\left(n^{-1}\right),$$

and the rate $O_P(n^{-1})$ is uniform in $h\in\mathcal{H}_n$. For S_2, take absolute values, apply Lemma A.1 with $\alpha = 1/4$ and use $n^{1/4}h_{min}^{p/2}\to\infty$ to deduce $\sup_{h\in\mathcal{H}_n}|h^{p/2}S_2| = o_P(n^{-1})$. ■

Lemma A.10 *Let the assumptions of Theorem 4.1 hold. Under H_0, for $\beta = 0$ or 1*

$$\sup_{h\in\mathcal{H}_n}\left|\frac{\tilde{V}_n^\beta(\theta_0)}{\hat{V}_n^\beta} - 1\right| = o_P(1).$$

Proof. Recall that

$$\left[\hat{V}_n^\beta\right]^2 = \hat{V}_n^\beta(\theta)^2 = \frac{2}{n(n-1)h^p}\sum_{i\neq j}\hat{U}_i^\beta(\theta)^2\hat{U}_j^\beta(\theta)^2 K_h^2(X_i - X_j).$$

The result is implied by the following statements:

$$\sup_{\theta\in\Theta,h\in\mathcal{H}_n}\left|\tilde{V}_n^{\beta}(\theta)^2-\hat{V}_n^{\beta}(\theta)^2\right|=o_P(1), \tag{A.10}$$

$$\sup_{h\in\mathcal{H}_n}\left|\tilde{V}_n^{\beta}(\theta)^2-\tilde{V}_n^{\beta}(\theta_0)^2\right|\leq\|\theta-\theta_0\|\times O_P(1) \tag{A.11}$$

with $O_P(1)$ independent of $\theta\in\Theta$, and

$$\tilde{V}_n^{\beta}(\theta_0)^2\to 2\int K^2(u)du\mathbb{E}\left\{\mathbb{E}^2\left[U^{\beta}(\theta_0)^2\mid X\right]g(X)\right\} \tag{A.12}$$

in probability, uniformly in $h\in\mathcal{H}_n$. The limit is finite and strictly positive since Assumption 4(iii) imply

$$\text{for each } x\in\mathcal{X},\qquad 0<c_1\leq\mathbb{E}\left[U^{\beta}(\theta_0)^2\mid X=x\right]\leq c_2<\infty,$$

for some constants c_1,c_2. The convergence (A.12) is quite standard for any sequence of h such that $h\to 0$ and $nh^{2p}\to\infty$. The expectation of $\tilde{V}_n^{\beta}(\theta_0)$ tends to the limit in (A.12) (see also the proof of Lemma A.4 above) while the variance of $\tilde{V}_n^{\beta}(\theta_0)$ vanishes. To obtain the convergence uniformly in $h\in\mathcal{H}_n$, use $\mathbb{E}[U_i^{\beta}(\theta_0)^4]<\infty$ to deduce

$$\frac{K^2(0)}{n^2h^p}\sum_{i=1}^{n}U_i^{\beta}(\theta_0)^4=o_P(1)$$

for $h=h_{min}$ and $h=h_{max}$ and apply Lemma A.2. To check (A.11), use a Taylor expansion in θ, Lemma A.3 and the fact that $\gamma^{\beta}(T_i)$ and $U_i^{\beta}(\theta_0)$ have finite moments of order 4. Finally, for proving (A.10) notice that by Assumption 5-(i) and Lemma A.1, for each $0\leq\alpha\leq 1/2$ and $\eta>0$

$$\begin{aligned}|\hat{U}_i^{\beta}(\theta)-U_i^{\beta}(\theta)| &\leq c\,|nW_{in}-\gamma(T_i)|\,(|T_i|+1) \qquad \text{(for some constant } c\text{)}\\ &= O_P\left(n^{-\alpha}\right)\gamma(T_i)\{C^{(n)}(T_i)\}^{\alpha+\eta}(|T_i|+1)\end{aligned}$$

and since $|U_i^{\beta}(\theta)|\leq c\gamma(T_i)(|T_i|+1)$, we also have

$$\left|\hat{U}_i^{\beta}(\theta)^2-U_i^{\beta}(\theta)^2\right|=O_P\left(n^{-\alpha}\right)\gamma^2(T_i)\{C^{(n)}(T_i)\}^{\alpha+\eta}(|T_i|+1)^2.$$

Taking α and η sufficiently small, by elementary algebra, Lemma A.3, Cauchy-Schwarz inequality and Assumptions 4-(iv) and 7 we obtain

$$\sup_{\theta\in\Theta,h\in\mathcal{H}_n}\left|\tilde{V}_n^{\beta}(\theta)^2-\hat{V}_n^{\beta}(\theta)^2\right|=o_P(1).$$

■

Like for Lemma A.1, the following result is designed to be applied under H_0 and under the alternatives and therefore the law of $(T_1, \delta_1), ..., (T_n, \delta_n)$ may depend on n.

Lemma A.11 *Suppose the conditions of Lemma A.1. Let $X_1, X_2, ...$ be an independent sample from the random vector X with support $\mathcal{X} \subset \mathbb{R}^p$ and bounded density g. Moreover, g is bounded away from zero on $\mathcal{X}$. There exist positive constants a and M (independent of n) such that for each n*

$$\mathbb{E}\left[\frac{T_{1n}^4 \gamma^4(T_{1n})}{C^{(n)}(T_{1n})^{-2a}}\right] + \sup_{x\in\mathcal{X}} \mathbb{E}\left[\frac{T_{1n}^2 \gamma^2(T_{1n})}{C^{(n)}(T_{1n})^{-a}} \mid X_1 = x\right] \leq M < \infty. \qquad \text{(A.13)}$$

Consider a kernel $L(x_1, ..., x_p) = \tilde{L}(x_1)...\tilde{L}(x_p)$ where $\tilde{L}$ is a symmetric density of bounded variation on the real line. Consider also a sequence of bandwidths $b_n \to 0$ such that $nb_n^{2p} \to \infty$. Let $Y_{in}^* = \delta_{in} T_{in}[1 - G(T_{in})]^{-1}$, $i = 1, ..., n$, and define

$$\sigma_n^{*2}(x) = \frac{\sum_{i=1}^n Y_{in}^{*2} L((X_i - x)/b_n)}{\sum_{i=1}^n L((X_i - x)/b_n)} - \left(\frac{\sum_{i=1}^n Y_{in}^* L((X_i - x)/b_n)}{\sum_{i=1}^n L((X_i - x)/b_n)}\right)^2, x \in \mathcal{X},$$

an estimate of $Var(Y_{1n}^* \mid X_1 = x)$. Define $\hat{\sigma}_n^{*2}(x)$ similarly but with $\hat{Y}_{in}^* = \delta_{in} T_{in}[1 - \hat{G}(T_{in})]^{-1}$ instead of Y_{in}^*. Then, $\sup_{x\in\mathcal{X}} |\hat{\sigma}_n^{*2}(x) - \sigma_n^{*2}(x)| \to 0$ in probability.

Proof. For simplicity we focus on the case of the null hypothesis, the arguments under the alternatives being similar. By the rate of convergence of an empirical process indexed by an Euclidean family for a constant envelope and the condition $n^{1/2} b_n^p \to \infty$,

$$\sup_{x\in\mathcal{X}} \left| \frac{1}{nb_n^p} \sum_{i=1}^n L((X_i - x)/b_n) - \mathbb{E}\left[b_n^{-p} L((X_i - x)/b_n)\right] \right| \to 0,$$

in probability. Moreover, by a change of variable and the properties of the density $g(\cdot)$, for all n, $0 < c_1 \leq \mathbb{E}\left[b_n^{-p} L((X_i - x)/b_n)\right] \leq c_2 < \infty$ for some constants c_1, c_2. Thus, to prove the result it remains to show that

$$\text{for } k = 1, 2 \ , \quad \sup_{x\in\mathcal{X}} \left| \frac{1}{nb_n^p} \sum_{i=1}^n \left[\hat{Y}_i^{*k} - Y_i^{*k}\right] L((X_i - x)/b_n) \right| \to 0,$$

in probability. We only consider the case $k = 2$ as the other case can be treated similarly. By Lemma A.1, for any $\alpha \in [0, 1/2]$ and $\eta > 0$

$$\left|\hat{Y}_i^* - Y_i^*\right| \leq O_P(n^{-\alpha}) \times |T_i| \gamma(T_i) C(T_i)^{\alpha+\eta}, \qquad 1 \leq i \leq n,$$

with the $O_P(n^{-\alpha})$ factor independent of i. In view of this bound and using the identity $b^2 - c^2 = (b-c)^2 + 2c(b-c)$, it is easy to see that the most difficult part is to show

$$\sup_{x\in\mathcal{X}} \left| \frac{1}{n} \sum_{i=1}^{n} T_i^2 \gamma^2 (T_i) \, C(T_i)^{2\alpha+2\eta} L((X_i - x)/b_n) \right| = O_P(b_n^p). \tag{A.14}$$

For $2\alpha + 2\eta \leq a$, taking conditional expectations we deduce that

$$\sup_{x\in\mathcal{X}} \mathbb{E}\left[T_n^2 \gamma^2 (T_n) \, C(T_n)^{2\alpha+2\eta} \, L((X-x)/b_n)\right] \leq c_3 b_n^p$$

for some finite c_3 independent of n. Now, center the sum in (A.14) to obtain an empirical process indexed by an Euclidean family of functions for a square integrable envelope. (The Euclidean property is given by the bounded variation property of L and the first part of condition (A.13).) Hence, after centering inside the absolute value in (A.14) we obtain the (uniform) rate $O_P(n^{-1/2})$. Finally, use the condition $n^{1/2} b_n^p \to \infty$ to derive the result under H_0. The uniform rate $O_P(n^{-1/2})$ obtained after centering the sum in (A.14) can also be derived when the law of the independent responses $Y_1, ..., Y_n$ is the same but depends on n. For this purpose use, for instance, the Main Corollary of Sherman (1994) with $k = 1$ and for each $n \geq 1$. The details are omitted. ■

Proof of Lemma 4.3. With the same convention for the superscripts (SD and WLS replaced with 0 and 1, respectively) and omitting θ, let

$$\begin{aligned} U_{in}^0 &= \frac{\delta_{in} T_{in}}{1 - G(T_{in})} - f(\theta, X_i), \qquad \hat{U}_{in}^0 = \frac{\delta_{in} T_{in}}{1 - \hat{G}(T_{in}-)} - f(\theta, X_i), \\ U_{in}^1 &= \frac{\delta_{in}\left[T_{in} - f(\theta, X_i)\right]}{1 - G(T_{in})}, \qquad \hat{U}_{in}^1 = \frac{\delta_{in}\left[T_{in} - f(\theta, X_i)\right]}{1 - \hat{G}(T_{in}-)}, \end{aligned}$$

$i = 1, ..., n$. Applying Lemma A.1 with $\alpha = 1/2$ and using the boundedness of $f(\cdot, \cdot)$, for $\beta = 0$ or 1

$$|\hat{U}_{in}^\beta - U_{in}^\beta| = |R_{in}^\beta| \leq O_P(n^{-1/2}) \frac{\delta_{in}}{1 - G\left(T_{in}\right)} \{|T_{in}| + 1\} [C^{(n)}\left(T_{in}\right)]^{1/2+\eta}.$$

Now, simplify the notation $K_h(X_i - X_j)$ to K_{ij} and write

$$\frac{1}{n^2h^p}\sum_{i\neq j}\left\{\hat{U}_{in}^{\beta}\hat{U}_{jn}^{\beta}-U_{in}^{\beta}U_{jn}^{\beta}\right\}K_{ij}=\frac{2}{n^2h^p}\sum_{i\neq j}R_{in}^{\beta}U_{jn}^{\beta}K_{ij}+\frac{1}{n^2h^p}\sum_{i\neq j}R_{in}^{\beta}R_{jn}^{\beta}K_{ij}$$
$$=2\int\hat{K}\left(hu\right)\left(\frac{1}{n}\sum_{j=1}^{n}U_{jn}^{\beta}\exp\left(2i\pi u'X_j\right)\right)\left(\frac{1}{n}\sum_{j=1}^{n}R_{jn}^{\beta}\exp\left(-2i\pi u'X_j\right)\right)du$$
$$-\frac{2K\left(0\right)}{n^2h^p}\sum_{j=1}^{n}R_{jn}^{\beta}U_{jn}^{\beta}$$
$$+\int\hat{K}\left(hu\right)\left|\frac{1}{n}\sum_{j=1}^{n}R_{jn}^{\beta}\exp\left(2i\pi u'X_j\right)\right|^2du-\frac{K\left(0\right)}{n^2h^p}\sum_{j=1}^{n}[R_{jn}^{\beta}]^2.$$

The first integral can be bounded using Cauchy-Schwarz inequality and the bound of the second integral. To show that the second integral is of order $O_P(n^{-1})$, apply Lemma A.1 with $\alpha=1/2$ and check that the expectation

$$\mathbb{E}\left[\frac{1}{h^p}\;K_{12}\;\frac{\gamma(T_{1n})\{|T_{1n}|+1\}}{\left[C^{(n)}\left(T_{1n}\right)\right]^{-(1/2+\eta)}}\;\frac{\gamma(T_{2n})\{|T_{2n}|+1\}}{\left[C^{(n)}\left(T_{2n}\right)\right]^{-(1/2+\eta)}}\right] \tag{A.15}$$

is bounded, where $\gamma(T_{1n})=\delta_{1n}[1-G\left(T_{1n}\right)]^{-1}$. From Assumption 8-(ii), deduce that this expectation equals

$$\mathbb{E}\left[\frac{1}{h^p}K_{12}\mathbb{E}\left[\frac{|Y_{1n}|+1}{\left[C^{(n)}\left(Y_{1n}\right)\right]^{-(1/2+\eta)}}\mid X_1\right]\mathbb{E}\left[\frac{|Y_{2n}|+1}{\left[C^{(n)}\left(Y_{2n}\right)\right]^{-(1/2+\eta)}}\mid X_2\right]\right]$$
$$=\mathbb{E}\left[h^{-p}K_{12}q_{\rho}^{(n)}(X_1)q_{\rho}^{(n)}(X_2)\right]$$

and the last expectation is bounded by Assumption 9 and Lemma A.4. The rest of the proof continues with arguments that we already used in the previous proofs. ■

Lemma A.12 *Let Assumptions 4-(i) to (iii), 5, 6, 8, 9 hold true and let $\hat{\theta}$ denote either θ^{SD} or θ^{WLS}.*

i) If for all $n\geq 1$, $\mathbb{E}\left[\lambda_n(X)\nabla_{\theta}f\left(\theta_0,X\right)\right]=0$ and $0\leq|\lambda_n(\cdot)|\leq M_{\lambda}<\infty$ for some constant M_{λ} and if $\mathbb{E}\left|\lambda_n(X)\right|\to 0$, under the alternatives H_{1n} defined in (4.28), $\hat{\theta}-\theta_0=O_P(n^{-1/2})$.

ii) If Assumption 10 hold, under the alternative H_1, $\hat{\theta}-\bar{\theta}=O_P(n^{-1/2})$.

Proof. *i)* Consider the SD approach and decompose

$$\begin{aligned} M_n^{SD}(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\left[Y_{in}^{*} - f(\theta, X_i)\right]^2 - \frac{2}{n}\sum_{i=1}^{n}\left[\hat{Y}_{in}^{*} - Y_{in}^{*}\right] f(\theta, X_i) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\left[\hat{Y}_{in}^{*\,2} - Y_{in}^{*\,2}\right] \\ &= M_{1n}^{SD}(\theta) + M_{2n}^{SD}(\theta) + R_n^{SD} \end{aligned}$$

where $\hat{Y}_{in}^{*} = \delta_{in} T_{in}\left[1 - \hat{G}(T_{in}-)\right]^{-1}$ and Y_{in}^{*} is defined similarly but $\hat{G}$ is replaced with G. First, we prove the consistency of $\hat{\theta}^{SD}$. Opening the brackets and using elementary arguments and condition $\mathbb{E}\left|\lambda_n(X)\right| \to 0$, deduce that

$$\sup_{\theta\in\Theta}\left| M_{1n}^{SD}(\theta) - \frac{1}{n}\sum_{i=1}^{n}\left[\gamma(T_{in})\left\{f(\theta_0, X_i) + \varepsilon_i\right\} - f(\theta, X_i)\right]^2\right| = o_P(1).$$

Next, by the Main Corollary of Sherman (1994) applied for each n

$$\begin{aligned} \sup_{\theta\in\Theta}\Bigg| \frac{1}{n}\sum_{i=1}^{n}&\left[\gamma(T_{in})\left\{f(\theta_0, X_i) + \varepsilon_i\right\} - f(\theta, X_i)\right]^2 \\ &- I - \mathbb{E}\left[\left\{\varepsilon + f(\theta_0, X) - f(\theta, X)\right\}^2\right]\Bigg| = o_P(1), \end{aligned}$$

where $I = \mathbb{E}[\{\gamma^2(T_{in}) - 1\}\{f(\theta_0, X) + \varepsilon\}^2]$. The required Euclidean property is ensured by the boundedness of $\nabla_\theta f(\cdot,\cdot)$ (see Assumption 5-(i)) and Lemma 2.14 of Pakes and Pollard (1989) and the finite second order moment of ε. On the other hand, by Lemma A.1 applied with some $\alpha \in (0, 1/2]$, $\sup_{\theta\in\Theta}\left|M_{2n}^{SD}(\theta)\right| = o_P(1)$. Deduce

$$\sup_{\theta\in\Theta}\left| M_{1n}^{SD}(\theta) + M_{2n}^{SD}(\theta) - I - \mathbb{E}\left[\left\{\varepsilon + f(\theta_0, X) - f(\theta, X)\right\}^2\right]\right| = o_P(1).$$

Since θ_0 is the unique maximizer of the expectation in the last display, deduce that $\hat{\theta}^{SD} - \theta_0 = o_P(1)$. To obtain the rate of convergence, let $\mu_{in} = Y_{in}^{*} - f(\theta_0, X_i)$ and notice that $\mathbb{E}\left[\mu_{in} \mid X_i\right] = \lambda_n(X_i)$ and $\mathbb{E}\left[\mu_{in}\nabla_\theta f(\theta_0, X_i)\right] = 0$. By the definition of $\hat{\theta}^{SD}$ and the second order Taylor expansion

$$\begin{aligned}
& M_{1n}^{SD}(\theta_0) - M_{1n}^{SD}(\hat{\theta}^{SD}) \\
= \; & \frac{1}{n}\sum_{i=1}^{n}\mu_{in}^2 - \frac{1}{n}\sum_{i=1}^{n}\left[\mu_{in} - \left\{f\left(\hat{\theta}^{SD}, X_i\right) - f(\theta_0, X_i)\right\}\right]^2 \\
= \; & -\frac{1}{n}\sum_{i=1}^{n}\left\{f\left(\hat{\theta}^{SD}, X_i\right) - f(\theta_0, X_i)\right\}^2 + \frac{2}{n}\sum_{i=1}^{n}\mu_{in}\left\{f\left(\hat{\theta}^{SD}, X_i\right) - f(\theta_0, X_i)\right\} \\
\leq \; & -\|\hat{\theta}^{SD} - \theta_0\|^2\left\{\frac{1}{n}\sum_{i=1}^{n}\Phi^2(X_i)\right\} + (\hat{\theta}^{SD} - \theta_0)'\left\{\frac{2}{n}\sum_{i=1}^{n}\mu_{in}\nabla_\theta f(\theta_0, X_i)\right\} \\
& + (\hat{\theta}^{SD} - \theta_0)'\left\{o_P(1) + \frac{1}{n}\sum_{i=1}^{n}\left[\mu_{in} - \lambda_n(X_i)\right]\nabla_\theta^2 f(\theta_0, X_i)\right. \\
& \qquad\qquad \left. + \frac{1}{n}\sum_{i=1}^{n}\lambda_n(X_i)\nabla_\theta^2 f(\theta_0, X_i)\right\}(\hat{\theta}^{SD} - \theta_0) \\
= \; & -A_n\|\hat{\theta}^{SD} - \theta_0\|^2 + (\hat{\theta}^{SD} - \theta_0)'B_{1n} + (\hat{\theta}^{SD} - \theta_0)'C_{1n}(\hat{\theta}^{SD} - \theta_0).
\end{aligned}$$

Notice that $A_n - A = o_P(1)$, where $A = \mathbb{E}\left[\Phi^2(X)\right] > 0$. On the other hand, by a classical central limit theorem for triangular arrays we obtain $\|B_{1n}\| = O_P(n^{-1/2})$. Similar arguments and condition $\mathbb{E}\left|\lambda_n(X)\right| \to 0$ yield $\|C_{1n}\|_2 = o_P(1)$. To handle $M_{2n}^{SD}(\theta)$, let us write

$$\begin{aligned}
M_{2n}^{SD}(\theta_0) - M_{2n}^{SD}(\hat{\theta}^{SD}) \; &= \; (\hat{\theta}^{SD} - \theta_0)'\frac{2}{n}\sum_{i=1}^{n}\left[\hat{Y}_{in}^* - Y_{in}^*\right]\nabla_\theta f(\theta_0, X_i) \\
& \quad + (\hat{\theta}^{SD} - \theta_0)'\left[\frac{1}{n}\sum_{i=1}^{n}\left[\hat{Y}_{in}^* - Y_{in}^*\right]\nabla_\theta^2 f(\theta_0, X_i)\right](\hat{\theta}^{SD} - \theta_0) \\
&= \; (\hat{\theta}^{SD} - \theta_0)'B_{2n} + (\hat{\theta}^{SD} - \theta_0)'C_{2n}(\hat{\theta}^{SD} - \theta_0).
\end{aligned}$$

By Lemma A.1 applied with $\alpha = 1/2$, Assumption 9 and the fact $\nabla_\theta^2 f(\cdot,\cdot)$ is bounded (see Assumption 5-(i)), we obtain $\|B_{2n}\| = O_P(n^{-1/2})$ and $\|C_{2n}\|_2 = o_P(1)$. Collecting results and using the definition of $\hat{\theta}^{SD}$

$$\begin{aligned}
0 \; &\leq \; M_n^{SD}(\theta_0) - M_n^{SD}\left(\hat{\theta}^{SD}\right) \\
&\leq \; -A_n\|\hat{\theta}^{SD} - \theta_0\|^2 + (\hat{\theta}^{SD} - \theta_0)'\left\{B_{1n} + B_{2n}\right\} \\
& \quad + (\hat{\theta}^{SD} - \theta_0)'\left\{C_{1n} + C_{2n}\right\}(\hat{\theta}^{SD} - \theta_0).
\end{aligned}$$

Now, consider the event $E_n = \{A_n \geq 3A/4\} \cap \{\|C_{1n} + C_{2n}\|_2 \leq A/4\}$ and notice that $\mathbb{P}(E_n) \to 1$. On the event E_n we have

$$A\|\hat{\theta}^{SD} - \theta_0\|^2 - 2\|B_{1n} + B_{2n}\| \|\hat{\theta}^{SD} - \theta_0\| \leq 0 \ ,$$

that is $\|\hat{\theta}^{SD} - \theta_0\| \leq 2A^{-1}\|B_{1n} + B_{2n}\|$. Deduce that $\hat{\theta}^{SD} - \theta_0 = O_P(n^{-1/2})$.

For the WLS approach, we write

$$\begin{aligned} M_n^{WLS}(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\gamma(T_{in})[Y_{in} - f(\theta, X_i)]^2 - \frac{2}{n}\sum_{i=1}^{n}[W_{in} - \gamma(T_{in})]\, Y_{in} f(\theta, X_i) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}[W_{in} - \gamma(T_{in})]\, f^2(\theta, X_i) + \frac{1}{n}\sum_{i=1}^{n}[W_{in} - \gamma(T_{in})]\, Y_{in}^2 \\ &= M_{1n}^{WLS}(\theta) + M_{2n}^{WLS}(\theta) + M_{3n}^{WLS}(\theta) + R_n^{WLS}. \end{aligned}$$

The term $M_{1n}^{WLS}(\theta)$ can be handled like $M_{1n}^{SD}(\theta)$, while for $M_{2n}^{WLS}(\theta)$ and $M_{3n}^{WLS}(\theta)$ we use Lemma A.1 with $\alpha = 1/2$ and the Taylor expansion like we did for $M_{2n}^{SD}(\theta)$. Adapting the arguments used for the SD case, deduce $\hat{\theta}^{WLS} - \theta_0 = O_P(n^{-1/2})$.

ii) The case of a fixed alternative H_1 can be treated using the same ingredients: Lemma A.1 applied with suitable $\alpha \in (0, 1/2]$ and Taylor expansions. The details are omitted. ■

Proof of Theorem 4.4. By Lemma A.3 and the assumptions, $\sup_{\theta\in\Theta}\left|Q_n^\beta(\theta)\right|$ is bounded in probability. Then Lemma 4.3 indicates that it remains to look at the limit of $\tilde{Q}_n^\beta(\hat{\theta})$. By Taylor expansion, arguments like those used in Lemma A.6 above and the fact that $\hat{\theta} - \bar{\theta} = O_P(n^{-1/2})$, we obtain $\sup_{h\in\mathcal{H}_n}\left|\tilde{Q}_n^\beta(\hat{\theta}) - \tilde{Q}_n^\beta(\bar{\theta})\right| = o_P(1)$. Now, since

$$U^\beta(\bar{\theta}) = \left\{[\gamma(T_i) - 1]\left[m(X_i) + \varepsilon_i - \beta f\left(\bar{\theta}, X_i\right)\right] + \varepsilon_i\right\} + \left\{m(X_i) - f\left(\bar{\theta}, X_i\right)\right\}$$

and $\mathbb{E}[\gamma(T_i) \mid X_i] = 1$, we can decompose $\tilde{Q}_n^\beta(\bar{\theta})$ in three parts, a degenerate and a zero-mean $U-$process of order 2 (indexed by h) and

$$\frac{1}{n(n-1)h^p}\sum_{i\neq j}\left[m(X_i) - f\left(\bar{\theta}, X_i\right)\right]\left[m(X_j) - f\left(\bar{\theta}, X_j\right)\right] K_h(X_i - X_j)\,.$$

By the arguments used for (A.12), this last part tends to $\mathbb{E}\left[\left\{m(X) - f\left(\bar{\theta}, X\right)\right\}^2 g(X)\right]$ and for $\beta = 0$ or 1 the variances $[\hat{V}_n^\beta]^2$ converge to

$$2\int K^2(u)du\,\mathbb{E}\left\{\mathbb{E}^2\left[U^\beta(\bar{\theta})^2 \mid X\right] g(X)\right\}, \tag{A.16}$$

uniformly in $h \in \mathcal{H}_n$. It is easy to see that for $\beta = 0$ or $\beta = 1$,

$$\begin{aligned}\mathbb{E}\left[U^{\beta}(\bar{\theta})^2 \mid X\right] &= \mathbb{E}\left[\left\{Y - \beta f(\bar{\theta}, X)\right\}^2 \frac{G\left(Y\right)}{1 - G\left(Y\right)} \mid X\right] + \mathbb{E}\left[\varepsilon^2 \mid X\right] \\ &\quad + \left[m(X) - f\left(\bar{\theta}, X\right)\right]^2,\end{aligned}$$

and thus there is no general order relationship between the limits in equation (A.16). ■

Proof of Theorems 4.5 and 4.6. Once again, Lemma 4.3 shows that we only need to look at $\tilde{Q}_n^{\beta}(\hat{\theta})$. Write $U_i^{\beta}(\theta) = u_{in} + v_{in} + w_{in} + \lambda_n\left(X_i\right) + \left\{f(\theta_0, X_i) - f\left(\theta, X_i\right)\right\}$ where

$$\begin{aligned}u_{in} &= \left[\gamma\left(T_{in}\right) - 1\right]\lambda_n\left(X_i\right) \\ v_{in} &= \beta\left\{\gamma\left(T_{in}\right) - 1\right\}\left\{f(\theta_0, X_i) - f\left(\theta, X_i\right)\right\} \\ w_{in} &= \gamma\left(T_{in}\right)\varepsilon_i + (1 - \beta)\left[\gamma\left(T_{in}\right) - 1\right] f(\theta_0, X_i)\end{aligned}$$

and notice that $\mathbb{E}\left(u_{in} \mid X_i\right) = \mathbb{E}\left(v_{in} \mid X_i\right) = \mathbb{E}\left(w_{in} \mid X_i\right) = 0$ a.s. and there exists a sequence of real numbers $\overline{\sigma}_n^2$ tending to zero such that for each $n \geq 1$, $\mathbb{E}\left(u_{in}^2 \mid X_i\right) \leq \overline{\sigma}_n^2$. Using this decomposition of $U_i^{\beta}(\theta)$ we can split $\tilde{Q}_n^{\beta}(\hat{\theta})$ in several $U-$statistics of order 2. By repeated applications of Taylor expansion and Lemma A.5, and using the fact that $\hat{\theta} - \theta_0 = O_P\left(n^{-1/2}\right)$ deduce that

$$\begin{aligned}\tilde{Q}_n^{\beta}(\hat{\theta}) &= \frac{1}{n\left(n-1\right)}\sum_{i \neq j} w_{in} w_{jn} K_h\left(X_i - X_j\right) \\ &\quad + \frac{r_n^2}{n\left(n-1\right)}\sum_{i \neq j}\lambda\left(X_i\right)\lambda\left(X_j\right)\frac{1}{h^p}K_h\left(X_i - X_j\right) \\ &\quad + O_P\left(r_n n^{-1/2}\right) + o_P\left(n^{-1}h^{-p/2}\right),\end{aligned} \tag{A.17}$$

provided that $\lambda_n\left(\cdot\right) = r_n\lambda\left(\cdot\right).$ Moreover, since $\left|U_i^{\beta}(\hat{\theta}) - w_{in}\right| \leq o_P\left(1\right)\left[\gamma\left(T_{in}\right) + 1\right]$ with $o_P\left(1\right)$ independent of i,

$$\left[\hat{V}_n^{\beta}(\hat{\theta})\right]^2 - \frac{2}{n\left(n-1\right)h^p}\sum_{i \neq j}^{2} w_{in}^2 w_{jn}^2 K_h^2\left(X_i - X_j\right) = o_P\left(1\right). \tag{A.18}$$

From this and Lemma 2.1-(i) of Guerre and Lavergne (2005), the first $U-$statistic on the right-hand side of (A.17) multiplied by $nh^{p/2}$ and divided by $\hat{V}_n^{\beta}(\hat{\theta})$ converges in law to a standard normal distribution. Since the second $U-$statistic in (A.17) (without the r_n^2 factor) converges to $\mathbb{E}\left[\lambda^2(X)g(X)\right]$ in probability, and $\hat{V}_n^{\beta}(\hat{\theta})$ converges to a positive finite constant in probability, the proof of Theorem 4.5 is complete. Under the condition (4.29) in Theorem 4.6, the arguments used for (A.17) indicate that $\tilde{Q}_n^\beta(\hat{\theta})$ can be decomposed

$$\begin{aligned}
\tilde{Q}_n^\beta(\hat{\theta}) &= \frac{1}{n(n-1)} \sum_{i \neq j} w_{in} w_{jn} K_h(X_i - X_j) \\
&\quad + (\hat{\theta} - \theta_0)' \frac{2}{n(n-1)} \sum_{i \neq j} \lambda_n(X_i) \nabla_\theta f(\theta_0, X_j) \frac{1}{h^p} K_h(X_i - X_j) \\
&\quad + \frac{2}{n(n-1)} \sum_{i \neq j} \lambda_n(X_i) w_{jn} \frac{1}{h^p} K_h(X_i - X_j) \\
&\quad + \frac{1}{n(n-1)} \sum_{i \neq j} \lambda_n(X_i) \lambda_n(X_j) \frac{1}{h^p} K_h(X_i - X_j) \\
&\quad + \{\text{terms of smaller order}\} \\
&= \tilde{Q}_{na}^\beta + 2(\hat{\theta} - \theta_0)' \tilde{Q}_{nb}^\beta + 2\tilde{Q}_{nc}^\beta + \tilde{Q}_{nd}^\beta + \{\text{terms of smaller order}\}.
\end{aligned}$$

By Lemma A.5, $\tilde{Q}_{na}^\beta = O_P(n^{-1}h^{-p})$ and $|\tilde{Q}_{nc}^\beta| \leq O_P(n^{-1/2})\|\lambda_n\|_n$, while Lemma A.2 implies $|\tilde{Q}_{nb}^\beta| = O_P(1)\|\lambda_n\|_n$. Next, to obtain the rate of $\tilde{Q}_{nd}^\beta$, we follow the lines of the proof of Theorem 4 of Horowitz and Spokoiny (2001). See also Guerre and Lavergne (2005) and Lavergne and Patilea (2006). That is, approximating $\lambda_n(\cdot)$ by a piecewise polynomial function, we deduce

$$\tilde{Q}_{nd}^\beta \geq c\{1 + o_P(1)\} \left[\|\lambda_n\|_n - h^s\right]^2,$$

for some positive constant c, provided that $\lambda_n(\cdot) \in C(L, s)$ and the density $g(\cdot)$ is bounded away from zero. For the standard deviation, use (A.18) to deduce that $\hat{V}_n^\beta(\hat{\theta}) = O_P(1)$. Collecting results and taking h of order $n^{-2/(4s+p)}$, deduce that for any positive constant c, $\mathbb{P}(T_n^\beta(\hat{\theta}) > c) \to 1$ and this proves Theorem 4.6. ■

References

[1] Bose, A. and Sen, A. (2002). Asymptotic Distribution of the Kaplan–Meier U-Statistics. *J. Multivariate Anal.* **83**, 84–123.

[2] Buckley, J. and James, I. R. (1979). Linear regression with censored data. *Biometrika* **66**, 429–436.

[3] Chow, Y.S. and Teicher, H. (1997). *Probability Theory : Independence, Interchangeability, Martingales.* 3rd ed., Springer-Verlag, New-York.

[4] Delecroix, M., Lopez, O. and Patilea, V. (2006). Nonlinear censored regression using synthetic data. Working Paper No. 2006-10, CREST-INSEE.

[5] Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *Ann. Statist.* **27**, 1012–1040.

[6] Fan, J. and Gijbels, I. (1994). Censored regression: local linear approximations and their applications. *J. Amer. Statist. Assoc.* **89**, 560–570.

[7] Fan, J. and Huang, L.S. (2001). Goodness-of-fit tests for parametric regression models. *J. Amer. Statist. Assoc.* **96**, 640–652.

[8] Gill, R. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.* **11**, 49–58.

[9] Gray, R.J. and Pierce, D.A. (1985). Goodness-of-fit for censored data. *Ann. Statist.* **13**, 552–563.

[10] Guerre, E. and Lavergne, P. (2005). Data-driven rate-optimal specification testing in regression models. *Ann. Statist.* **33**, 840–870.

[11] Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1296–1947.

[12] Hjort, N.L. (1990). Goodness of fit test in models for life history data based on cumulative hazard rates. *Ann. Statist.* **18** , 1221–1258.

[13] Horowitz, J.L. and Spokoiny, V.G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69**, 599–631.

[14] Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly right censored data. *Ann. Statist.* **9**, 1276–1288.

[15] Lavergne, P. and Patilea, V. (2006). Breaking the curse of dimensionality in nonparametric testing. Working Paper 2006-24, CREST-INSEE (www.crest.fr)

[16] Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data,* 2nd ed. John Wiley & Sons, Inc., Hoboken.

[17] Lin, D.Y. and Spiekerman, C.F. (1996). Model checking techniques for parametric regression with censored data. *Scand. J. Statist.* **23**, 157–177.

[18] Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika* **74**, 301–309.

[19] Nolan, D. and Pollard, D. (1987). $U-$processes : Rates of convergence. *Ann. Statist.* **15**, 780–799.

[20] Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027–1057.

[21] Pardo-Fernández, J.C., van Keilegom, I. and González-Manteiga, W. (2005). Goodness-of-fit tests for parametric models in censored regression. DP 0519, Institut de Statistique, Louvain-la-Neuve. http://www.stat.ucl.ac.be

[22] Rudin, W. (1987). *Real and complex analysis.* McGraw-Hill.

[23] Sánchez-Sellero, C., González-Manteiga, W. and van Keilegom, I. (2005). Uniform representation of product-limit integrals with applications. *Scand. J. Statist.* **32**, 563–581.

[24] Sherman, R.P. (1994). Maximal inequalities for degenerate $U-$processes with applications to optimization estimators. *Ann. Statist.* **22**, 439–459.

[25] Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *J. Multivariate Anal.* **45**, 89–103.

[26] Stute, W. (1995). The central limit theorem under random censorship. *Ann. Statist.* **23**, 422–439.

[27] Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scand. J. Statist.* **23**, 461–471.

[28] Stute, W. (1997). Nonparametric models checks for regression. *Ann. Statist.* **25**, 613–641.

[29] Stute, W. (1999). Nonlinear censored regression. *Statistica Sinica* **9**, 1089–1102.

[30] Stute, W., González-Manteiga, W., and Sánchez-Sellero, C. (2000). Nonparametric model checks in censored regression. *Comm. Statist. Theory Methods* **29**, 1611–1629.

[31] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New-York.

[32] Zheng, Z. (1987). A class of estimators of the parameters in linear regression with censored data. *Acta Mathematicae Applicatae Sinica* **3**, 231–241.

[33] Zheng, J.X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75**, 263–289.

[34] Zhou, M. (1991). Some properties of the Kaplan-Meier estimator for independent nonidentically distributed random variables. *Ann. Statist.* **19**, 2266–2274.

[35] Zhou, M. (1992a). M-estimation in censored linear models. *Biometrika* **79**, 837–841.

[36] Zhou, M. (1992b). Asymptotic normality of the "synthetic data" regression estimator for censored survival data. *Ann. Statist.* **20**, 1002–1021.