Goodness-of-fit tests for conditional models under censoring and truncation

Ricardo Cao, Wenceslao González-Manteiga

Goodness-of-fit tests for conditional models under censoring and truncation

Ricardo Cao, Wenceslao González-Manteiga

PII: S0304-4076(07)00166-2
DOI: doi:10.1016/j.jeconom.2007.08.011
Reference: ECONOM 2976

To appear in: *Journal of Econometrics*

Cite this article as: Ricardo Cao and Wenceslao González-Manteiga, Goodness-of-fit tests for conditional models under censoring and truncation, *Journal of Econometrics* (2007), doi:10.1016/j.jeconom.2007.08.011

Goodness-of-fit tests for conditional models under censoring and truncation

Ricardo Cao*

Departamento de Matemáticas, Universidade da Coruña

Wenceslao González-Manteiga

Departamento de Estadística e Investigación Operativa,

Universidade de Santiago de Compostela

August 21, 2007

ABSTRACT

The problem of specification tests for conditional models is studied when the data are subject to left truncation and right censoring. A general method is applied to derive tests for the polynomial regression, the proportional hazards, the additive risks and the proportional odds models. Bootstrap versions to approximate the critical values of the test are introduced and proved to work both from a theoretical viewpoint as well as in a small simulation study.

JEL Classification: C12, C15, C24

Keywords: additive risks model, duration data, kernel method, polynomial regression, proportional hazards model, proportional odds, survival analysis.

*Departamento de Matemáticas, Universidade da Coruña, A Coruña 15071, Spain, phone: +34-981 167 000 - 1225, fax: +34-981 167 160, e-mail: rcao@udc.es

1 Introduction

During the last two decades, an increasing interest in the so-called duration data models has been detected in the econometric literature. A small sample of papers in this large field is Chamberlain (1986), Kiefer (1988) or Lewbel and Linton (2002). The main motivation for this interest comes from the fact that in many settings, an economic response variable, typically a duration, can be subject to censoring and/or truncation. Plenty of econometric studies with duration data try to explain this response in terms of a vector of covariates. In these circumstances, the presence of truncation or censoring is rather common. The first is due to individuals that cannot enter the study because their duration time has ended before the follow-up period. This is the well known left truncation (LT) phenomenon. The second problem appears when the duration time is not completely observed since, for instance, the evolution of the individual could not be followed till the end of his duration for whatever reason. Typically, only some lower bound for this duration time is available (the random time to a previous event that avoids observation of the duration end point). This is known as the right censoring (RC) mechanism.

To be precise let us assume that our interest vector is (Y, T, C), where Y is the duration time, T is the left truncation time and C is the right censoring time. Let us denote by F, G, H and L the distribution functions of Y, C, $Z = \min\{Y, C\}$ and T, respectively. A very common assumption that we also make for this model is that Y is independent of (T, C). In the left truncation and right censoring model (LTRC) we are only able to observe (Z, T, δ) with $\delta = 1_{\{Y \leq C\}}$, whenever $Z \geq T$. We will denote by H_1 the subdistribution function of Z given that $\delta = 1$. If $Z < T$ we are not able to observe anything. Of course, a reasonable assumption in this setup is $\alpha = P(T \leq Z) > 0$.

Using some initial sample, $((Z_1, T_1, \delta_1), \ldots, (Z_n, T_n, \delta_n))$, where $T_i \leq Z_i$ for

every $i = 1, 2, \ldots, n$, the distribution function F can be estimated by means of the well known product-limit estimate (see Tsai, Jewell and Wang (1987)):

$$1 - \hat{F}_n(y) = \prod_{i=1}^{n} \left(1 - \frac{1_{\{Z_i \leq y, \delta_i = 1\}}}{nC_n(Z_i)}\right), \tag{1}$$

where $C_n(y) = \frac{1}{n}\sum_{j=1}^{n} 1_{\{T_j \leq y \leq Z_j\}}$ is the empirical estimator of the underlying function $C(y) = P\left(T \leq y \leq Z|_{T \leq Z}\right)$.

An important setup in duration analysis is the conditional case. Now we are interested in how an explanatory vector, X, typically formed with risk factors, affects the duration Y. In this case we will assume that Y is independent of (T, C) conditionally on X. A general estimator in the single covariate context has been introduced by Iglesias-Pérez and González-Manteiga (1999):

$$1 - \hat{F}_n(y|_x) = \prod_{i=1}^{n} \left(1 - \frac{1_{\{Z_i \leq y, \delta_i = 1\}} B_{ni}(x)}{\sum_{j=1}^{n} 1_{\{T_j \leq Z_i \leq Z_j\}} B_{nj}(x)}\right), \tag{2}$$

where $\hat{F}_n(\bullet|_x)$ is an estimator of $F(\bullet|_x)$, conditional distribution function of $Y|_{X=x}$, $\{(X_i, Z_i, T_i, \delta_i)\}_{i=1}^{n}$ is the original sample and $\{B_{ni}\}_{i=1}^{n}$ is a nonparametric weight sequence, for instance, the Nadaraya-Watson weights (see Nadaraya (1964) and Watson (1964)), the k-nearest neighbours weights, local linear weights, etc. A particular case of the estimator (2) is the estimator introduced by Beran (1981), which applies to the case where only censoring is present. Some theoretical properties of the Beran estimator, including an asymptotic representation, have been studied in the series of papers by Van Keilegom and Veraverbeke (1996, 1997a, 1997b). Another important particular case is the estimator proposed by Lavalley and Akritas (1994) for the truncation setup, These two situations, but without covariates, correspond to the well-known estimators introduced by Kaplan and Meier (1958) and Lynden-Bell (1971). Finally (1) is a particular case of (2) when no covariates are present, by just defining $B_{ni}(x) = \frac{1}{n}$ for $i = 1, 2, \ldots, n$.

The existing literature for testing Cox model is abundant. See, for instance, Gray (1990), Lin and Wei (1991), Burke and Yuen (1995), McKeague and Sun (1996), Marzec and Marzec (1997), Verweij, Van Houwelingen and Stijnen (1998) and Peña (1998). Several methods are presented in these papers based on different approaches: cumulative hazard function estimation (smoothed or not), information matrix score tests and extensions of smooth specification tests. Some alternative methods for testing other models, as the accelerated lifetime model or related models, can be found in Lin and Spiekerman (1996) and Stute, González-Manteiga and Sánchez-Sellero (2000), but always in the context of censored data. Some other relevant references in this context are Stute (1999) and Sánchez-Sellero, González-Manteiga and Van Keilegom (2005). No one of these papers deals with the general left censoring and right truncation situation that is considered in the present paper.

In the rest of the paper we will concentrate on the estimator given in (2) and use it to define some goodness-of-fit test statistics for different conditional curves in this LTRC setup. General polynomial regression models (PR), proportional hazards models (PH), additive risks models (AR) and proportional odds models (PO) are introduced in Section 2, where it is also indicated how the specification test statistic is defined. The asymptotic properties of the tests are presented in Section 3, while Section 4 contains the definition and theoretical properties of some bootstrap versions of these tests. The practical behaviour of these tests is analyzed through the simulation study included in Section 5. Finally, Section 6 contains the proofs.

2 Remarkable conditional models

One of the main aims of duration analysis, in a conditional setup, is to analyze how an explanatory q-dimensional covariate vector, $\mathbf{X}$, influences the duration,

Y. This dependency may be modelled in a different number of ways but, typically, the key idea is to assume some kind of functional relationship for the conditional distribution function or some other conditional curve. In the following we make a short overview of some of the most popular conditional models in this setup. Models PH, AR and PO have been studied in detail by Grigoletto and Akritas (1999).

2.1 General regression (GR) and polynomial regression (PR) models

We assume that the conditional distribution function, $F\left(\bullet|_{\mathbf{x}}\right)$, satisfies

$$T\left(F\left(\bullet|_{\mathbf{x}}\right)\right) = A^t\left(\mathbf{x}\right)\boldsymbol{\beta}$$

where $A : \mathbb{R}^q \to \mathbb{R}^p$ is a know function, $\boldsymbol{\beta} = \left(\beta_0, \beta_1, \ldots, \beta_p\right)^t$ is the vector of unknown parameters, the functional T is given by $T\left(N\right) = \int_0^1 N^{-1}\left(s\right) J\left(s\right) ds$, for any distribution function, N, $N^{-1}\left(s\right) = \inf\left\{u : N\left(u\right) \geq s\right\}$ is the quantile function and J is some nonnegative real function satisfying $\int_0^1 J\left(s\right) ds = 1$. In other words, T is an L-functional in the terminology of Serfling (1980, p. 265). If J is the density function of a $U\left[0,1\right]$ distribution, it is easy to check that $T\left(F\left(\bullet|_{\mathbf{x}}\right)\right) = E\left(Y|_{\mathbb{X}=\mathbf{x}}\right)$, i.e. we are imposing a polynomial structure to the regression function of Y given $\mathbf{X}$. In duration analysis, where the duration is typically asymmetric, it seems more reasonable to set $J\left(s\right) = \frac{1}{b-a} I_{\{a \leq s \leq b\}}$ for some $[a,b] \subsetneqq [0,1]$, which gives rise to trimmed conditional means or, in an extreme case, the conditional median.

In the special case of a one-dimensional covariate ($q = 1$), a popular choice is $A\left(x\right) = \left(1, x, \ldots, x^p\right)^t$, which leads to a polynomial regression model:

$$T\left(F\left(\bullet|_{\mathbf{x}}\right)\right) = \beta_0 + \beta_1 x + \ldots + \beta_p x^p. \tag{3}$$

For a detailed study of model (3) see Akritas (1996).

2.2 Proportional hazards model (PH)

The idea behind this model is two split the conditional hazard rate as a product of two factors:

$$\lambda\left(t|_{\mathbf{x}}\right) = \lambda_0\left(t\right)\exp\left(A^t\left(\mathbf{x}\right)\boldsymbol{\beta}\right), \tag{4}$$

where λ_0 is the so-called baseline hazard rate and $\boldsymbol{\beta} = \left(\beta_0, \beta_1, \ldots, \beta_p\right)^t$ is a vector of unknown real constants. This popular model, also known as Cox regression model, was proposed by Cox (1972) and can also be expressed in terms of the cumulative hazard function. If we then let the coefficients β_j, in equation (4), depend on the duration, we get the general Cox model with time depending coefficients:

$$\Lambda\left(t|_{\mathbf{x}}\right) = \Lambda_0\left(t\right)\exp\left(A^t\left(\mathbf{x}\right)\boldsymbol{\beta}\left(t\right)\right). \tag{5}$$

This model has been studied in Marzec and Marzec (1997).

2.3 Additive risks model (AR)

An alternative approach to model the conditional hazard rate is

$$\lambda\left(t|_{\mathbf{x}}\right) = \lambda_0\left(t\right) + A^t\left(\mathbf{x}\right)\boldsymbol{\beta}. \tag{6}$$

It is clear that some conditions on the term $A^t\left(\mathbf{x}\right)\boldsymbol{\beta}$ are needed now in order for $\lambda\left(t|_{\mathbf{x}}\right)$ to be a hazard rate. This model is a special case of the multiplicative intensity model suggested by Aalen (1980).

2.4 Proportional odds model (PO)

The conditional cumulative hazard function of Y given $\mathbf{X} = \mathbf{x}$, $\Lambda\left(t|_{\mathbf{x}}\right)$ is assumed to satisfy the following equation

$$\operatorname{logit}\left(1 - \exp\left(-\Lambda\left(t|_{\mathbf{x}}\right)\right)\right) = \alpha\left(t\right) + A^t\left(\mathbf{x}\right)\boldsymbol{\beta}, \tag{7}$$

where $\alpha(t)$ is an increasing function and $\text{logit}(u) = \ln\left(\frac{u}{1-u}\right)$. Observe that this model can be expressed in terms of the odds ratio:

$$\frac{P\left(Y \leq t|_{\mathbf{X}=\mathbf{x}}\right)}{P\left(Y > t|_{\mathbf{X}=\mathbf{x}}\right)} = \exp\left(\alpha(t) + A^t(\mathbf{x})\boldsymbol{\beta}\right).$$

2.5 Parameter estimation

For the multivariate case ($q > 1$), a typical selection is $A(\mathbf{x}) = \mathbf{x}$. To avoid the problems coming from the curse of dimensionality one needs to use preliminary $\sqrt{n}$-consistent estimators of $\boldsymbol{\beta}$ under the model in order to estimate quantities of the form $F\left(t|_{\mathbf{X}^t\boldsymbol{\beta}=u}\right)$. Along this subsection only the one-dimensional covariate case ($q = 1$) will be considered in detail. In such a case the curse of dimensionality is no longer a problem and, for the choice $A(x) = (1, x, \ldots, x^p)^t$, we will propose $\sqrt{n}$-consistent estimators of $\boldsymbol{\beta}$ that will be incorporated into the test statistic.

In Akritas (1996) some properties are studied for a general least squares estimator, $\hat{\boldsymbol{\beta}}$, of the parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^t$ in the PR model for complete, censored or truncated data, but not when both mechanisms are present. This estimator is defined as $\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \hat{\psi}_n(\boldsymbol{\beta})$, where

$$\hat{\psi}_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{r=1}^{n}\left(\hat{m}_r - (\beta_0 + \beta_1 X_r + \cdots + \beta_p X_r^p)\right)^2 \tag{8}$$

and $\hat{m}_r = T\left(\hat{F}_n\left(\bullet|_{X_r}\right)\right)$ is an estimator of $m_r = T\left(F\left(\bullet|_{X_r}\right)\right)$ for $r = 1, 2, \ldots, n$. To avoid definiteness problems with $T\left(\hat{F}_n\left(\bullet|_{X_r}\right)\right) = \int_0^1 \hat{F}_n^{-1}\left(s|_{X_r}\right) J(s)\, ds$ we modify the estimator $\hat{F}_n\left(\bullet|_{X_r}\right)$, if necessary, forcing it to attain the value 1 in the largest point with positive probability mass. This results in

$$\hat{\boldsymbol{\beta}} = \left(\chi^t\chi\right)^{-1}\chi^t\hat{m}, \tag{9}$$

where $\hat{m} = (\hat{m}_1, \hat{m}_2, \ldots, \hat{m}_n)^t$ and $\chi = \left(X_i^{k-1}\right)_{\substack{i=1,2,\ldots,n \\ k=1,2,\ldots p+1}}$ is the design matrix.

Grigoletto and Akritas (1999) define $\hat{\boldsymbol{\beta}}$, some version of the least squares estimator of the true parameter, $\boldsymbol{\beta}$, in models PH, AR and PO:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \bar{\psi}_n(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\beta}} \frac{1}{n}\sum_{r=1}^{n}\left(\hat{\Omega}_r - (\beta_0 + \beta_1 X_r + \cdots + \beta_p X_r^p)\right)^2, \tag{10}$$

where $\hat{\Omega}_r$ is some estimation of Ω_r, a suitable transformation of $\Lambda(\bullet|_{X_r})$ that will change from one model to other. An explicit expression for the estimator in (10) can be easily derived $\hat{\boldsymbol{\beta}} = (\chi^t\chi)^{-1}\chi^t\hat{\Omega}$, with $\hat{\Omega} = \left(\hat{\Omega}_1, \hat{\Omega}_2, \ldots, \hat{\Omega}_n\right)^t$. These authors give some asymptotic properties of $\hat{\boldsymbol{\beta}}$ under either censoring or truncation but not for both simultaneously.

For model PH, using a weight function, W, satisfying $W(s) \geq 0$ for $s \in [0,\infty)$ and $\int_0^\infty dW(s) = 1$, straight forward manipulations of equation (4) give

$$\begin{aligned}\int_0^\infty \ln\Lambda(s|_x)\, dW(s) &= \int_0^\infty \ln\Lambda_0(s)\, dW(s) + \beta_0 + \beta_1 x + \cdots + \beta_p x^p \\ &= \beta_0' + \beta_1 x + \cdots + \beta_p x^p,\end{aligned}$$

with $\beta_0' = \beta_0 + \int_0^\infty \ln\Lambda_0(s)\, dW(s)$. Parallel calculations for model (5) give:

$$\begin{aligned}\int_0^\infty \ln\Lambda(s|_x)\, dW(s) &= \tilde{\beta}_0 + \tilde{\beta}_1 x + \cdots + \tilde{\beta}_p x^p, \text{ with} \\ \tilde{\beta}_0 &= \int_0^\infty \ln\Lambda_0(s)\, dW(s) + \int_0^\infty \beta_0(s)\, dW(s), \\ \tilde{\beta}_j &= \int_0^\infty \beta_j(s)\, dW(s), \text{ for } j = 1, 2, \ldots, p,\end{aligned}$$

which essentially tells that this model can also be transformed to a polynomial regression model. Although the functions $\beta_j(t)$ are not identifiable then, this idea can be used for testing (5).

Now, it is clear then that $\Omega_x = \int_0^\infty \ln\Lambda(s|_x)\, dW(s)$, $\Omega_r = \Omega_{X_r}$ and $\hat{\Omega}_r = \int_0^\infty \ln\hat{\Lambda}_n(s|_{X_r})\, dW(s)$, where $\hat{\Lambda}_n(t|_x)$ is the conditional cumulative hazard function estimator connected to (2), i.e.

$$\hat{\Lambda}_n(t|_x) = \int_{-\infty}^{t} \frac{d\hat{H}_{1n}^*(s|_x)}{\hat{C}_n(s|_x)}, \tag{11}$$

where $\hat{H}^*_{1n}(s|_x)$ and $\hat{C}_n(s|_x)$ are nonparametric estimators of the conditional subdistribution function $H^*_1(s|_x) = P(Z \leq s, \delta = 1|_{T \leq Z, X=x})$ and the populational function $C(s|_x) = P(T \leq s \leq Z|_{T \leq Z, X=x})$.

For model AR simple algebra gives

$$\int_0^\infty \Lambda(s|_x)\, d\tilde{W}(s) = \beta_0'' + \beta_1 x + \cdots + \beta_p x^p,$$

with

$$\beta_0'' = \beta_0 + \frac{\int_0^\infty \Lambda_0(s)\, dW(s)}{\int_0^\infty s dW(s)}$$

and

$$\tilde{W} = \frac{W}{\int_0^\infty s dW(s)}.$$

As a consequence

$$\Omega_x = \int_0^\infty \Lambda(s|_x)\, d\tilde{W}(s),\ \Omega_r = \Omega_{X_r} \text{ and } \hat{\Omega}_r = \int_0^\infty \hat{\Lambda}_n(s|_{X_r})\, d\tilde{W}(s),$$

where $\hat{\Lambda}_n(t|_x)$ is the estimator defined in (11).

Under model PO, integrating both terms in (7) gives

$$\int_0^\infty \text{logit}\left(1 - \exp\left(-\Lambda(s|_x)\right)\right) dW(s) = \beta_0''' + \beta_1 x + \cdots + \beta_p x^p,$$

with $\beta_0''' = \beta_0 + \int_0^\infty \alpha(s)\, dW(s)$. Hence, $\Omega_r = \Omega_{X_r}$, with

$$\Omega_x = \int_0^\infty \text{logit}\left(1 - \exp\left(-\Lambda(s|_x)\right)\right) dW(s)$$

and

$$\hat{\Omega}_r = \int_0^\infty \text{logit}\left(1 - \exp\left(-\hat{\Lambda}_n(s|_{X_r})\right)\right) dW(s).$$

3 The test statistic

Let us first focus in model PR. The problem under study is to test

$$\begin{aligned} H_0 &: \exists \boldsymbol{\beta} \in \mathbb{R}^{p+1} \text{ such that equation (3) holds} \\ &\quad \text{versus the alternative} \\ H_1 &: \text{equation (3) does not hold for any } \boldsymbol{\beta} \in \mathbb{R}^{p+1}. \end{aligned} \tag{12}$$

An intuitive way of measuring the discrepancy between the hypothesized model and the data is to consider the stochastic process (8) at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$:

$$\hat{\psi}_n\left(\hat{\boldsymbol{\beta}}\right) = \frac{1}{n}\sum_{r=1}^{n}\left(\hat{m}_r - \left(\hat{\beta}_0 + \hat{\beta}_1 X_r + \cdots + \hat{\beta}_p X_r^p\right)\right)^2.$$

This is a kind of L_2-distance between the transformed data and the fitted model. Hence, H_0 should be rejected when $\hat{\psi}_n\left(\hat{\boldsymbol{\beta}}\right)$ is large.

In models PH, AR and PO, the hypothesis testing is completely similar to (12) but replacing equation (3) by (4), (6) or (7), respectively. The definition of the statistic is also parallel for models PH, AR and PO:

$$\bar{\psi}_n\left(\hat{\boldsymbol{\beta}}\right) = \frac{1}{n}\sum_{r=1}^{n}\left(\hat{\Omega}_r - \left(\hat{\beta}_0 + \hat{\beta}_1 X_r + \cdots + \hat{\beta}_p X_r^p\right)\right)^2.$$

These test statistics have to be multiplied by a normalizing sequence in order to have a limit distribution. This leads to the test statistics

$$\begin{aligned} T_n^{(1)} &= n\sqrt{h}\hat{\psi}_n\left(\hat{\boldsymbol{\beta}}\right) \text{ for model PR,} \\ T_n^{(2)} &= n\sqrt{h}\bar{\psi}_n\left(\hat{\boldsymbol{\beta}}\right) \text{ for models PH, AR and PO.} \end{aligned}$$

In order to obtain the limit distribution of these statistics under the null hypothesis some conditions have to be assumed. To do this we introduce some notation. We denote by $L\left(\bullet|_x\right)$ and $H\left(\bullet|_x\right)$ the distribution functions of T and Z, conditionally on $X = x$. In a similar way, $H_1\left(\bullet|_x\right)$ is the conditional subdistribution function of Z for $\delta = 1$, $L^*\left(\bullet|_x\right)$ is the distribution function of T, given $T \leq Z$ and $X = x$ and $H_1^*\left(\bullet|_x\right)$ is the subdistribution function of the variable Z for uncensored data conditionally on $T \leq Z$ and $X = x$, with $h_1^*\left(\bullet|_x\right)$ its pertaining conditional subdensity. The conditions needed, C1–C14, are collected in Section 6.

Let us state our main result.

Theorem 1. *Assume conditions C1-C14. Then, under* H_0,

$$T_n^{(1)} - b_{0h} \xrightarrow{\mathrm{d}} N\left(0, V\right),$$

with $b_{0h} = h^{-\frac{1}{2}} K^{(2)}(0) \int \sigma^2(x)\,dx$ *and* $V = 2K^{(4)}(0) \int \sigma^4(x)\,dx$. *Here* $K^{(2)}(u) = K * K(u)$ *and* $K^{(4)}(u) = K * K * K * K(u)$, *where* $*$ *denotes convolution and the definition of* $\sigma^2(x)$ is in C11.

In order to deal with some of the terms coming up in the proof of Theorem 1, we state the following lemma. Its proof is also included in Section 6.

Lemma 2. *Assume the conditions in Theorem 1 and* H_0, *then* $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = O_P\left(n^{-\frac{1}{2}}\right)$.

We now present an asymptotic distribution result (under the null hypothesis) for the statistic $T_n^{(2)}$ used for testing models PH, AR and PO. The new conditions needed here are presented in Section 6.

The main result for the test statistic $T_n^{(2)}$ is the following.

Theorem 3. *Assume conditions C1', C3-C10, C11', C12, C13' and C14'. Then, under* H_0,

$$T_n^{(2)} - b_{0h} \xrightarrow{\mathrm{d}} N(0, V),$$

with $b_{0h} = h^{-\frac{1}{2}} K^{(2)}(0) \int \sigma^2(x)\,dx$ *and* $V = 2K^{(4)}(0) \int \sigma^4(x)\,dx$, *where*

$$\sigma^2(x) = \begin{cases} \sigma^2_{PH}(x) = Var\left(\eta_{PH}(Z,T,\delta,x)|_{X=x}\right) & \text{for model PH} \\ \sigma^2_{AR}(x) = Var\left(\eta_{AR}(Z,T,\delta,x)|_{X=x}\right) & \text{for model AR} \\ \sigma^2_{PO}(x) = Var\left(\eta_{PO}(Z,T,\delta,x)|_{X=x}\right) & \text{for model PO} \end{cases}$$

and $\eta_{PH}(Z,T,\delta,x)$, $\eta_{AR}(Z,T,\delta,x)$ *and* $\eta_{PO}(Z,T,\delta,x)$ *will be defined in (16).*

Now, some comments related to the previous results follow.

Remark 4. Condition $\frac{nh^3}{(\ln n)^3} \to \infty$ in C12 could be replaced by the somewhat weaker condition of the form $\frac{nh^2}{(\ln n)^2} \to \infty$ if the almost sure order, $O\left(\frac{\ln n}{nh}\right)^{\frac{3}{4}}$, for the negligible term in Iglesias-Pérez and González-Manteiga (1999) could be improved, for instance, to $O\left(\frac{\ln n}{nh}\right)$.

Remark 5. If we choose J to be the uniform density in $(0,1)$, the functional T used in model PR becomes

$$T(N) = \int_0^1 N^{-1}(s)\,ds = \int_0^\infty t\; dN(t).$$

As a consequence, $T\left(F\left(\bullet|_x\right)\right) = E\left(Y|_{X=x}\right)$. In the case of no censoring and no truncation the estimator (2) reduces to

$$\hat{F}_n\left(y|_x\right) = \frac{\sum_{j=1}^n 1_{\{Z_j \leq y\}} B_{nj}\left(x\right)}{\sum_{j=1}^n B_{nj}\left(x\right)}$$

and $T\left(\hat{F}_n\left(\bullet|_x\right)\right)$ is the classical Nadaraya-Watson kernel estimator of the regression function. Under these circumstances, the test statistic $T_n^{(1)}$ coincides with the test statistic ΔASE proposed by González-Manteiga and Cao (1993) except for the fixed design used in that paper. The limit distribution of ΔASE in that paper is also a particular case of our Theorem 1 above, without censoring and truncation. In the same sense $T_n^{(1)}$ can be also viewed as a generalization of the test proposed in Härdle and Mammen (1993) to the case with censoring and truncation. However, these authors did not consider least squares parameter estimators with the smoothed responses.

Remark 6. As soon as one transforms the goodness-of-fit problem of interest into the model check for linearity of the values $(X_r, \hat{m}_r)$ it is clear that plenty of the alternative approaches for constructing specification tests in regression can be directly used in this setup. For instance, the ideas in Dette (1999) can be translated to this setup by of constructing a test based on the difference between two estimators of the integrated conditional variance. One based on the linearity assumption of the $(X_r, \hat{m}_r)$ and the other that is purely nonparametric. Similar ideas could be exploited for models PH, AR and PO using the data $\left(X_r, \hat{\Omega}_r\right)$.

Remark 7. Let us consider some local alternative to model PR of the following form

$$T\left(F\left(\bullet|_x\right)\right) = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + n^{-\frac{1}{2}} h^{-\frac{1}{4}} g\left(x\right),$$

where $g\left(x\right)$ is a squared integrable function orthogonal to $\beta_0 + \beta_1 x + \cdots + \beta_p x^p$.

A careful inspection of the proof of Theorem 1 gives

$$T_n^{(1)} - b_{0h} \xrightarrow{\mathrm{d}} N\left(\int g(x)^2 m(x)\, dx, V\right),$$

under this local alternative model, where m is the density function of X. This implies that the test is able to detect local alternatives that approach the hypothesized model at any rate slower or equal to $n^{-\frac{1}{2}} h^{-\frac{1}{4}}$.

A similar property can be derived for the test $T_n^{(2)}$ for suitable local alternatives to models PH, AR and PO. It is clear that such local alternatives have to be formulated in terms of

$$\Omega_x = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + n^{-\frac{1}{2}} h^{-\frac{1}{4}} g(x),$$

where Ω_x is the appropriate transformation introduced for each of these.

Remark 8. As stated in the previous section in order to extend the above techniques to the GR model, in the multivariate case, some preliminary $\sqrt{n}$-consistent estimator, $\tilde{\boldsymbol{\beta}}$, under the null model, is needed. Using the data $\left\{\left(\mathbf{X}_i^t \tilde{\boldsymbol{\beta}}, Z_i, T_i, \delta_i\right)\right\}_{i=1}^n$ one can estimate the conditional distribution function $F\left(t|_{\mathbf{X}^t\boldsymbol{\beta}=u}\right)$ using (2), already introduced for the single covariate case. Now the values $\hat{m}_r = T\left(\hat{F}_n\left(\bullet|_{\mathbf{X}_r^t\tilde{\boldsymbol{\beta}}}\right)\right)$ can be used in (8) to define $\hat{\boldsymbol{\beta}}$, a second stage estimator of $\boldsymbol{\beta}$, and to construct the test statistic in a parallel way to what has been proposed for model PR. Parallel procedures can be designed for models PH, AR and PO to multivariate settings. Similar arguments have been applied in previous papers in order to avoid the curse of dimensionality, but only for the complete data case (see Hall and Yao (2005)). The asymptotic null distribution in the general LTRC case is an open question.

4 Bootstrap version

The quadratic form structure of the dominant terms in the test statistics $T_n^{(1)}$ and $T_n^{(2)}$ suggest that the convergence of their distribution to the normal limit

will be slow. To remedy this, the bootstrap method can be used to approximate the sampling distribution of these statistics by the resampling distribution of some bootstrap versions. Different bootstrap resampling plans have been introduced to mimic the sampling distribution of some statistic for conditional models in duration analysis. Some of these are presented in the papers by Burke and Yuen (1995), Burr and Doss (1993), Burr (1994) and Doss and Chiang (1994). All these papers use bootstrap resampling plans that mimic the hypothesized model, which is not always possible in our context, since model PR do not fully specify the conditional distribution of the duration. For this reason we adopt a residual-based bootstrap resampling in the line of that proposed in Stute, González-Manteiga and Sánchez-Sellero (2000) and Zhu, Yuen and Tang (2002) in different contexts. The bootstrap method that will be proposed next can also be regarded just as a Monte Carlo approximation (see the book by Zhu (2005) for details).

We first deal with the goodness-of-fit problem for model PR. Since the conditional distribution $F\left(\bullet|_x\right)$ cannot be uniquely determined by equation (3), a reasonable way of bootstrapping the test statistic $T_n^{(1)}$ is to mimic the dominant terms of this statistic studied in the proof of Theorem 1. It is clear from (41) that the limit distribution of $T_n^{(1)}$ is the same as that of

$$\sqrt{n^2 h}\frac{1}{n}\sum_{r=1}^{n}\left(\sum_{i=1}^{n} B_{ni}\left(X_r\right)\varepsilon_i\right)^2.$$

We now use the bootstrap version of this statistic to approximate the null distribution of $T_n^{(1)}$. Let us define some estimators of the errors of the model as defined in (31):

$$\hat{\varepsilon}_i = \int \left(1-\hat{F}_g\left(y|_{X_i}\right)\right)\hat{\xi}\left(Z_i, T_i, \delta_i, X_i, y\right) J\left(\hat{F}_g\left(y|_{X_i}\right)\right) dy, \qquad (13)$$

where $\hat{F}_g\left(y|_x\right)$ is the estimator defined in (2) but using a pilot bandwidth g

instead of the original bandwidth h,

$$\hat{\xi}\left(Z,T,\delta,x,y\right)=\frac{1_{\{Z\leq y,\delta=1\}}}{\hat{C}_g\left(Z|_x\right)}-\int_0^y\frac{1_{\{T\leq u\leq Z\}}}{\hat{C}_g^2\left(u|_x\right)}d\hat{H}_{1g}^*\left(u|_x\right)$$

is the empirical version of (15), with $\hat{C}_g\left(z|_x\right)$ the kernel estimator of $C\left(z|_x\right)$ given by

$$\hat{C}_g\left(z|_x\right)=\frac{\frac{1}{ng}\sum_{i=1}^n 1_{\{T_i\leq z\leq Z_i\}}K\left(\frac{x-X_i}{g}\right)}{\frac{1}{ng}\sum_{i=1}^n K\left(\frac{x-X_i}{g}\right)}=\sum_{i=1}^n B_{gi}\left(x\right)1_{\{T_i\leq z\leq Z_i\}},$$

with

$$B_{gi}\left(x\right)=\frac{K\left(\frac{x-X_i}{g}\right)}{\sum_{j=1}^n K\left(\frac{x-X_i}{g}\right)} \tag{14}$$

and the kernel estimator of $H_1^*\left(u|_x\right)$ is $\hat{H}_{1g}^*\left(u|_x\right)=\sum_{i=1}^n B_{gi}\left(x\right)1_{\{Z_i\leq u\}}\delta_i$. Now, the bootstrap version of $T_n^{(1)}$ is defined as

$$T_n^{(1)*}=\sqrt{n^2h}\frac{1}{n}\sum_{r=1}^n\left(\sum_{i=1}^n B_{ni}\left(X_r\right)\varepsilon_i^*\right)^2,$$

where the error resample is given by $\varepsilon_i^*=W_i^*\hat{\varepsilon}_i$, $i=1,2,\ldots,n$ and the W_i^* are random observations (independent on the original sample) satisfying $E^*\left(W_i^*\right)=0$ and $E^*\left(W_i^{*2}\right)=1$, where E^* is a notation for bootstrap expectation, i.e., conditional expectation given the initial sample.

In order to state the main result that warranties the validity of the bootstrap version we need further integrability assumptions and some limit conditions on the pilot bandwidth. These are collected in Section 6.

The asymptotic validity of the bootstrap version of $T_n^{(1)}$ is stated next.

Theorem 9. *Assume conditions C1-C16. Then, under* H_0*,*

$$T_n^{(1)*}-b_{0h}\xrightarrow{\mathrm{d}^*}N\left(0,V\right),\ \textit{in probability.}$$

In order to construct bootstrap resampling plans for models PH, AR and PO we need to define some residuals, parallel to (13), but according to the

error structure of every model in (50). We first define the residuals

$$\hat{\varepsilon}_i' = \begin{cases} \hat{\varepsilon}_i^{PH} = \int \hat{\xi}\left(Z_i, T_i, \delta_i, X_i, y\right) \frac{w(y)}{\hat{\Lambda}_g\left(y|_{X_i}\right)} dy & \text{for model PH} \\ \hat{\varepsilon}_i^{AR} = \int \hat{\xi}\left(Z_i, T_i, \delta_i, X_i, y\right) \tilde{w}\left(y\right) dy & \text{for model AR} \\ \hat{\varepsilon}_i^{PO} = \int \hat{\xi}\left(Z_i, T_i, \delta_i, X_i, y\right) \frac{w(y)}{\hat{F}_g\left(y|_{X_i}\right)} dy & \text{for model PO} \end{cases}$$

where the functions w and $\tilde{w}$ are the densities pertaining to W and $\tilde{W}$,

$$\hat{\Lambda}_g\left(y|_{X_i}\right) = \sum_{r=1}^{n} \frac{1_{\{Z_r \leq y, \delta_r = 1\}} B_{gr}\left(X_i\right)}{\sum_{j=1}^{n} 1_{\{T_j \leq Z_r \leq Z_j\}} B_{gj}\left(X_i\right)},$$

is the Iglesias-Pérez and González-Manteiga (1999) estimator and $B_{gr}\left(x\right)$ has been defined in (14). The bootstrap version of the test statistic is then

$$T_n^{(2)*} = \sqrt{n^2 h} \frac{1}{n} \sum_{r=1}^{n} \left(\sum_{i=1}^{n} B_{ni}\left(X_r\right) \varepsilon_i'^{*} \right)^2,$$

where the bootstrap observations are given by $\varepsilon_i'^{*} = W_i^* \hat{\varepsilon}_i'$, $i = 1, 2, \ldots, n$ and the W_i^* are iid random observations with the properties mentioned above.

Some new assumption is needed to prove the consistency of the bootstrap version for models PH, AR and PO. This is condition C15', in Section 6. Next, we prove that $T_n^{(2)*}$ has the same limit behaviour as its counterpart $T_n^{(2)}$.

Theorem 10. *Assume conditions C1', C3-C10, C11', C12, C13'-C15' and C16. Then, under H_0, with probability one*

$$T_n^{(2)*} - b_{0h} \xrightarrow{\mathrm{d}^*} N\left(0, V\right).$$

The proof of this result is a straightforward modification of the proof of Theorem 9. For this reason it is not included in Section 6.

Remark 11. The comments in Remark 8 are also valid for the bootstrap version presented in this section.

5 Simulations

We have used a Koziol-Green Weibull conditional model in the simulation study. Let us consider $Y|_{X=x} \overset{\mathrm{d}}{=} W(\alpha(x), \beta(x))$, i.e.

$$1 - F(y|_x) = \begin{cases} \exp\left(-(\beta(x)\, y)^{\alpha(x)}\right) & \text{if } y \geq 0 \\ 1 & \text{if } y < 0 \end{cases}$$

for some functions $\alpha(x)$ and $\beta(x)$ to be defined later. The censoring conditional survival function is defined in terms of the interest one as in Koziol-Green model: $1 - G(y|_x) = (1 - F(y|_x))^{\eta(x)}$ for some function $\eta(x)$. The conditional distribution of the truncation variable is also of Koziol-Green type: $1 - L(y|_x) = (1 - F(y|_x))^{\nu(x)}$. Under this model, standard calculations lead to the following expression for the conditional hazard rate of interest:

$$\lambda(y|_x) = \alpha(x)\, \beta(x)^{\alpha(x)}\, y^{\alpha(x)-1}.$$

The function β that will be used in our simulation study is of the form $\beta(x) = \exp(\beta_0 + \beta_1 x + \cdots + \beta_p x^p)$.

Setting $\alpha(x) = \alpha_0$ (a constant) we are in Cox proportional hazard model, since the hazard rate can be written as a product:

$$\lambda(y|_x) = \lambda_0(y) \exp\left(\beta_0' + \beta_1' x + \cdots + \beta_p' x^p\right),$$

where $\lambda_0(y) = \alpha_0 y^{\alpha_0 - 1}$ and $\beta_j' = \alpha_0 \beta_j$, $j = 1, 2, \ldots, p$. To examine the power of our test we considered quadratic alternatives of the form $\alpha(x) = \alpha_0 + cx^2$. This gives the conditional hazard rate

$$\lambda(y|_x) = \left(\alpha_0 + cx^2\right) y^{\alpha_0 - 1 + cx^2} \exp\left(\left(\beta_0 + \beta_1 x + \cdots + \beta_p x^p\right)\left(\alpha_0 + cx^2\right)\right).$$

Along the simulation study we considered $2\beta(2,2)$ for the marginal distribution of X. The probability of uncensoring and the probability of observation

(absence of truncation) can be easily calculated when the functions η and ν are constants: $\eta(x) \equiv \eta$ and $\nu(x) \equiv \nu$. These two probabilities are then

$$\begin{aligned} \gamma &= P(Y \leq C) = E(\delta) = \frac{1}{1+\eta}, \\ \alpha &= P(T \leq Z) = \frac{\nu}{1+\eta+\nu}. \end{aligned}$$

Three models, with different values for η and ν, have been considered to deal with different degrees of censoring and truncation. These are reported in Table 1. The nominal significant level fixed for the test was 0.05.

Model	η	ν	$1-\gamma$	$1-\alpha$
1	0.25	5	0.2000	0.2000
2	0.50	4	0.3333	0.2727
3	0.75	3	0.4286	0.3684

Table 1. Probabilities of censoring $(1-\gamma)$ and truncation $(1-\alpha)$.

In order to compute the test statistic, the values $\hat{\Omega}_r$ have to be obtained first. To do this we need to choose some weight function, W, and to compute the estimator in (11). The integral that defines $\hat{\Omega}_r$ can be written as an empirical sum to avoid numerical integration. Our choice for W has been a uniform distribution in some interval, $[L_1, L_2]$, whose limits have been chosen to prevent boundary effects problems. The estimator in (11) has been computed using the gaussian kernel. In order to examine the effect of the choice of the smoothing parameter, h, several values have been tried along the simulations. This approach has been used as well for the pilot bandwidth, g, in the bootstrap resampling. Throughout the simulations we fixed 1000 trials and 500 bootstrap replications.

For the whole simulation batch the case with $p = 1$, $\beta_0 = -1$ and $\beta_1 = 1$ was considered. The support of the weight function in this case has set to $[L_1, L_2] = [0.43, 0.83]$. We chose $\alpha_0 = 2$ and tried the values $c = 0$ (H_0) and

$c = 1, 2, 3, 4$ (H_1) in the expression $\alpha(x) = \alpha_0 + cx^2$. Visual examination of the logarithm of the conditional hazard functions for this models shows that the choices $c = 1, 2$ are very hard to distinguish from the null hypothesis.

Tables 2-4 collect the percentages of acceptance of the test using the bootstrap critical value for the three models and sample size $n = 200$. Different values of the bandwidths have been tried under the constraint $g = 2h$. This is a reasonable choice since condition C16 implies that g has to be asymptotically larger than h^2 (see also Härdle and Mammen (1993) for some insight about possible choices of g in the regression case). Tables 5-6 show similar results for model 1 and samples sizes: $n = 50$, 100, while Table 7 reports on the sensitivity of the test to the choice of the pilot bandwidth, g, when h is fixed.

h	0.2	0.225	0.25	0.275	0.3
g	0.4	0.45	0.5	0.55	0.6
$c = 0$	75.8	84.7	90.3	94.7	98.0
$c = 1$	69.9	78.5	84.9	91.0	94.9
$c = 2$	48.6	56.9	65.3	72.9	79.9
$c = 3$	31.0	35.1	40.1	44.6	50.2
$c = 4$	19.0	21.6	23.4	26.2	29.8

Table 2. Percentages of acceptance for model 1 with $n = 200$, when $g = 2h$.

h	0.2	0.225	0.25	0.275	0.3
g	0.4	0.45	0.5	0.55	0.6
$c = 0$	71.2	80.7	88.7	94.0	97.3
$c = 1$	63.9	72.8	82.3	89.6	94.9
$c = 2$	43.0	52.6	61.5	69.2	75.6
$c = 3$	24.6	30.3	35.9	41.1	48.2
$c = 4$	15.6	18.4	21.1	23.0	26.4

Table 3. Percentages of acceptance for model 2 with $n = 200$, when $g = 2h$.

h	0.2	0.225	0.25	0.275	0.3
g	0.4	0.45	0.5	0.55	0.6
$c = 0$	71.9	82.5	89.7	94.8	98.2
$c = 1$	59.9	72.6	82.0	88.8	94.7
$c = 2$	40.4	49.8	58.7	67.2	76.3
$c = 3$	26.1	30.9	35.8	41.7	49.1
$c = 4$	16.1	19.4	21.3	25.3	29.4

Table 4. Percentages of acceptance for model 3 with $n = 200$, when $g = 2h$.

h	0.2	0.225	0.25	0.275	0.3
g	0.4	0.45	0.5	0.55	0.6
$c = 0$	64.9	78.2	89.4	95.8	98.7
$c = 1$	45.1	57.3	71.8	85.5	94.4
$c = 2$	27.1	36.9	50.5	63.6	80.8
$c = 3$	18.0	24.6	34.3	47.6	63.3
$c = 4$	16.0	21.6	28.5	38.9	52.6

Table 5. Percentages of acceptance for model 1 with $n = 50$, when $g = 2h$.

h	0.2	0.225	0.25	0.275	0.3
g	0.4	0.45	0.5	0.55	0.6
$c = 0$	70.6	82.5	90.8	95.4	98.6
$c = 1$	55.2	67.9	79.4	87.9	93.5
$c = 2$	32.1	42.2	53.1	65.6	75.1
$c = 3$	17.8	23.1	30.5	39.1	49.4
$c = 4$	10.4	13.5	17.7	22.5	29.2

Table 6. Percentages of acceptance for model 1 with $n = 100$, when $g = 2h$.

g	0.4	0.47	0.55	0.65	0.8
$c = 0$	94.1	94.2	94.7	95.7	96.6
$c = 1$	90.1	90.5	91.0	91.8	92.8
$c = 2$	72.8	72.7	72.9	73.4	77.1
$c = 3$	45.5	44.3	44.6	45.5	48.6
$c = 4$	27.4	26.9	26.2	27.4	29.4

Table 7. Percentages of acceptance for model 1 with $n = 200$ and $h = 0.275$.

The results in Tables 2-7 lead to some conclusions. First of all, the choice of the smoothing parameter, h, seems not to be very critical for the null hypothesis. However it should not be too small. For alternatives that are really close to Cox model one should prevent against taking the bandwidth h too large (see Tables 2-6), specially for small sample size ($n = 50$). All in all, it is clear that there exist some range of bandwidths for which the percentage of acceptation is close to the nominal 95%, for the null hypothesis, and reasonably small (specially when n increases) under the alternatives. The pilot bandwidth, g, used in the bootstrap resampling to approximate the critical value, has a secondary importance in view of Table 7. This is more evident for the null hypothesis, but the results do not fluctuate much for the alternative too. Finally, the results are not much affected by the actual proportion of censoring and truncation (see Tables 2-4).

6 Conditions and proofs

6.1 Conditions

We now state the conditions used in the results of Sections 3 and 4.

Conditions for Theorem 1.

C1 The function J is bounded and differentiable in $(0,1)$ with $J'(u) = 0$ for $u \in (0, \alpha_*) \cup (\beta_*, 1)$, for some $0 < \alpha_* < \beta_* < 1$. Furthermore, the following limit exists: $\lim_{u \to 1^-} J(u)$.

C2 For every point x, in the support of the random variable X,

$$\lim_{y \to \infty} y \left[1 - \left(F\left(y|_x\right)\right)\right] = 0.$$

C3 The covariate X has finite moments of order $2p$ and $n^{-1}\chi^t\chi \longrightarrow A$, almost surely, for some nonsingular matrix A.

C4 X, Y, T and C are absolutely continuous random variables.

C5 Let M^* be the conditional distribution of X given $T \leq Z$ and m^* its pertaining density. Let $I = [x_1, x_2]$ be an interval contained in the support of m^* such that

$$0 < \gamma = \inf_{x \in I_\delta} m^*(x) < \sup_{x \in I_\delta} m^*(x) = \Gamma < \infty,$$

for some $I_\delta = [x_1 - \delta, x_2 + \delta]$ with $\delta > 0$ and $0 < \delta\Gamma < 1$. The random variables Y, T and C are conditionally independent given $X = x$, for every $x \in I$ and $a_{L(\bullet|_x)} \leq a_{H(\bullet|_x)}$ and $b_{L(\bullet|_x)} \leq b_{H(\bullet|_x)}$, for every $x \in I_\delta$, where a_G and b_G are the left and right endpoints of the support of the distribution G. Finally, there exist some real numbers $a < b$ satisfying

$$\begin{aligned} \inf_{x \in I_\delta} \alpha(x)^{-1} \left(1 - H\left(b|_x\right)\right) L\left(a|_x\right) &\geq \theta > 0, \\ \bigcup_{x \in I} \left(F^{-1}\left(\frac{\alpha_*}{2}\Big|_x\right), F^{-1}\left(\frac{\beta_* + 1}{2}\Big|_x\right)\right) &\subset [a, b]. \end{aligned}$$

C6 The density of X, $m(x)$, and the function $\alpha(x)$ are twice continuously differentiable in I_δ. The functions $L(y|_x)$, $H(y|_x)$ and $H_1(y|_x)$ are twice continuously differentiable, with respect to x, and their first two derivatives are bounded for $(x,y) \in I_\delta \times [0,\infty)$.

C7 The functions $L(y|_x)$, $H(y|_x)$ and $H_1(y|_x)$ are twice continuously differentiable, with respect to y, for $(x,y) \in I_\delta \times [a,b]$.

C8 The partial second derivatives of the functions $L(y|_x)$, $H(y|_x)$ and $H_1(y|_x)$, with respect to x and y, exist and are continuous for $(x,y) \in I_\delta \times [a,b]$.

C9 The densities (or subdensity) pertaining to $L(y)$, $H(y)$ and $H_1(y)$ are bounded away from zero in $[a,b]$.

C10 The kernel function, K, is a bounded variation symmetric density with support contained in $(-1,1)$.

C11 Let's define

$$\eta(Z,T,\delta,x) = \int (1 - F(y|_x))\, \xi(Z,T,\delta,x,y)\, J(F(y|_x))\, dy,$$

with

$$\xi(Z,T,\delta,x,y) = \frac{1_{\{Z \le y, \delta=1\}}}{C(Z|_x)} - \int_0^y \frac{1_{\{T \le u \le Z\}}}{C^2(u|_x)} dH_1^*(u|_x). \qquad (15)$$

Then, the functions

$$\begin{aligned} \sigma^2(x) &= Var\left(\eta(Z,T,\delta,x)|_{X=x}\right) \\ \mu_4(x) &= E\left(\eta(Z,T,\delta,x)^4\big|_{X=x}\right) \end{aligned}$$

satisfy

$$\int \frac{\sigma^4(x)}{m^*(x)} dx < \infty \text{ and } \int \frac{\mu_4(x)}{m^*(x)} dx < \infty.$$

C12 The smoothing parameter fulfills the following limit conditions: $h \to 0$, $\frac{nh^3}{(\ln n)^3} \to \infty$ and $nh^4 \to 0$.

C13 The first and second derivatives of $C\left(y|_x\right)$ and $h_1^*\left(y|_x\right)$ with respect to x, $\dot{C}\left(y|_x\right)$, $\ddot{C}\left(y|_x\right)$, $\dot{h}_1^*\left(y|_x\right)$ and $\ddot{h}_1^*\left(y|_x\right)$ are bounded (in both variables x and y), $m^{*\prime}$ and $m^{*\prime\prime}$ are bounded and integrable with

$$\int\int\int\left|\nu\left(x,y\right)\nu\left(x,z\right)\varphi_1\left(x,y\right)\varphi_2\left(x,z\right)\frac{m^{*(i)}\left(x\right)m^{*(j)}\left(x\right)}{m^*\left(x\right)}dxdydz\right|<\infty,$$

for every $i,j=0,1,2$ and where $\nu\left(x,y\right)=\left(1-F\left(y|_x\right)\right)J\left(F\left(y|_x\right)\right)$ and $\varphi_1\left(x,y\right)$ and $\varphi_2\left(x,y\right)$ are any of the functions

$$\int_0^y\frac{\dot{h}_1^*\left(s|_x\right)}{C\left(s|_x\right)^2}ds,\ \int_0^y\frac{\ddot{h}_1^*\left(s|_x\right)}{C\left(s|_x\right)^2}ds,$$
$$\int_0^y\frac{\dot{C}\left(s|_x\right)}{C\left(s|_x\right)^2}dH_1^*\left(s|_x\right)\ \text{and}\ \int_0^y\frac{\ddot{C}\left(s|_x\right)}{C\left(s|_x\right)^2}dH_1^*\left(s|_x\right).$$

C14 The function

$$\begin{aligned}\varphi\left(x_1,x_2,x_3,y_1,y_2\right)=&\int_0^{y_1\wedge y_2}\frac{dH_1^*\left(s|_{x_3}\right)}{C\left(s|_{x_1}\right)C\left(s|_{x_2}\right)}\\&-\int_0^{y_2}\int_s^{y_1}\frac{dH_1^*\left(z|_{x_3}\right)}{C\left(z|_{x_1}\right)}\frac{L^*\left(s|_{x_3}\right)dH_1^*\left(s|_{x_2}\right)}{C\left(s|_{x_2}\right)^2}\\&-\int_0^{y_1}\int_s^{y_2}\frac{dH_1^*\left(z|_{x_3}\right)}{C\left(z|_{x_2}\right)}\frac{L^*\left(s|_{x_3}\right)dH_1^*\left(s|_{x_1}\right)}{C\left(s|_{x_1}\right)^2}\\+\int_0^{y_1}\int_0^{y_2}&\frac{L^*\left(s_1\wedge s_2|_{x_3}\right)\left(1-H_1^*\left(s_1\vee s_2|_{x_3}\right)\right)}{C\left(s_1|_{x_1}\right)^2C\left(s_2|_{x_2}\right)^2}dH_1^*\left(s_2|_{x_2}\right)dH_1^*\left(s_1|_{x_1}\right)\end{aligned}$$

is twice differentiable with respect to its first three variables and satisfies

$$\int\int\int\frac{\nu\left(x_1,y_1\right)\nu\left(x_2,y_2\right)}{m^*\left(x_1\right)}$$
$$\times\left|\frac{\partial^2}{\partial x_2^2}\left[\varphi\left(x_1,x_1,x_2,y_1,y_2\right)m^*\left(x_2\right)\right]\right|_{x_2=x_1}\Bigg|dy_1dy_2dx_1<\infty,$$

$$\int\int\int\frac{\nu\left(x_1,y_1\right)}{m^*\left(x_1\right)}$$
$$\times\left|\frac{\partial^2}{\partial x_2^2}\left[\nu\left(x_2,y_2\right)\varphi\left(x_1,x_2,x_2,y_1,y_2\right)m^*\left(x_2\right)\right]\right|_{x_2=x_1}\Bigg|dy_1dy_2dx_1<\infty,$$

$$\int\int\int \frac{\nu(x_1, y_2)}{m^*(x_1)}$$
$$\times \left| \frac{\partial^2}{\partial x_2^2} \left[\nu(x_2, y_1) \varphi(x_2, x_1, x_2, y_1, y_2) m^*(x_2) \right] \Big|_{x_2 = x_1} \right| dy_1 dy_2 dx_1 < \infty \text{ and}$$

$$\int\int\int \frac{1}{m^*(x_1)}$$
$$\times \left| \frac{\partial^2}{\partial x_2^2} \left[\nu(x_2, y_1) \nu(x_2, y_2) \varphi(x_2, x_2, x_2, y_1, y_2) m^*(x_2) \right] \Big|_{x_2 = x_1} \right| dy_1 dy_2 dx_1 < \infty$$

Conditions for Theorem 3.

C1' The distribution function W has a bounded density w. As a consequence, the derivative of $\tilde{W}$, $\tilde{w}$, is also bounded.

C11' Condition C11 holds when redefining

$$\eta(Z, T, \delta, x) = \begin{cases} \eta_{PH}(Z, T, \delta, x) = \int \xi(Z, T, \delta, x, y) \frac{dW(y)}{\Lambda(y|_x)} & \text{for model PH} \\ \eta_{AR}(Z, T, \delta, x) = \int \xi(Z, T, \delta, x, y) d\tilde{W}(y) & \text{for model AR} \\ \eta_{PO}(Z, T, \delta, x) = \int \xi(Z, T, \delta, x, y) \frac{dW(y)}{F(y|_x)} & \text{for model PO} \end{cases} \tag{16}$$

with $\xi(Z, T, \delta, x, y)$ as in (15).

C13' Let us change the definition of the function ν:

$$\nu(x, y) = \begin{cases} \nu_{PH}(x, y) = \frac{w(y)}{\Lambda(y|_x)} & \text{for model PH} \\ \nu_{AR}(x, y) = \tilde{w}(y) & \text{for model AR} \\ \nu_{PO}(x, y) = \frac{w(y)}{F(y|_x)} & \text{for model PO} \end{cases}$$

Then condition C13 is satisfied for this function.

C14' Condition C14 holds for the new definition of $\nu(x, y)$.

Conditions for Theorem 9.

C15 Let us define the function $\mu_1(x) = E\left(|\eta(Z,T,\delta,x)||_{X=x}\right)$. Then

$$\begin{aligned}\int\int (1-F(y|_x))\left(\int_0^y \frac{h_1^*(u|_x)}{C(u|_x)}du\right) J(F(y|_x))\,dydx &< \infty,\\ \int\int \frac{(1-F(y|_x))}{m^*(x)}\left(\int_0^y \frac{h_1^*(u|_x)}{C(u|_x)}du\right) J(F(y|_x))\,dydx &< \infty,\end{aligned}$$

$$\begin{aligned}\int \frac{\sigma^2(x)}{m^*(x)^\ell}dx &< \infty,\ \int \frac{\mu_1(x)^2}{m^*(x)^\ell}dx < \infty,\\ \int \frac{\mu_1(x)}{m^*(x)^\ell}dx &< \infty,\ \int \frac{\mu_1(x)\sigma^2(x)}{m^*(x)^\ell}dx < \infty,\end{aligned}$$

for $\ell = 0,1,2$ and

$$\int \frac{\mu_4(x)\sigma^4(x)}{m^*(x)^\ell}dx < \infty,$$

for $\ell = 1,2,3,4,5$.

C16 The pilot bandwidth g satisfies $g \to 0$, $ng \to \infty$ and $\frac{g}{h^2} \to \infty$.

Condition for Theorem 10.

C15' C15 holds for the definition of $\eta(Z,T,\delta,x)$ given in C11'.

6.2 Proofs

We now proceed with the proofs of the main results.

Proof of Theorem 1: Using (8),

$$\hat{\psi}_n\left(\hat{\boldsymbol{\beta}}\right) = \frac{1}{n}\sum_{r=1}^{n}\left(\hat{m}_r - m_r + \Delta_r\right)^2 = \hat{\psi}_n^{(1)}\left(\hat{\boldsymbol{\beta}}\right) + \Delta^{(1)} + \Delta^{(2)}, \qquad (17)$$

with $\hat{\psi}_n^{(1)}\left(\hat{\boldsymbol{\beta}}\right) = \frac{1}{n}\sum_{r=1}^{n}\left(\hat{m}_r - m_r\right)^2$, $\Delta_r = m_r - \left(\hat{\beta}_0 + \hat{\beta}_1 X_r + \cdots + \hat{\beta}_p X_r^p\right)$, $\Delta^{(1)} = \frac{1}{n}\sum_{r=1}^{n}\Delta_r^2$ and $\Delta^{(2)} = 2\frac{1}{n}\sum_{r=1}^{n}\left(\hat{m}_r - m_r\right)\Delta_r$.

Using Lemma 2 it is straightforward to prove that $\Delta_r = O_P\left(n^{-\frac{1}{2}}\right)$,uniformly in $r = 1, 2, \ldots, n$. As a consequence

$$\Delta^{(1)} = O_P\left(n^{-1}\right) \tag{18}$$

$$\Delta^{(2)} = 2\left(\frac{1}{n}\sum_{r=1}^{n}\left(\hat{m}_r - m_r\right), \ldots, \frac{1}{n}\sum_{r=1}^{n}\left(\hat{m}_r - m_r\right) X_r^p\right)\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right), \tag{19}$$

where the term (19) will be studied much later.

Let $\mathbb{J}(t) = \int_0^t J(u)\,du$, then, using condition C2, partial integration gives

$$\hat{m}_r - m_r = \int_0^\infty \left[\mathbb{J}\left(F\left(y|_{X_r}\right)\right) - \mathbb{J}\left(\hat{F}_n\left(y|_{X_r}\right)\right)\right] dy. \tag{20}$$

Conditions C4-C10 and expression (15) warranty that part (c) of Theorem 2 in Iglesias-Pérez and González-Manteiga (1999) can be applied to obtain

$$\begin{aligned} &\hat{F}_n\left(y|_{X_r}\right) - F\left(y|_{X_r}\right) \\ = \;& \left(1 - F\left(y|_{X_r}\right)\right)\sum_{i=1}^{n} B_{ni}\left(X_r\right)\xi\left(Z_i, T_i, \delta_i, X_r, y\right) + R_n'\left(y|_{X_r}\right), \end{aligned} \tag{21}$$

where

$$\sup_{y\in[a,b], r=1,2,\ldots,n}\left|R_n'\left(y|_{X_r}\right)\right| = O\left(\left(\frac{\ln n}{nh}\right)^{\frac{3}{4}}\right) \text{ a.s.}$$

Starting from these results, some extra work can be done to prove that

$$\sup_{y\in[a,b], x\in I}\left|\hat{F}_n\left(y|_x\right) - F\left(y|_x\right)\right| = O\left(\left(\frac{\ln n}{nh}\right)^{\frac{1}{2}}\right) \text{ a.s.} \tag{22}$$

Using (22) and conditions C1, C5 and C12 it is easy to verify that the inequality

$$\begin{aligned} &\left|\int_0^\infty \left[\mathbb{J}\left(F\left(y|_{X_r}\right)\right) - \mathbb{J}\left(\hat{F}_n\left(y|_{X_r}\right)\right)\right] dy \right. \\ &\left. - \int_0^\infty \left[F\left(y|_{X_r}\right) - \hat{F}_n\left(y|_{X_r}\right)\right] J\left(F\left(y|_{X_r}\right)\right) dy\right| = O_P\left(\frac{\ln n}{nh}\right) = o_P\left(n^{-\frac{1}{2}}\right) \end{aligned} \tag{23}$$

holds uniformly in $r = 1, 2, \ldots, n$.

Now, the leading term of equation (17) can be easily handled using (20), (23) and Cauchy-Schwarz inequality:

$$\begin{aligned}\hat{\psi}_n^{(1)}\left(\hat{\boldsymbol{\beta}}\right) &= \frac{1}{n}\sum_{r=1}^{n}\left(\int\left[\mathbb{J}\left(F\left(y|_{X_r}\right)\right)-\mathbb{J}\left(\hat{F}_n\left(y|_{X_r}\right)\right)\right]dy\right)^2 \\ &= \hat{\psi}_n^{(2)}\left(\hat{\boldsymbol{\beta}}\right)+\Delta^{(3)}+\Delta^{(4)}, \end{aligned} \tag{24}$$

with

$$\begin{aligned}\hat{\psi}_n^{(2)}\left(\hat{\boldsymbol{\beta}}\right) &= \frac{1}{n}\sum_{r=1}^{n}\left(\int\left[F\left(y|_{X_r}\right)-\hat{F}_n\left(y|_{X_r}\right)\right]J\left(F\left(y|_{X_r}\right)\right)dy\right)^2, \\ \Delta^{(3)} &= o_P\left(n^{-1}\right) \text{ and} \\ \Delta^{(4)} &= o_P\left(\left(\hat{\psi}_n^{(2)}\left(\hat{\boldsymbol{\beta}}\right)\right)^{\frac{1}{2}}n^{-\frac{1}{2}}\right). \end{aligned} \tag{25}$$

Now (25) and (21) imply that

$$\hat{\psi}_n^{(2)}\left(\hat{\boldsymbol{\beta}}\right)=\hat{\psi}_n^{(3)}\left(\hat{\boldsymbol{\beta}}\right)+\Delta^{(5)}+\Delta^{(6)}, \tag{26}$$

with

$$\begin{aligned}\hat{\psi}_n^{(3)}\left(\hat{\boldsymbol{\beta}}\right) &= \frac{1}{n}\sum_{r=1}^{n}\left(\int\left(1-F\left(y|_{X_r}\right)\right)\right. \\ &\quad \left.\sum_{i=1}^{n}B_{ni}\left(X_r\right)\xi\left(Z_i,T_i,\delta_i,X_r,y\right)J\left(F\left(y|_{X_r}\right)\right)dy\right)^2, \end{aligned} \tag{27}$$

$$\Delta^{(5)} = O_P\left(\left(\frac{\ln n}{nh}\right)^{\frac{3}{2}}\right) \text{ and} \tag{28}$$

$$\Delta^{(6)} = O_P\left(\left(\hat{\psi}_n^{(3)}\left(\hat{\boldsymbol{\beta}}\right)\right)^{\frac{1}{2}}\left(\frac{\ln n}{nh}\right)^{\frac{3}{4}}\right), \tag{29}$$

where, for $i=1,2,\ldots,n$, $B_{ni}\left(x\right)=K\left(\frac{x-X_i}{h}\right)\left[\sum_{j=1}^{n}K\left(\frac{x-X_j}{h}\right)\right]^{-1}$ are the Nadaraya-Watson weights. The term (27) may be expanded in the following

way

$$\begin{aligned}\hat{\psi}_n^{(3)}\left(\hat{\boldsymbol{\beta}}\right) &= \frac{1}{n}\sum_{r=1}^{n}\left(\sum_{i=1}^{n}B_{ni}\left(X_r\right)\varepsilon_i+\Delta_r^{(7)}\right)^2 \\ &= \frac{1}{n}\sum_{r=1}^{n}\left(\sum_{i=1}^{n}B_{ni}\left(X_r\right)\varepsilon_i\right)^2+\frac{1}{n}\sum_{r=1}^{n}\left(\Delta_r^{(7)}\right)^2 \\ &\quad +2\frac{1}{n}\sum_{r=1}^{n}\sum_{i=1}^{n}B_{ni}\left(X_r\right)\varepsilon_i\Delta_r^{(7)},\end{aligned} \tag{30}$$

with

$$\varepsilon_i = \int\left(1-F\left(y|_{X_i}\right)\right)\xi\left(Z_i,T_i,\delta_i,X_i,y\right)J\left(F\left(y|_{X_i}\right)\right)dy \text{ and} \tag{31}$$

$$\begin{aligned}\Delta_r^{(7)} &= \sum_{i=1}^{n}B_{ni}\left(X_r\right) \\ &\quad \left[\int\left(1-F\left(y|_{X_r}\right)\right)\xi\left(Z_i,T_i,\delta_i,X_r,y\right)J\left(F\left(y|_{X_r}\right)\right)dy-\right. \\ &\quad \left.\int\left(1-F\left(y|_{X_i}\right)\right)\xi\left(Z_i,T_i,\delta_i,X_i,y\right)J\left(F\left(y|_{X_i}\right)\right)dy\right].\end{aligned}$$

By defining $b_{nij} = \frac{1}{n}\sum_{r=1}^{n}B_{ni}\left(X_r\right)B_{nj}\left(X_r\right)$, the first term in the right handside of (30) can be decomposed into two new terms:

$$\frac{1}{n}\sum_{r=1}^{n}\left(\sum_{i=1}^{n}B_{ni}\left(X_r\right)\varepsilon_i\right)^2 = \Delta^{(8)}+\Delta^{(9)}, \tag{32}$$

$$\text{with } \Delta^{(8)} = \sum_{i=1}^{n}b_{nii}\varepsilon_i^2 \text{ and } \Delta^{(9)} = 2\sum_{i<j}b_{nij}\varepsilon_i\varepsilon_j$$

which can be dealt with in the same way as the term Δ_{121} and Δ_{122} in the proof of Theorem 2.1 in González-Manteiga and Cao (1993). Indeed,

$$\begin{aligned}\Delta^{(8)} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{n}B_{ni}\left(X_r\right)^2\varepsilon_i^2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{n}\left(\frac{K\left(\frac{X_r-X_i}{h}\right)}{\sum_{j=1}^{n}K\left(\frac{X_r-X_j}{h}\right)}\right)^2\varepsilon_i^2 \\ &= \frac{1}{n^3h^2}\sum_{r=1}^{n}\frac{1}{\hat{m}^*\left(X_r\right)^2}\sum_{i=1}^{n}K\left(\frac{X_r-X_i}{h}\right)^2\varepsilon_i^2 \\ &= \Delta_1^{(8)}+\Delta_2^{(8)}\end{aligned} \tag{33}$$

where $\hat{m}^*(x) = \frac{1}{nh}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)$ is the Parzen-Rosenblatt estimator of $m(x)$ and the terms $\Delta_1^{(8)}$ and $\Delta_2^{(8)}$ are given by

$$\begin{aligned}\Delta_1^{(8)} &= \frac{1}{n^3h^2}\sum_{r=1}^n \frac{1}{m^*(X_r)^2}\sum_{i=1}^n K\left(\frac{X_r-X_i}{h}\right)^2\varepsilon_i^2 \\ \Delta_2^{(8)} &= \frac{1}{n^3h^2}\sum_{r=1}^n \frac{m^*(X_r)^2-\hat{m}^*(X_r)^2}{\hat{m}^*(X_r)^2 m^*(X_r)^2}\sum_{i=1}^n K\left(\frac{X_r-X_i}{h}\right)^2\varepsilon_i^2.\end{aligned}$$

Standard arguments, condition C12 and Theorem B Silverman (1978) imply

$$\begin{aligned}\left|\Delta_2^{(8)}\right| &\leq \frac{\Gamma(1+D)D}{\gamma^2}\Delta^{(8)} \qquad (34)\\ \text{where } D &= \sup_{x\in\mathbb{R}}|m^*(x)-\hat{m}^*(x)| = O_P\left(\left(\frac{\ln h^{-1}}{nh}\right)^{\frac{1}{2}}+h^2\right).\end{aligned}$$

Standard mean and variance calculations for $\Delta_1^{(8)}$, Taylor expansions and condition C11 lead to

$$\Delta_1^{(8)} = \frac{1}{nh}\left(\int\sigma^2(x)\,dx\right)K^{(2)}(0)+O_P\left(n^{-2}h^{-2}+n^{-1}h+n^{-\frac{3}{2}}h^{-1}\right). \qquad (35)$$

We now use (33), (34) and (35) to conclude

$$\begin{aligned}\Delta^{(8)} &= \frac{1}{nh}\left(\int\sigma^2(x)\,dx\right)K^{(2)}(0) \qquad (36)\\ &+O_P\left(n^{-2}h^{-2}+n^{-1}h+n^{-\frac{3}{2}}h^{-1}+n^{-\frac{3}{2}}h^{-\frac{3}{2}}\left(\ln h^{-1}\right)^{\frac{1}{2}}\right)\end{aligned}$$

The term $\Delta^{(9)}$ can be approximated by a somewhat simpler expression

$$\tilde{\Delta}^{(9)} = 2\sum_{i<j}\tilde{b}_{nij}\varepsilon_i\varepsilon_j, \qquad (37)$$

where $\tilde{b}_{nij} = \frac{1}{n}\sum_{r=1}^n \tilde{B}_{ni}(X_r)\tilde{B}_{nj}(X_r)$ and $\tilde{B}_{ni}(X_r) = \frac{K\left(\frac{X_r-X_i}{h}\right)}{nh\cdot m^*(X_r)}$. Straight

forward calculations give

$$\begin{aligned}
\Delta^{(9)} &= \tilde{\Delta}^{(9)} + \Delta_1^{(9)} + \Delta_2^{(9)} + \Delta_3^{(9)} \text{ with} \\
\Delta_1^{(9)} &= \frac{2}{n}\sum_{i<j}\sum_{r=1}^{n} \tilde{B}_{ni}(X_r) B_{nj}(X_r) \frac{m^*(X_r) - \hat{m}^*(X_r)}{m^*(X_r)} \varepsilon_i \varepsilon_j, \qquad (38) \\
\Delta_2^{(9)} &= \frac{2}{n}\sum_{i<j}\sum_{r=1}^{n} B_{ni}(X_r) \tilde{B}_{nj}(X_r) \frac{m^*(X_r) - \hat{m}^*(X_r)}{m^*(X_r)} \varepsilon_i \varepsilon_j, \qquad (39) \\
\Delta_2^{(9)} &= \frac{2}{n}\sum_{i<j}\sum_{r=1}^{n} B_{ni}(X_r) B_{nj}(X_r) \frac{(m^*(X_r) - \hat{m}^*(X_r))^2}{m^*(X_r)^2} \varepsilon_i \varepsilon_j. \qquad (40)
\end{aligned}$$

Expressions (38), (39) and (40) can be proved to be $o_P\left(n^{-1}h^{-\frac{1}{2}}\right)$ by using Cauchy-Schwarz inequality, the fact that $D = O_P\left(\left(\frac{\ln h^{-1}}{nh}\right)^{\frac{1}{2}} + h^2\right)$ and variance calculations similar to what will be done next for $\tilde{\Delta}^{(9)}$.

A central limit theorem for quadratic forms (Theorem (2.1) in De Jong (1987)) can be applied to conclude that $\tilde{\Delta}^{(9)}$ in (37) is asymptotically normal with zero mean. The conditions on that theorem can be checked in a completely parallel way to the proof of Theorem 2.1 in González-Manteiga and Cao (1993). For this reason we skip it.

Standard U-statistic calculations, changes of variable, Taylor expansions and condition C12 can be used to obtain an asymptotic expression for the variance of $\tilde{\Delta}^{(9)}$:

$$Var\left(\tilde{\Delta}^{(9)}\right) = \frac{2}{n^2 h}\left(\int \sigma^4(x)\,dx\right) K^{(4)}(0) + o\left(n^{-2}h^{-1}\right),$$

which implies $\sqrt{n^2 h}\tilde{\Delta}^{(9)} \xrightarrow{\mathrm{d}} N(0, V)$. Now using (32), (36), condition C12 and the fact that (38), (39) and (40) are $o_P\left(n^{-1}h^{-\frac{1}{2}}\right)$ we get

$$\sqrt{n^2 h}\left[\frac{1}{n}\sum_{r=1}^{n}\left(\sum_{i=1}^{n} B_{ni}(X_r)\varepsilon_i\right)^2 - \frac{1}{nh}\left(\int \sigma^2(x)\,dx\right) K^{(2)}(0)\right] \xrightarrow{\mathrm{d}} N(0, V). \qquad (41)$$

As done for the term $\Delta^{(9)}$ we may get rid off the random denominator in $B_{ni}(X_r)$ for the second term in the right handside of (30):

$$\frac{1}{n}\sum_{r=1}^{n}\left(\Delta_r^{(7)}\right)^2 = \left(1+o_P\left(1\right)\right)\frac{1}{n}\sum_{r=1}^{n}\sum_{i,j=1}^{n}\tilde{\Delta}_{rij}^{(7)} \tag{42}$$

with

$$\begin{aligned}\tilde{\Delta}_{rij}^{(7)} &= \frac{K\left(\frac{X_r-X_i}{h}\right)K\left(\frac{X_r-X_j}{h}\right)}{n^2h^2m^*\left(X_r\right)^2}\\ &\int\int\left[\left(1-F\left(y_1|_{X_r}\right)\right)\xi\left(Z_i,T_i,\delta_i,X_r,y_1\right)J\left(F\left(y_1|_{X_r}\right)\right)-\right.\\ &\left(1-F\left(y_1|_{X_i}\right)\right)\xi\left(Z_i,T_i,\delta_i,X_i,y_1\right)J\left(F\left(y_1|_{X_i}\right)\right)]\\ &\times\left[\left(1-F\left(y_2|_{X_r}\right)\right)\xi\left(Z_j,T_j,\delta_j,X_r,y_2\right)J\left(F\left(y_2|_{X_r}\right)\right)-\right.\\ &\left.\left(1-F\left(y_2|_{X_j}\right)\right)\xi\left(Z_j,T_j,\delta_j,X_j,y_2\right)J\left(F\left(y_2|_{Xj}\right)\right)\right]dy_1dy_2.\end{aligned}$$

It is clear that $\tilde{\Delta}_{rij}^{(7)}=0$ when $r=i$ or $r=j$. On the other hand, using (15), we have

$$E\left(\xi\left(Z,T,\delta,x,y\right)|_{X=u}\right)=\int_0^y\frac{dH_1^*\left(s|_u\right)}{C\left(s|_x\right)}-\int_0^y\frac{C\left(s|_u\right)}{C\left(s|_x\right)^2}dH_1^*\left(s|_x\right)$$

and

$$E\left(\xi\left(Z,T,\delta,x_1,y_1\right)\xi\left(Z,T,\delta,x_2,y_2\right)|_{X=x_3}\right)=\varphi\left(x_1,x_2,x_3,y_1,y_2\right)$$

with φ defined in condition C14. It is straight forward but long and tedious to compute the order of $E\left(\tilde{\Delta}_{rij}^{(7)}\right)$. Standard arguments as changes of variables and Taylor expansions, as well as conditions C13 and C14 lead to

$$\begin{aligned}E\left(\tilde{\Delta}_{rij}^{(7)}\right) &= 0, \text{ if } r=i \text{ or } r=j,\\ E\left(\tilde{\Delta}_{rij}^{(7)}\right) &= E\left(\tilde{\Delta}_{123}^{(7)}\right)=O\left(\frac{h^4}{n^2}\right) \text{ if the indices } i,\ j \text{ and } r \text{ are different},\\ E\left(\tilde{\Delta}_{rij}^{(7)}\right) &= E\left(\tilde{\Delta}_{122}^{(7)}\right)=O\left(\frac{h}{n^2}\right) \text{ if } i=j\neq r.\end{aligned}$$

These results and (42) imply

$$\frac{1}{n}\sum_{r=1}^{n}\left(\Delta_r^{(7)}\right)^2=O_P\left(h^4+\frac{h}{n}\right). \tag{43}$$

Using (41) and (43), the last term in (30) may be bounded by means of Cauchy-Schwarz inequality

$$\left| 2\frac{1}{n}\sum_{r=1}^{n}\sum_{i=1}^{n} B_{ni}\left(X_r\right)\varepsilon_i\Delta_r^{(7)} \right| = O_P\left(n^{-\frac{1}{2}}h^{\frac{3}{2}} + n^{-1}\right).$$

This expression, as well as condition C12 and (43), (41) and (30) imply that

$$\sqrt{n^2h}\hat{\psi}_n^{(3)}\left(\hat{\boldsymbol{\beta}}\right) - h^{-\frac{1}{2}}K^{(2)}\left(0\right)\int \sigma^2\left(x\right)dx \xrightarrow{\mathrm{d}} N\left(0,V\right). \tag{44}$$

As a consequence, $\hat{\psi}_n^{(3)}\left(\hat{\boldsymbol{\beta}}\right) = O_P\left(n^{-1}h^{-1}\right)$ and, using (29) and condition C11, we conclude

$$\Delta^{(6)} = O_P\left((nh)^{-\frac{5}{4}}\left(\ln n\right)^{\frac{3}{4}}\right) = O_p\left(n^{-1}h^{-\frac{1}{2}}\right).$$

This rate and those in (28) and (29) apply to (26) and (44) to directly obtain

$$\sqrt{n^2h}\hat{\psi}_n^{(2)}\left(\hat{\boldsymbol{\beta}}\right) - h^{-\frac{1}{2}}K^{(2)}\left(0\right)\int \sigma^2\left(x\right)dx \xrightarrow{\mathrm{d}} N\left(0,V\right). \tag{45}$$

Now (25) imply that $\Delta^{(4)} = o_P\left(n^{-1}h^{-\frac{1}{2}}\right)$ and, consequently, (24) and (45) can be used to derive

$$\sqrt{n^2h}\hat{\psi}_n^{(1)}\left(\hat{\boldsymbol{\beta}}\right) - h^{-\frac{1}{2}}K^{(2)}\left(0\right)\int \sigma^2\left(x\right)dx \xrightarrow{\mathrm{d}} N\left(0,V\right). \tag{46}$$

In order to derive the asymptotic distribution of $\hat{\psi}_n\left(\hat{\boldsymbol{\beta}}\right)$ using (17), we need to study the term $\Delta^{(2)}$ in (19) which may be treated as done in (24)-(40). As a consequence

$$\begin{aligned}\frac{1}{n}\sum_{r=1}^{n}\left(\hat{m}_r - m_r\right)X_r^k &= \frac{1}{n}\sum_{r=1}^{n} X_r^k\int\left[\mathbb{J}\left(F\left(y|_{X_r}\right)\right) - \mathbb{J}\left(\hat{F}_n\left(y|_{X_r}\right)\right)\right]dy \\ &= \Delta_k^{(10)} + \Delta_k^{(11)} + O_P\left(\left(\frac{\ln n}{nh}\right)^{\frac{3}{4}}\right) + o_P\left(n^{-\frac{1}{2}}h^{-\frac{1}{2}}\right),\end{aligned} \tag{47}$$

with $\Delta_k^{(10)} = \frac{1}{n}\sum_{r=1}^{n} X_r^j\sum_{i=1}^{n}\tilde{B}_{ni}\left(X_r\right)\varepsilon_i$ and $\Delta_k^{(11)} = \frac{1}{n}\sum_{r=1}^{n} X_r^j\Delta_r^{(7)}$.

Using standard expectation and variance calculations and Taylor expansions, it is straightforward to prove that

$$\begin{aligned}\Delta_k^{(10)} &= O_P\left(n^{-\frac{1}{2}}+n^{-1}h^{\frac{1}{2}}\right) \text{ and}\\ \Delta_k^{(11)} &= O_P\left(h^2+n^{-\frac{1}{2}}h^2+n^{-1}h^{\frac{1}{2}}\right).\end{aligned}$$

Now (19), (47) and Lemma 2 lead to

$$\Delta^{(2)} = O_P\left(n^{-1}+n^{-\frac{3}{2}}h^{\frac{1}{2}}+n^{-\frac{1}{2}}h^2+n^{-\frac{5}{4}}h^{-\frac{3}{4}}\left(\ln n\right)^{\frac{3}{4}}\right)+o_P\left(n^{-1}h^{-\frac{1}{2}}\right).$$

and the proof concludes by using expressions (17), (18) and (46).■

Proof of Lemma 2: We will follow the lines of Theorem 2.1 in Akritas (1996). First of all we write $\hat{\boldsymbol{\beta}}-\boldsymbol{\beta} = \hat{A}^{-1}\hat{b}$ with $\hat{A} = n^{-1}\chi^t\chi$ and $\hat{b} = n^{-1}\chi^t\left(\hat{m}-m\right)$. Using (20), (23), condition C12, the representation (21) and standard arguments, the k-th coordinate of the vector $\hat{b}$ is

$$\begin{aligned}\hat{b}_k &= n^{-1}\sum_{i=1}^{n}X_i^{k-1}\int\left[\mathbb{J}\left(F\left(y|_{X_i}\right)\right)-\mathbb{J}\left(\hat{F}_n\left(y|_{X_i}\right)\right)\right]dy\\ &= \Delta^{(12)}+\Delta^{(13)}+o_P\left(n^{-\frac{1}{2}}\right) \qquad (48)\end{aligned}$$

with

$$\begin{aligned}\Delta^{(12)} &= -n^{-2}h^{-1}K\left(0\right)\sum_{i=1}^{n}X_i^{k-1}\\ &\quad\times\int\frac{\left(1-F\left(y|_{X_i}\right)\right)}{m^*\left(X_i\right)}\xi\left(Z_i,T_i,\delta_i,X_i,y\right)J\left(F\left(y|_{X_i}\right)\right)dy,\\ \Delta^{(13)} &= -n^{-2}h^{-1}\sum_{i,j=1,i\neq j}^{n}X_i^{k-1}\frac{K\left(\frac{X_i-X_j}{h}\right)}{m^*\left(X_i\right)}\\ &\quad\times\int\left[\left(1-F\left(y|_{X_i}\right)\right)\xi\left(Z_j,T_j,\delta_j,X_i,y\right)\right]J\left(F\left(y|_{X_i}\right)\right)dy.\end{aligned}$$

Usual mean nad variance calculations lead to

$$\Delta^{(12)} = O_P\left(n^{-\frac{3}{2}}h^{-1}\right) \text{ and } \Delta^{(13)} = O_P\left(h^2+n^{-1}h^{-\frac{1}{2}}\right). \qquad (49)$$

Now (48) and (49) imply that $\hat{b} = O_P\left(n^{-\frac{1}{2}}\right)$. Condition C3 can be used to control the limit behaviour of $\hat{A}$ and hence to conclude the proof.■

Sketch of the proof of Theorem 3:

Direct inspection enables to reach similar expressions to (24) and (25), but replacing the integral $\int \left[F\left(y|_{X_r}\right) - \hat{F}_n\left(y|_{X_r}\right)\right] J\left(F\left(y|_{X_r}\right)\right) dy$ with

$$\begin{aligned} & \int \left[\ln \hat{\Lambda}_n\left(y|_{X_r}\right) - \ln \Lambda\left(y|_{X_r}\right)\right] dW\left(y\right) \\ \simeq & \sum_{i=1}^{n} B_{ni}\left(X_r\right) \int \xi\left(Z_i, T_i, \delta_i, X_r, y\right) \frac{w\left(y\right)}{\Lambda\left(y|_{X_r}\right)} dy, \end{aligned}$$

for model PH,

$$\begin{aligned} & \int \left[\hat{\Lambda}_n\left(y|_{X_r}\right) - \Lambda\left(y|_{X_r}\right)\right] dW\left(y\right) \\ \simeq & \sum_{i=1}^{n} B_{ni}\left(X_r\right) \int \xi\left(Z_i, T_i, \delta_i, X_r, y\right) \tilde{w}\left(y\right) dy, \end{aligned}$$

for model AR and

$$\begin{aligned} & \int \left\{\text{logit}\left[1 - \exp\left(\hat{\Lambda}_n\left(y|_{X_r}\right)\right)\right] - \text{logit}\left[1 - \exp\left(\Lambda\left(y|_{X_r}\right)\right)\right]\right\} dW\left(y\right) \\ \simeq & \sum_{i=1}^{n} B_{ni}\left(X_r\right) \int \xi\left(Z_i, T_i, \delta_i, X_r, y\right) \frac{w\left(y\right)}{F\left(y|_{X_r}\right)} dy, \end{aligned}$$

for model PO, where the negligible terms in the approximations above may be handled using Taylor expansions, condition C1' and part (b) of Theorem 2 in Iglesias-Pérez and González-Manteiga (1999), as in the proof of Theorem 1.

Replacing the former ε_i in (31) by

$$\varepsilon_i' = \begin{cases} \varepsilon_i^{PH} = \int \xi\left(Z_i, T_i, \delta_i, X_i, y\right) \frac{w(y)}{\Lambda\left(y|_{X_i}\right)} dy & \text{for model PH} \\ \varepsilon_i^{AR} = \int \xi\left(Z_i, T_i, \delta_i, X_i, y\right) \tilde{w}\left(y\right) dy & \text{for model AR} \\ \varepsilon_i^{PO} = \int \xi\left(Z_i, T_i, \delta_i, X_i, y\right) \frac{w(y)}{F\left(y|_{X_i}\right)} dy & \text{for model PO} \end{cases} \tag{50}$$

the rest of the proof proceeds parallel to the proof of Theorem 1 and for this reason we skip it.■

Sketch of the proof of Theorem 9:

The proof proceeds parallel to the study of the terms $\Delta^{(8)}$ and $\Delta^{(9)}$ in the proof of Theorem 1. First of all we split the statistic $T_n^{(1)*}$ into two terms as done in (32):

$$T_n^{(1)*} = \sqrt{n^2 h}\Delta^{*(8)} + \sqrt{n^2 h}\Delta^{*(9)}, \tag{51}$$

with $\Delta^{*(8)} = \sum_{i=1}^{n} b_{nii}\varepsilon_i^{*2}$ and $\Delta^{*(9)} = 2\sum_{i<j} b_{nij}\varepsilon_i^*\varepsilon_j^*$.

Recall (33), then it is straightforward to prove

$$E^*\left(\Delta^{*(8)}\right) = \Delta^{(8)} + \Delta_1^{*(8)}, \tag{52}$$

with $\Delta_1^{*(8)} = \frac{1}{n}\sum_{i,r=1}^{n} B_{ni}\left(X_r\right)^2\left(\hat{\varepsilon}_i^2 - \varepsilon_i^2\right)$. Since the term $\Delta^{(8)}$ has already been studied in (36) we only need to check $\Delta_1^{*(8)}$. To do this we define

$$\Delta_{11}^{*(8)} = \frac{2}{n}\sum_{i,r=1}^{n} B_{ni}\left(X_r\right)^2\varepsilon_i\left(\hat{\varepsilon}_i - \varepsilon_i\right), \tag{53}$$

$$\Delta_{12}^{*(8)} = \frac{1}{n}\sum_{i,r=1}^{n} B_{ni}\left(X_r\right)^2\left(\hat{\varepsilon}_i - \varepsilon_i\right)^2 \tag{54}$$

and write $\Delta_1^{*(8)} = \Delta_{11}^{*(8)} + \Delta_{12}^{*(8)}$. In order to study these terms it is important to find some order for the maximal absolute difference $|\hat{\varepsilon}_i - \varepsilon_i|$.

Similar arguments to those leading to (22) can be used to obtain uniform consistency rates for the kernel estimators $\hat{C}_g\left(z|_x\right)$ and $\hat{H}_{1g}^*\left(u|_x\right)$:

$$\sup_{y\in[a,b],x\in I}\left|\hat{C}_g\left(y|_x\right) - C\left(y|_x\right)\right| = O\left(\left(\frac{\ln n}{ng}\right)^{\frac{1}{2}}\right) \text{ a.s.} \tag{55}$$

$$\sup_{y\in[a,b],x\in I}\left|\hat{H}_{1g}^*\left(y|_x\right) - H_{1g}^*\left(y|_x\right)\right| = O\left(\left(\frac{\ln n}{ng}\right)^{\frac{1}{2}}\right) \text{ a.s.} \tag{56}$$

Now conditions C1, C5, and expressions (22), (55) and (56) imply that

$$\max_{1\le i\le n}|\hat{\varepsilon}_i - \varepsilon_i| = O\left(\left(\frac{\ln n}{ng}\right)^{\frac{1}{2}}\right) \text{ a.s.} \tag{57}$$

This fact together with expectation calculations, Taylor expansions and condition C15 can be used to obtain

$$\begin{aligned}\left|\Delta_{11}^{*(8)}\right| &= O_P\left(\frac{(\ln n)^{\frac{1}{2}}}{n^{\frac{3}{2}}g^{\frac{1}{2}}h}\right),\\ \left|\Delta_{12}^{*(8)}\right| &= O_P\left(\frac{\ln n}{n^2gh}\right).\end{aligned}$$

Now (52) and (36) imply

$$E^*\left(\Delta^{*(8)}\right) = \frac{1}{nh}\left(\int \sigma^2(x)\,dx\right)K^{(2)}(0) \tag{58}$$
$$+O_P\left(n^{-2}h^{-2}+n^{-1}h+n^{-\frac{3}{2}}h^{-1}+n^{-\frac{3}{2}}h^{-\frac{3}{2}}\left(\ln h^{-1}\right)^{\frac{1}{2}}+n^{-\frac{3}{2}}g^{-\frac{1}{2}}h^{-1}(\ln n)^{\frac{1}{2}}\right)$$

To study the bootstrap variance of $\Delta^{*(8)}$ let's define $\gamma = Var^*(W_i^*)$. Now,

$$Var^*\left(\Delta^{*(8)}\right) = \gamma\sum_{i=1}^n b_{nii}^2\hat{\varepsilon}_i^4 = \Delta_2^{*(8)}+\Delta_3^{*(8)}, \tag{59}$$

with

$$\begin{aligned}\Delta_2^{*(8)} &= \gamma\sum_{i=1}^n b_{nii}^2\varepsilon_i^4 = \Delta_{21}^{*(8)}+o_P\left(\Delta_{21}^{*(8)}\right),\\ \Delta_3^{*(8)} &= \gamma\sum_{i=1}^n b_{nii}^2\left(\hat{\varepsilon}_i^4-\varepsilon_i^4\right),\end{aligned}$$

where the order of $\Delta_{21}^{*(8)} = \gamma\sum_{i=1}^n \tilde{b}_{nii}^2\varepsilon_i^4$ can be analyzed by standard expectation calculations: $E\left(\Delta_{21}^{*(8)}\right) = O\left(n^{-3}h^{-2}\right)$. Consequently,

$$\Delta_2^{*(8)} = O_P\left(n^{-3}h^{-2}\right). \tag{60}$$

The term $\Delta_3^{*(8)}$ can be easily bounded:

$$\left|\Delta_3^{*(8)}\right| \le \gamma\left(\max_{1\le i\le n}\left|\hat{\varepsilon}_i-\varepsilon_i\right|\right)\left(\Delta_{31}^{*(8)}+\Delta_{32}^{*(8)}+\Delta_{33}^{*(8)}+\Delta_{34}^{*(8)}\right), \tag{61}$$

where

$$\Delta_{31}^{*(8)} = 8\sum_{i=1}^n b_{nii}^2\left|\varepsilon_i\right|^3,\ \Delta_{32}^{*(8)} = 12\sum_{i=1}^n b_{nii}^2\varepsilon_i^2\left|\hat{\varepsilon}_i-\varepsilon_i\right|,\ \Delta_{33}^{*(8)} = 6\sum_{i=1}^n b_{nii}^2\left|\varepsilon_i\right|\left(\hat{\varepsilon}_i-\varepsilon_i\right)^2$$

and $\Delta_{34}^{*(8)} = \sum_{i=1}^{n} b_{nii}^2 \left|\hat{\varepsilon}_i - \varepsilon_i\right|^3$ can be treated in a parallel way to what has been done before for (53) and (54). Such a study leads to

$$\begin{aligned}
\Delta_{31}^{*(8)} &= O_P\left(n^{-3}h^{-2}\right), \Delta_{32}^{*(8)} = O_P\left(\left(\frac{\ln n}{ng}\right)^{\frac{1}{2}} n^{-3}h^{-2}\right), \\
\Delta_{33}^{*(8)} &= O_P\left(\frac{\ln n}{ng} n^{-3}h^{-2}\right), \Delta_{34}^{*(8)} = O_P\left(\left(\frac{\ln n}{ng}\right)^{\frac{3}{2}} n^{-3}h^{-2}\right),
\end{aligned}$$

which, together with (61), (57) and condition C16 imply

$$\Delta_{3}^{*(8)} = O_P\left(\left(\frac{\ln n}{ng}\right)^{\frac{1}{2}} n^{-3}h^{-2}\right).$$

This expression, (60) and (59) can be used to derive a final order for the bootstrap variance:

$$Var^*\left(\Delta^{*(8)}\right) = O_P\left(n^{-3}h^{-2}\right) \tag{62}$$

The asymptotic expressions found in (58) and (62) for the bootstrap mean and variance of $\Delta^{*(8)}$ can be readily used to conclude

$$\begin{aligned}
\Delta^{*(8)} &= \frac{1}{nh}\left(\int \sigma^2(x)\,dx\right) K^{(2)}(0) \qquad (63) \\
&+ O_{P^*}\left(n^{-2}h^{-2} + n^{-1}h + n^{-\frac{3}{2}}h^{-1} + n^{-\frac{3}{2}}h^{-\frac{3}{2}}\left(\ln h^{-1}\right)^{\frac{1}{2}} + n^{-\frac{3}{2}}g^{-\frac{1}{2}}h^{-1}\left(\ln n\right)^{\frac{1}{2}}\right),
\end{aligned}$$

in probability.

We now proceed by studying the term $\Delta^{*(9)} = 2\sum_{i<j} b_{nij}\varepsilon_i^*\varepsilon_j^*$ in a completely parallel way to what was done for the term $\tilde{\Delta}^{(9)}$ in the proof of Theorem 1. First of all, it is straightforward to prove that $E^*\left(\Delta^{*(9)}\right) = 0$ and

$$Var^*\left(\Delta^{*(9)}\right) = \sum_{j=1}^{9} \Delta_j^{*(9)}, \tag{64}$$

with

$$\begin{aligned}
\Delta_1^{*(9)} &= 4\sum_{i<j} b_{nij}^2 \varepsilon_i^2 \varepsilon_j^2,\ \Delta_2^{*(9)} = 8\sum_{i<j} b_{nij}^2 \varepsilon_i \varepsilon_j^2 \left(\hat{\varepsilon}_i - \varepsilon_i\right), \\
\Delta_3^{*(9)} &= 4\sum_{i<j} b_{nij}^2 \varepsilon_j^2 \left(\hat{\varepsilon}_i - \varepsilon_i\right)^2,\ \Delta_4^{*(9)} = 16\sum_{i<j} b_{nij}^2 \varepsilon_i \varepsilon_j \left(\hat{\varepsilon}_i - \varepsilon_i\right)\left(\hat{\varepsilon}_j - \varepsilon_j\right), \\
\Delta_5^{*(9)} &= 8\sum_{i<j} b_{nij}^2 \varepsilon_i \left(\hat{\varepsilon}_i - \varepsilon_i\right)\left(\hat{\varepsilon}_j - \varepsilon_j\right)^2,\ \Delta_6^{*(9)} = 8\sum_{i<j} b_{nij}^2 \varepsilon_j \left(\hat{\varepsilon}_i - \varepsilon_i\right)^2\left(\hat{\varepsilon}_j - \varepsilon_j\right), \\
\Delta_7^{*(9)} &= 4\sum_{i<j} b_{nij}^2 \left(\hat{\varepsilon}_i - \varepsilon_i\right)^2\left(\hat{\varepsilon}_j - \varepsilon_j\right)^2,\ \Delta_8^{*(9)} = 8\sum_{i<j} b_{nij}^2 \varepsilon_i^2 \varepsilon_j \left(\hat{\varepsilon}_j - \varepsilon_j\right), \\
\Delta_9^{*(9)} &= 4\sum_{i<j} b_{nij}^2 \varepsilon_i^2 \left(\hat{\varepsilon}_j - \varepsilon_j\right)^2.
\end{aligned}$$

To study $\Delta_1^{*(9)}$, which is the dominant term in the bootstrap variance, we use $\Delta_1^{*(9)} = \Delta_{11}^{*(9)} + o_P\left(\Delta_{11}^{*(9)}\right)$,where $\Delta_{11}^{*(9)} = 4\sum_{i<j} \tilde{b}_{nij}^2 \varepsilon_i^2 \varepsilon_j^2$.Routine expectation and variance calculations and conditions C12, C15 and C16 lead to

$$\Delta_{11}^{*(9)} = \frac{2}{n^2 h}\left(\int \sigma^4(x)\,dx\right) K^{(4)}(0) + o_P\left(n^{-2}h^{-1}\right). \tag{65}$$

The study of the terms $\Delta_j^{*(9)}$, for $j = 2, 3, \ldots, 9$ proceeds by using standard bounds, expectation calculations, Taylor expansions and integrability conditions in C15 to conclude

$$\Delta_j^{*(9)} = o_P\left(n^{-2}h^{-1}\right), \text{ for } j = 2, 3, \ldots, 9. \tag{66}$$

Now, (64), (65) and (66) imply

$$Var^*\left(\Delta^{*(9)}\right) = \frac{2}{n^2 h}\left(\int \sigma^4(x)\,dx\right) K^{(4)}(0) + o_P\left(n^{-2}h^{-1}\right)$$

and De Jong's (1987) central limit theorem applies to conclude that

$$\sqrt{n^2 h}\,\Delta^{*(9)} \xrightarrow{\mathrm{d}^*} N(0, V), \tag{67}$$

in probability. The proof concludes by using (51), (63) and (67).■

Acknowledgments

Research supported in part by MEC Grant MTM2005-00429 (ERDF support included) and XUGA Grant PGIDIT03 PXIC10505PN, for the first author, and MEC Grant MTM2005-08020 (ERDF support included) and XUGA Grant PGIDIT03 PXIC20702PN, for the second author. Michael Akritas pointed out several interesting comments and ideas about a first draft of this paper. Several comments and suggestions from the editor and two anonymous referees helped to improved the final version. The authors woul like to thank all of them.

References

Aalen, O.O., 1980, A model for nonparametric regression analysis of counting processes, in: W. Klonecki, A. Kozek and J. Rosinski, eds., Lecture Notes in Statistics 2 (Springer-Verlag) 1-25.

Akritas, M.G., 1996, On the use of nonparametric regression techniques for fitting parametric regression models, Biometrics 52, 1342-1362.

Beran, R., 1981, Nonparametric regression with randomly censored data, Technical report, University of California, Berkeley.

Burke, M.D. and K.C. Yuen, 1995, Goodness-of-fit test for the Cox model via bootstrap method, Journal of Statistical Planning and Infererence 47, 237-256.

Burr, D., 1994, A comparison of certain bootstrap confidence intervals in the Cox model, Journal of the American Statistical Association 89, 1290-1302.

Burr, D. and H. Doss, 1993, Confidence bands for the median survival time as a function of the covariates in the Cox Model, Journal of the American Statistical Association 88, 1330-1340.

Chamberlain, G., 1986, Efficiency in semiparametric models with censoring, Journal of Econometrics 32, 189-218.

Cox, D.R., 1972, Regression models and life tables, Journal of the Royal Statistical Society, Series B 34, 187-220.

De Jong, P., 1987, A central limit theorem for generalized quadratic forms, Probability Theory and Related Fields 75, 261-277.

Dette, H., 1999, A consistent test for the functional form of a regression based on a difference of variance estimators, The Annals of Statistics 27, 1012-1040.

Doss, H. and Y-Ch. Chiang, 1994, Choosing the resampling scheme when bootstrapping: a case study in reliability, Journal of the American Statistical Association 89, 298-308.

González-Manteiga, W. and R. Cao, 1993, Testing the hypothesis of a general linear model using nonparametric regression estimation, Test 2, 161-188.

Gray, R., 1990, Some diagnostic methods for Cox regression models through hazard smoothing, Biometrics 46, 93-102.

Grigoletto, M. and M.G. Akritas, 1999, Analysis of covariance with incomplete data via semiparametric model transformations, Biometrics 55, 1177-1362.

Hall, P. and Q. Yao, 2005, Approximating conditional distribution functions using dimension reduction, The Annals of Statistics 33, 1404-1421.

Härdle, W. and E. Mammen, 1993, Comparing nonparametric versus parametric regression fits, The Annals of Statistics 21, 1926-1947.

Iglesias-Pérez, C. and W. González-Manteiga, 1999, Strong representation of a generalized product-limit estimator for truncated and censored data with some applications, Journal of Nonparametric Statistics 10, 213-244.

Kaplan, E.L. and P. Meier, 1958, Nonparametric estimation from incomplete observations, Journal of American Statistical Association 53, 457-481.

Kiefer, N.M., 1988. Economic duration data and hazard functions, Journal

of Economic Literature 26, 646–679.

Lavalley, M.P. and M.G. Akritas, 1994, Extensions of the Lynden-Bell-Woodroofe method for truncated data, Unpublished manuscript.

Lewbel, A. and Linton, O., 2002, Nonparametric Censored and Truncated Regression, Econometrica 70, 765-779

Lin, D.Y. and C.F. Spiekerman, 1996, Model checking techniques for parametric regression with censored data, Scandinavian Journal of Statistics 23, 157-177.

Lin, D.Y. and L.J. Wei, 1991, Goodness-of-fit tests for the general Cox regression model, Statistica Sinica 1, 1-17

Lynden-Bell, D., 1971, A method of allowing for non observational selection in small samples applied to 3CR quasars, Monthly Notices of the Royal Astronomical Society 155, 95-118.

Marzec, L. and P. Marzec, 1997, On fitting Cox's regression model with time-dependent coefficients, Biometrika 84, 901-908.

McKeague, I.W. and Y. Sun, 1996, Towards an omnibus distribution-free goodness-of-fit test for the Cox model, Statistica Sinica 6, 579-588.

Nadaraya, E.A., 1964, On estimating regression, Theory of Probability and its Applications 9, 141-142.

Peña, E.A., 1998, Smooth goodness-of-fit tests for the baseline hazard in Cox's proportional hazard models, Journal of the American Statistical Association 93, 673-692.

Sánchez-Sellero, C., W. González-Manteiga and I. Van Keilegom, 2005, Uniform representation of product-limit integrals with applications, Scandinavian Journal of Statistics 32, 563–581.

Serfling, R.J., 1980, Approximation Theorems of Mathematical Statistics (John Wiley & Sons, New York).

Silverman, B.W., 1978, Weak and strong uniform consistency of the kernel

estimate of a density and its derivatives, The Annals of Statistics 6, 177-184.

Stute, W., 1999, Nonlinear censored regression, Statistica Sinica 9, 1089–1102.

Stute, W., W. González-Manteiga and C. Sánchez-Sellero, 2000, Nonparametric model checks in censored regression, Communications in Statistics: Theory and Methods 29, 1611-1629.

Tsai, W.Y., N.P. Jewell and M.C. Wang, 1987, A note on the product-limit estimator under right censoring and left truncation, Biometrika 74, 883-886.

Van Keilegom, I. and N. Veraverbeke, 1996, Uniform strong convergence results for the conditional Kaplan-Meier estimator and its quantiles, Communications in Statistics: Theory and Methods 25, 2251-2265.

Van Keilegom, I. and N. Veraverbeke, 1997a, Weak convergence of the bootstrapped conditional Kaplan-Meier process and its quantile process, Communications in Statistics: Theory and Methods 26, 853-869.

Van Keilegom, I. and N. Veraverbeke, 1997b, Estimation and bootstrap with censored data in fixed design nonparametric regression, Annals of the Institute of Statistical Mathematics 49, 467-491.

Verweij, P.J.M., H.C. Van Houwelingen and T. Stijnen, 1998, A goodness-of-fit test for Cox's proportional hazards model based on martingale residuals, Biometrics 54, 1517-1526.

Watson, G.S. (1964), Smooth regression analysis, Sankhyā, Series A 26, 359-372.

Zhu, L.X., 2005, Nonparametric Monte Carlo tests and their applications, Lecture Notes in Statistics, 182 (Springer, New York).

Zhu, L.X., K.C. Yuen and N.Y. Tang, 2002, Resampling methods for testing a semiparametric random censorship model, Scandinavian Journal of Statistics 29, 111-123.