ANNALS OF ECONOMICS AND FINANCE **2**, 77–96 (2001)

A Consistent Test for the Parametric Specification of the Hazard Function *

Yanqin Fan

Department of Economics
University of Windsor
Windsor, Ontario N9B 3P4 Canada
E-mail: Fan1@uwindsor.ca

and

Paul Rilstone

Department of Economics
York University
North York, Ontario M3J 1P3 Canada
E-mail: Pril@haavelmo.econ.yorku.ca

This paper develops a consistent test for the correct hazard rate specification within the context of random right hand censoring of the dependent variable. The test is based on comparing a parametric estimate with a kernel estimate of the hazard rate. We establish the asymptotic distribution of the test statistic under the null hypothesis of correct parametric specification of the hazard rate and establish the consistency of the test. © 2001 Peking University Press

Key Words: Consistent test; Hazard rate; Random censoring; Kernel estimation; Boundary kernel.
JEL Classification Numbers: C14, C52.

1. INTRODUCTION

Most of the current research into consistent model specification testing has focused on density and regression functions and on situations where the sampling is assumed to be either random or at least stationary, see Hart (1997) for a detailed discussion on nonparametric methods of func-

* Research of both authors was supported by Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada. Corresponding author: Yanqin Fan.

1529-7373/2001
Copyright © 2001 by Peking University Press
All rights of reproduction in any form reserved.

tion estimation and their use in testing the adequacy of parametric function specifications. This paper takes a similar approach but is concerned with developing a consistent test for correct specification of hazard rates. Hazard rate estimation is a very common task of applied econometricians. At present there does not really exist a suitable method for consistently testing if a parametric hazard rate has been correctly specified. As with regression and density analysis, a misspecified model can easily lead to incorrect inferences.

In this paper we adapt some of the recent results of Fan and Li (1996) to consistent model specification testing within the context of random right hand censoring of the dependent variable. Since this sort of problem is most often encountered in the context of duration analysis we will generally assume that the focus of inference is a hazard rate with covariates. We implicitly allow for situations where some of these may be unobserved. In principle, one could focus on, say specification of the survivor function or the integrated hazard. However, since it is usually the hazard rate which is directly specified in duration analysis, it is reasonable to concentrate on it.

There are currently several methods for testing these models. Nakamura and Walker (1994) provide an overview. One can use traditional LM, LR, and Wald tests. These are useful since they are based on maximum likelihood estimates and estimation in duration models is generally based on maximizing a likelihood function. They obviously are not robust. It is also popular to use Conditional Moment (CM) tests. They are often more robust than the likelihood-based tests. However, since they are based on a finite number of moments, it is generally possible to find alternatives against which they have little or no power. We shall return to these again below. Horowitz and Neumann (1992) have a variant of this based on moment restrictions. Perhaps the most popular form of model free testing is a form of residual analysis based on the Kaplan-Meier estimate of the survivor function. While in some cases this form of residual analysis may be informative, it is not in itself a rigorous test. Although the Kaplan-Meier estimator may be interpreted as a maximum likelihood estimator and has certain optimality features, its statistical properties are quite cumbersome to work out. It is also designed for unconditional models and is quite awkward to adapt to conditional models with covariates.

In this paper we develop a method for systematically evaluating the distance between a parametric estimated hazard rate and its nonparametric counterpart. This approach to testing model specification has become very popular recently, see e.g. Aït-Sahalia, Bickel, and Stoker (1994), Fan and Li (1996), Gozalo (1993), Härdle and Mammen (1993), Hong (2000), Hong and White (1995), Horowitz and Härdle (1994), Li and Wang (1998), Wooldridge (1992), Yatchew (1992), Zheng (1996) to mention only a few.

However, the existing tests are designed for testing the specification of a density function or of a regression function. They are not directly applicable to duration models for several reasons: First, in duration analysis, one typically specifies the hazard function rather than the density function; Second, there is often censoring in duration data which is not allowed in existing work; Third, the duration variable is non-negative. Hence, the kernel estimate suffers from the well known boundary effect. To the best of our knowledge, this has not been taken into account explicitly in existing work on model specification testing. The current paper attempts to bridge this gap. Specifically, we establish a consistent test for the parametric specification of the hazard rate allowing for the presence of censoring in the data. To overcome the boundary effect, we use the class of boundary kernels introduced in Müller (1991).

The remainder of this paper is organized as follows. In the next section we introduce the kernel estimate and the parametric estimate of the hazard function and present a measure of distance between the two estimates. This measure forms the basis of our test. In Section 3 we first establish the asymptotic distribution of the measure introduced in Section 2 under the null hypothesis and then construct our test. The last section concludes. The technical proofs are postponed to the Appendix.

2. THE NULL HYPOTHESIS AND THE KERNEL ESTIMATE

Let T^* be a duration variable with the conditional density function $f(\cdot|x)$, the conditional survivor function $F(\cdot|x)$, and the conditional hazard function $\lambda(\cdot|x)$, where the covariate X takes values in R^l. Let τ be a censoring variable with the conditional density function $g(\cdot|x)$ and the conditional survivor function $G(\cdot|x)$. For simplicity, we consider random censorship in this paper so that T^* and τ are independent.

Suppose n i.i.d. observations $\{t_i, d_i, x_i\}_{i=1}^n$ are available, where $t_i = \min(t_i^*, \tau_i)$ and $d_i = I_{\{t_i = t_i^*\}} = I_{\{t_i^* \leq \tau_i\}}$, where I_A is the indicator function of the set A. We are interested in testing the parametric functional form of the conditional hazard function $\lambda(\cdot|x)$. Namely, if $\{\lambda_0(\cdot|x,\beta) : \beta \in \mathcal{B} \in R^p\}$ is a family of parametric hazard functions, then the hypotheses of interest can be formulated as

$$H_0: \quad P(\lambda(T^*|X) = \lambda_0(T^*|X, \beta_0)) = 1 \text{ for some } \beta_0 \in \mathcal{B},$$

$$H_A: \quad P(\lambda(T^*|X) = \lambda_0(T^*|X, \beta)) < 1 \text{ for all } \beta \in \mathcal{B}.$$

Note that under H_0, the conditional density function and the conditional survivor function of the duration variable T^* take respectively the parametric forms, $f_0(\cdot|x,\beta_0)$ and $F_0(\cdot|x,\beta_0)$ (say) such that $\lambda_0(\cdot|x,\beta_0) = f_0(\cdot|x,\beta_0)/F_0(\cdot|x,\beta_0)$, where

$$f_0(t|x,\beta_0) = \lambda_0(t|x,\beta_0)\exp(-\int_0^t \lambda_0(s|x,\beta_0)ds).$$

To test H_0 versus H_A, we take a similar approach to that in Fan (1994) by comparing a kernel estimate of $\lambda(\cdot|\cdot)$ with a parametric estimate of $\lambda_0(\cdot|\cdot,\beta_0)$. By choosing an appropriate measure between the two estimates, we will develop a consistent test for H_0.

2.1. The Nonparametric and Parametric Estimates

Let T be the random variable that is i.i.d. as t_i and let $h_1(t,x)$ denote the joint probability density function of T, X and $d = 1$. Then it can be shown that $h_1(t,x) = f(t|x)G(t|x)f(x)$, where $f(x)$ is the density function of X. Similarly one can show that the conditional survivor function of T given x is $F(t|x)G(t|x)$. Set $h_2(t,x) = F(t|x)G(t|x)f(x)$. Then the conditional hazard function $\lambda(t|x)$ of T^* has the following expression

$$\lambda(t|x) = \frac{h_1(t,x)}{h_2(t,x)} = \frac{f(t|x)}{F(t|x)}. \tag{1}$$

Although $f(t|x)$ and $F(t|x)$ are not directly estimable, the functions $h_1(t,x)$ and $h_2(t,x)$ can be consistently estimated from the random sample $\{t_i, d_i, x_i\}_{i=1}^n$. Specifically, a kernel estimator of $h_1(t,x)$ is given by

$$\hat{h}_1(t,x) = \frac{1}{(n-1)\gamma^{l+1}}\sum_{j\neq i} d_j K_{1t}(\frac{t-t_j}{\gamma})K_2(\frac{x-x_j}{\gamma}), \tag{2}$$

where $\gamma = \gamma_n \to 0$ is a smoothing parameter, $K_2(\cdot)$ is an l dimensional kernel function, and

$$K_{1t}(z) = \begin{cases} K_{1+}(1,z) & \text{if } \gamma \le t < \infty \\ K_{1+}(\frac{t}{\gamma},z) & \text{if } 0 \le t < \gamma \end{cases} \tag{3}$$

with K_{1+} a boundary kernel satisfying Assumption (K1) introduced in Section 3. Here and in (4) below, the boundary kernel K_{1+} is used for t in the boundary region $[0,\gamma]$ to overcome the boundary effect associated with the duration variable T. Similarly, a kernel estimator of $h_2(t,x)$ is given by

$$\hat{h}_2(t,x) = \frac{1}{(n-1)\gamma^{l+1}}\sum_{j\neq i}[\int_t^\infty K_{1t}(\frac{u-t_j}{\gamma})du]K_2(\frac{x-x_j}{\gamma}). \tag{4}$$

The nonparametric estimator of the hazard function is defined as

$$\hat{\lambda}(t|x) = \frac{\hat{h}_1(t,x)}{\hat{h}_2(t,x)}. \tag{5}$$

Note that under regularity conditions, it can be shown that $\hat{h}_1(t,x)$ is a consistent estimator of $h_1(t,x)$ and $\hat{h}_2(t,x)$ is a consistent estimator of $h_2(t,x)$. Hence $\hat{\lambda}(t|x)$ is a consistent estimator of the hazard function of T^*. In fact, one can show that $\hat{h}_2(t,x)$ converges faster than $\hat{h}_1(t,x)$ because of the integration involved in the definition of $\hat{h}_2(t,x)$. This resembles the well known result that the kernel estimator of a distribution function converges faster than the corresponding kernel estimator of the density function.

The kernel estimator $\hat{\lambda}$ of the hazard function is to be compared with a parametric estimator obtained under H_0. Since the hazard function takes the parametric form $\lambda_0(t|x,\beta_0)$ under H_0, the conditional density function of T^* is given by $f_0(t|x,\beta_0) = \lambda_0(t|x,\beta_0)\exp(-\int_0^t \lambda_0(s|x,\beta_0)ds)$. Suppose that the density function of the covariate X does not depend on β_0. Then under H_0, β_0 can be root-n consistently estimated by the maximum likelihood estimator $\hat{\beta}$ (say). The corresponding parametric estimator of the hazard function is $\lambda_0(t|x,\hat{\beta})$. Given the parametric estimator of the hazard function, we can obtain a parametric estimator of the density function of T^*, $f_0(t|x,\hat{\beta})$ and of the survivor function $F_0(t|x,\hat{\beta})$.

2.2. The Basis of the Test

For any β, define

$$S(\beta) = \frac{1}{n}\sum_{i=1}^{n}[\lambda_0(t_i|x_i,\beta) - \hat{\lambda}_i]^2[\hat{h}_2(t_i,x_i)]^2 w_i d_i, \tag{6}$$

where $\hat{\lambda}_i = \hat{\lambda}(t_i|x_i)$ is the nonparametric estimator of the hazard function defined in (5), $\hat{h}_2(t_i,x_i)$ is given in (4), and $w_i = w(t_i,x_i)$ is a positive weighting function which can be used to direct power of the test towards different directions.

Our test for H_0 will be based on $S(\hat{\beta})$. Note that by using a weighted average squared difference between the two estimates in $S(\hat{\beta})$ instead of the integrated squared difference as in Fan (1994), we avoid having to evaluate an $(l+1)$ dimensional integral numerically. The multiplication by $[\hat{h}_2(t_i,x_i)]^2$ in (6) gets rid of the denominator in $\hat{\lambda}_i$. This greatly simplifies the technical analysis.

Intuitively, one would expect that under certain regularity conditions,

$$S(\hat{\beta}) \to \int\int[\lambda_0(t|x,\beta_*) - \lambda(t|x)]^2 h_2^2(t,x)w(t,x)h_1(t,x)dtdx \text{ in probability,}$$

where $\beta_* = \beta_0$ under H_0 (see White (1982) or Assumption (P) introduced in Section 3). Since the latter term is non-negative and is zero if and only if the null hypothesis holds, the test based on $S(\hat{\beta})$ proposed in the next section will be consistent for testing H_0 against H_A.

3. THE TEST AND ITS ASYMPTOTIC PROPERTIES

The derivation of the asymptotic null distribution of $S(\hat{\beta})$ is very tedious algebraically, because it depends on three estimators $\hat{h}_1(t_i, x_i)$, $\hat{h}_2(t_i, x_i)$, and $\hat{\beta}$. However, the idea underlying the derivation is not difficult to understand. To see this, we introduce

$$\bar{S}(\beta) = \frac{1}{n}\sum_{i=1}^{n}[\lambda_0(t_i|x_i,\beta) - \frac{\hat{h}_1(t_i,x_i)}{h_2(t_i,x_i)}]^2[h_2(t_i,x_i)]^2 w_i d_i. \tag{7}$$

Note from (5), (6), and (7) that the only difference between $S(\beta)$ and $\bar{S}(\beta)$ is the replacement of $\hat{h}_2(t_i, x_i)$ in $S(\beta)$ by $h_2(t_i, x_i)$ in $\bar{S}(\beta)$. Heuristically, since $\hat{h}_2(t, x)$ converges at a faster rate than $\hat{h}_1(t, x)$, under certain conditions, the asymptotic null distribution of $S(\beta_0)$ is the same as that of $\bar{S}(\beta_0)$ apart from the center terms. By the same token, one can show that the asymptotic null distribution of $S(\hat{\beta})$ is the same as that of $S(\beta_0)$, because $\hat{\beta}$ converges faster than both $\hat{h}_1$ and $\hat{h}_2$. Consequently, the asymptotic null distribution of $S(\hat{\beta})$ is given by that of $\bar{S}(\beta_0)$ apart from the center term.

3.1. Assumptions

Throughout this section, we will work with the following assumptions.

(f) The functions $F(t|x)$, $G(t|x)$, and $f(x)$ and their m-th order partial derivatives with respect to t and/or x are bounded and uniformly continuous on $R_+ \times R^l$, where m is a positive integer. The weight function $w(t, x)$ is Lipschitz continuous.

(K1) The support of $K_{1+}(q, z)$ is $[0, 1] \times [-1, q]$. For a fixed q, $K_{1+}(q, \cdot)$ is of order m on $[-1, q]$, that is

$$\int_{-1}^{q} z^i K_{1+}(q,z)dz = \begin{cases} 1, & i = 0, \\ 0, & 0 < i < m, \\ (-1)^m m! k_{mq}, & i = m. \end{cases}$$

For some finite constants $L, C > 0$, $sup_{z,q}|K_{1+}(q, z)| < C$ and $sup_q|K_{1+}(q, z_1) - K_{1+}(q, z_2)| \le L|z_1 - z_2|$ for all $z_1, z_2 \in [-1, q]$.

(K2) The kernel function $K_2(\cdot)$ is a bounded, symmetric function on R^l that satisfies $\int |K_2(u)|du < \infty$, $||u||^l|K_2(u)| \to 0$ as $||u|| \to \infty$, and is of

order m. Specifically, we assume

$$\int u_1^{i_1} u_2^{i_2} \ldots u_l^{i_l} K_2(u) du = \begin{cases} 1, & i_1 = \ldots = i_l = 0, \\ 0, & 0 < \sum_{j=1}^{l} i_j < m, \text{ or } \sum_{j=1}^{l} i_j = m \\ & \text{ and } i_j < m \text{ for all } j = 1, 2, \ldots, l, \\ (-1)^m m! k_m, & \sum_{j=1}^{l} i_j = m \text{ and } i_j = m \text{ for some } j, \end{cases}$$

and $\int |u_1^{i_1} \ldots u_l^{i_l} K_2(u)| du < \infty$ for $\sum_{j=1}^{l} i_j = m$, where $i_1, \ldots, i_l$ are non-negative integers, $||\cdot||$ is the Euclidean norm, and k_m does not depend on j.

(G) The smoothing parameter satisfies $\gamma \to 0$, and $n\gamma^{l+1} \to \infty$, and $n\gamma^{(l+1)/2+2m} \to 0$.

(P) There exists $\beta_* \in \mathcal{B}$ such that $\hat{\beta} \to \beta_*$ almost surely, and

$$\hat{\beta} - \beta_* = \frac{1}{n} A(\beta_*)^{-1} \sum_{i=1}^{n} D \log f(t_i | x_i, \beta_*) + o_p(n^{-1/2}),$$

where $D \log f(t_i|x_i, \beta_*)$ is the $p \times 1$ vector of first order partial derivatives of $\log f(t_i|x_i, \beta)$ with respect to β evaluated at $\beta = \beta_*$, and $A(\beta_*) = E[D^2 \log f(t_i|x_i, \beta_*)]$.

Assumption (f) imposes smoothness conditions on the conditional survivor functions of the duration variable and the censoring variable, as well as the density function of the covariate. Assumptions (K1) and (K2) specify conditions on the kernel functions associated with the duration variable and the covariate. Since the duration variable is non-negative, assumption (K1) requires that the kernel function K_{1+} be a boundary kernel of order m. Note that $K_{1+}(1, z)$ is a standard kernel function of order m on $[-1, 1]$. For more details on boundary kernels, see Müller (1991). Assumption (G) requires that the smoothing parameter γ undersmooth the kernel estimate $\hat{h}_1(t, x)$ of $h_1(t, x)$. Fan (1994) considers three cases corresponding to undersmoothing, oversmoothing, and optimal smoothing, and develops three different tests for the parametric specification of a density function accordingly. Hong (2000) develops a test for the parametric specification of a regression function using optimal smoothing. It is worth pointing out here that the classification of smoothing here is with respect to kernel estimation instead of testing, i.e., optimal smoothing for estimation may not be optimal for testing. Assumption (P) is introduced to examine the effect of estimating β_0 by $\hat{\beta}$ on the asymptotic null distribution of $S(\hat{\beta})$. For primitive conditions under which this assumption holds, see White (1982).

3.2. The Asymptotic Null Distribution of $S(\hat{\beta})$

We are now ready to establish the asymptotic null distribution of $S(\hat{\beta})$. Some details of the technical proofs are postponed to the Appendix. We provide an outline here.

As explained at the beginning of this section, the asymptotic null distribution of $S(\hat{\beta})$ is determined by that of $\bar{S}(\beta_0)$ apart from the center term. Hence we first establish the asymptotic null distribution of $\bar{S}(\beta_0)$.

Let $h_1(t,x,\beta_0) = f_0(t|x,\beta_0)G(t|x)f(x)$ and $h_2(t,x,\beta_0) = F_0(t|x,\beta_0)G(t|x)f(x)$. Noting that under H_0, $\lambda(t|x) = \lambda_0(t|x,\beta_0) = h_1(t,x,\beta_0)/h_2(t,x,\beta_0)$, one can decompose $\bar{S}(\beta_0)$ into the sum of three terms as in (8) below. Specifically, let E_i denote the conditional expectation given (t_i,x_i). Then we have from (7)

$$\begin{aligned}\bar{S}(\beta_0) &= \frac{1}{n}\sum_i[\hat{h}_1(t_i,x_i) - E_i\hat{h}_1(t_i,x_i)]^2 w_i d_i \\ &+\frac{2}{n}\sum_i[\hat{h}_1(t_i,x_i) - E_i\hat{h}_1(t_i,x_i)][E_i\hat{h}_1(t_i,x_i) - h_1(t_i,x_i,\beta_0)] \\ &+\frac{1}{n}\sum_i[E_i\hat{h}_1(t_i,x_i) - h_1(t_i,x_i,\beta_0)]^2 w_i d_i \\ &\equiv S_1 + 2S_2 + S_3. \end{aligned} \tag{8}$$

Each of the three terms S_1, S_2, and S_3 in (8) is an example of the numerous terms that we will need to handle in the derivation of the asymptotic null distribution of $S(\hat{\beta})$. Hence we will analyze S_1, S_2, and S_3 in detail, and only provide the final results for the rest of the terms in the paper. For clarity, we will classify these terms into three categories:

Category 1. Random variation only: S_1 results from the random variation of $\hat{h}_1(t_i,x_i)$;

Category 2. Random and deterministic variations: S_2 consists of the interaction between the random variation and the bias of $\hat{h}_1(t_i,x_i)$;

Category 3. Deterministic variation only: S_3 is due to the bias of $\hat{h}_1(t_i,x_i)$ only.

Depending on the smoothing parameter γ, both S_1 and S_2 may contribute to the asymptotic variance of $\bar{S}(\beta_0)$ as in Fan (1994). Under assumption (G), i.e., undersmoothing, we will show that S_1 dominates S_2 asymptotically. Hence the asymptotic variance of $\bar{S}(\beta_0)$ is given by that of S_1. The last term S_3 contributes to the center of the asymptotic distribution of $\bar{S}(\beta_0)$. In summary, we have

PROPOSITION 3.1. *Under assumptions (f), (K1), (K2), and (G), if H_0 holds, then*

$$n\gamma^{(l+1)/2}[\bar{S}(\beta_0) - c_1(n)] \to N(0, 2\sigma^2) \text{ in distribution },$$

where

$$c_1(n) = \frac{1}{n\gamma^{(l+1)}}[\int_{-\infty}^{\infty} K_2^2(x)dx] \int_0^{\infty} \{[\int_{-1}^{t/\gamma} K_{1t}^2(s)ds][\int w(t,x)h_1^2(t,x)dx]\}dt,$$

$$\sigma^2 = \{\int\int w^2(t,x)h_1^4(t,x)dtdx\}\{\int_0^{\infty}[K_{1+}{*}K_{1+}(1,s)]^2 ds\}\{\int [K_2{*}K_2(y)]^2 dy\},$$

with $K_{1+} * K_{1+}(1,s) = \int_{-1}^{1} K_{1+}(1,t_2)K_{1+}(1,s+t_2)dt_2$ *and* $K_2 * K_2(y) = \int K_2(x)K_2(y+x)dx$.

Proof. The structure of the proof is similar to, but more complicated than, that of Corollary 2.4 (c2) in Fan (1994). It consists of three steps: (i) the derivation of the asymptotic distribution of S_1; (ii) the derivation of the order of S_2; (iii) the derivation of the order of S_3.

(i) Let $K_{1i,ij} = K_{1t_i}(\frac{t_i - t_j}{\gamma})$, $K_{2ij} = K_2(\frac{x_i - x_j}{\gamma})$, and $K_{i,ij} = K_{1i,ij}K_{2ij}$. Then it follows from (8) that

$$\begin{aligned} S_1 &= \frac{1}{n(n-1)^2\gamma^{2(l+1)}} \sum\sum\sum_{i\neq j\neq k} [d_jK_{i,ij} - E_i(d_jK_{i,ij})][d_kK_{i,ik} \\ &- E_i(d_kK_{i,ik})]w_id_i + \frac{1}{n(n-1)^2\gamma^{2(l+1)}} \sum\sum_{i\neq j}[d_jK_{i,ij} - E_i(d_jK_{i,ij})]^2 w_id_i \\ &\equiv S_{11} + S_{12}. \end{aligned} \tag{9}$$

The first term S_{11} can be rewritten in terms of a U-statistic:

$$S_{11} = \frac{1}{3\gamma^{2(l+1)}}U_{n1}, \tag{10}$$

where

$$U_{n1} = \binom{n}{3}^{-1} \sum\sum\sum_{i<j<k} H_{n1}(z_i, z_j, z_k), \tag{11}$$

with $z_i = (t_i, x_i, d_i)$ and

$$\begin{aligned} H_{n1}(z_i,z_j,z_k) &= [d_jK_{i,ij} - E_i(d_jK_{i,ij})][d_kK_{i,ik} - E_i(d_kK_{i,ik})]w_id_i \\ &+[d_iK_{j,ji} - E_j(d_iK_{j,ji})][d_kK_{j,jk} - E_j(d_kK_{j,jk})]w_jd_j \\ &+[d_jK_{k,kj} - E_k(d_jK_{k,kj})][d_iK_{k,ki} - E_k(d_iK_{k,ki})]w_kd_k. \end{aligned} \tag{12}$$

It is easy to show that $E[H_{n1}(z_1, z_2, z_3)|z_1] = 0$, implying that U_{n1} is a degenerate U-statistic. By the proof of Lemma B.4 in Fan and Li (1996), it follows that under easily verifiable conditions (see Fan and Li (1996) for details), one gets from (10) and (11):

$$\begin{aligned}
S_{11} &= \frac{1}{3\gamma^{2(l+1)}}[\frac{6}{n(n-1)}\sum\sum_{i<j} E\{H_{n1}(z_i, z_j, z_k)|z_i, z_j\}] + o_p(\frac{1}{n(\gamma^{l+1})^{1/2}}) \\
&= \frac{2}{n(n-1)\gamma^{2(l+1)}}\sum\sum_{i<j} E\{[d_j K_{k,kj} - E_k(d_j K_{k,kj})][d_i K_{k,ki} \\
&- E_k(d_i K_{k,ki})]w_k d_k|z_i, z_j\} + o_p(\frac{1}{n(\gamma^{l+1})^{1/2}}) \\
&= \frac{2}{n(n-1)\gamma^{2(l+1)}}\sum\sum_{i<j}\int_0^\infty\int_{-\infty}^\infty [d_j K_{1t}(\frac{t-t_j}{\gamma})K_2(\frac{x-x_j}{\gamma}) - e_1(t,x)] \\
&\times [d_i K_{1t}(\frac{t-t_i}{\gamma})K_2(\frac{x-x_i}{\gamma}) - e_1(t,x)]w(t,x)h_1(t,x)dxdt + o_p(\frac{1}{n(\gamma^{l+1})^{1/2}}) \\
&= \frac{2}{n(n-1)\gamma^{2(l+1)}}\sum\sum_{i<j}\bar{H}_{n1}(z_i, z_j) + o_p(\frac{1}{n(\gamma^{l+1})^{1/2}}), \qquad (13)
\end{aligned}$$

where $e_1(t,x) = E[d_1 K_{2,21}|t_2 = t, x_2 = x] = E[d_1 K_{1t}(\frac{t-t_1}{\gamma})K_2(\frac{x-x_1}{\gamma})]$ and the definition of $\bar{H}_{n1}(z_i, z_j)$ should be obvious from (13).

Since $E[\bar{H}_{n1}(z_i, z_j)|z_i] = 0$, it follows from Theorem 1 in Hall (1984) that $\sum\sum_{i<j}\bar{H}_{n1}(z_i, z_j)$ is asymptotically normally distributed with zero mean and variance given by $2^{-1}n^2 E[\bar{H}_{n1}^2(z_1, z_2)]$, provided the following condition holds (The proof of this is similar to that in Hall (1984) and is thus omitted):

$$\frac{E[\bar{G}_{n1}^2(z_1, z_2)] + n^{-1}E[\bar{H}_{n1}^4(z_1, z_2)]}{\{E[\bar{H}_{n1}^2(z_1, z_2)]\}^2} \to 0,$$

where $\bar{G}_{n1}(x, y) = E[\bar{H}_{n1}(x, z_1)\bar{H}_{n1}(y, z_1)]$. In Lemma A.1 in the Appendix, we show that $E[\bar{H}_{n1}^2(z_1, z_2)] = \gamma^{3(l+1)}\sigma^2 + o((\gamma^{3(l+1)}))$. Hence S_{11} is asymptotically normal with zero mean and variance given by $(n^2\gamma^{l+1})^{-1}[2\sigma^2 + o(1)]$.

Similar to Fan (1994), it is straightforward to show that $S_{12} = c_1(n) + o_p((n\gamma^{(l+1)/2})^{-1})$. Hence

$$n\gamma^{(l+1)/2}(S_1 - c_1(n)) \to N(0, 2\sigma^2) \text{ in distribution.}$$

(ii) To analyze S_2, we need to know the bias structure of $\hat{h}_1$. This is given in Lemma A.2 (i) in the Appendix. Let

$$b_1(t,x) = [k_{mq}\frac{\partial^m h_1(t,x)}{\partial t^m} + k_m \sum_{i=1}^{l} \frac{\partial^m h_1(t,x)}{\partial x_i^m}].$$

Using Lemma A.2 (i), we get

$$\begin{aligned} S_2 &= \frac{\gamma^m}{n}\sum_{i=1}^{n} b_1(t_i,x_i)[\hat{h}_1(t_i,x_i) - E_i(\hat{h}_1(t_i,x_i))] \\ &= \frac{\gamma^m}{n(n-1)\gamma^{l+1}}\sum_i\sum_{j\neq i}[d_jK_{i,ij} - E_i(d_jK_{i,ij})] \\ &= \gamma^m U_{n2}, \end{aligned} \tag{14}$$

where

$$U_{n2} = \binom{n}{2}^{-1}\sum_i\sum_{j<i}\{[d_jK_{i,ij} - E_i(d_jK_{i,ij})] + [d_iK_{j,ji} - E_j(d_iK_{j,ji})]\}/\gamma^{l+1}.$$

Note that unlike U_{n1}, U_{n2} is a non-degenerate U-statistic which is similar to the U-statistic resulting from the weighted average derivative estimation in Powell, Stock, and Stoker (1989). Using Lemma 2.1 in Powell, Stock, and Stoker (1989), one can easily show that $U_{n2} = O_p(n^{-1/2})$ and hence $S_2 = O_p(\gamma^m n^{-1/2}) = o_p((n\gamma^{(l+1)/2})^{-1})$ under Assumption (G).

(iii) Based on Lemma A.2(i), one can easily show that $S_3 = O_p(\gamma^{2m}) = o_p((n\gamma^{(l+1)/2})^{-1})$ under Assumption (G).

The conclusion of Proposition 3.1 follows immediately from (8) and the results in (i)-(iii) above. ▌

We now show that apart from the center terms, $S(\beta_0)$ and $\bar{S}(\beta_0)$ are of the same asymptotic null distribution.

Proposition 3.2. *Under H_0, assumptions (f), (K1), (K2), and (G), we have*

$$n\gamma^{(l+1)/2}(S(\beta_0) - c(n)) \to N(0, 2\sigma^2) \text{ in distribution },$$

where $c(n) = c_1(n) + c_2(n) - 2c_3(n)$ with $c_1(n)$ and σ^2 defined in Proposition 3.1, $c_2(n)$ defined in (16), and $c_3(n)$ defined in (17).

Proof. It is easy to see that the following decomposition holds:

$$S(\beta_0) - \bar{S}(\beta_0) = \frac{1}{n}\sum_i[\hat{h}_2(t_i,x_i) - E_i\hat{h}_2(t_i,x_i)]^2\lambda^2(t_i|x_i,\beta_0)w_id_i$$

$$-\frac{2}{n}\sum_i[\hat{h}_1(t_i,x_i)-E_i\hat{h}_1(t_i,x_i)][\hat{h}_2(t_i,x_i)-E_i\hat{h}_2(t_i,x_i)]\lambda(t_i|x_i,\beta_0)w_id_i$$
$$-\frac{2}{n}\sum_i[\hat{h}_1(t_i,x_i)-E_i\hat{h}_1(t_i,x_i)][E_i\hat{h}_2(t_i,x_i)-h_2(t_i,x_i,\beta_0)]\lambda(t_i|x_i,\beta_0)w_id_i$$
$$-\frac{2}{n}\sum_i[\hat{h}_2(t_i,x_i)-E_i\hat{h}_2(t_i,x_i)]\{[E_i\hat{h}_1(t_i,x_i)-h_1(t_i,x_i,\beta_0)]$$
$$-[E_i\hat{h}_2(t_i,x_i)-h_2(t_i,x_i,\beta_0)]\lambda(t_i|x_i,\beta_0)\}\lambda(t_i|x_i,\beta_0)w_id_i$$
$$+\frac{1}{n}\sum_i[E_i\hat{h}_2(t_i,x_i)-h_2(t_i,x_i,\beta_0)]^2\lambda^2(t_i|x_i,\beta_0)w_id_i$$
$$-\frac{2}{n}\sum_i[E_i\hat{h}_1(t_i,x_i)-h_1(t_i,x_i,\beta_0)][E_i\hat{h}_2(t_i,x_i)-h_2(t_i,x_i,\beta_0)]\lambda(t_i|x_i,\beta_0)w_id_i$$
$$\equiv[S_4-2S_5]-2[S_6+S_7]+[S_8-2S_9], \tag{15}$$

where the definitions of S_4-S_9 should be clear from (15).

Although the above decomposition looks complicated, the terms on the right hand side of (15) are similar to S_1, S_2, and S_3 in (8) which are analyzed in the proof of Proposition 3.1. Specifically, S_4 and S_5 are similar to S_1; S_6 and S_7 are similar to S_2; and S_8, S_9 are similar to S_3. By following the same arguments as in the proof of Proposition 3.1, one can show that the results below are correct.

Category 1. (i) $S_4-c_2(n)=O_p(\gamma^2(n\gamma^{(l+1)/2})^{-1})=o_p((n\gamma^{(l+1)/2})^{-1})$, where

$$c_2(n)=\frac{1}{n\gamma^l}[\int K_2^2(x)dx]E[\{f(x_1)-h_2(t_1,x_1)\}\lambda^2(t_1|x_1,\beta_0)w_1d_1]. \tag{16}$$

Heuristically, S_4 is due to the random variation of $\hat{h}_2$ only and is thus similar to S_1. However because of the integration involved in the definition of $\hat{h}_2$, one obtains an extra γ^2 in the order of $[S_4-c_2(n)]$ which makes it of smaller order than $[S_1-c_1(n)]$. A proof of this result is provided in the Appendix, see Lemma A.3.

(ii) $S_5-c_3(n)=O_p(\gamma(n\gamma^{(l+1)/2})^{-1})=o_p((n\gamma^{(l+1)/2})^{-1})$, where

$$c_3(n)=\frac{1}{2n\gamma^l}[\int K_2^2(x)dx]E[h_1(t_1,x_1)\lambda(t_1|x_1,\beta_0)w_1d_1]. \tag{17}$$

The term S_5 arises from the interaction between the random variation of $\hat{h}_1$ and that of $\hat{h}_2$. Like S_4, the extra γ in the order of $[S_5-c_3(n)]$ is due to the integration involved in the expression for $\hat{h}_2$.

As a result of (i) and (ii) above, S_4 and S_5 only contribute to the center of the asymptotic null distribution of $S(\beta_0)$ by the term $[c_2(n) - 2c_3(n)]$.

Category 2. (i) $S_6 = O_p(\gamma^m n^{-1/2}) = o_p((n\gamma^{(l+1)/2})^{-1})$.

It is easy to see that S_6 is due to the interaction between the random variation of $\hat{h}_1$ and the bias of $\hat{h}_2$. Similar to the analysis of S_2, one can establish the stated order of S_6 by using Lemma A.2 (ii) and Lemma 2.1 in Powell, Stock, and Stoker (1989).

(ii) $S_7 = O_p(\gamma^m n^{-1/2}) = o_p((n\gamma^{(l+1)/2})^{-1})$.

Similar to S_6, S_7 is due to the interaction between the random variation of $\hat{h}_2$ and the bias of $\hat{h}_1$.

Category 3. (i) $S_8 = O_p(\gamma^{2m}) = o_p((n\gamma^{(l+1)/2})^{-1})$.

(ii) $S_9 = O_p(\gamma^{2m}) = o_p((n\gamma^{(l+1)/2})^{-1})$.

Both S_8 and S_9 involve only the bias of $\hat{h}_1$ and $\hat{h}_2$. Similar to S_3, one can establish the stated order of S_8 and S_9 by using Lemma A.2.

The conclusion in Proposition 3.2 follows immediately from (15), the above results, and Proposition 3.1. ▮

Finally, under the additional assumption (P), one can show by following Fan (1994) that $S(\hat{\beta})$ and $S(\beta_0)$ have the same asymptotic null distribution. Namely, we have

THEOREM 3.1. *Under (H_0), the assumptions (f), (K1), (K2), (G), and (P), the asymptotic distribution of* $n\gamma^{(l+1)/2}[S(\hat{\beta}) - c(n)]$ *is* $N(0, 2\sigma^2)$, *where* $c(n)$ *and* σ^2 *are defined in Propositions 3.1 and 3.2. In addition,* $\hat{c}(n) - c(n) = o_p(1)$ *and* $\hat{\sigma}^2 - \sigma^2 = o_p(1)$, *where* $\hat{c}(n) = \hat{c}_1(n) + \hat{c}_2(n) - 2\hat{c}_3(n)$ *with*

$$\begin{aligned}
\hat{c}_1(n) &= \frac{1}{n^2\gamma^{(l+1)}}\left[\int K_2^2(x)dx]\sum_{i=1}^n[\int_{-1}^{t_i/\gamma} K_{1t_i}^2(s)ds\right] w(t_i, x_i)\hat{h}_1(t_i, x_i)d_i, \\
\hat{c}_2(n) &= \frac{1}{n^2\gamma^l}\left[\int K_2^2(x)dx\right]\sum_{i=1}^n[\{\hat{f}(x_i) - \hat{h}_2(t_i, x_i)\}\hat{\lambda}^2(t_i|x_i)w_i d_i], \\
\hat{c}_3(n) &= \frac{1}{2n^2\gamma^l}\left[\int K_2^2(x)dx\right]\sum_{i=1}^n[\hat{h}_1(t_i, x_i)\hat{\lambda}(t_i|x_i)w_i d_i], \\
\hat{\sigma}^2 &= \left\{\frac{1}{n}\sum_{i=1}^n w^2(t_i, x_i)\hat{h}_1^3(t_i, x_i)d_i\right\}\left\{\int_0^\infty [K_{1+} * K_{1+}(1, s)]^2 ds\right\} \\
&\times \left\{\int [K_2 * K_2(y)]^2 dy\right\},
\end{aligned}$$

in which $\hat{f}(x_i)$ is the kernel estimator of the density function $f(x_i)$ of the covariate.

3.3. The Test Statistic

Based on Theorem 3.1, one can construct the following test statistic:

$$\hat{T} = \frac{n\gamma^{(l+1)/2}[S(\hat{\beta}) - \hat{c}(n)]}{\sqrt{2}\hat{\sigma}}. \tag{18}$$

Theorem 3.1 implies that under H_0, $\hat{T} \to N(0,1)$ in distribution. This forms the basis for the following one-sided asymptotic test for H_0: reject H_0 at significance level α if $\hat{T} > z_\alpha$, where z_α is the upper α-percentile of the standard normal distribution.

The last result of this section states the consistency of the above test.

THEOREM 3.2. *Suppose assumptions (f), (K1), (K2), (G), and (P) hold. Then the above test is consistent.*

Theorem 3.2 follows from the fact that under H_A, it holds that $\hat{\sigma} = O_p(1)$ and

$S(\hat{\beta}) \to \int\int [\lambda_0(t|x,\beta_*) - \lambda(t|x)]^2 h_2^2(t,x) w(t,x) h_1(t,x) dt dx$ in probability

which is positive. The proof of this is straightforward and thus omitted.

4. CONCLUSIONS

In this paper, we have proposed a consistent model specification test for the hazard function in the context of random right hand censoring of the dependent variable. It is based on the comparison of a kernel estimate and a parametric estimate of the hazard rate and hence belongs to the class of smoothing tests. Like most existing smoothing tests for the parametric specification of density and regression functions, the proposed test depends on the choice of the smoothing parameter. Versions of the test that are adaptive and optimal for the hazard rate might be constructed along the lines of adaptive and optimal tests for regression function in Horowitz and Spokoiny (2000). This is left for future research.

APPENDIX: ASSUMPTIONS

In this Appendix, we provide several lemmas that are used in the proof of the main result in the paper. Throughout, we assume that the assumptions in Section 3 hold. To simplify various expressions, we use $A \approx B$ to denote two quantities A and B satisfying $A/B = 1 + o_p(1)$.

LEMMA A.1. *$E[\bar{H}_{n1}^2(z_1, z_2)] = \gamma^{3(l+1)}\sigma^2 + o((\gamma^{3(l+1)}))$, where σ^2 is defined in Proposition 3.1 and $\bar{H}_{n1}(z_1, z_2)$ is defined via (13).*

Proof. From the definition of $\bar{H}_{n1}(z_1, z_2)$, it follows that

$$
\begin{aligned}
& E[\bar{H}_{n1}^2(z_1, z_2)] = \int_0^\infty \int_{-\infty}^\infty \int_0^\infty \int_{-\infty}^\infty \\
& E\{[d_2 K_{1t}(\frac{t-t_2}{\gamma})K_2(\frac{x-x_2}{\gamma}) - e_1(t,x)][d_1 K_{1t}(\frac{t-t_1}{\gamma})K_2(\frac{x-x_1}{\gamma}) - e_1(t,x)] \\
\times \; & [d_2 K_{1s}(\frac{s-t_2}{\gamma})K_2(\frac{y-x_2}{\gamma}) - e_1(s,y)][d_1 K_{1s}(\frac{s-t_1}{\gamma})K_2(\frac{y-x_1}{\gamma}) - e_1(s,y)]\} \\
\times \; & w(t,x)h_1(t,x)w(s,y)h_1(s,y)dxdtdyds \\
= \; & \int_0^\infty \int_{-\infty}^\infty \int_0^\infty \int_{-\infty}^\infty \\
& \{E[d_1 K_{1t}(\frac{t-t_1}{\gamma})K_2(\frac{x-x_1}{\gamma}) - e_1(t,x)][d_1 K_{1s}(\frac{s-t_1}{\gamma})K_2(\frac{y-x_1}{\gamma}) - e_1(s,y)]\}^2 \\
\times \; & w(t,x)h_1(t,x)w(s,y)h_1(s,y)dxdtdyds.
\end{aligned}
$$

Noting that $e_1(t,x) = E[d_1 K_{2,21}|t_2 = t, x_2 = x] = O(\gamma^{l+1})$, one can show based on the above expression that

$$
\begin{aligned}
& E[\bar{H}_{n1}(z_1, z_2)] \approx \int_0^\infty \int_{-\infty}^\infty \int_0^\infty \int_{-\infty}^\infty \\
& \{\int_0^\infty \int_{-\infty}^\infty [K_{1t}(\frac{t-t_1}{\gamma})K_2(\frac{x-x_1}{\gamma})][K_{1s}(\frac{s-t_1}{\gamma})K_2(\frac{y-x_1}{\gamma})] \\
\times \; & h_1(t_1,x_1)dx_1dt_1\}^2 w(t,x)h_1(t,x)w(s,y)h_1(s,y)dxdtdyds \\
= \; & \gamma^{2(l+1)} \int_0^\infty \int_{-\infty}^\infty \int_0^\infty \int_{-\infty}^\infty \\
& \{\int_{-1}^{t/\gamma} \int_{-\infty}^\infty K_{1t}(t_1)K_2(x_1)K_{1s}(\frac{s-t}{\gamma} + t_1)K_2(\frac{y-x}{\gamma} + x_1) \\
\times \; & h_1(t - \gamma t_1, x - \gamma x_1)dx_1dt_1\}^2 w(t,x)h_1(t,x)w(s,y)h_1(s,y)dxdtdyds \\
= \; & \gamma^{3(l+1)} \int_0^\infty \int_{-\infty}^\infty \int_{-1}^{s/\gamma} \int_{-\infty}^\infty \\
& \{\int_{-1}^{t/\gamma} \int_{-\infty}^\infty K_{1,s-\gamma t}(t_1)K_2(x_1)K_{1s}(t+t_1)K_2(x+x_1) \\
\times \; & h_1(s - \gamma t - \gamma t_1, y - \gamma x - \gamma x_1)dx_1dt_1\}^2 \\
\times \; & w(s-\gamma t, y - \gamma x)h_1(s - \gamma t, y - \gamma x)w(s,y)h_1(s,y)dxdtdyds \\
\approx \; & \gamma^{3(l+1)} \int_0^\infty \int_{-\infty}^\infty \int_{-1}^{s/\gamma} \int_{-\infty}^\infty \{\int_{-\infty}^\infty K_2(x_1)K_2(x+x_1)dx_1\}^2
\end{aligned}
$$

$$
\begin{aligned}
&\times \ \{\int_{-1}^{t/\gamma} K_{1,s-\gamma t}(t_1)K_{1s}(t+t_1)dt_1\}^2 w^2(s,y)h_1^4(s,y)dxdtdyds \\
&= \ \gamma^{3(l+1)}[\int_{-\infty}^{\infty}\{\int_{-\infty}^{\infty} K_2(x_1)K_2(x+x_1)dx_1\}^2dx] \\
&\times \ [\int_0^{\infty}\{\int_{-\infty}^{\infty} w^2(s,y)h_1^4(s,y)dy\}\int_{-1}^{s/\gamma}\{\int_{-1}^{t/\gamma} K_{1,s-\gamma t}(t_1)K_{1s}(t+t_1)dt_1\}^2dtds] \\
&= \ \gamma^{3(l+1)}[\int_{-\infty}^{\infty}\{\int_{-\infty}^{\infty} K_2(x_1)K_2(x+x_1)dx_1\}^2dx] \\
&\times \ [\int_0^{\infty}\int_{t\gamma}^{\infty}\{\int_{-\infty}^{\infty} w^2(s,y)h_1^4(s,y)dy\}\{\int_{-1}^{t/\gamma} K_{1,s-\gamma t}(t_1)K_{1s}(t+t_1)dt_1\}^2dsdt] \\
&= \ \gamma^{3(l+1)}[\int_{-\infty}^{\infty}\{\int_{-\infty}^{\infty} K_2(x_1)K_2(x+x_1)dx_1\}^2dx] \\
&\times \ [\int_0^{\infty}[\int_{t\gamma}^{t(\gamma+1)}\{\int_{-\infty}^{\infty} w^2(s,y)h_1^4(s,y)dy\}\{\int_{-1}^{t/\gamma} K_{1,s-\gamma t}(t_1)K_{1s}(t+t_1)dt_1\}^2ds \\
&+ \ \int_{t(\gamma+1)}^{\infty}\{\int_{-\infty}^{\infty} w^2(s,y)h_1^4(s,y)dy\}\{\int_{-1}^{t/\gamma} K_{1+}(1,t_1)K_{1+}(1,t+t_1)dt_1\}^2dsdt \\
&\approx \ \gamma^{3(l+1)}[\int_{-\infty}^{\infty}\{\int_{-\infty}^{\infty} K_2(x_1)K_2(x+x_1)dx_1\}^2dx] \\
&\times \ [\int_0^{\infty}\{\int_{-1}^{1} K_{1+}(1,t_1)K_{1+}(1,t+t_1)dt_1\}^2dt[\int_0^{\infty}\int_{-\infty}^{\infty} w^2(s,y)h_1^4(s,y)dyds\}
\end{aligned}
$$

∎

To derive the order of S_2, S_6, and S_7, we need to know the bias structures of $\hat{h}_1$ and $\hat{h}_2$. These are given in the following lemma.

LEMMA A.2. *Let $q = t/\gamma$ for $0 \le t < \gamma$ and $q = 1$ for $\gamma \le t < \infty$. Under (K1) and (K2), we get*

(i) $E[\hat{h}_1(t,x)] - h_1(t,x) = \gamma^m[k_{mq}\frac{\partial^m h_1(t,x)}{\partial t^m} + k_m\sum_{i=1}^{l}\frac{\partial^m h_1(t,x)}{\partial x_i^m}] + o(\gamma^m)$;

(ii) $E[\hat{h}_2(t,x)] - h_2(t,x) = \gamma^m[k_{mq}\frac{\partial^m h_2(t,x)}{\partial t^m} + k_m\sum_{i=1}^{l}\frac{\partial^m h_2(t,x)}{\partial x_i^m}] + o(\gamma^m)$.

Proof. The proof of (i) is straightforward. We will only prove (ii).

Let $\bar{K}_{1t}(\frac{t-t_j}{\gamma}) = \int_{-\infty}^{\frac{t-t_j}{\gamma}} K_{1t}(u)du$. Then $\bar{K}_{1t}(-\infty) = 0$ and

$$\hat{h}_2(t,x) = \frac{1}{(n-1)\gamma^l}\sum_{j\neq i}[1-\bar{K}_{1t}(\frac{t-t_j}{\gamma})]K_2(\frac{x-x_j}{\gamma}). \qquad \text{(A.1)}$$

Hence

$$E\hat{h}_2(t,x) = \frac{1}{\gamma^l}E\{[1-\bar{K}_{1t}(\frac{t-t_1}{\gamma})]K_2(\frac{x-x_1}{\gamma})\}$$

$$
\begin{aligned}
&= \frac{1}{\gamma^l} E[K_2(\frac{x-x_1}{\gamma})] - \frac{1}{\gamma^l} E[\bar{K}_{1t}(\frac{t-t_1}{\gamma}) K_2(\frac{x-x_1}{\gamma})] \\
&= f(x) + \gamma^m k_m \sum_{i=1}^{l} \frac{\partial^m f(x)}{\partial x_i^m} + o(\gamma^m) \\
&- \int \int_0^\infty \bar{K}_{1t}(\frac{t-t_1}{\gamma}) K_2(x_1)[d\{1 - F(t_1|x-\gamma x_1)G(t_1|x-\gamma x_1)\}] f(x-\gamma x_1) dx_1 \\
&= f(x) + \gamma^m k_m \sum_{i=1}^{l} \frac{\partial^m f(x)}{\partial x_i^m} + o(\gamma^m) \\
&- \frac{1}{\gamma} \int \int_0^\infty [1 - F(t_1|x-\gamma x_1)G(t_1|x-\gamma x_1)] K_{1t}(\frac{t-t_1}{\gamma}) K_2(x_1) f(x-\gamma x_1) dt_1 dx_1 \\
&= f(x) + \gamma^m k_m \sum_{i=1}^{l} \frac{\partial^m f(x)}{\partial x_i^m} + o(\gamma^m) \qquad \text{(A.2)} \\
&- [1 - F(t|x)G(t|x)] f(x) + \gamma^m [k_{mq} \frac{\partial^m h_2(t,x)}{\partial t^m} - k_m \sum_{i=1}^{l} \frac{\partial^m \{f(x) - h_2(t,x)\}}{\partial x_i^m}].
\end{aligned}
$$

The result follows from (A.2). ▌

Lemma A.3. *$S_4 - c_2(n) = O_p(\gamma^2 (n\gamma^{(l+1)/2})^{-1})$, where S_4 is defined via (15) and $c_2(n)$ is defined in (16).*

Proof. From (A.1), we get

$$
\hat{h}_2(t_i, x_i) = \frac{1}{(n-1)\gamma^l} \sum_{j \neq i} [1 - \bar{K}_{1t_i}(\frac{t_i - t_j}{\gamma})] K_2(\frac{x_i - x_j}{\gamma}).
$$

Let $\psi(t,x) = \gamma^l E[\hat{h}_2(t,x)]$. Then

$$
\begin{aligned}
S_4 &= \frac{1}{n} \sum_i [\hat{h}_2(t_i,x_i) - E_i \hat{h}_2(t_i,x_i)]^2 \lambda^2(t_i|x_i,\beta_0) w_i d_i \\
&= \frac{1}{n(n-1)^2\gamma^{2l}} \sum \sum \sum_{j \neq k \neq i} [\{1 - \bar{K}_{1t_i}(\frac{t_i - t_j}{\gamma})\} K_2(\frac{x_i - x_j}{\gamma}) - \psi(t_i,x_i)] \\
&\times [\{1 - \bar{K}_{1t_i}(\frac{t_i - t_k}{\gamma})\} K_2(\frac{x_i - x_k}{\gamma}) - \psi(t_i,x_i)] \lambda^2(t_i|x_i,\beta_0) w_i d_i \\
&+ \frac{1}{n(n-1)^2\gamma^{2l}} \sum \sum_{j \neq i} [\{1 - \bar{K}_{1t_i}(\frac{t_i - t_j}{\gamma})\} K_2(\frac{x_i - x_j}{\gamma}) \\
&- \psi(t_i,x_i)]^2 \lambda^2(t_i|x_i,\beta_0) w_i d_i \\
&= S_{41} + S_{42}, \qquad \text{(A.3)}
\end{aligned}
$$

where the definitions of S_{41} and S_{42} should be apparent from (A.3). Like S_{21}, one can rewrite S_{41} in terms of a U-statistic: $S_{41} = (3\gamma^{2l})^{-1} U_{n4}$, where

$$U_{n4} = \binom{n}{3}^{-1} \sum\sum\sum_{i<j<k} H_{n4}(z_i, z_j, z_k),$$

in which

$$\begin{aligned}
& H_{n4}(z_i, z_j, z_k) \\
= \ & [\{1 - \bar{K}_{1t_i}(\frac{t_i - t_j}{\gamma})\}K_2(\frac{x_i - x_j}{\gamma}) - \psi(t_i, x_i)] \\
\times \ & [\{1 - \bar{K}_{1t_i}(\frac{t_i - t_k}{\gamma})\}K_2(\frac{x_i - x_k}{\gamma}) - \psi(t_i, x_i)]\lambda^2(t_i|x_i, \beta_0)w_i d_i \\
+ \ & [\{1 - \bar{K}_{1t_j}(\frac{t_j - t_i}{\gamma})\}K_2(\frac{x_j - x_i}{\gamma}) - \psi(t_j, x_j)] \\
\times \ & [\{1 - \bar{K}_{1t_j}(\frac{t_j - t_k}{\gamma})\}K_2(\frac{x_j - x_k}{\gamma}) - \psi(t_j, x_j)]\lambda^2(t_j|x_j, \beta_0)w_j d_j \\
+ \ & [\{1 - \bar{K}_{1t_k}(\frac{t_k - t_i}{\gamma})\}K_2(\frac{x_k - x_i}{\gamma}) - \psi(t_k, x_k)] \\
\times \ & [\{1 - \bar{K}_{1t_k}(\frac{t_k - t_j}{\gamma})\}K_2(\frac{x_k - x_j}{\gamma}) - \psi(t_k, x_k)]\lambda^2(t_k|x_k, \beta_0)w_k d_k.
\end{aligned}$$

It is easy to show that $E[H_{n4}(z_i, z_j, z_k)|z_i] = 0$. Hence U_{n4} is a degenerate U-statistic. Similar to the analysis of U_{n1}, one has

$$\begin{aligned}
S_{41} &= \frac{1}{3\gamma^{2l}}[\frac{6}{n(n-1)}\sum\sum_{i<j} E\{H_{n4}(z_i, z_j, z_k)|z_i, z_j\}] \\
&= \frac{2}{n(n-1)\gamma^{2l}}\sum\sum_{i<j} E([\{1 - \bar{K}_{1t_k}(\frac{t_k - t_i}{\gamma})\}K_2(\frac{x_k - x_i}{\gamma}) - \psi(t_k, x_k)] \\
&\times [\{1 - \bar{K}_{1t_k}(\frac{t_k - t_j}{\gamma})\}K_2(\frac{x_k - x_j}{\gamma}) - \psi(t_k, x_k)]\lambda^2(t_k|x_k, \beta_0)w_k d_k|z_i, z_j) \\
&= \frac{2}{n(n-1)\gamma^{2l}}\sum\sum_{i<j}\int_0^\infty\int_{-\infty}^\infty[\{1 - \bar{K}_{1t}(\frac{t - t_i}{\gamma})\}K_2(\frac{x - x_i}{\gamma}) - \psi(t, x)] \\
&\times [\{1 - \bar{K}_{1t}(\frac{t - t_j}{\gamma})\}K_2(\frac{x - x_j}{\gamma}) - \psi(t, x)] \\
&\times \lambda^2(t|x, \beta_0)w(t, x)h_1(t, x)dxdt \\
&= \frac{2}{n(n-1)\gamma^{2l}}\sum\sum_{i<j}\bar{H}_{n4}(z_i, z_j). \qquad \text{(A.4)}
\end{aligned}$$

Similar to the proof of Lemma A.1, one gets

$$E[\bar{H}_{n4}^2(z_i, z_j)] = \int_0^\infty\int_{-\infty}^\infty\int_0^\infty\int_{-\infty}^\infty$$

$$
\begin{aligned}
& \Big[E[\{1-\bar{K}_{1t}(\frac{t-t_i}{\gamma})\}K_2(\frac{x-x_i}{\gamma})-\psi(t,x)] \\
\times\ & [\{1-\bar{K}_{1s}(\frac{s-t_i}{\gamma})\}K_2(\frac{y-x_i}{\gamma})-\psi(s,y)]]\Big]^2 \\
\times\ & \lambda^2(t|x,\beta_0)w(t,x)h_1(t,x)\lambda^2(s|y,\beta_0)w(s,y)h_1(s,y)dxdtdyds \\
=\ & \int_0^\infty\int_{-\infty}^\infty\int_0^\infty\int_{-\infty}^\infty[\int_0^\infty\int_{-\infty}^\infty \\
& [\{1-\bar{K}_{1t}(\frac{t-t_i}{\gamma})\}K_2(\frac{x-x_i}{\gamma})-\psi(t,x)] \\
\times\ & [\{1-\bar{K}_{1s}(\frac{s-t_i}{\gamma})\}K_2(\frac{y-x_i}{\gamma})-\psi(s,y)]]f(x_1)dx_1 \\
\times\ & d\{1-F(t_1|x_1)G(t_1|x_1)\}]^2\lambda^2(t|x,\beta_0)w(t,x)h_1(t,x) \\
\times\ & \lambda^2(s|y,\beta_0)w(s,y)h_1(s,y)dxdtdyds \\
=\ & \gamma^{2(l+1)}\int_0^\infty\int_{-\infty}^\infty\int_0^\infty\int_{-\infty}^\infty[\int_{-1}^{t/\gamma}\int_{-\infty}^\infty \\
& [\{1-\bar{K}_{1t}(t_1)\}K_2(x_1)-\psi(t,x)] \\
\times\ & [\{1-\bar{K}_{1s}(\frac{s-t}{\gamma}+t_1)\}K_2(\frac{y-x}{\gamma}+x_1)-\psi(s,y)]] \\
\times\ & f(x-\gamma x_1)dx_1d\{1-F(t-\gamma t_1|x-\gamma x_1)G(t-\gamma t_1|x-\gamma x_1)\}]^2 \\
\times\ & \lambda^2(t|x,\beta_0)w(t,x)h_1(t,x)\lambda^2(s|y,\beta_0)w(s,y)h_1(s,y)dxdtdyds \\
=\ & O(\gamma^{3(l+1)}). \qquad \text{(A.5)}
\end{aligned}
$$

Hence $Var(S_{41})=O((n^4\gamma^{4l})^{-1}n^2\gamma^{3(l+1)})=O(\gamma^4(n^2\gamma^{l+1})^{-1})$.

The center of S_4 is given by the dominating term of $E(S_{42})$. We now show that $E(S_{42})\approx c_2(n)$. Noting that $\psi(t,x)=O(\gamma^l)$, we have

$$
E(S_{42})\approx\frac{1}{(n-1)\gamma^{2l}}E\{[1-\bar{K}_{1t_1}(\frac{t_1-t_2}{\gamma})]^2K_2^2(\frac{x_1-x_2}{\gamma})\lambda^2(t_1|x_1,\beta_0)w_1d_1\}. \qquad \text{(A.6)}
$$

Now,

$$
\begin{aligned}
& E_1\{[1-\bar{K}_{1t_1}(\frac{t_1-t_2}{\gamma})]^2K_2^2(\frac{x_1-x_2}{\gamma})]= \\
=\ & \int_{-\infty}^\infty\int_0^\infty[1-\bar{K}_{1t_1}(\frac{t_1-t}{\gamma})]^2d\{1-F(t|x)G(t|x)\}K_2^2(\frac{x_1-x}{\gamma})f(x)dx \\
\approx\ & 2\int_{-\infty}^\infty\{\int_{-\infty}^{t/\gamma}[1-\bar{K}_{1t_1}(t)]K_{1t_1}(t)dt\}\{1-F(t_1|x)G(t_1|x)\}K_2^2(\frac{x_1-x}{\gamma})f(x)dx \\
\approx\ & \gamma^l\int_{-\infty}^\infty\{1-F(t_1|x_1-\gamma x)G(t_1|x_1-\gamma x)\}K_2^2(x)f(x_1-\gamma x)dx
\end{aligned}
$$

$$\approx \gamma^l[1 - F(t_1|x_1)G(t_1|x_1)]f(x_1) \int K_2^2(x)dx. \tag{A.7}$$

The conclusion follows immediately from (A.6) and (A.7). ∎

REFERENCES

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

Fan, Y., 1994, Testing the goodness-of-fit of a parametric density function by kernel method. *Econometric Theory* **10 (2)**, 316-356.

Fan, Y. and Q. Li, 1996, Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* **64**, 865-890.

Gozalo, P. L., 1993, A consistent model specification test for nonparametric estimation of regression function models. *Econometric Theory* **9**, 451-477.

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

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

Hart, J. D., 1997, *Nonparametric smoothing and lack-of-fit tests.* New York: Springer-Verlag.

Hong, Y., 2000, Consistent specification testing using optimal nonparametric kernel estimation. Working paper, Cornell University.

Hong, Y. and H. White, 1995, Consistent specification testing via nonparametric series regression. *Econometrica* **63**, 1133-1160.

Horowitz, J. T. and W. Härdle, 1994, Testing a parametric model against a semiparametric alternative. *Econometric Theory* **10**, 821-848.

Horowitz, J. L. and G. R. Neumann, 1992, A generalized moments specification test of the proportional hazards model. *Journal of the American Statistical Association* **87**, 234-240.

Horowitz, J. L. and V. G. Spokoiny, 2000, An adaptive, rate-optimal test of a parametric model against a nonparametric alternative. Forthcoming in *Econometrica.*

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

Müller, Hans-Georg, 1991, Smooth optimum kernel estimators near endpoints. *Biometrika* **78**, 521-530.

Nakamura, A. and J. R. Walker, 1994, Model evaluation and choice. *The Journal of Human Resources* **XXIX**, 223-247.

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

White, H., 1982, Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1-25.

Wooldridge, J., 1992, A test for functional form against nonparametric alternatives. *Econometric Theory* **8**, 452-475.

Yatchew, A. J., 1992, Nonparametric regression tests based on least squares. *Econometric Theory* **8**, 435-451.

Zheng, J. X. , 1996, A consistent test of functional form via nonparametric estimation technique. *Journal of Econometrics* **75**, 263-289.