Partial Identification and Inference in Censored Quantile Regression: A Sensitivity Analysis

Yanqin Fan and Ruixuan Liu[*]
University of Washington, Seattle

Working Paper no. 141
Center for Statistics and the Social Sciences
University of Washington

First version: June 2013
This version: December 2013

Abstract

In this paper we characterize the identified set and construct asymptotically valid and non-conservative confidence sets for the quantile regression coefficient in a linear quantile regression model, where the dependent variable is subject to possibly dependent censoring. The underlying censoring mechanism is characterized by an Archimedean copula for the dependent variable and the censoring variable. For a broad class of Archimedean copulas, we characterize an outer set of the corresponding identified set for the quantile regression coefficient via inequality constraints. For one-parameter ordered families of Archimedean copulas, we construct simple confidence sets by inverting asymptotically pivotal statistics related to kernel-based model specification testing. The methodology we develop in this paper allows practitioners to conduct sensitivity analysis of the robustness of conclusions on the quantile regression coefficient to the independent censoring mechanism. Bootstrap confidence sets are also constructed. Interpreting the dependent variable and the censoring variable in our censored quantile regression model as two competing risks, our methodology is useful in duration analysis with possibly dependent competing risks. We present an empirical application to the survival time after acute myocardial infarction.

Keywords: *Archimedean Copula, Competing Risks Model, Confidence Set, Dependent Censoring, Degenerate U-statistics, Independent Censoring, Mixed Type Regressor*
JEL Codes: C12, C14, C34, C41, C51

1 Introduction

1.1 Quantile Regression With Dependent Censoring

Since the seminal work of Koenker and Bassett (1978) who propose to use linear quantile regression to examine effects of an observable covariate on the distribution of a dependent variable other than the mean, linear quantile regression has become a dominant approach in empirical work in economics, see e.g., Buchinsky (1994) and Koenker (2005). For $q \in (0,1)$, a linear q-th quantile regression model takes the following form:

$$Q_{Yo}(q|x) = x'\beta_o \tag{1.1}$$

where $Q_{Yo}(q|x)$ denotes the q-th conditional quantile of the dependent variable Y given $X = x$ with X the observable vector of covariates. In many applications in economics, the dependent variable Y is censored by a censoring variable denoted as C. So instead of observing the variable Y, the econometrician observes the triplet (V, X, D) with $V \equiv \min(Y, C)$ and $D \equiv I\{Y < C\}$.

Existing work in the literature on the identification and inference in censored linear quantile regression models either assume the independent censoring mechanism[1] — that is, Y and C are independent (conditional on the covariate X), or make no assumption on the true censoring mechanism at all. Work in the former category include Buchinsky and Hahn (1998), Honore, Khan and Powell (2002), and Chernozhukov and Hong (2002), Portnoy (2003), Peng and Huang (2008) and Wang and Wang (2009), among others;[2] and Powell (1984, 1986) and Khan and Powell (2001) who adopt a special case of the independent censoring, i.e., the fixed known censoring mechanism. Under additional conditions including a rank condition, β_o is point identified in the case of independent censoring and the aforementioned work develop estimation and inference procedures for it. Work in the latter category include Khan and Tamer (2009), Khan, Ponomareva and Tamer (2011) who show that the quantile coefficient β_o is not point identified when no assumption is made on the true censoring mechanism and establish the identified set for β_o. In addition, Khan and Tamer (2009) develop confidence sets (CSs) for β_o when it is point identified.

The independent censoring mechanism is often violated in empirical applications, but on the other hand, the researcher typically has some information on the true censoring mechanism (e.g., Y and C may be known to be positively dependent), or may want to check robustness of conclusions to moderate deviations from independent censoring. The first objective of this paper is to develop inference procedures for the quantile coefficient β_o when partial information on the true censoring mechanism such as positive dependence is available. The second objective is to develop methods for examining sensitivity of conclusions on β_o reached under the independent censoring mechanism to deviations from it. To accomplish both objectives in a unified framework, we model the true censoring mechanism via an Archimedean copula for Y and C (conditional on X) and allow its generator function to vary in a pre-specified class. For a given class of Archimedean copulas, we propose a two-step approach to the identification of β_o. The first step extends the existing result

---

[1] Throughout this paper, we use the independent censoring mechanism to denote the conditional independent censoring mechanism which reduces to the unconditional independent censoring mechanism when there is no covariate.

[2] Some such as Chernozhukov and Hong (2002) assume that the censoring variable C is always observed.

1

in Rivest and Wells (2001) and Braekers and Veraverbeke (2005) which expresses the conditional distribution function of Y given X in terms of the generator function and functions that are identified from the sample information. In the second step, we make use of this result and the linear quantile regression (1.1) to establish the identified set for β_o when the generator function varies in the pre-specified class of generator functions. One interesting finding is that for a broad class of Archimedean copulas, β_o satisfies inequality constraints characterized by functionals of the conditional distribution function of V and the sub-distribution function of V.

Our identification strategy is well suited for sensitivity analysis to a *known* censoring mechanism. For example, to conduct a sensitivity analysis to independent censoring, we make use of global measures of dependence such as Kendall's τ to measure deviations from the independent censoring mechanism; A value of zero for Kendall's τ corresponds to independent censoring, while a value of one for the magnitude of Kendall's τ corresponds to perfectly dependent censoring—Y and C are perfectly dependent conditional on X. We develop a simple sensitivity analysis to independent censoring by adopting parametric Archimedean copulas. Most parametric Archimedean copulas are characterized by a single parameter and are ordered according to the concordance ordering. Because of the one-to-one relation between Kendall's τ and the copula parameter, the identified sets for β_o corresponding to bounded ranges of the copula parameter of such ordered parametric Archimedean copulas provide bases for examining the sensitivity of conclusions on β_o to the independent censoring mechanism. To formalize this procedure, we construct asymptotically valid and non-conservative confidence sets (CSs) for β_o for any pre-specified range of values of the copula parameter. The general idea underlying our CSs comes from the observation that for a given generator function, the closed-form expression for the conditional distribution function of Y given X established in the first step and (1.1) imply that β_o must satisfy some equality constraints. Although the true generator function is unknown, for any β in the identified set, there must be at least one generator in the pre-specified class that ensures such equality constraints to hold. The problem of constructing CSs for β_o is thus equivalent to a series of 'specification testing' problems; for each β in the parameter space, we test the correct specification of the copula or generator class and if the copula class is correctly specified in the sense that there exists at least one copula or generator such that the equality constraints hold, then β is in the CS for β_o; otherwise it is not. We construct two test statistics similar to test statistics for consistent model specification testing based on kernel estimators, see Fan (1994), Fan and Li (1996), and Zheng (1996) and many subsequent works in the literature. For most one-parameter ordered families of Archimedean copulas, we show that under an appropriate condition, for each β in the identified set, there exists a *unique* value of the copula parameter that ensures the equality constraints to hold. This ensures that for each β in the identified set, our test statistics are asymptotically normally distributed leading to asymptotically valid and non-conservative CSs that are easy to implement. We also develop bootstrap CSs and present an empirical application to the survival time after acute myocardial infarction.

1.2 A Semiparametric Competing Risks Model and Some Related Works

Interpreting Y and the censoring variable C in our model as two competing risks, this paper proposes a new semiparametric competing risks model. Applications of the competing risks model in economics include Flinn and Heckman (1982) who investigate the duration of unemployment where an individual can exit unemployment either by finding a job or by leaving the labor market; Katz and Meyer (1990) who study the probability of leaving unemployment through recalls and new jobs; Berrington and Diamond (2000) who study age at marriage or cohabitation; Booth and Satchell (1995) who study Ph.D. completion; Deng, Quigley, and Van Order (2000) who investigate mortgage termination; and Honore and Lleras-Muney (2006) who study changes in cancer and cardiovascular mortality since 1970.

Identification analysis of competing risks models has a long history dating back to Cox (1959). Tsiatis (1975) uses an explicit construction to demonstrate the non-identifiability of the marginal distribution function of Y once the independent censoring mechanism is dispensed with. Crowder (1991) further amplifies this identifiability crisis by proving that even if the two marginal distribution functions of (Y, C) are given, their joint distribution still remains unidentified generally. Peterson (1976) obtains the worst-case bounds for both the marginal distribution function of Y and the joint distribution function of Y and C.

In response to the identifiability crisis, two general approaches have been taken in the literature to achieve point identification of a competing risks model. First, covariate information and specific model structures imposed on the marginal distributions may restore point identification, see e.g., the proportional hazards and accelerated failure time models in Heckman and Honore (1989), Abbring and van den Berg (2003), Lee (2006), and Lee and Lewbel (2012); Second, assuming a *known* copula for the individual risks, Zheng and Klein (1995) first extend point identification results for independent risks to dependent risks and propose a copula-graphic estimator of the marginal survival function. When the copula function is Archimedean and *known*, Rivest and Wells (2001) first derive an explicit expression for the copula-graphic estimator of the survival function proposed in Zheng and Klein (1995). In addition to establishing uniform consistency and asymptotic normality of the copula-graphic estimator, Rivest and Wells (2001) also study asymptotic properties of the copula-graphic estimator under misspecification of the true Archimedean copula. Braekers and Veraverbeke (2005) extend Rivest and Wells (2001) to the fixed design regression.

This paper contributes to the competing risks literature in several ways. First, the duration of the competing risk C is left unspecified in our model[3] and thus inference on the conditional quantile of Y is robust to possible misspecification of the marginal model for the competing risk C. Moreover our inference procedures do not reply on conditions ensuring point identification of β_o and thus allow for the presence of general covariates in the marginal model for the risk of interest Y; Second, we don't impose a *known* copula on the individual risks, instead we allow the true copula to vary in a prespecified class of Archimedean copulas

---

[3]Independently, Szydlowski (2013) studies partial identification of the proportional hazards model for the risk of interest in a competing risks model without specifying the marginal model for the competing risk. Like Khan and Tamer (2011), Szydlowski (2013) makes no assumption on the true censoring mechanism. Using an outer set of the identified set for the parameter in a parametric proportional hazards model, Szydlowski (2013) applies existing inference procedures to constructing CSs for the parameter in his model.

and develop a formal approach to conducting inference and sensitivity analysis to the independent censoring mechanism. Informal sensitivity analysis has been performed in the context of competing risks models including marginal survival or hazard function (Slud and Rubinstein, 1983; Zheng and Klein, 1995; Rivest and Wells, 2001; Klein and Moeschberger, 1988); Cox regression (Huang and Zhang, 2008); and a general semiparametric transformation model (Chen, 2010). In contrast to the current paper, however, existing work first establish the consistency and asymptotic normality of the proposed estimators for a given dependence structure between Y and C conditional on X and then examine the sensitivity of the proposed estimators or inference procedures to independent censoring by selecting a few prespecified dependence structures. Lastly, we propose a novel two-step identification strategy for β_o or the marginal model for the risk of interest Y. Our identification strategy is very general and not specific to the linear quantile model (1.1), instead it is applicable to any parametric model for Y including the proportional harzeds model and marginal regression model, see Remark 2.7 for a detailed discussion.

Besides Khan and Tamer (2009), our paper is related to Honore and Lleras-Muney (2006) and Kline and Santos (2013). Assuming accelerated failure time models for each risk, Honore and Lleras-Muney (2006) derive bounds for aspects of the underlying distributions allowing for dependent risks with interval outcome data and discrete covariates. Both Khan and Tamer (2009) and Honore and Lleras-Muney (2006) are agnostic about the true censoring mechanism. Kline and Santos (2013) develop methods for conducting a sensitivity analysis in the context of missing data. They measure the degree of departure from missing-at-random by using the maximal Kolmogorov-Smirnov distance between the distributions of missing and observed outcomes across all values of the covariates. We refer readers to Henry and Mourifie (2012) for a partial identification analysis of the binary Roy model and other work on Roy models.

The subsequent sections are organized as the following: Section 2 first introduces our identification strategy for β_o when the true copula belongs to a given class of Archimedean copulas and then presents the identified set for β_o when the class of Archimedean copulas is ordered. In Section 3, we present two asymptotic CSs for β_o and their asymptotic validity is shown in Section 4. We also construct bootstrap CSs in Section 4. Section 5 presents an empirical application on the survival time after acute myocardial infarction. The Appendices containing all the proofs are further divided into three sections. Appendix A shows the asymptotic linear representation of the plug-in estimator of the conditional distribution function of Y given X for a given generator function. The main theorems and the validity of our confidence sets are proved in Appendix B. In Appendix C we collect a variety of auxiliary results used in Appendices A and B.

2 The General Framework and Partial Identification of β_o

We first introduce some notations used throughout the paper. Let $F_{Y_o}(y|x)$, $F_C(c|x)$, and $F_{Y,C}(y,c|x)$ denote respectively the conditional marginal and joint distribution functions of (Y,C) given $X=x$, with the corresponding conditional survival functions $S_{Y_o}(y|x)$, $S_C(c|x)$, and $S_{Y,C}(y,c|x)$. Also let $F_{V,D=1}(v|x)$ and $F_{V,D=0}(v|x)$ denote the two conditional sub distribution functions, summing up to $F_V(v|x)$. The marginal

distribution function of X is denoted as $F_X(x)$, supported on $\mathcal{X}$.

Let $\mathcal{C}_{xo}(u,v) : [0,1]^2 \mapsto [0,1]$ denote the conditional survival copula of (Y,C) given $X = x$. Following Braekers and Veraverbeke (2005), we assume that $\mathcal{C}_{xo}$ is Archimedean with generator function φ_{xo}, i.e., for $(u,v) \in [0,1]^2$,

$$\mathcal{C}_{xo}(u,v) = \varphi_{xo}^{[-1]}\left[\varphi_{xo}(u) + \varphi_{xo}(v)\right],$$

where $\varphi_{xo} : [0,1] \to [0,\infty]$ is a continuous, convex, strictly decreasing function with $\varphi_{xo}(1) = 0$. Here, $\varphi_{xo}^{[-1]}$ denotes the pseudo-inverse of φ_{xo} defined by

$$\varphi_{xo}^{[-1]}(u) = \left\{ \begin{array}{ll} \varphi_{xo}^{-1}(u), & 0 \le u \le \varphi_{xo}(0) \\ 0, & \varphi_{xo}(0) \le u \le +\infty \end{array} \right..$$

We say $\mathcal{C}_{xo}$ is strict if its generator function φ_{xo} is strict, i.e., $\varphi_{xo}(0) = +\infty$.

Archimedean copulas have many nice properties, see Joe (1997) and Nelsen (2006). They arise naturally from shared frailty models as in Clayton and Cuzick (1985), Heckman and Honore (1989). Specifically, if the conditional hazard density functions of Y and C denoted as λ_{Yo} and λ_{Co} are specified by the corresponding conditional baseline hazard functions and a multiplicative frailty term ω as:

$$\begin{aligned} \lambda_{Yo}(t|x,\omega) &= \omega\lambda_{Yo}(t|x) \text{ and} \\ \lambda_{Co}(t|x,\omega) &= \omega\lambda_{Co}(t|x), \end{aligned} \tag{2.1}$$

then it is well known that the conditional survival copula would be Archimedean with the (inverse of) generator $\varphi_{xo}^{-1} = \mathcal{L} \circ F_{\omega|x}$, the Laplace transform of the conditional distribution of frailty denoted as $F_{\omega|x}$. The complete monotonicity induced by the Laplace transform ensures that the generator function φ_{xo} satisfies the requirement to produce a copula function (Joe, 1997).

Braekers and Veraverbeke (2005) show that in competing risks models where Y and C are survival variables with support $(0,\infty)$, if $\mathcal{C}_{xo}$ is *known*, then $F_{Yo}(\cdot|x)$ is identified from the sample information extending the well-known identification result of competing risks models under the independent censoring mechanism. The latter is obtained when $\varphi_{xo}(u) = \log(1/u)$ for all x under consideration. More importantly they establish a closed-form expression for $F_{Yo}(y|x)$ in terms of φ_{xo} and functions that are identified from the sample information, see their Lemma 1 or Lemma 2.1 below. Based on this expression, they construct an estimator of $F_{Yo}(y|x)$ referred to as the copula-graphic estimator and establish its asymptotic properties. When the true copula is a Clayton copula, Klein and Moeschberger (1988) establish this result and derive bounds on $F_{Yo}(y|x)$ for a specified range of values for the copula parameter.

Our identification analysis of β_o builds on a slight extension of Lemma 1 in Braekers and Veraverbeke (2005) which will be presented in the subsection below followed by a detailed analysis of identification of β_o.

2.1 A Two-Step Approach to the Identification of β_o

2.1.1 Step 1. Identification of $F_{Yo}(\cdot|x)$

Throughout this section, we assume that $x \in \mathcal{X}$ is fixed.

Assumption (AC). The true copula $\mathcal{C}_{xo}$ is a strict Archimedean copula.

Assumption (SY). (i) Let the support of $F_{Yo}(\cdot|x)$ be $[y_{lx}, y_{ux}] \subseteq \mathcal{R}$. The functions $F_{V,D=1}(\cdot|x)$, $F_V(\cdot|x)$ and $F_{Yo}(\cdot|x)$ have continuous (sub) densities in $[y_{lx}, v_{ux}]$, where v_{ux} is the right end point of the support of $F_V(\cdot|x)$; (ii) $y_{ux} = v_{ux}$.

Assumption (AC) is an assumption on the true copula. Assumption (SY) imposes support assumptions on Y apart from some smoothness assumptions on the stated distribution functions. Assumption (SY) (ii) is needed only when one is interested in identifying the whole distribution function $F_{Yo}(\cdot|x)$. If $v_{ux} < y_{ux}$, we never observe anything beyond v_{ux} for Y, so would not expect to identify $F_{Yo}(\cdot|x)$ on $[v_{ux}, y_{ux}]$ even when φ_{xo} is known. This potential tail problem is also present under the independent censoring assumption (see Fleming and Harrington, 1991; Gine and Guillou, 2001) and with general copula graphic identification (see Corollary 3.3 in Zheng and Klein, 1995). In Braekers and Veraverbeke (2005), Y and C are survival variables assumed to have common support $(0, \infty)$ so that $y_{lx} = 0$ and $y_{ux} = v_{ux} = \infty$. We allow for finite y_{ux}, but assume $y_{ux} = v_{ux}$. When one is only interested in some functional or feature of the distribution function $F_{Yo}(\cdot|x)$ such as the quantile coefficient in (1.1), identification of the whole distribution function $F_{Yo}(\cdot|x)$ may not be needed and Assumption (SY) (ii) may thus be dropped, see Remark 2.5 for further elaboration.

Under Assumption (AC), $S_{Y,C}(y, c|x)$ can be written as

$$S_{Y,C}(y, c|x) = \varphi_{xo}^{-1}\left[\varphi_{xo}\{S_{Yo}(y|x)\} + \varphi_{xo}\{S_C(c|x)\}\right]. \tag{2.2}$$

By setting $y = c = v$ in (2.2), we get

$$S_V(v|x) = \varphi_{xo}^{-1}\left[\varphi_{xo}\{S_{Yo}(v|x)\} + \varphi_{xo}\{S_C(v|x)\}\right]. \tag{2.3}$$

Using (2.2) and (2.3), Braekers and Veraverbeke (2005) show in their Lemma 1 that when $y_{lx} = 0$ and $y_{ux} = v_{ux} = \infty$, under mild conditions, the conditional cdf $F_{Yo}(\cdot|x)$ is point identified from the sample information as long as the generator function φ_{xo} is *known* and more importantly they provide a closed-form expression for $F_{Yo}(y|x)$. For completeness, we will restate their result in the lemma below under Assumptions (AC) and (SY). Since the proof is short, we will reproduce it as well to illustrate the roles of Assumptions (AC) and (SY).

Lemma 2.1 *(Braekers and Veraverbeke, Lemma 1) Suppose Assumptions (AC) and (SY) hold. If φ'_{xo} exists and is continuous on $(0, 1]$, then $\forall y \in [y_{lx}, y_{ux}]$, we have:*

$$F_{Yo}(y|x) = 1 - \varphi_{xo}^{-1}\left(-\int_{y_{lx}}^{y} \varphi'_{xo}\{S_V(v|x)\} dF_{V,D=1}(v|x)\right). \tag{2.4}$$

Proof. First, we take care of the two boundary points. When $y = y_{lx}$, both sides of (2.4) will equal to zero. When $y = y_{ux}$, the left hand side of (2.4) will be equal to 1. We distinguish between two cases for the right hand side of (2.4). First, if $\int_{y_{lx}}^{y_{ux}} \varphi'_{xo}\{S_V(v|x)\} dF_{V,D=1}(v|x) = -\infty$, then $\varphi_{xo}^{-1}(\infty) = 0$ by definition and

the right hand side of (2.4) equals 1 as well. Second, if $-\int_{y_{lx}}^{y_{ux}} \varphi'_{xo} \{S_V(v|x)\} dF_{V,D=1}(v|x) < \infty$, then the same derivation below for $y \in (y_{lx}, y_{ux})$ will apply here.

It follows from Tsiatis (1975) that

$$F'_{V,D=1}(v|x) = -\frac{\partial}{\partial y} S_{Y,C}(y,c|x)|_{y=c=v}. \tag{2.5}$$

For any $y \in (y_{lx}, y_{ux})$, $S_V(y|x) > 0$ by the continuity property assumed in Assumption (AC). Hence φ_{x_o} and φ'_{x_o} come into play over their properly defined domain $(0,1]$. By (2.2) and the simple fact that $\frac{\partial}{\partial u} \mathcal{C}_{xo}(u,v) = \varphi'_{x_o}(u)/\varphi'_{x_o}(\mathcal{C}_{xo}(u,v))$ (see Chapter 5 in Nelsen, 2006), we get

$$\frac{\partial}{\partial y} S_{Y,C}(y,c|x)|_{y=c=v} = -\frac{\partial}{\partial t} \mathcal{C}_{xo}(S_{Yo}(y|x), S_C(c|x))|_{y=c=v}$$

$$= -\frac{\varphi'_{xo}\{S_{Yo}(v|x)\} F'_{Yo}(v|x)}{\varphi'_{xo}\{S_V(v|x)\}}.$$

So

$$F'_{V,D=1}(v|x) = \frac{\varphi'_{xo}\{S_{Yo}(v|x)\} F'_{Yo}(v|x)}{\varphi'_{xo}\{S_V(v|x)\}} \tag{2.6}$$

leading to

$$\int_{y_{lx}}^{y} \varphi'_{xo}\{S_{Yo}(v|x)\} F'_{Yo}(v|x) dv = \int_{y_{lx}}^{y} \varphi'_{xo}\{S_V(v|x)\} dF_{V,D=1}(v|x)$$

or

$$-\int_{y_{lx}}^{y} d\varphi_{xo}\{S_{Yo}(v|x)\} = \int_{y_{lx}}^{y} \varphi'_{xo}\{S_V(v|x)\} dF_{V,D=1}(v|x)$$

or

$$-\varphi_{xo}\{S_{Yo}(y|x)\} + \varphi_{xo}\{S_{Yo}(y_{lx}|x)\} = \int_{y_{lx}}^{y} \varphi'_{xo}\{S_V(v|x)\} dF_{V,D=1}(v|x).$$

The result or (2.4) follows from the above equation by noting that $S_{Yo}(y_{lx}|x) = 1$, $\varphi_{xo}(1) = 0$, and φ_{xo} is strict. ∎

Remark 2.1 *Braekers and Veraverbeke (2005) assume that φ'_{xo} exists on $[0,1]$ in their Lemma 1. However most commonly used generator functions do not have a finite φ'_{xo} at 0, see for example those listed in Table 1. The additional continuity assumption we impose on φ'_{xo} in Lemma 2.1 guarantees that the Stieltjes integral in (2.4) is well defined.*

Remark 2.2 *Lemma 2.1 implies that if φ_{xo} is known, then $F_{Yo}(\cdot|x)$ is point identified and has a closed-form expression. When the true copula function $\mathcal{C}_{xo}(u,v)$ is not known to be Archimedean, a straightforward extension of Theorem 3.1 and Corollary 3.2 in Zheng and Klein (1995) to allow for the covariate X implies that under mild conditions, $F_{Yo}(\cdot|x)$ is point identified from the sample information as well. However, no explicit expression for $F_{Yo}(y|x)$ is available at such a general level.*

2.1.2 Step 2. Identification of β_o

Suppose Assumptions (AC) and (SY) hold for all $x \in \mathcal{J} \subseteq \mathcal{X}$. For each $x \in \mathcal{J}$, Lemma 2.1 expresses the true cdf $F_{Yo}(y|x)$ in terms of the copula generator function φ_{xo} and functions that are identified from the sample information. In practice, the true copula or generator function is *unknown*. Let Φ_x denote a prespecified class of strict generator functions. Lemma 2.1 or (2.4) allows us to establish the identified set for $F_{Yo}(\cdot|x)$. Specifically, for a strict generator function $\varphi_x \in \Phi_x$, let

$$F_Y(y|x;\varphi_x) = 1 - \varphi_x^{-1}\left(-\int_{y_{lx}}^{y} \varphi_x'\{S_V(v|x)\}\, dF_{V,D=1}(v|x)\right), \ y \in [y_{lx}, y_{ux}] \tag{2.7}$$

and $\mathcal{F}_I(x)$ denote the identified set for $F_{Yo}(\cdot|x)$. Then it follows immediately from Lemma 2.1 or (2.4) that

$$\mathcal{F}_I(x) = \{F_Y(\cdot|x) : F_Y(\cdot|x) = F_Y(\cdot|x;\varphi_x) \text{ for some } \varphi_x \in \Phi_x\}. \tag{2.8}$$

The identified set for the quantile regression coefficient β_o can be deduced from the identified set for $F_{Yo}(\cdot|x)$ and (1.1). Let $\mathcal{B}_I$ denote the identified set for β_o. Then

$$\mathcal{B}_I = \{\beta \in \mathcal{B} : F_Y(x'\beta|x;\varphi_x) = q \text{ for some } \varphi_x \in \Phi_x \text{ and all } x \in \mathcal{J}\}. \tag{2.9}$$

Different choices of the generator class Φ_x reflect either the researcher's prior knowledge of the true censoring mechanism or represent deviations from independent censoring in a sensitivity analysis. The identified set $\mathcal{B}_I$ depends not only on Φ_x but also on the subset $\mathcal{J}$. For a given subset $\mathcal{J}$, the smaller the class of generator functions Φ_x, the smaller is the identified set $\mathcal{B}_I$. For a given class Φ_x, the identified set depends critically on the property of $\mathcal{J}$. Below we present two examples illustrating these two difference sources of identifying power.

Example 2.3 *Suppose for all $x \in \mathcal{J}$, the generator function φ_{xo} is known so $\Phi_x = \{\varphi_{xo}\}$. For example, under independent censoring, $\varphi_{xo}(u) = \log(1/u)$ for all $x \in \mathcal{X}$. Since the conditional distribution function in this case is point identified as $F_Y(\cdot|x;\varphi_{xo})$ for $x \in \mathcal{J}$, rank conditions similar to those in Koenker and Bassett (1978) and Wang and Wang (2009) would lead to point identification of β_o.*

Example 2.4 *Suppose Φ_x is the whole class of strict generator functions. Let*

$$\mathcal{J} = \left\{x \in \mathcal{X} : \Pr\left(C_i \geq X_i'\beta_o | X_i = x\right) = 1\right\}.$$

Suppose Assumption (A2) in Khan and Tamer (2009) holds, i.e., $\mathcal{X}_B$ does not lie in a proper linear subspace of $\mathcal{R}^d$. Then the identified set $\mathcal{B}_I$ is singleton. Notice that for $\forall x \in \mathcal{J}$, we have

$$
\begin{aligned}
F_V(x'\beta_o|x) &= \Pr(Y_i \leq x'\beta_o, Y_i \leq C_i|x) + \Pr(C_i \leq x'\beta_o, Y_i > C_i|x) \tag{2.10}\\
&= \Pr(Y_i \leq x'\beta_o|x) = q.
\end{aligned}
$$

Alternatively, using the expression in (2.7), we get that for all $x \in \mathcal{J}$,

$$
\begin{aligned}
F_Y\left(x'\beta_o|x;\varphi_x\right) &= 1 - \varphi_x^{-1}\left(-\int_{y_{lx}}^{x'\beta_o}\varphi_x'\left\{S_V\left(v|x\right)\right\}dF_V\left(v|x\right)\right) \\
&= 1 - \varphi_x^{-1}\left(\varphi_x\left\{S_V\left(x'\beta_o|x\right)\right\}\right) \\
&= F_V\left(x'\beta_o|x\right),
\end{aligned}
$$

where we have used the fact that $F_V\left(v|x\right) = F_{V,D=1}\left(v|x\right)$ for $v \leq x'\beta_o$ as derived in (2.10). As noted in Khan and Tamer (2009), this identification strategy has also been employed in Powell (1986), Honore, Khan and Powell (2002) in one way or another.

Remark 2.5 *Without Assumption (SY) (ii), for a given generator function φ_x, $F_{Yo}\left(y|x\right)$ is identified for all $y \in [y_{lx}, v_{ux}]$ which may be used to establish the identified set for β_o.*

2.2 A Characterization of an Outer Set of the Identified Set

In this section, we consider one class of generator functions denoted as Φ_x^O and provide a nice characterization of an outer set of the identified set for β_o via inequality constraints.

Assumption (O). The class of generator functions Φ_x^O is composed of continuously differentiable generator functions on $(0,1]$ and is indexed by a unique pair of generators $\left(\varphi_{x,L}, \varphi_{x,U}\right)$ such that $\varphi_{x,L}'\left(u\right)/\varphi_x'\left(u\right)$ and $\varphi_x'\left(u\right)/\varphi_{x,U}'\left(u\right)$ are both non-decreasing on $(0,1)$ for all $\varphi_x \in \Phi_x^O$.

The class of copulas generated by Φ_x^O has a convenient/nice property facilitating a sensitivity analysis. To describe it, let $\mathcal{C}_{xL}$ denote the Archimedean copula with generator function $\varphi_{x,L}$ and $\mathcal{C}_{xU}$ denote the Archimedean copula with generator function $\varphi_{x,U}$. Under Assumption (O), Corollary 4.4.6 in Nelsen (2006) implies that $\mathcal{C}_{xL} \prec \mathcal{C}_x \prec \mathcal{C}_{xU}$ for any Archimedean copula $\mathcal{C}_x$ generated by $\varphi_x \in \Phi_x^O$. Thus in terms of concordance ordering, $\mathcal{C}_{xL}$ is the smallest and $\mathcal{C}_{xU}$ is the largest in the class of Archimedean copulas with generators in Φ_x^O. Thus letting $\varphi_{x,L}\left(u\right) = \log\left(1/u\right)$ or $\mathcal{C}_{xL}\left(u,v\right) = uv$, a sensitivity analysis can be conducted by varying $\mathcal{C}_{xU}$ according to increasing or decreasing concordance ordering representing more strongly dependent censoring mechanisms. Dependence measures such as Kendall's τ and Spearman's ρ are natural measures of deviation from independent censoring.

Let $\mathcal{F}_I^O\left(x\right)$ and $\mathcal{B}_I^O$ denote the identified sets for $F_{Yo}\left(\cdot|x\right)$ and β_o corresponding to the class of generators Φ_x^O defined in Assumption (O). Further for $y \in [y_{lx}, y_{ux}]$, let

$$
F_L\left(y|x\right) = F_Y\left(y|x;\varphi_{x,L}\right) \text{ and } F_U\left(y|x\right) = F_Y\left(y|x;\varphi_{x,U}\right).
$$

We show below that elements of $\mathcal{F}_I^O\left(x\right)$ are bounded by $F_L\left(y|x\right)$ from below and $F_U\left(y|x\right)$ from above (see (2.12)) which leads to nice inequality constraints characterizing an outer set of $\mathcal{B}_I^O$.

Proposition 2.6 *Suppose Assumptions (AC), (SY), and (O) hold for all $x \in \mathcal{J}$. Then*

$$
F_U^{-1}\left(q|x\right) \leq x'\beta_o \leq F_L^{-1}\left(q|x\right) \text{ for all } x \in \mathcal{J}. \tag{2.11}
$$

Proof. We will complete the proof in two steps.

Step 1. We show that F_{Yo} satisfies:[4]

$$F_L\left(y|x\right) \leq F_{Yo}\left(y|x\right) \leq F_U\left(y|x\right), \forall y \in [y_{lx}, y_{ux}). \tag{2.12}$$

It follows from (2.6) that

$$F'_{Yo}\left(v|x\right) = \frac{\varphi'_{xo}\left\{S_V\left(v|x\right)\right\} F'_{V,D=1}\left(v|x\right)}{\varphi'_{xo}\left\{S_{Yo}\left(v|x\right)\right\}}.$$

Multiplying both sides of the above equation by $\varphi'_{x,L}\{S_{Yo}\left(v|x\right)\}$ and integrating from y_{lx} to y lead to

$$\varphi_{x,L}\{S_{Yo}\left(y|x\right)\} = -\int_{y_{lx}}^{y} \frac{\varphi'_{x,L}\{S_{Yo}\left(v|x\right)\}\varphi'_{xo}\left\{S_V\left(v|x\right)\right\}}{\varphi'_{xo}\left\{S_{Yo}\left(v|x\right)\right\}} dF_{V,D=1}\left(v|x\right). \tag{2.13}$$

Because $S_{Yo}\left(v|x\right) \geq S_V\left(v|x\right)$, given the monotonicity of $\varphi'_{x,L}\left(t\right)/\varphi'_{xo}\left(t\right)$, we have

$$\frac{\varphi'_{x,L}\{S_{Yo}\left(v|x\right)\}}{\varphi'_{xo}\left\{S_{Yo}\left(v|x\right)\right\}} \geq \frac{\varphi'_{x,L}\left\{S_V\left(v|x\right)\right\}}{\varphi'_{xo}\left\{S_V\left(v|x\right)\right\}}.$$

As $\varphi'_x\left(\cdot\right)$ is negative, we get

$$-\frac{\varphi'_{x,L}\{S_{Yo}\left(v|x\right)\}\varphi'_{xo}\left\{S_V\left(v|x\right)\right\}}{\varphi'_{xo}\left\{S_{Yo}\left(v|x\right)\right\}} \geq -\varphi'_{x,L}\left\{S_V\left(v|x\right)\right\}. \tag{2.14}$$

Hence (2.13) and (2.14) imply that

$$\begin{aligned} \varphi_{x,L}\{S_{Yo}\left(y|x\right)\} &\geq -\int_{y_{lx}}^{y} \varphi'_{x,L}\left\{S_V\left(v|x\right)\right\} dF_{V,D=1}\left(v|x\right) \\ &= \varphi_{x,L}\{S_Y\left(y|x, \varphi_{x,L}\right)\}, \end{aligned}$$

where we have used (2.7). The desired result follows from the above inequality and the decreasing property of $\varphi_{x,L}$. Flipping the sign to conclude the corresponding bounds for the distribution function.

Step 2. Since $F_{Yo}\left(x'\beta_o|x\right) = q$ holds for almost all x, we obtain: $F_L\left(x'\beta_o|x\right) \leq q \leq F_U\left(x'\beta_o|x\right)$. The claimed result follows the definition of the conditional quantile functions. ∎

Remark 2.7 *Interpreting Y and C as two competing risks, the model defined in (1.1) and (2.2) is a new semiparametric competing risks model where the marginal model for Y conditional on the covariate X is specified by the linear quantile model, the marginal model for C conditional on the covariate is unspecified, and the conditional copula function of Y and C is Archimedean. Our model and inference procedures are potentially useful in duration analysis where the researcher is only interested in one of the competing risks denoted by Y. Specifically the interest is in effects of some observable covariate X on the q-th conditional quantile of Y in the presence of a possibly dependent competing risk C. In fact the identification strategy and the subsequent inference procedures developed in this paper are not restricted to the marginal quantile model for Y. Given $F_Y\left(y|x; \varphi_x\right)$ in (2.7), one can easily write down the identified set for the parameter in*

---

[4] The proof of Step 1 is a slight modification of that of Proposition 2 in Rivest and Wells (2001) which measures the maximal bias of the copula-graphic estimator of the survival function due to a misspecified Archimedian copula generator. We include a proof for completeness.

any parametric model for Y including the proportional hazard model and any parametric regression model. The reason is that the true parameter in all these models satisfies equality constraints on known functionals of the true conditional distribution function of Y given X. With slight abuse of notation, denote, e.g., the true parameter as $\beta_o \in \mathcal{B}$ and the functional constraints as $G\left(F_{Yo}\left(\cdot|x\right);\beta_o\right) = 0$ for a known possibly vector-valued functional G. Then the identified set for β_o is

$$\left\{\beta \in \mathcal{B} : G\left(F_Y\left(\cdot|x;\varphi_x\right);\beta\right) = 0 \text{ for some } \varphi_x \in \Phi_x \text{ and all } x \in \mathcal{J}\right\}.$$

For example, $G\left(F_Y\left(\cdot|x;\varphi_x\right);\beta\right) = F_Y\left(x'\beta|x;\varphi_x\right) - q$ for the quantile model;

$$G\left(F_Y\left(\cdot|x;\varphi_x\right);\beta\right) = \int yF_Y\left(y|x;\varphi_x\right)dy - x'\beta$$

for the linear regression model; and

$$G\left(F_Y\left(\cdot|x;\varphi_x\right);\beta\right) = F_Y\left(\cdot|x;\varphi_x\right) - 1 + \mathcal{L}\left(\Lambda_0\left(\cdot;\beta_b\right)\exp\left(x'\beta_x\right);\beta_\omega\right), \beta = \left(\beta_b', \beta_x', \beta_\omega'\right)',$$

for the following parametric version of the mixed proportional hazard model in (2.1):

$$\lambda_{Yo}\left(t|x,\omega\right) = \omega\lambda_0\left(t;\beta_{ob}\right)\exp\left(x'\beta_{ox}\right),$$

where the conditional distribution function of ω given $X = x$ is denoted as $F_{\omega|x}\left(\cdot;\beta_{o\omega}\right)$ with the corresponding Laplace transform $\mathcal{L}\left(\cdot;\beta_{o\omega}\right)$, where $\Lambda_0\left(t;\beta_{ob}\right)$ is the integrated baseline hazard. Provided that the functional $G\left(\cdot;\beta\right)$ is smooth enough, the CSs developed in the rest of this paper could be easily extended to any parametric marginal model for Y.

2.3   Ordered Parametric Families of Invariant Copulas

To simplify the asymptotic analysis and the subsequent inference procedure, we introduce two more assumptions below, Assumptions (O-P-I) and (SC).

Assumption (O-P-I). (i) The true copula is invariant w.r.t x; (ii) It belongs to a one-parameter family of Archimedean copulas denoted as $\mathcal{C}\left(\cdot,\cdot;\alpha\right)$ with generator $\varphi\left(\cdot;\alpha\right)$ indexed by $\alpha \in \mathcal{A} \equiv [\alpha_L, \alpha_U]$; and (iii) for any $\alpha_1 < \alpha_2$ from $\mathcal{A}$, it holds that

$$\frac{\varphi'\left(u;\alpha_1\right)}{\varphi'\left(u;\alpha_2\right)} \text{ is strictly increasing } \forall u \in (0,1), \tag{2.15}$$

where $\varphi'\left(u;\alpha\right)$ denotes the partial derivative of $\varphi\left(u;\alpha\right)$ with respect to u.

Assumption (SC). Suppose there exists some $x_0 \in \mathcal{J}$, s.t. $S_Y\left(y|x_0;\alpha_U\right) > S_V\left(y|x_0\right)$ for all $y \in \left(y_{lx_0}, y_{ux_0}\right)$.

Assumption (O-P-I) (i) states that the true conditional copula function $\mathcal{C}_{xo}$ is invariant with respect to x and Assumption (O-P-I) (ii) parametrizes the generator function by some parameter $\alpha \in \mathcal{A} \subset \mathcal{R}$. The copula invariance assumption has been adopted in other contexts, see e.g., Chen and Fan (2006) for semiparametric copula-based multivariate dynamic models and Torgovitsky (2011) in nonseparable structural models. In the

context of competing risks, Bond and Shaw (2006) have studied the so-called covariate-time transformation model in which the modelling assumption directly implies the copula invariance. Bond and Shaw (2006) show that classical competing risks models including the accelerated failure-time model and the proportional hazard model fall into their framework, see also Clayton and Cuzick (1985), Heckman and Honore (1989). Assumption (O-P-I) (ii) restricts the class of copula functions to be in a given parametric class. Informal sensitivity analysis in Zheng and Klein (1995), Huang and Zhang (2008), and Chen (2010) suggest that the bias of estimates of the marginal survival function of Y is negligible when the parametric copula misspecifies the true copula as long as the dependence range (such as Kendall's tau) is correctly specified. This is also confirmed in our numerical analysis in Example 2.9 below. Many one-parameter families of Archimedean copulas including Frank or Clayton copulas satisfy Assumption (O-P-I) (iii). They are either positively ordered (increasing in concordance order as the parameter α increases) or negatively ordered (decreasing in concordance order as the parameter α increases), see Joe (1997), Nelsen (2006), and Rivest and Wells (2001). In Table 1 below, we list a number of one-parameter families[5] of Archimedean copulas that are ordered and satisfy Assumption (O-P-I) (iii).

Table 1: One-Parameter Archimedean Copulas

Copulas	$\varphi\left(u;\alpha\right)$	α's Range	$\varphi'\left(u;\alpha\right)$	$\varphi'\left(u;\alpha_L\right)/\varphi'\left(u;\alpha_U\right)$
Clayton	$\frac{u^{-\alpha}-1}{\alpha}$	$(0,\infty)$	$-u^{-\alpha-1}$	$u^{\alpha_U-\alpha_L}$
Frank	$\log\left(\frac{1-e^{-\alpha}}{1-e^{-\alpha u}}\right)$	$(-\infty,\infty)$	$-\frac{\alpha e^{-\alpha u}}{1-e^{-\alpha u}}$	$\frac{\alpha_L(1-e^{\alpha_U u})}{\alpha_U(1-e^{\alpha_L u})}$
Gumbel	$(-\log u)^{\alpha}$	$[1,\infty)$	$-\frac{\alpha}{u}\left(\log\frac{1}{u}\right)^{\alpha-1}$	$\frac{\alpha_L}{\alpha_U}\left(\log\frac{1}{u}\right)^{\alpha_L-\alpha_U}$
Gumbel-Hougaard	$\log\left(1-\alpha\log u\right)$	$(0,1]$	$-\frac{\alpha}{u-\alpha u\log u}$	$1+\frac{1/\alpha_L-1/\alpha_U}{u\log u-1/\alpha_L}$
Nelsen #12	$\left(\frac{1}{u}-1\right)^{\alpha}$	$[1,\infty)$	$-\frac{\alpha}{u^2}\left(\frac{1-u}{u}\right)^{\alpha-1}$	$\frac{\alpha_L}{\alpha_U}\left(\frac{1}{u}-1\right)^{\alpha_L-\alpha_U}$
Nelsen #16	$\left(\frac{\alpha}{u}+1\right)(1-u)$	$(0,\infty)$	$-\frac{\alpha}{u^2}-1$	$1+\frac{\alpha_L-\alpha_U}{u^2+\alpha_U}$
Nelsen #19	$e^{\alpha/u}-e^{\alpha}$	$(0,\infty)$	$-\frac{\alpha}{u^2}e^{\alpha/u}$	$\frac{\alpha_L}{\alpha_U}e^{(\alpha_L-\alpha_U)u}$
Nelsen #20	$e^{u^{-\alpha}}-e$	$(0,\infty)$	$-\frac{\alpha}{u}u^{-\alpha}$	$\frac{\alpha_L}{\alpha_U}e^{u^{-\alpha_L}-u^{-\alpha_U}}$

Under Assumption (O-P-I), the class of generator functions is given by

$$\Phi_x = \left\{\varphi_x : \varphi_x\left(\cdot\right) = \varphi\left(\cdot;\alpha\right) \text{ for some } \alpha \in \mathcal{A}\right\},$$

where the functional form of $\varphi\left(\cdot;\alpha\right)$ is known. So for a given family of ordered parametric copulas, the choice of Φ_x is equivalent to the choice of $\mathcal{A} \equiv [\alpha_L, \alpha_U]$. Users could specify α_L, α_U reflecting their prior knowledge on the possible dependence range giving rise to $\mathcal{C}_{xL}$ and $\mathcal{C}_{xU}$. In a sensitivity analysis, users could take α_L corresponding to the independence copula and specify several values for α_U reflecting different strengths of dependence between Y and C; the larger α_U is, the larger is the deviation of the true censoring mechanism from independence censoring. Under Assumption (O-P-I), the following equality holds:

$$\tau\left(\alpha\right) = 1 + 4\int_0^1 \frac{\varphi\left(u;\alpha\right)}{\varphi'\left(u;\alpha\right)}du. \tag{2.16}$$

---

[5]Some of the copulas in Table 1 do not have names (or not widely known among researchers), we simply attribute them as *Nelsen's #*, as those are found by Table 4.1 appearing in Chapter 4, Nelsen (2006).

It is evident from (2.16) that perturbation on τ could be performed on the copula's natural parameter α. For Clayton and Gumbel copulas, it is known that Kendall's $\tau = \frac{\alpha}{\alpha+2}$ and $\frac{\alpha-1}{\alpha}$ respectively, so $\alpha = 0$ for Clayton copula and $\alpha = 1$ for Gumbel copula correspond to independent censoring and as α increases the censoring mechanism deviates more from independent censoring. Assumption (SC) is a separation condition. It excludes cases where our copula lower bound $S_Y(\cdot|x; \alpha_U)$ is identical to Peterson's lower bound $S_V(\cdot|x)$ for all $x \in \mathcal{J}$.

Under Assumption (O-P-I), for $\varphi_x \in \Phi_x$, $F_Y(y|x; \varphi_x)$ depends on φ_x only through α. So we denote it as $F_Y(y|x; \alpha)$ in the rest of this paper.

Proposition 2.8 *Suppose Assumptions (AC), (SY), and (O-P-I) hold for all $x \in \mathcal{J}$. Then (i) the identified set for β_o is*

$$\mathcal{B}_I^O = \{\beta \in \mathcal{B} : F_Y(x'\beta|x; \alpha) = q \text{ for all } x \in \mathcal{J} \text{ and some } \alpha \in \mathcal{A}\}; \tag{2.17}$$

(ii) if Assumption (SC) holds as well, then given any β in $\mathcal{B}_I^O$, there is a unique $\alpha(\beta)$ such that

$$F_Y(x'\beta|x; \alpha(\beta)) = q \text{ for all } x \in \mathcal{J}. \tag{2.18}$$

Proof. (i) is obvious. Now given (i), it suffices to show for any $\alpha_1, \alpha_2 \in \mathcal{A}$ (w.l.o.g we take $\alpha_1 < \alpha_2$), $F_Y(x_0'\beta|x_0; \alpha_1) < F_Y(x_0'\beta|x_0; \alpha_2)$ holds for the particular x_0 in the separation assumption (SC). From the conclusion in Proposition 2.3 we know that $S_Y(\cdot|x_0, \alpha_2) \geq S_Y(\cdot|x_0, \alpha_U) > S_V(\cdot|x_0)$ holds in terms of the invariant generator and at location x_0. The proof follows almost verbatim from the proof of Prop. 2.3. After equation (2.9) we shall proceed with those strict inequalities:

$$\frac{\varphi'\{S_Y(v|x_0, \alpha_2); \alpha_1\}}{\varphi'\{S_Y(v|x_0, \alpha_2); \alpha_2\}} > \frac{\varphi'\{S_V(v|x_0); \alpha_1\}}{\varphi'\{S_V(v|x_0); \alpha_2\}}, \text{ with } v \in (y_{lx_0}, y_{ux_0})$$

for $\alpha_1 < \alpha_2$ by (2.15) . Similar manipulation leads to $\varphi\{S_Y(y|x_0, \varphi_{\alpha_2}); \alpha_1\} > \varphi\{S_Y(y|x_0, \varphi_{\alpha_1}); \alpha_1\}$, and the copula generator is strictly decreasing, thus $F_Y(y|x_0, \alpha_1) < F_Y(y|x_0, \alpha_2)$. Therefore given any β in $\mathcal{B}_I^O$ when $F_Y(x'\beta|x; \alpha(\beta)) = q \in (0, 1)$, we could only have a unique $\alpha(\beta)$ for $x \in \mathcal{J}$. ∎

Example 2.9 (Gumbel Copula) *Suppose Assumption (O-P-I) (i) and (ii) hold with the family of Gumbel copulas so*

$$\varphi(u; \alpha) = (-\log u)^\alpha, \ \alpha \in [1, \infty).$$

Let α_o denote the true value of the copula parameter. Suppose the true conditional marginal survival functions are $S_{Yo|X}(y|x) = e^{-y/x}$ and $S_{C|X}(c|x) = e^{-c/x}$ for $y \geq 0$, $c \geq 0$, and $x > 0$. It is easy to show that the conditional survival and sub-survival functions of the observable V are given by:

$$S_V(v|x) = \exp\left[-2^{1/\alpha_o}\frac{v}{x}\right] \text{ and}$$
$$S_{V,D=1}(v|x) = \frac{1}{2}\exp\left[-2^{1/\alpha_o}\frac{v}{x}\right], \ x > 0.$$
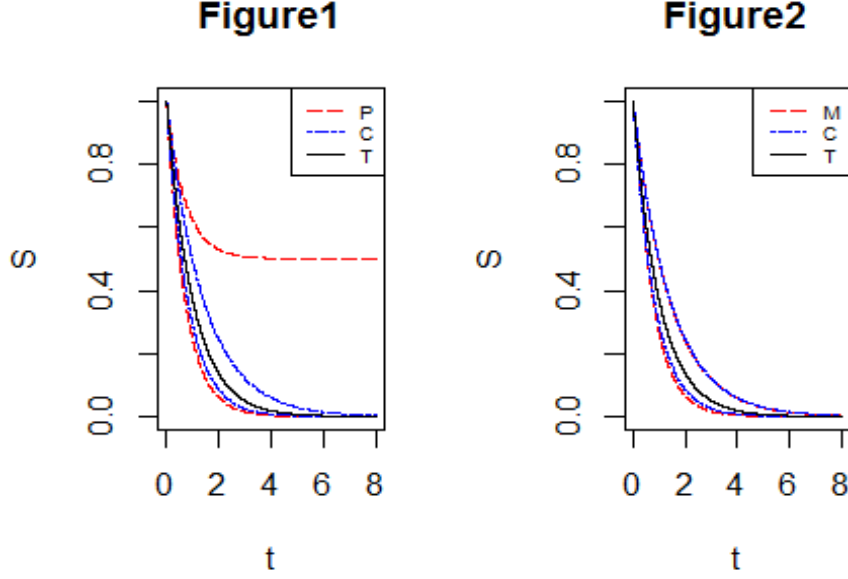
Figure1   **Figure2**

Suppose we know that $\alpha_o \in [\alpha_L, \alpha_U]$ or equivalently $\tau_o \in [\tau_L, \tau_U] = \left[\frac{\alpha_L - 1}{\alpha_L}, \frac{\alpha_U - 1}{\alpha_U}\right]$ (see Example 5.4 in Nelson, 2006). For any $\alpha \in [\alpha_L, \alpha_U]$, (2.7) implies that for $y > 0$,

$$S_Y(y|x; \alpha) = \exp\left[-2^{1/\alpha_o - 1/\alpha}\frac{y}{x}\right] \tag{2.19}$$

yielding the bounds $S_Y(y|x; \alpha_L)$ and $S_Y(y|x; \alpha_U)$ for the true survival function $S_{Y_o}(y|x)$.

Let $x = 1$, $\alpha_o = 2$ ($\tau_o = 0.5$), and $\alpha_L = 1, \alpha_U = 5$ ($\tau \in [0, 0.8]$). In Figure 1, we plot the true survival function $S_{Y_o}(y|1)$ (black solid curve), our copula bounds $S_Y(y|1; \alpha_L)$ and $S_Y(y|1; \alpha_U)$ (two blue curves), and the worst-case Peterson bounds (two red curves):

$$S_V(y|1) = \exp\left(-2^{1/\alpha_o}y\right) \text{ and}$$
$$S_{V,D=1}(y|1) + S_{V,D=0}(0|1) = \frac{1}{2}\exp\left(-2^{1/\alpha_o}y\right) + \frac{1}{2}.$$

Some observations follow immediately. First the Peterson's upper bound has some unpleasant feature, namely it is only pointwise sharp not functionally sharp. The upper bound on the survival function is not a proper survival function itself, more specifically, $\lim_{y \to \infty}[S_{V,D=1}(y|1) + S_{V,D=0}(0|1)] = \Pr(D = 0|1)$, which is strictly bigger than 0 in nontrivial cases (see Crowder 1991; Bedford and Meilijson 1997). Second, Peterson's bounds can be tightened significantly when prior knowledge on the censoring mechanism is available. Finally, the deviation from the independent censoring assumption may not be negligible, making the sensitivity analysis necessary.

Next we illustrate the effect of misspecification in the generator function (while fixing the dependence range) on the copula bounds. So instead of the Gumbel copula, we use the Clayton copula:

$$\widetilde{\varphi}(u; \widetilde{\alpha}) = \frac{u^{-\widetilde{\alpha}} - 1}{\widetilde{\alpha}}, \ \widetilde{\alpha} > 0$$

14

in (2.7) leading to

$$S_Y(y|x; \widetilde{\alpha}) = \left\{ \frac{1}{2^{1/\alpha_o}} \left[\exp\left[2^{1/\alpha_o} \widetilde{\alpha} \frac{y}{x} \right] - 1 \right] + 1 \right\}^{-1/\widetilde{\alpha}}. \tag{2.20}$$

Example 5.4 in Nelson (2006) shows that for the Clayton copula, $\tau = \frac{\widetilde{\alpha}}{\widetilde{\alpha}+2}$. The range for Kendall's τ varying in $[0, 0.8]$ would translate to $\widetilde{\alpha} \in [0, 8]$. In Figure 2 we again plot the true survival function $S_{Y_o}(y|1)$ (black solid curve) and the copula bounds $S_Y(y|1; \alpha_L)$ and $S_Y(y|1; \alpha_U)$ (two blue curves) using the correctly specified Gumbel copula. In addition, we also plot the misspecified copula bounds $S_Y(y|1; \widetilde{\alpha}_L)$ and $S_Y(y|1; \widetilde{\alpha}_U)$ (the two red curves) computed using the Clayton copula, where $\widetilde{\alpha}_L = 0$ and $\widetilde{\alpha}_U = 8$. Notice that the two sets of copula bounds are almost identical. This observation is consistent with existing simulation results showing the negligible bias in the estimated bounds with misspecified copula function as long as the dependence range is correctly specified, see Zheng and Klein (1995), Huang and Zhang (2008), and Chen (2010).

Finally we complete this example by deriving the identified set for β_o. By the restriction that $S_Y(\beta_o|x; \alpha) = 1 - q$, we get:

$$\beta_o \in \left[\log\left(\frac{1}{1-q} \right) 2^{1/\alpha_U - 1/\alpha_o}, \log\left(\frac{1}{1-q} \right) 2^{1/\alpha_L - 1/\alpha_o} \right]. \tag{2.21}$$

In terms of the corresponding $[\tau_L, \tau_U]$, we get

$$\beta_o \in \left[\log\left(\frac{1}{1-q} \right) 2^{1-\tau_U - 1/\alpha_o}, \log\left(\frac{1}{1-q} \right) 2^{1-\tau_L - 1/\alpha_o} \right].$$

In this example, the quantile regression coefficient is interval identified (Manski, 2003) and there is one-to-one correspondence between the quantile regression coefficient and the dependence level characterized by Kendall's tau.

It is obvious from the expression for $S_Y(y|x; \alpha)$ in (2.19) that Assumption (SC) holds for all $x > 0$ and all finite α_U.

3   Asymptotic Confidence Sets for β_o

In the rest of this paper, we suppose Assumptions (AC), (SY), and (O-P-I) hold for all $x \in \mathcal{J}$. In this section, we construct two asymptotic confidence sets for β_o based on the identified set $\mathcal{B}_I^O$ in (2.17):

$$\mathcal{B}_I^O = \left\{ \beta \in \mathcal{B} : F_Y(x'\beta|x; \alpha) = q \text{ for all } x \in \mathcal{J} \text{ and some } \alpha \in \mathcal{A} \right\}.$$

The identified set $\mathcal{B}_I^O$ allows X to be any random variable, discrete or continuous or mixed. In what follows, we work explicitly with mixed type regressors, so $X \equiv (X^c, X^d)$ with both continuous component $X^c = (X_1^c, \cdots X_p^c)$ and discrete component $X^d = (X_1^d, \cdots X_r^d)$. Furthermore, X_j^d takes the values $0, 1, ..., c_j - 1$ for $j = 1, ..., r$.

Define the population criterion function as

$$T(\beta) = \min_{\alpha \in [\alpha_L, \alpha_U]} T(\beta; \alpha) = \min_{\alpha \in [\alpha_L, \alpha_U]} \int_{\mathcal{J}} [F_Y(x'\beta|x; \alpha) - q]^2 f_X^2(x) \, dx, \tag{3.1}$$

where $\mathcal{J} = \mathcal{J}^c \times \mathcal{J}_1^d \times ... \times \mathcal{J}_r^d$, with $\mathcal{J}^c \subset \text{int}(\mathcal{X}^c)$ being compact and $\mathcal{J}_j^d = \{0, 1, ..., c_j\}$ for $j = 1, ..., r$. Also the integral $\int dx$ is understood to be $\sum_{x^d} \int d\mu(x^c)$, integrating over the counting measure on $\mathcal{J}_1^d \times ... \times \mathcal{J}_r^d$ and Lebesgue measure on $\mathcal{X}^c$. Our CSs are based on the sample version of $T(\beta; \alpha)$ defined as:

$$T_n(\beta; \alpha) = \int_{\mathcal{J}} \left[\widehat{F}(x'\beta | x; \alpha) - q\right]^2 \widehat{f}_X^2(x) \, dx, \tag{3.2}$$

where $\widehat{F}(\cdot | x; \alpha)$ is the plug-in estimator of $F_Y(\cdot | x; \alpha)$ introduced in the subsection below and $\widehat{f}_X(x)$ is the kernel-type density estimator of $f_X(x)$ defined below:

$$\widehat{f}_X(x) = \frac{1}{n} \sum_{i=1}^n W_\gamma(x, X_i) \tag{3.3}$$

where $W_\gamma(x, X_i) = K_h(x^c - X_i^c) L(x^d, X_i^d, \lambda)$, $\gamma = (h, \lambda) = (h, \lambda_1, \cdots, \lambda_r)$, $K_h(\cdot) = h^{-p} K(\cdot/h)$ denotes the standard kernel function for continuous regressors,[6] whereas $L(\cdot, \cdot, \lambda)$ is the Aitchison and Aitken (1976) kernel:

$$L(x^d, X_i^d, \lambda) = \prod_{j=1}^r (\lambda_j / (c_j - 1))^{N_{ij}(x)} (1 - \lambda_j)^{1 - N_{ij}(x)}$$

with $N_{ij}(x) = I\left[X_{ij}^d \neq x_j^d\right]$ for $j = 1, ..., r$. For the advantage of smoothing discrete regressors over the standard frequency approach, see Hall, Racine and Li (2004), Li and Racine (2007).

We propose two test statistics from which we could draw inference on β_o:

$$\widehat{T}_n(\beta) = T_n(\beta; \widehat{\alpha}(\beta)) \text{ and} \tag{3.4}$$

$$\widetilde{T}_n(\beta) = T_n(\beta; \widetilde{\alpha}(\beta)), \tag{3.5}$$

where

$$\widehat{\alpha}(\beta) \in \arg \min_{\alpha \in [\alpha_L, \alpha_U]} T_n(\beta; \alpha) \text{ and}$$

$$\widetilde{\alpha}(\beta) \in \arg \min_{\alpha \in [\alpha_L, \alpha_U]} \left| \frac{T_n(\beta; \alpha) - \widehat{B}_n(\beta; \alpha)}{\sqrt{\widehat{\Sigma}(\beta; \alpha)}} \right|$$

with $\widehat{B}_n(\beta; \alpha)$ and $\widehat{\Sigma}(\beta; \alpha)$ being uniformly consistent estimators of $B_n(\beta; \alpha)$ and $\Sigma(\beta; \alpha)$ defined in (B.6) (B.8) in Appendix B. Our CSs for β_o with asymptotic level $(1 - \epsilon)$ are defined as

$$CS^N_{1-\epsilon, \widehat{T}_n} = \left\{ \beta \in \mathcal{B} : \frac{nh^{p/2} \left|\widehat{T}_n(\beta) - \widehat{B}_n(\beta; \widehat{\alpha}(\beta))\right|}{\sqrt{\widehat{\Sigma}(\beta; \widehat{\alpha}(\beta))}} \leq z_{1-\epsilon/2} \right\} \text{ and} \tag{3.6}$$

$$CS^N_{1-\epsilon, \widetilde{T}_n} = \left\{ \beta \in \mathcal{B} : \frac{nh^{p/2} \left|\widetilde{T}_n(\beta) - \widehat{B}_n(\beta; \widetilde{\alpha}(\beta))\right|}{\sqrt{\widehat{\Sigma}(\beta; \widetilde{\alpha}(\beta))}} \leq z_{1-\epsilon/2} \right\}, \tag{3.7}$$

where $z_{1-\epsilon/2}$ is the $(1 - \epsilon/2)$-th percentile of the standard normal distribution.

---

[6] In typical applications, discrete regressors would have different support and cardinality, so we let λ change with r; for notational brevity we use single bandwidth for all continuous regressors.

We show in the next section that under conditions including Assumption (O-P-I) and the conditions of Proposition 2.6, both $CS^N_{1-\epsilon,\widehat{T}_n}$ and $CS^N_{1-\epsilon,\widetilde{T}_n}$ are asymptotically valid and non-conservative CSs for β_o.

By varying α_L and α_U, both CSs can be used to check sensitivity of inference for β_o to the independent censoring assumption. In contrast to most CSs for partially identified parameters, the CSs $CS^N_{1-\epsilon,\widehat{T}_n}$ and $CS^N_{1-\epsilon,\widetilde{T}_n}$ are straightforward to implement. This is especially important in the context of a sensitivity analysis since they are typically computed several times for different ranges of the copula parameter α.

Remark 3.1 *For a given $\beta \in \mathcal{B}^O_I$, the test statistics $\widehat{T}_n(\beta)$ and $\widetilde{T}_n(\beta)$ in (3.4) and (3.5) resemble the test statistics for consistent model specification testing based on kernel estimators, see Fan (1994), Fan and Li (1996), and Zheng (1996) and many subsequent works in the literature.[7] Indeed, as in these papers, we show later that under suitable conditions including the separation assumption (SC), the asymptotic distributions of $\widehat{T}_n(\beta)$ and $\widetilde{T}_n(\beta)$ are normal justifying the CSs $CS^N_{1-\epsilon,\widehat{T}_n}$ and $CS^N_{1-\epsilon,\widetilde{T}_n}$ defined in (3.6) and (3.7). Thus the CSs $CS^N_{1-\epsilon,\widehat{T}_n}$ and $CS^N_{1-\epsilon,\widetilde{T}_n}$ for β_o are intrinsically linked to specification tests for the class of parametric copulas with generator function φ_α, $\alpha \in \mathcal{A}$.*

Remark 3.2 *An alternative approach to constructing CS for β_o is to make use of the inequality constraints on β_o in Proposition 2.6: $F_L(x'\beta_o|x) \leq q \leq F_U(x'\beta_o|x)$ for all $x \in \mathcal{J}$. For instance, one could adopt the following criterion function:*

$$\int_\mathcal{J} \left[\widehat{F}(x'\beta|x;\alpha_L) - q \right]^2_- \widehat{f}^2_X(x)\, dx + \int_\mathcal{J} \left[\widehat{F}(x'\beta|x;\alpha_U) - q \right]^2_+ \widehat{f}^2_X(x)\, dx. \qquad (3.8)$$

Compared with $\widehat{T}_n(\beta)$ or $\widetilde{T}_n(\beta)$, this approach suffers from several drawbacks. First, the asymptotic distribution of the statistic in (3.8) is difficult to establish; Second, similar to existing work on inference for parameters defined by moment inequalities such as Andrews and Shi (2013), variants of the 'generalized moment selection' may be needed introducing additional parameters that practitioners have to select. In contrast, CSs based on $\widehat{T}_n(\beta)$ or $\widetilde{T}_n(\beta)$ circumvent this because they rely on equality constraints only; Third, let

$$\mathcal{B}_O = \{\beta \in \mathcal{B} : F_L(x'\beta_o|x) \leq q \leq F_U(x'\beta_o|x) \text{ for all } x \in \mathcal{J}\}.$$

Proposition 2.6 only shows that $\mathcal{B}_O$ is an outer set of the identified set $\mathcal{B}^O_I$, i.e., $\mathcal{B}^O_I \subseteq \mathcal{B}_O$, but it is not clear whether $\mathcal{B}_O \subseteq \mathcal{B}^O_I$.

3.1 The Plug-in Estimator of $F_Y(y|x;\alpha)$

Our test statistics depend on an estimator of $F_Y(x'\beta|x;\alpha)$ or generally of $F_Y(y|x;\alpha)$ defined in (2.7). Throughout this section we will suppress the subscript Y in its conditional distribution or survival functions unless otherwise emphasized. When the censoring mechanism is independent conditional on covariates, Dabrowska (1987, 1989) studies the consistency and weak convergence of the so-called conditional Kaplan-Meier estimator originally proposed by Beran in an unpublished report. Under dependent censoring mechanism, Braekers and Veraverbeke (2005) generalize the copula-graphic methods in Rivest and Wells (2001)

---

[7]Similar specification testing procedures with mixed type regressors could be found in Fan, Li and Min (2006) and Hsiao, Li and Racine (2007).

17

to the case where X is univariate and non-stochastic. In this section we propose a plug-in estimator of $F_Y (y|x; \alpha)$ using its expression in (2.7).

We first introduce the Nadaraya-Watson kernel estimators of $F_{V,D=1} (v|x)$ and $F_V (v|x)$:

$$\widehat{F}_{V,D=1} (v|x) = \sum_{i=1}^{n} w_{n\gamma} (x, X_i) I [V_i \leq v, D_i = 1] \text{ and}$$

$$\widehat{F}_V (v|x) = \sum_{i=1}^{n} w_{n\gamma} (x, X_i) I [V_i \leq v],$$

where $w_{n\gamma} (x, X_i) = W_\gamma (x, X_i) / \sum_{j=1}^{n} W_\gamma (x, X_j)$, with $W_\gamma (\cdot, \cdot)$ defined in the previous subsection. In view of our Lemma 2.1, it is natural to work with the plug-in type estimator for the conditional distribution functions indexed by α:

$$
\begin{aligned}
\widehat{F} (y|x; \alpha) &= 1 - \varphi^{-1} \left(- \int_{y_{lx}}^{y} \varphi' \left\{ \widehat{S}_V (s|x) ; \alpha \right\} d\widehat{F}_{V,D=1} (s|x) ; \alpha \right) \\
&= 1 - \varphi^{-1} \left[- \sum_{V_i \leq y, D_i = 1} \varphi' \left(\widehat{S}_V (V_i|x) ; \alpha \right) w_{n\gamma} (x, X_i) ; \alpha \right].
\end{aligned}
\tag{3.9}
$$

Remark 3.3 *An alternative estimator of $F_Y (y|x; \alpha)$ is the copula graphic estimator introduced in Braekers and Veraverbeke (2005) denoted as $\widetilde{F} (y|x; \alpha) = 1 - \widetilde{S} (y|x; \alpha)$, where*

$$
\begin{aligned}
&\widetilde{S} (y|x; \alpha) \\
&= = \varphi^{-1} \left\{ - \sum_{V_i \leq y, D_i = 1} \left[\varphi \left(\widehat{S}_V (V_i^-|x) ; \alpha \right) - \varphi \left(\widehat{S}_V (V_i^-|x) - w_{n\gamma} (x, X_i) ; \alpha \right) \right] ; \alpha \right\}.
\end{aligned}
\tag{3.10}
$$

The estimator $\widetilde{F} (y|x; \alpha)$ generalizes the conditional kernel Kaplan-Meier estimator proposed in Dabrowska (1987, 1989) to allow for conditional dependent censoring characterized by the generator function $\varphi (\cdot; \alpha)$. When $\varphi (t; \alpha) = \log (1/t)$, (Y, C) are independent conditional on $X = x$ and $\widetilde{F} (y|x; \alpha)$ reduces to the conditional kernel Kaplan-Meier estimator in Dabrowska (1987, 1989),

$$
\widetilde{F}_{Ind} (y|x) = 1 - \left[\prod_{V_{(i)} \leq y} \left(1 - \frac{w_{n\gamma} (x, X_{[i]})}{1 - \sum_{j=1}^{i-1} w_{n\gamma} (x, X_{[j]})} \right)^{D_{[i]}} \right]
\tag{3.11}
$$

where $\left(V_{(i)} \right)_{i=1}^{n}$ denote the order statistics and $\left(D_{[i]}, X_{[i]} \right)_{i=1}^{n}$ denote the induced order statistics of the sample. The above estimator resembles the traditional Kaplan-Meier estimator closely, replacing the empirical weight n^{-1} with the kernel weight $w_{n\gamma} (x, X_i)$. As shown in Lemma A.3, the two estimators $\widehat{F} (y|x; \alpha)$ and $\widetilde{F} (y|x; \alpha)$ are first order asymptotically equivalent.

4 Asymptotic Validity and Bootstrap Confidence Sets

In this section, we first establish a uniform asymptotic linear representation of the plug-in estimator of $F_Y (y|x; \alpha)$, then establish asymptotic validity of the CSs $CS^N_{1-\epsilon, \widehat{T}_n}$ and $CS^N_{1-\epsilon, \widetilde{T}_n}$, and lastly construct bootstrap CSs.

4.1 Asymptotic Linear Representation of $\widehat{F}(y|x,\alpha)$

We first present regularity assumptions used to establish the asymptotic linear representation of $\widehat{F}(y|x,\alpha)$. The random vector $Z_i = (V_i, D_i, X_i)$ stacks all the observable random variables. To ease the notational burden, we assume that the support of the conditional distribution function of Y is fixed at $[y_l, y_u]$, invariant with respect to x. In addition, we let $\varphi_\alpha(u) = \varphi(u; \alpha)$ throughout the rest of this paper and let

$$\dot{\varphi}'_\alpha(u) \equiv \frac{\partial}{\partial \alpha} \varphi'_\alpha(u) \text{ and } \dot{\varphi}_\alpha^{-1}(u) \equiv \frac{\partial}{\partial \alpha} \varphi_\alpha^{-1}(u).$$

Assumption (D). (i) The random variable X^c has an absolutely continuous and bounded density w.r.t the Lebsegue measure μ in $\mathcal{R}^p$, and $\inf_{x \in \mathcal{J}} f_X(x) > 0$ for the compact subset $\mathcal{J}$ in (3.1); (ii) The marginal density function $f_X(x) = f_X(x^c, x^d)$ satisfies $\forall x^d$, $x^c \to f_X(x^c, x^d)$ is s-order continuously differentiable over the set $\mathcal{J}^c$ and the s-order derivatives are bounded; (iii) There exists y_u^0 in the support of Y and $\delta_0 > 0$ such that

$$S_V(y_u^0|x) \geq \delta_0 \text{ a.s. } x \in \mathcal{J}. \tag{3.12}$$

Assumption (F). (i) The two conditional sub-distribution functions have continuous bounded conditional sub-density functions $f_{V,D=j}(v|x)$, $j = 0, 1$ uniformly for $x \in \mathcal{J}$; (ii) Along the x^c-axis the conditional sub distribution functions satisfy:

$\forall v \in [y_l, y_u^0]$ and $\forall x^d$, $x^c \to F_{V,D=1}(v|x^c, x^d)$, $x \to F_{V,D=0}(v|x^c, x^d)$ are s-order continuously differentiable over the set $\mathcal{J}^c$, with bounded s-order derivatives.

Assumption (G). (i) Along the u-axis, the generator function $\varphi_\alpha(\cdot)$ is third order continuously differentiable with $\varphi_\alpha^{(3)}(\cdot) \leq 0$ and $\varphi_\alpha^{(3)}(\cdot)$ remains bounded uniformly for $\forall \alpha \in \mathcal{A}$ and for $\forall u \in [\delta_0, 1]$. Moreover $\varphi'_\alpha(\cdot)$ is bounded away from 0 uniformly for $\alpha \in \mathcal{A}$ and $u \in [\delta_0, 1]$ for the δ_0 defined in (3.12);

(ii) The Lipschitz continuity property with respect to α holds for $\psi_\alpha(\cdot) = 1/\varphi'_\alpha(\cdot)$ or $\psi_\alpha(\cdot) = \varphi''_\alpha(\cdot)$ with positive constant L:

$$\sup_{u \in [\delta_0, 1]} \left| \psi_{\alpha_1}(u) - \psi_{\alpha_2}(u) \right| \leq L |\alpha_1 - \alpha_2|.$$

Assumption (H). (i) The bandwidth satisfies the following conditions: $h \to 0, \frac{nh^p}{\log n} \to \infty$, $nh^{2s} \to 0$, $\frac{nh^{2s+p}}{\log n} \to 0$ and $\frac{(\log n)^2}{nh^{3p/2}} \to 0$ as $n \to \infty$;

(ii) For all $j = 1, \cdots, r$, $\lambda_j \to 0$ and $\frac{nh^p \lambda_j^2}{\log n} \to 0$, as $n \to \infty$.

Assumption (K). Let $K(u) = \prod_{j=1}^p k(u_j)$, where $k(\cdot)$ is a bounded s-order kernel function with compact support, i.e.,

$$\int k(u)\, du = 1 \text{ and } \int u^j k(u)\, du = 0 \text{ for } j = 1, ..., s - 1.$$

Moreover it can be written as $\Psi(p(x))$, with $\Psi(\cdot)$ being of bounded variation and $p(x)$ a real polynomial on $\mathcal{R}$.

Assumptions (D)(i), (ii) and (F) are standard assumptions used to establish asymptotic properties of estimators or test statistics that are functionals of kernel type regression estimators (see Li and Racine,

2007). Assumption (D)(iii) plays a similar role to Assumption (D)(i), because most generator functions have unbounded derivatives at 0, see Table 1. Assumption (D)(iii) allows us to avoid having to deal with such explosive behavior of the generator function and require Assumption (G) (i) only when establishing the linear representation of $\widehat{F}(y|x, \alpha)$, see the expression in (3.9).[8] It can be relaxed by imposing suitable restrictions on the tail behavior of the generator function at the expense of more tedious proofs. Apropos of the requirement on the copula generator, the differentiability and non-vanishing first order derivative are almost necessary in view of the following uniform asymptotic linear representation. The Lipschitz continuity in Assumption (G)(ii) is used to prove uniformity of the linear representation over α and it also simplifies the convergence argument for $\widehat{\alpha}(\beta)$ or $\widetilde{\alpha}(\beta)$ when we apply the local U-process machinery. The condition on the bandwidth is standard in kernel estimation problem, and we undersmooth a bit to kill the bias term, facilitating the inference procedure. Under Assumption (K), we have the following VC-type functional class due to Nolan and Pollard (1987):

$$\mathcal{K} = \left\{ K\left(h^{-1}\left(x - \cdot \right) \right) : x \in \mathcal{R}^p,\, h > 0 \right\}.$$

An explicit construction of $k(\cdot)$ satisfying the above requirement could be found in Section 1.2.2 in Tsybakov (2008) based on Legendre polynomials.

Theorem 4.1 *Under Assumptions (D)-(K), it holds that*

$$\widehat{F}(y|x; \alpha) - F_Y(y|x; \alpha) = \frac{1}{n f_X(x)} \sum_{i=1}^{n} W_\gamma(x, X_i)\, g(y|Z_i, x; \alpha) + R_n(y, x; \alpha), \tag{3.13}$$

where $g(y|Z_i, x; \alpha) = c(y|Z_i, x; \alpha) + b(y|Z_i, x; \alpha)$ in which

$$c(y|Z_i, x; \alpha) = \frac{-1}{\varphi'_\alpha \{S_Y(y|x; \alpha)\}} \Big[\int_{y_l}^{y} \varphi''_\alpha \{S_V(v|x)\} \left[I(V_i \leq v) - F_V(v|X_i) \right] dF_{V, D=1}(v|x) \tag{3.14}$$
$$- \varphi'_\alpha \{S_V(y|x)\} \left[I(V_i \leq v, D_i = 1) - F_{V, D=1}(y|X_i) \right]$$
$$- \int_{y_l}^{y} \varphi''_\alpha \{S_V(v|x)\} \left[I(V_i \leq v, D_i = 1) - F_{V, D=1}(v|X_i) \right] dF_V(v|x) \Big],$$

$$b(y|Z_i, x; \alpha) = \frac{-1}{\varphi'_\alpha \{S_Y(y|x; \alpha)\}} \Big[\int_{y_l}^{y} \varphi''_\alpha \{S_V(v|x)\} \left[F_V(v|X_i) - F_V(v|x) \right] dF_{V, D=1}(v|x) \tag{3.15}$$
$$- \varphi'_\alpha \{S_V(y|x)\} \left[F_{V, D=1}(y|X_i) - F_{V, D=1}(y|x) \right]$$
$$- \int_{y_l}^{y} \varphi''_\alpha \{S_V(v|x)\} \left[F_{V, D=1}(y|X_i) - F_{V, D=1}(v|x) \right] dF_V(v|x) \Big]$$

and $R_n(y, x; \alpha)$ satisfies that

$$\sup_{\alpha \in \mathcal{A}} \sup_{x \in \mathcal{J}} \sup_{y \in [y_l, y_u^0]} |R_n(y, x; \alpha)| = O_p\left(\frac{\log n}{n h^p} \right).$$

---

[8] At the right end point 1, only Gumbel copula generator has $\varphi'(1) = 0$ in our Table 1, one could simply modify the above requirement for $t \in [\delta_0, \delta_1]$, with some approporiate $\delta_1 < 1$.

Compared with the result for the copula-graphic estimator $\widetilde{F}(y|x;\alpha)$ in Braekers and Veraverbeke (2005), (3.13) holds uniformly for $x \in \mathcal{J} \subset \mathcal{X}$ with a better rate for the remainder term, where the density of X stays away from 0 on $\mathcal{J}$, and our covariate is a multivariate random variable rather than univariate fixed design.

Remark 4.2 *Holding $X = x$ fixed and φ_α fixed at some $\alpha \in \mathcal{A}$, we could also establish the weak convergence of the conditional empirical process: $\left\{ \sqrt{nh^p} \left[\widehat{F}(y|x;\alpha) - F_Y(y|x;\alpha) \right] : y \in [y_l, y_u^0] \right\}$. It can be shown that the process is stochastically equicontinuous w.r.t. certain pseudo metric. We refer the readers to Braekers and Veraverbeke (2005) for a detailed proof for the copula-graphic estimator $\widetilde{F}(y|x;\alpha)$.*

We can also establish the uniform consistency of $\widehat{F}(y|x;\alpha)$, which we record as a corollary below. Its proof is actually shown in Lemma A.3 when characterizing the order of $R_{n3}(y,x;\alpha)$ defined in Appendix A.

Corollary 4.3 *Under the Assumptions (D)-(K), it holds that*

$$
\sup_{\alpha \in \mathcal{A}} \sup_{x \in \mathcal{J}} \sup_{y \in [y_l, y_u^0]} |\widehat{F}(y|x;\alpha) - F_Y(y|x;\alpha)| = O_p\left(\sqrt{\frac{\log n}{nh^p}} \right).
$$

In particular, it holds if we set $y = x'\beta$ for those $x'\beta \in [y_l, y_u^0]$.

4.2 Validity of the Asymptotic Confidence Sets

In order to prove the asymptotic exactness of the confidence sets defined in (3.6) and (3.7), we show that $\widehat{T}_n(\beta)$ and $\widetilde{T}_n(\beta)$ are both asymptotically normal upon proper centering and normalization. Noting that those two statistics resemble the population criterion function closely, we show below that $\widehat{\alpha}(\beta)$ and $\widetilde{\alpha}(\beta)$ converge in probability to $\alpha(\beta)$ and prove the stochastic equicontinuity (SE) of $[T_n(\beta;\alpha) - B_n(\beta;\alpha)]$ w.r.t α in the local neighborhood of $\alpha(\beta)$ whose radius is determined by the convergence rate of $\widehat{\alpha}(\beta)$ or $\widetilde{\alpha}(\beta)$ to $\alpha(\beta)$. Proving consistency and getting convergence rate for estimators obtained from minimizing a kernel based criterion function is akin to a problem from *smooth* minimum distance estimation, as shown in Linton (1997, 1998), also see Lavergne and Patilea (2013) on a recent account. For $\alpha = \alpha(\beta)$, the asymptotic distribution of $[T_n(\beta;\alpha) - B_n(\beta;\alpha)]$ is determined by a degenerate U-statistic similar to the test statistics in Hardle and Mammen (1993), Fan (1994), Fan and Li (1996), and Zheng (1996); when $\alpha \neq \alpha(\beta)$, $[T_n(\beta;\alpha) - B_n(\beta;\alpha)]$ could be decomposed as the degenerate U-statistic, a non-degenerate U-statistic and the deterministic drifting term: $\int_{\mathcal{J}} [F_Y(x'\beta|x;\alpha) - q]^2 f_X^2(x) dx$. The SE of $[T_n(\beta;\alpha) - B_n(\beta;\alpha)]$ would be proved by showing the SE of the degenerate U-process and negligibility of the other two terms when α approaches $\alpha(\beta)$ sufficiently fast.

We need two more sets of assumptions to show the validity of our confidence sets, one (Assumptions (V0) and (V1)) for $CS^N_{1-\epsilon, \widehat{T}_n}$ and one (Assumptions (V0) and (V2)) for $CS^N_{1-\epsilon, \widetilde{T}_n}$.

Assumption (V0). For all $\beta \in \mathcal{B}$, $x'\beta \in [y_l, y_u^0]$ for all $x \in \mathcal{J}$.

Assumption (V1). (i) For any $\beta \in \mathcal{B}_I^O$, the corresponding $\alpha(\beta)$ belongs to the interior of $\mathcal{A}$; (ii) In addition to Assumption (H), we assume that $nh^{2s} \to 0$; (iii) In addition to Assumption (G), those functions $\dot{\varphi}_\alpha'(u)$, $\dot{\varphi}_\alpha^{-1}(u)$ exist and are continuous and bounded in the range of $[\delta_0, 1]$ and $[0, \infty)$ respectively.

Assumption (V2). (i) In addition to Assumption (H), there exists a sequence $\varepsilon_n \to 0$ such that $\sqrt{\frac{n}{\log n}}\varepsilon_n \to \infty$ and $nh^p\varepsilon_n \to 0$; (ii) In addition to Assumption (G), the Lipschitz continuity property with respect to α holds for $\psi_\alpha(\cdot) = 1/\varphi_\alpha'(\cdot)$ or $\psi_\alpha(\cdot) = \varphi_\alpha''(\cdot)$ with positive constant L from below:

$$L\,|\alpha_1 - \alpha_2| \leq \sup_{u \in [\delta_0, 1]} \left|\psi_{\alpha_1}(u) - \psi_{\alpha_2}(u)\right|.$$

Assumption (V0) is used for both CSs. It ensures that all the conditional quantiles of potential interest stay sufficiently far away from the right end support point of V. When independent censoring is assumed, similar restrictions have also appeared in Peng and Huang (2008), Wang and Wang (2009) for β in a neighborhood of the (point identified) true β_o. There is a distinction between Assumption (V1) and Assumption (V2) because of the slightly different arguments used in the proofs for $CS_{1-\epsilon,\widehat{T}_n}^N$ and $CS_{1-\epsilon,\widetilde{T}_n}^N$. The consistency of $\widehat{\alpha}(\beta)$ follows the standard way to contrast sample criterion function and population criterion function, viewed as a minimum distance estimator. Its rate of convergence is shown once the requirement that $\alpha(\beta)$ stays in the interior and enough smoothness (w.r.t α) on the generator function are satisfied. In comparison, a different route is taken for $\widetilde{\alpha}(\beta)$ as in Santos (2006). Its consistency and rate of convergence will be achieved through the different convergence stochastic orders of the test statistic and a careful study of the local neighborhood of the 'null' set for α (see Santos, 2006):

$$\mathcal{A}_o^{\varepsilon_n} = \left\{\alpha \in \mathcal{A} : \int_{\mathcal{J}} \left[F_Y(x'\beta|x;\alpha) - q\right]^2 f_X^2(x)\,dx \leq \varepsilon_n, \text{ with } \varepsilon_n \to 0\right\}. \tag{3.16}$$

Notice that when $\varepsilon_n \to 0$, $\mathcal{A}_o^{\varepsilon_n}$ will shrink to the singleton $\{\alpha(\beta)\}$; on the other hand, when $\alpha \notin \mathcal{A}_o^{\varepsilon_n}$, the sample criterion function would be shown to be explosive.

The first main result in this section establishes the asymptotic distributions of the test statistics $\widehat{T}_n(\beta)$ and $\widetilde{T}_n(\beta)$ for $\beta \in \mathcal{B}_I^O$, thereafter the asymptotic size property of our confidence sets follows immediately.

Proposition 4.4 *Suppose Assumptions (SC), (D)-(K), and (V0) hold, then for $\beta \in \mathcal{B}_I^O$ with the unique $\alpha(\beta)$,*

$$\frac{nh^{p/2}\left[T_n(\beta; \alpha(\beta)) - B_n(\beta; \alpha(\beta))\right]}{\sqrt{\Sigma(\beta; \alpha(\beta))}} \Longrightarrow N(0, 1). \tag{3.17}$$

In addition if (V1) holds, we have:

$$\frac{nh^{p/2}\left[T_n(\beta; \widehat{\alpha}(\beta)) - \widehat{B}_n(\beta; \widehat{\alpha}(\beta))\right]}{\sqrt{\widehat{\Sigma}(\beta; \widehat{\alpha}(\beta))}} \Longrightarrow N(0, 1); \tag{3.18}$$

if (V2) holds, we have:

$$\frac{nh^{p/2}\left[T_n(\beta; \widetilde{\alpha}(\beta)) - \widehat{B}_n(\beta; \widetilde{\alpha}(\beta))\right]}{\sqrt{\widehat{\Sigma}(\beta; \widetilde{\alpha}(\beta))}} \Longrightarrow N(0, 1), \tag{3.19}$$

22

where $B_n(\beta;\alpha) = n^{-1}h^{-p}\int K^2(u)\,du\int \sigma^2(x'\beta|x;\alpha)\,f_X(x)\,dx$ in which $\sigma^2(x'\beta|x;\alpha)$ is defined in (C.2) in Appendix C, and $\Sigma(\beta;\alpha)$ is defined in (B.8) in Appendix B with uniformly consistent estimators $\widehat{B}_n(\beta,\alpha)$ and $\widehat{\Sigma}(\beta;\alpha)$ respectively.

Theorem 4.5 *Under the assumptions (SC), (D)-(K), and (V0), our confidence sets have pointwisely asymptotic exact size: for $\forall\beta\in\mathcal{B}_I^O$, if (V1) holds, we get: $\lim_{n\to\infty}\Pr\left\{\beta\in CS_{1-\epsilon,\widehat{T}_n}^N\right\} = 1-\epsilon$; or if (V2) holds, we get: $\lim_{n\to\infty}\Pr\left\{\beta\in CS_{1-\epsilon,\widetilde{T}_n}^N\right\} = 1-\epsilon$.*

Remark 4.6 *Both test statistics and confidence sets have their own merits. $\widehat{T}_n(\beta)$ circumvents the need to estimate the complicated drifting term and asymptotic variance term when the minimization over $\alpha\in\mathcal{A}$ is conducted; on the other hand, even without the separation assumption (SC), the procedure based on $\widetilde{T}_n(\beta)$ is still asymptotically valid although may be conservative, similar to Jun and Pinske (2009) in a different context:*

$$\limsup_n \Pr\left\{\min_{\alpha\in\mathcal{A}}\left|\frac{nh^{p/2}\left[T_n(\beta;\alpha)-\widehat{B}_n(\beta,\alpha)\right]}{\sqrt{\widehat{\Sigma}(\beta;\alpha(\beta))}}\right|\geq z_{1-\epsilon/2}\right\}$$

$$\leq \lim_n \Pr\left\{\left|\frac{nh^{p/2}\left[T_n(\beta;\alpha(\beta))-\widehat{B}_n(\beta,\alpha(\beta))\right]}{\sqrt{\widehat{\Sigma}(\beta;\alpha(\beta))}}\right|\geq z_{1-\epsilon/2}\right\}$$

$$= 1-\epsilon.$$

4.3 Bootstrap Confidence Sets

It is well documented in the literature on model specification testing that the normal approximation of the distribution of the kernel-based test statistics may not work well in small samples, see Hardle and Mammen (1993) and resampling methods such as bootstrap may be used. Below we present bootstrap analogues of the asymptotic confidence sets $CS_{1-\epsilon,\widehat{T}_n}^N$ and $CS_{1-\epsilon,\widetilde{T}_n}^N$.

Let

$$T_{n,b}^*(\beta;\alpha) \equiv \int_{\mathcal{J}}\left[\frac{1}{n}\sum_{i=1}^n W_\gamma(x,X_i)\,c_b^*\left(x'\beta|Z_i,x;\alpha\right)\right]^2 dx,$$

where

$$c_b^*(y|Z_i,x;\alpha) \tag{3.20}$$

$$= \frac{-M_{i,b}^*}{\varphi_\alpha'\left\{\widehat{S}(y|x;\alpha)\right\}}\Big[\int_{y_l}^y \varphi_\alpha''\left\{\widehat{S}_V(v|x)\right\}\left[I(V_i\leq v)-\widehat{F}_V(v|X_i)\right]d\widehat{F}_{V,D=1}(v|x)$$

$$-\varphi_\alpha'\left\{\widehat{S}_V(y|x)\right\}\left[I(V_i\leq v,D_i=1)-\widehat{F}_{V,D=1}(y|X_i)\right]$$

$$-\int_{y_l}^y \varphi_\alpha''\left\{\widehat{S}_V(v|x)\right\}\left[I(V_i\leq v,D_i=1)-\widehat{F}_{V,D=1}(v|X_i)\right]d\widehat{F}_V(v|x)\Big]$$

in which the perturbation variables $\left\{M_{i,b}^*\right\}_{i=1}^n$ are independently generated with zero mean and unit variance from, for example, the standard normal distribution or centered unit exponential distribution. Thereafter

we define $T^*_{n,b}(\beta; \widehat{\alpha}(\beta))$ and $T^*_{n,b}(\beta; \widetilde{\alpha}(\beta))$ accordingly.[9]

We could generate $\left\{ M^*_{i,b} \right\}^n_{i=1}$ for $b = 1, ..., B$ and obtain $\left\{ T^*_{n,b}(\beta; \widehat{\alpha}(\beta)) \right\}^B_{b=1}$ or $\left\{ T^*_{n,b}(\beta; \widetilde{\alpha}(\beta)) \right\}^B_{b=1}$. The bootstrap critical values are defined as

$$c^B_{n,\widehat{T}_n}(\beta, 1-\epsilon) = \inf \left\{ t : \frac{1}{B} \sum^B_{b=1} I \left\{ nh^{p/2} T^*_{n,b}(\beta; \widehat{\alpha}(\beta)) \le t \right\} \ge 1 - \epsilon \right\} \text{ and}$$

$$c^B_{n,\widetilde{T}_n}(\beta, 1-\epsilon) = \inf \left\{ t : \frac{1}{B} \sum^B_{b=1} I \left\{ nh^{p/2} T^*_{n,b}(\beta; \widetilde{\alpha}(\beta)) \le t \right\} \ge 1 - \epsilon \right\}.$$

Hence the following two bootstrap confidence sets are immediate:

$$\begin{aligned} CS^B_{1-\epsilon,\widehat{T}_n} &= \left\{ \beta \in \mathcal{B} : nh^{p/2}\widehat{T}_n(\beta) \le c^B_{n,\widehat{T}_n}(\beta, 1-\epsilon) \right\} \text{ and} \\ CS^B_{1-\epsilon,\widetilde{T}_n} &= \left\{ \beta \in \mathcal{B} : nh^{p/2}\widetilde{T}_n(\beta) \le c^B_{n,\widetilde{T}_n}(\beta, 1-\epsilon) \right\}. \end{aligned}$$

Theorem 4.7 *Under the assumptions (AC), (D)-(K), and (V0), our bootstrap confidence sets have pointwisely asymptotic exact size: for $\forall \beta \in \mathcal{B}^O_I$, if (V1) holds, we get: $\lim_{n\to\infty} \Pr \left\{ \beta \in CS^B_{1-\epsilon,\widehat{T}_n} \right\} = 1 - \epsilon$; or if (V2) holds, we get: $\lim_{n\to\infty} \Pr \left\{ \beta \in CS^B_{1-\epsilon,\widetilde{T}_n} \right\} = 1 - \epsilon$.*

5 An Empirical Application

In this section, we illustrate our methodology on a real data set used in Wang and Wang (2009). The data comes from a study on the survival of patients after acute myocardial infarction conducted at the University Clinical Center in Ljubljana and is publicly available in R package *relsurv*. It consists of $n = 1,040$ observations with 493 censored observations. The variable of interest, i.e., the survival time, is recorded in days and we transform it into the unit scale $[0, 1]$ by the empirical probability integral transformation. There are two regressors of mixed type; the discrete regressor is Gender (with 751 observations from Male vs. 289 observations from Female) and the continuous regressor is Age (we again transform the original data into the unit scale between $[0, 1]$).[10] The exact cause of censoring is unknown in this data, however in typical clinical studies censoring is not merely of administrative type (censoring occurs because the study simply terminates). Patients might be removed if there is evidence that treatment is ineffective, or patients withdraw themselves because of side effects or they die due to other causes (Fleming and Harrington, 1991). Hence it is reasonable to expect some positive dependence between Y and C in those situations.

We compare our confidence set $CS^B_{95\%,\widehat{T}_n}$ with two bootstrap confidence intervals in Portnoy (2003), Peng and Huang (2008) (*Por, PH* in Table 2 below respectively) where conditional independent censoring is assumed. Those two approaches could be automatically implemented in Roger Koenker's R package *quantreg*. Linear quantile regressions, with intercept $\overline{\beta}$ and slope β, are fitted on subsets splitted according

---

[9] One could bootstrap the drifting term when calculating $\widetilde{T}_n(\beta)$

[10] In comparison, Wang and Wang (2009) take the log transform of the original survival time. Also, even though the procedure in Wang and Wang (2009) calls for ordinary kernel smoothing across the variable Age, they still work with the original one, which is integer-valued.

to gender groups and we report the bootstrap confidence intervals on the slope coefficient β. Notice that the data between different gender groups are very unbalanced, which justifies the smoothing across the discrete regressor here (Li and Racine, 2007). Referring to the actual implementation of our approach, since the continuous regressor is univariate, we let $K\left(u\right) = \frac{15}{16}\left(1 - u^2\right)^2$, the bisquare kernel which is a second order kernel. Moreover the two tuning parameters are set to be $h = 2n^{-1/4}$, $\lambda = n^{-1/2}$, and the truncation of the integral is restricted to be $\mathcal{J} = [0.1, 0.9]$. Also the perturbation variable $M_{i,b}^*$ in the bootstrap weight is taken to be standard normal. Needless to say, our procedure is computationally more intensive as for every regression coefficient under consideration, a minimization over α is carried out and bootstrap is needed to obtain the critical value. To reduce computational cost, we simply set the the number of bootstrap replications at 100 and do a grid search over $\beta \in [-5, 1]$ with grid length equal to 0.01. Notice that our construction of confidence set leads to simultaneous inference on both the slope β and the intercept $\overline{\beta}$. However, the bootstrap intervals in Portnoy (2003), Peng and Huang (2008) are for the slope and intercept separately. For fair comparison, we have picked a projection based version of ours by considering all β not rejected while $\overline{\beta}$ runs over in $[0, 1]$ with a grid length 0.01.[11] To check the sensitivity of conclusions from maintaining ICM assumption, we consider two scenarios, small vs. moderate deviations. In the former case, we set $\tau \in [0, 0.2]$ whereas in the latter case $\tau \in [0, 0.5]$. Both confidence sets based on Clayton copula and Gumbel copula are reported to examine the effect of employing different copula generator functions when the specified dependence level τ coincides. The results are reported in Table 2 below, where DCM denotes dependent censoring mechanism.

Some remarks follow from Table 2. First of all, our confidence sets turn out to be intervals for this particular case, so there are no holes in between. Second, despite we choose the projection based inference (which might be conservative) and allow for a wider range of dependence, our confidence intervals are not necessarily wider than those in *Por, PH* for the female group. The frequency approach by splitting the data leaves too few observations in the female group and as noted in Wang and Wang (2009), *Por, PH* tend to be unstable for small samples. Third, the conclusion on negative effect (the sign) of aging on the survival time is robust even when we allow for $\tau \in [0, 0.5]$, but the one on the exact magnitude might change. For example, in the male group when $q = 3/4$, the ICM intervals lead to the accelerating effect ($|\beta| > 1$) from aging, but this could be overturned when we allow for moderate positive dependence. Finally, the difference between fitting a Clayton and Gumbel is almost negligible, never larger than 0.08. So the exact shape of a generator function plays only a minor role.

6   Concluding Remarks

Assuming an Archimedean copula for the dependent variable Y and the censoring variable C, we have proposed a two-step method for studying partial identification of the quantile coefficient β_o in quantile regression

---

[11] The reason we set the parameter space $\mathcal{B}$ to be $[0, 1] \times [-5, 1]$ is that it includes the widest interval coming from Portnoy (2003), Peng and Huang (2008) and is slightly enlarged.

Table 2: Confidence Sets

q	Male			Female		
	1/4	1/2	3/4	1/4	1/2	3/4
ICM	$\tau = 0$					
Por	$[-0.73, -0.49]$	$[-1.69, -0.73]$	$[-2.26, -1.34]$	$[-0.72, -0.28]$	$[-2.07, -0.82]$	$[-3.73, -0.80]$
PH	$[-0.75, -0.47]$	$[-2.18, -0.46]$	$[-2.38, -1.30]$	$[-0.76, -0.19]$	$[-2.12, -0.93]$	$[-4.39, -0.84]$
DCM	$\tau \in [0, 0.2]$					
Clay	$[-0.70, -0.44]$	$[-1.83, -0.51]$	$[-2.33, -1.31]$	$[-0.76, -0.40]$	$[-2.14, -1.05]$	$[-2.52, -1.29]$
Gum	$[-0.71, -0.41]$	$[-1.85, -0.59]$	$[-2.30, -1.28]$	$[-0.73, -0.38]$	$[-2.18, -1.07]$	$[-2.50, -1.24]$
DCM	$\tau \in [0, 0.5]$					
Clay	$[-0.80, -0.23]$	$[-1.94, -0.24]$	$[-2.47, -0.82]$	$[-0.84, -0.21]$	$[-2.21, -0.87]$	$[-2.73, -0.75]$
Gum	$[-0.78, -0.27]$	$[-1.98, -0.25]$	$[-2.44, -0.86]$	$[-0.84, -0.15]$	$[-2.20, -0.84]$	$[-2.76, -0.71]$

with possibly dependent censoring. For a broad class of Archimedean copulas, we have characterized an outer set of the identified set of β_o via inequality constraints. Most Archimedean copulas are characterized by a single parameter and are also ordered. Using such ordered parametric Archimedean copulas, we have developed an econometric method for conducting sensitivity analysis to examine the sensitivity of conclusions on β_o to the independent censoring mechanism commonly adopted in empirical work. Interpreting Y and the censoring variable C in our model as two competing risks, our methodology should be useful in duration analysis with possibly dependent competing risks.

As a first step towards developing formal sensitivity analysis in censored quantile regression models, we have opted for simplicity instead of generality in this paper. Many important extensions are worthwhile. First, in practice, it is also of interest to test certain linear restriction on the parameter in the identified set, e.g., whether a particular component equals to zero. Without point identification, such testing problems can be formulated as in Santos (2006, 2012), i.e., we check whether there is at least one β_o satisfying the linear restriction under the null,

$$H_0 : \mathcal{B}_o \cap R_\beta \neq \emptyset, \text{ vs. } H_1 : \mathcal{B}_o \cap R_\beta = \emptyset,$$

where $R_\beta = \{\beta \in \mathcal{B} : R\beta = r\}$, with a given matrix R and vector r. Second, endogenous regressors can be incorporated in our framework as in Khan and Tamer (2009) or one may extend Manski (1994) to allow for censored outcome variables in quantile selection models.

7 Appendix A: Asymptotic Linear Representation of the Plug-in Estimator $\widehat{F}(y|x; \alpha)$

In this section, let M be a universal finite constant and Δ be an intermediate value appearing in a Taylor expansion. Their specific values are of no importance, so may vary from line to line. We will need to handle various functional classes $\mathcal{F}$ using local U-process techniques collected in Appendix C, and will refer to the term σ^2, s.t. $\left\| P^m f^2 \right\|_{\mathcal{F}} \leq \sigma^2$ for $m = 1$ or 2 as the maximal variance (see Appendix C). The following

notations would facilitate the proofs:

$$H_V(v|x) \equiv S_V(v|x)/f_X(x), \quad H_{V,D=1}(v|x) \equiv S_{V,D=1}(v|x)/f_X(x), \tag{A.1}$$

$$\widehat{H}_V(v|x) \equiv \widehat{S}_V(v|x)/\widehat{f}_X(x), \quad \widehat{H}_{V,D=1}(v|x) \equiv \widehat{S}_{V,D=1}(v|x)/\widehat{f}_X(x),$$

where the various estimators are all of kernel type introduced in the main text.

The following facts would be used repeatedly in the proofs, so we list them below for easy reference.

Facts: Under the Assumptions (D)-(K), the following results hold:

$$\sup_{x \in \mathcal{J}} \left| \widehat{f}_X(x) - f_X(x) \right| = O_p\left(\sqrt{\frac{\log n}{nh^p}} \right), \tag{A.2}$$

$$\sup_{x \in \mathcal{J}} \sup_{y \in [y_l, y_u^0]} \left| \widehat{F}_V(y|x) - F_V(y|x) \right| = O_p\left(\sqrt{\frac{\log n}{nh^p}} \right), \text{ and} \tag{A.3}$$

$$\sup_{x \in \mathcal{J}} \sup_{y \in [y_l, y_u^0]} \left| \widehat{F}_{V,D=1}(y|x) - F_{V,D=1}(y|x) \right| = O_p\left(\sqrt{\frac{\log n}{nh^p}} \right). \tag{A.4}$$

The proofs of the above facts would follow from Theorems 1 and 3 in Einmahl and Mason (2005), combining the arguments dealing with discrete regressors as in Li and Racine (2007). For completeness, we sketch a proof of (A.3) in Appendix C.

We now give a main proof of Theorem 4.1, where the convergence rates of various terms used in the proof would be collected in a series of lemmas following the main proof.

Proof. Recall that

$$\widehat{F}(y|x;\alpha) = 1 - \varphi_\alpha^{-1}\left[-\int_{y_l}^y \varphi_\alpha'\left(\widehat{S}_V(v|x)\right) d\widehat{F}_{V,D=1}(v|x) \right]$$

$$= 1 - \varphi_\alpha^{-1}\left[-\sum_{V_i \leq y, D_i=1} \varphi_\alpha'\left(\widehat{S}_V(V_i|x)\right) w_{n\gamma}(x, X_i) \right].$$

Straightforward algebra shows that

$$-\int_{y_l}^y \varphi_\alpha'\left(\widehat{S}_V(v|x)\right) d\widehat{F}_{V,D=1}(v|x) + \int_{y_l}^y \varphi_\alpha'(S_V(v|x)) dF_{V,D=1}(v|x)$$

$$= -\int_{y_l}^y \left[\varphi_\alpha'\left(\widehat{S}_V(v|x)\right) - \varphi_x'(S_V(v|x)) \right] dF_{V,D=1}(v|x)$$

$$- \int_{y_l}^y \varphi_\alpha'(S_V(v|x)) d\left[\widehat{F}_{V,D=1}(v|x) - F_{V,D=1}(v|x) \right]$$

$$- \int_{y_l}^y \left[\varphi_\alpha'\left(\widehat{S}_V(v|x)\right) - \varphi_\alpha'(S_V(v|x)) \right] d\left[\widehat{F}_{V,D=1}(v|x) - F_{V,D=1}(v|x) \right].$$

Regarding the above expression, we now take a second order Taylor expansion on the first term and integrate by parts on the second term on the right hand side:

$$-\int_{y_l}^y \varphi_\alpha'\left(\widehat{S}_V(v|x)\right) d\widehat{F}_{V,D=1}(v|x) + \int_{y_l}^y \varphi_\alpha'(S_V(v|x)) dF_{V,D=1}(v|x)$$

$$= \int_{y_l}^y \varphi_\alpha''\{S_V(v|x)\} \left[\widehat{F}_V(v|x) - F_V(v|x) \right] dF_{V,D=1}(v|x) - \varphi_\alpha'\{S_V(v|x)\} \left[\widehat{F}_{V,D=1}(v|x) - F_{V,D=1}(v|x) \right]$$

$$- \int_{y_l}^y \varphi_\alpha''\{S_V(v|x)\} \left[\widehat{F}_{V,D=1}(v|x) - F_{V,D=1}(v|x) \right] dF_V(v|x)] + R_{n1}(y, x; \alpha) + R_{n2}(y, x; \alpha)$$

27

where

$$R_{n1}\left(y,x;\alpha\right) = -\int_{y_l}^{y} \frac{\varphi_{\alpha}^{(3)}\left\{\Delta\right\}}{2}\left[\widehat{F}_V\left(v|x\right) - F_V\left(v|x\right)\right]^2 dF_{V,D=1}\left(v|x\right) \text{ and} \tag{A.5}$$

$$R_{n2}(y,x;\alpha) = -\int_{y_l}^{y}\left[\varphi_{\alpha}^{'}\left(\widehat{S}_V\left(v|x\right)\right) - \varphi_{x}^{'}\left(S_V\left(v|x\right)\right)\right] d\left[\widehat{F}_{V,D=1}\left(v|x\right) - F_{V,D=1}\left(v|x\right)\right]. \tag{A.6}$$

By Lemmas A.1 and A.2, we get

$$\sup_{\alpha\in\mathcal{A}}\sup_{x\in\mathcal{J}}\sup_{y\in[y_l,y_u^0]} R_{n1}\left(y,x;\alpha\right) = O_p\left(\frac{\log n}{nh^p}\right) \text{ and}$$

$$\sup_{\alpha\in\mathcal{A}}\sup_{x\in\mathcal{J}}\sup_{y\in[y_l,y_u^0]} R_{n2}(y,x;\alpha) = O_p\left(\frac{\log n}{nh^p}\right).$$

Hence the result follows once we show the order of those remainder terms.

Finally to see why (3.14) and (3.15) hold, it suffices to illustrate on the first term inside the bracket of (3.13):

$$\int_{y_l}^{y} \varphi_{\alpha}^{''}\left\{S_V\left(v|x\right)\right\}\left[\widehat{F}_V\left(v|x\right) - F_V\left(v|x\right)\right] dF_{V,D=1}\left(v|x\right)$$

$$= \frac{1}{\widehat{f}_X\left(x\right)n}\sum_{i=1}^{n} W_{\gamma}\left(x,X_i\right)\int_{y_l}^{y}\varphi_{\alpha}^{''}\left\{S_V\left(v|x\right)\right\}\left[I\left(V_i\leq v\right) - F_V\left(v|x\right)\right] dF_{V,D=1}\left(v|x\right)$$

$$= \frac{1}{f_X\left(x\right)n}\sum_{i=1}^{n} W_{\gamma}\left(x,X_i\right)\int_{y_l}^{y}\varphi_{\alpha}^{''}\left\{S_V\left(v|x\right)\right\}\left[I\left(V_i\leq v\right) - F_V\left(v|x\right)\right] dF_{V,D=1}\left(v|x\right)$$

$$+ \frac{f_X\left(x\right) - \widehat{f}_X\left(x\right)}{\widehat{f}_X\left(x\right)f_X\left(x\right)n}\sum_{i=1}^{n} W_{\gamma}\left(x,X_i\right)\int_{y_l}^{y}\varphi_{\alpha}^{''}\left\{S_V\left(v|x\right)\right\}\left[I\left(V_i\leq v\right) - F_V\left(v|x\right)\right] dF_{V,D=1}\left(v|x\right),$$

where the second term is bounded from above uniformly in $\alpha \in \mathcal{A}$ by

$$M\sup_{x\in\mathcal{J}}\left|f_X\left(x\right) - \widehat{f}_X\left(x\right)\right| \times \sup_{x\in\mathcal{J}}\sup_{y\in[y_l,y_u^0]}\left|\widehat{F}_V\left(y|x\right) - F_V\left(y|x\right)\right| = O_p\left(\frac{\log n}{nh^p}\right),$$

whose rate of convergence follows from the stated facts (A.2) and (A.3). ∎

Lemma A.1 *Under Assumptions (D)-(K), we have:* $\sup_{\alpha\in\mathcal{A}}\sup_{x\in\mathcal{J}}\sup_{y\in[y_l,y_u^0]}\left|R_{n1}\left(y,x;\alpha\right)\right| = O_p\left(\frac{\log n}{nh^p}\right).$

Proof. We pull out the integrand on the right hand side of (A.5) and notice that $dF_{V,D=1}\left(\cdot|x\right)$ is a finite measure for *a.s.* $x\in\mathcal{J}$:

$$\sup_{\alpha\in\mathcal{A}}\sup_{x\in\mathcal{J}}\sup_{y\in[y_l,y_u^0]}\left|R_{n1}\left(y,x;\alpha\right)\right| \leq \frac{1}{2}\sup_{\alpha\in\mathcal{A}}\sup_{x\in\mathcal{J}}\sup_{y\in[y_l,y_u^0]}\left|\varphi_{\alpha}^{(3)}\left(\Delta\right)\right|\left[\widehat{F}_V\left(y|x\right) - F_V\left(y|x\right)\right]^2.$$

Now Assumption (G1) on the copula generator and the stated fact (A.3) give the desired rate. ∎

The next lemma characterizing the order of $R_{n2}\left(y,x;\alpha\right)$ is most cumbersome. Braekers and Veraverbeke (2005) first discretize along the y-axis and then bound the local oscillation uniformly for $x\in\mathcal{J}$ invoking certain maximal inequalities to arrive at the rate of $\left(\frac{\log n}{nh^p}\right)^{3/4}$. We shall improve this rate to $\frac{\log n}{nh^p}$ because the dominating term is a second order U-process as in Major (2006). A similar proof appears in Lemma 3.1 of Lopez (2011) dealing with the conditional Kaplan-Meier estimator.

Lemma A.2 *Under Assumptions (D)-(K), we have:* $\sup_{\alpha \in \mathcal{A}} \sup_{x \in \mathcal{J}} \sup_{y \in [y_l, y_u^0]} |R_{n2}(y, x; \alpha)| = O_p\left(\frac{\log n}{n h^p}\right).$

Proof. Given Assumption (G), we can decompose the expression for $R_{n2}(y, x; \alpha)$ in (A.6) into two terms as:

$$-\int_{y_l}^{y} \left[\varphi_\alpha'\left(\widehat{S}_V(v|x)\right) - \varphi_\alpha'(S_V(v|x))\right] d\left[\widehat{F}_{V,D=1}(v|x) - F_{V,D=1}(v|x)\right]$$

$$= \int_{y_l}^{y} \varphi_\alpha''(S_V(v|x)) \left[\widehat{S}_V(v|x) - S_V(v|x)\right] d\left[\widehat{S}_{V,D=1}(v|x) - S_{V,D=1}(v|x)\right]$$

$$+ \int_{y_l}^{y} \varphi_\alpha^{(3)}(\Delta) \left[\widehat{S}_V(v|x) - S_V(v|x)\right]^2 d\left[\widehat{S}_{V,D=1}(v|x) - S_{V,D=1}(v|x)\right].$$

Apparently the second term is negligible and of order $\left(\frac{\log n}{n h^p}\right)^{3/2}$ by (A.3) and (A.4). Some simplification occurs in handling the first term. Recalling the definitions given in (A.1), we have:

$$\widehat{S}_V(v|x) - S_V(v|x) = \frac{1}{f_X(x)} \left[\widehat{H}_V(v|x) - H_V(v|x)\right] + \frac{\left[f_X(x) - \widehat{f}_X(x)\right] H_V(v|x)}{f_X^2(x)}$$

$$+ \frac{\left[f_X(x) - \widehat{f}_X(x)\right]\left[\widehat{H}_V(v|x) - H_V(v|x)\right]}{f_X(x) \widehat{f}_X(x)} + \frac{\left[f_X(x) - \widehat{f}_X(x)\right]^2 H_V(v|x)}{f_X^2(x) \widehat{f}_X(x)}.$$

A similar expression holds for $\left[\widehat{S}_{V,D=1}(v|x) - S_{V,D=1}(v|x)\right]$. The latter two terms (also those from conditional sub-survival functions) would be of smaller order by (A.2), (A.3), (A.4), and Assumption (F). Now it reduces to bound the dominating terms as

$$\int_{y_l}^{y} \frac{\varphi_\alpha''(S_V(v|x))}{f_X^2(x)} \left[\widehat{H}_V(v|x) - H_V(v|x)\right] d\left[\widehat{H}_{V,D=1}(v|x) - H_{V,D=1}(v|x)\right] \text{ and}$$

$$\int_{y_l}^{y} \frac{\varphi_\alpha''(S_V(v|x))}{f_X^2(x)} \left[f_X(x) - \widehat{f}_X(x)\right] d\left[\widehat{H}_{V,D=1}(v|x) - H_{V,D=1}(v|x)\right].$$

It suffices to demonstrate with the first term. We introduce further notations as

$$\widetilde{H}_V(v|x) = E\left[\widehat{H}_V(v|x)\right] \text{ and } \widetilde{H}_{V,D=1}(v|x) = E\left[\widehat{H}_{V,D=1}(v|x)\right].$$

Standard bias-variance decomposition leads to

$$\int_{y_l}^{y} \frac{\varphi_\alpha''(S_V(v|x))}{f_X^2(x)} \left[\widehat{H}_V(v|x) - \widetilde{H}_V(v|x)\right] d\left[\widehat{H}_{V,D=1}(v|x) - H_{V,D=1}(v|x)\right]$$

$$+ \int_{y_l}^{y} \frac{\varphi_\alpha''(S_V(v|x))}{f_X^2(x)} \left[\widetilde{H}_V(v|x) - H_V(v|x)\right] d\left[\widehat{H}_{V,D=1}(v|x) - H_{V,D=1}(v|x)\right]$$

$$= \int_{y_l}^{y} \frac{\varphi_\alpha''(S_V(v|x))}{f_X^2(x)} \left[\widehat{H}_V(v|x) - \widetilde{H}_V(v|x)\right] d\left[\widehat{H}_{V,D=1}(v|x) - \widetilde{H}_{V,D=1}(v|x)\right]$$

$$+ \int_{y_l}^{y} \frac{\varphi_\alpha''(S_V(v|x))}{f_X^2(x)} \left[\widehat{H}_V(v|x) - \widetilde{H}_V(v|x)\right] d\left[\widetilde{H}_{V,D=1}(v|x) - H_{V,D=1}(v|x)\right]$$

$$+ \int_{y_l}^{y} \frac{\varphi_\alpha''(S_V(v|x))}{f_X^2(x)} \left[\widetilde{H}_V(v|x) - H_V(v|x)\right] d\left[\widehat{H}_{V,D=1}(v|x) - \widetilde{H}_{V,D=1}(v|x)\right]$$

$$+ \int_{y_l}^{y} \frac{\varphi_\alpha''(S_V(v|x))}{f_X^2(x)} \left[\widetilde{H}_V(v|x) - H_V(v|x)\right] d\left[\widetilde{H}_{V,D=1}(v|x) - H_{V,D=1}(v|x)\right].$$

Analogous expansion as in Lopez (2011) could be taken to bound the last three terms involving bias, and the order is of $\left(\frac{\log n}{nh^p}\right)^{1/2}\left(h^s + \sum\lambda_j\right) + \left(h^s + \sum\lambda_j\right)^2$, smaller than the dominating one under Assumption (H). In the end, it boils down to bound the first term which is a second order degenerate U-statistic plus the diagonal term:

$$\int_{y_l}^y \frac{\varphi_\alpha''(S_V(v|x))}{f_X^2(x)}\left[\widehat{H}_V(v|x) - \widetilde{H}_V(v|x)\right]d\left[\widehat{H}_{V,D=1}(v|x) - \widetilde{H}_{V,D=1}(v|x)\right]$$

$$= \frac{1}{n^2h^{2p}}\sum_{i=1}^n f(Z_i, Z_i) + \frac{1}{n^2h^{2p}}\sum_{i\neq j} f(Z_i, Z_j)$$

where

$$f(Z_1, Z_2) = D_1\frac{\varphi_\alpha''(S_V(V_1|x))}{f_X^2(x)}1\{V_1 > y\}h^p W_\gamma(x, X_1) H(X_2, V_2, V_1)$$

$$- \int d_1\frac{\varphi_\alpha''(S_V(v_1|x))}{f_X^2(x)}1\{v_1 > y\}h^p W_\gamma(x, x_1) H(X_2, V_2, v_1)dP_Z(v_1, d_1, x_1)$$

with

$$H(X_2, V_2, u) = 1\{V_2 > u\}h^p W_\gamma(x - X_2) - \int 1\{v_2 > u\}h^p W_\gamma(x, x_2)dP_Z(v_2, x_2).$$

The diagonal term could be bounded from above uniformly by

$$M\sup_x\left[\frac{1}{n^2}\sum_{i=1}^n W_\gamma^2(x, X_i)\right] \leq M\sup_x\left[\frac{1}{n^2}\sum_{i=1}^n K_h^2(x - X_i)\right] = O_p\left(\frac{1}{nh^p}\right).$$

When it comes to the degenerate U-process indexed by (y, x, α, γ), the maximal variance for the kernel function $f(\cdot, \cdot)$ is of order h^{2p} as in Lopez (2011). Hence the application of (C.4) gives the desired rate $O_p\left(\frac{\log n}{nh^p}\right)$. ∎

Lemma A.3 *Under Assumptions (D)-(K), we have:* $\sup_{\alpha\in\mathcal{A}}\sup_{x\in\mathcal{J}}\sup_{y\in[y_l,y_u^0]}\left|\widetilde{F}(y|x;\alpha) - \widehat{F}_Y(y|x;\alpha)\right| = O_p\left(\frac{\log n}{nh^p}\right).$

Proof. By the definition of these two estimators, we get:

$$\widetilde{F}(y|x;\alpha) - \widehat{F}_Y(y|x;\alpha)$$

$$= \left(\begin{array}{c} -\varphi_\alpha^{-1}\left[-\sum_{V_i\leq y, D_i=1}\varphi_\alpha\left(\widehat{S}_V(V_i^-|x)\right) - \varphi_\alpha\left(\widehat{S}_V(V_i^-|x) - w_{n\gamma}(x, X_i)\right)\right] \\ +\varphi_\alpha^{-1}\left[-\sum_{V_i\leq y, D_i=1}\varphi_\alpha'\left(\widehat{S}_V(V_i|x)\right)w_{n\gamma}(x, X_i)\right] \end{array}\right)$$

$$= R_{n3}(y, x; \alpha) + R_{n4}(y, x; \alpha),$$

where

$$R_{n3}(y, x; \alpha) = \frac{\varphi_\alpha''\{\varphi_\alpha^{-1}(\Delta)\}}{2\varphi_\alpha'\{\varphi_\alpha^{-1}(\Delta)\}^3}\left[\begin{array}{c} -\int_{y_l}^y\varphi_\alpha'\left(\widehat{S}_V(v|x)\right)d\widehat{F}_{V,D=1}(v|x) \\ +\int_{y_l}^y\varphi_\alpha'(S_V(v|x))dF_{V,D=1}(v|x) \end{array}\right]^2 \text{ and}$$

$$R_{n4}(y, x; \alpha) = \frac{-1}{\varphi_\alpha'\{\varphi_\alpha^{-1}(\Delta)\}}\left[\begin{array}{c} -\sum_{V_i\leq y, D_i=1}\varphi_\alpha\left(\widehat{S}_V(V_i^-|x)\right) - \varphi_\alpha\left(\widehat{S}_V(V_i^-|x) - w_{n\gamma}(x, X_i)\right) \\ +\sum_{V_i\leq y, D_i=1}\varphi_\alpha'\left(\widehat{S}_V(V_i|x)\right)w_{n\gamma}(x, X_i) \end{array}\right].$$

Hence it suffices to show that

$$\sup_{\alpha \in \mathcal{A}} \sup_{x \in \mathcal{J}} \sup_{y \in [y_l, y_u^0]} |R_{n3}(y, x; \alpha)| = O_p\left(\frac{\log n}{nh^p}\right) \text{ and}$$

$$\sup_{\alpha \in \mathcal{A}} \sup_{x \in \mathcal{J}} \sup_{y \in [y_l, y_u^0]} |R_{n4}(y, x; \alpha)| = O_p\left(\frac{\log n}{nh^p}\right).$$

In the expression for $R_{n3}(y, x; \alpha)$, the multiplier $\varphi_\alpha''/2\varphi_\alpha'$ remains uniformly bounded under our assumption (G). Hence it suffices to show that the following holds:

$$\left[-\int_{y_l}^y \varphi_\alpha'\left(\widehat{S}_V(v|x)\right) d\widehat{F}_{V,D=1}(v|x) + \int_{y_l}^y \varphi_\alpha'(S_V(v|x)) dF_{V,D=1}(v|x)\right] = O_p\left(\frac{\log n}{nh^p}\right)^{1/2}.$$

Omitting the smaller than $O_p\left(\frac{\log n}{nh^p}\right)$ terms, we obtain that

$$\left|-\int_{y_l}^y \varphi_\alpha'\left(\widehat{S}_V(v|x)\right) d\widehat{F}_{V,D=1}(v|x) + \int_{y_l}^y \varphi_\alpha'(S_V(v|x)) dF_{V,D=1}(v|x)\right|$$

$$\leq \left|\int_{y_l}^y \varphi_\alpha''\{S_V(v|x)\}\left[\widehat{F}_V(v|x) - F_V(v|x)\right] dF_{V,D=1}(v|x)\right|$$

$$+ \left|\varphi_\alpha'\{S_V(v|x)\}\left[\widehat{F}_{V,D=1}(v|x) - F_{V,D=1}(v|x)\right]\right|$$

$$+ \left|\int_{y_l}^y \varphi_\alpha''\{S_V(v|x)\}\left[\widehat{F}_{V,D=1}(v|x) - F_{V,D=1}(v|x)\right] dF_V(v|x)]\right|$$

$$\leq M\left(|\widehat{F}_V(v|x) - F_V(v|x)| + |\widehat{F}_{V,D=1}(v|x) - F_{V,D=1}(v|x)|\right).$$

The result follows again from the stated facts (A.3) and (A.4).

As for $R_{n4}(y, x; \alpha)$, taking second order Taylor expansion of

$$\sum_{V_i \leq y, D_i=1} \varphi_\alpha\left(\widehat{S}_V\left(V_i^-|x\right)\right) - \varphi_\alpha\left(\widehat{S}_V\left(V_i^-|x\right) - w_{n\gamma}(x, X_i)\right),$$

we get

$$R_{n4}(y, x; \alpha) = -\frac{1}{2}\sum_{V_i \leq y, D_i=1} \varphi_\alpha''(\Delta) w_{n\gamma}^2(x, X_i),$$

and by the decreasing property of $\varphi_\alpha''(\cdot)$ assumed in Assumption (G1)(i) and the fact that for large enough n, $\sup_{x \in \mathcal{J}} \sup_{y \in [y_l, y_u^0]} \widehat{S}_V(\cdot|x) \geq \delta_0$, this term is bounded from above by

$$\sup_{\alpha \in \mathcal{A}} \sup_{x \in \mathcal{J}} \sup_{y \in [y_l, y_u^0]} |R_{n4}(y, x; \alpha)| \leq \sup_{x \in \mathcal{J}} \left(\frac{1}{2}|\varphi_\alpha''(\delta_0)|\sum_{i=1}^n w_{n\gamma}^2(x, X_i)\right).$$

The conclusion follows from the boundedness of $\varphi_\alpha''(\cdot)$ at point δ_0 and the standard kernel density argument showing $\sum_{i=1}^n w_{n\gamma}^2(x, X_i) = O_p\left(\frac{\log n}{nh^p}\right)$. ∎

8 Appendix B: Asymptotic Validity of the Confidence Sets

To ease the notational burden and because we are fixing β at β_o under the null, we will denote the unique $\alpha(\beta_o)$ simply as α_o for the given β_o, i.e., $F(x'\beta_o|x; \alpha_o) = q$ a.s. Similarly we use notations $\widehat{\alpha}_o$ and $\widetilde{\alpha}_o$

instead of $\widehat{\alpha}(\beta_o)$ and $\widetilde{\alpha}(\beta_o)$. We will first present a proof of the main theorem which makes use of results in the subsequent lemmas.

Proof of Theorem 4.5. It suffices to prove those weak convergence results in Prop. 4.4. As already explained in the main text, there are both common and distinct parts in proving (3.18) and (3.19). $\widehat{T}_n(\beta_o)$ resembles the population criterion function closely we would argue below that $\widehat{\alpha}_o$ converges in probability to α_o and prove the stochastic equicontinuity of the process $nh^{p/2}[T_n(\beta_o;\alpha) - B_n(\beta_o;\alpha)]$ w.r.t α in the local neighborhood whose radius is determined by the convergence rate of $\widehat{\alpha}_o$ to α_o. When it comes to $\widetilde{T}_n(\beta_o)$, we will explore the fact that $[T_n(\beta_o;\alpha) - B_n(\beta_o;\alpha)]$ has different asymptotic behaviors for $\alpha = \alpha_o$ and for $\alpha \neq \alpha_o$ as in Santos (2006).

The proof of (3.18) will be accomplished in Steps 1-3 below and the proof of (3.19) will be accomplished in Steps 1'-3' below.

Step 1. We show the asymptotic normality of the degenerate U-statistic at α_o:

$$\frac{nh^{p/2}}{\sqrt{\Sigma(\beta_o;\alpha_o)}}[T_n(\beta_o;\alpha_o) - B_n(\beta_o;\alpha_o)] \Longrightarrow N(0,1), \tag{B.1}$$

where $\Sigma(\beta;\alpha)$ is defined in (B.8).

Step 2. We show $\widehat{\alpha}_o \rightarrow_p \alpha_o$ and characterize the convergence rate, $\widehat{\alpha}_o - \alpha_o = O_p(\delta_n)$, with $\delta_n = \frac{1}{\sqrt{n}} \vee \frac{\log n}{nh^p}$. Moreover, $\widehat{\Sigma}(\beta_o;\alpha_n) - \Sigma(\beta_o;\alpha_n) = o_p(1)$ for $\forall \alpha_n \rightarrow_p \alpha_o$.

Step 3. We show the stochastic equicontinuity of the process $[T_n(\beta_o;\alpha) - B_n(\beta_o;\alpha)]$ indexed by α in a neighborhood of α_o, i.e. $\forall \varepsilon > 0$, we could find a $\delta \leq O(\delta_n)$, s.t.

$$\lim_{n\to\infty} \sup \Pr\left(\sup_{|\alpha-\alpha_o|<\delta} nh^{p/2}|[T_n(\beta_o;\alpha) - B_n(\beta_o;\alpha)] - [T_n(\beta_o;\alpha_o) - B_n(\beta_o;\alpha_o)]| \geq \varepsilon \right) < \varepsilon. \tag{B.2}$$

Step 1$'$ is the same as Step 1.

Step 2$'$. We show that outside the neighborhood $\mathcal{A}_o^{\varepsilon_n}$ defined in (3.16), the test statistic will diverge to positive infinity:

$$\min_{\alpha \in \mathcal{A} \backslash \mathcal{A}_o^{\varepsilon_n}} \frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}(\beta_o;\alpha)}}|T_n(\beta_o;\alpha) - B_n(\beta_o;\alpha)| \rightarrow +\infty. \tag{B.3}$$

Step 3$'$. We argue the convergence of $\widetilde{\alpha}_o \rightarrow_p \alpha_o$ due to (B.3) and the uniqueness of α_o. Finally the result follows from the stochastic equicontinuity of the process $[T_n(\beta_o;\alpha) - B_n(\beta_o;\alpha)]$ indexed by α via (B.2) in the neighborhood $\mathcal{A}_o^{\varepsilon_n}$.

Notice that Step 1 leads to (3.17) while Steps 1-3 give us (3.18). Steps 1$'$-3$'$ lead to (3.19), see Santos (2006). Once we verify the claims in those three steps, combining them together leads to the conclusion that both confidence sets would be of exact size asymptotically. ∎

Proof of Step 1. We show that the asymptotic behavior of $T_n(\beta_o;\alpha_o)$ is given by a degenerate U-statistic.

Note that

$$T_n(\beta_o;\alpha_o) = \int_{\mathcal{J}} \left[\widehat{F}(x'\beta_o|x;\alpha_o) - F_Y(x'\beta_o|x;\alpha_o)\right]^2 f_X^2(x)\,dx + o_p\left(n^{-1}h^{-p/2}\right)$$

$$= \int_{\mathcal{J}} \left[\frac{1}{n}\sum_{i=1}^n W_\gamma(x,X_i)\left(c\left(x'\beta_o|Z_i,x;\alpha_o\right) + b\left(x'\beta_o|Z_i,x;\alpha_o\right)\right)\right]^2 dx$$

$$+ \int_{\mathcal{J}} R_n\left(x'\beta_o,x;\alpha\right)\left[\frac{1}{n}\sum_{i=1}^n W_\gamma(x,X_i)\left(c\left(x'\beta_o|Z_i,x;\alpha_o\right) + b\left(x'\beta_o|Z_i,x;\alpha_o\right)\right)\right]dx$$

$$+ \int_{\mathcal{J}} R_n^2\left(x'\beta_o,x;\alpha\right)dx + o_p\left(n^{-1}h^{-p/2}\right)$$

$$= \int_{\mathcal{J}} \left[\frac{1}{n}\sum_{i=1}^n W_\gamma(x,X_i)\left(c\left(x'\beta_o|Z_i,x;\alpha_o\right) + b\left(x'\beta_o|Z_i,x;\alpha_o\right)\right)\right]^2 dx + o_p\left(n^{-1}h^{-p/2}\right).$$

The first equality where we replaced $\widehat{f_X^2}(x)$ with $f_X^2(x)$ follows the same argument as in Prop.1 in Hardle and Mammen (1993). In the second equality we apply Cauchy-Shwartz inequality to the second term and use the result that $\sup|R_n(x'\beta_o,x;\alpha)| = O_p\left(\frac{\log n}{nh^p}\right)$. We need Assumption (H) which states: $\frac{(\log n)^2}{nh^{3p/2}} \to 0$, to show the negligibility of the third term. Also for simplicity, we omit the range of the integral, so

$$T_n(\beta_o;\alpha_o) = I_{n1} + 2I_{n2} + 2I_{n3} + I_{n4} + s.o. \tag{B.4}$$

where

$$I_{n1} = \frac{1}{n^2}\sum_{i=1}^n \int W_\gamma^2(x,X_i)\,c^2\left(x'\beta_o|Z_i,x;\alpha_o\right)dx,$$

$$I_{n2} = \frac{1}{n^2}\sum_{1\le i<j\le n} \int W_\gamma(x,X_i)\,W_\gamma(x,X_j)\,c\left(x'\beta_o|Z_i,x;\alpha_o\right)c\left(x'\beta_o|Z_j,x;\alpha_o\right)dx,$$

$$I_{n3} = \frac{1}{n}\sum_{i=1}^n \int W_\gamma(x,X_i)\,c\left(x'\beta_o|Z_i,x;\alpha_o\right)\left[\frac{1}{n}\sum_{j=1}^n W_\gamma(x,X_j)\,b\left(x'\beta_o|Z_j,x;\alpha_o\right)\right]dx,$$

$$I_{n4} = \int \left[\frac{1}{n}\sum_{i=1}^n W_\gamma(x,X_i)\,b\left(x'\beta_o|Z_i,x;\alpha_o\right)\right]^2 dx.$$

The above decomposition and the next sequence of lemmas parallel the results presented by Hall (1984ab), hence we follow his notation as closely as possible. In addition, we write conditional expectation given covariates as E', i.e. $E'[Z] = E[Z|X_1,\cdots,X_n]$, as in Hall (1984b). Recall from (C.2) in Appendix C, $\sigma^2(y|X_i,x;\alpha_o) = E'\left[c^2(y|Z_i,x;\alpha_o)\right]$ and $\sigma^2(y|x;\alpha_o) = \lim_{x_n\to x}\sigma^2(y|x_n,x;\alpha_o)$.

The calculation in the proof of Lemma C.3 in Appendix C gives $I_{n4} = O\left(h^{2s} + \sum_{j=1}^r \lambda_j^2\right)$ immediately, hence $nh^{p/2}I_{n4} = O\left(\frac{nh^{p+2s}+\sum_{j=1}^r \lambda_j^2 nh^p}{h^{p/2}}\right) = o\left(\frac{1}{h^{p/2}}\right)$ by Assumption (H), which is of smaller order than $nh^{p/2}I_{n1}$ demonstrated in Lemma B.1 below. The proof of Step 1 will be completed via Lemmas B.1-B.5. $\blacksquare$

Lemma B.1 *Under Assumptions (D)-(K), we have that*

$$nh^{p/2}I_{n1} = \frac{1}{h^{p/2}}\int K^2(u)\,du \int \sigma^2\left(x'\beta_o|x;\alpha_o\right)f_X(x)\,dx + s.o.$$

Proof. As in Hall, Li and Racine (2004), after taking conditional expectation one gets

$$E'[I_{n1}] = \frac{1}{n^2 h^{2p}} \sum_{i=1}^{n} \int K^2 \left(\frac{x^c - X_i^c}{h} \right) \sigma^2 \left(x'\beta_o | X_i, x; \alpha_o \right) dx + s.o.,$$

and by boundedness of $c \left(x'\beta_o | Z_i, x; \alpha_o \right)$, we obtain:

$$\left(nh^p \right)^4 E' \left[I_{n1} - E'[I_{n1}] \right]^2 \leq M \sum_{i=1}^{n} \left(\int K^2 \left(\frac{x^c - X_i^c}{h} \right) dx^c \right)^2 = O \left(nh^{2p} \right). \tag{B.5}$$

Hence by Markov's inequality, we get

$$\lim_{M \to \infty} \lim_{n \to \infty} \sup \Pr \left\{ \left| I_{n1} - E'[I_{n1}] \right| > M n^{-3/2} h^{-p} | X_1, \cdots, X_n \right\} = 0,$$

and thus the unconditional probability goes to zero too, i.e., $\left| I_{n1} - E'[I_{n1}] \right| = O_p \left(n^{-3/2} h^{-p} \right)$. Since

$$nh^{p/2} E \left[E'[I_{n1}] \right] = \frac{1}{h^{3p/2}} \int \sigma^2 \left(x'\beta_o | X_i, x; \alpha_o \right) K^2 \left(\frac{x^c - X_i^c}{h} \right) f_X (X_i) \, dX_i dx + s.o.$$

$$= \frac{1}{h^{p/2}} \int K^2 (u) \, du \int \sigma^2 \left(x'\beta_o | x; \alpha_o \right) f_X (x) \, dx + s.o.,$$

another application of Markov's inequality gives $nh^{p/2} I_{n1} = nh^{p/2} E'[I_{n1}] + s.o..$ ∎

Now we define

$$B_n (\beta; \alpha) = \frac{1}{nh^p} \int K^2 (u) \, du \int \sigma^2 \left(x'\beta | x; \alpha \right) f_X (x) \, dx. \tag{B.6}$$

Before dealing with I_{n2}, we introduce further notations needed in the martingale representation:

$$n^2 h^{2p} I_{n2} = \sum_{1 \leq j < i \leq n} W_{nij} = \sum_{i=2}^{n} Y_{ni}, \tag{B.7}$$

where $W_{nij} = \int h^{2p} W_\gamma (x, X_i) W_\gamma (x, X_j) c \left(x'\beta_o | Z_i, x; \alpha_o \right) c \left(x'\beta_o | Z_j, x; \alpha_o \right) dx$, and $Y_{ni} = \sum_{j=1}^{i-1} W_{nij}$. Now the partial sum $S_{ni} = \sum_{j=1}^{i} Y_{nj}$ is a martingale triangular array with respect to the filter $\mathcal{F}_{n,i}$ generated by $X_1, \cdots, X_n$ and $Z_1, \cdots, Z_i$. Let

$$V_n^2 = \sum_{i=2}^{n} E \left[Y_{ni}^2 | \mathcal{F}_{n,i-1} \right] = \sum_{i=2}^{n} \sum_{j=1}^{i-1} E \left[W_{nij}^2 | \mathcal{F}_{n,i-1} \right] + 2 \sum_{i=2}^{n} \sum_{1 \leq j < k \leq i-1} E \left[W_{nij} W_{nik} | \mathcal{F}_{n,i-1} \right]$$

$$= V_{n1} + V_{n2}.$$

Further let $\Sigma_o \equiv \Sigma (\beta_o; \alpha_o)$, where

$$\Sigma (\beta; \alpha) = 2 \int \sigma^4 \left(x'\beta | x; \alpha \right) f_X^2 (x) \, dx \int \left[\int K (u) K (v + u) \, du \right]^2 dv. \tag{B.8}$$

Lemma B.2 *Under Assumptions (D)-(K), we have that* $a_n \equiv E[V_{n1}] = \frac{n(n-1)}{4} h^{3d} \Sigma_o + s.o.$

Proof. To see why the above expression holds, we first focus on the conditional expectation given covariates:

$$E\left[\int h^{2p} W_\gamma\left(x, X_i\right) W_\gamma\left(x, X_j\right) c\left(x'\beta_o|Z_i, x; \alpha_o\right) c\left(x'\beta_o|Z_j, x; \alpha_o\right) dx\right]^2$$

$$= \iiiint \left[\begin{array}{c} \iint K\left(\frac{x^c - X_i^c}{h}\right) K\left(\frac{x^c - X_j^c}{h}\right) c\left(x'\beta_o|Z_i, x; \alpha_o\right) c\left(x'\beta_o|Z_j, x; \alpha_o\right) \times \\ K\left(\frac{y - X_i^c}{h}\right) K\left(\frac{y - X_j}{h}\right) c\left(y'\beta_o|Z_i, x; \alpha_o\right) c\left(y'\beta_o|Z_j, x; \alpha_o\right) dxdy \end{array}\right] \times$$

$$dQ\left(Z_i|X_i\right) dQ\left(Z_j|X_j\right) dF_X\left(X_i\right) dF_X\left(X_j\right) + s.o.$$

$$= \iint \left[\begin{array}{c} \int c\left(x'\beta_o|Z_i, x; \alpha_o\right) c\left(y'\beta_o|Z_i, x; \alpha_o\right) dQ\left(Z_i|X_i\right) \\ K\left(\frac{x - X_i^c}{h}\right) K\left(\frac{y - X_i^c}{h}\right) dF_X\left(X_i\right) \end{array}\right]^2 dxdy + s.o.$$

Also define

$$\eta\left(x'\beta_o, y'\beta_o|X_i, x; \alpha_o\right) = \int c\left(x'\beta_o|Z_i, x; \alpha_o\right) c\left(y'\beta_o|Z_i, x; \alpha_o\right) dQ\left(Z_i|X_i\right)$$

$$= E'\left[c\left(x'\beta_o|Z_i, x; \alpha_o\right) c\left(y'\beta_o|Z_i, x; \alpha_o\right)\right].$$

Finally, by omitting smaller order terms, we obtain:

$$\Sigma_o/2 = \frac{1}{h^{3d}} \int \left[\sum_{x^d, y^d} \eta\left(x'\beta_o, y'\beta_o|X_i, x; \alpha_o\right) K\left(\frac{x^c - X_i^c}{h}\right) K\left(\frac{y^c - X_i}{h}\right) f_X\left(X_i\right) dX_i\right]^2 dxdy + s.o.\text{(B.9)}$$

$$= \frac{1}{h^d} \int \left[\sum_{x^d, y^d} \eta\left(x'\beta_o, y'\beta_o|\left(x^c - uh, x^d\right), x; \alpha_o\right) K\left(u\right) K\left(\frac{y^c - x^c}{h} + u\right) f_X\left(x^c - uh, x^d\right) du\right]^2 dxdy + s.o.$$

$$= \int \sigma^4\left(x'\beta_o|x; \alpha_o\right) f_X^2\left(x\right) dx \int \left[\int K\left(u\right) K\left(v + u\right) du\right]^2 dv + s.o.$$

■

Now we are ready to present the result for I_{n2}.

Lemma B.3 *Under Assumptions (D)-(K), we have that $2nh^{p/2}I_{n2} \Longrightarrow N\left(0, \Sigma_o\right)$.*

Proof. The proof follows modification of Lemmas 1 & 2 in Hall (1984b), and we use the fact that the centering term $c\left(x'\beta_o|Z_i, x; \alpha_o\right)$ is uniformly bounded repeatedly without further mentioning.

First, we would show: $V_{n1} = a_n + o_p\left(1\right) = \frac{n(n-1)}{4} h^{3p} \Sigma_o + o_p\left(1\right)$ by controlling $E\left[V_{n1} - a_n\right]^2$: $E\left[V_{n1} - a_n\right]^2 \leq Mn^3 E\left[W_{n12}^4\right]$, and the fourth moment shall be bounded in a similar way as in Hall (1984a):

$$E\left[W_{n12}^4\right] \leq \iiiint \left\{E\left[\sum_{x_{(i)}^d} \prod_{i=1}^4 K\left(\frac{x_{(i)}^c - X_1^c}{h}\right) c\left(x^{(i)'}\beta_o|Z_1, x^{(i)}; \alpha_o\right)\right]\right\}^2 dx_{(1)}^c \cdots dx_{(4)}^c$$

$$\leq Mh^{2p} \iiiint \left\{\sum_{x_{(i)}^d} \int K\left(v\right) \left[\prod_{i=2}^4 K\left(v + \frac{x_{(i)}^c - x_{(1)}^c}{h}\right)\right] f\left(x_{(1)}^c - vh\right) dv\right\}^2 dx_{(1)}^c \cdots dx_{(4)}^c$$

$$= Mh^{5p} \iiiint \left\{\sum_{x_{(i)}^d} \int K\left(v\right) \left[\prod_{i=2}^4 K\left(v + w_{(i)}\right)\right] f\left(x^{(1)} - vh\right) dv\right\}^2 dx_{(1)}^c dw_{(2)} dw_{(3)} dw_{(4)}$$

$$= O\left(h^{5p}\right).$$

35

Hence $E\left[V_{n1} - a_n\right]^2 = O\left(n^3 h^{5p}\right)$, and $V_{n1} = a_n + O_p\left(n^{3/2}h^{5p/2}\right)$.

Analogously, one bounds the following higher order moment terms by routine yet tedious calculation as done on Page 251 in Hall (1984b),

$$E\left[W_{n13}W_{n14}W_{n23}W_{n24}\right] = O\left(h^{7p}\right) \text{ and } E\left[W_{n12}^2 W_{n13}^2\right] = O\left(h^{5p}\right).$$

From there we could show V_{n2} being negligible, since

$$
\begin{aligned}
E\left[V_{n2}^2\right] &= 4E \sum_{1 \leq j < k \leq n-1} \left(\sum_{i=k+1}^{n} W_{nij}W_{nik}\right)^2 \\
&\leq M \sum_{1 \leq j < k \leq n-1} \left\{nE\left[W_{n12}^2 W_{n13}^2\right] + n^2 E\left[W_{n13}W_{n14}W_{n23}W_{n24}\right]\right\} \\
&= O\left(n^3 h^{5p} + n^4 h^{7p}\right).
\end{aligned}
$$

Thus we get $E'\left[n^{-2}h^{-3p}V_{n2}\right]^2 = O_p\left(\frac{1}{nh^p} + h^p\right) = o_p\left(1\right)$.

Finally, we check the Lindeberg condition in order to apply Corollary 3.2 in Hall and Heyde (1980): $\forall \varepsilon > 0$, as $n \to \infty$,

$$n^{-2}h^{-3p}\sum_{i=1}^{n} E\left[Y_{ni}^2 I\left(|Y_{ni}| > \varepsilon n h^{3p/2}\right)\right] \to 0. \tag{B.10}$$

The proof shall be accomplished by Markov's inequality and computing the fourth moments: $\sum_{i=2}^{n} E\left[Y_{ni}^4\right] = \left(n^3 h^{5p}\right)$, since

$$n^{-2}h^{-3p}\sum_{i=1}^{n} E\left[Y_{ni}^2 I\left(|Y_{ni}| > \varepsilon n h^{3p/2}\right)\right] \leq \varepsilon^{-2}n^{-4}h^{-6p}\sum_{i=2}^{n} E\left[Y_{ni}^4\right] = O\left(\frac{1}{nh^p}\right).$$

Overall, $n^{-2}h^{-3p}V_n^2 \to \Sigma_o/4$ in probability, $n^{-2}h^{-3p}V_{n2} = o_p\left(1\right)$ and one would deduce

$$n^{-1}h^{-3p/2}\sum_{i=2}^{n} Y_{ni} \Longrightarrow N\left(0, \frac{\Sigma_o}{4}\right)$$

which is the desired result once rewritten in terms of $2I_{n2}$. $\blacksquare$

Next we characterize the stochastic order of I_{n3}, which is smaller than I_{n2} when we select an under-smoothing bandwidth.

Denote $b_n\left(x\right) = \frac{1}{n}\sum_{j=1}^{n} W_\gamma\left(x, X_i\right) b\left(x'\beta_o | Z_j, x; \alpha_o\right)$, and $\gamma_n\left(x\right) = E\left[b_n\left(x\right)\right]$. Break I_{n3} into $I_{n3} = J_{n1} + J_{n2}$, where

$$
\begin{aligned}
J_{n1} &= \frac{1}{n}\sum_{i=1}^{n} \int W_\gamma\left(x, X_i\right) c\left(x'\beta_o | Z_i, x; \alpha_o\right) \gamma_n\left(x\right) dx \text{ and} \\
J_{n2} &= \frac{1}{n}\sum_{i=1}^{n} \int W_\gamma\left(x, X_i\right) c\left(x'\beta_o | Z_i, x; \alpha_o\right) \left[b_n\left(x\right) - \gamma_n\left(x\right)\right] dx
\end{aligned}
$$

Lemma B.4 *Under Assumptions (D)-(K), $J_{n2} = o_p\left(n^{-1}h^{-p/2}\right)$.*

Proof. The following string of computation is routine:

$$
E'\left[J_{n2}^2\right] = \frac{1}{n^2}E'\left[\left(\sum_{i=1}^n \int W_\gamma\left(x, X_i\right)c\left(x'\beta_o|Z_i, x; \alpha_o\right)\left[b_n\left(x\right) - \gamma_n\left(x\right)\right]dx\right)^2\right]
$$

$$
= \frac{1}{n^2}\sum_{i=1}^n E'\left[\left(\int W_\gamma\left(x, X_i\right)c\left(x'\beta_o|Z_i, x; \alpha_o\right)\left[b_n\left(x\right) - \gamma_n\left(x\right)\right]dx\right)^2\right]
$$

$$
\leq \frac{1}{n^2}\sum_{i=1}^n E'\left[\int W_\gamma\left(x, X_i\right)c^2\left(x'\beta_o|Z_i, x; \alpha_o\right)dx\right] \times \int W_\gamma\left(x, X_i\right)\left[b_n\left(x\right) - \gamma_n\left(x\right)\right]^2 dx
$$

$$
\leq \frac{1}{n^2}\sup_i E'\left[\int W_\gamma\left(x, X_i\right)c^2\left(x'\beta_o|Z_i, x; \alpha_o\right)dx\right]\sum_{i=1}^n \int W_\gamma\left(x, X_i\right)\left[b_n\left(x\right) - \gamma_n\left(x\right)\right]^2 dx
$$

where the first inequality is merely Cauchy-Schwartz inequality and the claim would follow if one can show $E'\left[J_{n2}^2\right] = o_p\left(n^{-2}h^{-p}\right)$ via

$$
E\left[\sum_{i=1}^n \int h^p W_\gamma\left(x, X_i\right)\left[b_n\left(x\right) - \gamma_n\left(x\right)\right]^2 dx\right] = o_p\left(1\right).
$$

Indeed by expanding $\left[b_n\left(x\right) - \gamma_n\left(x\right)\right]^2$ into diagonal and cross product terms and recalling the expression of the bias term in the linear representation, we get

$$
E\left[\int h^p W_\gamma\left(x, X_i\right)\left[b_n\left(x\right) - \gamma_n\left(x\right)\right]^2 dx\right]
$$

$$
\leq M\left(nh^{2p}\right)^{-1}E\left[\int h^p W_\gamma\left(x, X_i\right)\right]dx \times
$$

$$
E\left[\int h^p W_\gamma\left(x, X_j\right)\left\{\left[F_{V, D=1}\left(\cdot|X_j\right) - F_{V, D=1}\left(\cdot|x\right)\right]^2 + \left[F_V\left(\cdot|X_j\right) - F_V\left(\cdot|x\right)\right]^2\right\}dx\right]
$$

$$
+ M\left(nh^p\right)^{-2}E\left[\int h^{2p}W_\gamma^2\left(x, X_i\right)\left\{\left[F_{V, D=1}\left(\cdot|X_i\right) - F_{V, D=1}\left(\cdot|x\right)\right]^2 + \left[F_V\left(\cdot|X_i\right) - F_V\left(\cdot|x\right)\right]^2\right\}dx\right]
$$

$$
\leq M\left(nh^p\right)^{-1}E\left[\int K\left(\frac{x^c - X_j^c}{h}\right)\left\{\left[F_{V, D=1}\left(\cdot|X_j\right) - F_{V, D=1}\left(\cdot|x\right)\right]^2 + \left[F_V\left(\cdot|X_j\right) - F_V\left(\cdot|x\right)\right]^2\right\}dx\right]
$$

$$
= o\left(\frac{1}{n}\right),
$$

where the last one follows from the standard kernel convergence result, thus the claim is proved. ∎

Lemma B.5 *Under assumptions (D)-(K), $\frac{\sqrt{n}}{h^s}J_{n1} = O_p\left(1\right)$.*

Proof. Conditional on the covariates, J_{n1} is a sum of centered independent random variables: $J_{n1} = \frac{1}{nh^p}\sum_{i=1}^n Z_{ni}$ with $Z_{ni} = \int h^p W_\gamma\left(x, X_i\right)c\left(x'\beta_o|Z_i, x; \alpha_o\right)\gamma_n\left(x\right)dx$. Actually one could show the CLT holds for $\frac{\sqrt{n}}{h^s}J_{n1}$ as in Hall (1984b) by verifying the Lindeberg condition: $\forall \varepsilon > 0$,

$$
\frac{1}{nh^{2(p+s)}}\sum_{i=1}^n E\left[Z_{ni}^2 I\left(|Z_{ni}| > \varepsilon\sqrt{n}h^{p+s}\right)\right] \to 0,
$$

and convergence in probability of $\frac{1}{nh^{2(p+s)}}E\left(\sum_{i=1}^n Z_{ni}\right)^2$. What we do here is much simpler by bounding the second moments of Z_{ni} since we are using undersmoothed bandwidth, only the weak result stated in the

lemma is of interest to us:

$$EZ_{ni}^2 \leq Mh^{2p} \int K(z) K(z+u) \, dz \iint \gamma_n(x) \gamma_n(x+uh) \, dxdu \leq Mh^{2(p+s)}.$$

The first inequality follows a change of variables and bounding the centering term, while the second one follows by expanding the bias term up to second order. Now

$$E\left[\frac{\sqrt{n}}{h^s} J_{n1}\right]^2 = E\left[\frac{1}{\sqrt{n}h^{p+s}} \sum_{i=1}^n Z_{ni}\right]^2 \leq \frac{M}{nh^{2(p+s)}} \times nh^{2(p+s)}.$$

Therefore under the Assumption (H) that we are using the bandwidth to kill the bias, I_{n3} is of smaller order.

Proof of Step 2. Recall the result in Prop. 2.6, which states that given any β_o under the null we have a unique α_o minimizing (3.1). By Assumption (G), $F_Y(y|x; \alpha)$ is continuous w.r.t α, so is the population criterion function (3.1). Now the compactness of $\mathcal{A}$ guarantees the well separation of this minimum point. Furthermore uniform almost surely convergence of $\widehat{F}_Y(y|x; \alpha)$ to $F_Y(y|x; \alpha)$ as shown in Cor. 3.8 would give the desired convergence of $\widehat{\alpha}_o$ to α_o by Theorem 5.7 in Van der Vaart (1998). Getting the rate of convergence is a bit more complicated and we separate it out in the next lemma. ∎

Referring to the claim that $\widehat{\Sigma}(\beta_o; \alpha_n) - \Sigma(\beta_o; \alpha_n) = o_p(1)$ for $\forall \alpha_n \to_p \alpha_o$, a close inspection of $\Sigma(\beta_o; \alpha)$'s expression from Appendix C shows that only φ_α', φ_α'' coupled with $F_V(v|x)$ and $F_{V,D=1}(v|x)$ are involved. By Assumption (G2), φ_α' and φ_α'' are uniformly continuous w.r.t α and the plug-in type estimator $\widehat{\Sigma}(\beta_o; \alpha)$ merely replaces $F_V(v|x)$ and $F_{V,D=1}(v|x)$ with their kernel estimators $\widehat{F}_V(v|x)$ and $\widehat{F}_{V,D=1}(v|x)$. Hence the result follows from the standard convergence result in kernel estimation. ∎

Lemma B.6 *Under Assumptions (D)-(K), (SC), (V0), and (V1), we have that $\widehat{\alpha}_o - \alpha_o = O_p(\delta_n)$, with $\delta_n = \frac{1}{\sqrt{n}} \vee \frac{\log n}{nh^p}$.*

Proof. To get the convergence rate of $\widehat{\alpha}_o$ to α_o, notice that the minimizer satisfies the following first order condition given the smoothness of generator function in Assumption (G):

$$\int \left[\widehat{F}(x'\beta_o|x; \widehat{\alpha}_o) - q\right] \widehat{f}_X^2(x) \left(\frac{\partial}{\partial \alpha} \widehat{F}(x'\beta_o|x; \widehat{\alpha}_o)\right) dx = 0.$$

Taking a first order expansion around α_o, we get

$$(\widehat{\alpha}_o - \alpha_o) \int \left\{\left(\frac{\partial}{\partial \alpha} \widehat{F}(x'\beta_o|x; \alpha_o)\right)^2 + \frac{\partial^2}{\partial \alpha^2} \widehat{F}(x'\beta_o|x; \alpha_o) \left[\widehat{F}(x'\beta_o|x; \alpha_o) - q\right]\right\} \widehat{f}_X^2(x) \, dx$$

$$= \int \left[\widehat{F}(x'\beta_o|x; \alpha_o) - q\right] \widehat{f}_X^2(x) \left(\frac{\partial}{\partial \alpha} \widehat{F}(x'\beta_o|x; \alpha_o)\right) dx + s.o.$$

$$= \int \left[\widehat{F}(x'\beta_o|x; \alpha_o) - q\right] \widehat{f}_X^2(x) \left(\frac{\partial}{\partial \alpha} F_Y(x'\beta_o|x; \alpha_o)\right) dx$$

$$+ \int \left[\widehat{F}(x'\beta_o|x; \alpha_o) - q\right] \widehat{f}_X^2(x) \left(\frac{\partial}{\partial \alpha} \widehat{F}(x'\beta_o|x; \alpha_o) - \frac{\partial}{\partial \alpha} F_Y(x'\beta_o|x; \alpha_o)\right) dx + s.o.$$

The integral on the LHS converges to the following term in probability:

$$\int \left\{ \left(\frac{\partial}{\partial \alpha} F\left(x'\beta_o|x;\alpha_o\right) \right)^2 + \frac{\partial^2}{\partial \alpha^2} F\left(x'\beta_o|x;\alpha_o\right) \left[F\left(x'\beta_o|x;\alpha_o\right) - q \right] \right\} f_X^2\left(x\right) dx$$

$$= \int \left(\frac{\partial}{\partial \alpha} F\left(x'\beta_o|x;\alpha_o\right) \right)^2 f_X^2\left(x\right) dx > 0,$$

where we have used the fact that $\left[F\left(x'\beta_o|x;\alpha_o\right) - q \right] = 0$ for any x.

Now we denote $\frac{\partial}{\partial \alpha} F_Y\left(x'\beta_o|x;\alpha_o\right) = G\left(x'\beta_o, x;\alpha_o\right)$, the first term on RHS could be written as

$$\frac{1}{h^{2p}} U_n^{(2)}\left(f\right) + \frac{1}{n} \sum_{i=1}^n \int W_\gamma\left(x, X_i\right) b\left(x'\beta_o|Z_i, x;\alpha_o\right) \widehat{f}_X\left(x\right) G\left(x'\beta_o, x;\alpha_o\right) dx + s.o.$$

$$= \frac{1}{h^{2p}} U_n^{(2)}\left(f\right) + O\left(h^s\right) + s.o.$$

where $f \in \mathcal{F}_3$ in Appendix C. After symmetrizations, the Hoeffding-Hajek decomposition shows that the leading term is $U_n^{(1)}\left(\pi_1 f\right)/h^p$, which is

$$\frac{1}{n} \sum_{i=1}^n \int W_\gamma\left(x, X_i\right) c\left(x'\beta_o|Z_i, x;\alpha_o\right) f_X\left(x\right) G\left(x'\beta_o, x;\alpha_o\right) dx = O_p\left(n^{-1/2}\right),$$

by standard Lindeberg central limit theorem, similar to the average derivative type result under Assumption (H). The second term on the RHS is

$$\int \left[\widehat{F}\left(x'\beta_o|x;\alpha_o\right) - q \right] \widehat{f}_X^2\left(x\right) \times \dot{\varphi}_\alpha^{-1} \left(\int \dot{\varphi}_\alpha'\left(S_V\left(v|x\right)\right) dF_{V,D=1}\left(v|x\right) \right) \times$$

$$\left[\int \dot{\varphi}_\alpha'\left(\widehat{S}_V\left(v|x\right)\right) d\widehat{F}_{V,D=1}\left(v|x\right) - \int \dot{\varphi}_\alpha'\left(S_V\left(v|x\right)\right) dF_{V,D=1}\left(v|x\right) \right] dx$$

$$\leq M \left(\int \left[\widehat{F}\left(x'\beta_o|x;\alpha_o\right) - q \right]^2 dx \right)^{1/2} \times$$

$$\left(\int \left[\int \dot{\varphi}_\alpha'\left(\widehat{S}_V\left(v|x\right)\right) d\widehat{F}_{V,D=1}\left(v|x\right) - \int \dot{\varphi}_\alpha'\left(S_V\left(v|x\right)\right) dF_{V,D=1}\left(v|x\right) \right]^2 dx \right)^{1/2}$$

$$= O_p\left(\frac{\log n}{nh^p} \right),$$

where the second term in the bracket could be handled just as the plug-in estimator. ∎

Proof of Step 3. The following decomposition holds for an arbitrary α:

$$\frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}\left(\beta_o;\alpha\right)}}\left[T_n\left(\beta_o;\alpha\right)-B_n\left(\beta_o;\alpha\right)\right] \tag{B.11}$$

$$= \frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}\left(\beta_o;\alpha\right)}}\left[\int\left[\widehat{F}\left(x'\beta_o|x;\alpha\right)-F_Y\left(x'\beta_o|x;\alpha\right)+F_Y\left(x'\beta_o|x;\alpha\right)-q\right]^2\widehat{f}_X^2\left(x\right)dx-B_n\left(\beta_o;\alpha\right)\right]$$

$$= \frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}\left(\beta_o;\alpha\right)}}\left[\frac{1}{n^2}\sum_{1\leq i<j\leq n}\int W_\gamma\left(x,X_i\right)W_\gamma\left(x,X_j\right)c\left(x'\beta_o|Z_i,x;\alpha_o\right)c\left(x'\beta_o|Z_j,x;\alpha_o\right)dx\right]$$

$$+\frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}\left(\beta_o;\alpha\right)}}\left[\frac{1}{nf_X\left(x\right)}\sum_{i=1}^n\int\left[F_Y\left(x'\beta_o|x;\alpha\right)-q\right]W_\gamma\left(x,X_i\right)c\left(x'\beta_o|Z_i,x;\alpha_o\right)\widehat{f}_X\left(x\right)dx\right]$$

$$+\frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}\left(\beta_o;\alpha\right)}}\int\left[F_Y\left(x'\beta_o|x;\alpha\right)-q\right]^2\widehat{f}_X^2\left(x\right)dx+s.o.$$

$$= \frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}\left(\beta_o;\alpha\right)}}\left[P_{n1}\left(\alpha\right)+P_{n2}\left(\alpha\right)+P_{n3}\left(\alpha\right)\right]+s.o.$$

The proof will be completed once we show that $\forall\varepsilon>0$, one could find a δ, s.t.

$$\limsup_{n\to\infty}\Pr\left(\sup_{|\alpha_1-\alpha_2|<\delta}nh^{p/2}\left|P_{n1}\left(\alpha_1\right)-P_{n1}\left(\alpha_2\right)\right|\geq\varepsilon\right)<\varepsilon,$$

and for $\forall\alpha$ s.t. $|\alpha-\alpha_o|\leq O\left(\delta_n\right)$, $nh^{p/2}\left[P_{n2}\left(\alpha\right)+P_{n3}\left(\alpha\right)\right]=o_p\left(1\right)$. It is rather easy to bound the deterministic term: $nh^{p/2}P_{n3}\left(\alpha\right)\leq nh^{p/2}M\sup_x\left[F_Y\left(x'\beta_o|x;\alpha\right)-q\right]^2=Mnh^{p/2}\times\left(\frac{1}{n}\vee\left(\frac{\log n}{nh^p}\right)^2\right)=o\left(1\right)$ under Assumption (H). The other two claims will be proved in the subsequent lemmas. ∎

Lemma B.7 *Under Assumptions (D)-(K), $\forall\varepsilon>0$, we could find a δ, s.t.*

$$\limsup_{n\to\infty}\Pr\left(\sup_{|\alpha_1-\alpha_2|<\delta}nh^{p/2}\left|P_{n1}\left(\alpha_1\right)-P_{n1}\left(\alpha_2\right)\right|\geq\varepsilon\right)<\varepsilon.$$

Proof. Recalling the expression for $P_{n1}\left(\alpha\right)$, the difference could be written as,

$$h^{2p}\left[P_{n1}\left(\alpha_1\right)-P_{n1}\left(\alpha_2\right)\right]=U_n^{(2)}\left(\overline{f}\right)$$

$$= \frac{1}{n^2}\sum_{1\leq i<j\leq n}\int h^{2p}W_\gamma\left(x,X_i\right)W_\gamma\left(x,X_j\right)\left[\begin{array}{c}c\left(x'\beta_o|Z_i,x;\alpha_1\right)c\left(x'\beta_o|Z_j,x;\alpha_1\right)-\\c\left(x'\beta_o|Z_i,x;\alpha_2\right)c\left(x'\beta_o|Z_j,x;\alpha_2\right)\end{array}\right]dx.$$

As the operation (taking difference of $f\in\mathcal{F}_4$ over α defined in Appendix C) still preserves the VC property, we have the following VC-type class of functions:

$$\overline{\mathcal{F}}_4=\left\{\overline{f}:h\geq0,\lambda_j\in\left[0,\frac{c_j-1}{c_j}\right],\left(\alpha_1,\alpha_2\right)\in\mathcal{A},\left|\alpha_1-\alpha_2\right|<\delta\right\}$$

with bounded envelope M and $\left\|P^2\overline{f}^2\right\|_{\overline{\mathcal{F}}_4}\leq\sigma^2$, where $\sigma^2=O\left(h^{3p}\delta^2\right)$ and using the fact that $1/\varphi'_\alpha$ and φ''_α are Lipschitz continuous with respect to α. Now we apply Major's (2006) tail bound with $k=2$ for a large

enough n s.t. the constraints are satisfied:

$$\Pr\left\{\sup_{|\alpha_1-\alpha_2|<\delta} nh^{p/2}\left|P_{n1}\left(\alpha_1\right)-P_{n1}\left(\alpha_2\right)\right|\geq\varepsilon\right\}=\Pr\left\{\sup_{\overline{f}\in\overline{\mathcal{F}}_4} nh^{-3p/2}\left|U_n^{(2)}\left(\overline{f}\right)\right|\geq\varepsilon\right\}$$

$$=\quad\Pr\left\{\sup_{\overline{f}\in\overline{\mathcal{F}}_4} n\left|U_n^{(2)}\left(\overline{f}\right)\right|\geq\varepsilon h^{3p/2}\right\}\leq M\exp\left[-M\left(\frac{\varepsilon h^{3p/2}}{\sigma}\right)\right]=M\exp\left[-M\left(\frac{\varepsilon}{\delta}\right)\right].$$

Hence the desired result follows as $\delta\to 0$. ∎

Lemma B.8 *Under Assumptions (D)-(K), we have*

$$\sup_{|\alpha-\alpha_o|\leq O(\delta_n)}\frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}\left(\beta_o;\alpha\right)}}P_{n2}\left(\alpha\right)=o_p\left(1\right).$$

Proof. Omiting the smaller order diagonal term in $P_{n2}\left(\alpha\right)$, we consider the U-process indexed by α :

$$U_n^{(2)}\left(f\right)=\frac{1}{n\left(n-1\right)}\sum_{i\neq j}\int\left[F_Y\left(x'\beta_o|x;\alpha\right)-q\right]W_\gamma\left(x,X_i\right)W_\gamma\left(x,X_j\right)c\left(x'\beta_o|Z_i,x;\alpha_o\right)dx$$

$$=\quad\frac{2}{n\left(n-1\right)}\sum_{i\neq j}\frac{1}{2}\left[\begin{array}{c}\int\left[F_Y\left(x'\beta_o|x;\alpha\right)-q\right]W_\gamma\left(x,X_i\right)W_\gamma\left(x,X_j\right)c\left(x'\beta_o|Z_i,x;\alpha_o\right)dx\\+\int\left[F_Y\left(x'\beta_o|x;\alpha\right)-q\right]W_\gamma\left(x,X_i\right)W_\gamma\left(x,X_j\right)c\left(x'\beta_o|Z_j,x;\alpha_o\right)dx\end{array}\right].$$

The function f belongs to $\mathcal{F}_3$ with $|\alpha-\alpha_o|\leq\delta_n$ in Appendix C and $G\left(y,x;\alpha\right)=\left[F_Y\left(y|x;\alpha\right)-q\right]$. By the Hoeffding decomposition, we have the first linear term being dominating since the second one is of smaller order following (C.4):

$$U_n^{(2)}\left(f\right)=\frac{2}{n}\sum_{i=1}^n\int\left[F_Y\left(x'\beta_o|x;\alpha\right)-q\right]\int K\left(u\right)f\left(u-x^ch\right)duW_\gamma\left(x,X_i\right)c\left(x'\beta_o|Z_i,x;\alpha_o\right)dx+U_n^{(2)}\left(\pi_2 f\right).$$

Under Assumption (G) we have $|F_Y\left(x'\beta_o|x;\alpha\right)-q|\leq M\delta_n$, hence both the square root of maximal variance and envelope functions are bounded up by δ_n in the first order Hoeffding decomposition. Hence by inequality (2.5) in Gine and Guillou (2001), we have

$$E\left\|U_n^{(1)}\left(\pi_1 f\right)\right\|_{|\alpha-\alpha_o|\leq O(\delta_n)}\leq M\sqrt{\frac{\log n}{n}}\delta_n=O\left(\sqrt{h^p\log n}\vee\sqrt{\frac{\log^3 n}{nh^p}}\right)=o\left(1\right),$$

which completes the proof. ∎

Proof of Step 2'. First of all, the normalized drifting term is still bounded from below:

$$\sqrt{n}P_{n3}\left(\alpha\right)\geq\sqrt{n}\int\left[F_Y\left(x'\beta_o|x;\alpha\right)-q\right]^2\widehat{f}_X^2\left(x\right)dx\geq\sqrt{n}\varepsilon_n+s.o.$$

By assumptions (K) and (V2) and compactness of $\mathcal{J}$, as in Appendix C we have the functional class $\mathcal{F}_4$ being of VC type with bounded envelope function and the maximal variance of order h^{3p} just as the calculation done in (B.9). Applying the maximal inequality by Gine and Mason (2007), we get that

$$h^{4p}E\left\|\frac{1}{n^2}\sum_{1\leq i<j\leq n}\int W_\gamma\left(x,X_i\right)W_\gamma\left(x,X_j\right)c\left(x'\beta_o|Z_i,x;\alpha_o\right)c\left(x'\beta_o|Z_j,x;\alpha_o\right)dx\right\|_{\mathcal{G}}^2\leq Mh^{3p}n^{-2}\left(\log n\right).$$

Hence by Markov inequality, uniformly we get

$$P_{n1}(\alpha) = \frac{1}{n^2} \sum_{1 \leq i < j \leq n} \int W_\gamma(x, X_i) W_\gamma(x, X_j) c(x'\beta_o, Z_i|x; \alpha) c(x'\beta_o|Z_i, x; \alpha_o) dx = O_p\left(\frac{\sqrt{\log n}}{nh^{p/2}}\right).$$

In order to deal with P_{n2}, we express it in terms of standard U-statistic with the negligible diagonal term:

$$P_{n2}(\alpha) = \frac{1}{n(n-1)} \sum_{i \neq j} \int [F_Y(x'\beta_o|x; \alpha) - q] W_\gamma(x, X_i) W_\gamma(x, X_j) c(x'\beta_o|Z_i, x; \alpha_o) dx + s.o.$$

After symmetrizations and by the inequality in Gine and Mason (2007), the above term is of order $O_p\left(\sqrt{\frac{\log n}{n}}\right)$. In sum, recalling Assumption (E), we have

$$\min_{\alpha \in \mathcal{A} \setminus \mathcal{A}_o^{\varepsilon n}} \frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}(\beta_o; \alpha)}} [T_n(\beta_o; \alpha) - B_n(\beta_o; \alpha)] \geq \frac{\sqrt{nh^p}}{\sqrt{\widehat{\Sigma}(\beta_o; \alpha)}} \left[O_p\left(\frac{\sqrt{\log n}}{\sqrt{n}}\right) + O_p\left(\sqrt{\log n}\right) + \sqrt{n}\varepsilon_n \right] \to +\infty,$$

because the final drifting term is positive and dominates others by Assumption (V2). ∎

Proof of Step 3'. Recall $\Sigma_o = \Sigma(\beta_o; \alpha_o)$, and the following string of inequalities bounding the studentized test statistic $\frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}(\beta_o; \alpha)}} [T_n(\beta_o; \alpha) - B_n(\beta_o; \alpha)]$ over the small neighborhood $\alpha \in \mathcal{A}_o^{\varepsilon n}$ is similar to Santos (2006) modulo the smaller order terms:

$$\sum_{j=1}^3 \inf_{\alpha \in \mathcal{A}_o^{\varepsilon n}} \frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}(\beta_o; \alpha)}} [P_{nj}(\alpha)] + s.o.$$

$$\leq \inf_{\alpha \in \mathcal{A}_o^{\varepsilon n}} \frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}(\beta_o; \alpha)}} [T_n(\beta_o; \alpha) - B_n(\beta_o; \alpha)]$$

$$\leq \frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}_o}} P_{n1}(\alpha_o) + \frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}_o}} (P_{n2}(\alpha_o) + P_{n3}(\alpha_o)) + s.o.$$

$$= \frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}_o}} P_{n1}(\alpha_o) + s.o.,$$

where the first inequality follows by taking infimum over three terms separately and the second inequality follows by the fact that $\alpha_o \in \mathcal{A}_o^{\varepsilon n}$. The last equality by $P_{n2}(\alpha_o) + P_{n3}(\alpha_o) = 0$, hence the studentized test statistic over $\alpha \in \mathcal{A}_o^{\varepsilon n}$ is bounded up by $\frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}_o}} P_{n1}(\alpha_o)$ plus smaller order term. Since that $\inf_{\alpha \in \mathcal{A}_o^{\varepsilon n}} P_{n3}(\alpha) = 0$ at α_o, we could bound the studentized test statistic from below by $\frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}_o}} P_{n1}(\alpha_o)$ once we show the following hold:

$$\sup_{\alpha \in \mathcal{A}_o^{\varepsilon n}} \left[\frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}(\beta_o; \alpha)}} |P_{n2}(\alpha)| \right] = o_p(1) \text{ and}$$

$$\inf_{\alpha \in \mathcal{A}_o^{\varepsilon n}} \left[\frac{nh^{p/2}}{\widehat{\Sigma}(\beta_o; \alpha)} P_{n1}(\alpha) \right] - \frac{nh^{p/2}}{\sqrt{\widehat{\Sigma}_o}} P_{n1}(\alpha_o) = o_p(1).$$

The last one is easy combining $\widetilde{\alpha}_o \to_p \alpha_o$ and the stochastic equicontinuity of $P_{n1}(\alpha)$ already proved in (B.2). Based on results in Step 2' we know $\widetilde{\alpha}_o$ must be in $\mathcal{A}_o^{\varepsilon n}$ for large enough n and Assumption (V2)

states the metric inducing the neighborhood $\mathcal{A}_o^{\varepsilon n}$ is equivalent as a standard neighborhood around α_o with radius $\sqrt{\varepsilon_n}$.

Also the proof handling $P_{n2}(\alpha)$ follows the complete analog in Lemma B.8 with a different radius $\sqrt{\varepsilon_n}$ around α_o:

$$E\left\|U_n^{(1)}(\pi_1 f)\right\|_{\alpha \in \mathcal{A}_o^{\varepsilon n}} \leq M\sqrt{\frac{\log n \times \varepsilon_n}{n}},$$

thereafter

$$\sup_{\alpha \in \mathcal{A}_o^{\varepsilon n}}\left[\frac{2nh^{p/2}}{\sqrt{\widehat{\Sigma}(\beta_o;\alpha)}}|P_{n2}(\alpha)|\right] = O_p\left(nh^{p/2} \times \sqrt{\frac{\varepsilon_n}{n}}\right) = O_p\left(\sqrt{nh^p\varepsilon_n}\right) = o_p(1).$$

∎

Proof of Bootstrap Consistency. Recall the expression for the bootstrap test statistic: for any α,

$$T_{n,b}^*(\beta_o;\alpha) \equiv \int\left[\frac{1}{n}\sum_{i=1}^n W_\gamma(x,X_i)c_b^*\left(x'\beta|Z_i,x,\alpha\right)\right]^2 dx$$

$$= \frac{1}{n^2}\sum_{i=1}^n \int W_\gamma^2(x,X_i)c_b^{*2}\left(x'\beta|Z_i,x,\alpha\right)dx +$$

$$\frac{2}{n^2}\sum_{1\leq i<j\leq n}\int W_\gamma(x,X_i)W_\gamma(x,X_j)c_b^*\left(x'\beta|Z_i,x,\alpha\right)c_b^*\left(x'\beta|Z_j,x,\alpha\right)dx$$

$$\equiv I_{n1}^*(\alpha) + 2I_{n2}^*(\alpha).$$

Let $\mathcal{Z}_n$ represent the whole sample. Apparently conditional on $\mathcal{Z}_n$, $I_{n2}^*(\alpha)$ is a degenerate U-statistic for any α, i.e. $E\left[I_{n2}^*(\alpha)|\mathcal{Z}_n, M_{i,b}^*\right] = 0$ by the construction of the multiplier bootstrap procedure. Referring to its conditional second moment, we have $\left(E\left[I_{n2}^{*2}(\alpha)|\mathcal{Z}_n\right] - \frac{n(n-1)}{4}h^{3p}\Sigma(\beta_o;\alpha)\right)$ converges to 0 for almost surely sample realization $\mathcal{Z}_n$. Hence continuing the same argument as in the proof of Lemma B.2, for almost surely sample realization $\mathcal{Z}_n$, we have

$$2nh^{p/2}I_{n2}^*(\alpha) \Longrightarrow N(0, \Sigma(\beta_o;\alpha)).$$

When it comes to $I_{n1}^*(\alpha)$, we get that $nh^{p/2}[I_{n1}^*(\alpha) - B_n(\alpha)]$ converges to zero in probability uniformly in α, for almost surely sample realization $\mathcal{Z}_n$. This is because $I_{n1}^*(\alpha)$ goes to its conditional expectation $E[I_{n1}^*(\alpha)|\mathcal{Z}_n] = I_{n1}(\alpha)$ defined in (B.4) by standard LLN and $nh^{p/2}[I_{n1}(\alpha) - B_n(\alpha)]$ goes to zero uniformly in α for almost surely sample realization $\mathcal{Z}_n$ because of (B.5) and the subsequent arguments. Also the negligibility of replacing the $\Sigma(\beta_o;\alpha)$ with its plug-in consistent estimator $\widehat{\Sigma}(\beta_o;\alpha)$ follows the same argument as in the proof of Step 2. In sum, we have shown the following conditional weak convergence:

$$\frac{nh^{p/2}}{\Sigma(\beta_o;\alpha)}\left[T_{n,b}^*(\beta_o;\alpha) - B_n(\alpha)\right] \Longrightarrow N(0,1),$$

uniformly in α. In particular the result holds for both $\widehat{\alpha}_o$ and $\widetilde{\alpha}_o$, hence the validity of bootstrap confidence sets follow immediately. ∎

9   Appendix C: Auxiliary Results

In this appendix, we present several useful results. The first one is the Hadamard differentiability of $S_Y \left(\cdot | x; \alpha \right)$ as a functional of $\left(S_V \left(\cdot | x\right), S_{V,D=1} \left(\cdot | x\right)\right)$, see Lemma C.1. Although we didn't actually use it in our proofs, as it is for a fixed α, we present it here since it generalizes a similar result for the Kaplan-Meier estimator in Van der Vaart and Wellner (1996) and the conditional Kaplan-Meier estimator in Dabraska (1987, 1989) assuming independent censoring. In Lemma C.2, we present the asymptotic variance and bias of the plug-in estimator $\widehat{F} \left(y | x; \alpha\right)$. Finally we collect some useful results on local U-processes used repeatedly in the proofs of the main results in Appendices A and B.

9.1   Hadamard Differentiability

Lemma C.1 *Suppose the copula generator $\varphi_\alpha \left(\cdot\right)$ is third order continuously differentiable and $\varphi_\alpha^{'} \left(\cdot\right)$ is not equal to zero. Then $S_Y \left(\cdot | x; \alpha\right)$ is Hadamard differentiable from the domain of $\mathcal{D}_- \left[y_l, y_u^0\right] \times \mathcal{BV}_1 \left[y_l, y_u^0\right]$ into $\mathcal{D}_- \left[y_l, y_u^0\right]$, where $\mathcal{D}_- \left[y_l, y_u^0\right]$ denote the space of càglàd functions on $\left[y_l, y_u^0\right]$ and $\mathcal{BV}_1 \left[y_l, y_u^0\right]$????*

Proof. First consider $\varphi_\alpha \circ S \left(\cdot | x; \varphi_x\right) = \theta \left(S_V \left(\cdot | x\right), S_{V,D=1} \left(\cdot | x\right)\right) \left(\cdot\right)$, where

$$\theta \left(S_V \left(\cdot | x\right), S_{V,D=1} \left(\cdot | x\right)\right) \left(\cdot\right) = \int_0^{\cdot} \varphi_\alpha^{'} \left\{S_V \left(s | x\right)\right\} dS_{V,D=1} \left(s | x\right).$$

Now we shall apply Lemma 20.10 in Van der Vaart (1998) to $\theta \left(S_V \left(\cdot | x\right), S_{V,D=1} \left(\cdot | x\right)\right)$. A close inspection shows that the proof does not rely on whether we are dealing with $\mathcal{D}_- \left[y_l, y_u^0\right]$ or $\mathcal{D} \left[y_l, y_u^0\right]$. The derivative is given combining Van der Vaart's formula and chain rule on φ_α^{-1}:

$$\frac{1}{\varphi_\alpha^{'} \circ S_Y \left(\cdot | x; \alpha\right)} \left(h_2 \varphi_\alpha^{'} \left(S_V \left(s | x\right)\right) |_{y_l}^{\cdot} - \int_{y_l}^{\cdot} h_{2-} d\varphi_\alpha^{'} \left(S_V \left(s | x\right)\right) + \int_{y_l}^{\cdot} \varphi_\alpha^{''} \left(S_V \left(s | x\right)\right) h_1 dS_{V,D=1} \left(s | x\right)\right) \text{(C.1)}$$

$$= \frac{1}{\varphi_\alpha^{'} \circ S_Y \left(\cdot | x; \alpha\right)} \left(h_2 \varphi_\alpha^{'} \left(S_V \left(\cdot | x\right)\right) - \int_{y_l}^{\cdot} h_{2-} \varphi_\alpha^{''} \left(S_V \left(s | x\right)\right) dS_V \left(s | x\right) + \int_{y_l}^{\cdot} \varphi_\alpha^{''} \left(S_V \left(s | x\right)\right) h_1 dS_{V,D=1} \left(s | x\right)\right)$$

Notice when we set $h_1 = F_V \left(\cdot | x\right) - \widehat{F}_V \left(\cdot | x\right)$ and $h_2 = F_{V,D=1} \left(\cdot | x\right) - \widehat{F}_{V,D=1} \left(\cdot | x\right)$, the above expression is consistent with our linear representation, although in a pointwise sense. ∎

Remark 9.1 *Given the weak convergence of kernel estimators of the (sub) survival functions of observable V and the above Hadamard differentiability, the weak convergence of the plug-in estimator is immediate (with a given α). Also, by Lemma 3.9.23 in van der Vaart and Wellner (1996), one could obtain the weak convergence of the conditional quantile process as long as the conditional density function exist and is bounded away from zero. Note merely upon change of notation, our Lemma C.1 also works for the marginal survival function $S_Y \left(\cdot | \alpha\right)$ when the covariate X is absent, which gives another proof of the weak convergence result in Rivest and Wells (2003), namely their Theorem 2. Moreover for the marginal survival function, since the empirical (sub) survival functions are efficient estimators, by the results in Section 25.7 in Van der Vaart (1998) we could conclude the efficiency of the plug-in estimator and the copula graphic estimator in Rivest and Wells (2003) given the above Hadamard differentiability. Again this generalizes the*

classical result showing the Kaplan-Meier estimator is efficient under independence censoring assumption, since now the dependence structure could be allowed to be different (but with a known generator φ and a fixed α).

9.2 Asymptotic Variance and Bias of $\widehat{F}(y|x;\alpha)$

Lemma C.2 *The asymptotic covariance of $\sqrt{nh^p}\left[\widehat{F}(t|x;\alpha) - F_Y(t|x;\alpha), \widehat{F}(s|x;\alpha) - F_Y(s|x;\alpha)\right]$ is given by*

$$
\Gamma(t,s|x;\alpha) = \frac{\|K\|_2^2}{f_X(x)\,\varphi_\alpha'\{S_Y(t|x;\alpha)\}\,\varphi_\alpha'\{S_Y(s|x;\alpha)\}}
$$

$$
\times \quad \left[\int_{y_l}^t \int_{y_l}^s \varphi_\alpha''\{S_V(u|x)\}\,\varphi_\alpha''\{S_V(v|x)\}\,[F_V(u\wedge v|x) - F_V(u|x)\,F_V(v|x)]\,dF_{V,D=1}(u|x)\,dF_{V,D=1}(v|x)\right.
$$

$$
- \int_{y_l}^t \varphi_\alpha''\{S_V(v|x)\}\,\varphi_\alpha'\{S_V(s|x)\}\,[F_{V,D=1}(v\wedge s|x) - F_V(v|x)\,F_{V,D=1}(s|x)]\,dF_{V,D=1}(v|x)
$$

$$
- \int_{y_l}^t \int_{y_l}^s \varphi_\alpha''\{S_V(u|x)\}\,\varphi_\alpha''\{S_V(v|x)\}\,[F_{V,D=1}(u\wedge v|x) - F_V(v|x)\,F_{V,D=1}(u|x)]\,dF_V(u|x)\,dF_{V,D=1}(v|x)
$$

$$
- \int_{y_l}^s \varphi_\alpha''\{S_V(v|x)\}\,\varphi_\alpha'\{S_V(t|x)\}\,[F_{V,D=1}(v\wedge t|x) - F_V(v|x)\,F_{V,D=1}(t|x)]\,dF_{V,D=1}(v|x)
$$

$$
+ \varphi_\alpha'\{S_V(t|x)\}\,\varphi_\alpha'\{S_V(s|x)\}\,[F_{V,D=1}(t\wedge s|x) - F_{V,D=1}(t|x)\,F_{V,D=1}(s|x)]
$$

$$
+ \int_{y_l}^s \varphi_\alpha''\{S_V(v|x)\}\,\varphi_\alpha'\{S_V(t|x)\}\,[F_{V,D=1}(v\wedge t|x) - F_V(t|x)\,F_{V,D=1}(v|x)]\,dF_V(v|x)
$$

$$
- \int_{y_l}^t \int_{y_l}^s \varphi_\alpha''\{S_V(u|x)\}\,\varphi_\alpha''\{S_V(v|x)\}\,[F_{V,D=1}(u\wedge v|x) - F_V(u|x)\,F_{V,D=1}(v|x)]\,dF_{V,D=1}(u|x)\,dF_V(v|x)
$$

$$
+ \int_{y_l}^t \varphi_\alpha''\{S_V(v|x)\}\,\varphi_\alpha'\{S_V(s|x)\}\,[F_{V,D=1}(v\wedge s|x) - F_{V,D=1}(v|x)\,F_{V,D=1}(s|x)]\,dF_V(v|x)
$$

$$
\left. + \int_{y_l}^t \int_{y_l}^s \varphi_\alpha''\{S_V(u|x)\}\,\varphi_\alpha''\{S_V(v|x)\}\,[F_{V,D=1}(u\wedge v|x) - F_V(u|x)\,F_{V,D=1}(v|x)]\,dF_V(u|x)\,dF_V(v|x)\right].
$$

Proof. We would first calculate the conditional covariance for the centering term $c(t|Z_i, x; \alpha)$ in the linear representation. Again we write conditional expectation given covariates as E', i.e. $E'[Z] = E[Z|X_1, \cdots, X_n]$ as in Hall (1984b). For any t, s, the following equalities need no further explanation:

$$
E'\left[(I[V_i \le t] - F_V(t|X_i))(I[V_i \le s] - F_V(s|X_i))\right] = F_V(t\wedge s|X_i) - F_V(t|X_i)\,F_V(s|X_i),
$$

$$
E'\left[(I[V_i \le t, D_i = 1] - F_{V,D=1}(t|X_i))(I[V_i \le s, D_i = 1] - F_{V,D=1}(s|X_i))\right]
$$
$$
= \quad F_{V,D=1}(t\wedge s|X_i) - F_{V,D=1}(t|X_i)\,F_{V,D=1}(s|X_i), \text{ and}
$$

$$
E'\left[(I[V_i \le t] - F_V(t|X_i))(I[V_i \le s, D_i = 1] - F_{V,D=1}(s|X_i))\right] = F_{V,D=1}(t\wedge s|X_i) - F_V(t|X_i)\,F_{V,D=1}(s|X_i).
$$

Therefore by straightforward yet tedious algebra, one gets

$$\eta\left(t,s|X_i,x;\alpha\right) = E^{'}\left[c\left(t|Z_i,x;\alpha\right)c\left(s|Z_i,x;\alpha\right)\right]$$

$$= \frac{1}{\varphi'_\alpha\left\{S_Y\left(t|x;\alpha\right)\right\}\varphi'_\alpha\left\{S_Y\left(s|x;\alpha\right)\right\}} \times$$

$$\left[\int_{y_l}^{t}\int_{y_l}^{s}\varphi''_\alpha\left\{S_V\left(u|x\right)\right\}\varphi''_\alpha\left\{S_V\left(v|x\right)\right\}\left[F_V\left(u\wedge v|X_i\right)-F_V\left(u|X_i\right)F_V\left(v|X_i\right)\right]dF_{V,D=1}\left(u|x\right)dF_{V,D=1}\left(v|x\right)\right.$$

$$-\int_{y_l}^{t}\varphi''_\alpha\left\{S_V\left(v|x\right)\right\}\varphi'_\alpha\left\{S_V\left(s|x\right)\right\}\left[F_{V,D=1}\left(v\wedge s|X_i\right)-F_V\left(v|X_i\right)F_{V,D=1}\left(s|X_i\right)\right]dF_{V,D=1}\left(v|x\right)$$

$$-\int_{y_l}^{t}\int_{y_l}^{s}\varphi''_\alpha\left\{S_V\left(u|x\right)\right\}\varphi''_\alpha\left\{S_V\left(v|x\right)\right\}\left[F_{V,D=1}\left(u\wedge v|X_i\right)-F_V\left(v|X_i\right)F_{V,D=1}\left(u|X_i\right)\right]dF_V\left(u|x\right)dF_{V,D=1}\left(v|x\right)$$

$$-\int_{y_l}^{s}\varphi''_\alpha\left\{S_V\left(v|x\right)\right\}\varphi'_\alpha\left\{S_V\left(t|x\right)\right\}\left[F_{V,D=1}\left(v\wedge t|X_i\right)-F_V\left(v|X_i\right)F_{V,D=1}\left(t|X_i\right)\right]dF_{V,D=1}\left(v|x\right)$$

$$+\varphi'_\alpha\left\{S_V\left(t|x\right)\right\}\varphi'_\alpha\left\{S_V\left(s|x\right)\right\}\left[F_{V,D=1}\left(t\wedge s|X_i\right)-F_{V,D=1}\left(t|X_i\right)F_{V,D=1}\left(s|X_i\right)\right]$$

$$+\int_{y_l}^{s}\varphi''_\alpha\left\{S_V\left(v|x\right)\right\}\varphi'_\alpha\left\{S_V\left(t|x\right)\right\}\left[F_{V,D=1}\left(v\wedge t|X_i\right)-F_V\left(t|X_i\right)F_{V,D=1}\left(v|X_i\right)\right]dF_V\left(v|x\right)$$

$$-\int_{y_l}^{t}\int_{y_l}^{s}\varphi''_\alpha\left\{S_V\left(u|x\right)\right\}\varphi''_\alpha\left\{S_V\left(v|x\right)\right\}\left[F_{V,D=1}\left(u\wedge v|X_i\right)-F_V\left(u|X_i\right)F_{V,D=1}\left(v|X_i\right)\right]dF_{V,D=1}\left(u|x\right)dF_V\left(v|x\right)$$

$$+\int_{y_l}^{t}\varphi''_\alpha\left\{S_V\left(v|x\right)\right\}\varphi'_\alpha\left\{S_V\left(s|x\right)\right\}\left[F_{V,D=1}\left(v\wedge s|X_i\right)-F_{V,D=1}\left(v|X_i\right)F_{V,D=1}\left(s|X_i\right)\right]dF_V\left(v|x\right)$$

$$+\int_{y_l}^{t}\int_{y_l}^{s}\varphi''_\alpha\left\{S_V\left(u|x\right)\right\}\varphi''_\alpha\left\{S_V\left(v|x\right)\right\}\left[F_{V,D=1}\left(u\wedge v|X_i\right)-F_V\left(u|X_i\right)F_{V,D=1}\left(v|X_i\right)\right]dF_V\left(u|x\right)dF_V\left(v|x\right)\right].$$

Hence we have the definition:

$$\sigma^2\left(y|X_i,x;\alpha\right)\equiv\eta\left(y,y|X_i,x;\alpha\right)\equiv E^{'}\left[c^2\left(y|Z_i,x;\alpha\right)\right]\text{ and }\sigma^2\left(y|x;\alpha\right)=\lim_{x_n\to x}\sigma^2\left(y|x_n,x;\alpha\right). \qquad (C.2)$$

An alternative way to proceed is to express the centering term as in Braekers and Veraverbeke (2005):

$$c\left(y|Z_i,x;\alpha\right)=\frac{-1}{\varphi'_\alpha\left\{S_Y\left(y|x;\alpha\right)\right\}}\left[\int_{y_l}^{y}\varphi''_\alpha\left\{S_V\left(v|x\right)\right\}\left[I\left(V_i\leq v\right)-F_V\left(v|X_i\right)\right]dF_{V,D=1}\left(v|x\right)\right. \quad (C.3)$$

$$\left.-\int_{y_l}^{y}\varphi'_\alpha\left\{S_V\left(v|x\right)\right\}d\left[I\left(V_i\leq v,D_i=1\right)-F_{V,D=1}\left(v|X_i\right)\right]\right],$$

by grouping the last two terms together. The result given by Braekers and Veraverbeke (2005) consists of seemingly four terms, yet a close inspection shows that there are still nine terms in total and upon integration by parts, and it is equivalent as what has been derived here. Now the asymptotic covariance follows the standard kernel estimation step:

$$\left(nh^p\right)cov\left[\widehat{F}\left(t|x;\alpha\right)-F_Y\left(t|x;\alpha\right),\widehat{F}\left(s|x;\alpha\right)-F_Y\left(s|x;\alpha\right)\right]$$

$$= h^p cov\left[w_{ni}\left(x,\gamma\right)c\left(t|Z_i,x;\alpha\right),w_{ni}\left(x,\gamma\right)c\left(s|Z_i,x;\alpha\right)\right]+o\left(1\right)$$

$$= \frac{1}{f_X^2\left(x\right)}E\left[h^{-p}K^2\left(\frac{X_i^c-x^c}{h}\right)E\left[c\left(t|Z_i,x;\alpha\right)c\left(s|Z_i,x;\alpha\right)|X_i\right]\right]+o\left(1\right)$$

$$= \frac{\|K\|^2}{f_X\left(x\right)}\eta\left(t,s|x;\alpha\right).$$

∎

Next lemma gives the order of the bias term.

Lemma C.3 *Under Assumptions (D)-(K), we have that*

$$\sup_{\alpha\in\mathcal{A}}\sup_{x\in\mathcal{J}}\sup_{y\in[y_l,y_u^0]}\left|\frac{1}{n}\sum_{i=1}^{n}W_\gamma\left(x,X_i\right)b\left(y|Z_i,x;\alpha\right)\right|=O_p\left(h^s+\sum_{j=1}^{r}\lambda_j\right).$$

Proof. Because $b\left(y|Z_i,x;\alpha\right)$ is linear in those three terms (see (3.15)), we shall only illustrate on the first one. Also the term $\frac{-1}{\varphi'_\alpha\{S_Y(y|x;\alpha)\}}$ remains uniformly bounded, hence it will be omitted in the derivation without affecting the stochastic order. We define the indicator functions $I_j\left(\widetilde{x}^d,x^d\right)$ for $j=1,...,r$ s.t. $I_j\left(\widetilde{x}^d,x^d\right)=1$ iff $\widetilde{x}^d$ and x^d differ only in the j-th component. Following Hall, Li and Racine (2004), we get

$$E\left[\sum_{i=1}^{n}W_\gamma\left(x,X_i\right)\int_{y_l}^{y}\varphi''_\alpha\{S_V\left(v|x\right)\}\left[F_V\left(v|X_i\right)-F_V\left(v|x\right)\right]dF_{V,D=1}\left(v|x\right)\right]$$

$$=\sum_{\widetilde{x}^d}\Pr\left(X^d=\widetilde{x}^d\right)\prod_{j=1}^{r}\left\{\left[\lambda_j/\left(1-\lambda_j\right)\left(c_j-1\right)\right]^{N_{ij}(x)}\left(1-\lambda_j\right)\right\}\times$$

$$\int_{y_l}^{y}\varphi''_\alpha\{S_V\left(v|x\right)\}\int K\left(u\right)\left[F_V\left(v|x^c-uh,\widetilde{x}^d\right)-F_V\left(v|x\right)\right]f_X\left(x-uh,\widetilde{x}^d\right)dudF_{V,D=1}\left(v|x\right)$$

$$=\int_{y_l}^{y}\sum_{j=1}^{r}\frac{\lambda_j}{c_j-1}\left\{\sum_{\widetilde{x}^d}I_j\left(\widetilde{x}^d,x^d\right)\left[F_V\left(v|x^c,\widetilde{x}^d\right)-F_V\left(v|x\right)\right]\right\}\varphi''_\alpha\{S_V\left(v|x\right)\}f_X\left(x^c,\widetilde{x}^d\right)dF_{V,D=1}\left(v|x\right)$$

$$+\frac{h^s}{s!}\int_{y_l}^{y}\sum_{j=1}^{p}\left[\int k\left(u_j\right)u_j^s du_j\sum_{i=0}^{s}\left(\frac{\partial^i F_V\left(v|x\right)}{\partial\left(x_j^c\right)^i}\frac{\partial^{s-i}f_X\left(x\right)}{\partial\left(x_j^c\right)^{s-i}}\right)\right]\varphi''_\alpha\{S_V\left(v|x\right)\}dF_{V,D=1}\left(v|x\right)+s.o.$$

$$\leq\sup_{\alpha\in\mathcal{A}}\sup_{x\in\mathcal{J}}\sup_{y\in[y_l,y_u]}\left|\int_{y_l}^{y}\varphi''_\alpha\{S_V\left(v|x\right)\}dF_{V,D=1}\left(v|x\right)\right|O\left(h^s+\sum_{j=1}^{r}\lambda_j\right)+s.o.$$

$$\leq O\left(h^s+\sum_{j=1}^{r}\lambda_j\right)+s.o.$$

Now the claim in the stated lemma follows from Markov inequality. ∎

9.3 Local U-processes

We borrow the following sequence of remarkable results on moment inequality in Gine and Mason (2007), and tail inequality in Major (2006). Compared with classical results on U-processes initiated by Nolan and Pollard (1987) and well summarized in de la Pena and Gine (1999), the following bounds allow the functional class to change with n and the (maximal) variance of individual function appears in the bound, as an analog of Bernstein type inequality. Hence they suit our purpose to handle the kernel type estimated function when bandwidth is changing with n and when the variance term is much smaller than the envelope function.

First some notations and terminologies will be collected here from Nolan and Pollard (1987), de la Pena and Gine (1999). We say a class of functions $\mathcal{F}$ is of VC type with respect to an envelope F if the covering number $N\left(\mathcal{F},L_2\left(Q\right),\varepsilon\right)$, the smallest number of $L_2\left(Q\right)$ open balls of radius ε required to cover $\mathcal{F}$, statisfies

$$N\left(\mathcal{F},L_2\left(Q\right),\varepsilon\right)\leq\left(\frac{M\left\|F\right\|_{L_2(Q)}}{\varepsilon}\right)^v\quad\text{for }0<\varepsilon\leq 2\left\|F\right\|_{L_2(Q)},$$

for some universal positive constants M, v and for every probability measure Q on the underlying space.

For a kernel function f of k variables, we denote

$$U_n^{(k)}(f) = \frac{(n-k)!}{n!} \sum_{i \in I_n^k} f(X_{i_1}, \cdots, X_{i_k}),$$

where $I_n^m = \{(i_1, \cdots, i_m) : 1 \leq i_j \leq n, i_j \neq i_k \text{ if } j \neq k\}$. Now suppose f is symmetric in its entries, we have the well-known Hoeffding decomposition:

$$U_n^{(m)}(f) - Ef = \sum_{k=1}^{m} U_n^{(k)}(\pi_k f),$$

where

$$\pi_k f = (\delta_{x_1} - P) \times \cdots \times (\delta_{x_k} - P) \times P^{m-k} f.$$

Moreover let σ^2 (which we call maximal variance) be any number satisfying

$$\left\| P^m f^2 \right\|_{\mathcal{F}} \leq \sigma^2 \leq M^2.$$

Lemma C.4 *(Gine and Mason, 2007, Theorem 8) Let $\mathcal{F}$ be a collection of measurable symmetric functions $f : S^m \to \mathcal{R}$, bounded up by M in absolute values, and let P be any probability measure on $(S, \mathcal{S})$. Assume $\mathcal{F}$ is of VC type with envelope function $F \equiv M$ and with characteristics A and v. Then for every $m \in \mathcal{N}$, and $A \geq e^m, v \geq 1$, there exist constants C_1, C_2, s.t. for any $k = 1, ..., m$,*

$$n^k E \left\| U_n^{(k)}(\pi_k f) \right\|_{\mathcal{F}}^2 \leq C_1 2^k \sigma^2 \left(\log \left(\frac{A}{\sigma} \right) \right)^k, \tag{C.4}$$

assuming $n\sigma^2 \geq C_2 \log \left(\frac{A}{\sigma} \right)$.

Lemma C.5 *(Major, 2006, Theorem 2) Let $\mathcal{F}$ be a collection of measurable symmetric functions $f : S^m \to \mathcal{R}$ satisfying the all assumptions stated in previous lemma, then we have*

$$\Pr \left\{ \sup_{f \in \mathcal{F}} \left| n^{k/2} U_n^{(k)}(\pi_k f) \right| \geq x \right\} \leq M \exp \left[-M \left(\frac{x}{\sigma} \right)^{2/k} \right], \tag{C.5}$$

if $n\sigma^2 \geq \left(\frac{x}{\sigma} \right)^{2/k} \geq M \left(\frac{1}{\log n} \right)^{3/2} \log \left(\frac{2}{\sigma} \right)$.

Remark 9.2 *The case that $m = 1$ corresponds to the usual empirical process result. Modulo the universal constant, the above lemmas are also in accordance with the rates obtained in Gine and Guillou (2001), Einmahl and Mason (2005) earlier, hence we will simply refer to the above two lemmas even in the case where we are dealing with the usual empirical process.*

Below we list those functional classes that have appeared in Appendices A and B, which fit into this local U-process (or empirical process) framework:

$$\mathcal{F}_1 \equiv \left\{ I[V \leq v] K \left(\frac{x^c - X^c}{h} \right) L(x^d, X^d, \lambda) : v \in \mathcal{R}^1, x \in \mathcal{J}, h \geq 0, \lambda_j \in \left[0, \frac{c_j - 1}{c_j} \right] \right\}, \tag{C.6}$$

$$
\mathcal{F}_2 \equiv \left\{ \begin{array}{c} D_1 \frac{\varphi_\alpha''(S_V(V_1|x))}{f_X^2(x)} 1\{V_1 > y\} h^p W_\gamma(x, X_1) H(X_2, V_2, V_1) - \\ \int d_1 \frac{\varphi_\alpha''(S_V(v_1|x))}{f_X^2(x)} 1\{v_1 > y\} h^p W_\gamma(x, x_1) H(X_2, V_2, v_1) dP_Z(v_1, d_1, x_1) \\ y \in [y_l, y_u^0], x \in \mathcal{J}, h \geq 0, \lambda_j \in [0, \frac{c_j - 1}{c_j}] \end{array} \right\}, \tag{C.7}
$$

$$
\mathcal{F}_3 \equiv \left\{ \begin{array}{c} \int c(y|Z_1, x; \alpha_o) K\left(\frac{x^c - X_1^c}{h}\right) L(x^d, X_1^d, \lambda) K\left(\frac{x^c - X_2^c}{h}\right) L(x^d, X_2^d, \lambda) G(y, x; \alpha_o) dx : \\ y \in [y_l, y_u^0], x \in \mathcal{J}, h \geq 0, \lambda_j \in [0, \frac{c_j - 1}{c_j}] \end{array} \right\}, \tag{C.8}
$$

$$
\mathcal{F}_4 \equiv \left\{ \begin{array}{c} \int c(y|Z_1, x; \alpha) c(y|Z_2, x; \alpha) K\left(\frac{x^c - X_1^c}{h}\right) L(x^d, X_1^d, \lambda) K\left(\frac{x^c - X_2^c}{h}\right) L(x^d, X_2^d, \lambda) dx : \\ y \in [y_l, y_u^0], h \geq 0, \lambda_j \in [0, \frac{c_j - 1}{c_j}] \end{array} \right\}, \tag{C.9}
$$

where the additional H and G functions appearing in $\mathcal{F}_2$ and $\mathcal{F}_3$ are defined in Lemma A.2, Lemma B.6, and Lemma B.8. To see that the entropy condition holds for all those four classes, note that the following three sub-classes are all uniformly bounded and of VC type:

$$
\begin{aligned}
\mathcal{F}_{1,1} &= \left\{ I[V \leq v] : v \in \mathcal{R}^1 \right\}, \\
\mathcal{F}_{1,2} &= \left\{ K\left(\frac{x^c - X^c}{h}\right) : x^c \in \mathcal{J}^c, h \geq 0 \right\}, \text{ and} \\
\mathcal{F}_{1,3} &= \left\{ L(x^d, X^d, \lambda) : \forall x^d, \lambda_j \in \left[0, \frac{c_j - 1}{c_j}\right] \right\}.
\end{aligned}
$$

Since the product is Lipschitz continuous with three individual terms using the boundedness of those functions, we have

$$
N\left(\mathcal{F}_1, L_2(Q), M\|F\|_{L_2(Q)}\varepsilon\right) \leq \prod_{i=1}^3 N\left(\mathcal{F}_{1,i}, L_2(Q), M\|F_i\|_{L_2(Q)}\varepsilon\right)
$$

by Theorem 2.10.20 in Van der Vaart and Wellner (1996), thus $\mathcal{F}_1$ is also of VC type. Complete analogy applies to $\mathcal{F}_2$ and $\mathcal{F}_3$. When it comes to $\mathcal{F}_4$, by our assumption (P1) on the generator functions, we obtain that $\forall f, f' \in \mathcal{F}_4, |f - f'| \leq M|\alpha - \alpha'|$, thus $\mathcal{F}_4$ is of VC type following Theorem 2.7.11 in Van der Vaart and Wellner (1996).

Now we illustrate with class $\mathcal{F}_1$ to give a quick proof of (A.3).

Proof of (A.3). For any $f_1 \in \mathcal{F}_1$, we have $Ef_1^2 = O\left(h^p\left(1 + \sum_{j=1}^r \lambda_j\right)\right) = O(h^p)$ as in Li and Racine (2007). Thereafter we could take the maximal variance for the current class as $\sigma_{\mathcal{F}_1}^2 = O(h^p)$.

$$
\begin{aligned}
&\left|\widehat{F}_V(y|x) - F_V(y|x)\right| \\
\leq\ & O_p\left(h^s + \sum_{j=1}^r \lambda_j\right) + O_p\left(\sqrt{E\left\|\frac{1}{h^p}U_n^{(1)}(\pi_1 f)\right\|_{\mathcal{F}_1}^2}\right) \\
\leq\ & O_p\left(h^s + \sum_{j=1}^r \lambda_j\right) + O_p\left(\frac{1}{h^p}\sqrt{\frac{h^p \log n}{n}}\right) = O_p\left(\sqrt{\frac{\log n}{nh^p}}\right),
\end{aligned}
$$

where the first inequality follows from standard bias-variance decomposition and in the second inequality we apply (C.4), and the final result follows our assumption (H). The strengthening to almost surely convergence is routine by the blocking device and Montgomeny-Smith inequality (see Theorem 1.1.5 in de la Pena and Gine, 1999) as presented in Gine and Guillou (2001), however, the weaker results are sufficient for our purpose. ∎

References

[1] Abbring, J.H., and G.J. van den Berg (2003), The identifiability of the mixed proportional hazards competing risks model, *Journal of the Royal Statistical Society B,* 65 701-710.

[2] Aitchison, J., and C.G.G. Aitken (1976), Multivariate binary discrimination by the kernel method, *Biometrika* 63, 413-420.

[3] Andrews, D.W.K., and X. Shi (2013), Inference based on conditional moment inequalities, *Econometrica,* 81, 609-666.

[4] Bedford, T., and I. Meilijson (1997), A characterization of marginal distributions of (possibly dependent) lifetime variables which right censor each other, *The Annals of Statistics,* 25 1622-1645.

[5] Berrington, A., and I. Diamond (2000), Marriage or cohabitation: A competing risks analysis of the first-partnership formation among the 1958 birth cohort, *Journal of the Royal Statistical Society A,* 163, 127-152.

[6] Bond, S., and J. Shaw (2006), Bounds on the covariate-time transformation for competing-risks survival analysis, *Lifetime Data Analysis* 12, 285–303.

[7] Booth, A. L., and S.E. Satchell (1995), The hazards of doing a PhD: An analysis of completion and withdrawal rates of British PhD Students in the 1980s, *Journal of the Royal Statistical Society A,* 158, 297-318.

[8] Braekers, R., and N. Veraverbeke (2005), A copula-graphic estimator for the conditional survival function under dependent censoring, *Canadian Journal of Statistics,* 33, 429–447.

[9] Buchinsky, M., and Hahn, J. (1998), An alternative estimator for the censored regression model, *Econometrica,* 66, 653–671.

[10] Chen, X., and Y. Fan (2006), Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification, *Journal of Econometrics,* 135, 125–154.

[11] Chen, Y.-H. (2010), Semiparametric marginal regression analysis for dependent competing risks under an assumed copula, *Journal of the Royal Statistical Society B,* 72 235-251.

[12] Chernozhukov, V., and H. Hong (2002), Three-step censored quantile regression and xxtramarital affairs, *Journal of The American Statistical Association,* 97, 872-882.

[13] Clayton, D., and J. Cuzick (1985), Multivariate generalizations of the proportional hazard model, *Journal of the Royal Statistical Society A,* 148, 82-117.

[14] Cox, D.R. (1959), The analysis of exponentially distributed life-times with two types of failure, *Journal of the Royal Statistical Society B,* 21, 411–421.

[15] Crowder, M. (1991), On the identifiability crisis in competing risk analysis, *Scandinavian Journal of Statistics* 18, 223-233.

[16] Dabrowska, D.M. (1987), Non-parametric regression with censored survival time data, *Scandinavian Journal of Statistics,* 14, 181-197.

[17] Dabrowska, D.M. (1989), Uniform consistency of kernel conditional Kaplan-Meier estimate, *The Annals of Statistics,* 17, 1157-1167.

[18] de la Pena, V.H., and E. Gine (1999), *Decoupling: From Dependence to Independence.* Springer.

[19] Deng, Y., Quigley, J.M., and R. Van Order (2000), Mortgage terminations, heterogeneity and the exercise of mortgage options, *Econometrica,* 68, 275-307.

[20] Einmahl, U., and D.M. Mason (2005), Uniform in bandwidth consistency of kernel-type function estimators, *The Annals of Statistics,* 33, 1380-1403.

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

[22] Fan, Y., and Q. Li (1996), Consistent model specification tests: omitted variables and semiparametric functional forms, *Econometrica,* 64, 865-890.

[23] Fan, Y., Li, Q., and I. Min (2006), A nonparametric bootstrap test of conditional distributions, *Econometric Theory,* 22, 587-613.

[24] Fermanian, J.-D. (2003), Nonparametric estimation of competing risks models with covariates, *Journal of Multivariate Analysis,* 85, 156–191.

[25] Fleming, T.R., and D.P. Harrington (1991), *Counting Processes and Survival Analysis,* Wiley, New York.

[26] Flinn, C.J., and J.J. Heckman (1982), New methods for analyzing structural models of labor force dynamics, *Journal of Econometrics,* 18, 115-168.

[27] Gine, E., and A. Guillou (2001), On consistency of kernel density estimators for randomly censored data: Rates holding uniformly over adaptive intervals, *Annales de l'Institut Henri Poincaré,* 37 503-522.

[28] Gine, E., and D.M. Mason (2007), On local U-statistic processes and the estimation of densities of functions of several sample variables, *The Annals of Statistics,* 35, 1105-1145.

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

[30] Hall, P. (1984b), Integrated square error properties of kernel estimators of regression functions, *The Annals of Statistics,* 12, 241-260.

[31] Hall, P. and C. Heyde (1980), *Martingale Limit Theory and Its applications,* Academic, New York.

[32] Hall, P., Racine, J., and Q. Li, (2004). Cross-validation and the estimation of conditional probability densities, *Journal of The American Statistical Association,* 99, 1015–1026.

[33] Hardle, W., and E. Mammen (1993), Comparing nonparametric versus parametric regression fits. *The Annals of Statistics,* 21, 1926-1947.

[34] Heckman, J.J., and B. Honore (1989), The identifiability of competing risks model, *Biometrika,* 76, 325-330.

[35] Henry, M., and I. Mourifie (2012), Sharp bounds in the binary Roy model, *Working Paper.*

[36] Honore, B., Khan, S., and J. Powell (2002), Quantile regression under random censoring, *Journal of Econometrics*, 109, 67–105.

[37] Honore, B., and A. Lleras-Muney (2006), Bounds in competing risks models and the war on cancer, *Econometrica,* 74, 1675-1698.

[38] Hsiao, C., Li, Q., and J.S. Racine (2007), A consistent model specification test with mixed discrete and continuous data, *Journal of Econometrics*, 140, 802–826.

[39] Huang, X., and N. Zhang (2008), Regression survival analysis with an assumed copula for dependent censoring: A sensitivity analysis approach. *Biometrics,* 64, 1090-1099.

[40] Joe, H. (1997), *Multivariate models and dependence concepts,* Chapman & Hall.

[41] Jun, S.J., and J. Pinske (2009), Semiparametric tests of conditional moment restrictions under weak or partial identification. *Journal of Econometrics,* 152, 3–18.

[42] Katz, L.E, and B.D. Meyer (1990), Unemployment insurance, recall expectations, and unemployment outcomes, *Quarterly Journal of Economics,* 105, 973-1002.

[43] Khan, S., and Powell, J.L. (2001), Two-Step Estimation of Semiparametric Censored Regression Models, *Journal of Econometrics*, 100, 319–355.

[44] Khan, S., and E. Tamer (2009), Inference on endogenously censored regression models using conditional moment inequalities, *Journal of Econometrics*, 152, 104–119.

[45] Khan, S., Ponomareva, M., and E. Tamer (2011), Sharpness in randomly censored linear models, *Economic Letters,* 113, 23-25.

[46] Klein, J.P., and M.L. Moeschberger (1988), Bounds on net survival probabilities for dependent competing risks, *Biometrics* 22, 529-538.

[47] Kline, P., and A. Santos (2013), Sensitivity to missing data assumptions: Theory and an evaluation of the U.S. wage structure, *Quantitative Economics,* 4, 231–267.

[48] Lavergne, P., and V. Patilea (2013), Smooth minimum distance estimation and testing with conditional estimating equations: Uniform in bandwidth theory, *Journal of Econometrics*, 177, 47–59.

[49] Lee, S. (2006), Identification of a competing risks model with unknown transformations of latent failure times, *Biometrika* 93, 996–1002.

[50] Lee, S., and A. Lewbel (2012), Nonparametric identification of accelerated failure time competing risks model, *Econometric Theory,* forthcoming.

[51] Li, Q., and J.S. Racine (2007), *Nonparametric Econometrics: Theory and Practice,* Princeton Press.

[52] Linton, O. (1997), Asymptotic inefficiency of an estimator derived from a kernel-based test statistic (Problem), *Econometric Theory,* 13, 306-307.

[53] Linton, O. (1998), Asymptotic inefficiency of an estimator derived from a kernel-based test statistic (Solution), *Econometric Theory,* 14, 153-154.

[54] Lopez, O. (2011), Nonparametric estimation of the multivariate distribution function in a censored regression model with applications, *Communications in Statistics: Theory and Methods,* 40, 2639-2660.

[55] Major, P. (2006), An estimate on the supremum of a nice class of stochastic integrals and U-statistics, *Probability Theory and Related Fields,* 134, 489-537.

[56] Manski, C. (1994), The selection problem, in C. Sims (editor) *Advances in Econometrics, Six World Congress,* Cambridge University Press, 143-170.

[57] Manski, C. (2003), *Partial identification of probability distributions,* Springer-Verlag.

[58] Nelson, R.B. (2006), *An Introduction to Copulas*, 2nd Edition, Springer.

[59] Nolan, D. and D. Pollard (1987), U-processes: rates of convergence, *The Annals of Statistics,* 15, 780-799.

[60] Peng, L., and Y. Huang, (2008), Survival analysis with quantile regression models, *Journal of the American Statistical Association,* 103, 637–649.

[61] Portnoy, S. (2003), Censored regression quantiles, *Journal of the American Statistical Association,* 98, 1001-1012.

[62] Powell, J. (1984), Least absolute deviation estimation for the censored regression model, *Journal of Econometrics*, 25, 303-325.

[63] Powell, J. (1986), Censored regression quantiles, *Journal of Econometrics*, 32, 143-155.

[64] Peterson, A.V. (1976), Bounds for a joint distribution function with fixed sub-distribution functions: applications to competing risks. *Proc. Natl. Acad. Sci. USA* 73**,** 11-13.

[65] Rivest, L.-P., and M.T. Wells (2001), A martingale approach to the copula-graphic estimator for the survival function under dependent censoring, *Journal Multivariate Analysis,* 79, 138-155.

[66] Santos, A. (2006), Inference in nonparametric instrumental variables with partial identification, working paper.

[67] Santos, A. (2012), Inference in nonparametric instrumental variables with partial identification, *Econometrica,* 80, 213-276.

[68] Slud, E.V., and L.V. Rubinstein (1983), Dependent competing risks and summary survival curves, *Biometrika,* 70, 643-649.

[69] Szydlowski, A. (2013), Endogenous censoring in the mixed proportional hazard model with an application to optimal unemployment insurance, *Job Market Paper*, Northwestern University.

[70] Torgovitsky, A. (2011), Identification of nonseparable models with general instruments, working paper.

[71] Tsiatis, A. (1975), A nonidentifiability aspect of the problem of competing risks. *Proc. Natl. Acad. Sci. USA* 72, 20-22.

[72] Tsybakov, A. (2008), *Introduction to Nonparametric Estimation,* Springer, New York.

[73] Van der Vaart A. (1998), *Asymptotic Statistics,* Cambridge University Press.

[74] Wang, H.J., and L. Wang (2009), Locally weighted censored quantile regression, *Journal of the American Statistical Association,* 104, 1117-1128.

[75] Zheng, M., and J.P. Klein (1995), Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika,* 82, 127-138.

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