UC Irvine
UC Irvine Previously Published Works

Title
Proportional cross-ratio model

Permalink
https://escholarship.org/uc/item/5sr566x6

Journal
Lifetime Data Analysis, 25(3)

ISSN
1380-7870

Authors
Hu, Tianle
Nan, Bin
Lin, Xihong

Publication Date
2019-07-01

DOI
10.1007/s10985-018-9451-6

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Lifetime Data Anal. Author manuscript; available in PMC 2020 July 01.

Published in final edited form as:

Lifetime Data Anal. 2019 July ; 25(3): 480–506. doi:10.1007/s10985-018-9451-6.

Proportional Cross-Ratio Model

Tianle Hu,

Eli Lilly and Company, Indianapolis, IN 46285, Tel.: +1-317-4420568

Bin Nan, and

Department of Statistics, University of California - Irvine, Irvine, CA 92697, Tel: +1-949-8242954

Xihong Lin

Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, Tel: +1-617-4322914

Abstract

Cross-ratio is an important local measure of the strength of dependence among correlated failure times. If a covariate is available, it may be of scientific interest to understand how the cross-ratio varies with the covariate as well as time components. Motivated by the Tremin study, where the dependence between age at a marker event reflecting early lengthening of menstrual cycles and age at menopause may be affected by age at menarche, we propose a proportional cross-ratio model through a baseline cross-ratio function and a multiplicative covariate effect. Assuming a parametric model for the baseline cross-ratio, we generalize the pseudo-partial likelihood approach of Hu et al. (2011) to the joint estimation of the baseline cross-ratio and the covariate effect. We show that the proposed parameter estimator is consistent and asymptotically normal. The performance of the proposed technique in finite samples is examined using simulation studies. In addition, the proposed method is applied to the Tremin study for the dependence between age at a marker event and age at menopause adjusting for age at menarche. The method is also applied to the Australian twin data for the estimation of zygosity effect on cross-ratio for age at appendicitis between twin pairs.

Keywords

Bivariate survival; Cross-ratio; Empirical process theory; Local pseudo-partial likelihood; U-process

1 Introduction

In female reproductive aging research, there has been considerable interest in identifying marker events for the onset of menopausal transition and investigating their utility for predicting the age at menopause. In the Tremin study, conducted as part of the Menstrual and Reproductive Health Study (Treloar, Boynton, Behn, and Brown 1967), scientists are interested in understanding several bleeding pattern change criteria that have been proposed as potential marker events for the early stage of menopausal transition. For instance, it has been suggested that the age at onset of experiencing a menstrual cycle of at least 45 days in length might be a good marker for the early menopausal transition (Lisabeth, Harlow,

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Gillespie, Lin, and Sowers 2004). However, the validity of these proposed bleeding markers and their associations with the age at menopause have not been adequately investigated, and sophisticated statistical analysis tools are lacking in this area.

To formally assess the utility of a proposed bleeding marker, Nan et al. (2006) analyzed the association between the age at a marker event (defined as the age at onset of a specific bleeding pattern change) and the age at natural menopause (defined as the final menstrual period (FMP), with FMP confirmed after at least 12 months of amenorrhea). They proposed using cross-ratio to measure the dependence by assuming the cross-ratio to be a piecewise constant function of the age at onset of the marker event. They focused on the age at which a woman first experienced a menstrual cycle of at least 45 days in length, which has been proposed as a marker event for entry into the early menopausal transition stage.

One advantage of using cross-ratio as the dependence measure is that it has an attractive hazard ratio interpretation comparing two groups of practical interest, which is simple to understand for practitioners and provides a convenient way to evaluate the marker. In particular, the cross-ratio can be interpreted as the relative hazard of menopause comparing women who have experienced the marker event at a certain age with women who have not yet experienced the marker event.

However, the piecewise constant model requires prior knowledge on cut-off points which is usually lacking in practice. To bypass the difficulty in determining the cut-off points in the piecewise constant model, we estimate the cross-ratio as a smooth function of t_1 and t_2. A similar idea was described in Hu et al (2011), where the cross-ratio is estimated by a flexible continuous function of both time components via a pseudo partial-likelihood approach without considering the covariate effect on the cross-ratio.

Moreover, in the Tremin Trust data, the cross-ratio of age at menopause and age at the 45-day cycle marker event may be affected by age at menarche, which motivates a model that explicitly characterizes the covariate effect on the cross-ratio function directly. It is well known that when a covariate exists, cross-ratio for the failure times of the two members of a pair should be estimated with some adjustment for known characteristics of the pair (Clayton 1978; Oakes 1982, 1986, 1989). For example, in the Australian twin study of appendicitis, Duffy et al. (1990) discovered significant concordance rate with respect to appendicitis within twin pairs. It was also found that monozygotic twins exhibited higher concordance rate than dizygotic twins, likely due to shared genetic factors. Therefore, it is of interest to quantify this genetic effect on cross-ratio within twin pairs.

In the literature, the covariate effect is often modeled through marginal distributions. Shih & Louis (1995) proposed a model that incorporates covariates via marginal Cox regression model, assuming constant cross-ratio θ. Likewise, when θ is piecewise constant on a grid of the sample space of (T_1, T_2), Nan et al. (2006) proposed a sequential two-stage method where covariates are modeled via marginal Cox regression model. Its estimation is similar to the two-stage method of Shih & Louis (1995) for the Clayton copula model, but with left truncation at the lower left corner of each rectangle. Fan and Prentice (2002) adjusted their

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

previously proposed class of weighted dependence measures for bivariate failure times to accommodate covariate effects on marginal hazard rates as well.

However, when the cross-ratio function itself is of major interest, modeling the covariate effect via marginal models does not answer explicitly how a covariate changes the cross-ratio or by how much. Mimicking the Cox proportional hazards model, we propose an analogous model where the covariate effect is multiplicative on cross-ratio. One novelty of this model lies in linking the covariate effect to the cross-ratio explicitly, by extending the model of Hu et al (2011) to the regression setting.

For estimation, we construct an objective function, which we call the local pseudo-partial likelihood, by mimicking the partial likelihood of the Cox proportional hazards model (Cox 1972). Specifically, when the covariate is discrete with finite levels, we group observations into distinct strata by covariate values. Within each stratum, we then treat whether an event happens at a time point or beyond along one time axis as a binary covariate and the other time component as the survival outcome variable, and construct the corresponding partial likelihood function. When the covariate is continuous, kernel smoothing is applied to the estimating equations. We obtain the parameter estimates by maximizing the local pseudo-partial likelihood function. This construction does not need any model for either the joint or the marginal survival function, and thus is robust against model mis-specification. We show that the proposed parameter estimator is consistent and asymptotically normal. The proposed method is readily extendable to the estimation of an arbitrary baseline cross-ratio function by using tensor product splines.

2 The conditional cross-ratio function given covariate

Let (T_1, T_2) be a pair of absolutely continuous failure times. In the Tremin Trust data, T_1 is time to the 45-day cycle marker and T_2 is time to menopause. Given covariate W, e.g., age at menarche, cross-ratio is a quantity conditional on W. Specifically, the definition of cross-ratio becomes:

$$\theta(t_1, t_2, w) = \frac{\lambda_2(t_2 | T_1 = t_1, W = w)}{\lambda_2(t_2 | T_1 > t_1, W = w)} = \frac{\lambda_1(t_1 | T_2 = t_2, W = w)}{\lambda_1(t_1 | T_2 > t_2, W = w)}, \quad (1)$$

where λ_1 and λ_2 are the conditional hazard functions of T_1 and T_2, respectively, given a common covariate W for both survival times. We propose an analogous model to the Cox proportional hazards model with multiplicative covariate effect on the cross-ratio:

$$\theta(t_1, t_2, w) = \theta_0(t_1, t_2) \exp(\alpha w), \quad (2)$$

where $\theta_0(t_1, t_2)$ is the baseline cross-ratio, i.e.

Author Manuscript

$$\theta_0(t_1, t_2) = \frac{\lambda_2(t_2 \mid T_1 = t_1, W = 0)}{\lambda_2(t_2 \mid T_1 > t_1, W = 0)} = \frac{\lambda_1(t_1 \mid T_2 = t_2, W = 0)}{\lambda_1(t_1 \mid T_2 > t_2, W = 0)}.$$

Model (2), which we call the proportional cross-ratio model, effectively separates the baseline cross-ratio function and the covariate effect, so that we can model each piece individually. We consider a parametric model $\beta_0(t_1, t_2; \gamma) = \log \theta_0(t_1, t_2)$ parameterized by a finite-dimensional Euclidean parameter γ. It is straightforward to extend the parametric model to a nonparametric model using tensor product splines. For covariate W, we consider a linear function parameterized by a Euclidean parameter α. Specifically, we assume

$$\begin{aligned} \beta(t_1, t_2, w; \xi) &= \beta_0(t_1, t_2; \gamma) + \alpha w \\ &= \sum_{k,l} \gamma_{kl} b_{kl}(t_1, t_2) + \alpha w, \end{aligned} \quad (3)$$

where ξ is the finite-dimensional vector of coefficients $\{\gamma_{kl}\}$ and α, and $\{b_{kl}\}$ are the basis functions of t_1 and t_2 that do not involve parameter ξ. For notational simplicity, we consider one-dimensional covariate W hereafter. Results developed in this article hold for any finite-dimensional discrete covariates, but need to be properly modified for multiple continuous covariates when a multi-dimensional kernel smoothing is implemented.

3 Regression parameter estimation

To estimate the baseline cross-ratio function and the covariate effect jointly, we first focus on a discrete covariate with a finite number of levels, by creating a dummy variable for each level or assuming a linear trend across levels. We then extend this method to continuous covariates using smoothing techniques, in particular, applying kernel smoothing to the estimating equation obtained for a discrete covariate.

Suppose we observe n independent and identically distributed copies of $(X_1, X_2, \Delta_1, \Delta_2, W)$, where $X_1 = \min(T_1, C_1)$, $X_2 = \min(T_2, C_2)$, $\Delta_1 = I(T_1 \le C_1)$, and $\Delta_2 = I(T_2 \le C_2)$. Here $I(\cdot)$ denotes the indicator function. The pair of continuous failure times (T_1, T_2) are subject to right censoring by a pair of censoring times (C_1, C_2). Assume censoring times are independent of failure times conditional on covariate W. We further assume that there are no ties among observed times for each of the two time components.

3.1 Discrete covariate with a finite number of levels

Borrowing the idea in Hu et al. (2011), we construct an objective function by treating $\{j: T_{1j} = t_1\}$ and $\{j: T_{1j} > t_1\}$ as the "exposure" group and the "non-exposure" group respectively. Then from the first equality in (1), the cross-ratio $\theta(t_1, t_2, w)$ becomes the hazard ratio of T_2 between these two groups within the stratum $W = w$. Denote $\lambda_2(X_{2j} | X_{1j} > X_{1i}, W_k = W_i)$ by A_{ij}^k and $\theta(X_{1i}, X_{2j}, W_i)^{I(X1k = X1i)}$ by B_{ij}^k respectively. By mimicking the partial likelihood idea, we can construct the objective function as follows based on these two groups categorized by $t_1 = X_{1i}$:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\prod_{j=1}^{n}\left[\frac{A_{ij}^{j}B_{ij}^{j}}{\sum_{X_{2k}\geq X_{2j}} I\left(W_k=W_i\right)I\left(X_{1k}\geq X_{1i}\right)A_{ij}^{k}B_{ij}^{k}}\right]^{I\left(W_j=W_i\right)I\left(X_{1j}\geq X_{1i}\right)\Delta_{2j}\Delta_{1i}}$$
$$=\prod_{j=1}^{n}\left[\frac{B_{ij}^{j}}{\sum_{X_{2k}\geq X_{2j}} I\left(W_k=W_i\right)I\left(X_{1k}\geq X_{1i}\right)B_{ij}^{k}}\right]^{I\left(W_j=W_i\right)I\left(X_{1j}\geq X_{1i}\right)\Delta_{2j}\Delta_{1i}},$$

where A_{ij}^{j} cancels with A_{ij}^{k} in the above equation because of the restriction $W = W_i$, which is achieved by indicators $I(W_j = W_i)$ in the outer exponent and $I(W_k = W_i)$ in the denominator. Following a similar argument to Hu et al. (2011), the denominator in the bracket can be simplified as $N(X_{1i}, X_{2j}; W_i) - I(X_{2j} \leq X_{2i})(1 - \theta(X_{1i}, X_{2j}; W_i))$, where $N\left(t_1, t_2, w\right) = \sum_{k=1}^{n} I\left(X_{1k} \geq t_1, X_{2k} \geq t_2, W_k = w\right)$. So we can rewrite the above objective function as

$$\prod_{j=1}^{n}\left[\frac{\theta\left(X_{1i}, X_{2j}, W_i\right)^{I\left(X_{1j}=X_{1i}\right)}}{N\left(X_{1i}, X_{2j}, W_i\right)-I\left(X_{2j}\leq X_{2i}\right)\left(1-\theta\left(X_{1i}, X_{2j}, W_i\right)\right)}\right]^{I\left(W_i=W_j\right)I\left(X_{1j}\geq X_{1i}\right)\Delta_{1i}\Delta_{2j}}. \quad (4)$$

Now denote (4) as $L_i^{(1)}$ Considering the symmetric structure of the definition of $\theta(t_1, t_2, w)$ determined by the second equality in (1), we can construct a similar objective function as (4) by switching the roles of X_1 and X_2, and denote it as $L_i^{(2)}$ By multiplying such constructed two objective functions over all possible ways of creating the “exposure” and “non-exposure” groups, i.e. all subjects, we obtain the following local pseudo-partial likelihood function:

$$L_n = \prod_{i=1}^{n} L_i^{(1)} L_i^{(2)}. \quad (5)$$

The estimator obtained by maximizing (5) is then called the maximum local pseudo-partial likelihood estimator.

Denote $l_n = n^{-1} \log L_n$, $\xi = (\gamma, \alpha)$ and $\dot{\beta}\left(t_1, t_2, w\right) = \partial\beta\left(t_1, t_2, w; \xi\right) / \partial\xi$ that is free of ξ. Differentiating $l_n(\xi)$ with respect to ξ and assuming no ties among observed times, we obtain the following estimating function for ξ:

$$\boldsymbol{U}_n(\xi) = \frac{\partial l_n(\xi)}{\partial \xi} = \boldsymbol{U}_n^{(1)}(\xi) - \boldsymbol{U}_n^{(2)}(\xi) + \boldsymbol{U}_n^{(3)}(\xi) - \boldsymbol{U}_n^{(4)}(\xi),$$

where

$$U_n^{(1)} = U_n^{(3)} = \frac{1}{n}\sum_{i=1}^{n} \Delta_{1i}\Delta_{2i}\dot{\beta}\left(X_{1i}, X_{2i}, W_i\right) \quad (6)$$

and

$$U_n^{(2)} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{I\left(W_j = W_i\right)\Delta_{1i}\Delta_{2j}I\left(X_{1j} \ge X_{1i}\right)I\left(X_{2j} \le X_{2i}\right)e^{\beta\left(X_{1i}, X_{2j}, W_i; \xi\right)}}{N\left(X_{1i}, X_{2j}, W_i\right) - I\left(X_{2j} \le X_{2i}\right)\left(1 - e^{\beta\left(X_{1i}, X_{2j}, W_i; \xi\right)}\right)} \quad (7)$$

$$\times \dot{\beta}\left(X_{1i}, X_{2j}, W_i\right),$$

$$U_n^{(4)} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{I\left(W_j = W_i\right)\Delta_{1j}\Delta_{2i}I\left(X_{2j} \ge X_{2i}\right)I\left(X_{1j} \le X_{1i}\right)e^{\beta\left(X_{1j}, X_{2i}, W_i; \xi\right)}}{N\left(X_{1j}, X_{2i}, W_i\right) - I\left(X_{1j} \le X_{1i}\right)\left(1 - e^{\beta\left(X_{1j}, X_{2i}, W_i; \xi\right)}\right)} \quad (8)$$

$$\times \dot{\beta}\left(X_{1j}, X_{2i}, W_i\right).$$

Note that by switching indices i and j, (7) and (8) only differ in the second term of their denominators, which is a negligible term asymptotically. Then an estimator $\hat{\boldsymbol{\xi}}_n$ can be obtained by solving the equation $\boldsymbol{U}_n(\boldsymbol{\xi}) = 0$ using Newton-Raphson algorithm.

3.2 Continuous covariate

When the covariate is continuous, the "grouping" idea by restricting observations with the same covariate values into distinct strata is no longer applicable. However, based on the estimating equations obtained for a discrete covariate, we replace the grouping indicator function $I(W_j = W_i)$ by a kernel function $\mathbf{K}_h(W_j - W_i)$ in (7) and (8), where $\mathbf{K}_h(\cdot) = 1/h\mathbf{K}(\cdot/h)$ and h is a bandwidth. Function $\mathbf{K}(\cdot)$ is usually chosen to be a symmetric probability density function. In the numerical study presented later, we use the standard normal kernel. Specifically, we propose the following estimating function for $\boldsymbol{\xi}$ when the covariate is continuous:

$$U_n(\xi) = U_n^{(1)}(\xi) - U_n^{(2)}(\xi) + U_n^{(3)}(\xi) - U_n^{(4)}(\xi),$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

where $\boldsymbol{U}_n^{(1)}$ and $\boldsymbol{U}_n^{(3)}$ are the same as in (6) and

$$\boldsymbol{U}_n^{(2)} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\mathbf{K}_h\left(W_j - W_i\right)\Delta_{1i}\Delta_{2j}I\left(X_{1j} \ge X_{1i}\right)I\left(X_{2j} \le X_{2i}\right)e^{\beta\left(X_{1i}, X_{2j}, W_i; \xi\right)}}{N\left(X_{1i}, X_{2j}, W_i\right) - \mathbf{K}_h(0)I\left(X_{2j} \le X_{2i}\right)\left(1 - e^{\beta\left(X_{1i}, X_{2j}, W_i; \xi\right)}\right)} \times \dot{\beta}\left(X_{1i}, X_{2j}, W_i\right),$$

$$\boldsymbol{U}_n^{(4)} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\mathbf{K}_h\left(W_j - W_i\right)\Delta_{1j}\Delta_{2i}I\left(X_{2j} \ge X_{2i}\right)I\left(X_{1j} \le X_{1i}\right)e^{\beta\left(X_{1j}, X_{2i}, W_i; \xi\right)}}{N\left(X_{1j}, X_{2i}, W_i\right) - \mathbf{K}_h(0)I\left(X_{1j} \le X_{1i}\right)\left(1 - e^{\beta\left(X_{1j}, X_{2i}, W_i; \xi\right)}\right)} \times \dot{\beta}\left(X_{1j}, X_{2i}, W_i\right),$$

where $N\left(t_1, t_2, w\right) = \sum_{k=1}^{n} I\left(X_{1k} \ge t_1, X_{2k} \ge t_2\right)\mathbf{K}_h\left(W_k - w\right)$. Then an estimator $\widehat{\boldsymbol{\xi}}_n$ can be obtained by solving the equation $\boldsymbol{U}_n(\boldsymbol{\xi}) = 0$ using Newton-Raphson algorithm.

4 Asymptotic properties

In this section, we provide asymptotic results for the estimation of $\boldsymbol{\xi}$ in (3). We consider the following regularity conditions for model (3):

C1. The covariate W is either continuous or discrete with finite levels, whose sample space $\mathcal{W}$ is bounded with $0 < \inf_{w \in \mathcal{W}} f(w)$ and $\sup_{w \in \mathcal{W}} f(w) < \infty$. Here f is the density function of W.

C2. Consider the support region $(t_1, t_2) \in [0, \tau_1) \times [0, \tau_2)$, $0 < \tau_1, \tau_2 <, \infty$ with $\inf_{w \in \mathcal{W}} Pr(T_1 > \tau_1, T_2 > \tau_2 | W = w) > 0$ and $\inf_{w \in \mathcal{W}} Pr(C_1 > \tau_1, C_2 > \tau_2 | W = w) > 0$.

C3. The parameter space of $\boldsymbol{\xi}$, denoted by Γ, is a compact set, and the true value $\boldsymbol{\xi}_0$ is an interior point of Γ.

C4. The matrix $E\left\{\Delta_1\Delta_2\dot{\beta}\left(X_1, X_2, W\right)^{\otimes 2}\right\}$ is positive definite. Here $\dot{\beta}^{\otimes 2} = \dot{\beta}\dot{\beta}'$.

C5. (T_1, T_2) and (C_1, C_2) are independent conditional on W.

For a continuous covariate, in order for the kernel smoothing technique to work, the following conditions are further warranted in addition to the above regularity conditions with functions $\tilde{h}()$, $b()$ and $S()$ defined in equations (14), (13) and (12) given in Appendix A:

C6. For some ϵ satisfying $0 < \epsilon \le 1, \tilde{h}(V; \boldsymbol{\xi}) < \infty$ is uniformly locally Lipschitz of order ϵ,

$$\operatorname*{Sup}_{x_1, x_2, \delta_1, \delta_2}\ \operatorname*{Sup}_{|W - W'| \le \delta_\epsilon}\left|\tilde{h}\left(x_1, x_2, \delta_1, \delta_2, W; \boldsymbol{\xi}\right) - \tilde{h}\left(x_1, x_2, \delta_1, \delta_2, W'; \boldsymbol{\xi}\right)\right| \le M_\epsilon |W - W'|^\epsilon,$$ where constant $M_\epsilon < \infty$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

where constant $M_\epsilon < \infty$.

C7. $E\left(\left|\frac{b(V^*, V; \xi)}{S(X_1^*, X_2, W^*)}\right|^\lambda\right)^{1/\lambda} < \infty$ for some λ, $2 < \lambda \le \infty$.

C8. Bandwidth h satisfies (i) $0 \le h \to 0$, (ii) $nh/\log n \to \infty$, (iii) $n^{1/4}h \to 0$, and (iv) $(n/\log n)^{1-2/\lambda}h \to \infty$.

C9. The kernel **K** is bounded and of bounded variation.

Details on conditions C6–C8 can be found in Härdle, Janssen and Serfling (1988), and conditions C8 (i), (ii) and C9 can be found in Nolan and Pollard (1987). It can be easily seen that $h \propto n^d$ with $2/\lambda - 1 < d < -1/4$ for some $8/3 < \lambda \le \infty$ satisfies conditions C7 and C8. In general, small λ is preferred, which means d is preferred to be close to $-1/4$. Therefore, for the simulations and data analysis in this paper, we chose $d = -1/3$.

Theorem 1 *Suppose that Conditions C1–C5 hold for discrete W and that Conditions C1–C8 hold for continuous W. Then the solution of* $\mathbf{U}_n(\xi) = 0$, *denoted by* $\hat{\xi}_n$ *is a consistent estimator of* ξ_0.

The proof of Theorem 1 is treated separately for discrete W with finite levels and continuous W, but follows similar steps. We first show that $\boldsymbol{U}_n(\xi)$ converges to a deterministic function $\boldsymbol{u}(\xi)$ uniformly, then show that $\boldsymbol{u}(\xi)$ is monotone and has a unique root at ξ_0. Then consistency follows easily. Details are provided in Appendix B.

Theorem 2 *Suppose that Conditions C1–C5 hold for discrete W and that Conditions C1–C9 hold for continuous W. Then we have that* $n^{1/2}(\hat{\xi}_n - \xi_0)$ *converges in distribution to a normal random variable with mean zero and variance* $\boldsymbol{I}(\xi_0)^{-1}\boldsymbol{\Sigma}(\xi_0)\boldsymbol{I}(\xi_0)^{-1}$, *where* $\boldsymbol{I}(\xi_0) = 2E\left\{\Delta_1\Delta_2\dot{\beta}(X_1, X_2, W)^{\otimes 2}\right\}$ *and* $\boldsymbol{\Sigma}(\xi_0)$ *is the asymptotic variance of* $\mathbf{U}_n(\xi_0)$, *whose estimator is described in* equation (19) *for discrete W and* equation (20) *for continuous W in* Appendix C.

The asymptotic normality in Theorem 2 can be achieved by using Taylor expansion of $\boldsymbol{U}_n(\hat{\xi}_n)$ around ξ_0. Again the detailed calculation which centers on the linearization of $U_n(\xi_0) - u(\xi_0)$ is deferred to Appendix C. A variance estimator of $n^{1/2}(\hat{\xi}_n - \xi_0)$ can be obtained by estimating $\boldsymbol{I}(\xi_0)$ directly from the data with ξ_0 substituted by $\hat{\xi}_n$ and by evaluating the asymptotic expression of $\boldsymbol{\Sigma}(\xi_0)$ from the data with details given at the end of Sections C.1 and C.2 in Appendix C.

5 Simulations

5.1 Discrete covariate

We conduct simulations to assess the performance of the proposed method. We generate data from Clayton model and the piecewise constant cross-ratio model, which accommodate multiplicative covariate effect easily. Generating data from a bivariate distribution with an

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

arbitrary cross-ratio function is almost impossible because there is no corresponding closed form survival function in general. For simplicity, we assume W is a binary random variable from *Bernoulli* (0.5). We generate data for $\beta_0(t_1, t_2; \gamma) = \gamma_0 = 0.25$ and $\alpha = 0.5$. Marginally, both T_1 and T_2 follow unit exponential distribution. We first generate T_1, then for a given covariate value, T_2 is generated from the conditional distribution of T_2 given T_1 derived from the corresponding Clayton model. This setup is equivalent to generating data from two Clayton models with $\theta = e^{0.25}$ when $W = 0$ and $\theta = e^{0.75}$ when $W = 1$. The censoring times C_1 and C_2 both follow a uniform (0, 3) distribution, resulting in a marginal censoring rate of about 30%. The basis functions used for the estimation are 1, t_1 and t_2, though only the intercept term is needed in the true model. The results based on sample sizes of 400 and 800 are summarized in Table 1, where α is the true covariate effect and γ's are the true coefficients for the basis functions 1, t_1 and t_2 respectively. Simulation results based on 1000 replications show that our estimators work well. The model-based variance estimator also works well since the empirical coverage probabilities are all close to the 95% nominal value.

To mimic the cross-ratio results of the Tremin Trust data, we also simulate data using algorithm in Nan et al. (2006) with a binary covariate $W \sim Bernoulli(0.5)$ and $\alpha = 0.5$. For $W = 0$, the cross-ratio is piecewise constant over four intervals: $\theta = .9$ when $t_1 \in [0, .25)$, $\theta = 2.0$ when $t_1 \in [.25, .5)$, $\theta = 4.0$ when $t_1 \in [.5, .75)$, and $\theta = 1.5$ when $t_1 > .75$. For $W = 1$, the cross-ratio θ is equal to $0.9 \times e^{0.5}$, $2.0 \times e^{0.5}$, $4.0 \times e^{0.5}$ and $1.5 \times e^{0.5}$ in the above intervals. Marginally, both event times T_1 and T_2 follow unit exponential distribution. The censoring times C_1 and C_2 both follow a uniform (0, 2) distribution, resulting in a marginal censoring rate of about 40%. Note that we do not intend to simulate the marginal distributions of the Tremin data that satisfy the constraint $T_1 < T_2$ due to the technical challenges of generating ordered bivariate survival times with a piecewise constant cross-ratio. We use the following indicator functions as basis functions $I(t_1 < 0.25)$, $I(0.25 \le t_1 < 0.5)$, $I(0.5 \le t_1 < 0.75)$, $I(0.75 \le t_1)$ together with a linear covariate w, assuming the cutoffs are known, i.e.,

$$\beta(t_1, t_2, w; \boldsymbol{\xi}) = \gamma_1 I(t_1 < 0.25) + \gamma_2 I(0.25 \le t_1 < 0.5) + \gamma_3 I(0.5 \le t_1 < 0.75) + \gamma_4 I(0.75 \le t_1) + \alpha w .$$

The results in Table 2 show that our estimators as well as their model based variance estimators all work well.

5.2 Continuous covariate

Like the simulations for the discrete covariate, we simulate data with $W \sim unif(-0.5, 0.5)$ and $\alpha = 0.5$ assuming the same Clayton model and piecewise constant model for the baseline cross-ratio. For Clayton model, the censoring times C_1 and C_2 both follow a uniform (0, 3) distribution, resulting in a marginal censoring rate of about 30%. For the piecewise constant model, the censoring times C_1 and C_2 both follow a uniform (0, 2) distribution, resulting in a marginal censoring rate of about 40%. A challenging issue in the continuous covariate case is the search for an optimal bandwidth for the kernel smoothing. Unfortunately, standard procedures for finding the optimal bandwidth such as cross validation are not applicable because we lack the proper objective function to optimize.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Therefore we recommend using $h = range(x) \times n^{-1/3}$. Simulation results for sample sizes of 400 and 800 summarized in Tables 3 and 4 have shown our recommended bandwidth works well.

6 Data analysis

6.1 The Tremin study

The Tremin Trust data were collected as part of the Menstrual and Reproductive Health Study (Treloar et al. 1967). This longitudinal cohort study followed participants throughout their reproductive life span. It provides a unique opportunity to investigate the process of female reproductive aging and menopausal transition. The study sample consisted of white college students enrolled at the University of Minnesota. Data collection started in 1935 and enrolled a sample of 1,997 women over 4 years. Study participants were followed for up to 40 years. Each woman was asked to use menstrual diary cards to record the days when bleeding was experienced. Some covariate information (e.g., age at menarche) was available.

Nan et al. (2006) used a subset of the Tremin Trust data to study the age at onset of a 45-day cycle as the bleeding pattern change criteria for the early and late stages of menopausal transition. They estimated the cross-ratio as a piecewise constant function. Here we analyze the same subset that consisted of 562 women in the original study cohort who were age 25 or younger at enrollment, had information on age at menarche, and were still participating in the study at age 35 (which they used as the baseline age in their study). Both time to a marker event and time to menopause were subject to right-censoring in the Tremin Trust data. For each individual, the censoring time was the same for both events. A total of 193 (34%) women were observed to experience natural menopause, and a total of 357 (64%) women were observed to experience a 45-day cycle marker. The median age at menopause was 51.7 years, the median age at the 45-day cycle marker was 42.7 years and the median age at menarche is 12 years. One woman had her 45-day cycle time censored at the time of menopause and was excluded from our analysis. Note that for this data example, the cross-ratio is only well-defined in the region $T_1 < T_2$, where T_1 is the age at onset of a 45-day cycle and T_2 is the age at menopause.

To be able to compare the results with Nan et al. (2006) and for the ease of interpretation, we model the cross-ratio as a quadratic function of t_1 only, i.e. the age at onset of a 45-day cycle, based on the same data. Assuming a multiplicative effect of menarche on cross-ratio, we model the log cross-ratio as:

$$\beta(t_1, t_2, w; \boldsymbol{\xi}) = \gamma_0 + \gamma_1 t_1 + \gamma_2 t_1^2 + \alpha w, \quad (9)$$

where w is the age at menarche. For model (9), we further consider two functional forms for the age at menarche: an ordinal age covariate with five levels ("≤ 10" = 1, "11" = 2, "12" = 3, "13" = 4, "≥ 14" = 5) where its linear trend is of interest; and a nominal age covariate with the same five levels where level 3 is the reference group, which ignores the ordering of these five levels.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We compare the baseline cross-ratio results of (9) fitted at the median age at menarche (w = 12) with results in Nan et al. (2006). The general pattern of the estimated cross-ratio curvature is an open-down parabola, consistent with piecewise-constant result in Nan et al. (2006). However, age at menarche is not significant in any covariate model, although Table 5 does suggest the larger the age at menarche, the smaller the log cross-ratio and hence the weaker correlation between marker event and menopause. We also fit the age at menarche as a continuous covariate which was recorded as integers in the Tremin dataset. Although applying kernel smoothing would not be appropriate, using $h = 1 \approx 9 \times 562^{-1/3}$, where 9 is the range of the covariate and 562 is the sample size, we obtain the covariate estimate of to be −0.10 (s.e.=0.23), close to the linear trend estimation ($\alpha = -0.12$) reported in Table 5.

6.2 The Australian twin study revisited

In the analysis of the Australian twin study of appendicitis in Hu et al. (2011), it was found that monozygotic twins exhibited higher concordance rate than dizygotic twins. It is therefore of interest to quantify the disparity between the different types of twin pairs. Additionally, it is desirable to characterize the dependence between twin pairs when the effect of zygocity is controlled for. Analyses presented here are based on 1953 female twin pairs with available appendectomy information. The data comprised 1218 monozygotic twin pairs and 735 dizygotic twin pairs. Out of the monozygotic twin pairs, there are 144 pairs in which both twins were appendectomized, 304 pairs in which one twin underwent appendectomy and 770 pairs in which neither twin received the procedure. The corresponding numbers for the dizygotic twin pairs are 63, 208 and 464, respectively.

Since the order of twin one and twin two is arbitrary in the Australian Twin Study, we can take advantage of such symmetry to improve the estimation e ciency. Assuming a multiplicative effect of zygocity on cross-ratio, we model the log cross-ratio as:

$$\begin{aligned}\beta(t_1, t_2, w; \boldsymbol{\xi}) &= \gamma_0 + \gamma_1(t_1 + t_2) + \gamma_2\left(t_1^2 + t_2^2\right) + \gamma_3 t_1 t_2 + \gamma_4\left(t_1^2 t_2 + t_1 t_2^2\right) \\ &\quad + \gamma_5\left(t_1^3 + t_2^3\right) + \alpha w,\end{aligned} \tag{10}$$

where w is a binary variable that encodes monozygotic twins vs dizygotic twins.

Implementing our proposed estimating method, we obtain an estimator of α at 0.39 (95% CI: 0.08 – 0.70), suggesting a genetic component to the disease. So the cross-ratio of monozygotic twins is estimated to be 1.47 times higher than that of dizygotic twins.

7 Discussion

We have developed a novel method for estimating the covariate effect on cross-ratio where we model the covariate effect parametrically and the baseline log cross-ratio as a linear model of polynomial basis functions of time. When the covariate is discrete of a few levels, the proposed method is a simple extension of Hu et al. (2011). When the covariate is continuous, kernel smoothing is applied to the estimating equations developed for a discrete covariate. A key contribution of this paper is that we have established consistency and asymptotic normality of the regression coefficient estimate, facilitated by theorems in Härdle et al. (1988) and Nolan and Pollard (1987).

We have considered in this paper a single covariate function. But in some situations, multiple covariates are to be accounted for. For example, in an oncology clinical trial setting, investigators may be interested in the correlation between overall survival and progression free survival after adjusting for treatment and disease stage. The current method can easily accommodate discrete multiple covariates by simply recoding the combination of multiple covariates into a single discrete covariate. However, extension to multiple continuous covariates is less straightforward. We suggest using a multi-dimensional kernel function, for example a multivariate normal probability density function, to smooth the estimating equations. A modified version of C8 is required for the asymptotic properties to continue to hold. Let p be the dimension of the continuous covariate, then bandwidth h should satisfy (i) $0 \le h \to 0$, (ii) $nh^p / \log n \to \infty$, (iii) $n^{1/4} h^p \to 0$, and (iv) $(n/\log n)^{1-2/\lambda} h^p \to \infty$.

Following Hu, Lin and Nan (2014), the objective function (4) can be easily modified to accommodate left truncation. We leave the details to interested readers.

Appendix A:: definitions

We extend the notation used in Hu et al. (2011) to accommodate covariates. Define the following simplified notation:

$$\partial_1 F(t_1, t_2 | w) = \frac{\partial F(t_1, t_2 | w)}{\partial t_1}, \qquad \partial_1 G(t_1, t_2 | w) = \frac{\partial G(t_1, t_2 | w)}{\partial t_1},$$

$$\partial_2 F(t_1, t_2 | w) = \frac{\partial F(t_1, t_2 | w)}{\partial t_2}, \qquad \partial_2 G(t_1, t_2 | w) = \frac{\partial G(t_1, t_2 | w)}{\partial t_2},$$

$$\partial_{1,2} F(t_1, t_2 | w) = \frac{\partial^2 F(t_1, t_2 | w)}{\partial t_1 \partial t_2}, \qquad \partial_{1,2} G(t_1, t_2 | w) = \frac{\partial^2 G(t_1, t_2 | w)}{\partial t_1 \partial t_2},$$

where F and G denote the survival functions of (T_1, T_2) and (C_1, C_2) conditional on $W = w$, respectively. Then the conditional density function of $(X_1, X_2, \Delta_1, \Delta_2)$ given $W = w$ can be written as

$$
\begin{aligned}
&q\left(t_1, t_2, \delta_1, \delta_2 \middle| w\right) \\
&= \partial_{1,2} F\left(t_1, t_2 \middle| w\right)^{\delta_1 \delta_2} \left\{-\partial_1 F\left(t_1, t_2 \middle| w\right)\right\}^{\delta_1\left(1-\delta_2\right)} \left\{-\partial_2 F\left(t_1, t_2 \middle| w\right)\right\}^{\left(1-\delta_1\right)\delta_2} \\
&F\left(t_1, t_2 \middle| w\right)^{\left(1-\delta_1\right)\left(1-\delta_2\right)} \partial_{1,2} G\left(t_1, t_2 \middle| w\right)^{\left(1-\delta_1\right)\left(1-\delta_2\right)} \\
&\left\{-\partial_1 G\left(t_1, t_2 \middle| w\right)\right\}^{\left(1-\delta_1\right)\delta_2} \left\{-\partial_2 G\left(t_1, t_2 \middle| w\right)\right\}^{\delta_1\left(1-\delta_2\right)} G\left(t_1, t_2 \middle| w\right)^{\delta_1 \delta_2},
\end{aligned}
$$

and the joint density of $(X_1, X_2, \Delta_1, \Delta_2, W)$ is

$$
p\left(t_1, t_2, \delta_1, \delta_2, w\right) = q\left(t_1, t_2, \delta_1, \delta_2 \middle| w\right) f_W(w), \quad (11)
$$

where $f_W(w)$ denotes the distribution function of W.

For a discrete covariate W, we introduce the following notation:

$$
\begin{aligned}
&\tilde{g}^d\left(\Delta_2, X_1, X_2, W, \Delta_1^*, X_1^*, X_2^*, W^*; \boldsymbol{\xi}\right) \\
&= \frac{I(W = W^*)\Delta_1^* \Delta_2 \dot{\beta}\left(X_1^*, X_2, W^*\right) I\left(X_1 \geq X_1^*\right) I\left(X_2 \leq X_2^*\right) \theta\left(X_1^*, X_2, W^*, \boldsymbol{\xi}\right)}{S\left(X_1^*, X_2, W^*\right)}
\end{aligned}
$$

where

$$
S\left(t_1, t_2, w\right) = Pr\left(X_1 \geq t_1, X_2 \geq t_2 \middle| W = w\right) f_W(w). \quad (12)
$$

By fixing $\left(\Delta_1^*, X_1^*, X_2^*, W^*\right)$ at (δ_1, x_1, x_2, w), we also define

$$
\tilde{h}_Q^d\left(\delta_1, x_1, x_2, w; \boldsymbol{\xi}\right) = Q\tilde{g}^d\left(\Delta_2, X_1, X_2, W, \delta_1, x_1, x_2, w; \boldsymbol{\xi}\right).
$$

Similarly, fixing (Δ_2, X_1, X_2, W) at (δ_2, x_1, x_2, w), define

$$
\tilde{h}_P^d\left(\delta_2, x_1, x_2, w; \boldsymbol{\xi}\right) = P\tilde{g}^d\left(\delta_2, x_1, x_2, w, \Delta_1^*, X_1^*, X_2^*, W^*; \boldsymbol{\xi}\right).
$$

For a continuous covariate W, define $V = (X_1, X_2, \Delta_1, \Delta_2, W)$ and

$$
b\left(V_i, V_j; \boldsymbol{\xi}\right) = \Delta_{1i} \Delta_{2j} I\left(X_{1j} \geq X_{1i}\right) I\left(X_{2j} \leq X_{2i}\right) e^{\beta\left(X_{1i}, X_{2j}, W_i; \xi\right)} \times \dot{\beta}\left(X_{1i}, X_{2j}, W_i\right), \quad (13)
$$

$$
v\left(V_i, V_j; \boldsymbol{\xi}\right) = \frac{1}{n}\left[N\left(X_{1i}, X_{2j}, W_i\right) - \mathbf{K}_h(0) I\left(X_{2j} \leq X_{2i}\right)\left(1 - \theta\left(X_{1i}, X_{2j}, W_i; \boldsymbol{\xi}\right)\right)\right],
$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$g^{(n)}\left(V_i, V_j; \boldsymbol{\xi}\right) = \frac{\mathbf{K}_h\left(W_j - W_i\right)b\left(V_i, V_j; \boldsymbol{\xi}\right)}{v\left(V_i, V_j; \boldsymbol{\xi}\right)},$$

$$g_h^{(n)}\left(V_i, V_j; \boldsymbol{\xi}\right) = \frac{\mathbf{K}_h\left(W_j - W_i\right)b\left(V_i, V_j; \boldsymbol{\xi}\right)}{S_h\left(X_{1i}, X_{2j}, W_i\right)}$$

$$\tilde{g}^{(n)}\left(V_i, V_j; \boldsymbol{\xi}\right) = \frac{\mathbf{K}_h\left(W_j - W_i\right)b\left(V_i, V_j; \boldsymbol{\xi}\right)}{S\left(X_{1i}, X_{2j}, W_i\right)}$$

$$\tilde{h}\left(V_i; \boldsymbol{\xi}\right) = E_{X_{1j}, X_{2j}, \Delta_{1j}, \Delta_{2j} \middle| W_j = W_i, X_{1i}, X_{2i}, \Delta_{1i}, \Delta_{2i}}\left[\frac{b\left(V_i, V_j; \boldsymbol{\xi}\right)}{S\left(X_{1i}, X_{2j} \middle| W_i\right)}\right] \quad (14)$$

$$\tilde{h}^*\left(V_j; \boldsymbol{\xi}\right) = E_{X_{1i}, X_{2i}, \Delta_{1i}, \Delta_{2i} \middle| W_i = W_j, X_{1j}, X_{2j}, \Delta_{1j}, \Delta_{2j}}\left[\frac{b\left(V_i, V_j; \boldsymbol{\xi}\right)}{S\left(X_{1i}, X_{2j} \middle| W_j\right)}\right]$$

$$u^{(2)}(\boldsymbol{\xi}) = E_{X_{1i}, X_{2i}, \Delta_{1i}, \Delta_{2i}, W_i}\tilde{h}\left(V_i; \boldsymbol{\xi}\right),$$

where

$$S_h\left(t_1, t_2, w\right) = E\left[I\left(X_1 \geq t_1, X_2 \geq t_2\right)\mathbf{K}_h(W - w)\right].$$

Clearly, we have $S(t_1, t_2, w) = \lim_{h \downarrow 0} S_h(t_1, t_2, w)$.

Appendix B:: proof of Theorem 1

For consistency, we will first show that $\boldsymbol{U}_n^{(k)}(\boldsymbol{\xi})$ converges uniformly to $\boldsymbol{u}^{(k)}$, $k = 1, 2$, then show that $\boldsymbol{u}(\boldsymbol{\xi}) = \mathbf{0}$ has the unique solution at $\boldsymbol{\xi}_0$, and finally show the consistency of $\hat{\boldsymbol{\xi}}_n$ satisfying $\boldsymbol{U}_n\left(\hat{\boldsymbol{\xi}}_n\right) = \boldsymbol{0}$.

The uniform convergence of $\boldsymbol{U}_n^{(1)}(\boldsymbol{\xi})$ to $\boldsymbol{u}^{(1)}(\boldsymbol{\xi})$ remains the same for both discrete and continuous covariates. However, for a continuous covariate, $\boldsymbol{U}_n^{(2)}(\boldsymbol{\xi})$ involves the kernel function which is unbounded as the bandwidth goes to 0, so the proof for the uniform

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

convergence of $\boldsymbol{U}_n^{(2)}(\boldsymbol{\xi})$ to $\boldsymbol{u}^{(2)}(\boldsymbol{\xi})$ is treated separately for discrete W with finite levels and continuous W. When W is discrete with finite levels, the proof is similar to that provided in Hu et al. (2011). So we focus on continuous W.

First, let $\left(X_1^*, X_2^*, \Delta_1^*, \Delta_2^*, W^*\right)$ be an identical copy of $(X_1, X_2, \Delta_1, \Delta_2, W)$. Define the deterministic function $\boldsymbol{u}(\boldsymbol{\xi}) = \boldsymbol{u}^{(1)}(\boldsymbol{\xi}) - \boldsymbol{u}^{(2)}(\boldsymbol{\xi}) + \boldsymbol{u}^{(3)}(\boldsymbol{\xi}) - u^{(4)}(\boldsymbol{\xi})$, with

$$\boldsymbol{u}^{(1)}(\boldsymbol{\xi}) = \boldsymbol{u}^{(3)}(\boldsymbol{\xi}) = E\left\{\Delta_1 \Delta_2 \dot{\beta}\left(X_1, X_2, W\right)\right\},$$

$$\begin{aligned}\boldsymbol{u}^{(2)}(\boldsymbol{\xi}) &= \boldsymbol{u}^{(4)}(\boldsymbol{\xi}) \\ &= E\left\{\Delta_1^* \Delta_2 \dot{\beta}\left(X_1^*, X_2, W\right) \frac{I\left(X_1 \geq X_1^*\right) I\left(X_2 \leq X_2^*\right) \theta\left(X_1^*, X_2, W; \boldsymbol{\xi}\right)}{S\left(X_1^*, X_2 \middle| W^*\right)}\right\},\end{aligned}$$

where $S(x_1, x_2|w) = Pr(X_1 > x_1, X_2 > x_2| W = w)$.

Similar to Hu et al. (2011), we use $\mathbb{P}_n$ and $\mathbb{Q}_n$ to denote the empirical measures of n independent copies of $\left(X_1^*, X_2^*, \Delta_1^*, \Delta_2^*, W^*\right)$ and $(X_1, X_2, \Delta_1, \Delta_2, W)$ that follow the distributions P and Q, respectively, which make the double summations more tractable. For model (3), $\boldsymbol{U}_n^{(1)}(\boldsymbol{\xi}) = \mathbb{Q}_n \Delta_1 \Delta_2 \dot{\beta}\left(X_1, X_2, W\right)$ is free of $\boldsymbol{\xi}$, and $\dot{\beta}\left(X_1, X_2, W\right)$ is bounded from Conditions C1, C2 and C5. Hence by the law of large numbers, we have

$$\sup_{\xi}\left|\boldsymbol{U}_n^{(1)}(\boldsymbol{\xi}) - \boldsymbol{u}^{(1)}(\boldsymbol{\xi})\right| = \left|\left(\mathbb{Q}_n - Q\right)\Delta_1 \Delta_2 \dot{\beta}\left(X_1, X_2, W\right)\right| \to 0$$

either almost surely or in probability. Convergence in probability should be adequate here for the proof.

By Härdle, Janssen and Serfling (1988),

$$\begin{aligned}&\frac{1}{n} N\left(t_1, t_2, w\right) \\ &= \frac{1}{n}\sum_{k=1}^{n} I\left(X_{1k} \geq t_1, X_{2k} \geq t_2\right)\mathbf{K}_h\left(W_k - w\right) \\ &= \frac{\sum_{k=1}^{n} I\left(X_{1k} \geq t_1, X_{2k} \geq t_2\right)\mathbf{K}_h\left(W_k - w\right)}{\sum_{k=1}^{n}\mathbf{K}_h\left(W_k - w\right)} \times \frac{\sum_{k=1}^{n}\mathbf{K}_h\left(W_k - w\right)}{n} \\ &= E\left(I\left(X_1 \geq t_1, X_2 \geq t_2\right)\middle| W = w\right) f(w) + o_p(1) \\ &= S\left(t_1, t_2, w\right) + o_p(1).\end{aligned}$$

Also note that the difference between $g^{(n)}$ and $\tilde{g}^{(n)}$ is their denominators wherein we replace the denominator of $g^{(n)}$ by its limit. We then have the following:

$$
\begin{aligned}
&\sup_{\xi}\left|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} g^{(n)}\left(V_i, V_j; \boldsymbol{\xi}\right) - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \tilde{g}^{(n)}\left(V_i, V_j; \boldsymbol{\xi}\right)\right| \\
&\quad \le \frac{1}{n}\sum_{i=1}^{n}\sup_{\xi}\left|\frac{1}{n}\sum_{j=1}^{n} g^{(n)}\left(V_i, V_j; \boldsymbol{\xi}\right) - \frac{1}{n}\sum_{j=1}^{n} \tilde{g}^{(n)}\left(V_i, V_j; \boldsymbol{\xi}\right)\right| \\
&\quad = \frac{1}{n}\sum_{i=1}^{n}\sup_{\xi}\left|\frac{1}{n}\sum_{j=1}^{n} \frac{\mathbf{K}_h\left(W_j - W_i\right) b\left(V_i, V_j; \boldsymbol{\xi}\right)}{v\left(V_i, V_j; \boldsymbol{\xi}\right) S\left(X_{1i}, X_{2j}, W_i\right)} \times \left(v\left(V_i, V_j; \boldsymbol{\xi}\right) - S\left(X_{1i}, X_{2j}, W_i\right)\right)\right| \\
&\quad = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n}\mathbf{K}_h\left(W_j - W_i\right)\sup_{\xi}\left|\frac{b\left(V_i, V_j; \xi\right)}{v\left(V_i, V_j; \xi\right) S\left(X_{1i}, X_{2j}, W_i\right)} \times \left(v\left(V_i, V_j; \boldsymbol{\xi}\right) - S\left(X_{1i}, X_{2j}, W_i\right)\right)\right| \\
&\quad \le \frac{1}{n}\sum_{i=1}^{n}\frac{\frac{1}{n}\sum_{j=1}^{n}\mathbf{K}_h\left(W_j - W_i\right)}{\frac{1}{n}N\left(X_{1i}, X_{2j}, W_i\right) + o_p(1)}\sup_{\xi}\left|\frac{b\left(V_i, V_j; \boldsymbol{\xi}\right)}{S\left(X_{1i}, X_{2j}, W_i\right)}\right| \times \left(\sup\left|\left(n^{-1}N\left(X_{1i}, X_{2j}, W_i\right) - S\left(X_{1i}, X_{2j}, W_i\right)\right)\right|\right. \\
&+ \left.\sup_{\xi}\left|n^{-1}\mathbf{K}_h(0) I\left(X_{2j} \le X_{2i}\right)\left(1 - \theta\left(X_{1i}, X_{2j}, W_i; \xi\right)\right)\right|\right) \\
&\quad \le \frac{1}{n}\sum_{i=1}^{n} O_p(1) O_p(1)\left(\sup\left|\left(n^{-1}N\left(X_{1i}, X_{2j}, W_i\right) - S\left(X_{1i}, X_{2j}, W_i\right)\right)\right| + O_p\left((nh)^{-1}\right)\right) \\
&\quad \le O_p(1)\left(O_p\left(\max\left\{(nh/\log n)^{-1/2}, h^{\epsilon}\right\}\right) + O_p\left((nh)^{-1}\right)\right) \\
&\quad = o_p(1).
\end{aligned}
$$

In the last inequality, we used the result of strong uniform consistency for conditional functional estimators of Härdle, Janssen and Serfling (1988).

Next, we want to show that the difference between $\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{g}_{ij}^{(n)}(\boldsymbol{\xi})$ and $\frac{1}{n}\sum_{i=1}^{n}\tilde{h}\left(V_i; \boldsymbol{\xi}\right)$ is $o_p(1)$. Again using the result of Härdle, Janssen and Serfling (1988) in the following calculation, we have

$$
\begin{aligned}
&\sup_{\xi}\left|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{g}^{(n)}\left(V_i, V_j; \boldsymbol{\xi}\right) - \frac{1}{n}\sum_{i=1}^{n}\tilde{h}\left(V_i; \boldsymbol{\xi}\right)\right| \\
&\quad \le \frac{1}{n}\sum_{i=1}^{n}\sup_{\xi}\left|\frac{1}{n}\sum_{i=1}^{n}\tilde{g}^{(n)}\left(V_i, V_j; \boldsymbol{\xi}\right) - \tilde{h}\left(V_i; \boldsymbol{\xi}\right)\right| \\
&\quad \le \frac{1}{n}\sum_{i=1}^{n}\sup_{\boldsymbol{\xi}, V_i}\left|\frac{1}{n}\sum_{i=1}^{n}\tilde{g}^{(n)}\left(V_i, V_j; \boldsymbol{\xi}\right) - \tilde{h}\left(V_i; \boldsymbol{\xi}\right)\right| \\
&\quad = \sup_{\boldsymbol{\xi}, V_i}\left|\frac{1}{n}\sum_{i=1}^{n}\tilde{g}^{(n)}\left(V_i, V_j; \boldsymbol{\xi}\right) - \tilde{h}\left(V_i; \boldsymbol{\xi}\right)\right| \\
&\quad = O_p\left(\max\left\{(nh/\log n)^{-1/2}, h^{\epsilon}\right\}\right) \\
&\quad = o_p(1).
\end{aligned}
$$

Last, we want to show that the difference between $\frac{1}{n}\sum_{i=1}^{n}\tilde{h}\left(V_i; \boldsymbol{\xi}\right)$ and its deterministic limit $u^{(2)}(\boldsymbol{\xi})$ is $o_p(1)$ uniformly in $\boldsymbol{\xi}$. For model (3) under C1–C3, it is straightforward to see that

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

all the component functions of $b(V_i, V_j; \xi)$ are Donsker. Thus $b(V_i, V_j; \xi)$ is Donsker. Then by Theorem 2.10.2 in van der Vaart and Wellner (1996), $\tilde{h}(V_i; \xi)$ is also Donsker. Hence, $\tilde{h}(V_i; \xi)$ is Glivenko-Cantelli. We then have

$$\sup_{\xi}\left|\frac{1}{n}\sum_{i=1}^{n}\tilde{h}(V_i; \xi) - u^{(2)}(\xi)\right| = o_p(1).$$

Thus we have shown that $\boldsymbol{U}_n(\xi)$ converges uniformly to $\boldsymbol{u}(\xi)$ in probability. Following a similar calculation in Hu et al. (2011), we can also show that ξ_0 is the unique solution of $\boldsymbol{u}(\xi) = 0$. The consistency of $\hat{\xi}_n$ follows immediately.

Appendix C:: proof of Theorem 2

For asymptotic normality, the goal is to write $\boldsymbol{U}_n(\xi_0)$ as an average of n *i.i.d.* terms plus a $o_p(n^{-1/2})$ term. The technical difficulty arises when $\boldsymbol{U}_n(\xi_0)$ involves the kernel function which is unbounded as the bandwidth goes to 0, so that we can no longer rely on the properties of Donsker functions. Here, we briefly give the results of a discrete covariate and then focus mainly on the linearization of $\boldsymbol{U}_n(\xi_0)$ for a continuous covariate.

Define $\dot{\boldsymbol{U}}_n(\xi) \equiv d\boldsymbol{U}_n(\xi)/d\xi$. By Taylor expansion of $\boldsymbol{U}_n(\hat{\xi}_n)$ around ξ_0, we have

$$n^{1/2}(\hat{\xi}_n - \xi_0) = -\left\{\dot{\boldsymbol{U}}_n(\xi^*)\right\}^{-1} n^{1/2}\boldsymbol{U}_n(\xi_0), \quad (15)$$

where ξ^* lies between $\hat{\xi}_n$ and ξ_0. By a similar calculation as in the proof of Theorem 1 showing the uniform consistency of $\boldsymbol{U}_n(\xi)$, we can show that $\sup\left|\dot{\boldsymbol{U}}_n(\xi) - \dot{\boldsymbol{u}}(\xi)\right| = o_p(1)$, Thus by the consistency of $\hat{\xi}_n$, which implies the consistency of ξ^*, and the continuity of $\dot{\boldsymbol{u}}(\xi)$, we obtain $\dot{\boldsymbol{U}}_n(\xi^*) = \dot{\boldsymbol{u}}(\xi_0) + o_p(1)$ where $\dot{\boldsymbol{u}}(\xi_0) = -2E\left\{\Delta_1\Delta_2\dot{\beta}(X_1, X_2, W)^{\otimes 2}\right\} = -\boldsymbol{I}(\xi_0)$ is invertible by Condition C4. Hence based on the fact that continuity holds for the inverse operator, (15) can be written as

$$n^{1/2}(\hat{\xi}_n - \xi_0) = \left\{\boldsymbol{I}(\xi_0)^{-1} + o_p(1)\right\} n^{1/2}\boldsymbol{U}_n(\xi_0). \quad (16)$$

We now need to find the asymptotic representation of $n^{1/2}\boldsymbol{U}_n(\xi_0)$. We only check it for $\boldsymbol{U}_n^{(1)}(\xi_0) - \boldsymbol{U}_n^{(2)}(\xi_0)$. The calculation for $\boldsymbol{U}_n^{(3)}(\xi_0) - \boldsymbol{U}_n^{(4)}(\xi_0)$ is virtually identical and yields the same asymptotic representation.

It is easily seen that

$$n^{1/2}\left(\boldsymbol{U}_n^{(1)}(\xi_0) - \boldsymbol{u}^{(1)}(\xi_0)\right) = \mathbb{G}_n\left\{\Delta_1\Delta_2\dot{\beta}(X_1, X_2, W)\right\}, \quad (17)$$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

where $\mathbb{G}_n = n^{1/2}(\mathbb{P}_n - P)$. We then focus on $n^{1/2}\left(\boldsymbol{U}_n^{(2)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{(2)}(\boldsymbol{\xi}_0)\right)$, whose linearization differs vastly for a discrete covariate and a continuous covariate, largely because we could no longer rely on Donsker Theorem for a continuous covariate case when kernel functions are involved. Thus the two cases are treated separately in the proof.

C.1. Linearization $n^{1/2}\left(\boldsymbol{U}_n^{(2)}\left(V_i, V_j; \boldsymbol{\xi}_0\right) - \boldsymbol{u}^{(2)}(\boldsymbol{\xi}_0)\right)$ of for a discrete covariate

Following similar calculation as in Hu et al. (2011), we can show that

$$\begin{aligned} n^{1/2}&\left\{\boldsymbol{U}_n^{(2)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{(2)}(\boldsymbol{\xi}_0)\right\} \\ &= \mathbb{G}_n\Big\{\tilde{h}_Q^d(\Delta_1, X_1, X_2, W; \boldsymbol{\xi}_0) + \tilde{h}_P^d(\Delta_2, X_1, X_2, W; \boldsymbol{\xi}_0) \\ &\quad - \iint I(X_1 \ge x_1^*, X_2 \ge x_2, W = w^*) r\left(\delta_1, x_1, x_2, w, \delta_2^*, x_1^*, x_2^*, w^*\right) \\ &\quad dP\left(\delta_1^*, \delta_2^*, x_1^*, x_2^*, w^*\right) dQ\left(\delta_1, \delta_2, x_1, x_2, w\right)\Big\} + o_p(1), \end{aligned} \tag{18}$$

where

$$r\left(\delta_1, x_1, x_2, w, \delta_2^*, x_1^*, x_2^*, w^*\right) = \frac{I(w = w^*)\delta_1^*\delta_2\dot{\beta}\left(x_1^*, x_2, w^*\right) I\left(x_1 \ge x_1^*\right) I\left(x_2 \le x_2^*\right) e^{\beta\left(x_1^*, x_2, w^*; \xi_0\right)}}{\left\{S\left(x_1^*, x_2, w^*\right)\right\}^2}.$$

Then we obtain

$$\begin{aligned} n^{1/2}\boldsymbol{U}_n(\boldsymbol{\xi}_0) &= 2\mathbb{G}_n\Big\{\Delta_1\Delta_2\dot{\beta}(X_1, X_2, W) - \tilde{h}_Q^d(\Delta_1, X_1, X_2, W; \boldsymbol{\xi}_0) \\ &\quad - \tilde{h}_P^d(\Delta_2, X_1, X_2, W; \boldsymbol{\xi}_0) \\ &\quad + \iint I\left(X_1 \ge x_1^*, X_2 \ge x_2, W = w^*\right) \times r\left(\delta_1, x_1, x_2, w, \delta_2^*, x_1^*, x_2^*, w^*\right) \\ &\qquad dP\left(\delta_1^*, \delta_2^*, x_1^*, x_2^*, w^*\right) dQ\left(\delta_1, \delta_2, x_1, x_2, w\right)\Big\} + o_p(1) \\ &\to_d N\left(\boldsymbol{0}, \boldsymbol{\Sigma}(\boldsymbol{\xi}_0)\right). \end{aligned} \tag{19}$$

Thus from (16) we obtain the desired asymptotic distribution of $n^{1/2}\left(\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0\right)$.

Let $\mathbf{Z}(\boldsymbol{\xi}_0)$ denote the expression inside {} in (19) for a generic data point. It is clear that each $\mathbf{Z}_i(\boldsymbol{\xi}_0)$ is a function of i-th observation, hence $\mathbf{Z}_i(\boldsymbol{\xi}_0)$'s are *i.i.d.* Then under the regularity conditions we have the weak convergence in (19) with $\boldsymbol{\Sigma}(\boldsymbol{\xi}_0) = 4E\left\{\boldsymbol{Z}(\boldsymbol{\xi}_0)^{\otimes 2}\right\}$, where $\mathbb{G}_n\left\{\boldsymbol{Z}(\boldsymbol{\xi}_0)\right\} = n^{-1}\sum_{i=1}^{n}\boldsymbol{Z}_i(\boldsymbol{\xi}_0)$. We estimate the covariance matrix of $\mathbf{Z}(\boldsymbol{\xi}_0)$ by its sample covariance matrix with $\tilde{h}_Q^d(\Delta_1, X_1, X_2, W; \boldsymbol{\xi}_0)$, $\tilde{h}_P^d(\Delta_1, X_1, X_2, W; \boldsymbol{\xi}_0)$ and the double integral substituted by their sample averages, and ξ_0 replaced by $\hat{\xi}_n$. After the approximation/substitution, quantities are no longer *i.i.d.* However, it can be shown that $\mathbf{Z}(\xi_0)$ and its sample approximation belong to Glivenko-Cantelli class of functions, which leads to an asymptotically valid covariance estimator. Our simulation has shown the empirical variance and the variance estimates are very close.

C.2. Linearization of $n^{1/2}\left(\boldsymbol{U}_n^{(2)}\left(V_i, V_j; \boldsymbol{\xi}_0\right) - \boldsymbol{u}^{(2)}\left(\boldsymbol{\xi}_0\right)\right)$ for a continuous covariate

We focus on $n^{1/2}\left(\boldsymbol{U}_n^{(2)}\left(V_i, V_j; \boldsymbol{\xi}_0\right) - \boldsymbol{u}^{(2)}\left(\boldsymbol{\xi}_0\right)\right)$ with the following decomposition:

$$
\begin{aligned}
& n^{1/2}\left(\boldsymbol{U}_n^{(2)}\left(V_i, V_j; \boldsymbol{\xi}_0\right) - u^{(2)}\left(\boldsymbol{\xi}_0\right)\right) \\
& = n^{1/2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(g^{(n)}\left(V_i, V_j; \boldsymbol{\xi}_0\right) - g_h^{(n)}\left(V_i, V_j; \boldsymbol{\xi}_0\right)\right) \\
& \quad + n^{1/2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(g_h^{(n)}\left(V_i, V_j; \boldsymbol{\xi}_0\right) - \tilde{g}^{(n)}\left(V_i, V_j; \boldsymbol{\xi}_0\right)\right) \\
& \quad + \frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{j=1}^{n}\tilde{g}^{(n)}\left(V_i, V_j; \boldsymbol{\xi}_0\right) - \tilde{h}\left(V_i; \boldsymbol{\xi}_0\right)\right) \\
& \quad + \frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(\tilde{h}\left(V_i; \boldsymbol{\xi}_0\right) - \boldsymbol{u}^{(2)}\left(\boldsymbol{\xi}_0\right)\right) \\
& = -A - B + C + D.
\end{aligned}
$$

Now we will look at the four terms separately. Firstly, term D is a sum of *i.i.d.* items, and thus $D = \mathbb{G}_n\left(\tilde{h}\left(V, \boldsymbol{\xi}_0\right)\right)$.

Secondly, term C can be decomposed as follows:

$$
\begin{aligned}
C & = \frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{j=1}^{n}\tilde{g}^{(n)}\left(V_i, V_j; \boldsymbol{\xi}_0\right) - \tilde{h}\left(V_i, \boldsymbol{\xi}_0\right)\right) \\
& = \mathbb{G}_n\left(\mathbb{P}_n^*\tilde{g}^{(n)}\left(V, V^*, \boldsymbol{\xi}_0\right) - P^*\tilde{g}^{(n)}\left(V, V^*, \boldsymbol{\xi}_0\right)\right) \\
& \quad + \mathbb{G}_n\left(P^*\tilde{g}^{(n)}\left(V, V^*, \boldsymbol{\xi}_0\right) - \tilde{h}^*\left(V, \boldsymbol{\xi}_0\right)\right) \\
& \quad + \mathbb{G}_n\left(\tilde{h}^*\left(V, \boldsymbol{\xi}_0\right)\right) + n^{1/2}\mathbb{P}_n^*\left(P\tilde{g}^{(n)}\left(V, V^*, \boldsymbol{\xi}_0\right) - \tilde{h}\left(V^*, \boldsymbol{\xi}_0\right)\right) \\
& = C_1 + C_2 + C_3 + C_4.
\end{aligned}
$$

For the last equality of the above equation, we want to show that $C_1 = o_p(1)$, $C_2 = o_p(1)$ and $C_4 = o_p(1)$, so that $C = C_3 + o_p(1)$. First, by lemma A.2 of Ichimura (1993) $P\tilde{g}^{(n)}\left(V, V^*, \boldsymbol{\xi}_0\right) - \tilde{h}\left(V^*, \boldsymbol{\xi}_0\right) = O\left(h^2\right)$. Thus $C_4 = n^{1/2}O(h^2) = o_p(1)$ for h satisfying C8. Likewise, $P^*\tilde{g}^{(n)}\left(V, V^*, \boldsymbol{\xi}_0\right) - \tilde{h}^*\left(V, \boldsymbol{\xi}_0\right) = O\left(h^2\right)$, and therefore $C_2 = n^{1/2}O(h^2) = o_p(1)$ for h satisfying C8. Finally, we need to show that

$$
C_1 = \mathbb{G}_n\left(\left(\mathbb{P}_n^* - P^*\right)\mathbf{K}_h(W^* - W)\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)}{S\left(X_1^*, X_2, W^*; \boldsymbol{\xi}_0\right)}\right) = o_p(1).
$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

First, set

$$r_h(V, V^*) = \mathbf{K}\left(\frac{W^* - W}{h}\right)\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)}{S\left(X_1^*, X_2, W^*; \boldsymbol{\xi}_0\right)},$$

$$\tilde{r}_h(V, V^*) = r_h(V, V^*) - Pr_h(V, V^*) - P^* r_h(V, V^*) + PP^* r_h(V, V^*),$$

and

$$T_n\left(\tilde{r}_h\right) = \sum_{1 \le i \ne j \le n} \tilde{r}_h\left(V_i, V_j\right).$$

Then we have

$$\begin{aligned} C_1 &= h^{-1}\mathbb{G}_n\left(\left(\mathbb{P}_n^* - P^*\right)r_h(V, V^*)\right) \\ &= h^{-1}\sqrt{n}\left(\left(\mathbb{P}_n - P\right)\left(\mathbb{P}_n^* - P^*\right)r_h(V, V^*)\right) \\ &= h^{-1}\sqrt{n}\frac{1}{n^2}\left(T_n\left(\tilde{r}_h\right) + \sum_{i=1}^{n}\tilde{r}_h\left(V_i, V_i\right)\right) \\ &= \frac{1}{\sqrt{n}hn}T_n\left(\tilde{r}_h\right) + \frac{1}{nh\sqrt{n}}\sum_{i=1}^{n}\tilde{r}_h\left(V_i, V_i\right) \\ &= C_{11} + C_{12}. \end{aligned}$$

Applying the central limit theorem, it is easy to see that $C_{12} = o_p(1)$. To show that $C_{11} = o_p(1)$, we need the following definition and theorem from Nolan and Pollard (1987). We keep the same numbering for the definition and theorem as in the original paper for the ease of reference.

Definition 8. Call a class of functions $\mathcal{F}$ Euclidean for the envelop F if there exist constants A and V such that

$$N_1(\epsilon, Q, \mathscr{F}, F) \le A\epsilon^{-V}, \text{ for } 0 < \epsilon \le 1,$$

whenever $0 < QF < \infty$, where N_1 denotes the covering number with L^1 norm.

Theorem 9. Let F be a Euclidean class of P-degenerate functions with envelope 1. Let $W(n, x)$ *be a bounded weight function that is decreasing in both arguments and satisfies*

$$\sum_{n=1}^{\infty}\int_0^1 n^{-1}W(n, x)(1 + \log(1/x))dx < \infty.$$

If $v(\cdot)$ *is a function on F for which* $v(f) \ge \sup_x P|f(x, \cdot)|$, *then*

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$n^{-1}\left\| W\left(n, v(f)^{1/2}\right) T_n(f) \right\| \to 0.$$

In our case, each $\tilde{r}_h$ is P-degenerate; that is $P\tilde{r}_h(V, \cdot) = 0$. The class of all $\tilde{r}_h$ is a candidate for the above theorem. Following Nolan and Pollard (1987) page 795, it is easy to check that there exists a constant C for which

$$\sup_{x, y, h} \left| \tilde{r}_h(x, y) \right| \le C \text{ and } \sup_x P^* \left| \tilde{r}_h(x, \cdot) \right| \le C(1 \wedge h)$$

for all $h > 0$. We can rescale to make C equal to 1.

If kernel **K** is of bounded variation, e.g. standard normal density, then $\{\tilde{r}_h\}$ is a Euclidean class. For details of establishing Euclidean property in a particular class, please refer to Section 5 of Nolan and Pollard (1987).

Invoking Theorem 9 of Nolan and Pollard (1987), we obtain

$$n^{-1}\left\| W\left(n, v(f)^{1/2}\right) T_n(f) \right\| = o_p(1),$$

where $v(\tilde{r}_h) = 1 \wedge h$ and $W(n, x) = (1 + nx^{10})^{-1}$. Since W is bounded by 1 and

$$\int_0^1 W(n, x)(1 + \log(1/x))dx = O\left(n^{-1/10} \log n\right)$$

the conditions of Theorem 9 are satisfied.

Returning to the calculation for C_{11},

$$\begin{aligned} C_{11} &= \frac{1}{\sqrt{nh}n} T_n\left(\tilde{r}_h\right) \\ &\le \frac{1}{\sqrt{nh} W\left(n, v(f)^{1/2}\right)} \left\| n^{-1} W\left(n, v(f)^{1/2}\right) T_n\left(\tilde{r}_h\right) \right\| \\ &= \frac{1 + n(1 \wedge h)^5}{\sqrt{nh}} o_p(1) \\ &\le \frac{1 + nh^5}{\sqrt{nh}} o_p(1) \\ &= o_p(1) + \sqrt{n}h^4 o_p(1) \end{aligned}$$

Thus $C_{11} = o_p(1)$ for h satisfying C8. Then we obtain $C_1 = C_{11} + C_{12} = o_p(1)$ and thus $C = \mathbb{G}_n\left(\tilde{h}^*\left(V, \boldsymbol{\xi}_0\right)\right) + o_p(1)$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Thirdly, we want to show B is $o_p(1)$ and hence negligible. Now

$$B = n^{\frac{1}{2}} n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\mathbf{K}_h\left(W_j - W_i\right) b\left(V_i, V_j; \boldsymbol{\xi}_0\right)}{S\left(X_{1i}, X_{2j}, W_i\right) S_h\left(X_{1i}, X_{2j}, W_i\right)} \times \left(S_h\left(X_{1i}, X_{2j}, W_i\right) - S\left(X_{1i}, X_{2j}, W_i\right)\right)$$
$$= n^{\frac{1}{2}} n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\mathbf{K}_h\left(W_j - W_i\right) b\left(V_i, V_j; \boldsymbol{\xi}_0\right)}{S\left(X_{1i}, X_{2j}, W_i\right) S_h\left(X_{1i}, X_{2j}, W_i\right)} O\left(h^2\right).$$

The inner summation divided by n is bounded by the density of W at W_i times $O(h^2)$, which is seen from the following:

$$n^{-1} \sum_{j=1}^{n} \frac{\mathbf{K}_h\left(W_j - W_i\right) b\left(V_i, V_j; \boldsymbol{\xi}_0\right)}{S\left(X_{1i}, X_{2j}, W_i\right) S_h\left(X_{1i}, X_{2j}, W_i\right)} O\left(h^2\right)$$
$$= n^{-1} \sum_{j=1}^{n} \frac{\mathbf{K}_h\left(W_j - W_i\right) b\left(V_i, V_j; \boldsymbol{\xi}_0\right)}{S\left(X_{1i}, X_{2j}, W_i\right)\left(S\left(X_{1i}, X_{2j}, W_i\right) + o(1)\right)} O\left(h^2\right)$$
$$\lesssim O\left(h^2\right) n^{-1} \sum_{j=1}^{n} \mathbf{K}_h\left(W_j - W_i\right)$$
$$\approx f\left(W_i\right) O\left(h^2\right)$$
$$= O\left(h^2\right)$$

where "$\lesssim$" denotes "less than up to some constant coefficient". Therefore, we have $B = n^{1/2} O(h^2) = o_p(1)$ for h satisfying C8.

Lastly, term A can be decomposed as

$$
\begin{aligned}
A &= n^{\frac{1}{2}} n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\mathbf{K}_h\left(W_j - W_i\right) b\left(V_i, V_j; \boldsymbol{\xi}_0\right)}{v\left(V_i, V_j; \boldsymbol{\xi}_0\right) S_h\left(X_{1i}, X_{2j}, W_i\right)} \times \left(v\left(V_i, V_j; \boldsymbol{\xi}_0\right) - S_h\left(X_{1i}, X_{2j}, W_i\right)\right) \\
&= n^{\frac{1}{2}} n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\mathbf{K}_h\left(W_j - W_i\right) b\left(V_i, V_j; \boldsymbol{\xi}_0\right)}{v\left(V_i, V_j; \boldsymbol{\xi}_0\right) S_h\left(X_{1i}, X_{2j}, W_i\right)} \times \left(\frac{1}{n} N\left(X_{1i}, X_{2j}, W_i\right) - S_h\left(X_{1i}, X_{2j}, W_i\right)\right) \\
&\quad + n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\mathbf{K}_h\left(W_j - W_i\right) b\left(V_i, V_j; \boldsymbol{\xi}_0\right)}{v\left(V_i, V_j; \boldsymbol{\xi}_0\right) S_h\left(X_{1i}, X_{2j}, W_i\right)} \frac{\mathbf{K}_h(0)}{n^{\frac{1}{2}}} \times I\left(X_{2j} \le X_{2i}\right)\left(1 - \theta\left(X_{1i}, X_{2j}, W_i; \boldsymbol{\xi}_0\right)\right) \\
&= n^{\frac{1}{2}} n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\mathbf{K}_h\left(W_j - W_i\right) b\left(V_i, V_j; \boldsymbol{\xi}_0\right)}{S_h\left(X_{1i}, X_{2j}, W_i\right)^2} \times \left(\frac{1}{n} N\left(X_{1i}, X_{2j}, W_i\right) - S_h\left(X_{1i}, X_{2j}, W_i\right)\right) + o_p(1) \\
&= n^{\frac{1}{2}} n^{-3} \sum_{k=1}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\mathbf{K}_h\left(W_j - W_i\right) b\left(V_i, V_j; \boldsymbol{\xi}_0\right)}{S_h\left(X_{1i}, X_{2j}, W_i\right)^2} \times \left(I\left(X_{1k} \ge X_{1i}, X_{2k} \ge X_{2j}\right) \mathbf{K}_h\left(W_k - W_i\right) - S_h\left(X_{1i}, X_{2j}, W_i\right)\right) \\
&\quad + o_p(1) \\
&= n^{\frac{1}{2}} \Bigg(\mathbb{P}_n^{\dagger} \mathbb{P}_n^{*} \mathbf{K}_h\left(W^{\dagger} - W^{*}\right) \mathbb{P}_n \frac{\mathbf{K}_h(W - W^{*}) b\left(V^{*}, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^{*}, X_2, W^{*}\right)^2} \times I\left(X_1^{\dagger} \ge X_1^{*}, X_2^{\dagger} \ge X_2\right) \\
&\quad - P^{\dagger} \mathbb{P}_n^{*} \mathbf{K}_h\left(W^{\dagger} - W^{*}\right) \mathbb{P}_n \frac{\mathbf{K}_h(W - W^{*}) b\left(V^{*}, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^{*}, X_2, W^{*}\right)^2} \times I\left(X_1^{\dagger} \ge X_1^{*}, X_2^{\dagger} \ge X_2\right) \\
&\quad + P^{\dagger} \mathbb{P}_n^{*} \mathbf{K}_h\left(W^{\dagger} - W^{*}\right) \mathbb{P}_n \frac{\mathbf{K}_h(W - W^{*}) b\left(V^{*}, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^{*}, X_2, W^{*}\right)^2} \times I\left(X_1^{\dagger} \ge X_1^{*}, X_2^{\dagger} \ge X_2\right) \\
&\quad - \mathbb{P}_n^{*} \mathbb{P}_n \frac{\mathbf{K}_h(W - W^{*}) b\left(V^{*}, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^{*}, X_2, W^{*}\right)} \Bigg) + o_p(1) \\
&= \mathbb{G}_n^{\dagger} \left(\mathbb{P}_n^{*} \mathbf{K}_h\left(W^{*} - W^{\dagger}\right) \mathbb{P}_n \mathbf{K}_h(W - W^{*}) \frac{b\left(V^{*}, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^{*}, X_2, W^{*}\right)^2} \times I\left(X_1^{\dagger} \ge X_1^{*}, X_2^{\dagger} \ge X_2\right) \right) \\
&\quad + n^{\frac{1}{2}} P^{\dagger} \mathbb{P}_n^{*} \mathbf{K}_h\left(W^{\dagger} - W^{*}\right) \mathbb{P}_n \frac{\mathbf{K}_h(W - W^{*}) b\left(V^{*}, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^{*}, X_2, W^{*}\right)^2} \times I\left(X_1^{\dagger} \ge X_1^{*}, X_2^{\dagger} \ge X_2\right) \\
&\quad - n^{\frac{1}{2}} \mathbb{P}_n^{*} \mathbb{P}_n \frac{\mathbf{K}_h(W - W^{*}) b\left(V^{*}, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^{*}, X_2, W^{*}\right)} + o_p(1) \\
&= A_1 + A_2 - A_3 + o_p(1)
\end{aligned}
$$

Term A_1 can be further decomposed as

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$
\begin{aligned}
A_1 &= \mathbb{G}_n^\dagger\left(\mathbb{P}_n^*\mathbf{K}_h\left(W^* - W^\dagger\right)\mathbb{P}_n\mathbf{K}_h(W - W^*)\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^*, X_2, W^*\right)^2} \times I\left(X_1^\dagger \geq X_1^*, X_2^\dagger \geq X_2\right)\right) \\
&= G_n^\dagger\left(\mathbb{P}_n^*\mathbf{K}_h\left(W^* - W^\dagger\right)\mathbb{P}_n\mathbf{K}_h(W - W^*)\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^*, X_2, W^*\right)^2} \times I\left(X_1^\dagger \geq X_1^*, X_2^\dagger \geq X_2\right)\right. \\
&\qquad \left. - P^*\mathbf{K}_h\left(W^* - W^\dagger\right)P\mathbf{K}_h(W - W^*)\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^*, X_2, W^*\right)^2} \times I\left(X_1^\dagger \geq X_1^*, X_2^\dagger \geq X_2\right)\right) \\
&\quad + \mathbb{G}_n^\dagger\left(P^*\mathbf{K}_h\left(W^* - W^\dagger\right)P\mathbf{K}_h(W - W^*)\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^*, X_2, W^*\right)^2} \times I\left(X_1^\dagger \geq X_1^*, X_2^\dagger \geq X_2\right)\right. \\
&\qquad \left. - E_{V^* \mid W^* = W^\dagger} E_{V \mid W = W^*}\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right) f(W^*)}{S_h\left(X_1^*, X_2, W^*\right)^2} \times I\left(X_1^\dagger \geq X_1^*, X_2^\dagger \geq X_2\right) f\left(W^\dagger\right)\right) \\
&\quad + \mathbb{G}_n^\dagger\left(E_{V^*}\middle| W^* = W^\dagger E_V \middle| W = W^* \frac{b\left(V^*, V; \boldsymbol{\xi}_0\right) f(W^*)}{S_h\left(X_1^*, X_2, W^*\right)^2} \times I\left(X_1^\dagger \geq X_1^*, X_2^\dagger \geq X_2\right) f\left(W^\dagger\right)\right) \\
&= A_{11} + A_{12} + A_{13}.
\end{aligned}
$$

We will show that $A_{12} = o_p(1)$ and $A_{11} = o_p(1)$ separately. First of all,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript

$$A_{12} = \mathbb{G}_n^{\dagger}\Bigg(P^*\mathbf{K}_h\left(W^* - W^{\dagger}\right)P\mathbf{K}_h(W - W^*)\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^*, X_2, W^*\right)^2} \times I\left(X_1^{\dagger} \geq X_1^*, X_2^{\dagger} \geq X_2\right)$$

$$-E_{V^*|W^* = W^{\dagger}}E_{V|W = W^*}\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)f(W^*)}{S_h\left(X_1^*, X_2, W^*\right)^2} \times I\left(X_1^{\dagger} \geq X_1^*, X_2^{\dagger} \geq X_2\right)f\left(W^{\dagger}\right)\Bigg)$$

$$= \mathbb{G}_n^{\dagger}\Bigg(P^*\mathbf{K}_h\left(W^* - W^{\dagger}\right)P\mathbf{K}_h(W - W^*)\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^*, X_2, W^*\right)^2} \times I\left(X_1^{\dagger} \geq X_1^*, X_2^{\dagger} \geq X_2\right)$$

$$-P^*\mathbf{K}_h\left(W^* - W^{\dagger}\right)E_{V|W = W^*}\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)f(W^*)}{S_h\left(X_1^*, X_2, W^*\right)^2} \times I\left(X_1^{\dagger} \geq X_1^*, X_2^{\dagger} \geq X_2\right)$$

$$+P^*\mathbf{K}_h\left(W^* - W^{\dagger}\right)E_{V|W = W^*}\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)f(W^*)}{S_h\left(X_1^*, X_2, W^*\right)^2} \times I\left(X_1^{\dagger} \geq X_1^*, X_2^{\dagger} \geq X_2\right)$$

$$-E_{V^*|W^* = W^{+}}E_{V|W = W^*}\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)f(W^*)}{S_h\left(X_1^*, X_2, W^*\right)^2} \times I\left(X_1^{\dagger} \geq X_1^*, X_2^{\dagger} \geq X_2\right)f\left(W^{\dagger}\right)\Bigg)$$

$$= \mathbb{G}_n^{\dagger}\Bigg(P^*\mathbf{K}_h\left(W^* - W^{\dagger}\right)\Bigg\{P\mathbf{K}_h(W - W^*)\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^*, X_2, W^*\right)^2} \times I\left(X_1^{\dagger} \geq X_1^*, X_2^{\dagger} \geq X_2\right)$$

$$-E_{V|W = W^*}\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)f(W^*)}{S_h\left(X_1^*, X_2, W^*\right)^2}I\left(X_1^{\dagger} \geq X_1^*, X_2^{\dagger} \geq X_2\right)\Bigg\}$$

$$+P^*\mathbf{K}_h\left(W^* - W^{\dagger}\right)E_{V|W = W^*}\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)f(W^*)}{S_h\left(X_1^*, X_2, W^*\right)^2} \times I\left(X_1^{\dagger} \geq X_1^*, X_2^{\dagger} \geq X_2\right)$$

$$-E_{V^*|W^* = W^{\dagger}}\Big|E_{V|W = W^*}\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)f(W^*)}{S_h\left(X_1^*, X_2, W^*\right)^2}I\left(X_1^{\dagger} \geq X_1^*, X_2^{\dagger} \geq X_2\right)f\left(W^{\dagger}\right)\Bigg).$$

Note that by Lemma A.2 of Ichimura (1993),

$$P\mathbf{K}_h(W - W^*)\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^*, X_2, W^*\right)^2}I\left(X_1^{\dagger} \geq X_1^*, X_2^{\dagger} \geq X_2\right)$$

$$= E_{V|W = W^*}\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)f(W^*)}{S_h\left(X_1^*, X_2, W^*\right)^2}I\left(X_1^{\dagger} \geq X_1^*, X_2^{\dagger} \geq X_2\right) + O\left(h^2\right),$$

and

$$P^*\mathbf{K}_h\left(W^* - W^{\dagger}\right)E_{V|W = W^*}\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)f(W^*)}{S_h\left(X_1^*, X_2, W^*\right)^2}I\left(X_1^{\dagger} \geq X_1^*, X_2^{\dagger} \geq X_2\right)$$

$$= E_{V^*|W^* = W^{\dagger}}E_{V|W = W^*}\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)f(W^*)}{S_h\left(X_1^*, X_2, W^*\right)^2}I\left(X_1^{\dagger} \geq X_1^*, X_2^{\dagger} \geq X_2\right)f\left(W^{\dagger}\right)\Bigg)$$

$$+O\left(h^2\right).$$

So term $A_{12} = n^{1/2}\, O(h^2) = o_p(1)$ for h satisfying C8.

To show term $A_{11} = o_p(1)$, first for fixed $V^\dagger$ set

$$m_h\left(V, V^*, V^\dagger\right) = h^{-1}\mathbf{K}\left(\frac{W^* - W^\dagger}{h}\right)\mathbf{K}\left(\frac{W - W^*}{h}\right)\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^*, X_2, W^*\right)^2} \times I\left(X_1^\dagger \ge X_1^*, X_2^\dagger \ge X_2\right),$$

$$\widetilde{m}_h\left(V, V^*, V^\dagger\right) = m_h\left(V, V^*, V^\dagger\right) - Pm_h\left(V, V^*, V^\dagger\right) - P^*m_h\left(V, V^*, V^\dagger\right) + PP^*m_h\left(V, V^*, V^\dagger\right).$$

Then term A_{11} can be decomposed into:

$$\mathbb{G}_n^\dagger\left(h^{-1}\mathbb{P}_n^*\mathbb{P}_n m_h - h^{-1}P^*Pm_h\right) = \mathbb{G}_n^\dagger\left(h^{-1}\mathbb{P}_n^*\mathbb{P}_n\widetilde{m}_h + \left(\mathbb{P}_n^* - P^*\right)h^{-1}Pm_h + \left(\mathbb{P}_n - P\right)h^{-1}P^*m_h\right).$$

Note that $\mathbb{P}_n^*\mathbb{P}_n\widetilde{m}_h$ is again a U-process. Using a proof similar to the one that shows $C_1 = o_p(1)$, we have $\mathbb{G}_n^\dagger\left(h^{-1}\mathbb{P}_n^*\mathbb{P}_n\widetilde{m}_h\right) = o_p(1)$, $\mathbb{G}_n^\dagger\left(\mathbb{P}_n^* - P^*\right)h^{-1}Pm_h = o_p(1)$, and $\mathbb{G}_n^\dagger\left(\mathbb{P}_n - P\right)h^{-1}P^*m_h = o_p(1)$ for h satisfying C8. Thus,

$$A_1 = A_{13} + o_p(1) = \mathbb{G}_n^\dagger\left(E_{V^*|W^* = W^\dagger}E_{V|W = W^*}\frac{b\left(V^*, V; \boldsymbol{\xi}_0\right)f(W^*)}{S_h\left(X_1^*, X_2, W^*\right)^2} \times I\left(X_1^\dagger \ge X_1^*, X_2^\dagger \ge X_2\right)f\left(W^\dagger\right)\right) + o_p(1).$$

Now focusing on $X_1^\dagger$, $X_2^\dagger$, $W^\dagger$ and their probability measure $P^\dagger$, we have

$$\begin{aligned}
A_2 &= P^\dagger\mathbb{P}_n^*\mathrm{K}_h\left(W^\dagger - W^*\right)\mathbb{P}_n\frac{\mathbf{K}_h(W - W^*)b\left(V^*, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^*, X_2, W^*\right)^2} \times I\left(X_1^\dagger \ge X_1^*, X_2^\dagger \ge X_2\right) \\
&= \mathbb{P}_n^*\mathbb{P}_n\frac{\mathbf{K}_h(W - W^*)b\left(V^*, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^*, X_2, W^*\right)^2} \times \left(S_h\left(X_1^*, X_2, W^*\right) + O\left(h^2\right)\right) \\
&= \mathbb{P}_n^*\mathbb{P}_n\frac{\mathbf{K}_h(W - W^*)b\left(V^*, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^*, X_2, W^*\right)} \\
&\quad + \mathbb{P}_n^*\mathbb{P}_n\frac{\mathbf{K}_h(W - W^*)b\left(V^*, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^*, X_2, W^*\right)^2}O\left(h^2\right) \\
&= \mathbb{P}_n^*\mathbb{P}_n\frac{\mathbf{K}_h(W - W^*)b\left(V^*, V; \boldsymbol{\xi}_0\right)}{S_h\left(X_1^*, X_2, W^*\right)} + O_p(1)O\left(h^2\right) \\
&= A_3 + o_p(1).
\end{aligned}$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Putting everything together, we obtain

$$\begin{aligned} n^{1/2}\boldsymbol{U}_n(\boldsymbol{\xi}_0) &= n^{1/2}\left\{\boldsymbol{U}_n(\boldsymbol{\xi}_0) - \boldsymbol{u}(\boldsymbol{\xi}_0)\right\} \\ &= n^{1/2}\left\{\boldsymbol{U}_n^{(1)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{(1)}(\boldsymbol{\xi}_0)\right\} - n^{1/2}\left\{\boldsymbol{U}_n^{(2)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{(2)}(\boldsymbol{\xi}_0)\right\} \\ &\quad + n^{1/2}\left\{\boldsymbol{U}_n^{(3)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{(3)}(\boldsymbol{\xi}_0)\right\} - n^{1/2}\left\{\boldsymbol{U}_n^{(4)}(\boldsymbol{\xi}_0) - \boldsymbol{u}^{(4)}(\boldsymbol{\xi}_0)\right\} \\ &= 2\mathbb{G}_n\Big\{\Delta_1\Delta_2\dot{\beta}(V) - \tilde{h}^{\star}(V;\boldsymbol{\xi}_0) - \tilde{h}(V;\boldsymbol{\xi}_0) \\ &\qquad + E_{V^*|W^* = W^\dagger} E_{V|W = W^*}\frac{b(V^*, V;\boldsymbol{\xi}_0)f(W^*)}{S_h(X_1^*, X_2, W^*)^2} \\ &\qquad \times I\left(X_1^\dagger \ge X_1^*, X_2^\dagger \ge X_2\right)f\left(W^\dagger\right)\Big\} + o_p(1) \\ &\rightarrow_d N(\boldsymbol{0}, \boldsymbol{\Sigma}(\boldsymbol{\xi}_0)). \end{aligned} \tag{20}$$

Thus from (16) we obtain the desired asymptotic distribution of $n^{1/2}(\hat{\xi}_n - \xi_0)$. The estimator of $\Sigma(\xi 0)$ can be obtained similarly to the case of a discrete covariate, with the conditional expectations evaluated using kernel smoothing.

References

Clayton DG (1978) A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65:141–151

Cox DR (1972) Regression models and life-tables (with discussion) J R Stat Soc Ser B 34:187–220

Duffy DL, Martin NG, Matthews JD (1990) Appendectomy in Australian twins. Am J Hum Genet 47:590–592. [PubMed: 2248678]

Fan J, Prentice RL (2002) Covariate-adjusted dependence estimation on a finite bivariate failure time region. Stat Sin 12:689–705

Härdle P, Janssen P, Serfling R (1988) Strong uniform consistency rates for estimators of conditional functionals. Ann Statist 16:1428–1449

Hu T, Lin X, Nan B (2014) Cross-ratio estimation for bivariate failure times with left truncation. Lifetime Data Anal 20:23–34 [PubMed: 23700275]

Hu T, Nan B, Lin X, Robins J (2011) Time-dependent cross-ratio estimation for bivariate failure times. Biometrika 98:341–354 [PubMed: 22822258]

Ichimura H (1993) Semiparametric least square (SLS) and weighted SLS estimation of single-index models. J Econometrics 58:71–120

Lisabeth LD, Harlow SD, Gillespie B, Lin X, and Sowers MF (2004) Staging Reproductive Aging: A Comparison of Proposed Bleeding Criteria for the Menopausal Transition. Menopause 11:186–197 [PubMed: 15021449]

Nan B, Lin X, Lisabeth LD, Harlow S (2006) Piecewise constant cross-ratio estimation for association of age at a marker event and age at menopause. J Am Stat Assoc 101:65–77

Nolan D, Pollard D (1987) U-process: rates of convergence. Ann Statist 15:780–799

Oakes D (1982) A model for association in bivariate survival data. J R Stat Soc Ser B 44:414–422

Oakes D (1986) Semiparametric inference in a model for association in bivariate survival data. Biometrika 73:353–361

Oakes D (1989) Bivariate survival models induced by frailties. J Am Stat Assoc 84:487–493

Shih J, Louis TA (1995) Inference on the association parameter in copula models for bivariate survival data. Biometrics 51:1384–1399 [PubMed: 8589230]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Treloar AE, Boynton RE, Behn BG, and Brown BW (1967) Variation of the Human Menstrual Cycle Through Reproductive Life. Int J Fertil 12:77–126 [PubMed: 5419031]

van der Vaart A and Wellner JA (1996) Weak convergence and empirical processes. Springer, New York

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1:

Cross-ratio regression for a discrete covariate with α = 0.5 and constant baseline cross-ratio with γ_0 = 0.25. $\hat{\alpha}$ and $\hat{\gamma}$, point estimate average; *E.SE*, the empirical standard error; *M.SE*, the average of the model based standard error estimates; *M.CP*, the 95% coverage probability.

	n=400				n=800			
α	$\hat{\alpha}$	*E.SE*	*M.SE*	*M.CP*	$\hat{\alpha}$	*E.SE*	*M.SE*	*M.CP*
0.50	0.51	0.19	0.18	95%	0.50	0.12	0.13	96%
γ	$\hat{\gamma}$	*E.SE*	*M.SE*	*M.CP*	$\hat{\gamma}$	*E.SE*	*M.SE*	*M.CP*
0.25	0.24	0.19	0.19	96%	0.25	0.13	0.13	96%
0	0.02	0.23	0.22	96%	0.01	0.15	0.15	96%
0	0.02	0.22	0.22	95%	0.01	0.14	0.15	95%

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2:

Cross-ratio regression for a discrete covariate with $\alpha = 0.5$ and the piecewise constant baseline cross-ratio. $\hat{\alpha}$ and $\hat{\gamma}$, point estimate average; *E.SE*, the empirical standard error; *M.SE*, the average of the model based standard error estimates; *M.CP*, the 95% coverage probability.

	n=400				n=800			
α	$\hat{\alpha}$	*E.SE*	*M.SE*	*M.CP*	$\hat{\alpha}$	*E.SE*	*M.SE*	*M.CP*
0.50	0.50	0.21	0.21	95%	0.50	0.14	0.14	96%
γ	$\hat{\gamma}$	*E.SE*	*M.SE*	*M.CP*	$\hat{\gamma}$	*E.SE*	*M.SE*	*M.CP*
−0.11	−0.10	0.18	0.18	95%	−0.10	0.13	0.13	94%
0.69	0.72	0.20	0.19	94%	0.70	0.13	0.13	95%
1.39	1.41	0.23	0.24	96%	1.41	0.16	0.16	95%
0.41	0.41	0.26	0.25	94%	0.42	0.17	0.17	94%

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3:

Cross-ratio regression for continuous covariate with $\alpha = 0.5$ and constant baseline cross-ratio with $\gamma_0 = 0.25$. $\hat{\alpha}$ and $\hat{\gamma}$, point estimate average; *E.SE*, the empirical standard error; *M.SE*, the average of the model based standard error estimates; *M.CP*, the 95% coverage probability.

	n=400				n=800			
α	$\hat{\alpha}$	*E.SE*	*M.SE*	*M.CP*	$\hat{\alpha}$	*E.SE*	*M.SE*	*M.CP*
0.50	0.49	0.30	0.30	95%	0.47	0.20	0.21	95%
γ	$\hat{\gamma}$	*E.SE*	*M.SE*	*M.CP*	$\hat{\gamma}$	*E.SE*	*M.SE*	*M.CP*
0.25	0.24	0.17	0.17	94%	0.25	0.11	0.12	95%
0	0.02	0.23	0.22	95%	0.01	0.15	0.15	95%
0	0.02	0.22	0.22	96%	0.00	0.14	0.15	95%

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 4:

Cross-ratio regression for continuous covariate with α = 0.5 and piecewise constant baseline cross-ratio. $\hat{\alpha}$ and $\hat{\gamma}$, point estimate average; *E.SE*, the empirical standard error; *M.SE*, the average of the model based standard error estimates; *M.CP*, the 95% coverage probability.

	n=400				n=800			
α	$\hat{\alpha}$	*E.SE*	*M.SE*	*M.CP*	$\hat{\alpha}$	*E.SE*	*M.SE*	*M.CP*
0.50	0.46	0.34	0.35	96%	0.46	0.24	0.25	94%
γ	$\hat{\gamma}$	*E.SE*	*M.SE*	*M.CP*	$\hat{\gamma}$	*E.SE*	*M.SE*	*M.CP*
−0.11	−0.10	0.15	0.15	95%	−0.10	0.11	0.10	94%
0.69	0.71	0.17	0.17	95%	0.70	0.12	0.12	96%
1.39	1.40	0.21	0.23	97%	1.40	0.15	0.16	97%
0.41	0.41	0.23	0.23	95%	0.41	0.16	0.16	96%

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Table 5:

Estimate of covariate effect and standard error of the effect of age at menarche in Tremin data. We consider two functional forms for the age at menarche: nominal (nominal covariate with 5 levels (≤10 = 1, 11 = 2, 12 = 3, 13 = 4, ≥ 14 = 5)) and linear trend (ordinal covariate with linear trend effect with the same 5 levels).

***w*: age at menarche**		$\hat{\alpha}$	**se**
	1: ≤ 10	0.10	0.59
	2: 11	0.15	0.52
Nominal	3: 12	−	−
	4: 13	0.03	0.54
	5: ≥ 14	−0.36	0.52
Linear trend		−0.12	0.17