Randomly censored partially linear single-index models

Xuewen Lu\textsuperscript{a, *}, Tsung-Lin Cheng\textsuperscript{b}

\textsuperscript{a} Department of Mathematics and Statistics, University of Calgary, Calgary, AB, Canada T2N 1N4
\textsuperscript{b} Department of Mathematics, National Changhua University of Education, Taiwan, ROC

Received 20 January 2006
Available online 16 January 2007

Abstract

This paper proposes a method for estimation of a class of partially linear single-index models with randomly censored samples. The method provides a flexible way for modelling the association between a response and a set of predictor variables when the response variable is randomly censored. It presents a technique for “dimension reduction” in semiparametric censored regression models and generalizes the existing accelerated failure-time models for survival analysis. The estimation procedure involves three stages: first, transform the censored data into synthetic data or pseudo-responses unbiasedly; second, obtain quasi-likelihood estimates of the regression coefficients in both linear and single-index components by an iteratively algorithm; finally, estimate the unknown nonparametric regression function using techniques for univariate censored nonparametric regression. The estimators for the regression coefficients are shown to be jointly root-\(n\) consistent and asymptotically normal. In addition, the estimator for the unknown regression function is a local linear kernel regression estimator and can be estimated with the same efficiency as all the parameters are known. Monte Carlo simulations are conducted to illustrate the proposed methodology.

© 2006 Elsevier Inc. All rights reserved.

AMS 2000 subject classification: 62N01; 62G08; 62E20

Keywords: Accelerated failure-time model; Asymptotic normality; Kernel smoothing; Local linear fit; Partially linear single-index model; Quasi likelihood; Random censoring; Synthetic data

1. Introduction

In studying the relationship between a response and a set of predictor variables or regressors, the mean response variable is often assumed to be a linear regression function of the regressors.
recent years, many models have been developed in studying high-dimensional data by nonpara-
metric or semiparametric regression models. To overcome the difficulties caused by the “curse of dimensionality” in smoothing, one of the approaches is to use single-index models or partially linear single-index models. For example, Härdle and Stoker [12], Powell et al. [28], Newey and Stoker [26] and Ichimura [15] investigated single-index models; Carroll et al. [4] and Xia and Härdle [35] studied generalized partially linear single-index models; Xia et al. [36] explored an extended version of Carroll et al.’s [4] models.

The aforementioned approaches are used to model the relationship between the response and the predictor variables when the data are fully observable. For the censored data, however, these techniques cannot be directly applied. A large number of estimators exist for parametric and semiparametric censored regression models. They mainly focus on the case where the censoring variable $C$ is random (which is a model adopted in many medical applications), or constant (which is a model adopted in many economic applications). To name a few, Buckley and James [3], Koul et al. [16], Lai et al. [17], Powell [27], Duncan [7], Fernandez [9], Horowitz [14], Powell et al. [28], Ritov [29], Ichimura [15], Lewbel [20], Buchinsky and Hahn [2] and Heuchenne and Van Keilegom [13], among others. Most of these models either assume a parametric regression form, or assume that the error distribution is parametric.

Not many estimators exist for censored nonparametric regression models. Under random censoring, Fan and Gijbels [8] proposed a censored nonparametric regression estimator based on a class of unbiased data transformations. While they considered only a univariate regressor, though they proclaimed their ideas hold for the case of two or more regressors. From a different motivation, Van Keilegom and Akritas [32] obtained the uniform consistency of the estimators for the unknown regression function and the heteroscedastic scale function and their derivatives. In the situation where there are multi-regressors, Wang and Zheng [33] and Liang and Zhou [23] investigated asymptotic properties in a semiparametric partial linear model. Li et al. [22] found ways of reducing the dimensionality of the regressor using the sliced inverse regression. Singh and Lu [30] studied censored nonparametric additive regression models based on some special data transformations. Recently, Lu and Burke [24] proposed a method called “censored average derivative estimation (CADE)” for studying the estimation of the unknown multiple regression function; GØrgens [11] developed semiparametric kernel-based estimators of risk-specific hazard functions for competing risk data; Lu et al. [25] investigated a class of partially linear single-index proportional hazards models for survival data. Under fixed censoring and truncation, Lewbel and Linton [21] proposed a novel technique of estimating nonparametric regression function and its derivatives in two stages. Recently, Chen et al. [6] considered identification and estimation of a nonparametric location-scale model under fixed censoring. Estimators under random censoring and estimators under fixed censoring cannot interchange, they are consistent only under different censoring schemes; moreover, their focuses are also different. The former focuses on the medical applications, while the latter focuses on the economic applications. The estimation of parametric and nonparametric effects of high-dimensional semiparametric regression models under random censoring can be applied to both economics and medical science.

In survival analysis (see [18]) and duration analysis (see [34]), an alternative model to the proportional hazards model or the multiplicative hazards model is the accelerated failure-time model. In contrast to the partially linear single-index proportional hazards model proposed by Lu et al. [25] for censored data, we consider the following randomly censored partially linear single-index model which is a class of accelerated failure-time models without the specification of the distribution function of the response variable,
defined by
\[ Y = \beta_0^T V + \lambda_0(\mathbf{z}_0^T X) + \sigma(V, X) \epsilon \quad \text{with} \quad \| \mathbf{z}_0 \| = 1, \]  

(1)

where \( Y \) is the survival time or some time-to-event outcome (usually on the log scale), \( V \) and \( X \) are the associated regressor \( q \) and \( p \) vectors, respectively, \( X \in \mathcal{X} \), \( \mathcal{X} \) is a compact subset of \( \mathbb{R}^p \), \( \mathbf{z}_0 \) and \( \beta_0 \) are regression coefficient parameter vectors, \( a^T \) denotes the transpose of a column vector \( a \), \( \lambda_0(\cdot) \) is a smooth function with an unspecified functional form, and \( \sigma(\cdot, \cdot) \) is the conditional variance representing the possible heteroscedacity, \( \| \cdot \| \) denotes the Euclidean norm. The constraint \( \| \mathbf{z}_0 \| = 1 \) on the single-index coefficient parameters is required for parameter identifiability. Assume that \( (V, X) \) and \( \epsilon \) are independent, \( E(\epsilon) = 0 \) and \( \text{Var}(\epsilon) = 1 \). Let \( C \) be the random censoring time associated with the survival time \( Y \). Assume \( C \) is independent of \( (V, X, Y) \). Denote \( Z = \min(Y, C) \) and \( \Delta = I(Y \leq C) \). The observations are \( \{(V_i, X_i, Z_i, \Delta_i) : i = 1, \ldots, n\} \), which is a random sample from the population \( (V, X, Z, \Delta) \).

Model (1) is an extended version of the generalized partially single-index model studied by Carroll et al. [4] to survival analysis, it also generalizes the existing censored linear regression models, censored nonparametric regression models, censored partially linear models and censored single-index models in the framework of accelerated failure-time models considered by many authors, for example, Koul et al. [16], Zhou [37], Lai et al. [17], Fan and Gijbels [8], Van Keilegom and Akritas [32], Wang and Zheng [33], Liang and Zhou [23], and Lu and Burke [24], among others. Our motivations come from all these censored regression models with many covariates, a part of them may have nonlinear effects on the response variable. In this case, the traditional linear models or kernel smoothing methods fail to incorporate both linear and nonlinear covariate effects. On the other hand, when a large amount of covariates have nonlinear effects, the multivariate kernel smooth suffers from the “curse of dimensionality”. The covariate effects in model (1) are addressed in a semiparametric fashion, which offers better flexibility in modelling the relationship between the failure time and the covariates than the existing models. Hence, model (1) is worthy of a full investigation. The main focus of this article is the estimation of parameter vectors \( \mathbf{z}_0 \) and \( \beta_0 \) under random censoring. Another objective is, when data are subject to censoring, to estimate the nonparametric regression function \( \lambda_0(u) \).

For application of the partially linear single-index model (1), a practical issue arises: Which covariates go into the \( V \) vector and which ones go into the \( X \) vector? There are a few strategies that may be applied for the division of the available covariates. The first is to utilize subject-matter knowledge related to the data collection experiment and the underlying physical mechanism. In model (1), the \( X \) vector serves primarily the role of dimension reduction, while the \( V \) vector may contain the major covariates of interest for the study. From this perspective, the selection of \( V \) and \( X \) can be readily made from the context of the study itself. For example, in a clinical study, the treatment effect of medicine is of interest and is coded as a categorical variable, it should be included in the \( V \) vector; other covariates such as patients’ age and blood pressure may be included in the \( X \) vector. The second is to carry out some simple analysis for covariates, which enables us to determine which covariates should be in the \( V \) or \( X \) vector. For example, for each covariate, we perform a simple regression analysis based on kernel smoothing such as in univariate nonparametric regression or partially linear models. If the fitted curve appears to be linear or approximately linear, we then assign this covariate into \( V \), otherwise, we assign it into \( X \). Examples for this with complete data can be found in Xia and Härdle [35]. In general, it is always helpful to consider any other strategies that would give rise to a more sensible and interpretable model.
Our model is analogous to that of Carroll et al. [4] except that the link function is the identity and the response $Y$ is randomly censored. We show that results similar to those of Carroll et al. [4] hold for censored data. In addition, observing that the optimal one-dimensional nonparametric rate of convergence is achievable in the estimation of $\lambda_0(u)$.

The plan of the paper is as follows: in Section 2, we will introduce our estimation method for model (1). Our main results will be presented in Section 3. In Section 4, we will report Monte Carlo simulation results. Some concluding remarks will be addressed in Section 5. The detailed proof of the main theorem is shown in Appendix A.

2. The procedure of estimation

This section presents a quasi-likelihood estimation procedure, which is implemented in an iterative minimization algorithm.

Let $\theta = (\alpha, \beta)$ be the vector of model parameters. If data are fully observed, i.e., $Z \equiv Y$, the quasi-likelihood estimators of $\theta_0 = (\alpha_0, \beta_0)$ and $\lambda_0$ are the minimizers of the following quasi log-likelihood function of a sample $\{(V_i, X_i, Y_i, \Delta_i \equiv 1), i = 1, \ldots, n\}$:

$$\ell_n(\theta, \lambda) = \sum_{i=1}^{n} \left[ Y_i - \{\beta^T V_i + \lambda(z^T X_i)\}^2 \right] \text{ with } \|z\| = 1.$$  (3)

This model is a special case of the generalized partially linear single-index models studied by Carroll et al. [4]. For the censored data, a difficulty for estimation arises due to censoring. Another difficulty common to single-index models in minimizing (3) is the involvement of the nonparametric function $\lambda$. To overcome these two difficulties, our solution is to use the synthetic data as well as the local linear fit. Our estimation procedure consists of the following steps. In the subsequent sections, we will discuss and explain each step in detail.

Transformation of the data: Let $F$ and $G$ be the distribution functions of $Y$ and $C$, respectively. That is, $F(x) = P(Y \leq x)$, $G(x) = P(C \leq x)$. Denote $\tau_F = \inf\{t: F(t) = 1\}$ and $\tau_G = \inf\{t: G(t) = 1\}$. We suppose $\tau_F \leq \tau_G$ throughout this paper. When $G(\cdot)$ is unknown, assume $1 - \hat{G}(\cdot)$ is an estimator of the survival function of random censoring variable $C$, for example, the Kaplan–Meier estimator. We construct the following synthetic data or pseudo-responses

$$Z_i\hat{G} = (1 + \phi)L_i\hat{G} - \phi K_i\hat{G},$$  (4)

where $L_i\hat{G} = \int_{-\infty}^{\infty} I[Z_i > s] / (1 - \hat{G}(s)) - I[s < 0]) ds$, $K_i\hat{G} = Z_i\Delta_i / (1 - \hat{G}(Z_i))$, $\phi$ is a tuning parameter which controls the weights put on the censored or uncensored observations, $1 - \hat{G}(\cdot)$ is the left-continuous version of the Kaplan–Meier estimator defined by

$$1 - \hat{G}(t) = \prod_{i=1}^{n} \left[ \frac{n - i}{n - i + 1} \right] I[Z_{(i)} \leq t, \Delta_{(i)} = 0].$$

$Z_{(1)} \leq Z_{(2)} \leq \cdots \leq Z_{(n)}$ are the order statistics of the Z-sample, and $\Delta_{(i)}$ is the $\Delta$ associated with $Z_{(i)}$, $i = 1, 2, \ldots, n$. We replace the observed data $\{(V_i, X_i, Z_i, \Delta_i)\}$ by $\{(V_i, X_i, Z_i\hat{G})\}$. Note
Component we apply the local linear fit to a quasi log-likelihood iteratively to estimate both parametric \( E \) that are unbiased. This class of transformations is introduced by Fan and Gijbels [8], \( \phi = -1 \) and \( \phi = 0 \) give the Koul, Susarla, and Van Ryzin [16] transformation \( K_{iG} \) (abbreviated as the KSV transformation) and the Leurgans [19] transformation \( L_{iG} \), respectively. An appropriate choice of \( \phi \) reduces the variability of the transformed data. Fan and Gijbels [8] recommend the following choice of \( \phi \):

\[
\hat{\phi} = \min_{\{i: \Delta_i = 1\}} \frac{\int_{-\infty}^{\infty} (I[Z_i \geq s]/(1 - \hat{G}(s-)) - I[s < 0]) ds - Z_i}{\int_{-\infty}^{\infty} (I[Z_i \geq s]/(1 - \hat{G}(s-)) - I[s < 0]) ds},
\]

which is the largest \( \phi \) such that the transformed response \( Z_{iG} \geq Z_i \) for the uncensored response, we will use it in our implementations. For ease of the technical proofs, we assume \( Y_i \geq 0 \) and \( C_i \geq 0 \) from now on; thus the integration in \( L_{iG} \) starts at 0 rather than \( -\infty \) such that \( L_{iG} = \int_{0}^{\infty} I[Z_i \geq s]/(1 - \hat{G}(s-)) ds \). The general case can be dealt with similarly at a cost of some more conditions on the left tail of the distributions.

**Application of the local quasi log-likelihood method:** After obtaining the transformed data, we apply the local linear fit to a quasi log-likelihood iteratively to estimate both parametric component \( \theta_0 \) and nonparametric component \( \lambda_0 \). Two well-known merits of the local linear fit are the reduction of the bias for the estimation of the nonparametric function and the avoidance of boundary effects. Suppose that \( \lambda(\cdot) \) is continuously differentiable. Then in a neighborhood of a fixed point \( u \), we can write \( \lambda'(v) \approx a_0 + a_1(v - u) \), where \( a_0 = \lambda(u) \) and \( a_1 = \lambda'(u) \). This is called the local linear fit.

Let \( W(\cdot) \) be a kernel. With a given bandwidth \( b \) and a given parameter vector \( \theta \), one can obtain local estimators \( \hat{a}_0 \equiv \widehat{a}_0(u; b, \theta) \) and \( \hat{a}_1 \equiv \widehat{a}_1(u; b, \theta) \) by minimizing the following local quasi log-likelihood

\[
\ell_n(a_0, a_1) = \sum_{i=1}^{n} \left[ Z_{iG} - \{b^T V_i + a_0 + a_1(x_i^T X_i - u)\} \right]^2 W_b(x_i^T X_i - u), \tag{5}
\]

where \( W_b(\cdot) = b^{-1} W(\cdot/b) \), and \( u \) is a fixed real number.

When the true parameter vector \( \theta_0 \) is unknown, to obtain estimators for model (1), we need to iteratively update the estimations of the nonparametric component \( \lambda_0(\cdot) \) and the parametric component \( \theta_0 = (z_0, \beta_0) \). Specifically, our iterative algorithm consists of the following steps:

**Step 1:** Treat the pseudo-responses \( Z_{iG} \) as complete data and apply an available method, such as Xia and Härdle’s [35] the minimum average (conditional) variance estimation (MAVE) method for the partially linear single-index models, to obtain initial estimates \( \tilde{z} \) and \( \tilde{\theta} \) of \( z_0 \) and \( \beta_0 \), respectively, with the restriction \( \|\tilde{z}\| = 1 \), and set \( \tilde{\theta} = (\tilde{z}, \tilde{\beta}) \).

**Step 2:** Find \( \hat{\lambda}(u; b, \tilde{\theta}) = \hat{a}_0 \) as a function of \( u \) by maximizing the local quasi log-likelihood (5) with respect to \( a_0 \) and \( a_1 \) with fixed \( \tilde{\theta} = \tilde{\theta} \) and a suitable bandwidth \( b \) (see bandwidth conditions given in Theorems 1 and 3).

**Step 3:** Update \( \tilde{\theta} \) by minimizing

\[
\sum_{i=1}^{n} \left[ Z_{iG} - \{b^T V_i + \hat{\lambda}(x_i^T X_i; b, \tilde{\theta})\} \right]^2 \tag{6}
\]

with respect to \( \theta = (z, \beta) \) and under \( \|z\| = 1 \).
Step 4: Iterate Steps 2 and 3 until convergence is achieved. Denote the final estimate of \( \theta_0 \) by \( \hat{\theta} = (\hat{z}, \hat{\beta}) \) and the final estimate of \( \lambda_0(u) \) by \( \hat{\lambda}(u; \hat{\theta}) = \hat{a}_0 \), where \( (\hat{a}_0, \hat{a}_1) \) is obtained by solving (5). At this final step, we take \( b \) to be an estimate of the bandwidth that is optimal for the estimation of \( \lambda_0(\cdot) \) when \( \theta_0 = (z_0, \beta_0) \) is known.

3. Asymptotic distribution theory for the estimators

The following conditions will be used to establish the asymptotic normality results of the quasi-likelihood estimators given in Theorems 1–3. Condition A is necessary for Theorem 1. For Theorems 2 and 3, both Conditions A and B are required.

**Condition A:**

(i) The kernel \( W \) is a symmetric density function on \([-1, 1]\), and satisfies uniform Lipschitz condition of order 1 on \( R \).

(ii) The random vectors \( V \) and \( X \) are bounded.

(iii) The marginal density \( f(u) \) of \( U = z_0^T X \) is positive, and has a continuous second derivative on its compact support \( D \subset R \).

(iv) The random vector \( X \) has a compact support \( \mathcal{X} \subset \mathbb{R}^p \), \( D_{z_0} \) is an open interval containing \( \{x^T X : \|x\| = 1, x \in \mathcal{X}\} \). The second derivative of \( \lambda_0(u) \) exists, is continuous and bounded on \( D_{z_0} \).

(v) The functions \( E[X|U = u] \) and \( E[V|U = u] \) are twice differentiable in \( u \in D \), and their second derivatives satisfy Lipschitz condition of order 1 on \( \mathcal{X} \). On the boundaries, the continuity and differentiability mean left or right continuity and differentiability.

(vi) There exists \( \kappa > 0 \) such that \( (1 - G)^{-1}(t) \leq C_G t^{-\kappa} \) for some constant \( C_G \), when \( t \to 0^+ \), where \( (1 - G)^{-1} \) denotes the inverse function of \( 1 - G \). There also exists \( 0 < \gamma < 1/(\kappa + 1) \) such that \( E[|Y||1 - G(|Y|)|^{2/\gamma}] < \infty \).

(vii) Functions \( E[|Y| |U = u] \), \( E[|Y|/(1 - G(Y))|U = u] \), \( E[Y^2/(1 - G(Y))|U = u] \), \( E[|Y|/(1 - G(Y))^2|U = u] \) and \( E[Y^2/(1 - G(Y))^2|U = u] \) have continuous derivatives on \( D \).

(viii) For a given \( \hat{\lambda} \), assume that \( \hat{z} - z_0 = O_P(n^{-1/2}) \) and \( \hat{\beta} - \beta_0 = O_P(n^{-1/2}) \) in (6), i.e. the initial estimates are in a \( \sqrt{n} \)-neighborhood of the true parameter values in probability, respectively.

(ix) For \( K_G = Z\Delta/(1 - G(Z-)) \), \( L_G = \int_0^\infty I[Z \geq s]/(1 - G(s-)) \, ds \), \( Z_G = (1 + \phi)L_G - \phi K_G \),

\[
\Psi = \begin{pmatrix} X \lambda_0'(U) \\ V \end{pmatrix}, \quad H = \Psi - E(\Psi|U), \quad \epsilon_G = Z_G - \{\beta_0^T V + \lambda_0(z_0^T X)\},
\]

both \( Q = E[HH^\otimes 2] \) and \( \Omega = E[(H\epsilon_G) \otimes 2] \) are positively definite, where \( a^{\otimes 2} = aa^T \) for a column vector \( a \).

**Condition B:**

(i) \( G \) is continuous.

(ii) When \( \tau_F < \tau_G \), \( \limsup_{t \to \tau_F} \left( G(1 - F(s)) dG(s) \right)^{1-\eta} \to < \infty \), for some \( \frac{1}{2} < \eta < \frac{1}{2} \).

(iii) When \( \tau_F = \tau_G \), for some \( 0 < \zeta < 1 \), \( (1 - G(t))^\zeta = O((1 - F(t-))) \) as \( t \to \tau_F \).
(iv) For $J(G, s, Z) = \int_{s}^{Z} \{1 - G(t -)\}^{-1} dt \mathbb{I}[s < Z]$, let $H(s) = \frac{E[H((1 + \phi)J(G, s, Z) - \phi K_G)\mathbb{I}[s < Z]]}{(1 - G(s))(1 - F(s -))}$, then
\[
\Xi(\tau_F) = \int_{0}^{\tau_F} \{H(s)\}^{2}(1 - F(s -)) dG(s) < \infty,
\]
\[
\Omega - \Xi(\tau_F) \text{ is positively definite.}
\]
(v) For every $\varepsilon > 0$, there exists $y(\varepsilon) < \tau_F$ such that
\[
\int_{s \in \mathbb{R}^{p+q}} \int_{y=y(\varepsilon)}^{\tau_F} \frac{\|s\|_y dF(s, y)}{(1 - G(y))(1 - F(y))^{1/2}} < \varepsilon,
\]
where $F(s, y) = \Pr(H \leq s, Y \leq y)$, both $H$ and $s$ are vectors.

Condition A is used for the case when $G$ is known. Conditions A and B are required for the case when $G$ is unknown and estimated from the Kaplan–Meier estimator. More specifically, Conditions A(i)–(viii) are used to give an asymptotic representation for the parametric component estimator, Condition A(ix) is for the asymptotic normality of the estimator. Conditions A(i)–(viii) and B(i)–(iii) are used to obtain an asymptotic representation for the parametric component estimator when $G$ is unknown and replaced by its estimator, while Conditions B(iv) and (v) are required to get the asymptotic normality of the estimator in this case. Under Conditions B(i)–(iii), Chen and Lo [5] and Gu and Lai [10] have shown that $\sup_{t \leq \tau_F} |\hat{G}(t) - G(t)| = O_p(n^{-\frac{1}{2}})$ and $\sup_{t \leq \tau_F} |\hat{G}(t) - G(t)| = O_p((n/ \log \log n)^{-1/2})$, respectively, in the cases of $\tau_F < \tau_G$ and $\tau_F = \tau_G$. From these conditions, we see that the rate of convergence of $\hat{G}$ is faster than that of ordinary nonparametric regression estimators, which is $O_p(n^{-2/5})$ when the optimal bandwidth is selected.

First, we assume that $G$ is given in the iterative algorithm, and we obtain the following result.

**Theorem 1.** Under Conditions A and the following conditions on the bandwidth: $nb^4 \to 0$ and $nb^3 = O(\log n)$, as $n \to \infty$, hold. Then, the estimator $\hat{\Theta} = (\hat{\alpha}, \hat{\beta})$ from the iterative algorithm satisfies
\[
n^{1/2} \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} \overset{D}{\to} N(0, Q^{-1}\Omega Q^{-1}),
\]
where $Q$ and $\Omega$ are defined in Condition A(ix), “$\overset{D}{\to}$” denotes convergence in distribution.

When the censoring distribution $G$ is unknown, we replace $G$ by the Kaplan–Meier estimator $\hat{G}$ in all terms associated with $G$ in Theorem 1. The effect of replacing $G$ by $\hat{G}$ is that it produces extra terms in the asymptotic representation of the parametric component estimator. These are studied in the Appendix, but we state our results first.

**Theorem 2.** Under Conditions A and B, we have
\[
n^{1/2} \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} \overset{D}{\to} N(0, Q^{-1}(\Omega - \Xi(\tau_F))Q^{-1}),
\]
where $\Xi(\tau_F)$ is defined in Condition B(iv).
Remark 1. When there is no censoring, i.e., \( 1 - G(\cdot) = 1 \), then \( \mathbb{E}(\tau_F) = 0 \), the asymptotic variance becomes \( Q^{-1}\Omega Q^{-1} \), where \( \Omega = E[(H\sigma e)^{\otimes 2}] \). Further, if \( \sigma \) is independent of \( (V, X) \), then \( Q^{-1}\Omega Q^{-1} = Q^{-1}\sigma^2 = [E[H^{\otimes 2}]]^{-1}\sigma^2 \), the result coincides with Theorem 5 for the normal response in Carroll et al. [4], the estimator is shown to achieve semiparametric efficiency.

Remark 2. The limiting variance and standard error of the estimator involve \( Q \), \( \mathbb{E}(\tau_F) \) and \( \Omega \). These quantities can be consistently estimated by their corresponding empirical versions as follows:

\[
\hat{Q}_n = \frac{1}{n} \sum_{i=1}^{n} [\hat{\Psi}_i - \hat{E}(\Psi|U_i)]^{\otimes 2},
\]

\[
\hat{\Omega}_n = \frac{1}{n} \sum_{i=1}^{n} [(\hat{\Psi}_i - \hat{E}(\Psi|U_i))[Z_i\hat{G} - (\hat{\beta}^T V_i + \hat{\lambda}(\hat{U}_i))]^{\otimes 2},
\]

\[
\hat{H}_n(s) = \frac{n^{-1}\sum_{i=1}^{n} [(\hat{\Psi}_i - \hat{E}(\Psi|U_i))((1 + \phi)J(\hat{G}, s, Z_i) - \phi K_i\hat{G})I[s < Z_i]]}{(1 - \hat{G}(s))(1 - \hat{F}(s))},
\]

\[
\hat{\lambda}_n(\tau_F) = \frac{1}{n} \sum_{i=1}^{n} (1 - \Delta_i)\{\hat{H}_n(Z_i)\}^{\otimes 2},
\]

where \( \hat{U}_i = \hat{\Delta}^T X_i, 1 - \hat{F}(s-) = n^{-1}\sum_{i=1}^{n} I[Z_i \geq s]/(1 - \hat{G}(s)) \), \( \hat{\Psi}_i = (X_i^T \hat{\lambda}_0(U_i), Z_i^T \hat{\lambda}_0(U_i))^T \) and \( \hat{E}(\Psi|U_i) = \sum_{j=1}^{n} \hat{\Psi}_j K((\hat{U}_j - \hat{U}_i)/h)/\sum_{j=1}^{n} K((\hat{U}_j - \hat{U}_i)/h) \) for some kernel function \( K(\cdot) \) and bandwidth \( h \).

An estimation procedure for the nonparametric component is suggested in the final step of the iterative algorithm given in Section 2. Using the results given by Fan and Gijbels [8] for univariate censored nonparametric regression, under some regularity conditions, we can obtain a consistent estimator \( \hat{\lambda}(u, \hat{\beta}) = \hat{\lambda}(u, \hat{\beta}) \) of \( \lambda_0 \). Moreover, in some cases, we have shown that the estimator is asymptotically normal.

Theorem 3. Let \( f(\cdot) \) be the density function of \( U = X^T \beta_0 \). If \( b = O(n^{-1/5}) \) and \( W(\cdot) \) has third-order continuous derivatives and its third-order derivative is bounded on \( D \), then under the conditions given in Theorem 2, conditioned on the covariates \{\( U_1, \ldots, U_n \)\}, for any interior point \( u \in D \),

\[
\sqrt{nb}(\hat{\lambda}(u; \hat{\beta}) - \lambda_0(u) - \lambda_0''(u)c_W b^2/2) \overset{D}{\Rightarrow} N(0, d_W\sigma^2(u)), \quad (7)
\]

where \( \sigma^2(u) = \text{Var}((Z_G - \beta_0^T V)|U = u) \), \( c_W = \int_{-\infty}^{+\infty} v^2 W(v) dv \), and \( d_W = \int_{-\infty}^{+\infty} W^2(v) dv \).

When \( \beta_0 \) and \( \beta_0 \) are known, we can easily prove the asymptotic normality of \( \hat{\lambda}(u; \beta_0) \) using the results in Fan and Gijbels [8] for univariate censored nonparametric regression. Therefore, to prove Theorem 3, it suffices to show that \( \hat{\lambda}(u; \hat{\beta}) - \hat{\lambda}(u; \beta_0) = O_p(n^{-1/2}) \). This is implied by the root-\( n \) consistency of \( (\hat{\beta}, \hat{\beta}) \) and the assumptions for the bandwidth \( b \) and the kernel function \( W(\cdot) \). Note that in this estimation, we can take an optimal bandwidth \( b = O(n^{-1/5}) \) to estimate \( \lambda_0(\cdot) \), so that the rate of convergence of \( \hat{\lambda}(u; \hat{\beta}) \) is \( O_p(n^{-2/5}) \).
4. Monte Carlo simulations

To check the finite sample behavior of our estimators, we have conducted some Monte Carlo simulations to estimate the regression coefficients in a censored partially single-index model with \( q = \text{dim}(V) = 2 \) and \( p = \text{dim}(X) = 2 \). We use covariates \( X^T = (X_1, X_2) \), \( X_1 \sim \text{Uniform}(-2, 2) \), \( X_2 \sim \text{Triangular}(-2, 2) \), \( X_2 = (\frac{1}{2})X_1 + (\frac{3}{2})X_2 \) and \( V^T = (V_1, V_2) \), \( V_1 \) and \( V_2 \sim \text{Bernoulli}(p = 0.5) \) and are independent. The covariate vectors \( V \) and \( X \) are independent, however, it is seen that \( X_1 \) and \( X_2 \) are dependent and their correlation coefficient \( \rho_{X_1, X_2} = 0.5 \). The model is stated as

\[
Y = \beta_0^T V + \lambda_0(z_0^T X) + \epsilon,
\]

where \( \epsilon \sim N(0, \sigma_{\epsilon}^2 = 0.5^2) \), the true parameters are set as \( \beta_0 = (-1, 2)^T \) and \( z_0 = (\sqrt{2}/2, \sqrt{2}/2)^T = (0.707, 0.707)^T \), the true nonparametric regression \( \lambda_0(u) = (-1/2)(u - \sqrt{2}/2)^2 + 6, u = z_0^T x \).

The censoring distribution is selected to be \( N(\mu, \sigma_\varepsilon^2 = 2^2) \). We have chosen different \( \mu \) to study the performance of our estimators with different censoring proportions.

In Table 1 we report the results over 1000 simulations. In this table the sample mean (MEAN), standard deviation (SD), and root-mean-squared-error (RMSE), as well as the median (MED) are given as a function of sample size \( n \) and censoring parameter \( \mu \) (corresponding different censoring proportions). We used the Epanechnikov kernel function \( W(v) = \left(\frac{3}{4}\right)(1 - v^2) I[|v| \leq 1] \) to compute the estimates. The bandwidth is selected to be \( b = 3\{\log(n)/n\}^{1/3} \) in estimating \( z_0 \) and \( \beta_0 \), where the constant 3 is about 2.2 × Var\((z_0^T X)\) and is obtained partly by trial and error. Using this constant, the bandwidth formula gives values in the range of \( (0.7 \times SD(z_0^T X), 1.2 \times SD(z_0^T X)) \) in the simulation studies. Although a data driven bandwidth selection is desirable, it is beyond the scope of the present study.

We can make the following conclusions from the simulation results in Table 1:

1. The median bias, SD and RMSE are small for moderate sample sizes or censoring proportions. It confirms that our method works quite well in these cases.
2. When the censoring proportion is high \( \rho \geq 0.40 \) and the sample size is small \( n \leq 100 \), the estimators for the single-index coefficients are not very well behaved. The estimates are biased away from their true values, but the median bias of \( \hat{\lambda} \) tends to be much smaller than the mean bias. In fact, the median of \( \hat{\lambda} \) is so close to the true value that the median bias is nearly zero. We notice that the larger mean bias is due to the asymmetry of the sampling distribution of \( \hat{\lambda} \), which is induced by the constraint imposed on \( z_0 \).
3. For the same sample size, when the censoring proportion increases, the (mean or median) bias, SD and RMSE increase.
4. For the same censoring proportion, when the sample size increases, the mean bias decreases, the median bias fluctuates around zero and is relatively stable, SD and RMSE decrease, which suggests that large samples would be required to reduce the mean bias.

5. Concluding remarks

A partially linear single-index model is proposed as a tool for studying relationships between a response and a set of predictor variables in survival analysis when the response variable is subject to random censorship. Our model can model flexible covariate effects when either pure parametric or nonparametric model is not appropriate to fit data. Since we do not assume the
distribution of the response variable, we use a quasi-likelihood approach. It is found this approach has similar features of the likelihood approaches investigated by Carroll et al. [4] for complete data in generalized linear single-index models and by Lu et al. [25] for censored data in partially linear single-index proportional hazards models. Asymptotic normality of the estimators for both parametric and nonparametric components are obtained. Monte Carlo simulation results are given, which demonstrate the accuracy and usefulness of the method. In our results, we have used the assumption that the censoring variable \( C \) and \((V, X, Y)\) are independent, a more realistic assumption is possible when the dimension of the covariate vector \((V, X)\) is low, for example, assuming that \( C \) is independent of \( Y \), given the covariates \((V, X)\). In this case, we can estimate the conditional distribution of \( C \) given \((V, X)\) to get \( \hat{G}(\cdot) = \hat{G}(\cdot|V, X) \). But with high-dimensional covariates, it is hard to get such an estimator and the possible estimator will not have the required rate of convergence to establish the asymptotic results. This issue would be an interesting problem in future research for multiple censored regression models.

Table 1
Descriptive statistics of estimated regression coefficients as a function of censoring proportion \( p \) (censoring distribution parameter \( \mu \)) and sample size \( n \)

<table>
<thead>
<tr>
<th>( x_0 )</th>
<th>( \beta_0 )</th>
<th>MEAN</th>
<th>SD</th>
<th>RMSE</th>
<th>MED</th>
<th>MEAN</th>
<th>SD</th>
<th>RMSE</th>
<th>MED</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p = 10% \ (\mu = 8.8), n = 50 )</td>
<td>( p = 10% \ (\mu = 8.8), n = 100 )</td>
<td>( p = 10% \ (\mu = 8.8), n = 200 )</td>
<td>( p = 25% \ (\mu = 7.4), n = 50 )</td>
<td>( p = 25% \ (\mu = 7.4), n = 100 )</td>
<td>( p = 25% \ (\mu = 7.4), n = 200 )</td>
<td>( p = 40% \ (\mu = 6.2), n = 50 )</td>
<td>( p = 40% \ (\mu = 6.2), n = 100 )</td>
<td>( p = 40% \ (\mu = 6.2), n = 200 )</td>
<td></td>
</tr>
<tr>
<td>0.707</td>
<td>0.706</td>
<td>0.078</td>
<td>0.078</td>
<td>0.707</td>
<td>−1</td>
<td>−0.999</td>
<td>0.213</td>
<td>0.213</td>
<td>−1.004</td>
</tr>
<tr>
<td>0.707</td>
<td>0.699</td>
<td>0.081</td>
<td>0.081</td>
<td>0.707</td>
<td>2</td>
<td>2.001</td>
<td>0.203</td>
<td>0.203</td>
<td>1.999</td>
</tr>
<tr>
<td>0.707</td>
<td>0.706</td>
<td>0.066</td>
<td>0.066</td>
<td>0.707</td>
<td>−1</td>
<td>−1.007</td>
<td>0.187</td>
<td>0.187</td>
<td>−0.998</td>
</tr>
<tr>
<td>0.707</td>
<td>0.701</td>
<td>0.070</td>
<td>0.070</td>
<td>0.707</td>
<td>2</td>
<td>2.001</td>
<td>0.145</td>
<td>0.145</td>
<td>2.004</td>
</tr>
<tr>
<td>0.707</td>
<td>0.705</td>
<td>0.061</td>
<td>0.061</td>
<td>0.707</td>
<td>−1</td>
<td>−1.003</td>
<td>0.129</td>
<td>0.129</td>
<td>−0.999</td>
</tr>
<tr>
<td>0.707</td>
<td>0.704</td>
<td>0.064</td>
<td>0.064</td>
<td>0.707</td>
<td>2</td>
<td>2.001</td>
<td>0.109</td>
<td>0.109</td>
<td>2.001</td>
</tr>
<tr>
<td>0.707</td>
<td>0.695</td>
<td>0.123</td>
<td>0.124</td>
<td>0.707</td>
<td>−1</td>
<td>−0.996</td>
<td>0.332</td>
<td>0.332</td>
<td>−0.999</td>
</tr>
<tr>
<td>0.707</td>
<td>0.698</td>
<td>0.124</td>
<td>0.125</td>
<td>0.707</td>
<td>2</td>
<td>1.990</td>
<td>0.303</td>
<td>0.303</td>
<td>1.996</td>
</tr>
<tr>
<td>0.707</td>
<td>0.704</td>
<td>0.084</td>
<td>0.084</td>
<td>0.707</td>
<td>−1</td>
<td>−1.003</td>
<td>0.220</td>
<td>0.220</td>
<td>−1.001</td>
</tr>
<tr>
<td>0.707</td>
<td>0.700</td>
<td>0.087</td>
<td>0.087</td>
<td>0.707</td>
<td>2</td>
<td>2.004</td>
<td>0.201</td>
<td>0.201</td>
<td>1.994</td>
</tr>
<tr>
<td>0.707</td>
<td>0.709</td>
<td>0.058</td>
<td>0.058</td>
<td>0.707</td>
<td>−1</td>
<td>−1.001</td>
<td>0.145</td>
<td>0.145</td>
<td>−0.995</td>
</tr>
<tr>
<td>0.707</td>
<td>0.701</td>
<td>0.060</td>
<td>0.060</td>
<td>0.707</td>
<td>2</td>
<td>1.999</td>
<td>0.141</td>
<td>0.141</td>
<td>1.997</td>
</tr>
<tr>
<td>0.707</td>
<td>0.680</td>
<td>0.192</td>
<td>0.193</td>
<td>0.707</td>
<td>−1</td>
<td>−0.975</td>
<td>0.522</td>
<td>0.522</td>
<td>−0.953</td>
</tr>
<tr>
<td>0.707</td>
<td>0.672</td>
<td>0.203</td>
<td>0.206</td>
<td>0.707</td>
<td>2</td>
<td>1.987</td>
<td>0.561</td>
<td>0.561</td>
<td>1.993</td>
</tr>
<tr>
<td>0.707</td>
<td>0.700</td>
<td>0.140</td>
<td>0.140</td>
<td>0.707</td>
<td>−1</td>
<td>−0.991</td>
<td>0.335</td>
<td>0.335</td>
<td>−0.980</td>
</tr>
<tr>
<td>0.707</td>
<td>0.686</td>
<td>0.144</td>
<td>0.144</td>
<td>0.707</td>
<td>2</td>
<td>1.984</td>
<td>0.305</td>
<td>0.305</td>
<td>1.983</td>
</tr>
<tr>
<td>0.707</td>
<td>0.704</td>
<td>0.091</td>
<td>0.091</td>
<td>0.707</td>
<td>−1</td>
<td>−0.997</td>
<td>0.233</td>
<td>0.233</td>
<td>−1.005</td>
</tr>
<tr>
<td>0.707</td>
<td>0.698</td>
<td>0.092</td>
<td>0.093</td>
<td>0.707</td>
<td>2</td>
<td>1.986</td>
<td>0.207</td>
<td>0.207</td>
<td>1.982</td>
</tr>
</tbody>
</table>
Acknowledgments

The authors are grateful to the Editor, the Associate Editor and two referees for their insightful and constructive comments that have helped greatly to improve the paper. Lu’s research is partly supported by Discovery Grant 73-1069 from the Natural Sciences and Engineering Research Council of Canada. Cheng’s research is supported in part by the NSC of Taiwan.

Appendix A. Proofs of Theorems 1 and 2

The proof of Theorem 1 is just a part of arguments used in the proof of Theorem 2, we omit it. Here, we give a detailed proof of Theorem 2 only.

Denote

\[ \Psi = \begin{pmatrix} X \lambda'_0(U) \\ V \end{pmatrix}, \quad \Lambda = \begin{pmatrix} X \lambda'_0(U) & 0 \\ 0 & V \end{pmatrix}, \]

and

\[ \Omega = E \left[ \left( \Psi - E(\Psi|U) \right) \epsilon_G \right] \left( \Psi - E(\Psi|U) \right)^T, \]

where \( U = X_0^T X \) and \( \epsilon_G = Z_G - \{ \lambda_0(U) + \beta_0^T V \} \). Let \( Q = B_{20, \beta_0} - A_{20, \beta_0} \), with

\[ A_{20, \beta_0} = -E \left[ \Psi \Psi^T \right], \quad B_{20, \beta_0} = -E \left[ E(\Psi|U) E(\Psi^T|U) \right]. \]

The proof consists of two steps. The first step is to obtain an expansion for \( \hat{\lambda} \). For simplicity, let \( a_0 = a_0(u) = \lambda_0(u), a_1 = a_1(u) = b\lambda'_0(u), \epsilon^*_G = Z_G - \{ a_0 + a_1(U_i - u)/b + \beta_0^T V_i \} \). Without loss of generality, suppose that \( D = [c, d] \) for \(-\infty < c < d < \infty\), and define \( D^0 = [c + b, d - b] \) and \( D^1 = D \setminus D^0 \), where \( b \) is the bandwidth. Let

\begin{align*}
L_n(u) &= n^{-1} \sum_{i=1}^n W_i(U_i - u) \frac{\epsilon^*_G}{f(u)} - (\hat{\lambda} - \lambda_0(U_i)|U = u \}
&\quad - (\hat{\beta} - \beta_0^T) E[V|U = u]. \tag{A.1}
\end{align*}

We will show that

\begin{align*}
\sup_{u \in D^0} |\hat{\lambda}(u; \hat{\lambda}, \hat{\beta}) - \lambda_0(u) - L_n(u)| &= o_p(n^{-1/2}) + O_p(b^2), \\
\sup_{u \in D^1} |\hat{\lambda}(u; \hat{\lambda}, \hat{\beta}) - \lambda_0(u) - L_n(u)| &= o_p(n^{-1/2}) + O_p(b^2) + O_p(b). \tag{A.2}
\end{align*}

Denote the \( k \times k \) identity matrix by \( I_k \) and \( P_{\infty_0} \)

\[ P_{20} = \begin{bmatrix} I_p - x_0 x_0^T & 0 \\ 0 & I_q \end{bmatrix}. \]
Then, we will obtain the following representation:

\[
P_{20} Q n^{1/2} \left( \hat{x} - x_0 \right) = n^{-1/2} \sum_{i=1}^{n} P_{20} \left[ \Psi_i - E \{ \Psi_i | U_i \} \right] \epsilon_i \hat{G} + o_p(1) \\
= n^{-1/2} \sum_{i=1}^{n} P_{20} \left[ \Psi_i - E \{ \Psi_i | U_i \} \right] \epsilon_i G \\
+ n^{-1/2} \sum_{i=1}^{n} P_{20} \left[ \Psi_i - E \{ \Psi_i | U_i \} \right] (\epsilon_i \hat{G} - \epsilon_i G) + o_p(1) \\
\triangleq S_{n1} + S_{n2} + o_p(1),
\]

(A.3)

where \( \epsilon_i \hat{G} = Z_i \hat{G} - \{ \hat{a}_0(U_i) + \hat{\beta}_0^T V_i \} \) and \( \epsilon_i G = Z_i G - \{ \hat{a}_0(U_i) + \hat{\beta}_0^T V_i \} \). The second step is to show that the first term on the right-hand side of (A.3) has an asymptotic variance–covariance matrix \( P_{20} \Omega P_{20} \), the second term has an asymptotic variance–covariance matrix \( P_{20} \Xi(\tau_F) P_{20} \), and the covariance matrix of these two terms is \(-2 P_{20} \Xi(\tau_F) P_{20} \). Therefore,

\[
n^{-1/2} \left( \hat{x} - x_0 \right) = (P_{20} Q)^{-1} (S_{n1} + S_{n2}) + (P_{20} Q)^{-1} o_p(1),
\]

where \( A^{-} \) denotes the generalized inverse of a square matrix \( A \), \( (P_{20} Q)^{-1} (S_{n1} + S_{n2}) \) has an asymptotic variance–covariance \( (P_{20} Q)^{-1} P_{20} (\Omega - \Xi(\tau_F)) P_{20} \left( (P_{20} Q)^{-1} \right)^T = Q^{-1} (\Omega - \Xi(\tau_F)) Q^{-1} = Q^{-1} (\Omega - \Xi(\tau_F)) Q^{-1} \), \( (P_{20} Q)^{-1} o_p(1) = o_p(1) \) since the elements of \( (P_{20} Q)^{-1} \) are finite. To the end, Theorem 2 is proved by applying the central limit theorem. Note that when \( G \) is known, we do not need to estimate it, hence, the second term on the right-hand side of (A.3) disappears, all the above arguments suffice to give a proof of Theorem 1. Now, we start to derive the desired results in each step.

**Proof of (A.2).** Recall \( a_0 = \hat{a}_0(u) \), \( a_1 = b \hat{a}_1(u) \). The local linear estimates of \( a_0 \) and \( a_1 \) are obtained from solving

\[
0 = n^{-1} \sum_{i=1}^{n} W_b(U_i - u) \left[ \frac{1}{(U_i - u)/b} \right] \hat{e}_{i \hat{G}}^*,
\]

where \( \hat{e}_{i \hat{G}}^* = Z_i \hat{G} - \{ \hat{a}_0 + \hat{a}_1(U_i - u)/b + \hat{\beta}_0^T V_i \} \), \( \hat{\cdot} \) indicates to use an estimated error and \( \hat{\cdot}^* \) indicates to use a local version of the estimated error. By this convention, we define \( \hat{e}_{i \hat{G}} = Z_i \hat{G} - \{ \hat{a}(\hat{x}_T X_i; \hat{x}, \hat{\beta}) + \hat{\beta}_0^T V_i \} \). Using a Taylor expansion approximately and eliminating higher order terms, we get uniformly for \( u \in D \),

\[
0 = n^{-1} \sum_{i=1}^{n} W_b(U_i - u) \left[ \frac{1}{(U_i - u)/b} \right] \{ \hat{e}_{i \hat{G}}^* - (\hat{a}_0 - a_0) - ((U_i - u)/b)(\hat{a}_1 - a_1) \\
- (a_1/b)(\hat{x}_T - x_0^T) X_i - (\hat{\beta}_0^T - \beta_0^T) V_i \} + o_p(n^{-1/2}) + O_p(b^2).
\]
Solving the above equation for \( \hat{a}_0 - a_0 \), we have uniformly for \( u \in D \),
\[
\hat{a}_0 - a_0 = \left[ 1 / \left\{ n^{-1} \sum_{i=1}^{n} W_b(U_i - u) \right\} \right] \left[ n^{-1} \sum_{i=1}^{n} W_b(U_i - u) \right. \\
\left. \epsilon_{i\hat{G}}^* - (a_1/b)(\hat{\beta}_T - \beta_0^T)X_i - (\hat{\beta}_{T}^T - \beta_0^T)V_i \right] + o_p(n^{-1/2}) + O_p(b^2) \]  
(A.4)

Let \( \hat{f}(u) = n^{-1} \sum_{i=1}^{n} W_b(U_i - u) \) be the kernel estimator of \( f(u) \), we have the following results about the kernel density estimator and the kernel regression estimators (proofs are put in Sections A.1 and A.2):

\[
\sup_{u \in D} \left| n^{-1} \sum_{i=1}^{n} W_b(U_i - u)(a_1/b)X_i / \hat{f}(u) - E\{X|U = u\} \right| = O_p(b),
\]

\[
\sup_{u \in D} \left| n^{-1} \sum_{i=1}^{n} W_b(U_i - u)V_i / \hat{f}(u) - E\{V|U = u\} \right| = O_p(b), \quad \text{(A.5)}
\]

\[
\sup_{u \in D} \left| n^{-1} \sum_{i=1}^{n} W_b(U_i - u)\epsilon_{i\hat{G}}^*/\hat{f}(u) - 0 \right| = O_p(b), \quad \text{(A.6)}
\]

and

\[
\sup_{u \in D^0} | \hat{f}(u) - f(u) | = O_p(b), \quad \sup_{u \in D^1} | \hat{f}(u) - f(u) | = O_p(1). \quad \text{(A.7)}
\]

Since
\[
\frac{n^{-1} \sum_{i=1}^{n} W_b(U_i - u)\epsilon_{i\hat{G}}^*}{n^{-1} \sum_{i=1}^{n} W_b(U_i - u)} - \frac{n^{-1} \sum_{i=1}^{n} W_b(U_i - u)\epsilon_{i\hat{G}}^*}{f(u)}
\]
\[
= \frac{n^{-1} \sum_{i=1}^{n} W_b(U_i - u)\epsilon_{i\hat{G}}^*}{\hat{f}(u)} \times \frac{f(u) - \hat{f}(u)}{f(u)},
\]

by (A.5) and (A.7), we obtain
\[
\sup_{u \in D^0} \left| \frac{n^{-1} \sum_{i=1}^{n} W_b(U_i - u)\epsilon_{i\hat{G}}^*}{n^{-1} \sum_{i=1}^{n} W_b(U_i - u)} - \frac{n^{-1} \sum_{i=1}^{n} W_b(U_i - u)\epsilon_{i\hat{G}}^*}{f(u)} \right| = O_p(b^2)
\]

and
\[
\sup_{u \in D^1} \left| \frac{n^{-1} \sum_{i=1}^{n} W_b(U_i - u)\epsilon_{i\hat{G}}^*}{n^{-1} \sum_{i=1}^{n} W_b(U_i - u)} - \frac{n^{-1} \sum_{i=1}^{n} W_b(U_i - u)\epsilon_{i\hat{G}}^*}{f(u)} \right| = O_p(b).
\]

Substituting the kernel terms in the linearized equation (A.4) by their asymptotic counterparts, we obtain (A.2).
Proof of (A.3). By a Taylor expansion, we have

\[
\hat{\lambda}(\hat{z}^T X_i; \hat{z}, \hat{\beta}) - \lambda_0(z_0^T X_i) \\
= \hat{\lambda}(\hat{z}^T X_i; \hat{z}, \hat{\beta}) - \hat{\lambda}(z_0^T X_i; \hat{z}, \hat{\beta}) + \hat{\lambda}(z_0^T X_i; \hat{z}, \hat{\beta}) - \lambda_0(z_0^T X_i) \\
= \hat{\lambda}'(z_0^T X_i; \hat{z}, \hat{\beta})(\hat{z}^T - z_0^T)X_i + \hat{\lambda}(z_0^T X_i; \hat{z}, \hat{\beta}) - \lambda_0(z_0^T X_i) + o_p(n^{-1/2}) \\
= \lambda_0'(z_0^T X_i)(\hat{z}^T - z_0^T)X_i + \hat{\lambda}(z_0^T X_i; \hat{z}, \hat{\beta}) - \lambda_0(z_0^T X_i) + o_p(n^{-1/2}).
\]  

(A.8)

With \( \xi \) being the Lagrange multiplier, we know that \((\hat{z}, \hat{\beta})\) is the solution to

\[
0 = \xi \begin{pmatrix} \hat{z} \\ 0 \end{pmatrix} + n^{-1/2} \sum_{i=1}^{n} \hat{\Lambda}_i D_i,
\]

where

\[
\hat{\Lambda}_i = \begin{pmatrix} X_i \hat{\lambda}'(\hat{z}^T X_i; \hat{z}, \hat{\beta}) & 0 \\ 0 & 0 \end{pmatrix}, \quad D_i = \begin{pmatrix} \hat{\epsilon}_i \hat{G} \\ \hat{\epsilon}_i \hat{G} \end{pmatrix},
\]

\[
\hat{\epsilon}_i \hat{G} = Z_i \hat{G} - \{\hat{\lambda}(\hat{z}^T X_i; \hat{z}, \hat{\beta}) + \hat{\beta}^T V_i\}. \text{ Let}
\]

\[
D_{0i} = \begin{pmatrix} \epsilon_i \hat{G} \\ \epsilon_i \hat{G} \end{pmatrix}.
\]

By Taylor expansions, we obtain

\[
D_i = D_{0i} + \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} \hat{\lambda}(\hat{z}^T X_i; \hat{z}, \hat{\beta}) - \lambda_0(z_0^T X_i) \\ V_i^T (\hat{\beta} - \beta_0) \end{pmatrix} + o_p(n^{-1/2}) \\
= D_{0i} + \begin{pmatrix} -1 \\ -1 \end{pmatrix} \{ V_i^T (\hat{\beta} - \beta_0) \} \\
+ \begin{pmatrix} -1 \\ -1 \end{pmatrix} \{ \hat{\lambda}(\hat{z}^T X_i; \hat{z}, \hat{\beta}) - \lambda_0(z_0^T X_i) \} + o_p(n^{-1/2}).
\]

Since \( \hat{\Lambda}_i = \Lambda_i + o_p(1) \), we have

\[
0 = \xi \begin{pmatrix} \hat{z} \\ 0 \end{pmatrix} + n^{-1/2} \sum_{i=1}^{n} \Lambda_i D_{0i} + n^{-1/2} \sum_{i=1}^{n} \Lambda_i \begin{pmatrix} -1 \\ -1 \end{pmatrix} \{ V_i^T (\hat{\beta} - \beta_0) \} \\
+ n^{-1/2} \sum_{i=1}^{n} \Lambda_i \begin{pmatrix} -1 \\ -1 \end{pmatrix} \{ \hat{\lambda}(\hat{z}^T X_i; \hat{z}, \hat{\beta}) - \lambda_0(z_0^T X_i) \} + o_p(1).
\]  

(A.9)
By (A.8), we get
\[ n^{-1/2} \sum_{i=1}^{n} \Lambda_i \left( \begin{array}{c} -1 \\ -1 \end{array} \right) \left( \lambda(\hat{x}_i^T X_i; \hat{x}, \hat{\beta}) - \lambda_0(x_0^T X_i) \right) \]
\[ = n^{-1/2} \sum_{i=1}^{n} \Lambda_i \left( \begin{array}{c} -1 \\ -1 \end{array} \right) \lambda_0(x_0^T X_i)X_i^T (\hat{x} - x_0) \]
\[ + n^{-1/2} \sum_{i=1}^{n} \Lambda_i \left( \begin{array}{c} -1 \\ -1 \end{array} \right) \{ \lambda(x_0^T X_i; \hat{x}, \hat{\beta}) - \lambda_0(x_0^T X_i) \} + o_p(1). \]

Plugging this into (A.9) gives
\[ 0 = \zeta \left( \begin{array}{c} \hat{x} \\ 0 \end{array} \right) + n^{-1/2} \sum_{i=1}^{n} \hat{\Lambda}_i D_{0i} + n^{-1/2} \sum_{i=1}^{n} \hat{\Lambda}_i \left( \begin{array}{c} -1 \\ -1 \end{array} \right) \{ V_i^T (\hat{\beta} - \beta_0) \} \]
\[ + n^{-1/2} \sum_{i=1}^{n} \hat{\Lambda}_i \left( \begin{array}{c} -1 \\ -1 \end{array} \right) \lambda_0(x_0^T X_i)X_i^T (\hat{x} - x_0) \]
\[ + n^{-1/2} \sum_{i=1}^{n} \hat{\Lambda}_i \left( \begin{array}{c} -1 \\ -1 \end{array} \right) \{ \lambda(x_0^T X_i; \hat{x}, \hat{\beta}) - \lambda_0(x_0^T X_i) \} + o_p(1). \]

This leads to
\[ 0 = \zeta \left( \begin{array}{c} \hat{x} \\ 0 \end{array} \right) + n^{-1/2} \sum_{i=1}^{n} \Lambda_i D_{0i} + n^{-1/2} \sum_{i=1}^{n} \Lambda_i \left( \begin{array}{cc} -1 & -1 \\ -1 & -1 \end{array} \right) \Lambda_i^T \left( \begin{array}{c} \hat{x} - x_0 \\ \hat{\beta} - \beta_0 \end{array} \right) \]
\[ + n^{-1/2} \sum_{i=1}^{n} \Lambda_i \left( \begin{array}{c} -1 \\ -1 \end{array} \right) \{ \lambda(x_0^T X_i; \hat{x}, \hat{\beta}) - \lambda_0(x_0^T X_i) \} + o_p(1). \]

Note that by using matrix notation, \( L_n(u) \) in (A.2) can be written as
\[ L_n(u) = n^{-1} \sum_{i=1}^{n} W_b(U_i - u) \frac{e_i^s}{f(u)} \]
\[ + E \left[ \left\{ \Lambda \left( \begin{array}{c} -1 \\ -1 \end{array} \right) \right\}^T U = u \right] \left( \begin{array}{c} \hat{x} - x_0 \\ \hat{\beta} - \beta_0 \end{array} \right). \]

Then from (A.2) and the definition of \( A_{x_0, \beta_0} \), we obtain
\[ 0 = \zeta \left( \begin{array}{c} \hat{x} \\ 0 \end{array} \right) + n^{-1/2} \sum_{i=1}^{n} \Lambda_i D_{0i} + A_{x_0, \beta_0} n^{1/2} \left( \begin{array}{c} \hat{x} - x_0 \\ \hat{\beta} - \beta_0 \end{array} \right) \]
\[ + n^{-1/2} \sum_{i=1}^{n} \Lambda_i \left( \begin{array}{c} -1 \\ -1 \end{array} \right) E \left[ \left\{ \Lambda_i \left( \begin{array}{c} -1 \\ -1 \end{array} \right) \right\}^T U = U_i \right] \left( \begin{array}{c} \hat{x} - x_0 \\ \hat{\beta} - \beta_0 \end{array} \right) \]
\[ + n^{-1/2} \sum_{i=1}^{n} \Lambda_i \left( \begin{array}{c} -1 \\ -1 \end{array} \right) \left[ \frac{n^{-1} \sum_{j=1}^{n} W_b(U_j - U_i) e_j^s}{f(U_i)} \right] \]
\[ n^{-1/2} \sum_{i=1}^{n} \Lambda_i \begin{pmatrix} -1 \\ -1 \end{pmatrix} \{ O_p(b^2) I[U_i \in D^0] + O_p(b) I[U_i \in D^1] \} \]

\[ + Op(n^{-1/2}) + Op(b^2) + o_p(1). \tag{A.10} \]

It is easy to see that the sixth term in (A.10) is \( Op(\sqrt{n}b^2) + op(1) = op(1). \) The fifth term in (A.10) is essentially the same as (a proof is given in Section A.3)

\[ \sum_{i=1}^{n} E \left[ \left\{ \Lambda_i \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\} \right] \epsilon \hat{G} + op(1). \tag{A.11} \]

From

\[ -n^{-1} \sum_{i=1}^{n} \Lambda_i \begin{pmatrix} -1 \\ -1 \end{pmatrix} \left[ \left\{ \Lambda_i \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\} \right] E \left[ \left\{ \Lambda_i \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\} \right] = B_{\alpha_0, \beta_0} \]

and the definition of \( Q, \) (A.10) can be written as

\[ 0 = \hat{\zeta} \begin{pmatrix} \hat{x} \\ 0 \end{pmatrix} + n^{-1/2} \sum_{i=1}^{n} \left[ \Lambda_i \begin{pmatrix} -1 \\ -1 \end{pmatrix} + E \left[ \Lambda_i \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right] \right] \epsilon \hat{G} - Qn^{1/2} \begin{pmatrix} \hat{x} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} + o_p(1). \]

Multiplying both sides by \( P_{\alpha_0} \) and noticing that \( \Lambda_i(1,1)^T = \Psi_i, \) we obtain the first equality in (A.3). We will use the martingale techniques for \( \hat{G} - G \) to deal with the second term in the second equality in (A.3). At the moment, we focus on those results required to establish the first equality.

### A.1. Proofs of (A.5) and (A.7)

**Proof of (A.5).** Let \( \psi^*(\cdot, \cdot) \) denote the quantity \( \psi(a_0(u) + a_1(u)(U_i - u) / b + \beta_0^T V_i) \) and let \( \psi(\cdot, \cdot) \) denote the similar quantity \( \psi(a_0(U_i) + \beta_0^T V_i) \) for some differential and bounded function \( \psi(\cdot) \) or one of the quantities \( V_i \) and \( X_i \) shown up in (A.5). We will show that

\[ \sup_{u \in D} \left| n^{-1} \sum_{i=1}^{n} W_b(U_i - u) \psi^*(\cdot, \cdot) / \hat{f}(u) - n^{-1} \sum_{i=1}^{n} W_b(U_i - u) \psi(\cdot, \cdot) / \hat{f}(u) \right| = O_p(b) \tag{A.12} \]

and

\[ \sup_{u \in D} \left| n^{-1} \sum_{i=1}^{n} W_b(U_i - u) \psi(\cdot, \cdot) / \hat{f}(u) - E[\psi(\cdot)|U = u] \right| = O_p(b). \tag{A.13} \]
Eq. (A.12) will be used in the proof of (A.6). First, we assume that (A.13) holds, we prove (A.12). Let \( \psi'(t) = \frac{d\psi(t)}{dt} \), then

\[
\psi^*(\cdot) - \psi(\cdot) = \psi(\lambda_0(u) + \lambda_0'(u)(U_i - u) + \beta_0^T V_i) - \psi(\lambda_0(U_i) + \beta_0^T V_i)
\]

\[
= \psi' \left( \xi_i(u) \right) \left[ \lambda_0(u) - \lambda_0(U_i) + \lambda_0'(u)(U_i - u) \right]
\]

\[
= \psi' \left( \xi_i(u) \right) \left[ \frac{\lambda_0''(\xi_i(u))}{2} (U_i - u)^2 \right]
\]

\[
= O_p((U_i - u)^2), \tag{A.14}
\]

where \( \xi_i(u) \) is between \( \lambda_0(U_i) + \beta_0^T V_i \) and \( \lambda_0(u) + \lambda_0'(u)(U_i - u) + \beta_0^T V_i, \xi_i(u) \) is between \( U_i \) and \( u \). Therefore

\[
\sup_{u \in D} \left| n^{-1} \sum_{i=1}^n W_b(U_i - u)\psi^*(\cdot) / \hat{f}(u) - n^{-1} \sum_{i=1}^n W_b(U_i - u)\psi(\cdot) / \hat{f}(u) \right| \leq O_p(1) \sup_{u \in D} \left| n^{-1} \sum_{i=1}^n W_b(U_i - u)(U_i - u)^2 / \hat{f}(u) \right| = O_p(b),
\]

using (A.13) by taking \( \psi(\cdot) = (U_i - u)^2 \) and noticing that \( E\{(U_i - u)^2|U_i = u\} = 0 \), this proves (A.12).

Now we prove (A.13). Let \( \hat{r}_b(u) = n^{-1} \sum_{i=1}^n W_b(U_i - u)\psi(\cdot) \), then

\[
n^{-1} \sum_{i=1}^n W_b(U_i - u)\psi(\cdot) / \hat{f}(u) - E\{\psi(\cdot)|U = u\} = \frac{[\hat{r}_b(u) - E\{\hat{r}_b(u)\}]E\{\hat{f}(u)\} - [\hat{f}(u) - E\{\hat{f}(u)\}]E\{\hat{r}_b(u)\}}{[\hat{f}(u) - E\{\hat{f}(u)\}]E\{\hat{f}(u)\}}
\]

\[
+ \left[ \frac{E\{\hat{r}_b(u)\}}{E\{\hat{f}(u)\}} - E\{\psi(\cdot)|U = u\} \right]
\]

\[
= I_1(u) + I_2(u), \tag{A.15}
\]

We consider \( I_2(u) \) first. Since

\[
E\{\hat{r}_b(u)\} = E\{W_b(U_i - u)\psi(\cdot)\} = E\{W_b(U_i - u)E\{\psi(\cdot)|U_i\}\}
\]

\[
= \frac{1}{b} \int_c^d W \left( \frac{y - u}{b} \right) E\{\psi(\cdot)|U = y\} f(y) dy
\]

\[
= \int_{c - u}^{d - u} W(t)E\{\psi(\cdot)|U = u + bt\} f(u + bt) dt
\]

\[
= \left\{ \int_{c - u}^{d - u} W(t) dt \right\} E\{\psi(\cdot)|U = u\} f(u) + O(b)
\]

and

\[
E\{\hat{f}(u)\} = \left\{ \int_{c - u}^{d - u} W(t) dt \right\} f(u) + O(b)
\]
hold uniformly for \( u \in D \), we have
\[
\sup_{u \in D} |I_2(u)| = O(b).
\]

To finish the proof, it suffices to show
\[
\sup_{u \in D} |\hat{r}_b(u) - E[\hat{r}_b(u)]| = O_p(b), \tag{A.16}
\]
\[
\sup_{u \in D} |\hat{f}(u) - E[\hat{f}(u)]| = O_p(b). \tag{A.17}
\]

We prove (A.16) here as (A.17) is very easy to prove. We consider a more general case where \( \psi(\cdot) \) might be unbounded but \( |\psi(\cdot)| \leq C_\psi C_T T_g^s \) for some constants \( C_\psi, C_T, g > 0 \) and some i.i.d. random variables \( T_i \) for which \( \sup_{i,u} E\{T_i (2s+1)g | U_i = u\} < \infty \) and \( \sup_i E\{T_i (2s+1)g\} < \infty \) for some \( s > 1 \). Taking \( N_n = b^{-1/s} \) and writing
\[
\hat{r}_b(u) = n^{-1} \sum_{i=1}^n W_b(U_i - u) \psi(\cdot) I(|\psi(\cdot)| \leq N_n)
+ n^{-1} \sum_{i=1}^n W_b(U_i - u) \psi(\cdot) I(|\psi(\cdot)| > N_n)
\equiv J_1(u) + J_2(u),
\]
it suffices to show
\[
\sup_{u \in D} |J_1(u) - E[J_1(u)]| = O_p(b), \tag{A.18}
\]
and
\[
\sup_{u \in D} |J_2(u) - E[J_2(u)]| = O_p(b). \tag{A.19}
\]

When \( \psi(\cdot) \) is bounded, (A.19) is trivial.

Suppose that \( M_n \) intervals \( \{u : |u - u_l| \leq \eta_n\}, l = 1, \ldots, M_n \), cover the compact set \( D \) and the union of these intervals equals \( D \). Then, for any \( \nabla > 0 \),
\[
P \left\{ \sup_{u \in D} |J_1(u) - E[J_1(u)]| > \nabla b \right\}
= P \left\{ \sup_{l=1,\ldots,M_n} \sup_{|u-u_l| \leq \eta_n} |J_1(u) - E[J_1(u)]| > \nabla b \right\}
\leq P \left\{ \sup_{l=1,\ldots,M_n} |J_1(u_l) - E[J_1(u_l)]| > \frac{\nabla}{2} b \right\}
\]
\[
+ P \left\{ \sup_{l=1,\ldots,M_n} \sup_{|u-u_l| \leq \eta_n} |J_1(u) - J_1(u_l) - (E[J_1(u)] - E[J_1(u_l)])| > \frac{\nabla}{2} b \right\}. \tag{A.20}
\]

By Condition A(ii), there exist some constants \( C_W > 0 \) and \( C_L > 0 \) such that \( |W(x)| \leq C_W \) and \( |W(x_1) - W(x_2)| \leq C_L |x_1 - x_2| \). Taking \( M_n = O(n^2) \) and \( \eta_n = O(n^{-2}) \), when \( |u - u_l| \leq \eta_n \).
we have
\[ |J_1(u) - J_1(u_l)| = \left( nb^{-1} \sum_{i=1}^{n} \left\{ W \left( \frac{U_i - u}{b} \right) - W \left( \frac{U_i - u_l}{b} \right) \right\} \psi(\cdot) I[|\psi(\cdot)| \leq N_n] \right) \]
\[ \leq (nb)^{-1} C_L \left| \frac{u - u_l}{b} \right| n N_n \]
\[ = C_L b^{-(2+1/s)} \eta_n = O((nb)^{-2}) \cdot O(b^{4-1/s}) = o_p(b). \]

Therefore, \( \sup_{l=1,\ldots,M_n} \sup_{|u-u_l| \leq \eta_n} |J_1(u) - J_1(u_l)| = o_p(b). \)

Similarly, \( \sup_{l=1,\ldots,M_n} \sup_{|u-u_l| \leq \eta_n} |E[J_1(u)] - E[J_1(u_l)]| = o(b). \) Hence, the second probability in (A.20) is negligible. Let \( d_l(u) = W(\frac{U_l - u}{b})\psi(\cdot) I[|\psi(\cdot)| \leq N_n] \) and \( S_n(u) = \sum_{l=1}^{n} [d_l(u) - E[d_l(u)]] \). Then, \( |d_l(u) - E[d_l(u)]| \leq 2C_W N_n \) and \( \sigma_n^2 = \text{Var}(S_n(u)) = n[E[W^2(\frac{U_l - u}{b})\psi^2(\cdot) I[|\psi(\cdot)| \leq N_n]]^2] = O(nb) - O(nb^2) = O(nb), \) because \( E[\psi^2(\cdot)|U_i = u] \leq C^2 \psi^2 \mathbb{E}[T_i^2|U_i = u] < M < \infty \) for some constants \( C_\psi > 0 \) and \( M > 0 \) by the preceding assumptions. Without loss of generality, we assume \( \sigma_n^2 = nb \). By Bernstein’s inequality (see [1]), for any \( \omega > 0 \), we get
\[
P(|S_n(u_l)| \geq \omega \sigma_n) \leq 2 \exp \left[-\frac{\omega^2}{2 + \frac{2}{3} \frac{\mathbb{E}[\nabla^2 S_n(u)]}{\sigma_n^2}}\right].
\]

Taking \( \omega = (\nabla b \sigma_n)/2 \) and noticing \( \sigma_n = \sqrt{nb} \) and \( N_n = b^{-1/s}, s > 1 \), we get
\[
P(|S_n(u_l)| \geq (nb)(\nabla/2)b) \leq 2 \exp \left[-\frac{\left( \frac{\nabla b \sigma_n}{2} \right)^2}{2 + \frac{2}{3} \frac{\mathbb{E}[\nabla^2 S_n(u)]}{\sigma_n^2}}\right]
\]
\[ = \exp \left[-\frac{\left( \frac{\nabla b \sigma_n}{2} \right)^2}{2 + \frac{2}{3} (C_W \nabla) b^{1-1/s}}\right] \]
\[ = \exp \left[-\frac{\left( \frac{\nabla b}{2} \right)^2 O(nb^3)}{2 + \frac{2}{3} (C_W \nabla) b^{1-1/s}}\right] \]
\[ \leq O(2n^{-(3/32)\nabla^2}) \quad (\text{Assume } C_W \nabla b^{1-1/s} < 1). \]

Since \( M_n = O(n^2) \), when \( \nabla \) is large enough so that \( (\frac{3}{32}) \nabla^2 > 2 \), we get
\[
P \left\{ \sup_{l,\ldots,M_n} \left| \sum_{i=1}^{n} [d_l(u_l) - E[d_l(u_l)]] \right| \geq \left( \frac{\nabla b}{2} \right) \sigma_n^2 \right\} \leq M_n O(2n^{-(3/32)\nabla^2}) \xrightarrow{n \to \infty} 0.
\]
This implies
\[ \sup_{i,\ldots,M_n} \left| \frac{1}{nb} \sum_{i=1}^{n} [d_i(u_i) - E\{d_i(u_i)\}] \right| = O_p(b). \] (A.21)

Combining (A.20) and (A.21) proves (A.18).

Now we prove (A.19). By Conditions A and B, \(|\psi(\cdot)| \leq C_\psi C_T T_i^g\) for some constant \(C_T\), \(C_\psi > 0\). From \(|W(x)| \leq C_W\), we have
\[ \sup_{u \in D} \left| J_2(u) - E\{J_2(u)\}\right| = \frac{2C_\psi C_W C_T}{b} \frac{1}{n} \sum_{i=1}^{n} T_i^g I[T_i^g > N_n] \] (A.22)

since \(E\{T_i^g I[T_i^g > N_n]\} = \int_{t > N_n^{1/g}} t^g dF_T(t)\), where \(F_T(t)\) is the c.d.f. of \(T\), and \(N_n = b^{-1/s}\), \(s > 1\). Let \(Q_n = N_n^{1/g} = b^{-1/(sg)}\), we get
\[ \int_{t > Q_n} t^g dF_T(t) = \frac{\int_{t > Q_n} t^g dF_T(t)}{Q_n^{2sg}} \leq \frac{\int_{t > Q_n} t^{(2s+1)g} dF_T(t)}{Q_n^{2sg}} \rightarrow 0, (n \to \infty), \]

because \(E\{T^{(2s+1)g}\} < \infty\) by the preceding assumptions. This implies \((1/nb) \sum_{i=1}^{n} T_i^g I[T_i^g > N_n] = O_p(b)\). Therefore, by (A.22), we obtain (A.19).

**Proof of (A.7).** Since
\[ \sup_{u \in D^0} |\hat{f}(u) - f(u)| \leq \sup_{u \in D^0} |\hat{f}(u) - E\{\hat{f}(u)\}| + \sup_{u \in D^0} |E\{\hat{f}(u)\} - f(u)|, \]
\[ \sup_{u \in D^1} |\hat{f}(u) - f(u)| \leq \sup_{u \in D^1} |\hat{f}(u) - E\{\hat{f}(u)\}| + \sup_{u \in D^1} |E\{\hat{f}(u)\} - f(u)|, \]
\[ \sup_{u \in D^1} |E\{\hat{f}(u)\} - f(u)| \leq \sup_{u \in D^1} |E\{\hat{f}(u)\}| + \sup_{u \in D^1} |f(u)| = O(1), \]
using (A.17) and noticing that \(\sup_{u \in D^0} |E\{\hat{f}(u)\} - f(u)| = O(b^2)\), we obtain (A.7).

**A.2. Proof of (A.6)**

Since \(\epsilon_{i\hat{G}} = Z_{i\hat{G}} - \{\lambda_0(u) + \lambda_0'(u)(U_i - u) + \beta_0^T V_i\}\), it suffices to show
\[ \sup_{u \in D} \left| n^{-1} \sum_{i=1}^{n} W_b(U_i - u) \epsilon_{i\hat{G}}/\hat{f}(u) - n^{-1} \sum_{i=1}^{n} W_b(U_i - u) \epsilon_{i\hat{G}}/\hat{f}(u)\right| = O_p(b) \] (A.23)
and
\[ \sup_{u \in D} \left| n^{-1} \sum_{i=1}^{n} W_b(U_i - u) \epsilon_{i\hat{G}}/\hat{f}(u) - 0 \right| = O_p(b). \] (A.24)
The proof of (A.23) is similar to that of (A.12), we omit it. The proof of (A.24) is equivalent to the proof of the following two equalities:

\[
\sup_{u \in D} \left| n^{-1} \sum_{i=1}^{n} W_b(U_i - u) (\epsilon_i \hat{G} - \epsilon_i G) / \hat{f}(u) \right| = O_p(b) \tag{A.25}
\]

and

\[
\sup_{u \in D} \left| n^{-1} \sum_{i=1}^{n} W_b(U_i - u) \epsilon_i G / \hat{f}(u) - 0 \right| = O_p(b). \tag{A.26}
\]

We prove (A.25) first. Observing that

\[
|K_{i \hat{G}} - K_{i G}| = \left| \frac{K_{i \hat{G}}}{1 - G(Z_i)} \left[ \frac{\hat{G}(Z_i) - G(Z_i)}{1 - \hat{G}(Z_i)} \right] \right|
\]

\[
\leq \left[ 1 + \sup_{t \leq \max_i |Z_i|} \frac{|\hat{G}(t) - G(t)|}{|1 - \hat{G}(t)|} \right] |\hat{G}(Z_i) - G(Z_i)| |K_{i G}| \left( \frac{\Delta Z_i}{1 - G(Z_i)} \right)^2,
\]

\[
\left| \int_{0}^{Z_i} \left\{ \frac{1}{1 - \hat{G}(t)} - \frac{1}{1 - G(t)} \right\} dt \right|
\]

\[
\leq \sup_{t \leq \tau_F} |\hat{G}(t) - G(t)| \left[ 1 + \sup_{t \leq \max_i |Z_i|} \frac{|\hat{G}(t) - G(t)|}{|1 - \hat{G}(t)|} \right] \left[ \frac{Z_i}{1 - G(Z_i)} \right]^2,
\]

using \(\sup_{t \leq \max_i |Z_i|} |\hat{G}(t) - G(t)| / |1 - \hat{G}(t)| = O_p(1)\) due to Srinivasan and Zhou [31], and

\[
\sup_{u \in D} \left| n^{-1} \sum_{i=1}^{n} W_b(U_i - u) \left[ |(1 + \phi)Z_i(1 - G(Z_i))^{-2}| + |\phi \Delta_i Z_i(1 - G(Z_i))^{-2}| / \hat{f}(u) \right] \right| = O_p(1), \tag{A.28}
\]

which is implied by Condition A(vii), we get

\[
\sup_{u \in D} \left| n^{-1} \sum_{i=1}^{n} W_b(U_i - u) (\epsilon_i \hat{G} - \epsilon_i G) / \hat{f}(u) \right|
\]

\[
\leq \sup_{t \leq \tau_F} |\hat{G}(t) - G(t)| \left[ 1 + \sup_{t \leq \max_i |Z_i|} \frac{|\hat{G}(t) - G(t)|}{|1 - \hat{G}(t)|} \right] \left| n^{-1} \sum_{i=1}^{n} W_b(U_i - u) \right| \left| |(1 + \phi)Z_i(1 - G(Z_i))^{-2}| + |\phi \Delta_i Z_i(1 - G(Z_i))^{-2}| / \hat{f}(u) \right|
\]

\[
= O_p(1) \sup_{t \leq \tau_F} |\hat{G}(t) - G(t)|. \tag{A.29}
\]
By Conditions B(i)–(iii), and the results obtained by Gu and Lai [10] and Chen and Lo [5],
\[\sup_{t \leq \varepsilon_T} |\hat{G}(t) - G(t)| = o_p(b).\] Hence, (A.25) holds.

Now we prove (A.26). Note that \(E\{n^{-1}\sum_{i=1}^{n} W_b(U_i - u)\epsilon_i G\} = 0\), by the same arguments used in the proof of (A.13), it suffices to show
\[
\sup_{u \in D} \left| n^{-1} \sum_{i=1}^{n} W_b(U_i - u)\epsilon_i G - E \left\{ n^{-1} \sum_{i=1}^{n} W_b(U_i - u)\epsilon_i G \right\} \right| = O_p(b). \tag{A.30}
\]

By decomposition, it suffices to show
\[
\sup_{u \in D} \left| n^{-1} \sum_{i=1}^{n} W_b(U_i - u)\lambda_0(U_i) - E \left\{ n^{-1} \sum_{i=1}^{n} W_b(U_i - u)\lambda_0(U_i) \right\} \right| = O_p(b), \tag{A.31}
\]
\[
\sup_{u \in D} \left| n^{-1} \sum_{i=1}^{n} W_b(U_i - u)\beta_0^T V_i - E \left\{ n^{-1} \sum_{i=1}^{n} W_b(U_i - u)\beta_0^T V_i \right\} \right| = O_p(b), \tag{A.32}
\]
and
\[
\sup_{u \in D} \left| n^{-1} \sum_{i=1}^{n} W_b(U_i - u)Z_i G - E \left\{ n^{-1} \sum_{i=1}^{n} W_b(U_i - u)Z_i G \right\} \right| = O_p(b). \tag{A.33}
\]

We shall apply the similar techniques used in the proof of (A.16) to prove the preceding three equalities. By Conditions A(ii) and (iv), \(\lambda_0(U_i)\) and \(\beta_0^T V_i\) are bounded random variables, the proofs of (A.31) and (A.32) are straightforward. We only need to prove (A.33). Assume \(\psi(\cdot) = Z_i G\) in the proof of (A.16). Note that, for any \(\rho_n \in (0, 1)\), \(\rho_n \to 0, n \to \infty\),
\[
I[|Z_i G| \geq N_n] \leq I[|Y_i| \geq N_n'(1 - G(Y_i))]
\]
\[
= I[|Y_i| \geq N_n'(1 - G(Y_i)), 1 - G(Y_i) > \rho_n]
\]
\[
+ I[|Y_i| \geq N_n'(1 - G(Y_i)), 1 - G(Y_i) \leq \rho_n]
\]
\[
\leq I[|Y_i| \geq N_n'\rho_n] + I[1 - G(Y_i) \leq \rho_n], \tag{A.34}
\]
where \(N_n' = (1 + \phi + |\phi|)^{-1}N_n\), for ease of presentation, we treat it as \(N_n\) without affecting the proof. Let \(\rho_n = b^\gamma\), for \(0 < \gamma < 1/(\kappa + 1)\), where \(\kappa\) and \(\gamma\) are given in Condition A(vi). Let \(N_n = \rho_n^{-1}(1 - G)^{-1}(\rho_n)\), \((1 - G)^{-1}(\cdot)\) denotes the inverse function of \(1 - G\). By Condition A(vi), \(N_n \leq C G b^{-\gamma/s}b^{-\kappa/2} = C G b^{-\gamma/(\kappa + 1)} \equiv C G b^{-1/s}, s = 1/\{\gamma(\kappa + 1)\} > 1\). For this \(N_n\), by Condition A(vii), using a similar argument to the proof of (A.18), we see that (A.18) also holds for \(\psi(\cdot) = Z_i G\). Next we show the result in (A.19). By such chosen \(\rho_n\) and \(N_n\) and (A.34), we have
\[
I[|Z_i G| \geq N_n] \leq I[|Y_i| \geq N_n(1 - G(Y_i))] \leq 2I[|Y_i| \geq (1 - G)^{-1}(b^\gamma)]. \tag{A.35}
\]

Therefore,
\[
\frac{E[|Z_i G|I[|Z_i G| \geq N_n]]}{b^2} \leq \frac{E[|Y_i|I[|Y_i| \geq N_n(1 - G(Y_i))]]}{b^2}
\]
\[
\leq 2 \frac{E[|Y_i|I[|Y_i| \geq (1 - G)^{-1}(\rho_n)]]}{b^2}.
\]
\[ \int_{|y| \geq (1 - G)^{-1}(b)} |y| b^{-2} \, dF(y) \leq 2 \int_{|y| \geq (1 - G)^{-1}(b)} |y| (1 - G(|y|))^{-2/\gamma} \, dF(y) \rightarrow 0 \text{ when } b \rightarrow 0, \quad (A.36) \]

since \( E[|Y| (1 - G(|Y|))^{-2/\gamma}] < \infty \) by Condition A(vi). Hence, (A.19) can be established for \( \psi(c) = Z_{iG} \) too. Finally, (A.18) and (A.19) imply (A.33).

### A.3. Proof of (A.11)

Let \( \epsilon_j(u) = \epsilon_j^*(u) = Z_j G - \{ \lambda_0(u) + \lambda'_0(u)(U_j - u) + \beta_0^T V_j \} \), then

\[ \epsilon_j(U_i) - \epsilon_j(U_j) = \frac{\lambda''_0(\zeta_i)/2}{f(U_i)} (U_j - U_i)^2 \]

where \( \zeta_i \) is between \( U_j \) and \( U_i \). Hence,

\[ \frac{n^{-1} \sum_{j=1}^{n} W_b(U_j - U_i) \epsilon_j(U_i)}{f(U_i)} = \frac{n^{-1} \sum_{j=1}^{n} W_b(U_j - U_i) \epsilon_j(U_j)}{f(U_i)} + O_p(b^2). \quad (A.38) \]

Each element in \( A_i(-1, -1)^T \) is a function of \( (X_i, V_i, U_i) \), we consider only one such element here, assuming that it is \( g(X_i, V_i, U_i) \). Let

\[ q_i = g(X_i, V_i, U_i) \]

and

\[ \Phi_{1n} = n^{-3/2} \sum_{i=1}^{n} \sum_{j=1}^{n} q_i f^{-1}(U_i) \epsilon_j G W_b(U_j - U_i), \quad (A.39) \]

then by (A.38),

\[ n^{-1/2} \sum_{i=1}^{n} g(X_i, V_i, U_i) \left[ \frac{n^{-1} \sum_{j=1}^{n} W_b(U_j - U_i) \epsilon_j(U_i)}{f(U_i)} \right] = \Phi_{1n} + O_p(\sqrt{nb^2}). \quad (A.40) \]

Now we show that

\[ \Phi_{1n} = \Phi_{2n} + o_p(1), \quad (A.41) \]

where \( \Phi_{2n} = n^{-1/2} \sum_{j=1}^{n} r(U_j) \epsilon_j G \) and \( r(U_j) = E[q_j|U_j] \).

For both \( \Phi_{1n} \) and \( \Phi_{2n} \), using some similar arguments to the proofs in Section A.4, we can show that

\[ \Phi_{1n} = \Phi_{1n} + \tilde{M}_n + o_p(1) \quad (A.42) \]
and
\[ \Phi_{2nG} = \Phi_{2nG} + \tilde{M}_n + o_p(1), \tag{A.43} \]
where
\[ \tilde{M}_n = n^{-1/2} \sum_{k=1}^{n} \left[ \int_0^{rF} E \left\{ \frac{r(U_1)((1 + \phi)J(G, s, Z_1) - \phi K_{1G}[I[s < Z_1]])}{(1 - G(s))(1 - F(s - ))} \right\} dM_k(s) \right], \]
which is a martingale and \( M_k(s) \) is defined in Section A.4. Hence, it suffices to show that
\[ \Phi_{1nG} = \Phi_{2nG} + o_p(1). \tag{A.44} \]
It is easy to see that for \( i \neq j \),
\[ E[q_i f^{-1}(U_i)W_b(U_j - U_i)|U_j] = r(U_j) + O_p(b^2). \]
From (A.39), it follows that
\[ E(\Phi_{1n} - \Phi_{2n})^2 = n^{-3} \sum_{i,j,k,l} E[\epsilon_j G \epsilon_l G \{q_i f^{-1}(U_i)W_b(U_j - U_i) - r(U_j)\}] \]
\[ \times \{q_k f^{-1}(U_k)W_b(U_l - U_k) - r(U_l)\}]\]
\[ = n^{-3} \left[ \sum_{\text{at most two indices equal}} + \sum_{\text{at least three indices equal}} \right] \]
\[ = n^{-3} \sum_{\text{at most two indices equal}} + O(n^{-1}). \tag{A.45} \]

There are three different cases of summands appearing in the above expression.

Case 1: All four indices are different: \( i \neq j \neq k \neq l \). Since \( E[\epsilon_j G | X_j, V_j] = 0 \), all terms in the summation are zero.

Case 2: \( i \neq k \), and precisely two indices are equal. For example, \( j = l \), conditioning on all random variables indexed by \( i \) and \( k \) yields that all these summands are of order \( O(b) \).

Case 3: \( i = k \), and all other indices are unequal. The summands become
\[ E[\epsilon_j G \epsilon_l G E[q_i^2 f^{-2}(U_i)W_b(U_j - U_i)W_b(U_l - U_i)|U_j, U_l]] = O(b). \]

All these imply that \( E(\Phi_{1n} - \Phi_{2n})^2 = O(b) = o(1) \). This establishes (A.44). Combining (A.40) and (A.41) leads to (A.11).

A.4. Using martingale techniques to prove Theorem 2

Now we return to the proof of Theorem 2. In (A.3), \( \Psi_i - E[\Psi_i|U_i] \) is a vector with \( p + q \) elements. Let \( H_i = \Psi_i - E[\Psi_i|U_i] \) and suppose its elements are \( H_{i,l} = H_{i,l}(X_i, V_i, U_i), l = 1, 2, \ldots, p + q \), then we consider
\[ n^{-1/2} \sum_{i=1}^{n} H_{i,l} \epsilon_{iG}, \quad l = 1, 2, \ldots, p + q. \tag{A.46} \]
Let
\[ G(t) = \int_0^t \frac{1}{1 - G(s)} \, dG(s), \]
\[ N_i(t) = I[Z_i \leq t, \delta_i = 0], \quad Y_i(t) = I[Z_i \geq t], \]
\[ M_i(t) = N_i(t) - \int_0^t I[Z_i \geq s] \, d\Lambda_i(s), \quad \Lambda_i(s) = \Lambda^G(s), \]
\[ Y_n(t) = \sum_{i=1}^n Y_i(t), \quad \bar{Y}_n = \frac{1}{n} Y_n(t). \]

By the fact
\[ \hat{G}(z-) - G(z-) = \int_{s<z} \frac{1 - \hat{G}(s) - \sum_{j=1}^n dM_j(s)}{1 - G(s)} \frac{1}{1 - G(s)} \bar{Y}_n dM_k(s), \]
we obtain
\[ K_i\hat{G} - K_iG = \frac{Z_i \Delta_i}{1 - \hat{G}(Z_i-)} \frac{\hat{G}(Z_i-) - G(Z_i-)}{1 - G(Z_i-)} \]
\[ = n^{-1} \sum_{k=1}^n \left[ \int_0^{Z_i} K_i\hat{G} I[s < Z_i] \frac{1 - \hat{G}(s)}{1 - G(s)} \frac{1}{\bar{Y}_n} dM_k(s) \right]. \]

Noticing \( J(\hat{G}, s, Z_i) = \int_{s}^{Z_i} 1/[1 - \hat{G}(t-)] \, dt \) \( I[s < Z_i] \), we decompose \( \epsilon_i \hat{G} \) in (A.46) so that we get
\[ H_{i,l} \epsilon_i \hat{G} = H_{i,l} \epsilon_i G + H_{i,l}(\epsilon_i \hat{G} - \epsilon_i G) \]
\[ = H_{i,l} \epsilon_i G + H_{i,l}(Z_i \hat{G} - Z_i G) \]
\[ = H_{i,l} \epsilon_i G + n^{-1} \sum_{k=1}^n \left[ \int_0^{Z_i} H_{i,l} \{ (1 + \phi) J(\hat{G}, s, Z_i) - \phi K_{i} \hat{G} \} \right] \]
\[ \times I[s < Z_i] \frac{1 - \hat{G}(s)}{1 - G(s)} \frac{1}{\bar{Y}_n} dM_k(s). \]
Fix \( v < \tau_F \), let
\[
M_{1n}^l = n^{-1/2} \sum_{i=1}^{n} H_i, i \in G
\]
and
\[
\hat{M}_{2n}^l(v) = n^{-1/2} \sum_{k=1}^{n} \left[ \int_0^v \left\{ n^{-1} \sum_{i=1}^{n} H_i, i \{(1 + \phi) J(\hat{G}, s, Z_i) - \phi K_{i_1} \} I[s < Z_i] \right\} \right. \\
\times \frac{1 - \hat{G}(s-)}{1 - G(s)} \frac{1}{Y_n} dM_k(s) \left. \right].
\]
Using arguments similar to the proofs of (2.28) and (2.29) in Lai et al. [17], it can be shown that
\[
\hat{M}_{2n}^l(\tau_F) = M_{2n}^l(\tau_F) + o_p(1),
\]
where
\[
M_{2n}^l(v) = n^{-1/2} \sum_{k=1}^{n} \left[ \int_0^v H_l(s) dM_k(s) \right].
\]
\[
H_l(s) = \frac{E[H_l, i\{(1+\phi)J(G, s, Z_i) - \phi K_{i_1}(G)I[s < Z_i]\}]}{(1-G(s))(1-F(s-))}.
\]
For \(-\infty < v \leq \tau_F\), the process \( M_{2n}(v) \) is a local martingale with respect to the right continuous filtration \( F_t \) defined by \( F_t = \sigma\{(V_i, X_i), N_i(s), Y_i(s+) : 0 \leq s \leq t, i = 1, \ldots, n\} \). Note that the predictable covariation process is given by
\[
\langle M_{1n}^l(v), M_{2n}^l(v) \rangle = \frac{1}{n} \sum_{k=1}^{n} \int_0^v H_l(s) H_{2n}^l(s) I[Z_k \geq s] (1 - \Delta \Lambda^G(s)) d\Lambda^G(s)
\]
\[
\geq \int_0^v H_l(s) H_{2n}^l(s) P[Z \geq s] (1 - \Delta \Lambda^G(s)) d\Lambda^G(s)
\]
\[
= \int_0^v H_l(s) H_{2n}^l(s) (1 - G(s-))(1 - F(s-)) \frac{1 - G(s)}{1 - G(s-)} \frac{dG(s)}{1 - G(s-)}
\]
\[
= \int_0^v H_l(s) H_{2n}^l(s) (1 - F(s-)) \frac{dG(s)}{1 - G(s-)}.
\]
Hence, we obtain
\[
\text{Cov}(M_{2n}^l(\tau_F), M_{2n}^l(\tau_F)) = \int_0^{\tau_F} H_l(s) H_{2n}^l(s) (1 - G(s-))(1 - F(s-)) \frac{dG(s)}{1 - G(s-)}. \quad (A.47)
\]
Fix \( v < \tau_F \), consider the joint distribution of \( (M_{1n}^l, M_{2n}^l(t)) \). Since \( dM_k(s) = dN_k(s) - I[Z_k \geq s] d\Lambda_k(s) = (1 - \Delta_k) I[Z_k \leq s] - I[Z_k \geq s] \frac{dG(s)}{1 - G(s-)} \), noticing that \( \Delta_i(1 - \Delta_i) = 0 \), we have
\[
E\{M_{1n}^l, M_{2n}^l(v)\}
\]
\[
= \frac{1}{n} E \sum_{i=1}^{n} \int_0^v H_{2n}(s) H_{i,l_1}(Z_{iG} - E(Y)) dM_i(s)
\]
\[
= E \int_0^v H_{2n}(s) H_{i,l_1}(Z_{1G} - E(Y)) dM_1(s)
\]
\[\begin{align*}
&= -E \int_0^v H_{1,l_1} Z_1 G I[Z_1 \geq s] \frac{H_{l_2}(s)}{1 - G(s)} dG(s) \\
&+ E \int_0^v H_{1,l_1} E(Y) I[Z_1 \geq s] \frac{H_{l_2}(s)}{1 - G(s)} dG(s) - E \{H_{l_2}(Z_1) I[Z_1 \leq v] H_{1,l_1} E(Y)(1 - \Delta_1)\} \\
&= -E \int_0^v H_{1,l_1} Z_1 G I[Z_1 \geq s] \frac{H_{l_2}(s)}{1 - G(s)} dG(s) \\
&+ E \int_0^v H_{l_2}(s) H_{1,l_1} E(Y) [1 - F_{V_1,X}(s)] dG(s) \\
&= -E \int_0^v H_{l_2}(s) H_{1,l_1} E(Y) [1 - F_{V_1,X}(s)] dG(s) \\
&= -E \int_0^v H_{l_2}(s) H_{1,l_1} E(Y) [1 - F_{V_1,X}(s)] dG(s) \\
&- E \int_0^v \frac{E[H_{l_2}(Z_1 G I[Z_1 > s])]}{(1 - G(s))(1 - F(s -))} (1 - G(s))(1 - F(s -)) \frac{dG(s)}{1 - G(s -)} \\
&= -\text{Cov}(M_{2n}^1(v), M_{2n}^2(v)),
\end{align*}\]

where \( F_{V,X}(s) = P(Y \geq s|V, X) \), we have used condition B(i) that \( G \) is continuous. Letting \( v \to \tau_F \) gives \( E\{M_{1n}^1 M_{2n}^2(\tau_F)\} = -\text{Cov}(M_{2n}^1(\tau_F), M_{2n}^2(\tau_F)) \). Therefore, by (A.46), we have shown that

\[n^{-1/2} \sum_{i=1}^n [\Psi_i - E(\Psi_i|U_i)] e_i \hat{G} = M_{1n} + M_{2n}(\tau_F) + o_p(1), \quad (A.48)\]

and

\[\lim_{n \to \infty} E[(M_{1n} + M_{2n}(\tau_F))^\otimes 2] = \Omega - \Xi(\tau_F), \quad (A.49)\]

where \( M_{1n} = (M_{1n}^1, \ldots, M_{1n}^{p+q})^T \) and \( M_{2n}(\tau_F) = (M_{2n}^1(\tau_F), \ldots, M_{2n}^{p+q}(\tau_F))^T \). Theorem 2 is proved by the central limit theorem for sums of independent random vectors.

**References**