Semiparametric Estimation of Single-Index Transition Intensities

Gørgens, Jakob Tue

DISCUSSION PAPERS
Department of Economics
University of Copenhagen

99-25

Semiparametric Estimation of Single-Index Transition Intensities

Tue Gørgens

Studiestræde 6, DK-1455 Copenhagen K., Denmark
Tel. +45 35 32 30 82 - Fax +45 35 32 30 00
http://www.econ.ku.dk

Semiparametric Estimation of Single-Index Transition Intensities

Tue Gørgens*
School of Economics
University of New South Wales
Sydney 2052 Australia
E-mail: T.Gorgens@unsw.edu.au

December 7, 1999

Abstract

This research develops semiparametric kernel-based estimators of state-specific conditional transition intensities, $h_s(y|x)$, for duration models with right-censoring and/or multiple destinations (competing risks). Both discrete and continuous duration data are considered. The maintained assumptions is that $h_s(y|x)$ depends on x only through an index $x'\beta_s$. In contrast to existing semiparametric estimators, proportional intensities is not assumed. The new estimators are asymptotically normally distributed. The estimator of β_s is root-n consistent. The estimator of $h_s(y|x)$ achieves the one-dimensional rate of convergence. Thus the single-index assumption eliminates the "curse of dimensionality". The estimators perform well in Monte Carlo experiments.

Keywords: Semiparametric estimation, kernel regression, duration analysis, competing risks, censoring.

JEL classification numbers: C14, C24, C41

*The author acknowledges the hospitality of the economics departments of the University of Aarhus and the University of Copenhagen, where some of this research was undertaken.

1 Introduction

Duration models with multiple destinations, also known as competing risks models, have long been one of the principal tools of applied econometrics and other fields such as biostatistics. These models seek to estimate changes in transition intensities as time progresses (duration dependence) and differences in transition intensities across subpopulations or treatments. This paper develops flexible semiparametric estimators of conditional transition intensities, assuming only that the intensities depend on regressors through a single index. In particular, it is not assumed, as the literature generally has, that duration dependence is separable from the effects of regressors.

Empirical durations model have traditionally been estimated by the maximum likelihood method. However, nonlinearities and asymmetries in the distribution of duration data complicate the choice of appropriate parametric functional forms, and extensive specification searches are often necessary. The abundance of parametric families used in duration models attests to the difficulty of finding appropriate forms.[1] Over the past two or three decades, growing concern about assumed parametric forms has lead to the development of nonparametric and semiparametric methods, which allow estimation of transition intensities under much weaker assumptions. Fully nonparametric estimators impose only mild smoothness assumptions.[2] Unfortunately, these estimators are notoriously unreliable because of the now-familiar "curse of dimensionality" and have not gained widespread acceptance.

Most duration models used in econometrics today are so-called proportional intensities models, which assume multiplicative separability between duration dependence and the effect of regressors. In other words, it is assumed that the transition intensity can be factored into a baseline intensity and a scale term, where the baseline intensity depends on duration but not on regressors and the scale term depends on regressors but not on duration. No economic justification for the proportionality restriction has (yet) been found, and the popularity of proportional intensities models is due to the fact that Cox (1972, 1975) developed a semiparametric estimator which allows the baseline intensities to be estimated without restricting their shape to particular parametric forms. Usually a parametric form is assumed for the scale term, the most popular of which is an exponential transformation of a single (linearly additive) index of the regressors. Nonparametric estimators of the scale term have been proposed by Hastie and Tibshirani (1986, 1990a), who considered a proportional intensities model where the scale term is an exponential transformation of a generalized (nonlinear) additive form.[3] By assuming that the regressors appear in an index form, the curse of dimensionality is eliminated. These estimators are an attractive compromise

[1] See for example chapter 3 of Lancaster (1990).

[2] Nonparametric estimators of conditional (integrated) transition intensities given regressors have been analyzed by for example Beran (1981), Dabrowska (1987), McKeague and Utikal (1990), and Nielsen and Linton (1995).

[3] See also chapter 8 of Hastie and Tibshirani (1990b) and the references therein.

between flexibility and reliability in applications where proportional intensities is a reasonable assumption.

This paper develops semiparametric estimators of transition intensities without assuming proportionality. The only maintained assumption is that the conditional transition intensities depend on regressors through a single index. The paper thus extends Cox's model in a different direction than do Hastie and Tibshirani (1986) and does not consider their generalized additive form. The single-index restriction is a valuable means of reducing the dimension of the regressors, given that the tighter structure (when valid) implies a faster rate of convergence to the asymptotic distribution and better performance in applications with moderate sample sizes and many regressors.

To fix ideas, suppose occupation of a given initial state can end with exit to one of several possible destination states. If the data is subject to censoring, the event of censoring is treated as a separate exit state without loss of generality. Let Y denote the duration of stay in the initial state, S the destination state, and X a q-vector of regressors. Let $h_s(y|x)$ be the conditional transition intensity of exiting to state s at time y given $X = x$ and let $H_s(y|x)$ be the corresponding integrated transition intensity. The maintained assumption is that there exist a vector β_s and functions $\hat{h}_s$ and $\hat{H}_s$ such that $h_s(y|x) = \hat{h}_s(y|x'\beta_s)$ and $H_s(y|x) = \hat{H}_s(y|x'\beta_s)$. This paper proposes estimators of β_s, $\hat{h}_s$, and $\hat{H}_s$. The estimation procedure consists of several steps. In the first step, the index coefficients are estimated for the continuous regressors. In the second step, the index coefficients are estimated for the discrete regressors. The third step is to estimate $\hat{h}_s$ and $\hat{H}_s$ using standard nonparametric kernel estimation techniques and $x'\beta_{sn}$ as a proxy for $x'\beta_s$, where β_{sn} denotes an estimator of β_s.

Uniform consistency and asymptotic normality are established in several theorems. The index coefficient estimator is shown to be root-n consistent. The rates of convergence of the transition intensity estimators are found to be independent of the dimension of X, and thus the curse of dimensionality is eliminated. The properties of the estimators are also investigated through Monte Carlo experimentation. The results suggest that the estimators perform well in small samples.

Existing estimators of index coefficients could be used in applications with uncensored single-destination data and applications with right-censored single-destination data when the censoring mechanism is independent of X.[4] However, the literature currently contains no semiparametric estimators for multiple-destination data, nor for single-destination data when the censoring mechanism depends on X. This research fills the gap by extending Powell, Stock, and Stoker's (1989) average derivative estimator and Horowitz and Härdle's (1996) direct semiparametric estimator to multiple-destination data.

Right-censored single-destination data where the censoring mechanism depends on the regressors is a special case of multiple-destination data, given that right-censoring formally is equivalent to an exit state. Therefore the new estimator of β_s has

[4]References to existing estimators are given in section 2.

wider applications in the literature on right-censored data. For example, it can also be used in the first stage of Gørgens and Horowitz' (1999) semiparametric estimator of the censored transformation (GAFT) model and Horowitz's (1999) semiparametric estimator of the mixed proportional intensities model.

The paper is organized as follows. Section 2 considers estimation of index coefficients. Section 3 discusses estimation of transition intensities. Monte Carlo results are presented in section 4 and conclusions in section 5. The appendices contain proofs of the theorems.

To estimate the transition intensity to a given state, the distinction between other states is not necessary. Therefore the subscript s is suppressed in the rest of the paper. Complicated notation is unavoidable, unfortunately, because of the need to distinguish between the first and the remaining components of the regressors, between continuous and discrete regressors, between linear combinations of continuous and discrete regressors, and between similarly defined subsets of index coefficients. Hopefully, the notation chosen is the least confusing. Throughout the paper v_1 and v_{-1} denote the first component and the vector of remaining components of any vector v. A bar indicates a variable related to the continuous regressors and double-dots variables related to the discrete regressors. In particular, $X = (\bar{X}', \ddot{X}')'$, where $\bar{X}$ the a $\bar{q}$-vector of continuous random variables and $\ddot{X}$ is the $\ddot{q}$-vector of discrete random variables. Similarly, partition x and β similarly so that $x'\beta = \bar{x}'\bar{\beta} + \ddot{x}'\ddot{\beta}$. Finally, define the random variables $Z = X'\beta$ and $\bar{Z} = \bar{X}'\bar{\beta}$.

2 Index Coefficient Estimation

Define the conditional distribution functions

$$F_1(y|x) = \Pr(Y \leq y, S = s|X = x), \tag{1}$$
$$F_2(y|x) = \Pr(Y \geq y|X = x). \tag{2}$$

By definition the integrated transition intensity is

$$H(y|x) = \int_0^y \frac{F_1(dv|x)}{F_2(v|x)}. \tag{3}$$

The key assumption of this paper is assumption 1.1. It is easy to show that if the transition intensity h is of single-index form then so is its integral H and vice versa.

Assumption 1

1. *There is a function $\hat{H}$ and a vector β such that $H(y|x) = \hat{H}(y|x'\beta)$ for all y and x.*

2. *There is at least one continuous regressor, so $\bar{q} \geq 1$.*

3. *The index $x'\beta$ contains no constant term and $\beta_1 = 1$.*

Assumption 1.2 is a necessary condition for identification, as shown by Ichimura (1993) among others. Assumption 1.3 contains location and scale normalizations, which are needed because the addition of any constant to $x'\beta$ or the multiplication by any constant can be subsumed into $\hat{H}$. Since $\bar{q} \geq 1$, the assumption that β_1 can be scaled to 1 is not restrictive.

A number of methods for estimating index coefficients in regression models have been proposed, including for example average derivatives (Härdle and Stoker 1989 and Powell, Stock, and Stoker 1989) for continuous regressors, direct semiparametric estimation for discrete regressors (Horowitz and Härdle 1996), semiparametric least squares (Ichimura 1993), maximum rank correlation (Han 1987 and Sherman 1993), and semiparametric maximum likelihood (Ai 1997). Any of these methods can be used here provided F_1 and F_2 are single-index functions. However, in general F_1 and F_2 are not single-index functions, even under assumption 1. Assuming that the density $\partial_1 F_1$ corresponding to F_1 exists everywhere, this can be seen from the familiar results that

$$F_2(y|x) = \exp\bigl(-H_0(y|x) - \hat{H}(y|x'\beta)\bigr) \tag{4}$$

and

$$\partial_1 F_1(y|x) = \hat{h}(y|x'\beta)\exp\bigl(-H_0(y|x) - \hat{H}(y|x'\beta)\bigr), \tag{5}$$

where H_0 denotes the conditional integrated intensity for transition to any destination other than s. In general, the transition intensities for the various destinations will not depend on the same index; indeed $H_0(y|x)$ may not satisfy any single-index restriction. Therefore, F_1 and F_2 will be single-index functions only in special cases, such as uncensored single-destination data where $H_0(y|x) = 0$ everywhere, or right-censored single-destination data where the censoring mechanism is independent of the regressors, that is, $H_0(y|x)$ is independent of x. In case of censored single-destination data where the censoring mechanism depends on the regressors or in case of multiple-destination data, it is unlikely that F_1 or F_2 are single-index functions, and existing estimators do not apply.

In the remainder of section 2 an estimator of β is proposed, which is applicable to multiple-destination data.

2.1 Estimation of $\bar{\beta}$

Let $\mathcal{X} = \{\chi_1, \ldots, \chi_m\}$ be the support of $\ddot{X}$, and let ξ denote the density of $\bar{X}$. Define

$$\bar{A}_1(y, \bar{x}, \ddot{x}) = \Pr(Y \leq y, S = s, \ddot{X} = \ddot{x}|\bar{X} = \bar{x})\xi(\bar{x}), \tag{6}$$

$$\bar{A}_2(y, \bar{x}, \ddot{x}) = \Pr(Y \geq y, \ddot{X} = \ddot{x}|\bar{X} = \bar{x})\xi(\bar{x}). \tag{7}$$

Writing $H(y|(\bar{x}', \ddot{x}')')$ as $H(y|\bar{x}, \ddot{x})$ for simplicity, note that by equation (3)

$$H(y|\bar{x}, \ddot{x}) = \int_0^y \frac{\bar{A}_1(dv, \bar{x}, \ddot{x})}{\bar{A}_2(v, \bar{x}, \ddot{x})}. \tag{8}$$

The new estimator of $\bar{\beta}$ is similar to the weighted average derivative estimator by Powell, Stock, and Stoker (1989). The idea is simple. If the conditional transition intensity is of single-index form, then[5] $\partial_{\bar{x}} H(dy|\bar{x}, \ddot{x}) = \partial_2 \hat{H}(dy|\bar{x}'\bar{\beta} + \ddot{x}'\ddot{\beta})\bar{\beta}$. Let W be a weight function. Then $\bar{\beta}$ is proportional to $\bar{\beta}^*$ defined by[6]

$$\bar{\beta}^* = \sum_{l=1}^{m} \iint W(y, \bar{x}, \chi_l) \partial_{\bar{x}} H(dy|\bar{x}, \chi_l)\, d\bar{x}, \tag{9}$$

provided

$$\sum_{l=1}^{m} \iint W(y, \bar{x}, \chi_l) \partial_2 \hat{H}(dy|\bar{x}'\bar{\beta} + \chi_l'\ddot{\beta})\, d\bar{x} \tag{10}$$

is finite and nonzero. Furthermore, $\bar{\beta}_{-1} = \bar{\beta}^*_{-1}/\bar{\beta}^*_1$ by the normalization in assumption 1.3. An estimator is defined below by replacing $\partial_{\bar{x}} H$ in (9) with a nonparametric estimator. This paper considers the case where the weight function W is

$$W(y, \bar{x}, \ddot{x}) = \bar{A}_2(y, \bar{x}, \ddot{x})^2 w(y, \bar{x}, \ddot{x}), \tag{11}$$

where w is another weight function. This choice is convenient because it avoids random denominators in the estimation formula for $\partial_{\bar{x}} H(dy|\bar{x}, \ddot{x})$. Since

$$\partial_{\bar{x}} H(dy|\bar{x}, \ddot{x}) = \frac{\partial_{\bar{x}} \bar{A}_1(dy, \bar{x}, \ddot{x})}{\bar{A}_2(y, \bar{x}, \ddot{x})} - \frac{\partial_{\bar{x}} \bar{A}_2(y, \bar{x}, \ddot{x}) \bar{A}_1(dy, \bar{x}, \ddot{x})}{\bar{A}_2(y, \bar{x}, \ddot{x})^2}, \tag{12}$$

it follows that

$$\bar{\beta}^* = \sum_{l=1}^{m} \iint w(y, \bar{x}, \chi_l) \bar{A}_2(y, \bar{x}, \chi_l) \partial_{\bar{x}} \bar{A}_1(dy, \bar{x}, \chi_l)\, d\bar{x} - \sum_{l=1}^{m} \iint w(y, \bar{x}, \chi_l) \partial_{\bar{x}} \bar{A}_2(y, \bar{x}, \chi_l) \bar{A}_1(dy, \bar{x}, \chi_l)\, d\bar{x}. \tag{13}$$

Choosing the weight function w is not complicated. The main purpose of the weight function is to provide a way of ensuring that the expression in (10), under definition (11), is finite and nonzero. In the Monte Carlo experiments presented in section 4 the weight function is simply $w(y, \bar{x}, \ddot{x}) = 1(y \leq y_{.85})(1/6)$, where $y_{.85}$ is the 85% quantile of the conditional distribution of Y given $S = s$.

The estimator proposed here consists of replacing unknown functions in (13) by sample analogs based on kernel estimation. The sample available for analysis is

[5] If f is a function, let $\partial_i^j f$ denote the jth order partial derivative of f with respect to its ith argument. With a bit notational abuse, let $\partial_{\bar{x}}^j f$ denote the $\bar{q}$-vector of jth order partial derivatives with respect to whichever argument represents the $\bar{q}$-vector of continuous regressors.

[6] Throughout the paper, the range is not indicated whenever integration is over an entire Euclidean space.

assumed to consist of n independent observations $(Y_i, S_i, \bar{X}_i', \ddot{X}_i')'$, $i = 1, 2, \ldots, n$. Let $\bar{b}$ be a bandwidth parameter, and let $\bar{K} : R^{\bar{q}} \to R$ be a kernel function. Define $\bar{K}_{\bar{b}}(\bar{x}) = \bar{b}^{-\bar{q}}\bar{K}(\bar{b}^{-1}\bar{x})$. Then define the estimators

$$\bar{A}_{1n}(y, \bar{x}, \ddot{x}) = \frac{1}{n}\sum_{i=1}^{n} \bar{K}_{\bar{b}}(\bar{x} - \bar{X}_i)1(\ddot{X}_i = \ddot{x})1(Y_i \leq y)1(S_i = s), \tag{14}$$

$$\bar{A}_{2n}(y, \bar{x}, \ddot{x}) = \frac{1}{n}\sum_{i=1}^{n} \bar{K}_{\bar{b}}(\bar{x} - \bar{X}_i)1(\ddot{X}_i = \ddot{x})1(Y_i \geq y). \tag{15}$$

The estimator of $\bar{\beta}^*$ is

$$\begin{aligned}
\bar{\beta}_n^* &= \sum_{l=1}^{m} \iint w(y, \bar{x}, \chi_l)\bar{A}_{2n}(y, \bar{x}, \chi_l)\partial_{\bar{x}}\bar{A}_{1n}(dy, \bar{x}, \chi_l)\, d\bar{x} \\
&\quad - \sum_{l=1}^{m} \iint w(y, \bar{x}, \chi_l)\partial_{\bar{x}}\bar{A}_{2n}(y, \bar{x}, \chi_l)\bar{A}_{1n}(dy, \bar{x}, \chi_l)\, d\bar{x} \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{m} \int \partial_{\bar{x}}\bar{K}_{\bar{b}}(\bar{x} - \bar{X}_i)\bar{K}_{\bar{b}}(\bar{x} - \bar{X}_j) \\
&\qquad \times w(Y_i, \bar{x}, \chi_l)1(\ddot{X}_i = \chi_l)1(\ddot{X}_i = \ddot{X}_j)1(Y_j \geq Y_i)1(S_i = s)\, d\bar{x} \\
&\quad - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{m} \int \bar{K}_{\bar{b}}(\bar{x} - \bar{X}_i)\partial_{\bar{x}}\bar{K}_{\bar{b}}(\bar{x} - \bar{X}_j) \\
&\qquad \times w(Y_i, \bar{x}, \chi_l)1(\ddot{X}_i = \chi_l)1(\ddot{X}_i = \ddot{X}_j)1(Y_j \geq Y_i)1(S_i = s)\, d\bar{x}.
\end{aligned} \tag{16}$$

Given $\bar{\beta}_n^*$, estimate $\bar{\beta}_{-1n}$ by $\bar{\beta}_{-1n}^*/\bar{\beta}_{1n}^*$. Uniform consistency and asymptotic normality of $\bar{\beta}_n^*$ and $\bar{\beta}_{-1n}$ are established in theorem 1 below. Computing $\bar{\beta}_n^*$ involves evaluating a $\bar{q}$-dimensional integral. It is possible to simplify this to $\bar{q}$ one-dimensional integrals with closed-form solutions by using a polynomial product kernel and weight function as in the Monte Carlo experiments in section 4.

Let P denote the distribution of $(Y, S, \bar{X}', \ddot{X}')'$, and let P_n denote the empirical measure formed from the n independent observations on P, that is, P_n puts probability $1/n$ on each of the observations.[7] Assumption 2 defines the sample.

Assumption 2 *The sequence* $\{(Y_i, S_i, \bar{X}_i', \ddot{X}_i')'\}_{i=1}^n$ *is a random sample from* P.

The derivation of the limiting distribution depends on applications of the mean value theorem and Taylor series expansions. Hence, the underlying functions must be smooth. Sufficient conditions are listed as assumption 3.[8]

---

[7] Linear functional notation is used throughout this paper. The expected value of a random variable V is denoted $\mathbb{E}V$, $Pf(t) = \mathbb{E}f(Y, S, \bar{X}, \ddot{X}, t)$, and $P_nf(t) = n^{-1}\sum_{i=1}^{n} f(Y_i, S_i, \bar{X}_i, \ddot{X}_i, t)$.

[8] For simplicity of exposition, derivatives are assumed to exist everywhere on the domain of the original functions. The result of theorem 1 continues to hold even if a function is not differentiable everywhere, provided w is chosen to avoid "edge-effects" in the kernel smoothing. That is, if the kernel estimates involve smoothing over $\bar{X}_i$ near $\bar{x}$ then $\bar{A}_1(y, \cdot, \ddot{x})$ $\bar{A}_2(y, \cdot, \ddot{x})$ must be smooth on $[\bar{x} - \bar{b}, \bar{x} + \bar{b}]$ for all $\bar{b}$ small. Similar remarks apply to the other theorems in the paper.

Assumption 3 *For $\bar{k} \in N_+$ given below:*

1. *The $\bar{q}$-vector $\bar{X}$ is absolutely continuous and has density ξ with respect to Lebesgue measure.*
2. *ξ is bounded.*
3. *$\int \left|\partial_{\bar{x}}^j \bar{A}_1(dy,\cdot,\cdot)\right|$ exist and are bounded and continuous for $j = 1, \ldots, 1+\bar{k}$.*
4. *$\partial_{\bar{x}}^j \bar{A}_2$ exist and are bounded and continuous for $j = 1, \ldots, 1+\bar{k}$.*

A researcher who wishes to use the estimators must choose a weight function, a bandwidth, and a kernel function. To establish consistency and asymptotic normality, it is necessary to restrict the choices. Sufficient conditions are given in assumptions 4 and 5 and in the theorem itself.

Assumption 4 *The weight function $w : R^{1+\bar{q}+\ddot{q}} \to R$ satisfies:*

1. *w is bounded.*
2. *The expression given in (10), with definition (11), is finite and nonzero.*
3. *$\partial_{\bar{x}} w$ and $\partial_{\bar{x}}^2 w$ exist and are bounded.*

Assumption 5 *For $\bar{k} \in N_+$ given below, the kernel function $\bar{K} : R^{\bar{q}} \to R$ satisfies:*

1. *$\bar{K}$ is a bounded $\bar{k}$-order kernel with support $[-1,1]^{\bar{q}}$.*
2. *$\partial_{\bar{x}} \bar{K}$ exists and is bounded and continuous on $R^{\bar{q}}$.*

These are standard assumptions in the literature on semiparametric estimation. In order to state the theorem, define[9]

$$\begin{aligned}
&\Phi_{\bar{\beta}^*}(\mathbf{y},\mathbf{s},\bar{\mathbf{x}},\ddot{\mathbf{x}}) \\
&\quad = 2\int w(v,\bar{\mathbf{x}},\ddot{\mathbf{x}})1(\mathbf{y} \geq v)\partial_{\bar{x}}\bar{A}_1(dv,\bar{\mathbf{x}},\ddot{\mathbf{x}}) - 2w(\mathbf{y},\bar{\mathbf{x}},\ddot{\mathbf{x}})\partial_{\bar{x}}\bar{A}_2(\mathbf{y},\bar{\mathbf{x}},\ddot{\mathbf{x}})1(\mathbf{s}=s) \\
&\qquad + \int \partial_{\bar{x}} w(v,\bar{\mathbf{x}},\ddot{\mathbf{x}})1(\mathbf{y} \geq v)\bar{A}_1(dv,\bar{\mathbf{x}},\ddot{\mathbf{x}}) - \partial_{\bar{x}} w(\mathbf{y},\bar{\mathbf{x}},\ddot{\mathbf{x}})\bar{A}_2(\mathbf{y},\bar{\mathbf{x}},\ddot{\mathbf{x}})1(\mathbf{s}=s) \\
&\qquad - 2\bar{\beta}^*
\end{aligned} \tag{17}$$

and, letting $\Phi_{1\bar{\beta}^*}(\mathbf{y},\mathbf{s},\bar{\mathbf{x}},\ddot{\mathbf{x}})$ and $\Phi_{-1\bar{\beta}^*}(\mathbf{y},\mathbf{s},\bar{\mathbf{x}},\ddot{\mathbf{x}})$ denote the first and the remaining components of the $\bar{q}$-vector $\Phi_{\bar{\beta}^*}(\mathbf{y},\mathbf{s},\bar{\mathbf{x}},\ddot{\mathbf{x}})$, define

$$\Phi_{\bar{\beta}_{-1}}(\mathbf{y},\mathbf{s},\bar{\mathbf{x}},\ddot{\mathbf{x}}) = \frac{\Phi_{-1\bar{\beta}^*}(\mathbf{y},\mathbf{s},\bar{\mathbf{x}},\ddot{\mathbf{x}}) - \bar{\beta}_{-1}\Phi_{1\bar{\beta}^*}(\mathbf{y},\mathbf{s},\bar{\mathbf{x}},\ddot{\mathbf{x}})}{\bar{\beta}_1^*}. \tag{18}$$

It is straightforward to verify that $P\Phi_{\bar{\beta}^*} = 0$ and $P\Phi_{\bar{\beta}_{-1}} = 0$. Define $\Sigma_{\bar{\beta}^*} = P\Phi_{\bar{\beta}^*}\Phi'_{\bar{\beta}^*}$ and $\Sigma_{\bar{\beta}_{-1}} = P\Phi_{\bar{\beta}_{-1}}\Phi'_{\bar{\beta}_{-1}}$.

[9] Throughout the paper boldface lowercase letters are used as placeholders and integration dummies for the corresponding uppercase random variables.

Theorem 1 *Suppose assumptions 1, 2, 3, 4, and 5 hold. Then:*

i. *If* $n\bar{b}^{2\bar{q}+2} \to \infty$, *then* $|\bar{\beta}_n^* - \mathbb{E}\bar{\beta}_n^* - P_n\Phi_{\bar{\beta}^*}| = o_p\big(n^{-1/2}\big)$ *and* $n^{1/2}(\bar{\beta}_n^* - \mathbb{E}\bar{\beta}_n^*) \to^d N(0, \Sigma_{\bar{\beta}^*})$ *as* $n \to \infty$.

ii. *If* $n\bar{b}^{2\bar{k}} \to 0$, *then* $n^{1/2}(\mathbb{E}\bar{\beta}_n^* - \bar{\beta}^*) \to 0$ *as* $n \to \infty$.

iii. *If* $n\bar{b}^{2\bar{q}+2} \to \infty$ *and* $n\bar{b}^{2\bar{k}} \to 0$, *then* $|\bar{\beta}_n^* - \bar{\beta}^* - P_n\Phi_{\bar{\beta}^*}| = o_p\big(n^{-1/2}\big)$ *and* $n^{1/2}(\bar{\beta}_n^* - \bar{\beta}^*) \to^d N(0, \Sigma_{\bar{\beta}^*})$ *as* $n \to \infty$.

Suppose the above-mentioned assumptions hold and that $n\bar{b}^{2\bar{q}+2} \to \infty$ *and* $n\bar{b}^{2\bar{k}} \to 0$ *as* $n \to \infty$. *Then:*

iv. $|\bar{\beta}_{-1n} - \bar{\beta}_{-1} - P_n\Phi_{\bar{\beta}_{-1}}| = o_p\big(n^{-1/2}\big)$ *as* $n \to \infty$.

v. $n^{1/2}(\bar{\beta}_{-1n} - \bar{\beta}_{-1}) \to^d N(0, \Sigma_{\bar{\beta}_{-1}})$ *as* $n \to \infty$.

Given the approximation results in part i of the theorem, asymptotic normality in part ii follows from the Lindeberg-Lévy central limit theorem and the Cramér-Wold theorem. Part iii follows immediately from parts i and ii. Parts iv and v follow by applying the delta method to part iii.

The most important conclusions of theorem 1 are that $\bar{\beta}_n$ converges at the root-n rate, which is the familiar rate from parametric estimation, and that $\bar{\beta}_n$ is asymptotically normally distributed. These nice properties are not unexpected, since they are shared with the index coefficient estimators listed in the beginning of section 2.

Before turning to estimation of $\ddot{\beta}$ is it worth pointing out that the condition $n\bar{b}^{2\bar{q}+2} \to \infty$ in the theorem is determined by the "diagonal" terms where $i = j$ in (16). Examination of the proof of the theorem shows that if these diagonal terms were omitted, the condition $n\bar{b}^{2\bar{q}+2} \to \infty$ could be weakened to $n\bar{b}^{\bar{q}+2} \to \infty$, which is the same as the requirement in Powell, Stock, and Stoker (1989).

2.2 Estimation of $\ddot{\beta}$

The method proposed here for estimating the coefficients of the discrete regressors is similar to Horowitz and Härdle's (1996) direct semiparametric estimator. Horowitz and Härdle were concerned with estimating index coefficients for a single-index regression function. Their ideas are here adapted to estimating $\ddot{\beta}$ from $H(\mu|\bar{x}, \ddot{x}) = \hat{H}(\mu|\bar{x}'\bar{\beta} + \ddot{x}'\ddot{\beta})$, where μ is some appropriately chosen constant.

Assume in this section that $\ddot{q} \geq 1$, so that some of the regressors are discrete. Define $\bar{Z} = \bar{X}'\bar{\beta}$ and let $\bar{\zeta}$ denote the density of $\bar{Z}$. Given μ, define G by

$$G(\bar{z}, \ddot{x}) = \int_0^\mu \frac{\ddot{A}_1(dv, \bar{z}, \ddot{x})}{\ddot{A}_2(v, \bar{z}, \ddot{x})}, \tag{19}$$

where

$$\ddot{A}_1(y, \bar{z}, \ddot{x}) = \Pr(Y \leq y, S = s, \ddot{X} = \ddot{x}|\bar{Z} = \bar{z})\bar{\zeta}(\bar{z}), \tag{20}$$

$$\ddot{A}_2(y, \bar{z}, \ddot{x}) = \Pr(Y \geq y, \ddot{X} = \ddot{x} | \bar{Z} = \bar{z})\bar{\zeta}(\bar{z}), \tag{21}$$

and note that the single-index assumption 1.1 implies

$$G(\bar{z}, \ddot{x}) = \hat{H}(\mu | \bar{z} + \ddot{x}'\ddot{\beta}). \tag{22}$$

To describe the idea of the estimator, assume, temporarily, that $G(\cdot, \ddot{x})$ is known and monotone. Then to estimate $\ddot{\beta}$, fix $\bar{z}$ and find the unique set of numbers $\Delta_2, \ldots, \Delta_m$ such that

$$G(\bar{z}, \chi_1) = G(\bar{z} + \Delta_2, \chi_2) = \cdots = G(\bar{z} + \Delta_m, \chi_m). \tag{23}$$

By equation (22), the numbers satisfy the equations $\Delta_l = (\chi_1' - \chi_l')\ddot{\beta}$, $l = 2, \ldots, m$, and, subject to a nonsingularity condition, it is straightforward to solve these $m-1$ linear equations for $\ddot{\beta}$. In the case where G is unknown (but known to be monotone), an estimator could be defined by replacing G in these arguments by a kernel estimator G_n. The only caveat is that G must be estimable at each of the points $\bar{z}$ and $\bar{z} + \Delta_l$, $l = 2, \ldots, m$. In other words, these points must be in the support of $\bar{Z}$.

Horowitz and Härdle (1996) show that it is possible to estimate $\ddot{\beta}$ under a much weaker monotonicity assumption. Specifically, assume that there is an interval $[\pi_0, \pi_1]$ in the support of $\bar{Z}$ and another interval $[c_0, c_1]$ such that $G(\bar{z}, \ddot{x}) < c_0$ whenever $\bar{z} < \pi_0$ and $G(\bar{z}, \ddot{x}) > c_1$ whenever $\bar{z} > \pi_1$. Let $J(\ddot{x})$ be the integral over $[\pi_0, \pi_1]$ of $G(\cdot, \ddot{x})$ "clipped" at c_0 and c_1. That is, define

$$J(\ddot{x}) = \int_{\pi_0}^{\pi_1} \Big(c_0 1\big(G(\bar{z}, \ddot{x}) < c_0\big) + c_1 1\big(G(\bar{z}, \ddot{x}) > c_1\big) + G(\bar{z}, \ddot{x}) 1\big(c_0 \leq G(\bar{z}, \ddot{x}) \leq c_1\big)\Big)\, d\bar{z}. \tag{24}$$

Using simple geometric arguments it can be shown that (Horowitz and Härdle, lemma 1)

$$J(\chi_l) - J(\chi_1) = (c_1 - c_0)(\chi_l' - \chi_1')\ddot{\beta}, \qquad l = 2, \ldots, m, \tag{25}$$

which constitutes $m-1$ linear equations in the $\ddot{q}$ unknown components of $\ddot{\beta}$. If a unique solution exists, these may be solved for $\ddot{\beta}$. To do so, define the $(m-1)$-vector

$$\Delta J = \begin{bmatrix} J(\chi_2) - J(\chi_1) \\ \vdots \\ J(\chi_m) - J(\chi_1) \end{bmatrix} \tag{26}$$

and the $(m-1) \times \ddot{q}$-matrix

$$W = \begin{bmatrix} \chi_2' - \chi_1' \\ \vdots \\ \chi_m' - \chi_1' \end{bmatrix}. \tag{27}$$

If $W'W$ is nonsingular, then

$$\ddot{\beta} = (c_1 - c_0)^{-1}(W'W)^{-1}W'\Delta J. \tag{28}$$

To estimate $\ddot{\beta}$, replace the unknown G by a nonparametric estimator. Many estimators could be used; the estimator used here is a kernel estimator.

Define $\bar{Z}_{in} = \bar{X}_i'\bar{\beta}_n$, where $\bar{\beta}_n$ is an estimator of $\bar{\beta}$ (for example the estimator developed in the previous section). Let $\ddot{b}$ be a bandwidth parameter, and let $\ddot{K} : R \to R$ be a kernel function. Define $\ddot{K}_{\ddot{b}}(\bar{z}) = \ddot{b}^{-1}\ddot{K}(\ddot{b}^{-1}\bar{z})$ and the estimators

$$\ddot{A}_{1n}(y, \bar{z}, \ddot{x}) = \frac{1}{n}\sum_{i=1}^{n} \ddot{K}_{\ddot{b}}(\bar{z} - \bar{Z}_{in})1(\ddot{X}_i = \ddot{x})1(Y_i \leq y)1(S_i = s), \tag{29}$$

$$\ddot{A}_{2n}(y, \bar{z}, \ddot{x}) = \frac{1}{n}\sum_{i=1}^{n} \ddot{K}_{\ddot{b}}(\bar{z} - \bar{Z}_{in})1(\ddot{X}_i = \ddot{x})1(Y_i \geq y), \tag{30}$$

and

$$G_n(\bar{z}, \ddot{x}) = \int_0^\mu \frac{\ddot{A}_{1n}(dv, \bar{z}, \ddot{x})}{\ddot{A}_{2n}(v, \bar{z}, \ddot{x})} = \frac{1}{n}\sum_{i=1}^{n} \frac{\ddot{K}_{\ddot{b}}(\bar{z} - \bar{Z}_{in})1(\ddot{X}_i = \ddot{x})1(Y_i \leq \mu)1(S_i = s)}{\ddot{A}_{2n}(Y_i, \bar{z}, \ddot{x})}. \tag{31}$$

Let J_n be J with G_n replacing G in (24), let ΔJ_n be ΔJ with J_n replacing J in (26), and define the estimator

$$\ddot{\beta}_n = (c_1 - c_0)^{-1}(W'W)^{-1}W'\Delta J_n. \tag{32}$$

Convergence of $\ddot{\beta}_n$ to $\ddot{\beta}$ is established in theorem 2 below.

Assumption 6.1 contains the weak monotonicity condition used to identify $\ddot{\beta}$.

Assumption 6 *Suppose $\ddot{q} \geq 1$. Fix μ and suppose there are intervals $[\pi_0, \pi_1]$ and $[c_0, c_1]$ such that:*

1. *$\hat{H}(\mu|\bar{z} + \ddot{x}'\ddot{\beta}) < c_0$ whenever $\bar{z} < \pi_0$ and $\hat{H}(\mu|\bar{z} + \ddot{x}'\ddot{\beta}) > c_1$ whenever $\bar{z} > \pi_1$ for all $\ddot{x} \in \mathcal{X}$.*
2. *$\hat{H}(\mu|\bar{z} + \ddot{x}'\ddot{\beta})$ equals c_0 or c_1 at only finitely many values of $\bar{z}$ in $[\pi_0, \pi_1]$.*

Assumption 6.2 is a minor technical assumption used in the proofs. The remaining assumptions used to prove asymptotic normality are standard in the literature.

Assumption 7 is used to bound remainder terms and to ensure that the bias vanishes sufficiently quickly. In particular, assumptions 7.3 and 7.4 are used to bound remainder terms arising from the fact that $\bar{\beta}_n$ is random. If necessary, assumption 7.3 can be satisfied by artificially truncating $\bar{X}_{-1}$. Define $\mathcal{M}_D = \{\mu\} \times [\pi_0, \pi_1] \times \mathcal{X}$ and

$$\ddot{A}_0^*(y, s, \bar{z}, \ddot{x}, \bar{x}_{-1}) = \Pr(Y \leq y, S \leq s, \ddot{X} \leq \ddot{x}, \bar{X}_{-1} \leq \bar{x}_{-1}|\bar{Z} = \bar{z})\bar{\zeta}(\bar{z}). \tag{33}$$

Assumption 7.1 ensures that there is sufficient data in $\mathcal{M}_D$ to estimate G.

Assumption 7 *For $\ddot{k} \in N_+$ given below:*

1. *There is a constant $C_4 > 0$ such that $\ddot{A}_2 > C_4$ on $\mathcal{M}_D$.*
2. *$\bar{\zeta}$ is bounded.*
3. *$\bar{X}_{-1}$ is bounded with probability one.*
4. *$\partial_3 \ddot{A}_0^*$ and $\partial_3^2 \ddot{A}_0^*$ exist and are bounded and continuous.*
5. *$\partial_2^j \ddot{A}_1$ and $\partial_2^j \ddot{A}_2$ exist and are bounded and continuous for $j = 1, \ldots, \ddot{k}$.*
6. *$\partial_2 \ddot{A}_2$ exists and is bounded and continuous. There is a constant C_1 such that*

$$\big|\ddot{A}_1(t) - \ddot{A}_1(t^*)\big| \leq C_1 \big\|t - t^*\big\| \quad \text{for all } t, t^* \in \mathcal{M}_D, \tag{34}$$

where t denotes $(y, \bar{z}, \ddot{x})'$.

Assumption 8 is similar to assumption 5; boundedness of the second-order derivative of $\ddot{K}$ is used to handle the fact that $\bar{\beta}_n$ is random.

Assumption 8 *For $\ddot{k} \in N_+$ given below, the kernel function $\ddot{K} : R \to R$ satisfies:*

1. *$\ddot{K}$ is a bounded $\ddot{k}$-order kernel with support $[-1, 1]$.*
2. *$\partial \ddot{K}$ and $\partial^2 \ddot{K}$ exist and are bounded and continuous on R.*

The following assumption concerns the properties of the preliminary estimator of $\bar{\beta}$. It is satisfied by the estimator discussed in the previous section. If $\bar{q} = 1$ or $\bar{\beta}$ is known, this assumption is not needed.

Assumption 9 *The estimator $\bar{\beta}_n$ satisfies:*

1. *By normalization $\bar{\beta}_{1n} = \bar{\beta}_1$.*
2. *There is a function $\Omega : R^{1+1+\bar{q}+\ddot{q}} \to R^{\bar{q}-1}$ such that $P\Omega = 0$, the components of $P\Omega\Omega'$ are finite, and $n^{1/2}(\bar{\beta}_{-1n} - \bar{\beta}_{-1}) = n^{1/2} P_n \Omega + o_p(1)$ as $n \to \infty$.*

In order to state the theorem, define

$$\ddot{f}(\mathbf{y}, \mathbf{s}, \ddot{\mathbf{x}}, \bar{z}, \ddot{x}) = 1(\ddot{\mathbf{x}} = \ddot{x})\left(\frac{1(\mathbf{y} \leq \mu)1(\mathbf{s} = s)}{\ddot{A}_2(\mathbf{y}, \bar{z}, \ddot{x})} - \int_0^{\mu} \frac{1(\mathbf{y} \geq v)\ddot{A}_1(dv, \bar{z}, \ddot{x})}{\ddot{A}_2(v, \bar{z}, \ddot{x})^2}\right), \tag{35}$$

$$\sigma(\mathbf{y}, \mathbf{s}, \ddot{\mathbf{x}}, \bar{z}, \ddot{x}) = 1(\pi_0 \leq \bar{z} \leq \pi_1)1\big(c_0 \leq G(\bar{z}, \ddot{x}) \leq c_1\big)\ddot{f}(\mathbf{y}, \mathbf{s}, \ddot{\mathbf{x}}, \bar{z}, \ddot{x}), \tag{36}$$

and

$$\Gamma(\ddot{x}) = \iiiint\!\!\int \sigma(\mathbf{y}, \mathbf{s}, \ddot{\mathbf{x}}, \bar{z}, \ddot{x})\bar{\mathbf{x}}_{-1}\partial_3 \ddot{A}_0^*(d\mathbf{y}, d\mathbf{s}, \bar{z}, d\ddot{\mathbf{x}}, d\bar{\mathbf{x}}_{-1})\, d\bar{z}. \tag{37}$$

Let $\Phi_{\ddot{\beta}}$ be the $\ddot{q}$-vector whose jth component is

$$\begin{aligned}\big[\Phi_{\ddot{\beta}}(\mathbf{y}, \mathbf{s}, \bar{\mathbf{x}}, \ddot{\mathbf{x}})\big]_j = (c_1 - c_0)^{-1} \sum_{l=2}^{m} \big[(W'W)^{-1}W'\big]_{j,l-1}\Big(\sigma(\mathbf{y}, \mathbf{s}, \ddot{\mathbf{x}}, \bar{\mathbf{x}}'\bar{\beta}, \chi_l) \\ - \sigma(\mathbf{y}, \mathbf{s}, \ddot{\mathbf{x}}, \bar{\mathbf{x}}'\bar{\beta}, \chi_1) + \big(\Gamma(\chi_l) - \Gamma(\chi_1)\big)'\Omega(\mathbf{y}, \mathbf{s}, \bar{\mathbf{x}}, \ddot{\mathbf{x}})\Big).\end{aligned} \tag{38}$$

It can be verified that $P\Phi_{\ddot{\beta}} = 0$. Let $\Sigma_{\ddot{\beta}} = P\Phi_{\ddot{\beta}}\Phi_{\ddot{\beta}}'$.

Theorem 2 *Suppose assumptions 1, 2, 6, 7, 8, and 9 hold with $\ddot{k} \geq 2$ and that $n^{-1/2}\ddot{b}^{-4} \to 0$ and $n^{1/4}\ddot{b}^{\ddot{k}} \to 0$ as $n \to \infty$. Then:*

i. $|\ddot{\beta}_n - \ddot{\beta} - P_n\Phi_{\ddot{\beta}}| = o_p\big(n^{-1/2}\big)$ *as* $n \to \infty$.

ii. $n^{1/2}(\ddot{\beta}_n - \ddot{\beta}) \to^d N(0, \Sigma_{\ddot{\beta}})$ *as* $n \to \infty$.

Theorem 2 confirms the expectation that $\ddot{\beta}_n$ converges at the root-n rate and is asymptotically normally distributed. As with theorem 1, given the approximation results in part i of the theorem, asymptotic normality in part ii follow from the Lindeberg-Lévy central limit theorem and the Cramér-Wold theorem. The bandwidth requirements are satisfied, for example, if $\ddot{b} \propto n^{-1/9}$ and $\ddot{k} = 4$.

Put $\beta_{-1n} = (\bar{\beta}_{-1}{}'_n, \ddot{\beta}'_n)'$ and $\Phi_{\beta_{-1}}(\mathbf{y}, \mathbf{s}, \bar{\mathbf{x}}, \ddot{\mathbf{x}}) = \big(\Phi_{\bar{\beta}_{-1}}(\mathbf{y}, \mathbf{s}, \bar{\mathbf{x}}, \ddot{\mathbf{x}})', \Phi_{\ddot{\beta}}(\mathbf{y}, \mathbf{s}, \bar{\mathbf{x}}, \ddot{\mathbf{x}})'\big)'$, and let $\Sigma_{\beta_{-1}} = P\Phi_{\beta_{-1}}\Phi'_{\beta_{-1}}$.

Theorem 3 *Under the assumptions of theorems 1 and 2:*

i. $|\beta_{-1n} - \beta_{-1} - P_n\Phi_{\beta_{-1}}| = o_p\big(n^{-1/2}\big)$ *as* $n \to \infty$.

ii. $n^{1/2}(\beta_{-1n} - \beta_{-1}) \to^d N(0, \Sigma_{\beta_{-1}})$ *as* $n \to \infty$.

Theorem 3 follows immediately from theorems 1 and 2.

3 Single-Index Transition Intensity Estimation

Define $Z = X'\beta$. As shown earlier, in general F_j and A_j are *not* single-index functions even if H is. It may therefore not be apparent that $\hat{H}(y|x'\beta)$ is not only the conditional integrated transition intensity of Y given $X = x$ but also the conditional integrated transition intensity given $Z = x'\beta$. That is,

$$\int_0^y \frac{F_1(dv|x)}{F_2(v|x)} = H(y|x) = \hat{H}(y|z) = \int_0^y \frac{\hat{F}_1(dv|z)}{\hat{F}_2(v|z)} \quad \text{for } z = x'\beta, \tag{39}$$

where

$$\hat{F}_1(y|z) = \Pr(Y \leq y, S = s|Z = z), \tag{40}$$

$$\hat{F}_2(y|z) = \Pr(Y \geq y|Z = z). \tag{41}$$

To see why, additional notation is needed. Let $\tilde{P}_z$ denote the conditional distribution of X_{-1} given $Z = z$. Define the function τ by

$$\tau(z, x_{-1}) = \left(\frac{z - x_{-1}{}'\beta_{-1}}{\beta_1}, x_{-1}\right). \tag{42}$$

Then $\tau(z, x_{-1})'\beta = z$ and

$$\hat{F}_j(y|z) = \int F_j\big(y|\tau(z, x_{-1})\big)\, \tilde{P}_z(dx_{-1}), \qquad j = 1, 2. \tag{43}$$

By definition $H(dy|x) = F_1(dy|x)/F_2(y|x)$ and by assumption $H(dy|x) = \hat{H}(dy|x'\beta)$. It follows that $F_1(dy|x) = \hat{H}(dy|x'\beta)F_2(y|x)$, whence

$$\begin{aligned}\hat{F}_1(dy|z) &= \int F_1(dy|\tau(z,x_{-1}))\,\tilde{P}_z(dx_{-1}) \\ &= \int \hat{H}(dy|\tau(z,x_{-1})'\beta)F_2(y|\tau(z,x_{-1}))\,\tilde{P}_z(dx_{-1}) \\ &= \hat{H}(dy|z)\int F_2(y|\tau(z,x_{-1}))\,\tilde{P}_z(dx_{-1}) \\ &= \hat{H}(dy|z)\hat{F}_2(y|z), \qquad (44)\end{aligned}$$

and equation (39) follows. A nice implication of (39) is that estimators of $\hat{H}$ and $\hat{h}$ can be based on probabilities conditional on Z instead of X.

3.1 Estimation of $\hat{H}$

First consider the estimation of the integrated transition intensity. Let ζ denote the density of Z, and define $\hat{A}_j(y,z) = \hat{F}_j(y|z)\zeta(z)$ for $j = 1,2$. Then by (39)

$$\hat{H}(y|z) = \int_0^y \frac{\hat{A}_1(dv,z)}{\hat{A}_2(v,z)}. \qquad (45)$$

Define $Z_{in} = X_i'\beta_n$, where β_n is an estimator of β (for example the estimator discussed in the previous section). Let b_z be a bandwidth parameter, and let $K_z : R \to R$ be a kernel function. Define $K_{zn}(z) = b_z^{-1}K_z(b_z^{-1}z)$ and the estimators

$$\hat{A}_{1n}(y,z) = \frac{1}{n}\sum_{i=1}^n K_{zn}(z - Z_{in})1(Y_i \le y)1(S_i = s), \qquad (46)$$

$$\hat{A}_{2n}(y,z) = \frac{1}{n}\sum_{i=1}^n K_{zn}(z - Z_{in})1(Y_i \ge y). \qquad (47)$$

Consider the estimator

$$\hat{H}_n(y|z) = \int_0^y \frac{\hat{A}_{1n}(dv,z)}{\hat{A}_{2n}(v,z)} = \frac{1}{n}\sum_{i=1}^n \frac{K_{zn}(z - Z_{in})1(Y_i \le y)1(S_i = s)}{\hat{A}_{2n}(Y_i,z)}. \qquad (48)$$

Under conditions given below, $\hat{H}_n(y|z)$ is uniformly consistent over a compact set, asymptotically normally distributed, and achieves the rate of convergence for a one-dimensional regressor. The precise results are stated in theorem 4.

The assumptions needed are similar to the assumptions used to establish convergence of $\ddot{\beta}_n$. Define

$$\hat{A}_0^*(y,s,z,x_{-1}) = \Pr(Y \le y, S \le s, X_{-1} \le x_{-1}|Z = z)\zeta(z). \qquad (49)$$

Suppose $\hat{H}(y|z)$ is to be estimated for (y,z) in a set $\hat{\mathcal{M}} \subset R \times R$.

Assumption 10 *For $k_z \in N_+$ given below, suppose:*

1. *The estimation set $\hat{\mathcal{M}} \subset R \times R$ is compact.*
2. *There is a constant $C_6 > 0$ such that $\hat{A}_2 > C_6$ on $\hat{\mathcal{M}}$.*
3. *ζ is bounded.*
4. *X_{-1} is bounded with probability one.*
5. *$\partial_3 \hat{A}_0^*$ and $\partial_3^2 \hat{A}_0^*$ exist and are bounded and continuous.*
6. *$\partial_2^j \hat{A}_1$ and $\partial_2^j \hat{A}_2$ exist and are bounded and continuous for $j = 1, \ldots, k_z$.*
7. *$\partial_2 \hat{A}_2$ exists and is bounded and continuous. There is a constant C_1 such that*

$$\left|\hat{A}_1(t) - \hat{A}_1(t^*)\right| \leq C_1 \left\|t - t^*\right\| \quad \text{for all } t, t^* \in \hat{\mathcal{M}}, \tag{50}$$

   where t denotes $(y, z)'$.

Note that assumption 10.7 does not require continuity of $\hat{A}_1$ since $\hat{\mathcal{M}}$ need not be connected.

Assumption 11 concerns the estimator of β. It is satisfied by the new estimator proposed in section 2.

Assumption 11 *The estimator β_n satisfies:*

1. *By normalization $\beta_{1n} = \beta_1$.*
2. *There is a function $\Omega : R^{1+1+q} \to R^{q-1}$ such that $P\Omega = 0$, the components of $P\Omega\Omega'$ are finite, and $n^{1/2}(\beta_{-1n} - \beta_{-1}) = n^{1/2} P_n \Omega + o_p(1)$ as $n \to \infty$.*

Assumption 12 *For $k_z \in N_+$ given below, the kernel function $K_z : R \to R$ satisfies:*

1. *K_z is a bounded k_z-order kernel with support $[-1, 1]$.*
2. *∂K_z and $\partial^2 K_z$ exist and are bounded and continuous on R.*

Theorem 4 *Suppose assumptions 1, 2, 10, 11, and 12 hold with $k_z \geq 2$, and suppose further that $n^{-1/2} b_z^{-7/2} \to 0$, and $n^{1/2} b_z^{1/2+k_z} \to 0$ as $n \to \infty$. Then*

i. *$\sup_{(y,z)\in\hat{\mathcal{M}}} \left|\hat{H}_n(y|z) - \hat{H}(y|z)\right| = O_p\left(n^{-1} b_z^{-3}\right) + O\left(b_z^{k_z}\right) + o\left(n^{-1/2} b_z^{-1/2} \log n\right)$ almost surely.*

For $j = 1, 2, \ldots, J$ let $(y_j, z_j) \in \hat{\mathcal{M}}$ and define

$$C_{jk} = 1(z_j = z_k) \int_0^{\min(y_j, y_k)} \frac{\hat{A}_1(dv, z_j)}{\hat{A}_2(v, z_j)^2} \int K_z(u)^2 \, du. \tag{51}$$

Then

ii. $\sqrt{nb_z}\big(\hat{H}_n(y_1|z_1) - \hat{H}(y_1|z_1), \ldots, \hat{H}_n(y_J|z_J) - \hat{H}(y_J|z_J)\big) \to^d N$, *where* N *is a multivariate normal random vector with mean 0 and covariance matrix* $C = [C_{jk}]$.

The three terms on the right-hand side in part i of the theorem represent the randomness in β_n, the bias, and the variance, respectively. The bandwidth conditions are satisfied, for example, if $b_z \propto n^{-1/8}$ and $k_z = 4$.

Two properties are worth emphasizing. First, the rate of convergence of $\hat{H}_n$ does not depend on the dimension of the regressor vector. The rate of convergence is the same as for a one-dimensional regressor. The single-index restriction thus eliminates the curse of dimensionality problem. Secondly, the fact that β is estimated does not affect the asymptotic distribution of $\hat{H}_n$. This is because β_n converges at the root-n rate which is faster than the nonparametric rate. Asymptotically, the variance inherent in the nonparametric estimator completely dominates the variance in β_n.

3.2 Estimation of $\hat{h}$

Having established the asymptotic properties of an estimator of the integrated transition intensity, this section consider estimators of the transition intensity itself. The results are very similar, especially for discrete duration data. For continuous duration data, additional smoothing is necessary which slightly complicates the notation and the derivations.

Suppose Y given $S = s$ is discrete with support $\mathcal{Y} = \big\{\gamma_l : l = 1, 2, 3, \ldots; \gamma_l > \gamma_{l-1}\big\}$. Then

$$\begin{aligned} \hat{h}(\gamma_l|z) &= \Pr(Y = \gamma_l, S = s | Y \geq \gamma_l, Z = z) \\ &= \begin{cases} \hat{H}(\gamma_1|z) & \text{if } l = 1 \\ \hat{H}(\gamma_l|z) - \hat{H}(\gamma_{l-1}|z) & \text{if } l > 1. \end{cases} \end{aligned} \tag{52}$$

The asymptotic distribution of the estimator

$$\hat{h}_n(\gamma_l|z) = \begin{cases} \hat{H}_n(\gamma_1|z) & \text{if } l = 1 \\ \hat{H}_n(\gamma_l|z) - \hat{H}_n(\gamma_{l-1}|z) & \text{if } l > 1 \end{cases} \tag{53}$$

follows immediately from theorem 4. For completeness the results are stated as theorem 5. Let $\hat{\mathcal{M}}$ be the (compact) subset of $\mathcal{Y} \times R$ on which $\hat{h}$ is to be estimated.

Theorem 5 *Suppose the assumptions of theorem 4 hold. Then*

i. $\sup_{(y,z)\in\hat{\mathcal{M}}} \big|\hat{h}_n(y|z) - \hat{h}(y|z)\big| = O_p\big(n^{-1}b_z^{-3}\big) + O\big(b_z^{k_z}\big) + o\big(n^{-1/2}b_z^{-1/2}\log n\big)$ *almost surely.*

For $y_j = \gamma_l$ let $y_{j'}$ denote γ_{l-1}, and define $\hat{A}_1(\gamma_{1-1}, \cdot) = 0$. For $j = 1, 2, \dots, J$ let $(y_j, z_j) \in \hat{\mathcal{M}}$ and define

$$c_{jk} = 1(z_j = z_k)1(y_j = y_k)\frac{\hat{A}_1(y_j, z_j) - \hat{A}_1(y_{j'}, z_j)}{\hat{A}_2(y_j, z_j)^2} \int K_z(u)^2\, du. \tag{54}$$

Then

ii. $\sqrt{nb_z}\big(\hat{h}_n(y_1|z_1) - \hat{h}(y_1|z_1), \dots, \hat{h}_n(y_J|z_J) - \hat{h}(y_J|z_J)\big) \to^d N$, *where N is a multivariate normal random vector with mean 0 and covariance matrix $c = [c_{jk}]$.*

The typical element in the covariance matrix is $c_{jk} = C_{jk} - C_{jk'} - C_{j'k} + C_{j'k'}$, where C_{jk} is defined in theorem 4. The proof of theorem 5 is trivial and hence omitted.

Now suppose Y given $S = s$ is continuous. A number of different intensity estimators have been proposed, going back to Watson and Leadbetter (1964a, 1964b). Rice and Rosenblatt (1976) showed that the asymptotic variance is the same for all these estimators, but the asymptotic bias may differ. Asymptotically unbiased versions of the estimators are available, however, using higher-order kernels. This paper considers the simplest, asymptotically unbiased, estimator.

Since Y given $S = s$ is continuous $\hat{F}_1(y|z)$ and $\hat{H}(y|z)$ are differentiable with respect to y. Define $\hat{a}_1$ by $\hat{a}_1(y, z) = \partial_1 \hat{A}_1(y, z) = \partial_1 \hat{F}_1(y|z)\zeta(z)$. Then

$$\hat{h}(y|z) = \frac{\partial_1 \hat{F}_1(y|x)}{\hat{F}_2(y|x)} = \frac{\hat{a}_1(y, z)}{\hat{A}_2(y, z)}. \tag{55}$$

To estimate $\hat{h}$, it is necessary to smooth $\hat{A}_{1n}(y, z)$ in the y-direction as well as the z-direction. Let b_y be yet another bandwidth parameter, let $K_y : R \to R$ be a kernel function, define $K_{yn}(y) = b_y^{-1} K_y(b_y^{-1} y)$ and the estimator

$$\hat{a}_{1n}(y, z) = \frac{1}{n}\sum_{i=1}^{n} K_{zn}(z - Z_{in})K_{yn}(y - Y_i)1(S_i = s). \tag{56}$$

The estimator of $\hat{h}(y|z)$ considered here is

$$\hat{h}_n(y|z) = \frac{\hat{a}_{1n}(y, z)}{\hat{A}_{2n}(y, z)}, \tag{57}$$

where $\hat{A}_{2n}$ is defined in equation (47).

Define

$$\hat{A}_1^*(y, z, x_{-1}) = \Pr(Y \le y, S = s, X_{-1} \le x_{-1}|Z = z)\zeta(z), \tag{58}$$

and $\hat{a}_1^*(y, z, x_{-1}) = \partial_1 \hat{A}_1^*(y, z, x_{-1})$.

Sufficient smoothness conditions in the y-direction, complementary to assumption 10, are listed as assumption 13. A special boundary problem arises because the support of Y is bounded below at 0. In general, the ordinary kernel density estimator of $\hat{a}_1$ defined above will have a bias of order $O_p(1)$ for y near 0. Following many other theoretical papers, the problem is sidestepped in this section simply by not considering estimating $\hat{h}$ for y near 0. Specifically, assumption 13.1 requires $y > \epsilon$ which ensures that the boundary effect vanishes asymptotically. In practice, however, one is often particularly interested in estimating $\hat{h}$ for y near 0 and a modification is needed to obtain acceptable estimates. A simple but effective boundary correction is discussed in section 4.

Assumption 13 *For $k_z \in N_+$ and $k_y \in N_+$ given below, suppose:*

1. *There is a constant $\epsilon > 0$ such that if $(y, z) \in \hat{\mathcal{M}}$, then $y > \epsilon$.*
2. *$\hat{a}_1$ is bounded on $\hat{\mathcal{M}}$.*
3. *$\partial_2 \hat{a}_1^*$, $\partial_1 \partial_2 \hat{a}_1^*$ and $\partial_2^2 \hat{a}_1^*$ exist and are continuous on $R_+ \times R$.*
4. *$\partial_1^j \hat{a}_1$ exist and are bounded and continuous on $R_+ \times R$ for $j = 1, \ldots, k_y$, and $\partial_2^j \hat{a}_1$ exist and are bounded and continuous on $R_+ \times R$ for $j = 1, \ldots, k_z$.*

Assumption 14 *For $k_y \in N_+$ given below, the kernel function $K_y : R \to R$ satisfies:*

1. *K_y is a bounded k_y-order kernel with support $[-1, 1]$.*
2. *∂K_y exists and is bounded and continuous on R.*

Theorem 6 *Suppose assumptions 1, 2, 10, 11, 12, 13, and 14 hold and that $n^{-1/2} \times b_y^{-1/2} b_z^{-3/2} \log n$ is bounded as $n \to \infty$. Then*

i. $\sup_{(y,z)\in\hat{\mathcal{M}}} \big|\hat{h}_n(y|z) - \hat{h}(y|z)\big| = O_p\big(n^{-1} b_y^{-1} b_z^{-3}\big) + o\big(n^{-1/2} b_y^{-1/2} b_z^{-1/2} \log n\big) + O\big(b_y^{k_y}\big) + O\big(b_z^{k_z}\big)$ *almost surely.*

For $j = 1, 2, \ldots, J$ let $(y_j, z_j) \in \hat{\mathcal{M}}$, and define

$$C_{jk} = 1(z_j = z_k) 1(y_j = y_k) \frac{\hat{a}_1(y_j, z_j)}{\hat{A}_2(y_j, z_j)^2} \int K_y(t)^2 \, dt \int K_z(u)^2 \, du. \tag{59}$$

If also $n^{-1/2} b_y^{-1/2} b_z^{-5/2} \to 0$, $n^{1/2} b_y^{1/2+k_y} b_z^{1/2} \to 0$, and $n^{1/2} b_y^{1/2} b_z^{1/2+k_z} \to 0$ as $n \to \infty$, then

ii. $\sqrt{n b_y b_z}\big(\hat{h}_n(y_1|z_1) - \hat{h}(y_1|z_1), \ldots, \hat{h}_n(y_J|z_J) - \hat{h}(y_J|z_J)\big) \to^d N$, *where N is a multivariate normal random vector with mean 0 and covariance matrix $C = [C_{jk}]$.*

Note again that the estimator achieves the rate of convergence for a one-dimensional regressor, although the rate is reduced from $n^{-1/2} b_z^{-1/2}$ to $n^{-1/2} b_z^{-1/2} b_y^{-1/2}$ because of the smoothing in the y-direction. The bandwidth requirements for part ii of the theorem are satisfied, for example, if $b_y \propto n^{-1/5}$, $b_z \propto n^{-1/7}$, $k_y = 2$, and $k_z = 4$.

4 Monte Carlo Experiments

This section reports the results of Monte Carlo experiments undertaken to investigate the properties of the new estimators in a sample of moderate size. There are two exit destinations in all experiments. It is convenient to refer to the event $S = s$ as "exit" and the event $S \neq s$ as "censoring". The focus is on estimating the index coefficients and the transition intensity for the exit destination only.

The data generating process is defined as follows. Both censoring and exit are governed by Weibull transition intensities

$$h(y|x) = \alpha y^{\alpha-1} \exp(x'\beta/2). \tag{60}$$

The parameters α and β for the censoring destination are $\alpha = 1$ and $\beta = (1, 0, 0, 0)$. The parameters for the exit destination varies across experiments as indicated in the tables. There are two continuous regressors and two discrete regressors, all independent. The continuous regressors are standard normally distributed. The first discrete regressor takes the values -1 and 1 each with probability 1/2. The second discrete regressor takes the values $-\sqrt{3/2}$, 0, and $\sqrt{3/2}$, each with probability 1/3. All four regressors have mean 0 and variance 1.

The sample size is 500. In the experiments where α is the same for both the censoring and the exit destinations, expected censoring is 50%. The effective sample size is therefore about 250. There are 500 Monte Carlo replications in each experiment. The computations are carried out using GAUSS and GAUSS' pseudo-random number generators.

The weight function used to calculate the coefficients of the continuous regressors, β_{1n} and β_{2n}, is $w(y, \bar{x}, \ddot{x}) = 1(y \leq y_{.85})(1/6)$, where $y_{.85}$ is the 85% quantile of the conditional distribution of Y given $S = s$. The value of $y_{.85}$ varies across experiments as indicated in table 1. The kernel function is the product kernel $\bar{K}((\bar{x}_1, \bar{x}_2)') = K(\bar{x}_1)K(\bar{x}_2)$, where K is the univariate fourth order ($k = 4$) kernel

$$K(u) = \frac{105}{64}\big(1 - 5u^2 + 7u^4 - 3u^6\big)1\big(|u| \leq 1\big). \tag{61}$$

The bandwidth is $\bar{b}_l = c_c n_l^{-1/7}$, where $c_c = 6$ and n_l is the number of observations for which $\ddot{X}_i = \chi_l$.

Developing an optimal bandwidth selection procedure for the new estimators is left for future research. Optimal bandwidth selection procedures were developed for Powell, Stock, and Stoker's (1989) original weighted average derivative estimator by Härdle and Tsybakov (1993) and Powell and Stoker (1996). They defined the optimal bandwidth as the bandwidth which minimizes the second-order terms in an expansion of the mean square error (the first-order terms do not depend on the bandwidth). In applications they suggest estimating the optimal bandwidth by "plugging-in" preliminary estimates of the bias and variance components in the formula. A similar approach is likely to be successful in the present context.

Table 1: Monte Carlo Designs

α	β	Censoring Min	Censoring Max	$y_{.85}$	$[\pi_0, \pi_1]$	$[c_0, c_1]$
1	(1,0,0,0)'	.50	.50	1.07	$[-2.50, 2.50]$	$[.31, 3.71]$
1	(1,1,0,0)'	.50	.50	.99	$[-3.25, 3.72]$	$[.20, 6.38]$
1	(1,2,0,0)'	.50	.50	.87	$[-4.39, 6.03]$	$[.10, 17.8]$
1	(1,0,.5,0)	.43	.56	1.04	$[-2.49, 2.50]$	$[.38, 2.82]$
1	(1,0,1,0)'	.38	.62	.99	$[-2.49, 2.51]$	$[.47, 2.09]$
1	(1,0,1.5,0)'	.32	.68	.91	$[-2.49, 2.50]$	$[.55, 1.50]$
1	(1,0,0,.5)'	.43	.58	1.06	$[-2.50, 2.50]$	$[.41, 2.72]$
1	(1,0,0,1)'	.35	.68	1.00	$[-2.50, 2.50]$	$[.53, 1.90]$
1	(1,0,0,1.5)'	.29	.71	.93	$[-2.51, 2.50]$	$[.66, 1.29]$
.5	(1,0,0,0)'	.45	.46	.59	$[-2.34, 2.63]$	$[.24, 2.87]$
2	(1,0,0,0)'	.54	.55	1.27	$[-2.63, 2.29]$	$[.43, 5.09]$

The columns labeled Censoring Min and Max show the minimum and the maximum of $\Pr(S \neq s | \ddot{X} = \chi_l)$ across the support points $\chi_1, \ldots, \chi_6$ of $\ddot{X}$.

To compute the coefficients of the discrete regressors, the constant μ is set equal to $y_{.85}$. Horowitz and Härdle (1996) proposed a data-driven method for selecting the intervals $[\pi_0, \pi_1]$ and $[c_0, c_1]$, which can be used without modification in the present context. However, in order to limit the computing time, the interval $[\pi_0, \pi_1]$ is chosen to be the intersection of the six intervals computed as 2.5 standard deviations on either side of the mean of the conditional distribution of $\bar{X}'\bar{\beta}_1$ given $S = s$ and $\ddot{X} = \chi_l$ for $l = 1, \ldots, 6$. Having determined $[\pi_0, \pi_1]$, $[c_0, c_1]$ is the widest possible interval satisfying assumption 6. The kernel function $\ddot{K}$ is set equal to K defined above, and the bandwidth is $\ddot{b}_l = c_d \operatorname{std}_l n_l^{-1/9}$, where $c_d = 5$ and std_l is the sample standard deviation of $\bar{X}_i'\bar{\beta}_n$ conditional on $\ddot{X}_i = \chi_l$. Again, addressing the question of optimal bandwidth selection is beyond the scope of this paper. No optimal bandwidth selection procedure has been developed for Horowitz and Härdle's (1996) original direct estimator either.

Experiments are carried out for different values of the parameters α and β for the exit destination. The values of the choice variables μ $(= y_{.85})$, $[\pi_0, \pi_1]$, and $[c_0, c_1]$ are shown in table 1, together with the minimum and maximum censoring rates for the six points in the support of $\ddot{X}$. Changing the parameters changes the distribution of the dependent variables Y and S. This is reflected in the variation in μ, $[\pi_0, \pi_1]$, $[c_0, c_1]$, and the censoring rates across experiments. Because some of these choice variables vary across experiments, the effect of changing the parameters on the distribution of the estimators is *ceteris paribus* only in a limited sense.

Results for the index coefficients estimators are reported in table 2. The first three experiments show that as β_2 increases relative to β_1, both the bias and the variance of β_{2n} increase considerably. The increase in β_2 has two opposing effects on β_{3n} and

Table 2: Monte Carlo Estimates: Index Coefficients

α	β	β_{2n}			β_{3n}			β_{4n}		
		Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
1	(1,0,0,0)'	.00	.19	.19	.00	.15	.15	.01	.15	.15
1	(1,1,0,0)'	.06	.30	.31	−.00	.14	.14	−.00	.14	.14
1	(1,2,0,0)'	.11	.52	.53	.00	.13	.13	.00	.13	.13
1	(1,0,.5,0)	.00	.19	.19	−.10	.16	.19	−.00	.17	.17
1	(1,0,1,0)'	.00	.18	.18	−.03	.20	.21	−.00	.20	.20
1	(1,0,1.5,0)'	.00	.18	.18	.11	.22	.24	−.00	.19	.19
1	(1,0,0,.5)'	.00	.19	.19	−.00	.17	.17	−.09	.18	.20
1	(1,0,0,1)'	−.00	.19	.19	−.03	.20	.21	.00	.22	.22
1	(1,0,0,1.5)'	−.00	.19	.19	−.04	.19	.19	−.01	.19	.19
.5	(1,0,0,0)'	−.00	.17	.17	.01	.13	.13	.00	.13	.13
2	(1,0,0,0)'	−.01	.23	.23	−.00	.16	.16	−.00	.16	.16

SD and RMSE indicate standard deviation and root mean square error, respectively.

β_{4n}. On the one hand, the accompanying increase in the variance of β_{2n} increases the variances of β_{3n} and β_{4n}. On the other hand, the increase in the variance of $\bar{X}'\bar{\beta}$ facilitates estimation of β_{3n} and β_{4n} by allowing the choice of wider intervals $[\pi_0, \pi_1]$ and $[c_0, c_1]$. In these experiments the net effect is a slight decrease in the variances of β_{3n} and β_{4n}.

In the next two sets of three experiments the coefficients β_3 and β_4 of the discrete regressors are increased. This widens the distance between $G(\cdot, \chi_l)$, $l = 1, 2, \dots, 6$, which in turn forces a reduction in the interval $[c_0, c_1]$. Here the effect on the bias of β_{3n} and β_{4n} is ambiguous, but the variances and the RMSE increase compared to the experiment with $\beta = (1, 0, 0, 0)'$. The bias and variance of β_{2n} are not sensitive to the values of β_3 and β_4 in these experiments, but this will not always be the case. Although the support of discrete regressors plays no direct role in estimating the coefficients of continuous regressors, there is an indirect effect, given that changing β_3 and β_4 shifts the degree censoring between the support points, as can be seen from table 1.

In the last two experiments α is varied. This causes the overall censoring rate, and hence the effective sample sizes, to vary as well. In the experiment where $\alpha = .5$ the overall censoring rate is 46% and when $\alpha = 2$ it is 55%. The results are that the RMSE is slightly lower when $\alpha = .5$ and slightly larger when $\alpha = 2$ compared with the experiment where $\beta = (1, 0, 0, 0)'$.

Consider now the new estimators of the integrated transition intensity and the transition intensity itself. To focus on the implications of the single-index restriction and the curse of dimensionality, the new semiparametric estimators are compared with estimators which require more or less information. The estimators which require less information are ordinary nonparametric estimators. Since there are only two continuous regressors in these experiments the curse of dimensionality is not severe, and the nonparametric estimators should be reasonably well-behaved. The estimators which require more information are semiparametric estimators which impose the single-index restriction, as does the new estimators, but which assume that the index coefficients β are known. Although infeasible in practice, these estimators provide a natural benchmark for the performance of the new estimators. Given the result that the need to estimate β doesn't affect the asymptotic distribution, the performance of the new estimators is expected to be similar to the performance when β is known. Let H_n^I, H_n^S, H_n^N denote the infeasible estimator, the new semiparametric estimator, and the nonparametric estimator of the integrated transition intensity, and let h_n^I, h_n^S, h_n^N denote similarly defined estimators of the transition intensity itself.

The kernel function K_z used to compute H_n^I, H_n^S, h_n^I, and h_n^S is equal to the fourth order kernel K given in equation (61). The kernel used to smooth in the $\bar{X}$-directions for H_n^N and h_n^N is the product kernel $\bar{K}((\bar{x}_1, \bar{x}_2)') = K(\bar{x}_1)K(\bar{x}_2)$. Since Y is continuous, smoothing is required also in the y-direction to estimate the transition intensity. As mentioned earlier, because of "boundary effects" unmodified estimates are often unacceptable for y near 0. A convenient way of reducing the bias is to use a boundary kernel. Many such kernels have been proposed. The one used here is taken from Müller and Wang (1994). The modification consists of replacing the term $K_{yn}(y - Y_i)$ in equation (56) with $K_{yn}(y, y - Y_i) = K_y(y/b_y, (y - Y_i)/b_y)/b_y$, where

$$K_y(q,u) = \begin{cases} \frac{15}{16}(1-u^2)^2 1\big(|u| \le 1\big), & q > 1, \\ \frac{15}{(1+q)^5}(u+1)^2(q-u) \\ \qquad \times \left[2u\left(5\frac{1-q}{1+q} - 1\right) + (3q-1) + 5\frac{(1-q)^2}{1+q}\right] 1\big(-1 \le u \le q\big), & q \le 1. \end{cases} \tag{62}$$

This kernel satisfies assumption 14. The bandwidths b_z and b_y are chosen to approximately minimize the mean square error at each estimation point in order to focus on the relative performance of the estimators. In applications, the data-based bandwidth selection procedure for transition intensity estimation developed by Müller and Wang (1994) can be used.

Estimation results are shown in tables 3 and 4 for the three experiments in table 2 where $\beta = (1, 0, 0, 0)$ and α is either 1, 1/2, or 2. The corresponding transition intensities are horizontal, downward sloping, and upward sloping. The functions are estimated at nine points for different values of (y, x). Since $\beta = (1, 0, 0, 0)$, the conditional distribution of Y given $X = x$ depends only on the first component of x. The evaluation points are therefore chosen such that y and the first component of x varies while the remaining components are constant.

The results confirm the expectations outlined above. In every case in table 3 is it true that $\text{RMSE}(H_n^I) \le \text{RMSE}(H_n^S) < \text{RMSE}(H_n^N)$, and in almost every case is $\text{RMSE}(H_n^S)$ very close to $\text{RMSE}(H_n^I)$ and much lower than $\text{RMSE}(H_n^N)$. Only when

Table 3: Monte Carlo Estimates: Integrated Transition Intensities

		Root Mean Square Errors								
		$\alpha = 1$			$\alpha = .5$			$\alpha = 2$		
y	x	H_n^I	H_n^S	H_n^N	H_n^I	H_n^S	H_n^N	H_n^I	H_n^S	H_n^N
.5	$(-1,0,0,0)$	.05	.07	.13	.06	.08	.16	.03	.04	.09
1.0	$(-1,0,0,0)$	.09	.11	.23	.09	.11	.22	.09	.12	.24
1.5	$(-1,0,0,0)$	.13	.17	.34	.10	.13	.27	.18	.24	.51
.5	$(0,0,0,0)$	.04	.04	.10	.05	.05	.12	.03	.03	.07
1.0	$(0,0,0,0)$	.10	.11	.20	.08	.09	.18	.11	.13	.23
1.5	$(0,0,0,0)$	.17	.20	.36	.12	.14	.25	.31	.40	.64
.5	$(1,0,0,0)$	.11	.15	.24	.15	.20	.31	.07	.08	.15
1.0	$(1,0,0,0)$	.28	.36	.56	.25	.35	.53	.30	.39	.60
1.5	$(1,0,0,0)$	.62	.75	.97	.40	.51	.71	1.27	1.41	1.71

Table 4: Monte Carlo Estimates: Transition Intensities

		Root Mean Square Errors								
		$\alpha = 1$			$\alpha = .5$			$\alpha = 2$		
y	x	h_n^I	h_n^S	h_n^N	h_n^I	h_n^S	h_n^N	h_n^I	h_n^S	h_n^N
.5	$(-1,0,0,0)$	.07	.08	.19	.07	.08	.19	.05	.05	.11
1.0	$(-1,0,0,0)$	.10	.10	.28	.07	.07	.17	.13	.15	.42
1.5	$(-1,0,0,0)$	.14	.14	.42	.07	.07	.17	.42	.43	1.16
.5	$(0,0,0,0)$	.10	.10	.23	.11	.11	.26	.13	.14	.24
1.0	$(0,0,0,0)$	.11	.11	.31	.09	.09	.24	.29	.35	.60
1.5	$(0,0,0,0)$	.15	.15	.61	.09	.09	.25	.54	.54	1.64
.5	$(1,0,0,0)$	.20	.27	.44	.13	.13	.30	.32	.39	.58
1.0	$(1,0,0,0)$	.24	.24	.49	.10	.10	.24	.70	.97	1.40
1.5	$(1,0,0,0)$	.25	.25	.77	.11	.11	.28	1.13	1.20	2.64

$y = 1.5$ and $x = (1,0,0,0)$ is the difference between RMSE(H_n^S) and RMSE(H_n^N) less striking, ranging from 21% to 39%. In all other cases, RMSE(H_n^N) is between 50% to 150% larger than RMSE(H_n^S). Similar results hold for the transition intensity estimators in table 4. Here the least favorable result is for $\alpha = 2$, $y = 1.0$, and $x = (1,0,0,0)$ where RMSE(h_n^S) is 39% larger than and only 44% lower than RMSE(h_n^N). In all other cases, RMSE(h_n^S) is only slightly larger than RMSE(h_n^I) and RMSE(h_n^N) is between 50% and 300% larger than RMSE(h_n^I) and RMSE(h_n^S).

5 Conclusions

This paper has proposed new semiparametric estimators for transition data where the transition intensity of interest satisfies a single-index restriction. In contrast to existing semiparametric estimators, proportional intensities (separability between duration dependence and regressors) is not assumed. Estimators were developed for index coefficients, the conditional transition intensity, and the integrated conditional transition intensity. The new estimators are "direct" in the sense that iteration is not required.

The asymptotic distributions of the estimators were derived. The index coefficient estimator is root-n consistent and asymptotically normally distributed, similarly to index coefficient estimators developed for other settings. The estimators of the conditional transition intensity and the integrated conditional transition intensity converge at the rate typical for the case of a univariate regressor. Thus the single-index restriction eliminates the curse of dimensionality. The intensity estimators are also asymptotically normally distributed, and the asymptotic distributions are as if the index coefficients were known.

The Monte Carlo experiments suggested that the estimators perform very well in samples of 500 observations with 45–55% censoring. The magnitudes of the root mean square errors are reasonable. In practical applications, one has to be careful when choosing the intervals $[\pi_0, \pi_1]$ and $[c_0, c_1]$. The automatic procedure described by Horowitz and Härdle (1996) may provide a good starting point. For the intensity estimates, the root mean square errors are only slightly larger than if the index coefficients were known and much smaller than for the nonparametric estimators which ignore the single-index restriction. Thus, the asymptotic results hold up in the Monte Carlo experiments.

One large problem, the question of optimal bandwidth selection for the index coefficient estimators, was left for future research. The approach taken by Härdle and Tsybakov (1993) and Powell and Stoker (1996) for Powell, Stock, and Stoker's (1989) weighted average derivative estimator seems promising. It may be worthwhile investigating, both for Horowitz and Härdle's (1996) estimator of the coefficients of the discrete regressors in a model where the regression function satisfy a single-index restriction, as well as for the present coefficient estimator in a model where transition intensity is of single-index form.

A Technical Appendix

A.1 Proof of Theorem 1

Proof of Theorem 1 Put $T = (Y, S, \bar{X}', \ddot{X}')'$ and $\mathbf{t} = (\mathbf{y}, \mathbf{s}, \bar{\mathbf{x}}', \ddot{\mathbf{x}}')'$. Then

$$\bar{\beta}_n^* = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \rho_0(T_i, T_j),$$

where

$$\begin{aligned}\rho_0(\mathbf{t}_i, \mathbf{t}_j) &\\ = \int & \partial_{\bar{x}} \bar{K}_{\bar{b}}(\bar{x} - \bar{\mathbf{x}}_i) \bar{K}_{\bar{b}}(\bar{x} - \bar{\mathbf{x}}_j) w(\mathbf{y}_i, \bar{x}, \ddot{\mathbf{x}}_i) 1(\ddot{\mathbf{x}}_i = \ddot{\mathbf{x}}_j) 1(\mathbf{y}_j \geq \mathbf{y}_i) 1(\mathbf{s}_i = s)\, d\bar{x} \\ & - \int \bar{K}_{\bar{b}}(\bar{x} - \bar{\mathbf{x}}_i) \partial_{\bar{x}} \bar{K}_{\bar{b}}(\bar{x} - \bar{\mathbf{x}}_j) w(\mathbf{y}_i, \bar{x}, \ddot{\mathbf{x}}_i) 1(\ddot{\mathbf{x}}_i = \ddot{\mathbf{x}}_j) 1(\mathbf{y}_j \geq \mathbf{y}_i) 1(\mathbf{s}_i = s)\, d\bar{x}.\end{aligned}$$

Define also

$$\rho_1(\mathbf{t}_i, \mathbf{t}_j) = \frac{\rho_0(\mathbf{t}_i, \mathbf{t}_j) + \rho_0(\mathbf{t}_j, \mathbf{t}_i)}{2},$$
$$V_n = \frac{1}{n^2} \sum_{i=1}^n \rho_0(T_i, T_i),$$
$$U_n = \frac{1}{n^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \rho_1(T_i, T_j),$$

and

$$\hat{U}_n = \frac{2}{n} \sum_{i=1}^n P\rho_1(T_i, \cdot) - P \otimes P\rho_0 = P_n \otimes P\rho_0 + P \otimes P_n\rho_0 - P \otimes P\rho_0.$$

To part i of the theorem, convergence is established for each term in the decomposition

$$\bar{\beta}_n^* - \mathbb{E}\bar{\beta}_n^* = V_n + (U_n - \hat{U}_n) + (\hat{U}_n - \mathbb{E}\hat{U}_n) + (\mathbb{E}\hat{U}_n - \mathbb{E}\bar{\beta}_n^*).$$

Note that $\mathbb{E}\bar{\beta}_n^* = P \otimes P\rho_0 + (1/n)\big(\mathbb{E}\rho_0(T, T) - P \otimes P\rho_0\big) = P \otimes P\rho_0 + O\big(n^{-1}\big)$, $\mathbb{E}U_n = (n(n-1)/n^2)P \otimes P\rho_0 = P \otimes P\rho_0 + O\big(n^{-1}\big)$, and $\mathbb{E}\hat{U}_n = P \otimes P\rho_0$. It follows immediately that $\mathbb{E}\hat{U}_n - \mathbb{E}\bar{\beta}_n^* = o\big(n^{-1/2}\big)$.

By change of variables and with $d = 0, 1$, the terms making up $\rho_0(\mathbf{t}_i, \mathbf{t}_j)$ have the form

$$\begin{aligned}Q_0(\mathbf{t}_i, \mathbf{t}_j) = \frac{(-1)^{d+1}}{2\bar{b}^{\bar{q}+1}} \int & \partial_{\bar{x}}^d \bar{K}(\bar{x}) \partial_{\bar{x}}^{1-d} \bar{K}\left(\frac{\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j + \bar{b}\bar{x}}{\bar{b}}\right) \\ & \times w(\mathbf{y}_i, \bar{\mathbf{x}}_i + \bar{b}\bar{x}, \ddot{\mathbf{x}}_i) 1(\ddot{\mathbf{x}}_i = \ddot{\mathbf{x}}_j) 1(\mathbf{y}_j \geq \mathbf{y}_i) 1(\mathbf{s}_i = s)\, d\bar{x}.\end{aligned}$$

Since $\partial_{\bar{x}}^{1-d}\bar{K}$ and w are bounded and $\partial_{\bar{x}}^{d}\bar{K}$ is integrable, $V_n = O\big(n^{-1}\bar{b}^{-\bar{q}-1}\big) = o\big(n^{-1/2}\big)$ almost surely provided $n^{-1/2}\bar{b}^{-\bar{q}-1} \to 0$, or $n\bar{b}^{2\bar{q}+2} \to \infty$, as $n \to \infty$.

By lemma 3.1 of Powell, Stock, and Stoker (1989), $U_n - \hat{U}_n = o_p\big(n^{-1/2}\big)$ provided[10] $\mathbb{E}\big(\|\rho_1(T_i, T_j)\|^2\big) = o_p(n)$ for $i \neq j$. To verify this condition, note that

$$\mathbb{E}\big(\|\rho_1(T_i,T_j)\|^2\big) = \frac{1}{2}\mathbb{E}\big(\rho_0(T_i,T_j)'\rho_0(T_i,T_j)\big) + \frac{1}{2}\mathbb{E}\big(\rho_0(T_i,T_j)'\rho_0(T_j,T_i)\big).$$

By change of variables, the terms making up $\rho_0(\mathbf{t}_i,\mathbf{t}_j)'\rho_0(\mathbf{t}_i,\mathbf{t}_j)$ have the form

$$\begin{aligned} Q_1(\mathbf{t}_i,\mathbf{t}_j) = \frac{(-1)^{d_1+d_2}}{4\bar{b}^{2\bar{q}+2}} \iint & \partial_{\bar{x}'}^{d_1}\bar{K}(\bar{x}_1)\partial_{\bar{x}'}^{1-d_1}\bar{K}\left(\frac{\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j + \bar{b}\bar{x}_1}{\bar{b}}\right) \\ & \times \partial_{\bar{x}}^{d_2}\bar{K}(\bar{x}_2)\partial_{\bar{x}}^{1-d_2}\bar{K}\left(\frac{\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j + \bar{b}\bar{x}_2}{\bar{b}}\right) w(\mathbf{y}_i, \bar{\mathbf{x}}_i + \bar{b}\bar{x}_1, \ddot{\mathbf{x}}_i) \\ & \times w(\mathbf{y}_i, \bar{\mathbf{x}}_i + \bar{b}\bar{x}_2, \ddot{\mathbf{x}}_i) 1(\ddot{\mathbf{x}}_i = \ddot{\mathbf{x}}_j) 1(\mathbf{y}_j \geq \mathbf{y}_i) 1(\mathbf{s}_i = s)\, d\bar{x}_1\, d\bar{x}_2. \end{aligned}$$

By further change of variables,

$$\begin{aligned} &\mathbb{E}\big(Q_1(T_i,T_j)\big) \\ &\quad = \frac{(-1)^{d_1+d_2}}{4\bar{b}^{\bar{q}+2}} \sum_{l=1}^{m} \iiiint\!\!\iint \partial_{\bar{x}'}^{d_1}\bar{K}(\bar{x}_1)\partial_{\bar{x}'}^{1-d_1}\bar{K}(\bar{\mathbf{x}}_j)\partial_{\bar{x}}^{d_2}\bar{K}(\bar{x}_2)\partial_{\bar{x}}^{1-d_2}\bar{K}(\bar{\mathbf{x}}_j - \bar{x}_1 + \bar{x}_2) \\ &\qquad \times w(v, \bar{\mathbf{x}}_i + \bar{b}\bar{x}_1, \chi_l) w(v, \bar{\mathbf{x}}_i + \bar{b}\bar{x}_2, \chi_l) \\ &\qquad \times \bar{A}_2(v, \bar{\mathbf{x}}_i + \bar{b}\bar{x}_1 - \bar{b}\bar{\mathbf{x}}_j, \chi_l)\bar{A}_1(dv, \bar{\mathbf{x}}_i, \chi_l)\, d\bar{x}_1\, d\bar{x}_2\, d\bar{\mathbf{x}}_i\, d\bar{\mathbf{x}}_j. \end{aligned}$$

Similarly, the terms making up $\rho_0(\mathbf{t}_i,\mathbf{t}_j)'\rho_0(\mathbf{t}_j,\mathbf{t}_i)$ have the form

$$\begin{aligned} Q_2(\mathbf{t}_i,\mathbf{t}_j) = \frac{(-1)^{d_1+d_2}}{4\bar{b}^{2\bar{q}+2}} \iint & \partial_{\bar{x}'}^{d_1}\bar{K}(\bar{x}_1)\partial_{\bar{x}'}^{1-d_1}\bar{K}\left(\frac{\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j + \bar{b}\bar{x}_1}{\bar{b}}\right)\partial_{\bar{x}}^{d_2}\bar{K}(\bar{x}_2) \\ & \times \partial_{\bar{x}}^{1-d_2}\bar{K}\left(\frac{\bar{\mathbf{x}}_j - \bar{\mathbf{x}}_i + \bar{b}\bar{x}_2}{\bar{b}}\right) w(\mathbf{y}_i, \bar{\mathbf{x}}_i + \bar{b}\bar{x}_1, \ddot{\mathbf{x}}_i) w(\mathbf{y}_j, \bar{\mathbf{x}}_j + \bar{b}\bar{x}_2, \ddot{\mathbf{x}}_i) \\ & \times 1(\ddot{\mathbf{x}}_i = \ddot{\mathbf{x}}_j) 1(\mathbf{y}_i = \mathbf{y}_j) 1(\mathbf{s}_i = s) 1(\mathbf{s}_j = s)\, d\bar{x}_1\, d\bar{x}_2, \end{aligned}$$

and

$$\begin{aligned} &\mathbb{E}\big(Q_2(T_i,T_j)\big) \\ &\quad = \frac{(-1)^{d_1+d_2}}{4\bar{b}^{\bar{q}+2}} \sum_{l=1}^{m} \iiiint\!\!\iint \partial_{\bar{x}'}^{d_1}\bar{K}(\bar{x}_1)\partial_{\bar{x}'}^{1-d_1}\bar{K}(\bar{\mathbf{x}}_j)\partial_{\bar{x}}^{d_2}\bar{K}(\bar{x}_2)\partial_{\bar{x}}^{1-d_2}\bar{K}(\bar{x}_1 + \bar{x}_2 - \bar{\mathbf{x}}_j) \\ &\qquad \times w(\mathbf{y}_i, \bar{\mathbf{x}}_i + \bar{b}\bar{x}_1, \chi_l) w(\mathbf{y}_j, \bar{\mathbf{x}}_i + \bar{b}\bar{x}_1 - \bar{b}\bar{\mathbf{x}}_j + \bar{b}\bar{x}_2, \chi_l) \\ &\qquad \times 1(\mathbf{y}_i = \mathbf{y}_j)\bar{A}_1(d\mathbf{y}_i, \bar{\mathbf{x}}_i, \chi_l)\bar{A}_1(d\mathbf{y}_j, \bar{\mathbf{x}}_i + \bar{b}\bar{x}_1 - \bar{b}\bar{\mathbf{x}}_j, \chi_l)\, d\bar{x}_1\, d\bar{x}_2\, d\bar{\mathbf{x}}_i\, d\bar{\mathbf{x}}_j. \end{aligned}$$

[10] $\|\cdot\|$ denotes the Euclidean norm $\|u\|^2 = \sum_i u_i^2 = u'u$.

Since $\partial_{\bar{x}}^{1-d_j}\bar{K}$, ξ, and w are bounded and $\partial_{\bar{x}}^{d_j}\bar{K}$ is integrable, these results imply $\mathbb{E}\big(\|\rho_1(T_i,T_j)\|^2\big) = O_p\big(\bar{b}^{-\bar{q}-2}\big) = o_p(n)$ provided that $n\bar{b}^{\bar{q}+2} \to \infty$ as $n \to \infty$. Note that this is exactly the same as the condition derived by Powell, Stock, and Stoker (1989). It follows that the limiting distribution of $\sqrt{n}(\bar{\beta}_n^* - \mathbb{E}\bar{\beta}_n^*)$ is the same as the limiting distribution of $\sqrt{n}(\hat{U}_n - \mathbb{E}\hat{U}_n)$.

By change of variables and integration by parts

$$\begin{aligned} P\rho_0(\mathbf{t},\cdot) = &-\iint \bar{K}(\bar{x})\bar{K}(\bar{\mathbf{x}}_j)\Big(\partial_{\bar{x}}w(\mathbf{y},\bar{\mathbf{x}}+\bar{b}\bar{x},\ddot{\mathbf{x}})\bar{A}_2(\mathbf{y},\bar{\mathbf{x}}+\bar{b}\bar{x}-\bar{b}\bar{\mathbf{x}}_j,\ddot{\mathbf{x}}) \\ &\qquad + w(\mathbf{y},\bar{\mathbf{x}}+\bar{b}\bar{x},\ddot{\mathbf{x}})\partial_{\bar{x}}\bar{A}_2(\mathbf{y},\bar{\mathbf{x}}+\bar{b}\bar{x}-\bar{b}\bar{\mathbf{x}}_j,\ddot{\mathbf{x}})\Big)1(\mathbf{s}=s)\,d\bar{x}\,d\bar{\mathbf{x}}_j \\ &-\iint \bar{K}(\bar{x})\bar{K}(\bar{\mathbf{x}}_j)w(\mathbf{y},\bar{\mathbf{x}}+\bar{b}\bar{x},\ddot{\mathbf{x}}) \\ &\quad \times \partial_{\bar{x}}\bar{A}_2(\mathbf{y},\bar{\mathbf{x}}+\bar{b}\bar{x}-\bar{b}\bar{\mathbf{x}}_j,\ddot{\mathbf{x}})1(\mathbf{s}=s)\,d\bar{x}\,d\bar{\mathbf{x}}_j \\ = &-\partial_{\bar{x}}w(\mathbf{y},\bar{\mathbf{x}},\ddot{\mathbf{x}})\bar{A}_2(\mathbf{y},\bar{\mathbf{x}},\ddot{\mathbf{x}})1(\mathbf{s}=s) \\ &- 2w(\mathbf{y},\bar{\mathbf{x}},\ddot{\mathbf{x}})\partial_{\bar{x}}\bar{A}_2(\mathbf{y},\bar{\mathbf{x}},\ddot{\mathbf{x}})1(\mathbf{s}=s) + r_1(\mathbf{t}), \end{aligned}$$

where $\sup|r_1| = O(\bar{b})$ because $\partial_{\bar{x}}w$, $\partial_{\bar{x}}^2 w$, $\partial_{\bar{x}}\bar{A}_2$, and $\partial_{\bar{x}}^2\bar{A}_2$ exist and are bounded and $\bar{K}$ is integrable. Similarly,

$$\begin{aligned} P\rho_0(\cdot,\mathbf{t}) = &\iiint \bar{K}(\bar{\mathbf{x}}_i)\bar{K}(\bar{x})w(\mathbf{y}_i,\bar{\mathbf{x}}+\bar{b}\bar{x},\ddot{\mathbf{x}})1(\mathbf{y}\ge\mathbf{y}_i) \\ &\quad \times \partial_{\bar{x}}\bar{A}_1(d\mathbf{y}_i,\bar{\mathbf{x}}+\bar{b}\bar{x}-\bar{b}\bar{\mathbf{x}}_i,\ddot{\mathbf{x}})\,d\bar{x}\,d\bar{\mathbf{x}}_i \\ &+\iiint \bar{K}(\bar{\mathbf{x}}_i)\bar{K}(\bar{x})\Big(\partial_{\bar{x}}w(\mathbf{y}_i,\bar{\mathbf{x}}+\bar{b}\bar{x},\ddot{\mathbf{x}})1(\mathbf{y}\ge\mathbf{y}_i) \\ &\quad \times \bar{A}_1(d\mathbf{y}_i,\bar{\mathbf{x}}+\bar{b}\bar{x}-\bar{b}\bar{\mathbf{x}}_i,\ddot{\mathbf{x}}) + w(\mathbf{y}_i,\bar{\mathbf{x}}+\bar{b}\bar{x},\ddot{\mathbf{x}})1(\mathbf{y}\ge\mathbf{y}_i) \\ &\quad \times \partial_{\bar{x}}\bar{A}_1(d\mathbf{y}_i,\bar{\mathbf{x}}+\bar{b}\bar{x}-\bar{b}\bar{\mathbf{x}}_i,\ddot{\mathbf{x}})\Big)\,d\bar{x}\,d\bar{\mathbf{x}}_i \\ = &\,2\int w(\mathbf{y}_i,\bar{\mathbf{x}},\ddot{\mathbf{x}})1(\mathbf{y}\ge\mathbf{y}_i)\partial_{\bar{x}}\bar{A}_1(d\mathbf{y}_i,\bar{\mathbf{x}},\ddot{\mathbf{x}}) \\ &+\int \partial_{\bar{x}}w(\mathbf{y}_i,\bar{\mathbf{x}},\ddot{\mathbf{x}})1(\mathbf{y}\ge\mathbf{y}_i)\bar{A}_1(d\mathbf{y}_i,\bar{\mathbf{x}},\ddot{\mathbf{x}}) + r_2(\mathbf{t}) \end{aligned}$$

where $\sup|r_2| = O(\bar{b})$ because $\partial_{\bar{x}}^j w$ and $\int\big|\partial_{\bar{x}}^j\bar{A}_1(dy,\cdot,\cdot)\big|$ exist and are bounded for $j=1,2$ and $\bar{K}$ is integrable. It follows that

$$P\rho_0(\mathbf{t},\cdot) + P\rho_0(\cdot,\mathbf{t}) = \Phi_{\bar{\beta}^*}(\mathbf{y},\mathbf{s},\bar{\mathbf{x}},\ddot{\mathbf{x}}) + 2\bar{\beta}^* + O(\bar{b})$$

and

$$\hat{U}_n - \mathbb{E}\hat{U}_n = (P_n - P)\Phi_{\bar{\beta}^*} + (P_n-P)r_1 + (P_n-P)r_2,$$

The second moment of $n^{1/2}(P_n-P)r_j$, $j=1,2$, is $P(r_j^2) - (Pr_j)^2$ which is bounded by $P(r_j^2) = O\big(\bar{b}^2\big)$ using $\sup|r_j| = O(\bar{b})$. It follows by Chebyshev's inequality that

$(P_n - P)r_j = o_p\big(n^{-1/2}\big)$, whence the limiting distribution of $\sqrt{n}(\bar{\beta}^*_n - \mathbb{E}\bar{\beta}^*_n)$ is the same as the limiting distribution of $\sqrt{n}(P_n\Phi_{\bar{\beta}^*} - P\Phi_{\bar{\beta}^*}) = \sqrt{n}P_n\Phi_{\bar{\beta}^*}$. Part i of the theorem now follows from the multivariate Lindeberg-Lévy central limit theorem.

Turn now to the bias term. By previous arguments

$$\mathbb{E}\bar{\beta}^*_n - \bar{\beta}^* = \mathbb{E}V_n + \mathbb{E}U_n = o\big(n^{-1/2}\big) + (n(n-1)/n^2)P \otimes P\rho_0 - \bar{\beta}^*.$$

By integration by parts and change of variables

$$\begin{aligned} P \otimes P\rho_0 = \sum_{l=1}^{m} \iiint & \bar{K}(\bar{\mathbf{x}}_i)\bar{K}(\bar{\mathbf{x}}_j)w(\mathbf{y}_i, \bar{x}, \chi_l) \\ & \times \bar{A}_2(\mathbf{y}_i, \bar{x} - \bar{b}\bar{\mathbf{x}}_j, \chi_l)\partial_{\bar{x}}\bar{A}_1(d\mathbf{y}_i, \bar{x} - \bar{b}\bar{\mathbf{x}}_i, \chi_l)\, d\bar{x}\, d\bar{\mathbf{x}}_i\, d\bar{\mathbf{x}}_j \\ - \sum_{l=1}^{m} \iiint & \bar{K}(\bar{\mathbf{x}}_i)\bar{K}(\bar{\mathbf{x}}_j)w(\mathbf{y}_i, \bar{x}, \chi_l) \\ & \times \partial_{\bar{x}}\bar{A}_2(\mathbf{y}_i, \bar{x} - \bar{b}\bar{\mathbf{x}}_j, \chi_l)\bar{A}_1(d\mathbf{y}_i, \bar{x} - \bar{b}\bar{\mathbf{x}}_i, \chi_l)\, d\bar{x}\, d\bar{\mathbf{x}}_i\, d\bar{\mathbf{x}}_j. \end{aligned}$$

Using the assumptions that $\int\big|\partial^j_{\bar{x}}\bar{A}_1(dy, \cdot, \cdot)\big|$ and $\partial^j_{\bar{x}}\bar{A}_2$ exist and are bounded and continuous for $j = 1, \dots, \bar{k}$ and that $\bar{K}$ is of order $\bar{k}$, a Taylor series expansion implies

$$\begin{aligned} P \otimes P\rho_0 &= \sum_{l=1}^{m} \int w(\mathbf{y}, \bar{x}, \chi_l)\big(\bar{A}_2(\mathbf{y}, \bar{x}, \chi_l)\partial_{\bar{x}}\bar{A}_1(d\mathbf{y}, \bar{x}, \chi_l) \\ &\qquad - \partial_{\bar{x}}\bar{A}_2(\mathbf{y}, \bar{x}, \chi_l)\bar{A}_1(d\mathbf{y}, \bar{x}, \chi_l)\big)\, d\bar{x} + O\big(\bar{b}^{2\bar{k}}\big) \\ &= \bar{\beta}^* + O\big(\bar{b}^{2\bar{k}}\big). \end{aligned}$$

Part ii of the theorem follows.

Recalling $\bar{\beta}_{-1} = \bar{\beta}^*_{-1}/\bar{\beta}^*_1$ linearizing $\bar{\beta}_{-1n} - \bar{\beta}_{-1}$ yields

$$\frac{\bar{\beta}^*_{-1n}}{\bar{\beta}^*_{1n}} - \frac{\bar{\beta}^*_{-1}}{\bar{\beta}^*_1} = \frac{\bar{\beta}^*_{-1n} - \bar{\beta}_{-1}\bar{\beta}^*_{1n}}{\bar{\beta}^*_1} - \frac{(\bar{\beta}^*_{-1n} - \bar{\beta}^*_{-1})(\bar{\beta}^*_{1n} - \bar{\beta}^*_1)}{\bar{\beta}^{*\,2}_{-1}} - \frac{(\bar{\beta}^*_{-1n} - \bar{\beta}^*_{-1})^2\bar{\beta}^*_{1n}}{\bar{\beta}^*_{-1n}\bar{\beta}^{*\,2}_{-1}}.$$

Parts iv and v of the theorem follow. ■

A.2 Proof of Theorem 2

Three lemmas are needed to prove theorem 2. Lemma 1, which is stated without proof, contains some standard results on rates of convergence for kernel estimators. Lemma 3 establishes convergence of G_n to G. Lemma 2 deals with convergence of a particular remainder term which appears in the proof of lemma 3.

Define

$$\ddot{A}_0(y, s, \bar{z}, \ddot{x}) = \Pr(Y \leq y, S \leq s, \ddot{X} \leq \ddot{x}|\bar{Z} = \bar{z})\bar{\zeta}(\bar{z}).$$

Assumptions 7.2 and 7.5 imply that $\ddot{A}_0$ is bounded and that the derivatives $\partial_2^j \ddot{A}_0$ exist and are bounded and continuous for $j = 0, 1, \ldots, \ddot{k}$. Given some function f and $(t, \bar{z}) \in \tilde{S}$, where $\tilde{S}$ is a bounded set, define

$$\phi_n(\mathbf{y}, \mathbf{s}, \ddot{\mathbf{x}}, \bar{\mathbf{x}}) = f(\mathbf{y}, \mathbf{s}, \ddot{\mathbf{x}}, \bar{z}, t)\frac{\ddot{K}\big(\ddot{b}^{-1}(\bar{z} - \bar{\mathbf{x}}'\bar{\beta}_n)\big)}{\ddot{b}},$$

$$\varphi_n(\mathbf{y}, \mathbf{s}, \ddot{\mathbf{x}}, \bar{\mathbf{x}}) = f(\mathbf{y}, \mathbf{s}, \ddot{\mathbf{x}}, \bar{z}, t)\frac{\ddot{K}\big(\ddot{b}^{-1}(\bar{z} - \bar{\mathbf{x}}'\bar{\beta})\big)}{\ddot{b}},$$

$$\phi = \iiint f(\mathbf{y}, \mathbf{s}, \ddot{\mathbf{x}}, \bar{z}, t)\ddot{A}_0(d\mathbf{y}, d\mathbf{s}, \bar{z}, d\ddot{\mathbf{x}}),$$

$$\tilde{\phi} = \iiiint f(\mathbf{y}, \mathbf{s}, \ddot{\mathbf{x}}, \bar{z}, t)\bar{\mathbf{x}}_{-1}\partial_3 \ddot{A}_0^*(d\mathbf{y}, d\mathbf{s}, \bar{z}, d\ddot{\mathbf{x}}, d\bar{\mathbf{x}}_{-1}),$$

where $\ddot{A}_0^*$ is defined in equation (33). Convergence of $P_n\phi_n$ is established in lemma 1.

Lemma 1 *Suppose assumptions 7, 8, and 9 hold and that f is bounded. Suppose also that the class $\mathcal{F} = \big\{(\mathbf{y}, \mathbf{s}, \ddot{\mathbf{x}}) \mapsto f(\mathbf{y}, \mathbf{s}, \ddot{\mathbf{x}}, \bar{z}, t) : (t, \bar{z}) \in \tilde{S}\big\}$ is Euclidean for a constant envelope (see Pakes and Pollard 1989).*

i. If $n^{-1/2}\ddot{b}^{-1/2}\log n$ is bounded as $n \to \infty$, then $\sup P_n|\phi_n| = O_p\big(n^{-1/2}\ddot{b}^{-2}\big) + O(1)$.

ii. If $n^{-1/2}\ddot{b}^{-3/2}\log n$ is bounded as $n \to \infty$, then $\sup\big|P_n\phi_n - P_n\varphi_n\big| = O_p\big(n^{-1/2}\big) + O_p\big(n^{-1}\ddot{b}^{-3}\big)$.

iii. If $n^{-1/2}\ddot{b}^{-3/2}\log n$ is bounded as $n \to \infty$, then $\sup\big|P_n\phi_n - P_n\varphi_n - \tilde{\phi}' P_n\Omega\big| = o_p\big(n^{-1/2}\big) + O_p\big(n^{-1}\ddot{b}^{-3}\big)$.

iv. $\sup\big|P_n\varphi_n - \phi\big| = O\big(\ddot{b}^{\ddot{k}}\big) + o\big(n^{-1/2}\ddot{b}^{-1/2}\log n\big)$ almost surely.

v. If $n^{-1/2}\ddot{b}^{-3/2}\log n$ is bounded as $n \to \infty$, then $\sup\big|P_n\phi_n - \phi\big| = O_p\big(n^{-1}\ddot{b}^{-3}\big) + O\big(\ddot{b}^{\ddot{k}}\big) + o\big(n^{-1/2}\ddot{b}^{-1/2}\log n\big)$ almost surely.

Let $(t_j, \bar{z}_j) \in \tilde{S}$, $j = 1, 2, \ldots, J$ and let ϕ_{nj} and ϕ_j denote the corresponding ϕ_n and ϕ. Define

$$C_{jk} = 1(\bar{z}_j = \bar{z}_k)\iint f(\mathbf{y}, \mathbf{s}, \ddot{\mathbf{x}}, \bar{z}_j, t_j) f(\mathbf{y}, \mathbf{s}, \ddot{\mathbf{x}}, \bar{z}_k, t_k)\ddot{A}_0(d\mathbf{y}, d\mathbf{s}, \bar{z}_j, d\ddot{\mathbf{x}})\int \ddot{K}(u)^2\, du.$$

vi. If $n^{-1/2}\ddot{b}^{-5/2} \to 0$ and $n^{1/2}\ddot{b}^{1/2+\ddot{k}} \to 0$ as $n \to \infty$, then $\sqrt{n\ddot{b}}\big(P_n\phi_{n1} - \phi_1, \ldots, P_n\phi_{nJ} - \phi_J\big) \to^d N$, where N is a multivariate normal random vector with mean 0 and covariance matrix $C = [C_{jk}]$.

The proof of lemma 1 is standard and hence omitted. Parts iii, iv and v are essentially lemmas A.6 and A.7 of Gørgens and Horowitz (1999); parts i and ii can be proved by similar methods. Part vi can be proved using methods similar to section 2.5 of

Pagan and Ullah (1998). In part vi, the condition $n^{-1/2}\ddot{b}^{-5/2} \to 0$ removes the effect of estimating rather than knowing $\bar{\beta}$, and the condition $n^{1/2}\ddot{b}^{1/2+\ddot{k}} \to 0$ removes the asymptotic bias.

Let $\ddot{\Pi}_{jn}$ be defined as $\ddot{A}_{jn}$ with $\bar{Z}_i = \bar{X}_i'\bar{\beta}$ replacing $\bar{Z}_{in} = \bar{X}_i'\bar{\beta}_n$. Let L be any function and define

$$r_n^{\Pi}(y,\bar{z},\ddot{x}) = \int_0^y L(v,\bar{z},\ddot{x})\big(\ddot{\Pi}_{2n}(v,\bar{z},\ddot{x}) - \ddot{A}_2(v,\bar{z},\ddot{x})\big)\big(\ddot{\Pi}_{1n}(dv,\bar{z},\ddot{x}) - \ddot{A}_1(dv,\bar{z},\ddot{x})\big)$$

and

$$r_n^{A}(y,\bar{z},\ddot{x}) = \int_0^y L(v,\bar{z},\ddot{x})\big(\ddot{A}_{2n}(v,\bar{z},\ddot{x}) - \ddot{A}_2(v,\bar{z},\ddot{x})\big)\big(\ddot{A}_{1n}(dv,\bar{z},\ddot{x}) - \ddot{A}_1(dv,\bar{z},\ddot{x})\big).$$

The integral $r_n^A(y,\bar{z},\ddot{x})$ arises as a remainder term in the proof of lemma 3, where $L(v,\bar{z},\ddot{x}) = 1/\ddot{A}_2(v,\bar{z},\ddot{x})^2$. It has the form of a U-process. A U-statistic can be written as a P-degenerate U-process plus an empirical process, and the rate of convergence is determined by the slower of the two, that is, by the empirical process. Lemma 2 shows that for this particular U-process the empirical process part converges faster than usual. The first part of the lemma deals with the case where $\bar{\beta}$ is known, whereas the general case where $\bar{\beta}$ is estimated is considered in the second part.

Lemma 2 *Suppose assumptions 7 and 8 hold and that L and $\partial_2 L$ are bounded on $\mathcal{M}_D$. Then*

i. $\sup_{\mathcal{M}_D} |r_n^{\Pi}| = o\big(n^{-1+\delta}\ddot{b}^{-2}\big) + o\big(n^{-1/2}\ddot{b}^{-1+\ddot{k}}\log n\big) + O\big(\ddot{b}^{2\ddot{k}}\big)$ *almost surely.*

Suppose in addition that assumption 9 holds and that $n^{-1+\delta}\ddot{b}^{-3}$ is bounded for some $\delta > 0$. Then

ii. $\sup_{\mathcal{M}_D} |r_n^{A}| = o\big(n^{-1+\delta}\ddot{b}^{-2}\big) + o\big(n^{-1/2}\ddot{b}^{-1+\ddot{k}}\log n\big) + O_p\big(n^{-1}\ddot{b}^{-4}\big) + o_p\big(n^{-1/2}\big) + O\big(\ddot{b}^{2\ddot{k}}\big)$ *almost surely.*

Proof of Lemma 2 Define $T = (Y,S,\ddot{X}',\bar{Z})'$ and $\mathbf{t} = (\mathbf{y},\mathbf{s},\ddot{\mathbf{x}}',\bar{\mathbf{z}})'$. Writing out the integral gives

$$r_n^{\Pi}(y,\bar{z},\ddot{x}) = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \rho_0(T_i,T_j),$$

where

$$\begin{aligned}\rho_0(\mathbf{t}_i,\mathbf{t}_j) = {}& \ddot{K}_{\ddot{b}}(\bar{z}-\bar{\mathbf{z}}_i)\ddot{K}_{\ddot{b}}(\bar{z}-\bar{\mathbf{z}}_j)\\ &\times L(\mathbf{y}_i,\bar{z},\ddot{x})1(\ddot{\mathbf{x}}_i=\ddot{x})1(\ddot{\mathbf{x}}_j=\ddot{x})1(\mathbf{y}_i\le y)1(\mathbf{s}_i=s)1(\mathbf{y}_i\le\mathbf{y}_j)\\ &- \ddot{K}_{\ddot{b}}(\bar{z}-\bar{\mathbf{z}}_i)\ddot{A}_2(\mathbf{y}_i,\bar{z},\ddot{x})L(\mathbf{y}_i,\bar{z},\ddot{x})1(\ddot{\mathbf{x}}_i=\ddot{x})1(\mathbf{y}_i\le y)1(\mathbf{s}_i=s)\end{aligned}$$

$$- \ddot{K}_{\ddot{b}}(\bar{z} - \bar{\mathbf{z}}_j)1(\ddot{\mathbf{x}}_j = \ddot{x}) \int_0^y L(v, \bar{z}, \ddot{x})1(v \leq \mathbf{y}_j)\ddot{A}_1(dv, \bar{z}, \ddot{x})$$
$$+ \int_0^y L(v, \bar{z}, \ddot{x})\ddot{A}_2(v, \bar{z}, \ddot{x})\ddot{A}_1(dv, \bar{z}, \ddot{x}).$$

Hence, except for diagonal terms and asymmetry, $r_n^{\Pi}(y, \bar{z}, \ddot{x})$ is a U-process.

Let $\ddot{P}$ denote the distribution of T and let $\ddot{P}_n$ be the corresponding empirical measure. Define

$$\tilde{\rho}(\mathbf{t}_i, \mathbf{t}_j) = \rho_0(\mathbf{t}_i, \mathbf{t}_j) - \ddot{P}\rho_0(\mathbf{t}_i, \cdot) - \ddot{P}\rho_0(\cdot, \mathbf{t}_j) + \ddot{P} \otimes \ddot{P}\rho_0$$

and the "projection" of $r_n^{\Pi}(y, \bar{z}, \ddot{x})$

$$\eta(\mathbf{t}) = \ddot{P}\rho_0(\mathbf{t}, \cdot) + \ddot{P}\rho_0(\cdot, \mathbf{t}) - \ddot{P} \otimes \ddot{P}\rho_0.$$

Then

$$r_n^{\Pi}(y, \bar{z}, \ddot{x}) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \tilde{\rho}(T_i, T_j) + (\ddot{P}_n - \ddot{P})\eta + \ddot{P} \otimes \ddot{P}\rho_0,$$

where the double-sum is a $\ddot{P}$-degenerate U-process since $\ddot{P} \otimes \ddot{P}\tilde{\rho} = 0$. Note also that $\ddot{P}\eta = \ddot{P} \otimes \ddot{P}\rho_0$. In the following, convergence of the U-process, the empirical process $(\ddot{P}_n - \ddot{P})\eta$, and the bias term $\ddot{P} \otimes \ddot{P}\rho_0$ are proved consecutively.

Consider the U-process. Convergence will follow from theorem 9 of Nolan and Pollard (1987). Symmetrize the U-process by defining

$$f(\mathbf{t}_i, \mathbf{t}_j) = \frac{\ddot{b}^2}{2}\big(\tilde{\rho}(\mathbf{t}_i, \mathbf{t}_j) + \tilde{\rho}(\mathbf{t}_j, \mathbf{t}_i)\big)$$

and put

$$S_n(f) = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} f(T_i, T_j).$$

Define $\mathcal{F} = \big\{f : (y, \bar{z}, \ddot{x}) \in \mathcal{M}_D,\ 0 < \ddot{b} < 1\big\}$. Note that f is uniformly bounded over $(y, \bar{z}, \ddot{x})$ and $\ddot{b}$ since L, $\ddot{K}$, and $\bar{\zeta}$ are bounded. By lemmas 2.13 and 2.14 of Pakes and Pollard (1989), the assumptions that $\ddot{K}$, $\partial\ddot{K}$, L, and $\partial_2 L$ are bounded, that $\ddot{A}_1$ is Lipschitz continuous on $\mathcal{M}_D$, and that the support of $\ddot{X}$ is finite imply that $\mathcal{F}$ is a Euclidean class for a constant envelope. Since L and $\ddot{K}$ are bounded, $\sup_{f \in \mathcal{F}} |f| \leq C_1$ and $\sup_{f \in \mathcal{F}} \ddot{P}\big|f(\mathbf{t}, \cdot)\big| \leq C_1$ for some constant C_1. Put $\nu(f) = 1$ and $W(n, x) = n^{-\delta}(1 + x)^{-1}$ where $\delta > 0$. Then W is bounded by 1, decreasing in both arguments, and

$$\int_0^1 W(n, x)\big(1 + \log(1/x)\big)\, dx = n^{-\delta}\left(\frac{\pi^2}{12} + \log(2)\right).$$

Rescale by C_1, then the conditions of theorem 9 of Nolan and Pollard (1987) are satisfied, and therefore

$$\sup_{f\in\mathcal{F}}\left|W(n,\nu(f)^{1/2})n^{-1}S_n(f)\right| = \sup_{f\in\mathcal{F}}\left|n^{-1-\delta}S_n(f)/2\right| = o(1) \quad \text{a.s.}$$

It follows that

$$\sup_{f\in\mathcal{F}}\left|n^{-2}\ddot{b}^{-2}S_n(f)\right| = \sup_{\substack{(y,\bar{z},\ddot{x})\in\mathcal{M}_D \\ 0<\ddot{b}<1}}\left|\frac{1}{n^2}\sum_{j=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n}\tilde{\rho}(T_i,T_j)\right| = o\big(n^{-1+\delta}\ddot{b}^{-2}\big) \quad \text{a.s.}$$

The sum of diagonal terms converges as well, because $|\tilde{\rho}| \le \ddot{b}^{-2}C_1$ implies

$$\sup_{(y,\bar{z},\ddot{x})\in\mathcal{M}_D}\left|\frac{1}{n^2}\sum_{i=1}^{n}\tilde{\rho}(T_i,T_i)\right| = O\big(n^{-1}\ddot{b}^{-2}\big) = o\big(n^{-1+\delta}\ddot{b}^{-2}\big).$$

It follows that the U-process is uniformly $o\big(n^{-1+\delta}\ddot{b}^{-2}\big)$ almost surely.

Turn to the empirical process. Note that $\ddot{P}\eta = \ddot{P}\otimes\ddot{P}\rho_0$ and that

$$\begin{aligned}
&\eta(\mathbf{t}) - \ddot{P}\otimes\ddot{P}\rho_0 \\
&\quad= \ddot{K}_{\ddot{b}}(\bar{z}-\bar{\mathbf{z}})1(\ddot{\mathbf{x}}=\ddot{x})L(\mathbf{y},\bar{z},\ddot{x})1(\mathbf{y}\le y)1(\mathbf{s}=s)\int \ddot{K}_{\ddot{b}}(\bar{z}-\bar{\mathbf{z}}_j)\ddot{A}_2(\mathbf{y},\bar{\mathbf{z}}_j,\ddot{x})\,d\bar{\mathbf{z}}_j \\
&\qquad - \ddot{K}_{\ddot{b}}(\bar{z}-\bar{\mathbf{z}})1(\ddot{\mathbf{x}}=\ddot{x})L(\mathbf{y},\bar{z},\ddot{x})1(\mathbf{y}\le y)1(\mathbf{s}=s)\ddot{A}_2(\mathbf{y},\bar{z},\ddot{x}) \\
&\qquad + \ddot{K}_{\ddot{b}}(\bar{z}-\bar{\mathbf{z}})1(\ddot{\mathbf{x}}=\ddot{x})\int \ddot{K}_{\ddot{b}}(\bar{z}-\bar{\mathbf{z}}_i)\int_0^y L(v,\bar{z},\ddot{x})1(v\le\mathbf{y})\ddot{A}_1(dv,\bar{\mathbf{z}}_i,\ddot{x})\,d\bar{\mathbf{z}}_i \\
&\qquad - \ddot{K}_{\ddot{b}}(\bar{z}-\bar{\mathbf{z}})1(\ddot{\mathbf{x}}=\ddot{x})\int_0^y L(v,\bar{z},\ddot{x})1(v\le\mathbf{y})\ddot{A}_1(dv,\bar{z},\ddot{x}) \\
&\qquad + \int \ddot{K}_{\ddot{b}}(\bar{z}-\bar{\mathbf{z}}_i)\int_0^y L(v,\bar{z},\ddot{x})\ddot{A}_2(v,\bar{z},\ddot{x})\ddot{A}_1(dv,\bar{\mathbf{z}}_i,\ddot{x})\,d\bar{\mathbf{z}}_i \\
&\qquad + \int \ddot{K}_{\ddot{b}}(\bar{z}-\bar{\mathbf{z}}_j)\int_0^y L(v,\bar{z},\ddot{x})\ddot{A}_2(v,\bar{\mathbf{z}}_j,\ddot{x})\ddot{A}_1(dv,\bar{z},\ddot{x})\,d\bar{\mathbf{z}}_j.
\end{aligned}$$

Convergence will follow from theorem II.37 of Pollard (1984). Define

$$\begin{aligned}
h_1(\mathbf{y},\ddot{\mathbf{x}},\bar{\mathbf{z}}_i) &= 1(\ddot{\mathbf{x}}=\ddot{x})\int_0^y L(v,\bar{z},\ddot{x})1(v\le\mathbf{y})\ddot{A}_1(dv,\bar{\mathbf{z}}_i,\ddot{x}), \\
h_2(\mathbf{y},\mathbf{s},\ddot{\mathbf{x}},\bar{\mathbf{z}}_j) &= L(\mathbf{y},\bar{z},\ddot{x})1(\ddot{\mathbf{x}}=\ddot{x})1(\mathbf{y}\le y)1(\mathbf{s}=s)\ddot{A}_2(\mathbf{y},\bar{\mathbf{z}}_j,\ddot{x}).
\end{aligned}$$

Then $(\ddot{P}_n-\ddot{P})\eta = (\ddot{P}_n-\ddot{P})f_{1n} + (\ddot{P}_n-\ddot{P})f_{2n}$, where by change of variables

$$f_{1n}(\mathbf{t}) = \ddot{K}_{\ddot{b}}(\bar{z}-\bar{\mathbf{z}})1(\ddot{\mathbf{x}}=\ddot{x})\int \ddot{K}(\bar{\mathbf{z}}_i)\big(h_1(\mathbf{y},\ddot{\mathbf{x}},\bar{z}+\ddot{b}\bar{\mathbf{z}}_i) - h_1(\mathbf{y},\ddot{\mathbf{x}},\bar{z})\big)\,d\bar{\mathbf{z}}_i,$$

$$f_{2n}(\mathbf{t}) = \ddot{K}_{\ddot{b}}(\bar{z}-\bar{\mathbf{z}})1(\ddot{\mathbf{x}}=\ddot{x})\int \ddot{K}(\bar{\mathbf{z}}_j)\big(h_2(\mathbf{y},\mathbf{s},\ddot{\mathbf{x}},\bar{z}+\ddot{b}\bar{\mathbf{z}}_j) - h_2(\mathbf{y},\mathbf{s},\ddot{\mathbf{x}},\bar{z})\big)\,d\bar{\mathbf{z}}_j.$$

The subscript n signifies dependence on the bandwidth sequence $\ddot{b} \to 0$ as $n \to \infty$. Using the assumption of a higher-order kernel and boundedness of $\partial_4^{\ddot{k}} h_1$ and $\partial_3^{\ddot{k}} h_2$, consecutive Taylor expansions yield $\ddot{P}(\ddot{b} f_{jn})^2 \le C_3 \ddot{b}^{2\ddot{k}}$ for all $(y, \bar{z}, \ddot{x}) \in \mathcal{M}_D$ and some constant C_3. Furthermore, the classes $\mathcal{F}_{jn} = \{\ddot{b} f_{jn} : (y, \bar{z}, \ddot{x}) \in \mathcal{M}_D\}$ are Euclidean for constant envelopes by lemmas 2.13 and 2.14 of Pakes and Pollard (1989) since $\ddot{K}$, $\partial \ddot{K}$, L, $\partial_2 L$, and $\partial_2 \ddot{A}_2$ are bounded, $\ddot{A}_1$ is Lipschitz continuous on $\mathcal{M}_D$, and the support of $\ddot{X}$ is finite. It follows by theorem II.37 of Pollard (1984) that

$$\ddot{b}^{-1} \sup_{f \in \mathcal{F}_{jn}} \left| (\ddot{P}_n - \ddot{P})(\ddot{b} f) \right| = \ddot{b}^{-1} o\big(n^{-1/2} \ddot{b}^{\ddot{k}} \log n\big) = o\big(n^{-1/2} \ddot{b}^{\ddot{k}} \log n\big) \quad \text{a.s.}$$

Now consider the bias term. Note that

$$\begin{aligned} \ddot{P} \otimes \ddot{P} \rho_0 = & \iint \ddot{K}_{\ddot{b}}(\bar{z} - \bar{\mathbf{z}}_i) \ddot{K}_{\ddot{b}}(\bar{z} - \bar{\mathbf{z}}_j) \int_0^y L(v, \bar{z}, \ddot{x}) \ddot{A}_2(v, \bar{\mathbf{z}}_j, \ddot{x}) \ddot{A}_1(dv, \bar{\mathbf{z}}_i, \ddot{x}) \, d\bar{\mathbf{z}}_j \, d\bar{\mathbf{z}}_i \\ & - \int \ddot{K}_{\ddot{b}}(\bar{z} - \bar{\mathbf{z}}_i) \int_0^y L(v, \bar{z}, \ddot{x}) \ddot{A}_2(v, \bar{z}, \ddot{x}) \ddot{A}_1(dv, \bar{\mathbf{z}}_i, \ddot{x}) \, d\bar{\mathbf{z}}_i \\ & - \int \ddot{K}_{\ddot{b}}(\bar{z} - \bar{\mathbf{z}}_j) \int_0^y L(v, \bar{z}, \ddot{x}) \ddot{A}_2(v, \bar{\mathbf{z}}_j, \ddot{x}) \ddot{A}_1(dv, \bar{z}, \ddot{x}) \, d\bar{\mathbf{z}}_j \\ & + \int_0^y L(v, \bar{z}, \ddot{x}) \ddot{A}_2(v, \bar{z}, \ddot{x}) \ddot{A}_1(dv, \bar{z}, \ddot{x}). \end{aligned}$$

By change of variables and Taylor expansions using the assumption that the kernel is of higher-order

$$\sup_{(y, \bar{z}, \ddot{x}) \in \mathcal{M}_D} \left| \ddot{P} \otimes \ddot{P} \rho_0 \right| = O\big(\ddot{b}^{2\ddot{k}}\big),$$

since $\partial_2^{\ddot{k}} \ddot{A}_1$ and $\partial_2^{\ddot{k}} \ddot{A}_2$ are bounded and continuous and $\ddot{K}$ is supported on $[-1, 1]$. This completes the proof of the the first conclusion of the lemma.

Consider now the case where $\bar{\beta}$ is estimated. Writing out the integral gives

$$\begin{aligned} r_n^A(y, \bar{z}, \ddot{x}) = & \frac{1}{n} \sum_{i=1}^n \Big(\frac{1}{n} \sum_{j=1}^n \ddot{K}_{\ddot{b}}(\bar{z} - \bar{Z}_{jn}) 1(\ddot{X}_j = \ddot{x}) 1(Y_j \ge Y_i) - \ddot{A}_2(Y_i, \bar{z}, \ddot{x}) \Big) \\ & \times L(Y_i, \bar{z}, \ddot{x}) \ddot{K}_{\ddot{b}}(\bar{z} - \bar{Z}_{in}) 1(\ddot{X}_i = \ddot{x}) 1(Y_i \le y) 1(S_i = s) \\ & - \int_0^y L(v, \bar{z}, \ddot{x}) \Big(\frac{1}{n} \sum_{j=1}^n \ddot{K}_{\ddot{b}}(\bar{z} - \bar{Z}_{jn}) 1(\ddot{X}_j = \ddot{x}) 1(Y_j \ge v) - \ddot{A}_2(v, \bar{z}, \ddot{x}) \Big) \\ & \times \ddot{A}_1(dv, \bar{z}, \ddot{x}). \end{aligned}$$

By the mean value theorem,

$$\ddot{K}_{\ddot{b}}(\bar{z} - \bar{X}_i' \bar{\beta}_n) = \ddot{K}_{\ddot{b}}(\bar{z} - \bar{X}_i' \bar{\beta}) - \partial \ddot{K}_{\ddot{b}}(\bar{z} - \bar{X}_i' \bar{\beta}) \bar{X}_{-1}{}_i' (\bar{\beta}_{-1n} - \bar{\beta}_{-1}) + \eta_{ni},$$

where

$$\eta_{ni} = \big(\partial \ddot{K}_{\ddot{b}}(\bar{z} - \bar{X}_i' \bar{\beta}) - \partial \ddot{K}_{\ddot{b}}(\bar{z} - \bar{X}_i' \bar{\beta}_n^*) \big) \bar{X}_{-1}{}_i' (\bar{\beta}_{-1n} - \bar{\beta}_{-1})$$

and $\bar{\beta}_n^*$ is between $\bar{\beta}_n$ and $\bar{\beta}$. Redefine $T = (Y, S, \ddot{X}', \bar{Z}, \bar{X}_{-1}')'$ and similarly let $\mathbf{t} = (\mathbf{y}, \mathbf{s}, \ddot{\mathbf{x}}', \bar{\mathbf{z}}, \bar{\mathbf{x}}_{-1}')'$. Substituting into the expression for r_n^A gives

$$r_n^A(y, \bar{z}, \ddot{x}) = r_n^\Pi(y, \bar{z}, \ddot{x}) - (\bar{\beta}_{-1n} - \bar{\beta}_{-1})' \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \rho_1(T_i, T_j) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n r_{nij},$$

where $r_n^\Pi(y, \bar{z}, \ddot{x})$ is defined above, where

$$\begin{aligned}
\rho_1(\mathbf{t}_i, \mathbf{t}_j) = {} & \partial \ddot{K}_{\ddot{b}}(\bar{z} - \bar{\mathbf{z}}_i) \ddot{K}_{\ddot{b}}(\bar{z} - \bar{\mathbf{z}}_j) \\
& \quad \times 1(\ddot{\mathbf{x}}_i = \ddot{x}) 1(\ddot{\mathbf{x}}_j = \ddot{x}) L(\mathbf{y}_i, \bar{z}, \ddot{x}) 1(\mathbf{y}_i \le y) 1(\mathbf{s}_i = s) 1(\mathbf{y}_i \le \mathbf{y}_j) \bar{\mathbf{x}}_{-1i} \\
& + \ddot{K}_{\ddot{b}}(\bar{z} - \bar{\mathbf{z}}_i) \partial \ddot{K}_{\ddot{b}}(\bar{z} - \bar{\mathbf{z}}_j) \\
& \quad \times 1(\ddot{\mathbf{x}}_i = \ddot{x}) 1(\ddot{\mathbf{x}}_j = \ddot{x}) L(\mathbf{y}_i, \bar{z}, \ddot{x}) 1(\mathbf{y}_i \le y) 1(\mathbf{s}_i = s) 1(\mathbf{y}_i \le \mathbf{y}_j) \bar{\mathbf{x}}_{-1j} \\
& - \partial \ddot{K}_{\ddot{b}}(\bar{z} - \bar{\mathbf{z}}_i) 1(\ddot{\mathbf{x}}_i = \ddot{x}) L(\mathbf{y}_i, \bar{z}, \ddot{x}) 1(\mathbf{y}_i \le y) 1(\mathbf{s}_i = s) \ddot{A}_2(\mathbf{y}_i, \bar{z}, \ddot{x}) \bar{\mathbf{x}}_{-1i} \\
& - \partial \ddot{K}_{\ddot{b}}(\bar{z} - \bar{\mathbf{z}}_j) 1(\ddot{\mathbf{x}}_j = \ddot{x}) \int_0^y L(v, \bar{z}, \ddot{x}) 1(v \le \mathbf{y}_j) \ddot{A}_1(dv, \bar{z}, \ddot{x}) \bar{\mathbf{x}}_{-1j},
\end{aligned}$$

and where the r_{nij}'s consist of all terms involving η_{ni} or η_{nj} plus the term

$$\begin{aligned}
\partial \ddot{K}_{\ddot{b}}(\bar{z} - \bar{Z}_i) \partial \ddot{K}_{\ddot{b}}(\bar{z} - \bar{Z}_j) 1(\ddot{X}_i = \ddot{x}) 1(\ddot{X}_j = \ddot{x}) L(Y_i, \bar{z}, \ddot{x}) 1(Y_i \le y) 1(S_i = s) 1(Y_i \le Y_j) \\
\times (\bar{\beta}_{-1n} - \bar{\beta}_{-1})' \bar{X}_{-1i} (\bar{\beta}_{-1n} - \bar{\beta}_{-1})' \bar{X}_{-1j}.
\end{aligned}$$

By $n^{-1/2}$-convergence of $\bar{\beta}_{-1n} - \bar{\beta}_{-1}$ and boundedness of $\partial^2 \ddot{K}$,

$$\sup_{(y, \bar{z}, \ddot{x}) \in \mathcal{M}_D} \left| \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n r_{nij} \right| = O_p\big(n^{-1} \ddot{b}^{-4}\big).$$

Convergence of $r_n^\Pi(y, \bar{z}, \ddot{x})$ has already been established. Similar arguments can be used to show that

$$\sup_{(y, \bar{z}, \ddot{x}) \in \mathcal{M}_D} \left| \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \rho_1(T_i, T_j) \right| = o_p(1),$$

provided $n^{-1+\delta} \ddot{b}^{-3}$ is bounded. The second conclusion of the lemma follows. ■

Define

$$\begin{aligned}
B_n^G(\bar{z}, \ddot{x}) &= \int_0^\mu \frac{\ddot{\Pi}_{1n}(dv, \bar{z}, \ddot{x})}{\ddot{A}_2(v, \bar{z}, \ddot{x})} - \int_0^\mu \frac{\ddot{\Pi}_{2n}(v, \bar{z}, \ddot{x}) \ddot{A}_1(dv, \bar{z}, \ddot{x})}{\ddot{A}_2(v, \bar{z}, \ddot{x})^2} \\
&= \frac{1}{n} \sum_{i=1}^n \ddot{K}_{\ddot{b}}(\bar{z} - \bar{Z}_i) \ddot{f}(Y_i, S_i, \ddot{X}_i, \bar{z}, \ddot{x})
\end{aligned}$$

and

$$\tilde{\Gamma}(\bar{z}, \ddot{x}) = \iiiint \ddot{f}(\mathbf{y}, \mathbf{s}, \ddot{\mathbf{x}}, \bar{z}, \ddot{x}) \bar{\mathbf{x}}_{-1} \partial_3 \ddot{A}_0^*(d\mathbf{y}, d\mathbf{s}, \bar{z}, d\ddot{\mathbf{x}}, d\bar{\mathbf{x}}_{-1}),$$

where $\ddot{f}$ is defined in equation (35).

Lemma 3 *Suppose assumptions 1, 2, 7, 8 and 9 hold with* $\ddot{k} \geq 2$. *If* $n^{-1+2\delta}\ddot{b}^{-4}$ *is bounded for some* $\delta > 0$, $n^{1/4}\ddot{b}^{\ddot{k}} \to 0$, *and* $n^{-1/2}\ddot{b}^{-2}$ *is bounded as* $n \to \infty$, *then for each* $\ddot{x} \in \mathcal{X}$*:*

i. $\sup_{\bar{z}\in[\pi_0,\pi_1]}\big|G_n(\bar{z},\ddot{x}) - G(\bar{z},\ddot{x}) - B_n^G(\bar{z},\ddot{x}) - \tilde{\Gamma}' P_n\Omega\big| = o_p(n^{-1/2}) + O_p\big(n^{-1}\ddot{b}^{-4}\big)$.

ii. $\sup_{\bar{z}\in[\pi_0,\pi_1]}\big|G_n(\bar{z},\ddot{x}) - G(\bar{z},\ddot{x})\big| = O_p\big(n^{-1}\ddot{b}^{-4}\big) + O\big(\ddot{b}^{\ddot{k}}\big) + o\big(n^{-1/2}\ddot{b}^{-1/2}\log n\big)$ *almost surely.*

Proof As a matter of algebra,

$$\begin{aligned} G_n(\bar{z},\ddot{x}) - G(\bar{z},\ddot{x}) &= \int_0^\mu \frac{\ddot{A}_{1n}(dv,\bar{z},\ddot{x})}{\ddot{A}_{2n}(v,\bar{z},\ddot{x})} - \int_0^\mu \frac{\ddot{A}_1(dv,\bar{z},\ddot{x})}{\ddot{A}_2(v,\bar{z},\ddot{x})} \\ &= E_n(\bar{z},\ddot{x}) + r_{1n}(\bar{z},\ddot{x}) + r_{2n}(\bar{z},\ddot{x}), \end{aligned}$$

where

$$\begin{aligned} E_n(\bar{z},\ddot{x}) &= \int_0^\mu \frac{\ddot{A}_{1n}(dv,\bar{z},\ddot{x})}{\ddot{A}_2(v,\bar{z},\ddot{x})} - \int_0^\mu \frac{\ddot{A}_{2n}(v,\bar{z},\ddot{x})\ddot{A}_1(dv,\bar{z},\ddot{x})}{\ddot{A}_2(v,\bar{z},\ddot{x})^2} \\ &= \frac{1}{n}\sum_{i=1}^n \ddot{K}_{\ddot{b}}(\bar{z} - \bar{Z}_{in})1(\ddot{X}_i = \ddot{x}) \\ &\quad \times \bigg(\frac{1(Y_i \leq \mu)1(S_i = s)}{\ddot{A}_2(Y_i,\bar{z},\ddot{x})} - \int_0^\mu \frac{1(Y_i \geq v)\ddot{A}_1(dv,\bar{z},\ddot{x})}{\ddot{A}_2(v,\bar{z},\ddot{x})^2}\bigg), \\ r_{1n}(\bar{z},\ddot{x}) &= -\int_0^\mu \frac{\big(\ddot{A}_{2n}(v,\bar{z},\ddot{x}) - \ddot{A}_2(v,\bar{z},\ddot{x})\big)\big(\ddot{A}_{1n}(dv,\bar{z},\ddot{x}) - \ddot{A}_1(dv,\bar{z},\ddot{x})\big)}{\ddot{A}_2(v,\bar{z},\ddot{x})^2}, \\ r_{2n}(\bar{z},\ddot{x}) &= \int_0^\mu \frac{\big(\ddot{A}_{2n}(v,\bar{z},\ddot{x}) - \ddot{A}_2(v,\bar{z},\ddot{x})\big)^2\ddot{A}_{1n}(dv,\bar{z},\ddot{x})}{\ddot{A}_{2n}(v,\bar{z},\ddot{x})\ddot{A}_2(v,\bar{z},\ddot{x})^2} \\ &= \frac{1}{n}\sum_{i=1}^n \frac{\big(\ddot{A}_{2n}(Y_i,\bar{z},\ddot{x}) - \ddot{A}_2(Y_i,\bar{z},\ddot{x})\big)^2}{\ddot{A}_{2n}(Y_i,\bar{z},\ddot{x})\ddot{A}_2(Y_i,\bar{z},\ddot{x})^2} \\ &\quad \times \ddot{K}_{\ddot{b}}(\bar{z} - \bar{Z}_{in})1(\ddot{X}_i = \ddot{x})1(Y_i \leq \mu)1(S_i = s). \end{aligned}$$

Since $n^{-1+2\delta}\ddot{b}^{-4}$ is bounded and $n^{1/2}\ddot{b}^{2\ddot{k}} \to 0$, lemma 2.ii implies

$$\begin{aligned} \sup_{(\mu,\bar{z},\ddot{x})\in\mathcal{M}_D} |r_{1n}(\bar{z},\ddot{x})| &= o\big(n^{-1+\delta}\ddot{b}^{-2}\big) + o\big(n^{-1/2}\ddot{b}^{-1+\ddot{k}}\log n\big) + o_p\big(n^{-1/2}\big) \\ &\quad + O_p\big(n^{-1}\ddot{b}^{-4}\big) + O\big(\ddot{b}^{2\ddot{k}}\big) \quad \text{a.s.} \\ &= o_p\big(n^{-1/2}\big) + O_p\big(n^{-1}\ddot{b}^{-4}\big). \end{aligned}$$

Boundedness of $n^{-1+2\delta}\ddot{b}^{-4}$ means that $n^{-1/2}\ddot{b}^{-3/2}\log n$ is bounded, whence lemma 1.v implies

$$\sup|\ddot{A}_{2n} - \ddot{A}_2| = O_p\big(n^{-1}\ddot{b}^{-3}\big) + O\big(\ddot{b}^{\ddot{k}}\big) + o\big(n^{-1/2}\ddot{b}^{-1/2}\log n\big) \quad \text{a.s.} \qquad \text{(A)}$$

Therefore, $\sup|\ddot{A}_{2n} - \ddot{A}_2| < \epsilon$ for n sufficiently large. Moreover, by assumption $\ddot{A}_2 > C_4$, so for n sufficiently large

$$|r_{2n}(\bar{z},\ddot{x})| \leq \frac{\sup|\ddot{A}_{2n} - \ddot{A}_2|^2}{(C_4-\epsilon)C_4^2}\frac{1}{n}\sum_{i=1}^{n}\bigl|\ddot{K}_{\ddot{b}}(\bar{z}-\bar{Z}_{in})\bigr|.$$

Since $n^{-1/2}\ddot{b}^{-2}$ is bounded, the sum is $O_p(1)$ by lemma 1.i. Therefore, since boundedness of $n^{-1+2\delta}\ddot{b}^{-4}$ implies boundedness of $n^{-1/2}\ddot{b}^{-3/2}\log n$ and since $n^{1/2}\ddot{b}^{2\ddot{k}} \to 0$, another application of (A) yields

$$\begin{aligned}\sup_{(\mu,\bar{z},\ddot{x})\in\mathcal{M}_D}|r_{2n}(\bar{z},\ddot{x})| &= O_p\bigl(n^{-2}\ddot{b}^{-6}\bigr) + O_p\bigl(\ddot{b}^{2\ddot{k}}\bigr) + o_p\bigl(n^{-1}\ddot{b}^{-1}(\log n)^2\bigr)\\ &= O_p\bigl(n^{-2}\ddot{b}^{-6}\bigr) + o_p\bigl(n^{-1/2}\bigr).\end{aligned}$$

It follows that for each $\ddot{x} \in \mathcal{X}$,

$$\sup_{\bar{z}\in[\pi_0,\pi_1]}\bigl|G_n(\bar{z},\ddot{x}) - G(\bar{z},\ddot{x}) - E_n(\bar{z},\ddot{x})\bigr| = o_p\bigl(n^{1/2}\bigr) + O_p\bigl(n^{-1}\ddot{b}^{-4}\bigr).$$

Now define $\phi_{\ddot{b}}(\mathbf{y},\mathbf{s},\ddot{\mathbf{x}},\bar{\mathbf{x}}) = \ddot{K}_{\ddot{b}}(\bar{z}-\bar{\mathbf{x}}'\bar{\beta}_n)\ddot{f}(\mathbf{y},\mathbf{s},\ddot{\mathbf{x}},\bar{z},\ddot{x})$, where $\ddot{f}$ is defined in equation (35). Then $E_n = P_n\phi_{\ddot{b}}$, and $\{(\mathbf{y},\mathbf{s},\ddot{\mathbf{x}}) \mapsto \ddot{f}(\mathbf{y},\mathbf{s},\ddot{\mathbf{x}},\bar{z},\ddot{x}) : \bar{z}\in[\pi_0,\pi_1],\ \ddot{x}\in\mathcal{X}\}$ is a Euclidean class because $1/\ddot{A}_2$ and $\partial_2\ddot{A}_2$ are bounded and $\ddot{A}_1$ is Lipschitz continuous on $\mathcal{M}_D$. Lemmas 1.iii and 1.v imply

$$\sup\bigl|P_n\phi_{\ddot{b}} - P_n\varphi_{\ddot{b}} - \tilde{\phi}'P_n\Omega\bigr| = o_p\bigl(n^{-1/2}\bigr) + O_p\bigl(n^{-1}\ddot{b}^{-3}\bigr)$$

and

$$\sup\bigl|P_n\phi_{\ddot{b}} - \phi\bigr| = O_p\bigl(n^{-1}\ddot{b}^{-3}\bigr) + O\bigl(\ddot{b}^{\ddot{k}}\bigr) + o\bigl(n^{-1/2}\ddot{b}^{-1/2}\log n\bigr) \quad \text{a.s.},$$

where $P_n\varphi_{\ddot{b}} = B_n^G$, $\phi = G$, and $\tilde{\phi} = \tilde{\Gamma}$. The conclusions of the lemma follows. ■

Proof of Theorem 2 Let $\ddot{P}$ denote the distribution of $(Y,S,\ddot{X}',\bar{Z})'$ and let $\ddot{P}_n$ denote the corresponding empirical measure. By lemma 3.ii G_n converges uniformly to G. Assumption 6.2 therefore implies

$$\int\Bigl|1\bigl(G_n(\bar{z},\ddot{x}) < c_0\bigr) - 1\bigl(G(\bar{z},\ddot{x}) < c_0\bigr)\Bigr|\,d\bar{z} = o_p\bigl(n^{-1/2}\bigr).$$

The integral is also $o_p\bigl(n^{-1/2}\bigr)$ if c_0 is replaced by c_1 or the directions of the inequalities are reversed. Since $n^{-1/2}\ddot{b}^{-4} \to 0$, this and lemma 3.i imply that

$$\begin{aligned}J_n(\ddot{x}) - J(\ddot{x}) &= \int_{\pi_0}^{\pi_1}\bigl(G_n(\bar{z},\ddot{x}) - G(\bar{z},\ddot{x})\bigr)1\bigl(c_0 \leq G(\bar{z},\ddot{x}) \leq c_1\bigr)\,d\bar{z} + o_p\bigl(n^{-1/2}\bigr)\\ &= \ddot{P}_n\sigma_{\ddot{b}}(\ddot{x}) + \Gamma(\ddot{x})'P_n\Omega + o_p\bigl(n^{-1/2}\bigr),\end{aligned}$$

where σ and Γ are defined in (36) and (37) and

$$\sigma_{\ddot{b}}(\mathbf{y},\mathbf{s},\ddot{\mathbf{x}},\bar{\mathbf{z}},\ddot{x}) = \int \sigma(\mathbf{y},\mathbf{s},\ddot{\mathbf{x}},\bar{z},\ddot{x})\ddot{K}_{\ddot{b}}(\bar{z}-\bar{\mathbf{z}})\,d\bar{z}.$$

A change of variables and a Taylor series expansion around $\bar{\mathbf{z}}=0$ (using the assumption of a $\ddot{k}$-order kernel and boundedness of $\partial_3^{\ddot{k}}\ddot{A}_0$) yield

$$\ddot{P}\sigma_{\ddot{b}}(\ddot{x}) = \ddot{P}\sigma(\ddot{x}) + O\big(\ddot{b}^{\ddot{k}}\big).$$

By change of variables

$$\begin{aligned}\sigma_{\ddot{b}}(\mathbf{y},\mathbf{s},\ddot{\mathbf{x}},\bar{\mathbf{z}},\ddot{x}) &- \sigma(\mathbf{y},\mathbf{s},\ddot{\mathbf{x}},\bar{\mathbf{z}},\ddot{x})\\ &= \int \big(\sigma(\mathbf{y},\mathbf{s},\ddot{\mathbf{x}},\bar{\mathbf{z}}-\ddot{b}\bar{z},\ddot{x}) - \sigma(\mathbf{y},\mathbf{s},\ddot{\mathbf{x}},\bar{\mathbf{z}},\ddot{x})\big)\ddot{K}(\bar{z})\,d\bar{z}.\end{aligned}$$

Since σ is bounded and $\ddot{K}$ is integrable, there is a constant C_5 such that $\sup|\sigma_{\ddot{b}}-\sigma| < C_5$. Given $\ddot{x}\in\mathcal{X}$, let $\mathcal{D}$ denote the set of the discontinuity points of the function $g(\bar{\mathbf{z}}) = 1(\pi_0 \le \bar{\mathbf{z}} \le \pi_1)1\big(c_0 \le G(\bar{\mathbf{z}},\ddot{x}) \le c_1\big)$. Assumption 6 implies that $\mathcal{D}$ is finite. Let $\mathcal{D}_{\ddot{b}}$ be a cover of $\mathcal{D}$ consisting of intervals of length $2\ddot{b}$ centered on each point in $\mathcal{D}$. Define the indicator function $Q_{\ddot{b}}(\mathbf{y},\mathbf{s},\ddot{\mathbf{x}},\bar{\mathbf{z}},\ddot{x}) = 1(\bar{\mathbf{z}}\in\mathcal{D}_{\ddot{b}})$. Then

$$\ddot{P}\big(Q_{\ddot{b}}(\ddot{x})|\sigma_{\ddot{b}}(\ddot{x})-\sigma(\ddot{x})|\big) \le C_5\ddot{P}Q_{\ddot{b}}(\ddot{x}) = O(\ddot{b}).$$

For $\bar{\mathbf{z}}$ not in $\mathcal{D}_{\ddot{b}}$ and n so large that $\ddot{b}<1/2$, there are no discontinuity points between $\bar{\mathbf{z}}$ and $\bar{\mathbf{z}}-\ddot{b}\bar{z}$. Boundedness of $\partial_4\sigma$ therefore implies

$$\ddot{P}\big(|1-Q_{\ddot{b}}(\ddot{x})||\sigma_{\ddot{b}}(\ddot{x})-\sigma(\ddot{x})|\big) = O(\ddot{b}).$$

Combining these results gives $\ddot{P}|\sigma_{\ddot{b}}(\ddot{x})-\sigma(\ddot{x})| = O(\ddot{b})$ and $\ddot{P}(\sigma_{\ddot{b}}(\ddot{x})-\sigma(\ddot{x}))^2 = O(\ddot{b})$, since $\sigma_{\ddot{b}}-\sigma$ is bounded. It follows by Chebyshev's inequality that $\ddot{P}_n\sigma_{\ddot{b}}(\ddot{x})-\ddot{P}\sigma_{\ddot{b}}(\ddot{x}) = O_p\big(n^{-1/2}\ddot{b}^{1/2}\big) = o_p\big(n^{-1/2}\big)$. Combining this with the previous results that $\ddot{P}\sigma_{\ddot{b}}(\ddot{x}) - \ddot{P}\sigma(\ddot{x}) = O\big(\ddot{b}^{\ddot{k}}\big)$ gives $|\ddot{P}_n\sigma_{\ddot{b}}(\ddot{x})-\ddot{P}_n\sigma(\ddot{x})| = o_p\big(n^{-1/2}\big)$. It follows that, for each $\ddot{x}\in\mathcal{X}$,

$$J_n(\ddot{x}) - J(\ddot{x}) = \ddot{P}_n\sigma(\ddot{x}) + \Gamma(\ddot{x})'P_n\Omega + o_p\big(n^{-1/2}\big).$$

Therefore,

$$\begin{aligned}\ddot{\beta}_n - \ddot{\beta} &= (c_1-c_0)^{-1}(W'W)^{-1}W'(\Delta J_n - \Delta J)\\ &= (c_1-c_0)^{-1}(W'W)^{-1}W'V_n + o_p\big(n^{-1/2}\big),\end{aligned}$$

where V_n is the $(m-1)$-vector whose $(l-1)$th, $l=2,\dots,m$, component is

$$[V_n]_{l-1} = \ddot{P}_nF(\chi_l) - \ddot{P}_nF(\chi_1) + \big(\Gamma(\chi_l)-\Gamma(\chi_1)\big)'P_n\Omega.$$

The conclusion of the theorem follows. ■

A.3 Proof of Theorem 4

Proof of Theorem 4 Define

$$E_n(y,z) = \int_0^y \frac{\hat{A}_{1n}(dv,z)}{\hat{A}_2(v,z)} - \int_0^y \frac{\hat{A}_{2n}(v,z)\hat{A}_1(dv,z)}{\hat{A}_2(v,z)^2}$$
$$= \frac{1}{n}\sum_{i=1}^n K_{zn}(z-Z_{in})\Big(\frac{1(Y_i \le y)1(S_i = s)}{\hat{A}_2(Y_i,z)} - \int_0^y \frac{1(Y_i \ge v)\,\hat{A}_1(dv,z)}{\hat{A}_2(v,z)^2}\Big).$$

Arguments similar to those given in the proof of lemma 3 show that

$$\sup_{(y,z)\in\hat{\mathcal{M}}} \big|\hat{H}_n(y|z) - \hat{H}(y|z) - E_n(y,z)\big| = o_p\big(n^{-1/2}b_z^{-1/2}\big)$$

using the assumptions that $n^{-1+2\delta}b_z^{-3}$ is bounded, $n^{-1/2}b_z^{-7/2} \to 0$, and $n^{1/2}b_z^{1/2+k_z} \to 0$ as $n \to \infty$. Define

$$f(\mathbf{y},\mathbf{s},z,y) = \frac{1(\mathbf{y} \le y)1(\mathbf{s} = s)}{\hat{A}_2(\mathbf{y},z)} - \int_0^y \frac{1(\mathbf{y} \ge v)\hat{A}_1(dv,z)}{\hat{A}_2(\mathbf{y},z)^2}.$$

Define $\phi_{b_z}(\mathbf{y},\mathbf{s},\mathbf{x}) = K_{zn}(z - \mathbf{x}'\beta_n)f(\mathbf{y},\mathbf{s},z,y)$. Then $E_n = P_n\phi_{b_z}$ and it is straightforward to show that the corresponding ϕ-function is $\phi = 0$. Consistency and asymptotic normality now follows from a result similar to part vi of lemma 1. Put $\mathcal{F} = \big\{(\mathbf{y},\mathbf{s}) \mapsto f(\mathbf{y},\mathbf{s},z,y) : (y,z) \in \hat{\mathcal{M}}\big\}$. Then $\mathcal{F}$ is a Euclidean class for a constant envelope by lemmas 2.13 and 2.14 of Pakes and Pollard (1989), since $1/\hat{A}_2$ and $\partial_2\hat{A}_2$ are bounded on $\hat{\mathcal{M}}$ and $\hat{A}_1$ is Lipschitz continuous on $\hat{\mathcal{M}}$. Note that

$$\iint f(\mathbf{y},\mathbf{s},z,y_j)f(\mathbf{y},\mathbf{s},z,y_k)A_0(d\mathbf{y},d\mathbf{s},z)$$
$$= \int_0^{\min(y_j,y_k)} \frac{\hat{A}_1(dv,z)}{\hat{A}_2(v,z)^2} - \int_0^{y_j}\int_0^{y_k} 1(v_1 \ge v_2)\frac{\hat{A}_1(dv_2,z)\hat{A}_1(dv_1,z)}{\hat{A}_2(v_2,z)^2\hat{A}_2(v_1,z)}$$
$$- \int_0^{y_j}\int_0^{y_k} 1(v_2 \ge v_1)\frac{\hat{A}_1(dv_2,z)\hat{A}_1(dv_1,z)}{\hat{A}_2(v_2,z)\hat{A}_2(v_1,z)^2}$$
$$+ \int_0^{y_j}\int_0^{y_k} \big(1(v_1 \ge v_2)\hat{A}_2(v_1,z) + 1(v_2 \ge v_1)\hat{A}_2(v_2,z)\big)\frac{\hat{A}_1(dv_2,z)\hat{A}_1(dv_1,z)}{\hat{A}_2(v_2,z)^2\hat{A}_2(v_1,z)^2}$$
$$= \int_0^{\min(y_j,y_k)} \frac{\hat{A}_1(dv,z)}{\hat{A}_2(v,z)^2},$$

and the covariance function given in the theorem follows. ∎

A.4 Proof of Theorem 6

Lemma 4 below extends lemma 1 to the case of kernel smoothing over both Z and Y. Like lemma 1 it is stated without proof.

Given some function f and $(y,z) \in \hat{\mathcal{M}}$, where $\hat{\mathcal{M}}$ is a bounded set, define

$$\phi_n(\mathbf{y},\mathbf{s},\mathbf{x}) = \frac{K_z\big(b_z^{-1}(z-\mathbf{x}'\beta_n)\big)}{b_z}\frac{K_y\big(b_y^{-1}(y-\mathbf{y})\big)}{b_y}1(\mathbf{s}=s)f(y,z),$$

$$\varphi_n(\mathbf{y},\mathbf{s},\mathbf{x}) = \frac{K_z\big(b_z^{-1}(z-\mathbf{x}'\beta)\big)}{b_z}\frac{K_y\big(b_y^{-1}(y-\mathbf{y})\big)}{b_y}1(\mathbf{s}=s)f(y,z),$$

$$\phi = f(y,z)\hat{a}_1(y,z),$$

$$\tilde{\phi} = f(y,z)\int \mathbf{x}_{-1}\partial_2\hat{A}_1^*(y,z,d\mathbf{x}_{-1}).$$

Convergence of $P_n\phi_n$ is established in lemma 4.

Lemma 4 *Suppose assumptions 11, 12, 13, and 14 hold, and that f is bounded. Then*

i. *If $n^{-1/2}b_y^{-1/2}b_z^{-1/2}\log n$ is bounded, then $\sup P_n|\phi_n| = O_p\big(n^{-1/2}b_y^{-1}b_z^{-2}\big)+O(1)$ as $n\to\infty$.*
ii. *If $n^{-1/2}b_y^{-1/2}b_z^{-3/2}\log n$ is bounded, then $\sup\big|P_n\phi_n - P_n\varphi_n\big| = O_p\big(n^{-1/2}\big) + O_p\big(n^{-1}b_y^{-1}b_z^{-3}\big)$ as $n\to\infty$.*
iii. *If $n^{-1/2}b_y^{-1/2}b_z^{-3/2}\log n$ is bounded as $n\to\infty$, then $\sup\big|P_n\phi_n - P_n\varphi_n - \tilde{\phi}'P_n\Omega\big| = o_p\big(n^{-1/2}\big)+O_p\big(n^{-1}b_y^{-1}b_z^{-3}\big)$.*
iv. *$\sup\big|P_n\varphi_n-\phi\big| = o\big(n^{-1/2}b_y^{-1/2}b_z^{-1/2}\log n\big)+O\big(b_y^{k_y}\big)+O\big(b_z^{k_z}\big)$ almost surely as $n\to\infty$.*
v. *If $n^{-1/2}b_y^{-1/2}b_z^{-3/2}\log n$ is bounded, then $\sup\big|P_n\phi_n-\phi\big| = O_p\big(n^{-1}b_y^{-1}b_z^{-3}\big) + o\big(n^{-1/2}b_y^{-1/2}b_z^{-1/2}\log n\big)+O\big(b_y^{k_y}\big)+O\big(b_z^{k_z}\big)$ almost surely as $n\to\infty$.*

Let $(y_j,z_j)\in\hat{\mathcal{M}}$, $j=1,2,\dots,J$ and let ϕ_{nj} and ϕ_j denote the corresponding ϕ_n and ϕ. Define

$$C_{jk} = 1(y_j=y_k)1(z_j=z_k)f(y_j,z_j)^2\hat{a}_1(y_j,z_j)\int K_y(t)^2\,dt\int K_z(u)^2\,du.$$

vi. *If $n^{-1/2}b_y^{-1/2}b_z^{-3/2}\log n$ is bounded, $n^{-1/2}b_y^{-1/2}b_z^{-5/2}\to 0$, and $n^{1/2}b_y^{1/2}b_z^{1/2+k_z}\to 0$ and $n^{1/2}b_y^{1/2+k_y}b_z^{1/2}\to 0$ as $n\to\infty$, then $\sqrt{nb_yb_z}\big(P_n\phi_{n1}-\phi_1,\dots,P_n\phi_{nJ}-\phi_J\big)\to^d N$, where N is a multivariate normal random vector with mean 0 and covariance matrix $C=[C_{jk}]$.*

Proof of Theorem 6 Let $\hat{\pi}_{1n}$ and $\hat{\Pi}_{2n}$ be defined as $\hat{a}_{1n}$ and $\hat{A}_{2n}$ with $Z_i = X_i'\beta$ replacing $Z_{in} = X_i'\beta_n$. Since $n^{-1/2}b_z^{-3/2}\log n$ is bounded as $n\to\infty$, lemmas 1.ii, 1.iv and 1.v imply

$$\sup_{\hat{\mathcal{M}}}\big|\hat{A}_{2n}-\hat{\Pi}_{2n}\big| = O_p\big(n^{-1/2}\big)+O_p\big(n^{-1}b_z^{-3}\big), \tag{A}$$

$$\sup_{\hat{\mathcal{M}}}\big|\hat{\Pi}_{2n} - \hat{A}_2\big| = O\big(b_z^{k_z}\big) + o\big(n^{-1/2}b_z^{-1/2}\log n\big) \quad \text{a.s.,} \tag{B}$$

$$\sup_{\hat{\mathcal{M}}}\big|\hat{A}_{2n} - \hat{A}_2\big| = O\big(b_z^{k_z}\big) + o\big(n^{-1/2}b_z^{-1/2}\log n\big) \quad \text{a.s.} \tag{C}$$

Similarly, since $n^{-1/2}b_y^{-1/2}b_z^{-3/2}\log n$ is bounded as $n \to \infty$, lemas 4.ii and 4.iv imply

$$\sup_{\hat{\mathcal{M}}}\big|\hat{a}_{1n} - \hat{\pi}_{1n}\big| = O_p\big(n^{-1/2}\big) + O_p\big(n^{-1}b_y^{-1}b_z^{-3}\big) \tag{D}$$

$$\sup_{\hat{\mathcal{M}}}\big|\hat{\pi}_{1n} - \hat{a}_1\big| = O\big(b_y^{k_y}\big) + O\big(b_z^{k_z}\big) + o\big(n^{-1/2}b_y^{-1/2}b_z^{-1/2}\log n\big) \quad \text{a.s.} \tag{E}$$

The first step is to prove that conditioning on β_n instead of β does not affect the asymptotic distribution. Let $\epsilon > 0$ be a small number. By (C) and (B) $\sup|\hat{A}_{2n} - \hat{A}_2| < \epsilon$ and $\sup|\hat{\Pi}_{2n} - \hat{A}_2| < \epsilon$ for all large n. By assumption $\hat{A}_2$ is bounded below on $\hat{\mathcal{M}}$. This means that $\hat{A}_{2n}$ and $\hat{\Pi}_{2n}$ are bounded below as well. Moreover, boundedness of $\hat{a}_1$ and (E) implies boundedness of $\hat{\pi}_{1n}$. Therefore, by (A) and (D),

$$\begin{aligned}\sup_{\hat{\mathcal{M}}}\left|\frac{\hat{a}_{1n}}{\hat{A}_{2n}} - \frac{\hat{\pi}_{1n}}{\hat{\Pi}_{2n}}\right| &\le \sup_{\hat{\mathcal{M}}}\left|\frac{\hat{a}_{1n} - \hat{\pi}_{1n}}{\hat{A}_{2n}}\right| + \sup_{\hat{\mathcal{M}}}\left|\frac{\big(\hat{A}_{2n} - \hat{\Pi}_{2n}\big)\hat{\pi}_{1n}}{\hat{A}_{2n}\hat{\Pi}_{2n}}\right| \\ &= O_p\big(n^{-1/2}\big) + O_p\big(n^{-1}b_y^{-1}b_z^{-3}\big).\end{aligned}$$

Next step is to show that asymptotically the double-smoothed $\hat{\pi}_{1n}$ dominates the single-smoothed $\hat{\Pi}_{2n}$. By (B),

$$\sup_{\hat{\mathcal{M}}}\left|\frac{\hat{\pi}_{1n}}{\hat{\Pi}_{2n}} - \frac{\hat{\pi}_{1n}}{\hat{A}_2}\right| = \sup_{\hat{\mathcal{M}}}\left|\frac{\hat{\pi}_{1n}\big(\hat{\Pi}_{2n} - \hat{A}_2\big)}{\hat{\Pi}_{2n}\hat{A}_2}\right| = O\big(b_z^{k_z}\big) + o\big(n^{-1/2}b_z^{-1/2}\log n\big) \quad \text{a.s.}$$

It follows that

$$\begin{aligned}&\sup_{\hat{\mathcal{M}}}\big|\hat{h}_n - \hat{h}\big| \\ &\qquad = \sup_{\hat{\mathcal{M}}}\left|\frac{\hat{\pi}_{1n}}{\hat{A}_2} - \frac{\hat{a}_1}{\hat{A}_2}\right| + O_p\big(n^{-1}b_y^{-1}b_z^{-3}\big) + O\big(b_z^{k_z}\big) + o\big(n^{-1/2}b_z^{-1/2}\log n\big) \quad \text{a.s.}\end{aligned}$$

Given (y,z), define $f(y,z) = 1/\hat{A}_2(y,z)$, $\varphi_n(\mathbf{s},\mathbf{y},\mathbf{x}) = f(y,z)K_{zn}(z - \mathbf{x}'\beta)K_{yn}(y - \mathbf{y})$ and $\phi = \hat{a}_1(y,z)/\hat{A}_2(y,z)$, then $P_n\varphi_n = \hat{\pi}_{1n}(y,z)/\hat{A}_2(y,z)$. Consistency and asymptotic normality now follows from lemmas 4.v and 4.vi. ■

References

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *Annals of Statistics 6*, 701–726.

Ai, C. (1997). A semiparametric maximum likelihood estimator. *Econometrica 65*(4), 933–963.

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, Department of Statistics, University of California, Berkeley.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society 34B*, 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika 62*, 269–276.

Dabrowska, D. M. (1987). Nonparametric regression with censored survival time data. *Scandinavian Journal of Statistics. Theory and Applications 14*(3), 181–197.

Gørgens, T. and J. Horowitz (1999). Semiparametric estimation of a censored regression model with an unknown transformation of the dependent variable. *Journal of Econometrics 90*(2), 155–191.

Han, A. K. (1987). Non-parametric analysis of a generalized regression model. *Journal of Econometrics 35*, 303–316.

Härdle, W. and T. M. Stoker (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association 84*, 986–995.

Härdle, W. and A. B. Tsybakov (1993). How sensitive are average derivatives? *Journal of Econometrics 58*(1-2), 31–48.

Hastie, T. and R. Tibshirani (1986). Generalized additive models. *Statistical Science 1*(3), 297–318. With discussion.

Hastie, T. and R. Tibshirani (1990a). Expoloring the nature of covariate effects in the proportional hazards model. *Biometrics 46*, 1005–1016.

Hastie, T. J. and R. J. Tibshirani (1990b). *Generalized Additive Models.* London: Chapman and Hall Ltd.

Horowitz, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica 64*(1), 103–137.

Horowitz, J. L. (1999). Semiparametric estimation of a proportional hazard model with unobserved heterogeneity. *Econometrica 67*(5), 1001–1028.

Horowitz, J. L. and W. Härdle (1996). Direct semiparametric estimation of single-index models with discrete covariates. *Journal of the American Statistical Association 91*(436), 1632–1640.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single index models. *Journal of Econometrics 58*, 71–120.

Kaplan, E. and P. Meier (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association 53*, 457–481.

Klein, R. W. and R. H. Spady (1993, March). An efficient semiparametric estimator for binary repsonse models. *Econometrica 61*(2), 387–421.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge; New York: Cambridge University Press.

Manski, C. F. (1988). Identification of binary response models. *Journal of the American Statistical Association 83*, 729–738.

McKeague, I. W. and K. J. Utikal (1990). Inference for a nonlinear counting process regression model. *Annals of Statistics 18*(3), 1172–1187.

Müller, H.-G. and J.-L. Wang (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics 50*(1), 61–76.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics 14*, 945–966.

Nielsen, J. P. and O. B. Linton (1995). Kernel estimation in a nonparametric marker dependent hazard model. *Annals of Statistics 23*(5), 1735–1748.

Nolan, D. and D. Pollard (1987). U-processes: Rates of convergence. *Annals of Statistics 15*(2), 780–799.

Pagan, A. and A. Ullah (1998). Non-parametric econometrics. Manuscript.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*, 1027–1057.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer-Verlag.

Powell, J. L., J. H. Stock, and T. M. Stoker (1989). Semiparametric estimation of index coefficients. *Econometrica 57*(6), 1403–1430.

Powell, J. L. and T. M. Stoker (1996). Optimal bandwidth choice for density-weighted averages. *Journal of Econometrics 75*(2), 291–316.

Ramlau-Hansen, H. (1983). Smoothing counting process intensitites by means of kernel functions. *Annals of Statistics 11*(2), 453–466.

Rice, J. and M. Rosenblatt (1976). Estimation of the log survivor function and hazard function. *Sankhya Series A 38*, 60–78.

Sherman, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica 61*(1), 123–137.

Watson, G. S. and M. R. Leadbetter (1964a). Hazard analysis I. *Biometrika 51*, 175–184.

Watson, G. S. and M. R. Leadbetter (1964b). Hazard analysis II. *Sankhyā Series A 26*, 110–116.

Yandell, B. S. (1983). Nonparametric inference for rates with censored survival data. *Annals of Statistics 11*(4), 1119–1135.