Tilburg University

Tilburg University

Jump-Preserving Varying-Coefficient Models for Nonlinear Time Series

Cizek, Pavel; Koo, Chao

Publication date:
2017

Document Version
Early version, also known as pre-print

Link to publication in Tilburg University Research Portal

JUMP-PRESERVING VARYING-COEFFICIENT MODELS FOR NONLINEAR TIME SERIES

By

Pavel Čížek, Chao Hui Koo

TILBURG ◆ UNIVERSITY

Jump-Preserving Varying-Coefficient Models for Nonlinear Time Series

Pavel Čížek[*] and Chao Hui Koo[†]

December, 2016

Abstract

An important and widely used class of semiparametric models is formed by the varying-coefficient models. Although the varying coefficients are traditionally assumed to be smooth functions, the varying-coefficient model is considered here with the coefficient functions containing a finite set of discontinuities. Contrary to the existing nonparametric and varying-coefficient estimation of piecewise smooth functions, the varying-coefficient models are considered here under dependence and are applicable in time series with heteroscedastic and serially correlated errors. Additionally, the conditional error variance is allowed to exhibit discontinuities at a finite set of points too. The (uniform) consistency and asymptotic normality of the proposed estimators are established and the finite-sample performance is tested via a simulation study.

Keywords: Change point; Heteroscedasticity; Local linear fitting; Nonlinear time series; Varying-coefficient models.

JEL Classification Numbers: C13, C14, C22.

---

[*]CentER, Department of Econometrics & OR, Tilburg University, P.O.Box 90153, 5000LE Tilburg, The Netherlands. E-mail: `P.Cizek@uvt.nl`.

[†]Department of Econometrics & OR, Tilburg University, P.O.Box 90153, 5000LE Tilburg, The Netherlands. E-mail: `C.Koo@uvt.nl`.

1   Introduction

The varying-coefficient models (VCM) form an important class of semiparametric models (see Hastie and Tibshirani, 1993; Cai et al., 2000) that assume the marginal effects of covariates to be an unknown function of an observable index variable. Practically, VCMs are formulated as linear models with coefficients being general functions of the index variable. Most existing literature assumes the coefficient functions to be continuous and smooth. In this paper, we however allow coefficient functions to contain a finite set of discontinuities; additionally, discontinuities can be present also in the conditional error variance. To the best of our knowledge, VCMs with discontinuities in coefficient functions have not been investigated before in heteroscedastic and time series setting. For independent and identically distributed data, Zhu et al. (2014) and Zhao et al. (2016) suggested methods for estimation of varying-coefficient models with discontinuities.

There is a vast amount of literature on VCMs when coefficients are smooth continuous functions. Recent works include Hoover et al. (1998), Wu et al. (1998), and Fan and Zhang (2000) on longitudinal data analysis, Cai et al. (2000) and Huang and Shen (2004) on nonlinear time series, and Cai and Li (2008) and Sun et al. (2009) on panel data analysis. Additionally, hybrids of varying-coefficient models have also been developed: for example, partial linearly varying-coefficient models where some coefficient functions are constant (Zhang et al., 2002; Fan and Huang, 2005; Ahmad et al., 2005; Lee and Mammen, 2016), generalized linear models with varying coefficients (Cai et al., 2000), and varying-coefficient models in which the varying index is latent and estimated as a linear combination of several observed variables (Fan et al., 2003).

Although very few studies on VCMs allow discontinuities in coefficient functions, literature on nonparametric estimation of discontinuous regression function is extensive. The classical estimation procedures usually consist of two stages. The locations of discontinuities are first estimated and then a conventional nonparametric estimator, which

assumes the underlying function to be continuous, is used within each segment between two consecutive discontinuities to estimate the regression function itself. Examples of this approach include Müller (1992), Wu and Chu (1993), Kang et al. (2000), and Gijbels and Goderniaux (2004).

There are other techniques that do not estimate first the locations of discontinuities in a nonparametric regression; see, for example Godtliebsen et al. (1997) on nonlinear Gaussian filtering and Spokoiny (1998) and Polzehl and Spokoiny (2000) on adaptive weights smoothing. Besides these approaches, Gijbels et al. (2007) recently proposed an estimation method based on three local linear estimators in the framework of fixed design and homoscedastic errors. At each design point z, they considered local linear estimates using data from the left-, right-, and two-sided neighborhoods of z. The final estimate of the conditional mean of the response equals one of these three local linear estimates chosen by comparing the weighted residual mean squared errors of three local linear fits. This approach was extended to conditional variance estimation by Casas and Gijbels (2012).

We generalize the estimation procedure by Gijbels et al. (2007) in two directions. First, we extend Gijbels et al. (2007) estimation method based on a comparison of the weighted residual mean squared errors to the VCMs, where discontinuities might occur only in one, few, or all coefficients. Although this has already been done by Zhao et al. (2016) in the case of independently and identically sampled observations, we analyze this method in the context of heteroscedastic and dependent data and provide additional asymptotic results such as the uniform convergence rate of the coefficient estimates. Second, as the method is shown to work well only if the conditional variance function of the error term is continuous, we propose an alternative measure of the three local linear fits based on the local Wald test statistics such that the proposed method is applicable even if the conditional variance function of the error term contains discontinuities.

This paper is structured as follows. In Section 2, the VCM is introduced and the jump-preserving estimation procedure is introduced based on Gijbels et al. (2007) and Zhao et al. (2016). In Section 3, we establish the consistency and asymptotic normality of this estimator. In Section 4, an alternative estimator that does not require the continuity of conditional error variance is proposed and its asymptotic properties are derived. Finally, the finite sample properties of the two proposed estimators are investigated by means of a simulation study in Section 5. Proofs can be found in Appendices A and B.

2 The discontinuous varying-coefficient model

The varying-coefficient regression model takes the following form:

$$Y_i = X_i^\top a(Z_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where Y_i is the response variable, X_i is a $p \times 1$ vector of covariates, Z_i is a scalar index variable, $a(\cdot)$ is a $p \times 1$ vector of unspecified coefficient functions, and ε_i is an error term such that $\mathrm{E}[\varepsilon_i|X_i, Z_i] = 0$ and $\mathrm{E}[\varepsilon_i^2|X_i, Z_i] = \sigma^2(X_i, Z_i)$. Note that both X_i and Z_i can contain lagged values of Y_i. In this paper, we consider piecewise-smooth coefficient functions $a(\cdot)$ that can exhibit a finite set of discontinuities located at points $\{s_q\}_{q=1}^Q$, where the number Q of jumps, the jump locations s_q, and the jump sizes d_q of the coefficient functions are all unknown. Contrary to Zhao et al. (2016), we assume that the conditional variance $\sigma^2(z) = \mathrm{E}[\sigma^2(X, Z)|Z = z]$ is not constant, but it is a continuous function of z in this section. The case with discontinuous $\sigma^2(z)$ will be investigated later in Section 4.

The semiparametric model (1) has been studied by Zhao et al. (2016) for the independent and identically distributed data, and in the present setting, it includes many popular time-series models. When X_i is a constant, the model is reduced to a nonpara-

metric jump-preserving model in Gijbels et al. (2007). If all coefficient functions are constant, the model becomes a linear (possibly autoregressive) model. If the coefficient functions have the form: $a(\cdot) = \beta_1 w(\cdot) + \beta_2 \{1 - w(\cdot)\}$ with $w(\cdot)$ being an unspecified scalar function, model (1) covers semiparametric transition models such as the one by Čížek and Koo (2015), who estimated $w(\cdot)$ by a jump-preserving estimation proposed in this work. Moreover, model (1) includes the threshold autoregressive model and the smooth transition autoregressive model when $w(\cdot)$ takes a particular parametric form.

To define first the estimator of coefficient functions $a(\cdot)$ analogous to Gijbels et al. (2007) and Zhao et al. (2016), we let $K^{(c)}(\cdot)$ be a conventional bounded symmetric kernel function with a compact support $[-1, 1]$ and define $K^{(l)}(\cdot)$ and $K^{(r)}(\cdot)$ to be the corresponding left-sided and right-sided kernels, respectively, given by

$$K^{(l)}(v) = K^{(c)}(v) \cdot \mathbf{1}\{v \in [-1, 0)\} \quad \text{and} \quad K^{(r)}(v) = K^{(c)}(v) \cdot \mathbf{1}\{v \in [0, 1]\}, \qquad (2)$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function. Using these kernels, we can define three pairs of local linear estimators $\hat{a}_n^{(\iota)}(z)$ and $\hat{b}_n^{(\iota)}(z)$ ($\iota = c, l, r$) of coefficient functions $a(\cdot)$ and its derivatives $a'(\cdot)$, respectively, at a fixed point z:

$$\left[\hat{a}_n^{(\iota)}(z), \hat{b}_n^{(\iota)}(z)\right] = \arg\min_{a,b} \sum_{i=1}^n \left\{Y_i - X_i^\top [a + b(Z_i - z)]\right\}^2 K_h^{(\iota)}(Z_i - z), \quad \iota = c, l, r, \quad (3)$$

where $K_h^{(\iota)}(\cdot) = h_n^{-1} K^{(\iota)}(\cdot / h_n)$, $h_n > 0$ is a bandwidth such that $h_n \to 0$ as $n \to \infty$, and the superscript $\iota = c, l, r$ indicates whether the conventional, left-sided, or right-sided

kernel is used. Solving the least-squares minimization problem (3) for $\iota = c, l, r$ yields

$$
\begin{pmatrix} \hat{a}_n^{(\iota)}(z) \\ \hat{b}_n^{(\iota)}(z) \end{pmatrix} = \left[\sum_{i=1}^{n} \begin{pmatrix} X_i \\ X_i(Z_i - z) \end{pmatrix} \begin{pmatrix} X_i \\ X_i(Z_i - z) \end{pmatrix}^{\top} K_h^{(\iota)}(Z_i - z) \right]^{-1}
$$
$$
\sum_{i=1}^{n} \begin{pmatrix} X_i \\ X_i(Z_i - z) \end{pmatrix} Y_i K_h^{(\iota)}(Z_i - z). \tag{4}
$$

To measure the quality of each local linear fit, Gijbels et al. (2007) and Zhao et al. (2016) advocate the use of the weighted residual mean squared error (WRMSE):

$$
\Psi_n^{(\iota)}(z) = \frac{\sum_{i=1}^{n} \hat{\varepsilon}_{n,i}^{(\iota)^2} K_h^{(\iota)}(Z_i - z)}{\sum_{i=1}^{n} K_h^{(\iota)}(Z_i - z)}, \quad \iota = c, l, r, \tag{5}
$$

where the estimated residual $\hat{\varepsilon}_{n,i}^{(\iota)} = Y_i - X_i^{\top}\{\hat{a}_n^{(\iota)}(z) + \hat{b}_n^{(\iota)}(z)(Z_i - z)\}$. WRMSE is an estimator of conditional error variance $\sigma^2(z)$, which is similar to the one proposed in Fan and Yao (1998) except that the local constant fitting of $\hat{\varepsilon}_{n,i}^{(\iota)^2}$ and same bandwidth h_n for the conditional variance are used here. Although employing a different bandwidth for the conditional variance would improve the finite sample performance, our aim is to compare performance of the three local estimates of $a(z)$ rather than providing a good estimate of $\sigma^2(z)$. To avoid technical complexity in the proofs, the same bandwidth is therefore applied for the coefficient functions and WRMSE estimates.

The WRMSE introduced in (5) can be now used to select the consistent estimator out of (3) and thus to define the jump-preserving estimator of $a(z)$, which will be proved

consistent if the conditional error variance $\sigma^2(z)$ is continuous (cf. Zhao et al., 2016):

$$\check{a}_n(z) = \begin{cases} \hat{a}_n^{(c)}(z), & \text{if } \text{diff}(z) \leq u_n, \\[2mm] \hat{a}_n^{(l)}(z), & \text{if } \text{diff}(z) > u_n \text{ and } \Psi_n^{(l)}(z) < \Psi_n^{(r)}(z), \\[2mm] \hat{a}_n^{(r)}(z), & \text{if } \text{diff}(z) > u_n \text{ and } \Psi_n^{(l)}(z) > \Psi_n^{(r)}(z), \\[2mm] \dfrac{\hat{a}_n^{(l)}(z) + \hat{a}_n^{(r)}(z)}{2}, & \text{if } \text{diff}(z) > u_n \text{ and } \Psi_n^{(l)}(z) = \Psi_n^{(r)}(z), \end{cases} \qquad (6)$$

where $\text{diff}(z) = \Psi_n^{(c)}(z) - \min\{\Psi_n^{(l)}(z), \Psi_n^{(r)}(z)\}$ and the auxiliary parameter $u_n > 0$ is tending to zero, $u_n \to 0$ as $n \to \infty$. The intuition behing this proposal is based on the fact that the conventional local estimate $\hat{a}_n^{(c)}(z)$ should be the most precise one as it uses all observations in the interval $[z - h_n, z + h_n]$, but it is consistent only if there are no discontinuities in $(z - h_n, z + h_n)$. If $a(\cdot)$ is discontinuous at some point of $(z - h_n, z + h_n)$, $\hat{a}_n^{(c)}(z)$ is generally inconsistent (and the same can be also true in the case of $\hat{a}_n^{(l)}(z)$ or $\hat{a}_n^{(r)}(z)$), which leads to an increase of the corresponding WRMSE value in (5) as we confirm later in Section 3. Consequently, only a consistent estimator will minimize (5) asymptotically and will be thus selected in (6). The existence of a consistent estimator among $\hat{a}_n^{(c)}(z)$, $\hat{a}_n^{(l)}(z)$, and $\hat{a}_n^{(r)}(z)$ can be however assumed as bandwidth $h_n \to 0$ as $n \to \infty$ and the interval $(z - h_n, z + h_n)$ thus contains at most one point of discontinuity for any z and a sufficiently large n. See Zhao et al. (2016) for more details.

3   Asymptotic results

To derive the asymptotic properties of the proposed jump-preserving estimator, the assumptions about the data generating process (1) have to be detailed first. Later, the requirements on the kernel function and bandwidth are specified too.

Let us now define the α-mixing and the assumptions on the model (1). Suppose that

$\mathcal{F}_a^b$ is the σ-algebra generated by $\{\xi_i; a \leq i \leq b\}$. The α-mixing coefficient of the process $\{\xi_i\}_{i=-\infty}^{\infty}$ is defined as

$$\alpha(m) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_m^{\infty}\}.$$

If $\alpha(m) \to 0$ as $m \to \infty$, then the process $\{\xi_i\}_{i=-\infty}^{\infty}$ is called strong mixing or α-mixing. In the following assumptions, we additionally denote by $f(\cdot, \cdot)$ the joint probability density function of variables X_i and Z_i and by $f_Z(\cdot)$ the marginal density function of Z_i.

Assumption A.

A1. The process $\{X_i, Z_i, \varepsilon_i\}$ is strictly stationary and strong mixing with α-mixing coefficients $\alpha(m)$, $m \in \mathbb{N}$, that satisfy $\alpha(m) \leq Cm^{-\gamma}$ with $0 < C < \infty$ and $\gamma > (2\delta - 2)/(\delta - 2)$ for some $\delta > 2$.

A2. There is a compact set $D = [s_0, s_{Q+1}]$ such that $\inf_{z \in D} f_Z(z) > 0$. The derivative of $f_Z(\cdot)$ is bounded and Lipschitz continuous for $z \in D$. The partial derivative of the joint density function $f(\cdot, \cdot)$ with respect to Z is bounded and continuous uniformly on the support of X and D except for the points $\{s_q\}_{q=0}^{Q+1}$, at which the left and right partial derivatives of $f(\cdot, \cdot)$ with respect to Z are bounded and left and right continuous, respectively.

A3. Let φ_i represent any element of matrix $X_i X_i^\top$, vector $X_i \varepsilon_i$, or variable ε_i^2. For δ given in Assumption A1,

   (i) $E|\varphi_i|^{\delta} < \infty$,

   (ii) $\sup_{z \in D} E(|\varphi_i|^{\delta}|Z_i = z)f_Z(z) < \infty$,

   (iii) for all integers $j > 1$,

$$\sup_{(z_1, z_j) \in D \times D} E(|\varphi_1 \varphi_j| | Z_1 = z_1, Z_j = z_j) f_{Z_1 Z_j}(z_1, z_j) < \infty,$$

8

where $f_{Z_1 Z_j}(z_1, z_j)$ denotes the joint density of (Z_1, Z_j).

A4. The variance matrix $\Omega(z) = \mathrm{E}[XX^\top | Z = z]$ is bounded and positive definite uniformly on D except for the discontinuities $\{s_q\}_{q=0}^{Q+1}$, at which $\Omega_-(s_q) = \lim_{z \uparrow s_q} \mathrm{E}[XX^\top | Z = z]$ and $\Omega_+(s_q) = \lim_{z \downarrow s_q} \mathrm{E}[XX^\top | Z = z]$ are bounded and positive definite.

A5. The second-order partial derivatives of $a(z)$ are bounded and Lipschitz continuous on D except for the discontinuities $\{s_q\}_{q=0}^{Q+1}$, at which the left and right second-order partial derivatives of $a(z)$ are bounded and left and right Lipschitz continuous, respectively.

A6. The partial derivative of $\sigma^2(X, Z)$ with respect to Z is bounded and continuous on D.

Assumptions A1–A5 are standard conditions for the VCMs with dependent data (see e.g. Conditions A.1 and A.2 in Cai et al. (2000) for the local linear estimation in VCMs and the assumptions in Hansen (2008) for a general nonparametric kernel estimator) adapted for discontinuities, at which we impose the corresponding conditions for the left and right limits. Further, Assumption A6 imposes that the conditional variance $\sigma^2(z) = \mathrm{E}[\sigma^2(X, Z) | Z = z]$ is continuous; the case with discontinuous $\sigma^2(z)$ is investigated in Section 4.

The following assumptions about the kernel K, bandwidth h_n, auxiliary parameter u_n, and mixing exponent γ are also needed to show the asymptotic results for the jump-preserving estimator $\breve{a}_n(z)$. First, standard assumptions on the kernel and bandwidth are given. After that, assumptions required by Hansen (2008) in the asymptotic analysis of the local linear regression estimators under dependence are introduced.

Assumption B.

B1. The kernel $K^{(c)}(\cdot)$ is a bounded symmetric continuous density function and has a compact support $[-1, 1]$. It is chosen so that the following constants are well defined and finite for $j = 0, 1, 2$ and $\iota = c, r, l$:

$$\mu_j^{(\iota)} = \int_{-1}^{1} v^j K^{(\iota)}(v)\mathrm{d}v, \qquad \nu_j^{(\iota)} = \int_{-1}^{1} v^j K^{(\iota)2}(v)\mathrm{d}v,$$

$$c_0^{(\iota)} = \frac{\mu_2^{(\iota)}}{\mu_2^{(\iota)}\mu_0^{(\iota)} - \mu_1^{(\iota)2}}, \quad \text{and} \quad c_1^{(\iota)} = \frac{-\mu_1^{(\iota)}}{\mu_2^{(\iota)}\mu_0^{(\iota)} - \mu_1^{(\iota)2}}. \tag{7}$$

B2. The bandwidths h_n and u_n satisfy $u_n \to 0$, $h_n \to 0$, and $nh_n \to \infty$ as $n \to \infty$.

B3. Additionally, $nh_n^5 \to \bar{c} \in [0, +\infty)$ as $n \to \infty$, where $\bar{c}$ is some non-negative constant.

Assumption C.

C1. The functions $\mathcal{K}_j^{(c)}(u) = u^j K^{(c)}(u)$ are Lipschitz continuous for all $j = 0, 1, 2, 3$.

C2. For some $\varsigma \geq 1$, the strong mixing exponent γ given in Assumption A1 satisfies

$$\gamma > \frac{1 + (\delta - 1)(2 + 1/\varsigma)}{\delta - 2}.$$

C3. The bandwidth h_n satisfies $\ln n/(nh_n^3) = o(1)$ and $\ln n/(n^\theta h_n) = o(1)$, where

$$\theta = \frac{\gamma - 2 - \dfrac{1}{\varsigma} - \dfrac{1+\gamma}{\delta - 1}}{\gamma + 2 - \dfrac{1+\gamma}{\delta - 1}}.$$

Before providing the asymptotic properties of the jump-preserving estimator $\check{a}_n(z)$, we study the behavior of the three local linear estimators (3) in the continuous region

10

and in the neighborhoods of discontinuities. The regions of continuity are defined by

$$D_{1n} = D_{1n}^{(c)} = D \setminus \bigcup_{q=0}^{Q+1} [s_q - h_n, s_q + h_n],$$

$$D_{1n}^{(l)} = D \setminus \bigcup_{q=0}^{Q+1} [s_q, s_q + h_n], \quad \text{and} \quad D_{1n}^{(r)} = D \setminus \bigcup_{q=0}^{Q+1} [s_q - h_n, s_q],$$

where $D = [s_0, s_{Q+1}]$ is a compact region with a positive probability.

Theorem 1. *Under Assumptions A1–A6, B, and C, it holds for $n \to \infty$ that*

$$\sup_{z \in D_{1n}^{(\iota)}} \left\| \hat{a}_n^{(\iota)}(z) - a(z) \right\| = O_p \left(\sqrt{\frac{\ln n}{nh_n}} \right), \quad \iota = c, l, r.$$

Theorem 2. *If Assumptions A1–A6 and B are satisfied and a fixed point $z \in D_{1n}^{(\iota)}$ for some $n \in \mathbb{N}$ and $\iota = c, l, r$, it holds that*

$$\sqrt{nh_n} \left[\hat{a}_n^{(\iota)}(z) - a(z) - \frac{h_n^2}{2} \left(c_0^{(\iota)} \mu_2^{(\iota)} + c_1^{(\iota)} \mu_3^{(\iota)} \right) a''(z) \right] \xrightarrow{d} N \left(0, \Phi^{(\iota)}(z) \right)$$

as $n \to \infty$, where

$$\Phi^{(\iota)}(z) = \frac{c_0^{(\iota)2} \nu_0^{(\iota)} + 2c_0^{(\iota)} c_1^{(\iota)} \nu_1^{(\iota)} + c_1^{(\iota)2} \mu_2^{(\iota)}}{f_Z(z)} \cdot \Omega^{-1}(z) \Theta(z) \Omega^{-1}(z), \quad (8)$$

$\Omega(z) = \mathrm{E}[XX^\top | Z = z]$, *and* $\Theta(z) = \mathrm{E}[XX^\top \sigma^2(X, Z) | Z = z]$.

Theorem 1 establishes the uniform consistency of the three local linear estimators in their corresponding continuous regions. Theorem 2 then specifies the asymptotic distributions of the estimators $\hat{a}_n^{(c)}(z)$, $\hat{a}_n^{(l)}(z)$, and $\hat{a}_n^{(r)}(z)$ in the regions, where $a(\cdot)$ is continuous, left-continuous, and right-continuous around z, respectively. Since all three local linear estimators are consistent in their corresponding regions of continuity according to Theorem 1, it is easy to see that their corresponding WRMSE estimates (5) consistently

converge to the conditional error variance $\sigma^2(z)$.

Theorem 3. *Let Assumptions A1–A6 and B hold. At any point $z \in D_{1n}^{(\iota)}$ for some $n \in \mathbb{N}$ and $\iota = c, l, r$, the mean squared error in (5) satisfies $\Psi_n^{(\iota)}(z) = \sigma^2(z) + \mathrm{o}_\mathrm{p}(1)$ as $n \to \infty$.*

Such a result does not however hold if the point z is close to a jump, that is, to a point of discontinuity. If a jump is located in the right neighborhood of z, only the left-sided local linear estimator $\hat{a}_n^{(l)}(z)$ is consistent. Similarly, the right-sided estimator $\hat{a}_n^{(r)}(z)$ is the only consistent estimator of $a(z)$ when there is a jump in the left neighborhood of z. Consequently, the three WRMSE estimates behave differently near a jump point. The next theorem describes the asymptotic behavior of WRMSE in a neighborhood of a jump s_q when the conditional error variance $\sigma^2(z)$ is continuous in z (cf. Zhao et al., 2016).

Theorem 4. *Let Assumptions A1–A6 and B hold. Then it holds as $n \to \infty$ that*

(i) *for any $z = s_q + \tau h_n \in D$ with $q = 1, \ldots, Q + 1$ and $\tau \in [-1, 0)$,*

$$\Psi_n^{(c)}(z) = \sigma^2(s_q) + d_q^\top C_\tau^{(c)} d_q + \mathrm{o}_\mathrm{p}(1),$$
$$\Psi_n^{(l)}(z) = \sigma^2(s_q) + \mathrm{o}_\mathrm{p}(1),$$
$$\Psi_n^{(r)}(z) = \sigma^2(s_q) + d_q^\top C_\tau^{(r)} d_q + \mathrm{o}_\mathrm{p}(1).$$

(ii) *for any $z = s_q + \tau h_n \in D$ with $q = 0, \ldots, Q$ and $\tau \in (0, 1]$,*

$$\Psi_n^{(c)}(z) = \sigma^2(s_q) + d_q^\top C_\tau^{(c)} d_q + \mathrm{o}_\mathrm{p}(1),$$
$$\Psi_n^{(l)}(z) = \sigma^2(s_q) + d_q^\top C_\tau^{(l)} d_q + \mathrm{o}_\mathrm{p}(1),$$
$$\Psi_n^{(r)}(z) = \sigma^2(s_q) + \mathrm{o}_\mathrm{p}(1).$$

In both cases, $d_q = \lim_{z \downarrow s_q} a(z) - \lim_{z \uparrow s_q} a(z)$ and $C_\tau^{(\iota)}, \iota = c, l, r$, represents a positive definite matrix defined in Appendix A, equation (40).

The above theorem shows that only the left-sided WRMSE is a consistent estimator of the conditional error variance $\sigma^2(z)$ if a jump in coefficients $a(z)$ occurs in the right neighborhood of z, while the other two WRMSE estimates contain strictly positive biases, which do not vanish asymptotically. Similarly, if a jump is in the left neighborhood of z, only the right-sided WRMSE leads to a consistent estimator of $\sigma^2(z)$. To sum up, the smallest WRMSE is – at least asymptotically – $\Psi_n^{(l)}(z)$ when a jump is in a right neighborhood of z and it is $\Psi_n^{(r)}(z)$ when a jump is in a left neighborhood of z. Hence, it is intuitively clear that the jump-preserving estimator $\check{a}_n(z)$ defined in (6) selects the appropriate local linear estimator at every point z for a sufficiently large n.

Based on this result, we will establish the consistency of $\check{a}_n(z)$ in the continuous region D_{1n}, in the neighborhoods of discontinuity points D_{2n}, and in the neighborhoods of discontinuity points excluding small regions around centers and around endpoints $D_{2n,\delta}$. These regions are defined as follows:

$$
D_{2n} = D \cap \bigcup_{q=0}^{Q+1} \{[s_q - h_n, s_q) \cup (s_q, s_q + h_n]\} \quad \text{and}
$$

$$
D_{2n,\delta} = D \cap \bigcup_{q=0}^{Q+1} \{[s_q - (1-\delta)h_n, s_q - \delta h_n] \cup [s_q + \delta h_n, s_q + (1-\delta)h_n]\} \tag{9}
$$

for some $\delta \in (0, 1/2)$.

Theorem 5. *If Assumptions A1–A6, B, and C are satisfied, it holds for $n \to \infty$ and some $\delta \in (0, 1/2)$ that*

(i)
$$
\sup_{z \in D_{1n}} \|\check{a}_n(z) - a(z)\| = O_p\left(\sqrt{\frac{\ln n}{nh_n}}\right),
$$

(ii)
$$
\sup_{z \in D_{2n,\delta}} \|\check{a}_n(z) - a(z)\| = O_p\left(\sqrt{\frac{\ln n}{nh_n}}\right), \quad and
$$

13

(iii) *for any $z \in D_{2n}$,*

$$\check{a}_n(z) = a(z) + \mathrm{O}_{\mathrm{p}}\left(\sqrt{\frac{\ln n}{nh_n}}\right).$$

Theorem 5 states that the jump-preserving estimator $\check{a}_n(z)$ is uniformly consistent on D_{1n} and $D_{2n,\delta}$ for some $\delta \in (0, 1/2)$. At a point $z \in D_{2n}$ arbitrarily close to a point of discontinuity, $\check{a}_n(z)$ is only pointwise consistent.

The jump-preserving estimator $\check{a}_n(z)$ selects consistently (i.e., with probability approaching to 1) the appropriate local linear estimator on D excluding the jump points, where each of these local linear estimators is asymptotically normal at any point $z \in D \setminus \{s_q\}_{q=0}^{Q+1}$ according to Theorem 2. The following theorem can therefore establish the asymptotic normality of the jump-preserving estimator $\check{a}_n(z)$ at $z \in D \setminus \{s_q\}_{q=0}^{Q+1}$ (see also Casas and Gijbels, 2012; Zhao et al., 2016, Theorems 3.1).

Theorem 6. *If Assumptions A1–A6, B, and C are satisfied and $z \in D \setminus \{s_0, \ldots, s_{Q+1}\}$, it holds that*

$$\sqrt{nh_n}\left[\check{a}_n(z) - a(z) - \frac{h_n^2}{2}\left(c_0^{(\iota)}\mu_2^{(\iota)} + c_1^{(\iota)}\mu_3^{(\iota)}\right)a''(z)\right] \xrightarrow{d} N\left(0, \Phi^{(\iota)}(z)\right)$$

as $n \to \infty$, where $\Phi^{(\iota)}(z)$ is defined in equation (8) and

$$\iota = \begin{cases} c, & \text{if } z \in D_{1n}, \\ l, & \text{if } z \in D \cap \bigcup_{q=0}^{Q+1} [s_q - h_n, \ s_q), \\ r, & \text{if } z \in D \cap \bigcup_{q=0}^{Q+1} (s_q, \ s_q + h_n]. \end{cases}$$

4 Discontinuous conditional variance function

In this section, the conditional variance function $\sigma^2(z)$ is also allowed to exhibit discontinuities. For this purpose, we replace Assumption A6 by the following condition.

Assumption A6'. The partial derivative of $\sigma^2(X, Z)$ with respect to Z is bounded and continuous on D except for the points of discontinuity $\{s_q^2\}_{q=0}^{Q2+1}$, at which the left and right partial derivatives of $\sigma^2(X, Z)$ with respect to Z are bounded and left and right continuous, respectively. Without loss of generality, let $\{s_q\}_{q=0}^{Q+1} \subseteq \{s_q^2\}_{q=0}^{Q2+1}$.

Although Assumption A6' does not influence the consistency and convergence rates of the three local estimators (3), it can adversely affect the selection rule (6) based on a comparison of the three WRMSE estimates. In particular, if $\sigma^2(z)$ exhibits a jump at (or nearby) s_q, the error variances and thus WRMSE estimates are different in the left and right neighborhoods of the estimation point z. Hence, the limits of $\Psi_n^{(c)}(z)$, $\Psi_n^{(l)}(z)$, and $\Psi_n^{(r)}(z)$ in Theorem 4 contain different variances – error variance to the left of s_q, to the right of s_q, or a combination of those – and it is no longer possible to claim that $\Psi_n^{(l)}(z)$ is minimal in Theorem 4(i) or that $\Psi_n^{(r)}(z)$ is minimal in Theorem 4(ii). In such cases, the selection method (6) fails to detect and preserve jumps. On the other hand, if $\sigma^2(z)$ exhibits a jump in the continuity region D_1, all local linear estimates are consistent, but for the reason stated above, the selection method (6) can still fail to select the best (conventional) estimate. Thus the consistency is not violated, but the variance of estimates can increase and the asymptotic distribution in Theorem 6 becomes incorrect.

To deal with the discontinuity of $\sigma^2(z)$, we introduce now an alternative jump-preserving estimator which does not require the continuity of conditional error variance. Let the left-, right-, and two-sided h_n-neighborhood of z be

$$D_{zn}^{(l)} = [z - h_n, z], \quad D_{zn}^{(r)} = [z, z + h_n], \quad \text{and} \quad D_{zn}^{(c)} = [z - h_n, z + h_n],$$

respectively. To motivate an alternative to the selection method (6), we first suppose that s_q is in the right neighborhood of z, i.e., $s_q \in D_{zn}^{(r)}$. In such a case, only the left-sided local linear estimates $\hat{a}_n^{(l)}(z)$ and $\hat{b}_n^{(l)}(z)$ converge to the true parameter values $a^{(l)}(z) = a(z)$ and $b^{(l)}(z) = a'(z)$, respectively. (We are again implicitly assuming that bandwidth h_n is so small that there is at most one jump in $(z - h_n, z + h_n)$ for a sufficiently large n.) By the Taylor expansion and $\mathrm{E}[\varepsilon_i | Z_i] = \mathrm{E}[\mathrm{E}[\varepsilon_i | X_i, Z_i] | Z_i] = 0$, we have

$$
\begin{aligned}
&\mathrm{E}[Y_i - X_i^\top a^{(l)}(z) | Z_i] \\
&= \mathrm{E}\left[X_i^\top \{a(Z_i) - a^{(l)}(z)\} | Z_i\right] \\
&\leq \mathrm{E}\left[\|X_i\| | Z_i\right] \mathrm{E}\left[\|a(Z_i) - a^{(l)}(z)\| | Z_i\right] \\
&= \mathrm{O}(Z_i - z) = \mathrm{O}(h_n) = \mathrm{o}(1).
\end{aligned}
$$

for $Z_i \in D_{zn}^{(l)}$. On the other hand, the above result does not hold for the limit values of the right-sided and two-sided local linear estimators, $a^{(c)}(z)$ and $a^{(r)}(z)$, which are different from $a(z)$. Thus as long as the coefficient functions $a(\cdot)$ are identified and $a^{(\iota)}(z) \neq a(z)$, $\iota = c, r$, it holds for $Z_i \in D_{zn}^{(\iota)}$ that $\mathrm{E}[Y_i - X_i^\top a^{(\iota)}(z) | Z_i] = \mathrm{E}[X_i^\top \{a(Z_i) - a^{(\iota)}(z)\} | Z_i] \neq \mathrm{o}(1)$ in general. Analogous claims can be made if s_q is in the left neighborhood of z.

Contrary to (6), the asymptotic conditional mean independence described above is a property independent of conditional error variance $\sigma^2(z)$. To select the consistent estimator out of the three local linear estimators (3), we therefore propose to test locally whether $\mathrm{E}[\varepsilon_i^{(\iota)} | Z_i] = 0$ for $Z_i \in D_{zn}^{(\iota)}$ and $\iota = c, l, r$, where $\varepsilon_i^{(\iota)} = Y_i - X_i^\top a^{(\iota)}(z)$:[1] rejection of $\mathrm{E}[\varepsilon_i^{(\iota)} | Z_i] = 0$ indicates that a given local linear estimator is not consistent and should not be used in a given neighborhood of z. According to Bierens (1982, Theorems 1 and 2), the conditional mean independence $\mathrm{E}[\varepsilon_i^{(\iota)} | Z_i] = 0$ is equivalent to zero correlation between $\varepsilon_i^{(\iota)}$ and $\exp(k Z_i)$ for all $k \in \mathbb{R}$, or alternatively, to zero correlation between

---

[1] A similar result holds also if the local linear approximation, $\varepsilon_i^{(\iota)} = Y_i - X_i^\top \{a(z) + b(z)(Z_i - z)\}$, is used.

$\varepsilon_i^{(\iota)}$ and Z_i^k for all $k \in \mathbb{N} \cup \{0\}$. To design a simple procedure with a good power, we therefore suggest to test zero correlation between $\varepsilon_i^{(\iota)}$ and Z_i^k for $k = 1, \ldots, m$, where m is a small finite number. Given the specific form of $\mathrm{E}[\varepsilon_i^{(\iota)}|Z_i] = \mathrm{E}[\varepsilon_i + X_i^\top \{a(z) - a^{(\iota)}(z)\}|Z_i]$ caused by an unaccounted discontinuity in $a(z)$, the cubic polynomial approximates this expectation well and $m = 3$ provides a sufficient power to detect its nonlinearity even in small intervals $(z - h_n, z + h_n)$; see Section 5.

To test for non-zero correlation of $\varepsilon_i^{(\iota)}$ and Z_i^j, $j = 1, \ldots, m$, we propose to regress the estimated residual $\hat{\varepsilon}_{n,i}^{(\iota)} = Y_i - X_i^\top \hat{a}_n^{(\iota)}(z)$ on $\rho\left(\frac{Z_i - z}{h_n}\right)$ for $Z_i \in D_{zn}^{(\iota)}$, where $\rho(v) = (1, v, \cdots, v^m)^\top$. The corresponding ordinary least-squares slope estimates $\hat{\gamma}_n^{(\iota)}(z)$ will converge to $\gamma^{(\iota)} = 0$ under the null hypothesis of $\mathrm{E}[\varepsilon_i^{(\iota)}|Z_i] = 0$ and to $\gamma^{(\iota)} \neq 0$ otherwise (for a sufficiently large m); $\iota = c, l, r$. More specifically, we test significance of the slope estimates $\hat{\gamma}_n^{(\iota)}(z)$ that are the minimizers of the following least square problem:

$$\min_\gamma \sum_{i=1}^n \left\{ \hat{\varepsilon}_{n,i}^{(\iota)} - \rho^\top\left(\frac{Z_i - z}{h_n}\right)\gamma \right\}^2 \tilde{K}_h^{(\iota)}(Z_i - z), \tag{10}$$

where $\tilde{K}_h^{(\iota)}(\cdot) = h_n^{-1}\tilde{K}^{(\iota)}(\cdot/h_n)$, $\tilde{K}^{(c)}(\cdot)$ is a uniform kernel function on $[-1, 1]$,

$$\tilde{K}^{(l)}(v) = \tilde{K}^{(c)}(v) \cdot \mathbf{1}\{v \in [-1, 0)\}, \quad \text{and} \quad \tilde{K}^{(r)}(v) = \tilde{K}^{(c)}(v) \cdot \mathbf{1}\{v \in [0, 1]\}.$$

Solving the minimization (10) leads to estimate $\hat{\gamma}_n^{(\iota)}(z) = \tilde{S}_n^{(\iota)-1}(z)\tilde{T}_n^{(\iota)}(z)$, where

$$\tilde{S}_n^{(\iota)}(z) = \frac{1}{n}\sum_{i=1}^n \rho\left(\frac{Z_i - z}{h_n}\right)\rho^\top\left(\frac{Z_i - z}{h_n}\right)\tilde{K}_h^{(\iota)}(Z_i - z) \quad \text{and}$$

$$\tilde{T}_n^{(\iota)}(z) = \frac{1}{n}\sum_{i=1}^n \rho\left(\frac{Z_i - z}{h_n}\right)\tilde{K}_h^{(\iota)}(Z_i - z)\hat{\varepsilon}_{n,i}^{(\iota)}.$$

In order to test the hypothesis $\gamma^{(\iota)} = 0$, the Wald test statistics is used here, which forms an alternative measure $\tilde{\Psi}_n^{(\iota)}(z)$ to the WRMSE $\Psi_n^{(\iota)}(z)$ introduced in (5) and pro-

vides an indication about the dependence between estimated residual and Z_i:

$$\tilde{\Psi}_n^{(\iota)}(z) = \hat{\gamma}_n^{(\iota)\top}(z) \left(\frac{\tilde{S}_n^{(\iota)}(z)}{\tilde{N}_n^{(\iota)}(z)} \right) \hat{\gamma}_n^{(\iota)\top}(z), \qquad (11)$$

where

$$\hat{e}_{n,i}^{(\iota)}(z) = \tilde{\varepsilon}_{n,i}^{(\iota)} - \rho^\top \left(\frac{Z_i - z}{h_n} \right) \hat{\gamma}_n^{(\iota)}(z) \quad \text{and}$$

$$\tilde{N}_n^{(\iota)}(z) = \frac{1}{n} \sum_{i=1}^n \hat{e}_{n,i}^{(\iota)2}(z) \tilde{K}_h^{(\iota)}(Z_i - z).$$

For this quantity (11), we derive now theorems analogous to Theorems 3 and 4 for the case of the Wald measure $\tilde{\Psi}_n^{(\iota)}(z)$ under the following condition.

Assumption D.

D1. The uniform kernel $\tilde{K}^{(c)}(\cdot)$ has support $[-1, 1]$ and the kernel moment matrix $\tilde{M}^{(\iota)} = \int_{-1}^1 \rho(u)\rho^\top(u)\tilde{K}^{(\iota)}(u)\mathrm{d}u$, $\iota = c, l, r$, is positive definite.

D2. The number m of powers used in the auxiliary regressions (10) is sufficiently large such that at least one of the slope coefficients $\gamma^{(\iota)}$, which has its explicit expression given in equation (68), is non-zero if $\mathrm{E}[\varepsilon_i^{(\iota)}|Z_i] \neq 0$ for $Z_i \in D_z^{(\iota)}$.

Theorem 7. *Suppose that Assumptions A1–A5, A6', B, and D hold. At any $z \in D_{1n}^{(\iota)}$ for some $n \in \mathbb{N}$ and $\iota = c, l, r$, it holds for $n \to \infty$ that $\tilde{\Psi}_n^{(\iota)}(z) = o_p(1)$.*

Theorem 8. *If Assumptions A1–A5, A6', B, and D are satisfied, the following results hold as $n \to \infty$.*

(i) *For any $z = s_q + \tau h_n \in D$ with $q = 1, \ldots, Q + 1$ and $\tau \in (-1, 0)$,*

$$\tilde{\Psi}_n^{(c)}(z) = \gamma^{(c)\top} \tilde{C}_\tau^{(c)} \gamma^{(c)} + \mathrm{o_p}(1),$$

$$\tilde{\Psi}_n^{(l)}(z) = \mathrm{o_p}(1),$$

$$\tilde{\Psi}_n^{(r)}(z) = \gamma^{(r)\top} \tilde{C}_\tau^{(r)} \gamma^{(r)} + \mathrm{o_p}(1).$$

(ii) *For any $z = s_q + \tau h_n \in D$ with $q = 0, \ldots, Q$ and $\tau \in (0, 1)$,*

$$\tilde{\Psi}_n^{(c)}(z) = \gamma^{(c)\top} \tilde{C}_\tau^{(c)} \gamma^{(c)} + \mathrm{o_p}(1),$$

$$\tilde{\Psi}_n^{(l)}(z) = \gamma^{(l)\top} \tilde{C}_\tau^{(l)} \gamma^{(l)} + \mathrm{o_p}(1),$$

$$\tilde{\Psi}_n^{(r)}(z) = \mathrm{o_p}(1).$$

In both cases, $\tilde{C}_\tau^{(\iota)}$, $\iota = c, l, r$, is a positive definite matrix defined in Appendix A, equation (71), and the explicit form of $\gamma^{(\iota)}$, $\iota = c, l, r$, is given in Appendix A, (68).

Given the above results, we can use the Wald statistics $\tilde{\Psi}_n^{(\iota)}(z)$ to again distinguish which local estimators $\hat{a}_n^{(\iota)}(z)$ are consistent or inconsistent due to a discontinuity of coefficient functions, but now without requiring that the conditional variance $\sigma^2(z)$ is continuous. We thus propose a new jump-preserving estimator $\tilde{a}_n(z)$ of coefficient functions $a(z)$ when the conditional error variance contains a finite set of discontinuities:

$$\tilde{a}_n(z) = \begin{cases} \hat{a}_n^{(c)}(z), & \text{if } \tilde{\mathrm{diff}}(z) \leq u_n, \\[2mm] \hat{a}_n^{(l)}(z), & \text{if } \tilde{\mathrm{diff}}(z) > u_n \text{ and } \tilde{\Psi}_n^{(r)}(z) > \tilde{\Psi}_n^{(l)}(z), \\[2mm] \hat{a}_n^{(r)}(z), & \text{if } \tilde{\mathrm{diff}}(z) > u_n \text{ and } \tilde{\Psi}_n^{(l)}(z) > \tilde{\Psi}_n^{(r)}(z), \\[2mm] \dfrac{\hat{a}_n^{(l)}(z) + \hat{a}_n^{(r)}(z)}{2}, & \text{if } \tilde{\mathrm{diff}}(z) > u_n \text{ and } \tilde{\Psi}_n^{(l)}(z) = \tilde{\Psi}_n^{(r)}(z), \end{cases} \tag{12}$$

where the auxiliary parameter $u_n > 0$ is again tending to zero with increasing n and

19

$\widetilde{\text{diff}}(z) = \tilde{\Psi}_n^{(c)}(z) - \min\{\tilde{\Psi}_n^{(l)}(z), \tilde{\Psi}_n^{(r)}(z)\}$. The consistency and asymptotic normality of the proposed jump-preserving estimator $\tilde{a}_n(z)$ are established in the following theorems.

Theorem 9. *Under Assumptions A1–A5, A6', B, C, and D, it holds for $n \to \infty$ and some $\delta \in (0, 1/2)$ that*

(i)
$$\sup_{z \in D_{1n}} \|\tilde{a}_n(z) - a(z)\| = O_p\left(\sqrt{\frac{\ln n}{nh_n}}\right),$$

(ii)
$$\sup_{z \in D_{2n,\delta}} \|\tilde{a}_n(z) - a(z)\| = O_p\left(\sqrt{\frac{\ln n}{nh_n}}\right), \quad and$$

(iii) *for any given $z \in D_{2n}$,*

$$\tilde{a}_n(z) = a(z) + O_p\left(\sqrt{\frac{\ln n}{nh_n}}\right).$$

Theorem 10. *If Assumptions A1–A5, A6', B, C, and D are satisfied and a point $z \in D \setminus \{s_0, \ldots, s_{Q+1}\}$, it holds that*

$$\sqrt{nh_n}\left[\tilde{a}_n(z) - a(z) - \frac{h_n^2}{2}\left(c_0^{(\iota)}\mu_2^{(\iota)} + c_1^{(\iota)}\mu_3^{(\iota)}\right)a''(z)\right] \xrightarrow{d} N\left(0, \Phi^{(\iota)}(z)\right)$$

as $n \to \infty$, where $\Phi^{(\iota)}(z)$ is defined in (8) and

$$\iota = \begin{cases} c, & \text{if } z \in D_{1n}, \\ l, & \text{if } z \in D \cap \bigcup_{q=0}^{Q+1} [s_q - h_n, \ s_q), \\ r, & \text{if } z \in D \cap \bigcup_{q=0}^{Q+1} (s_q, \ s_q + h_n]. \end{cases}$$

5 Simulations

In this section, we first discuss the selection procedure of the smoothing parameters h_n and u_n. Next, we examine the finite sample properties of the jump-preserving estimators $\breve{a}_n(\cdot)$ defined in (6) and $\tilde{a}_n(\cdot)$ given in (12) using two simulated examples.

Among many bandwidth selection procedures for nonparametric models, we opt for the cross-validation method similarly to Zhao et al. (2016). When covariates X_i and Z_i do not any contain lagged dependent variables, we select the smoothing parameters by the leave-one-out cross-validation. The selected smoothing parameters $\hat{h}_n$ and $\hat{u}_n$ are thus determined by

$$(\hat{h}_n, \hat{u}_n) = \arg \min_{h_n, u_n} \sum_{i=1}^{n} \left[Y_i - X_i^\top \mathring{a}_{n,-i}(Z_i) \right]^2,$$

where $\mathring{a}_{n,-i}(Z_i)$ represents a jump-preserving estimate $\breve{a}_n(\cdot)$ or $\tilde{a}_n(\cdot)$ based on all data except for the ith observation (Y_i, X_i, Z_i). If covariates X_i and Z_i do contain some lagged dependent variables with the lags up to order m, we suggest to apply the m-block-out cross-validation technique:

$$(\hat{h}_n, \hat{u}_n) = \arg \min_{h_n, u_n} \sum_{i=1}^{n} \left[Y_i - X_i^\top \mathring{a}_{n,-m_i}(Z_i) \right]^2,$$

where $\mathring{a}_{n,-m_i}(Z_i)$ is computed without using observations $\{Y_{i+j}, X_{i+j}, Z_{i+j}\}_{j=-m}^{m}$ (see Patton et al., 2009, for the data-dependent block-size selection).

To observe the estimation precision both in neighborhoods of change points and overall, we evaluate the performance of the proposed estimators via the global mean absolute deviation of errors (MADE) and local mean absolute deviation of errors ($\text{MADE}_{\text{local}}$):

$$\text{MADE} = \frac{1}{n_{\text{grid}}} \sum_{j=1}^{n_{\text{grid}}} \| \mathring{a}_n(z_j) - a(z_j) \|_1$$

and

$$\text{MADE}_{\text{local}} = \frac{1}{n_{\text{grid}}} \sum_{q=1}^{Q} \sum_{j=1}^{n_{\text{grid}}} \|\mathring{a}_n(z_j) - a(z_j)\|_1 \cdot \mathbf{1}\{z_j \in (s_q - 0.1, s_q + 0.1)\},$$

where $\mathring{a}_n(z_j)$ represents one of the considered estimators, $\{z_j\}_{j=1}^{n_{\text{grid}}}$ are the grid points, and $\| \cdot \|_1$ denotes the absolute value norm.

5.1 Experiment 1: Constant conditional variance function

First, we consider an AR(1) process:[2]

$$X_t = a_0(Z_t) + a_1(Z_t)X_{t-1} + \sigma(Z_t)\varepsilon_t, \quad t = 1, \ldots, n, \tag{13}$$

where the variable Z_t is drawn independently from the uniform distribution, $Z_t \sim U(0,1)$, the errors are independent standard normal, $\varepsilon_t \sim N(0,1)$, and the coefficient functions

$$a_0(Z_t) = 1.2\cos(Z_t) - 1.68 \cdot \mathbf{1}\{Z_t < 0.5\} - 0.66 \cdot \mathbf{1}\{Z_t \geq 0.5\} \quad \text{and}$$

$$a_1(Z_t) = \cos(Z_t) - \mathbf{1}\{Z_t < 0.5\} - 0.25 \cdot \mathbf{1}\{Z_t \geq 0.5\}.$$

In this first simulation experiment, the variance function is constant: $\sigma^2(Z_t) = 0.6^2$. The process (13) is evaluated at two sample sizes $n = 300$ and $n = 600$, and for each sample size, 1000 samples are simulated. We estimate the coefficient functions using local linear fitting on an equispaced grid of points $\{z_j\}_{j=1}^{n_{\text{grid}}}$ with $z_1 = 0$, $z_{n_{\text{grid}}} = 1$, and $n_{\text{grid}} = 200$. All nonparametric estimators employ the Epanechnikov kernel: $K^{(c)}(v) = 0.75(1 - v^2)\mathbf{1}\{|v| \leq 1\}$.

First, the bandwidth h_n is set to $0.54n^{-1/5}$ for all three local estimators, and u_n is se-

---

[2]We have also studied the same AR(1) process (13) with coefficients that are functions of time t/n. Though using a linear time trend t/n as Z_t might violate Assumption A1, the simulation results are similar to the case with a uniformly distributed Z_t.
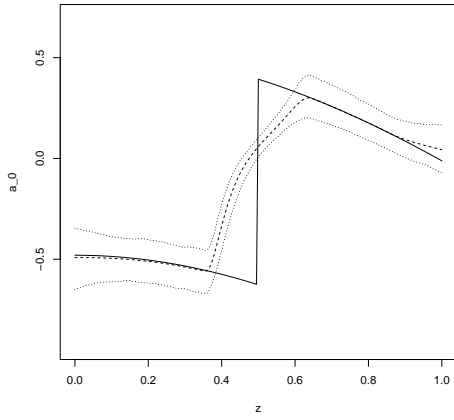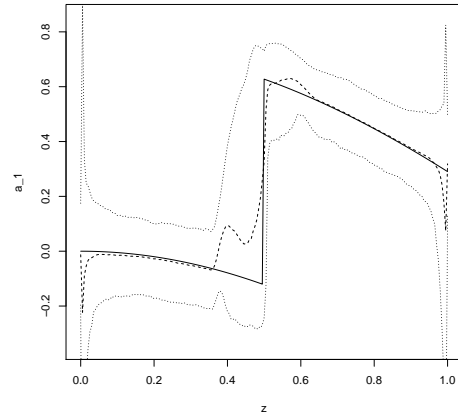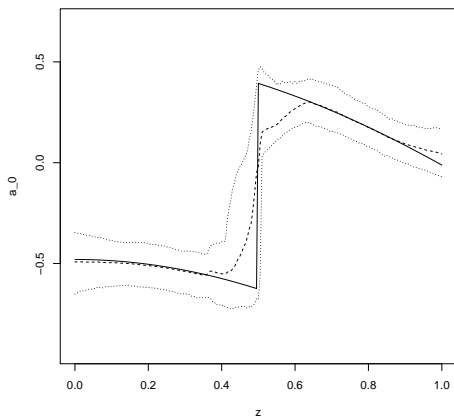
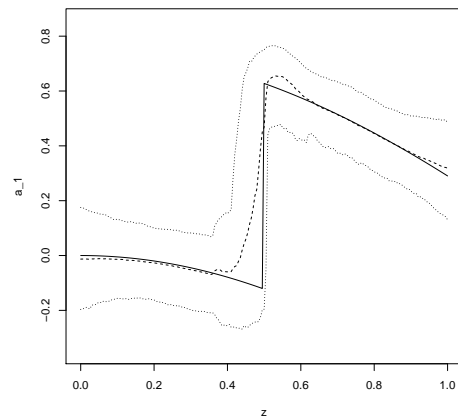(a) Conventional $\hat{a}_{n,0}^{(c)}$ (b) Conventional $\hat{a}_{n,1}^{(c)}$
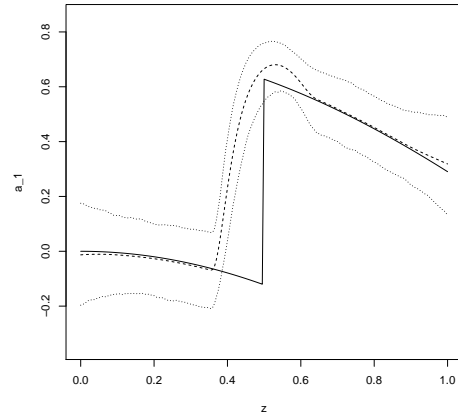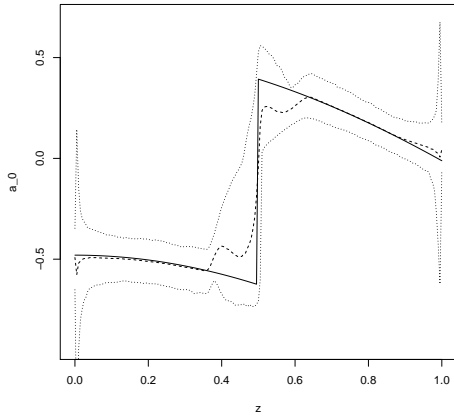
(c) Jump-preserving $\breve{a}_{n,0}$ (WRMSE)   (d) Jump-preserving $\breve{a}_{n,1}$ (WRMSE)
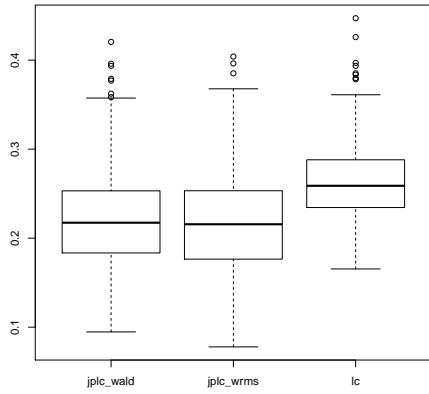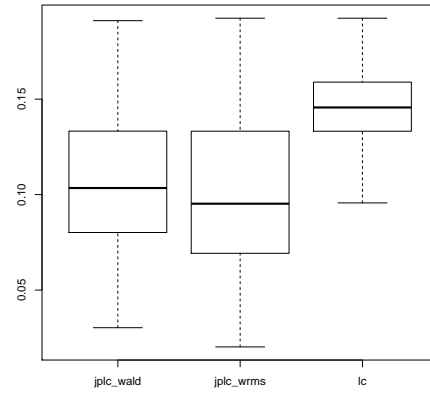
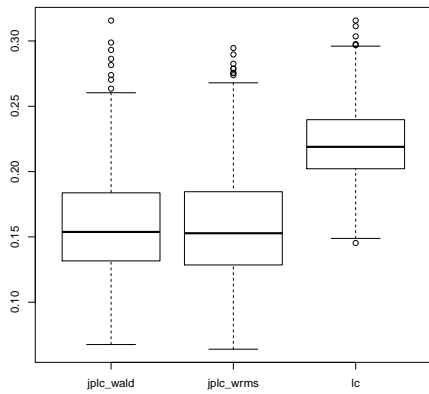(e) Jump-preserving $\tilde{a}_{n,0}$ (Wald)   (f) Jump-preserving $\tilde{a}_{n,1}$ (Wald)

Figure 1: Homoscedastic model with the fixed bandwidth and $n = 600$: the solid lines represent the true coefficient functions, the dashed lines are the average varying coefficient estimates, and the dotted lines are the 95% confidence bands.
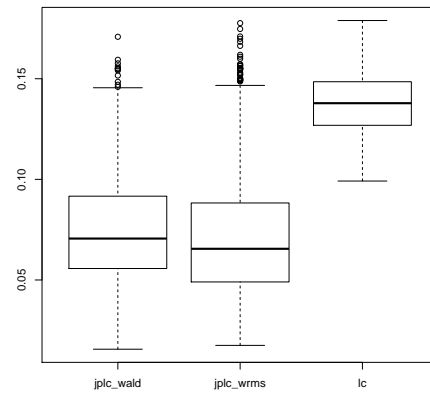
(a) MADE for $n = 300$ (b) $\text{MADE}_{\text{local}}$ for $n = 300$

(c) MADE for $n = 600$ (d) $\text{MADE}_{\text{local}}$ for $n = 600$

Figure 2: Homoscedastic model with the fixed bandwidth: global and local mean absolute deviations of the estimates. Each plot contains boxplots for (from left to right) the jump-preserving estimator based on the Wald statistics, the jump-preserving estimator based on WRMSE, and the conventional estimator.

24

lected by cross-validation. Figure 1 provides a graphical presentation of the performance of the two jump-preserving local linear estimators $\check{a}_n(z)$ (selection using WRMSE) and $\tilde{a}_n(z)$ (selection using the Wald statistics) and the conventional local linear estimator $\hat{a}_n^{(c)}(z)$ for $n = 600$. Both jump-preserving estimators track the true coefficient functions closely, while the conventional local linear estimator is inconsistent around the discontinuity $z = 0.5$ as the confidence intervals of $\hat{a}_n^{(c)}(z)$ do not contain the discontinuity. In addition, $\check{a}_n(z)$ compared to $\tilde{a}_n(z)$ has a wider confidence interval near the boundaries. The procedure of selecting the left-sided, right-sided, or conventional local estimators proposed for $\tilde{a}_n(z)$ in Section 4 still chooses $\hat{a}_n^{(c)}(z)$ around the boundary points and is thus less affected by the boundaries than $\check{a}_n(z)$.

Due to strong boundary effects in $\check{a}_n(z)$, the 1000 global and local MADE values for each sample size are computed for $z \in [0.05, 0.95]$. The boxplots are shown in Figure 2. The conventional local linear estimator has higher global and local MADE values compared to the jump-preserving estimators $\check{a}_n(z)$ and $\tilde{a}_n(z)$, where there is no significant difference in the MADEs of $\check{a}_n(z)$ and $\tilde{a}_n(z)$. Both jump-preserving estimators thus perform well in the case of the process with homoscedastic error. When the sample size becomes larger, all global and local MADEs decrease proportionally for all estimators.

Next, we repeat the experiment, but cross-validate both h_n and u_n for each replication; the results are shown in Figures 3 and 4. The interpretation of the results is similar as above. The main difference is that the MADE of the conventional local linear estimator is smaller than before since the bandwidth selected for $\hat{a}_n^{(c)}(z)$ is freely chosen and thus becomes smaller in an attempt to capture the discontinuity as good as possible, while decreasing the precision in the continuity region. Nevertheless, the discontinuity is not included in its confidence interval and its performance is still worse than that of the proposed jump-preserving estimators.
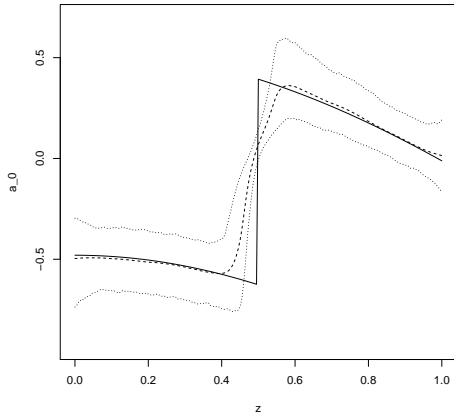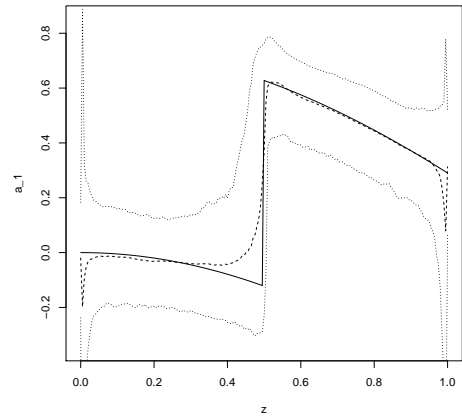
(a) Conventional $\hat{a}_{n,0}$ (b) Conventional $\hat{a}_{n,1}$
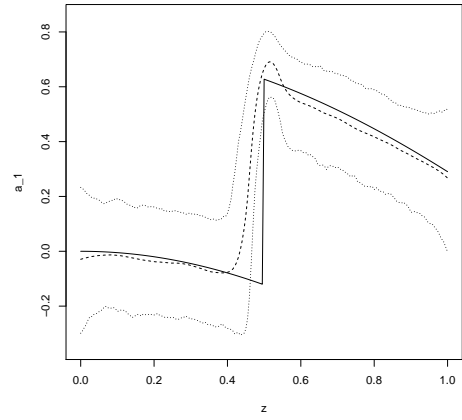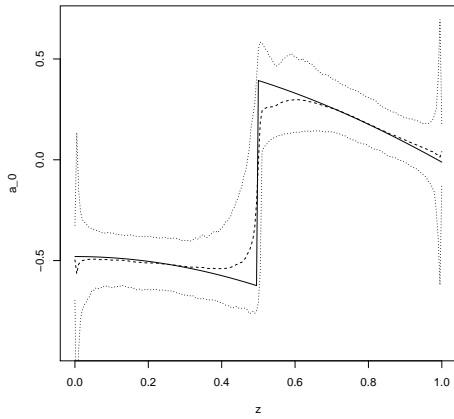
(c) Jump-preserving $\breve{a}_{n,0}$ (WRMSE)   (d) Jump-preserving $\breve{a}_{n,1}$ (WRMSE)
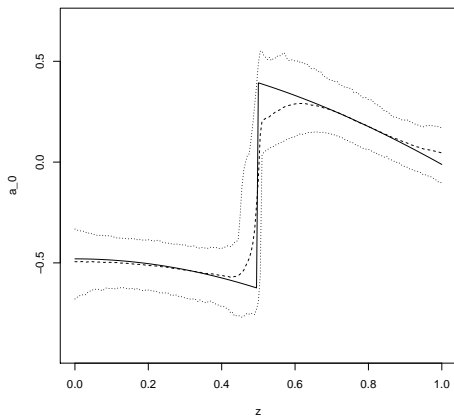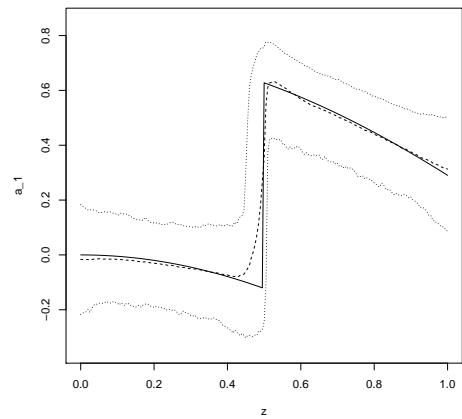
(e) Jump-preserving $\tilde{a}_{n,0}$ (Wald)   (f) Jump-preserving $\tilde{a}_{n,1}$ (Wald)

Figure 3: Homoscedastic model with the cross-validated bandwidth and $n = 600$: the solid lines represent the true coefficient functions, the dashed lines are the average varying coefficient estimates, and the dotted lines are the 95% confidence bands.
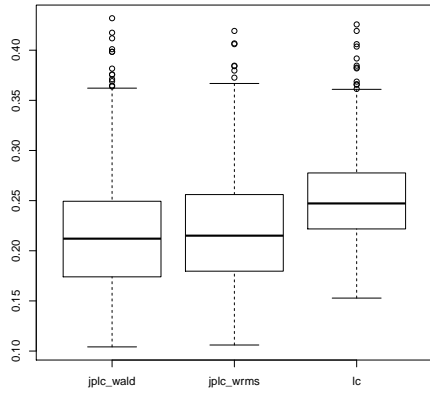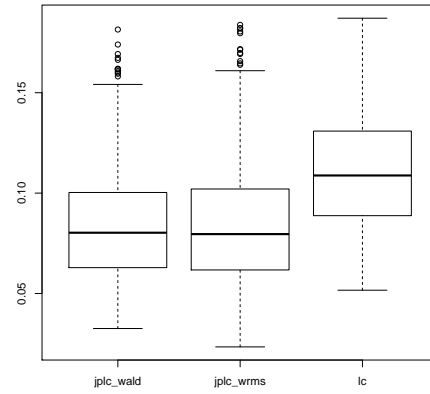
(a) MADE for $n = 300$

(b) $\text{MADE}_{\text{local}}$ for $n = 300$
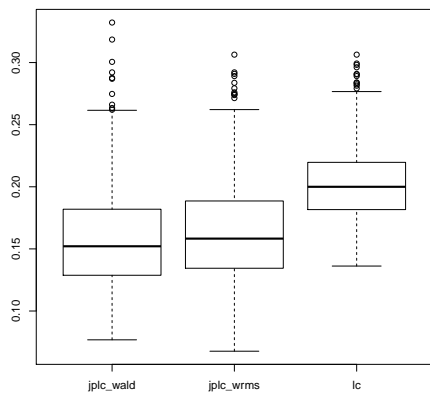
(c) MADE for $n = 600$

(d) $\text{MADE}_{\text{local}}$ for $n = 600$

Figure 4: Homoscedastic model with the cross-validated bandwidth: global and local mean absolute deviations of the estimates. Each plot contains boxplots for (from left to right) the jump-preserving estimator based on the Wald statistics, the jump-preserving estimator based on WRMSE, and the conventional estimator.

5.2 Experiment 2: discontinuous conditional variance function

Now we consider the same time-varying AR(1) process as in (13), but with a discontinuous conditional variance function:

$$\sigma^2(t) = \left(0.8 \cdot \mathbf{1}\{t < 0.5\} + 0.6 \cdot \mathbf{1}\{t \geq 0.5\}\right)^2. \tag{14}$$

The evaluation is performed in the same way as in the previous section.

Figure 5 provides a graphical presentation of the performance of the convetional estimator $\hat{a}_n^{(c)}(z)$, jump-preserving estimator $\check{a}_n(z)$ based on WRMSE, and jump-preserving estimator $\tilde{a}_n(z)$ based on the Wald statistics with a fixed bandwidth $h_n = 0.54n^{-1/5}$, whereas the results using the cross-validated bandwidth h_u and u_n are presented in Figure 7. In this case, only the proposed jump-preserving estimators $\tilde{a}_n(z)$ based on the Wald statistics preserve the discontinuity, whereas $\hat{a}_n^{(c)}(z)$ and $\check{a}_n(z)$ are both inconsistent as their confidence intervals do not contain the discontinuity for z's near the jump point; note that this is true even for the jump-preserving method based on WRMSE. The corresponding boxplots with MADE are shown in Figures 6 and 8. The proposed estimator $\tilde{a}_n(z)$ based on the Wald statistics has the lowest global and local MADE values compared to the other jump-preserving estimator $\hat{a}_n(z)$ and to the conventional local linear estimator $\check{a}_n(z)$. The differences become a bit smaller when we cross-validate both the bandwidths h_n and u_n (see Figure 8). In both cases, the jump-preserving estimator $\tilde{a}_n(\cdot)$ in (12) outperforms the existing method $\hat{a}_n(\cdot)$ in (6) in the presence of the discontinuity of conditional variance function, in particular in terms of MADE.

(a) Conventional $\hat{a}_{n,0}$
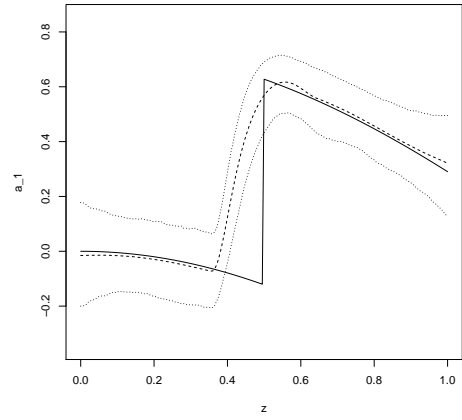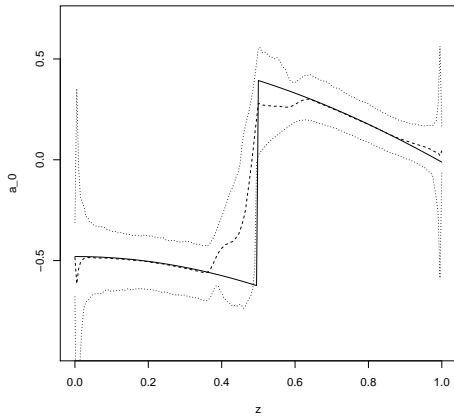
(b) Conventional $\hat{a}_{n,1}$

(c) Jump-preserving $\breve{a}_{n,0}$ (WRMSE)
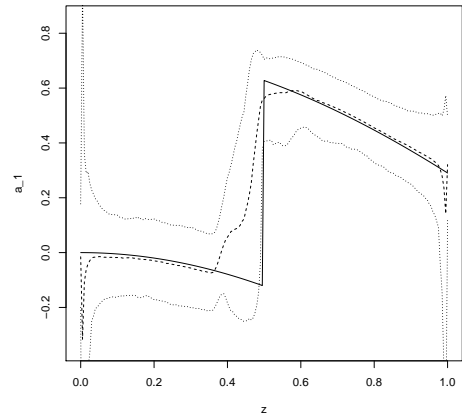
(d) Jump-preserving $\breve{a}_{n,1}$ (WRMSE)
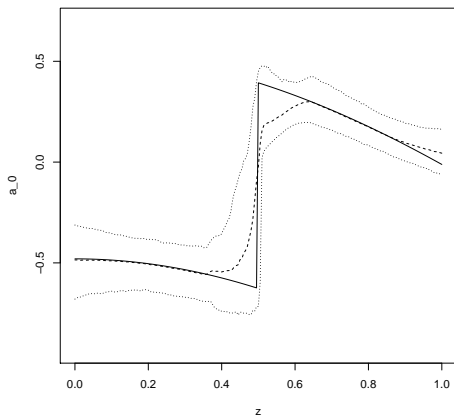
(e) Jump-preserving $\tilde{a}_{n,0}$ (Wald)
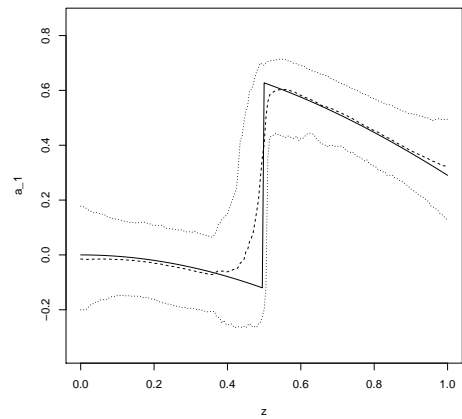
(f) Jump-preserving $\tilde{a}_{n,1}$ (Wald)

Figure 5: Heteroscedastic model with the fixed bandwidth: the solid lines represent the true coefficient functions, the dashed lines are the average varying coefficient estimates, and the dotted lines are the 95% confidence bands.
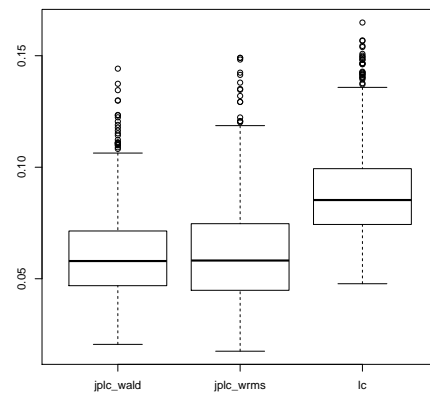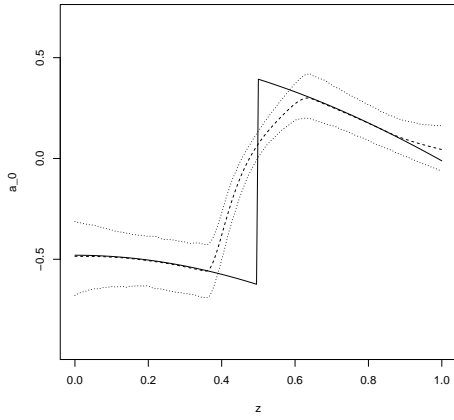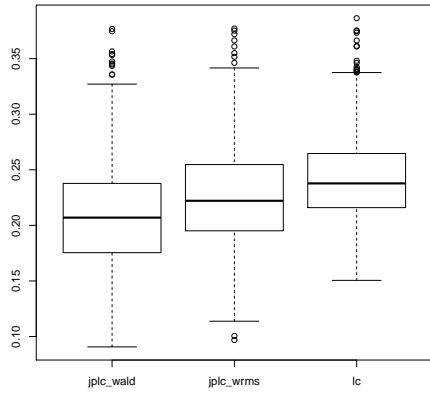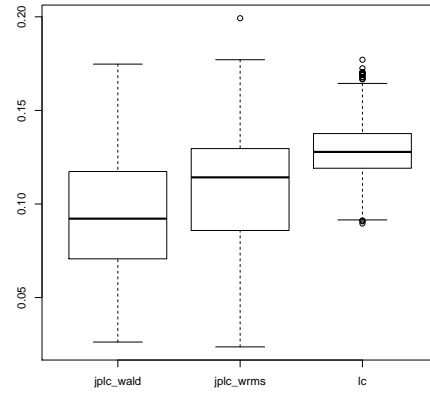
(a) MADE for $n = 300$

(b) MADE$_{local}$ for $n = 300$
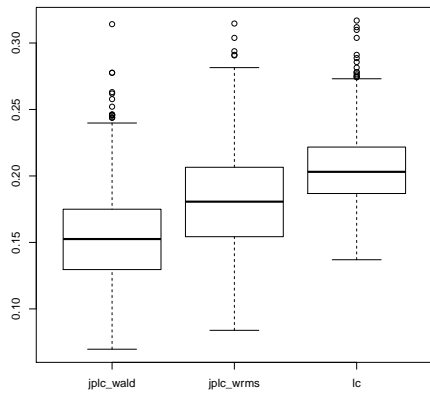
(c) MADE for $n = 600$
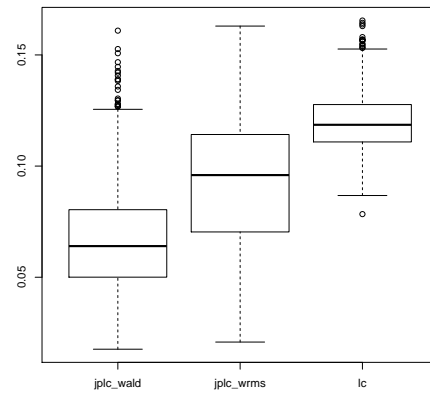
(d) MADE$_{local}$ for $n = 600$

Figure 6: Heteroscedastic model with the fixed bandwidth: global and local mean absolute deviations of the estimates. Each plot contains boxplots for (from left to right) the jump-preserving estimator based on the Wald statistics, the jump-preserving estimator based on WRMSE, and the conventional estimator.
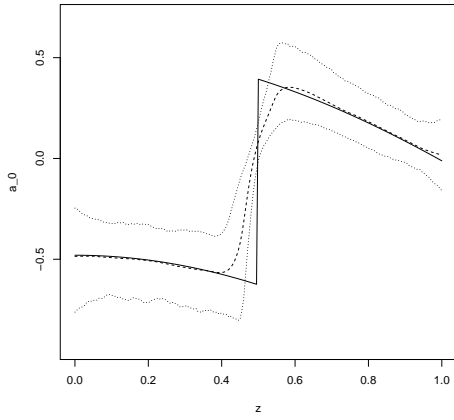
(a) Conventional $\hat{a}_{n,0}$ (b) Conventional $\hat{a}_{n,1}$
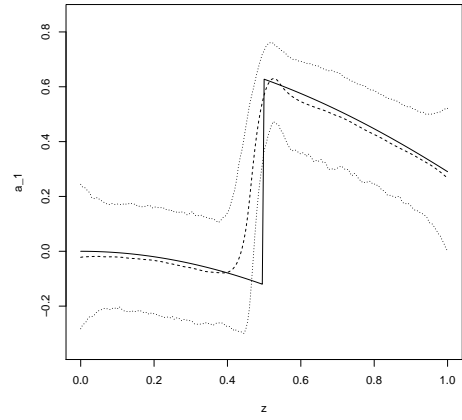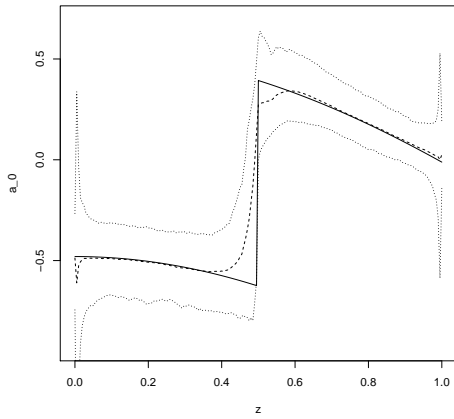
(c) Jump-preserving $\breve{a}_{n,0}$ (WRMSE)   (d) Jump-preserving $\breve{a}_{n,1}$ (WRMSE)
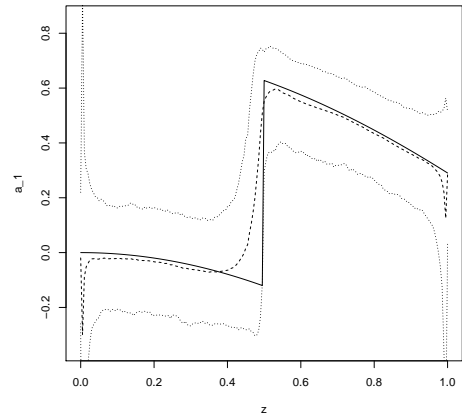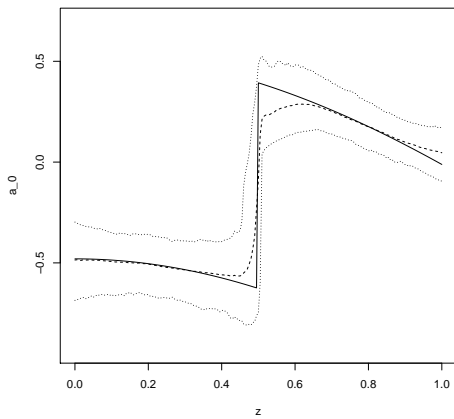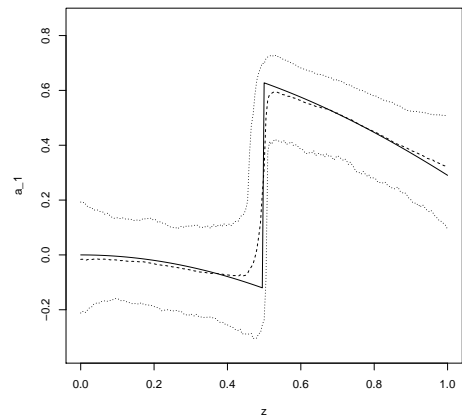
(e) Jump-preserving $\tilde{a}_{n,0}$ (Wald)   (f) Jump-preserving $\tilde{a}_{n,1}$ (Wald)

Figure 7: Heteroscedastic model with the cross-validated bandwidth: the solid lines represent the true coefficient functions, the dashed lines are the average varying coefficient estimates, and the dotted lines are the 95% confidence bands.
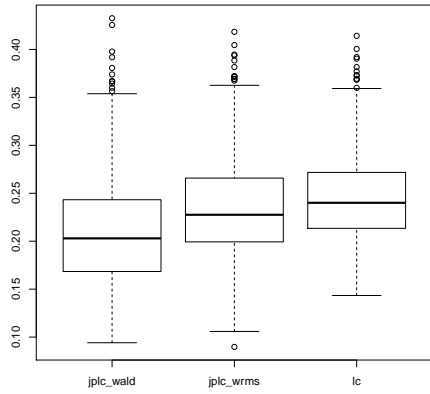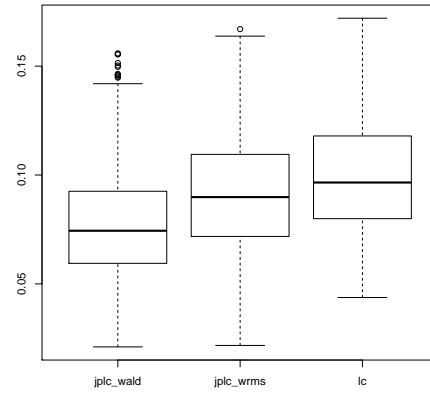
(a) MADE for $n = 300$

(b) MADE$_{\text{local}}$ for $n = 300$
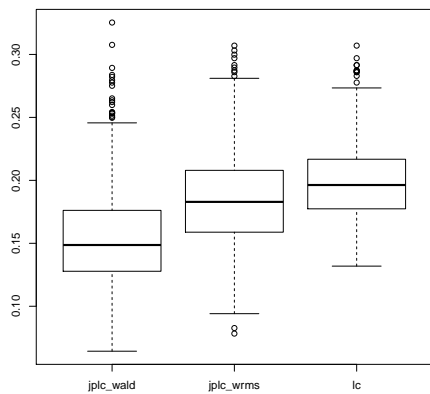
(c) MADE for $n = 600$
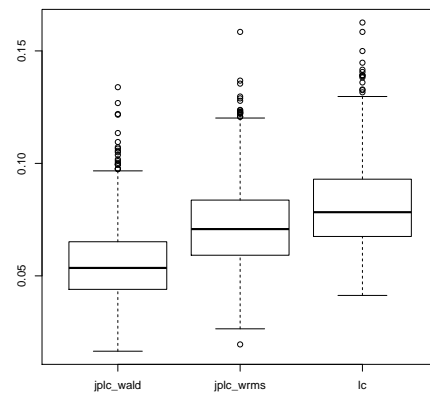
(d) MADE$_{\text{local}}$ for $n = 600$

Figure 8: Heteroscedastic model with the cross-validated bandwidth: global and local mean absolute deviations of the estimates. Each plot contains boxplots for (from left to right) the jump-preserving estimator based on the Wald statistics, the jump-preserving estimator based on WRMSE, and the conventional estimator.

6 Conclusions

In this paper, we propose estimators for varying-coefficient models with discontinuous coefficient functions. First, we adapt the local linear estimators of Gijbels et al. (2007) and Zhao et al. (2016), which select among the left-sided, right-sided, and conventional local linear estimators by comparing their weighted residual mean squared errors, to the time series setting. This approach works well when there are no discontinuities in the conditional error variance. To cope with the discontinuity problem in the conditional error variance, we propose a different "correctness" measure of the three local linear fits based on the Wald statistics. In all cases, the asymptotic properties including the uniform consistency and asymptotic normality are derived for both proposed estimators and their performance is tested with simulated examples.

A Proofs of the main results

In this section, we prove the theorems presented in Section 3 and 4. Auxiliary lemmas are collected in Appendix B. Throughout Appendices A and B, we let C be a generic positive constant, which may take different values at different places, and write $M \succ 0$ if matrix M is positive definite. All limiting expressions including $o_p(\cdot)$ and $O_p(\cdot)$ are taken for $n \to \infty$, unless stated otherwise. $\{\xi_i; a \leq i \leq b\}$. The α-mixing coefficient of the process $\{\xi_i\}_{i=-\infty}^{\infty}$ is defined as The dependence on z of the variables introduced in Appendices A and B is kept implicit in order to shorten the length of proofs.

First, we introduce some notation. Denote

$$S_n^{(\iota)} = \begin{pmatrix} S_{n,0}^{(\iota)} & S_{n,1}^{(\iota)} \\ S_{n,1}^{(\iota)} & S_{n,2}^{(\iota)} \end{pmatrix}, \quad T_n^{(\iota)} = \begin{pmatrix} T_{n,0}^{(\iota)} \\ T_{n,1}^{(\iota)} \end{pmatrix}, \quad \text{and} \quad F_n^{(\iota)} = \begin{pmatrix} F_{n,0}^{(\iota)} \\ F_{n,1}^{(\iota)} \end{pmatrix},$$

where

$$S_{n,j}^{(\iota)} = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^{\top} \left(\frac{Z_i - z}{h_n} \right)^j K_h^{(\iota)}(Z_i - z), \quad j = 0, 1, 2, 3, \tag{15}$$

$$T_{n,j}^{(\iota)} = \frac{1}{n} \sum_{i=1}^{n} X_i \left(\frac{Z_i - z}{h_n} \right)^j K_h^{(\iota)}(Z_i - z) Y_i, \quad j = 0, 1, \quad \text{and} \tag{16}$$

$$F_{n,j}^{(\iota)} = \frac{1}{n} \sum_{i=1}^{n} X_i \left(\frac{Z_i - z}{h_n} \right)^j K_h^{(\iota)}(Z_i - z) \varepsilon_i, \quad j = 0, 1. \tag{17}$$

Using the above notation, the local linear estimators of $a(\cdot)$ and $a'(\cdot)$ in (4) can be written as

$$
\begin{aligned}
\hat{\beta}_n^{(\iota)} &= \begin{pmatrix} \hat{a}_n^{(\iota)}(z) \\ \hat{b}_n^{(\iota)}(z) \end{pmatrix} \\
&= H_n^{-1} \left[\sum_{i=1}^{n} H_n^{-1} \begin{pmatrix} X_i \\ X_i(Z_i - z) \end{pmatrix} \begin{pmatrix} X_i \\ X_i(Z_i - z) \end{pmatrix}^{\top} H_n^{-1} K_h^{(\iota)}(Z_i - z) \right]^{-1} \\
&\quad \sum_{i=1}^{n} H_n^{-1} \begin{pmatrix} X_i \\ X_i(Z_i - z) \end{pmatrix} Y_i K_h^{(\iota)}(Z_i - z) \\
&= H_n^{-1} S_n^{(\iota)-1} T_n^{(\iota)}, \tag{18}
\end{aligned}
$$

where H_n is a $2p \times 2p$ diagonal matrix with its first p diagonal elements equal to 1's and its last p elements equal to h_n's.

Since the coefficient functions $a(z)$ are twice continuously differentiable except for the discontinuities $\{s_q\}_{q=0}^{Q+1}$ (Assumption A5), it follows from the Taylor expansion for $Z_i \in D_{zn}^{(\iota)}$ that

$$a(Z_i) = a(z) + h_n \left(\frac{Z_i - z}{h_n} \right) a'(z) + \frac{h_n^2}{2} \left(\frac{Z_i - z}{h_n} \right)^2 a''(z) + \mathrm{o}(Z_i - z)^2 \tag{19}$$

uniformly in $z \in D_{1n}^{(\iota)}$, which implies

$$T_{n,0}^{(\iota)} - F_{n,0}^{(\iota)} = \frac{1}{n} \sum_{i=1}^{n} K_h^{(\iota)}(Z_i - z) X_i X_i^\top a(Z_i)$$

$$= S_{n,0}^{(\iota)} a(z) + h_n S_{n,1}^{(\iota)} a'(z) + \frac{h_n^2}{2} S_{n,2}^{(\iota)} a''(z) + S_{n,0}^{(\iota)} \cdot o_p(h_n^2)$$

and

$$T_{n,1}^{(\iota)} - F_{n,1}^{(\iota)} = S_{n,1}^{(\iota)} a(z) + h_n S_{n,2}^{(\iota)} a'(z) + \frac{h_n^2}{2} S_{n,3}^{(\iota)} a''(z) + S_{n,1}^{(\iota)} \cdot o_p(h_n^2).$$

Consequently for $\beta = [a^\top(z), a'^\top(z)]^\top$, it holds that

$$T_n^{(\iota)} - F_n^{(\iota)} = S_n^{(\iota)} H_n \beta + \frac{h_n^2}{2} \begin{pmatrix} S_{n,2}^{(\iota)} \\ S_{n,3}^{(\iota)} \end{pmatrix} a''(z) + \begin{pmatrix} S_{n,0}^{(\iota)} \\ S_{n,1}^{(\iota)} \end{pmatrix} \cdot o_p(h_n^2). \tag{20}$$

Using (18), (20), and Lemma 8(ii), we finally obtain

$$H_n(\hat{\beta}_n^{(\iota)} - \beta) = S_n^{(\iota)-1} T_n^{(\iota)} - H_n \beta$$

$$= S_n^{(\iota)-1} F_n^{(\iota)} + \frac{h_n^2}{2} S_n^{(\iota)-1} \begin{pmatrix} S_{n,2}^{(\iota)} \\ S_{n,3}^{(\iota)} \end{pmatrix} a''(z) + o_p(h_n^2) \tag{21}$$

uniformly in $z \in D_{1n}^{(\iota)}$.

Proof of Theorem 1.

According to Lemma 3, the terms $S_{n,j}^{(\iota)}$, $S_n^{(\iota)-1}$, and $F_{n,j}^{(\iota)}$ uniformly converge on $D_{1n}^{(\iota)}$ to their corresponding expected values at rates $(nh_n/\ln n)^{-1/2} + h_n$ and $(nh_n/\ln n)^{-1/2}$,

respectively. It follows from (21) and Assumptions A2, A3(ii), and A4 that

$$
\sup_{z\in D_{1n}^{(\iota)}} \left\| H_n(\hat{\beta}_n^{(\iota)} - \beta) \right\| \leq \sup_{z\in D_{1n}^{(\iota)}} \left\| S_n^{(\iota)-1} \right\| \left\{ \sup_{z\in D_{1n}^{(\iota)}} \left\| F_n^{(\iota)} \right\| + \sup_{z\in D_{1n}^{(\iota)}} \left\| \frac{h_n^2}{2} \begin{pmatrix} S_{n,2}^{(\iota)} \\ S_{n,3}^{(\iota)} \end{pmatrix} \right\| \right\}
$$

$$
\cdot \max_{z\in D_{1n}^{(\iota)}} \left\| a''(z) \right\| + o_p(h_n^2)
$$

$$
\leq C_1 \cdot \frac{\sup_{z\in D_{1n}^{(\iota)}} \left\| \Omega^{-1}(z) \right\|}{\inf_{z\in D} f_Z(z)} \left\{ 1 + O_p\left(\sqrt{\frac{\ln n}{nh_n}} + h_n \right) \right\}
$$

$$
\cdot \left[O_p\left(\sqrt{\frac{\ln n}{nh_n}} \right) + C_2 h_n^2 \left\{ \sup_{z\in D_{1n}^{(\iota)}} \left\| f_Z(z)\Omega(z) \right\| + O_p\left(\sqrt{\frac{\ln n}{nh_n}} + h_n \right) \right\} \right]
$$

$$
+ o_p(h_n^2)
$$

$$
\leq C_3 \cdot \left\{ 1 + O_p\left(\sqrt{\frac{\ln n}{nh_n}} + h_n \right) \right\} \cdot O_p\left(\sqrt{\frac{\ln n}{nh_n}} + h_n^2 + h_n^3 \right) + o_p(h_n^2)
$$

$$
= O_p\left(\sqrt{\frac{\ln n}{nh_n}} \right) + O_p(h_n^2), \quad \iota = c, l, r,
$$

where C_1, C_2, and C_3 represent some positive constants and $\Omega(z) = \mathrm{E}[XX^\top | Z = z]$. As a result, we have

$$
\sup_{z\in D_{1n}^{(\iota)}} \left\| \hat{a}_n^{(\iota)}(z) - a(z) \right\| = O_p\left(\sqrt{\frac{\ln n}{nh_n}} \right) + O_p(h_n^2), \quad \iota = c, l, r,
$$

and

$$
\sup_{z\in D_{1n}^{(\iota)}} \left\| \hat{b}_n^{(\iota)}(z) - a'(z) \right\| = O_p\left(h_n^{-1} \sqrt{\frac{\ln n}{nh_n}} \right) + O_p(h_n), \quad \iota = c, l, r.
$$

The claim follows by noting that $h_n^2 = o(\sqrt{\ln n/(nh_n)})$ by Assumption B3. $\qquad\square$

Proof of Theorem 2.

By the weak convergence results for $S_{n,j}^{(\iota)}$ and $S_n^{(\iota)-1}$ in Lemmas 1(i) and 1(ii) and equation (21),

$$\hat{a}_n^{(\iota)}(z) - a(z) = \left[\frac{\Omega^{-1}(z)}{f_Z(z)} \left(c_0^{(\iota)} F_{n,0}^{(\iota)} + c_1^{(\iota)} F_{n,1}^{(\iota)} \right) + \frac{h_n^2}{2} \left(c_0^{(\iota)} \mu_2^{(\iota)} + c_1^{(\iota)} \mu_3^{(\iota)} \right) a''(z) \right] \quad (22)$$
$$\cdot (1 + o_p(1)) + o_p(h_n^2),$$

where $c_j^{(\iota)}$ and $\mu_j^{(\iota)}$ are defined in (7). The stochastic term in (22) can be analyzed in the following way. Let

$$U_n^{(\iota)} = c_0^{(\iota)} F_{n,0}^{(\iota)} + c_1^{(\iota)} F_{n,1}^{(\iota)} = \frac{1}{n} \sum_{i=1}^n W_i^{(\iota)}, \quad (23)$$

where

$$W_i^{(\iota)} = X_i \left[c_0^{(\iota)} + c_1^{(\iota)} \left(\frac{Z_i - z}{h_n} \right) \right] K_h^{(\iota)}(Z_i - z_0)\varepsilon_i. \quad (24)$$

By applying the central limit theorem for strong mixing process (Fan and Yao, 2003, Theorem 2.21) under the mixing condition in Assumption A1 and the moment condition in Assumption A3(i), $\sqrt{nh_n}U_n^{(\iota)}$ is asymptotically normal with mean 0 (due to the law of iterated expectation) and variance (by Lemma 2)

$$nh_n \text{var}(U_n^{(\iota)}) = f_Z(z)\Theta(z) \left[c_0^{(\iota)2} \nu_0^{(\iota)} + 2c_0^{(\iota)} c_1^{(\iota)} \nu_1^{(\iota)} + c_1^{(\iota)2} \nu_2^{(\iota)} \right] + o(1),$$

where $\Theta(z) = \text{E}[XX^\top \sigma^2(X, Z)|Z = z]$. As the remaining term in (22) is deterministic, we obtain

$$\sqrt{nh_n} \left[\hat{a}_n^{(\iota)}(z) - a(z) - \frac{h_n^2}{2} \left(c_0^{(\iota)} \mu_2^{(\iota)} + c_1^{(\iota)} \mu_3^{(\iota)} \right) a''(z) \right] = \frac{\Omega^{-1}(z)}{f_Z(z)} \sqrt{nh_n}U_n^{(\iota)} + o_p(1),$$

37

where the leading term is asymptotically normal with mean 0 and variance $\Phi^{(\iota)}(z)$ given in Theorem 2. $\qquad\square$

Proof of Theorem 3.

It follows from the definition of WRMSE $\Psi_n^{(\iota)}(z)$ in (5) that

$$
\Psi_n^{(\iota)}(z) = \frac{N_n^{(\iota)}}{K_n^{(\iota)}},
$$

where the denominator

$$
K_n^{(\iota)} = \frac{1}{n} \sum_{i=1}^{n} K_h^{(\iota)}(Z_i - z) \tag{25}
$$

and the numerator $N_n^{(\iota)}$, which can be decomposed into three terms, is given by

$$
\begin{aligned}
N_n^{(\iota)} =& \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_{n,i}^{(\iota)2} K_h^{(\iota)}(Z_i - z) \\
=& \frac{1}{n} \sum_{i=1}^{n} \left[Y_i - X_i^{\top} \{\hat{a}_n^{(\iota)}(z) + \hat{b}_n^{(\iota)}(z)(Z_i - z)\} \right]^2 K_h^{(\iota)}(Z_i - z) \\
=& \frac{1}{n} \sum_{i=1}^{n} \left[\varepsilon_i + X_i^{\top} \{a(Z_i) - \hat{a}_n^{(\iota)}(z) - \hat{b}_n^{(\iota)}(z)(Z_i - z)\} \right]^2 K_h^{(\iota)}(Z_i - z) \\
=& \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 K_h^{(\iota)}(Z_i - z) \\
& + \frac{2}{n} \sum_{i=1}^{n} \varepsilon_i \left[X_i^{\top} \{a(Z_i) - \hat{a}_n^{(\iota)}(z) - \hat{b}_n^{(\iota)}(z)(Z_i - z)\} \right] K_h^{(\iota)}(Z_i - z) \tag{26} \\
& + \frac{1}{n} \sum_{i=1}^{n} \left[X_i^{\top} \{a(Z_i) - \hat{a}_n^{(\iota)}(z) - \hat{b}_n^{(\iota)}(z)(Z_i - z)\} \right]^2 K_h^{(\iota)}(Z_i - z) \\
=& N_{n,1}^{(\iota)} + N_{n,2}^{(\iota)} + N_{n,3}^{(\iota)}
\end{aligned}
$$

with $N_{n,1}^{(\iota)}$, $N_{n,2}^{(\iota)}$, and $N_{n,3}^{(\iota)}$ being the first, second, and third terms in (26), respectively. According to Lemmas 1(iv) and 1(v), $N_{n,1}^{(\iota)}/K_n^{(\iota)} = \sigma^2(z) + o_p(1)$ for $z \in D_{1n}^{(\iota)}$. It remains to show $N_{n,2}^{(\iota)} = o_p(1)$ and $N_{n,3}^{(\iota)} = o_p(1)$. By the Taylor expansion of $a(Z_i)$ and the weak consistency results for $F_{n,j}^{(\iota)}$, $\hat{a}_n^{(\iota)}(z)$, and $\hat{b}_n^{(\iota)}(z)$ in Lemmas 1(iii), 1(vi), and 1(vii), respectively, we have

$$
\begin{aligned}
N_{n,2}^{(\iota)} =& \frac{2}{n} \sum_{i=1}^{n} \varepsilon_i \left[X_i^\top \{a(z) + a'(z)(Z_i - z) + o(Z_i - z)\} \right. \\
& \left. - X_i^\top \{\hat{a}_n^{(\iota)}(z) + \hat{b}_n^{(\iota)}(z)(Z_i - z)\} \right] K_h^{(\iota)}(Z_i - z) \\
=& 2\{a(z) - \hat{a}_n^{(\iota)}(z)\}^\top F_{n,0}^{(\iota)} + 2h_n\{a'(z) - \hat{b}_n^{(\iota)}(z)\}^\top F_{n,1}^{(\iota)} + o_p(h_n) \\
=& 2o_p(1) \cdot o_p(1) + 2h_n \cdot o_p(h_n^{-1}) \cdot o_p(1) + o_p(h_n) \\
=& o_p(1).
\end{aligned}
$$

Similarly by the Taylor expansion of $a(Z_i)$, Lemmas 1(i), 1(vi), and 1(vii), and the boundedness condition on $f_Z(z)\Omega(z)$ in Assumption A3(ii), it follows that

$$
\begin{aligned}
N_{n,3}^{(\iota)} =& \frac{1}{n}\sum_{i=1}^{n}\left[X_i^\top\{a(z) + a'(z)(Z_i - z) + \mathrm{o}(Z_i - z)\}\right.\\
&\left.-X_i^\top\{\hat{a}_n^{(\iota)}(z) + \hat{b}_n^{(\iota)}(z)(Z_i - z)\}\right]^2 K_h^{(\iota)}(Z_i - z)\\
=&\{a(z) - \hat{a}_n^{(\iota)}(z)\}^\top S_{n,0}^{(\iota)}\{a(z) - \hat{a}_n^{(\iota)}(z)\}\\
&+ 2h_n\{a(z) - \hat{a}_n^{(\iota)}(z)\}^\top S_{n,1}^{(\iota)}\{a'(z) - \hat{b}_n^{(\iota)}(z)\}\\
&+ h_n^2\{a'(z) - \hat{b}_n^{(\iota)}(z)\}^\top S_{n,2}^{(\iota)}\{a'(z) - \hat{b}_n^{(\iota)}(z)\} + \mathrm{o_p}(h_n)\\
\leq& \mathrm{o_p}(1)\cdot\mathrm{O}\left\{\sup_{z\in D}\|f_Z(z)\Omega(z)\| + \mathrm{o_p}(1)\right\}\cdot\mathrm{o_p}(1)\\
&+ 2h_n\cdot\mathrm{o_p}(1)\cdot\mathrm{O}\left\{\sup_{z\in D}\|f_Z(z)\Omega(z)\| + \mathrm{o_p}(1)\right\}\cdot\mathrm{o_p}(h_n^{-1})\\
&+ h_n^2\cdot\mathrm{o_p}(h_n^{-1})\cdot\mathrm{O}\left\{\sup_{z\in D}\|f_Z(z)\Omega(z)\| + \mathrm{o_p}(1)\right\}\cdot\mathrm{o_p}(h_n^{-1}) + \mathrm{o_p}(h_n)\\
=&\mathrm{o_p}(1).
\end{aligned}
$$

This completes the proof of Theorem 3. $\qquad\square$

Before investigating the limiting behavior of the jump-preserving estimator, we introduce additional notation. For any $z = s_q + \tau h_n$ with $\tau \in (-1, 1)$, we denote random variables

$$
\acute{S}_{n,j}^{(\iota)} = \frac{1}{n}\sum_{i:Z_i<s_q} X_i X_i^\top \left(\frac{Z_i - z}{h_n}\right)^j K_h^{(\iota)}(Z_i - z), \quad j = 0, 1, 2, \tag{27}
$$

$$
\grave{S}_{n,j}^{(\iota)} = \frac{1}{n}\sum_{i:Z_i\geq s_q} X_i X_i^\top \left(\frac{Z_i - z}{h_n}\right)^j K_h^{(\iota)}(Z_i - z), \quad j = 0, 1, 2, \tag{28}
$$

$$
\acute{F}_{n,j}^{(\iota)} = \frac{1}{n}\sum_{i:Z_i<s_q} X_i \left(\frac{Z_i - z}{h_n}\right)^j K_h^{(\iota)}(Z_i - z)\varepsilon_i, \quad j = 0, 1, \tag{29}
$$

and

$$\grave{F}_{n,j}^{(\iota)} = \frac{1}{n} \sum_{i:Z_i \geq s_q} X_i \left(\frac{Z_i - z}{h_n}\right)^j K_h^{(\iota)}(Z_i - z)\varepsilon_i, \quad j = 0,1. \tag{30}$$

Further, let

$$\acute{\mu}_{j,\tau}^{(\iota)} = \int_{-1}^{-\tau} u^j K^{(\iota)}(u)\mathrm{d}u, \quad \grave{\mu}_{j,\tau}^{(\iota)} = \int_{-\tau}^{1} u^j K^{(\iota)}(u)\mathrm{d}u, \tag{31}$$

$$\Omega_-(s_q) = \lim_{z\uparrow s_q} \mathrm{E}[XX^\top|Z = z], \quad \Omega_+(s_q) = \lim_{z\downarrow s_q} \mathrm{E}[XX^\top|Z = z], \tag{32}$$

$$\acute{\Omega}_{-,\tau}^{(\iota)}(s_q) = \begin{pmatrix} \acute{\mu}_{0,\tau}^{(\iota)}\Omega_-(s_q) & \acute{\mu}_{1,\tau}^{(\iota)}\Omega_-(s_q) \\ \acute{\mu}_{1,\tau}^{(\iota)}\Omega_-(s_q) & \acute{\mu}_{2,\tau}^{(\iota)}\Omega_-(s_q) \end{pmatrix},$$

$$\grave{\Omega}_{+,\tau}^{(\iota)}(s_q) = \begin{pmatrix} \grave{\mu}_{0,\tau}^{(\iota)}\Omega_+(s_q) & \grave{\mu}_{1,\tau}^{(\iota)}\Omega_+(s_q) \\ \grave{\mu}_{1,\tau}^{(\iota)}\Omega_+(s_q) & \grave{\mu}_{2,\tau}^{(\iota)}\Omega_+(s_q) \end{pmatrix},$$

$$a_-(s_q) = \lim_{z\uparrow s_q} a(z), \quad \text{and} \quad a_+(s_q) = \lim_{z\downarrow s_q} a(z) = a_-(s_q) + d_q.$$

Without loss of generality, we assume that $a(\cdot)$ is right continuous, i.e., $a(s_q) = a_+(s_q)$ for $q = 0,\ldots,Q$. By the mean value theorem and boundedness of the (left) partial derivatives of $a(\cdot)$ (Assumption A5), it holds for $Z_i \in [s_q - (1 - \tau)h_n, s_q)$ that

$$a(Z_i) = a_-(s_q) + \mathrm{O}(Z_i - s_q). \tag{33}$$

Similarly, we have for $Z_i \in (s_q, s_q + (1 + \tau)h_n]$,

$$a(Z_i) = a_+(s_q) + \mathrm{O}(Z_i - s_q) = a_-(s_q) + d_q + \mathrm{O}(Z_i - s_q). \tag{34}$$

Using equations (33) and (34) and the consistency results for $\acute{F}_{n,j}^{(\iota)}$, $\grave{F}_{n,j}^{(\iota)}$, $\acute{S}_{n,j}^{(\iota)}$, and $\grave{S}_{n,j}^{(\iota)}$ in

41

Lemmas 4(i) and 4(ii), we have for $j = 0, 1$,

$$
\begin{aligned}
T_{n,j}^{(\iota)} =& \frac{1}{n} \sum_{i=1}^{n} X_i \left[X_i^\top a(Z_i) + \varepsilon_i \right] \left(\frac{Z_i - z}{h_n} \right)^j K_h^{(\iota)}(Z_i - z) \\
=& \frac{1}{n} \sum_{i:Z_i < s_q} X_i X_i^\top \left(\frac{Z_i - z}{h_n} \right)^j K_h^{(\iota)}(Z_i - z) a(Z_i) + \acute{F}_{n,j}^{(\iota)} \\
&+ \frac{1}{n} \sum_{i:Z_i \geq s_q} X_i X_i^\top \left(\frac{Z_i - z}{h_n} \right)^j K_h^{(\iota)}(Z_i - z) a(Z_i) + \grave{F}_{n,j}^{(\iota)} \\
=& \frac{1}{n} \sum_{i:Z_i < s_q} X_i X_i^\top \left(\frac{Z_i - z}{h_n} \right)^j K_h^{(\iota)}(Z_i - z) \left\{ a_-(s_q) + \mathrm{O}(Z_i - s_q) \right\} + \mathrm{o_p}(1) \\
&+ \frac{1}{n} \sum_{i:Z_i \geq s_q} X_i X_i^\top \left(\frac{Z_i - z}{h_n} \right)^j K_h^{(\iota)}(Z_i - z) \left\{ a_+(s_q) + \mathrm{O}(Z_i - s_q) \right\} + \mathrm{o_p}(1) \\
=& \acute{S}_{n,j}^{(\iota)} a_-(s_q) + \grave{S}_{n,j}^{(\iota)} \left\{ a_-(s_q) + d_q \right\} + \mathrm{O_p}(h_n) + \mathrm{o_p}(1) \\
=& f_Z(s_q) \left[\left\{ \acute{\mu}_{j,\tau}^{(\iota)} \Omega_-(s_q) + \grave{\mu}_{j,\tau}^{(\iota)} \Omega_+(s_q) \right\} a_-(s_q) + \grave{\mu}_{j,\tau}^{(\iota)} \Omega_+(s_q) d_q \right] + \mathrm{o_p}(1).
\end{aligned}
$$

Hence, by Lemmas 4(i), 8(ii), and 9, the local linear estimator in (18) can be expressed for $z = s_q + \tau h_n$ with $\tau \in (-1, 1)$ as

$$
\begin{aligned}
H_n \hat{\beta}_n^{(\iota)} =& S_n^{(\iota)^{-1}} T_n^{(\iota)} \\
=& \left[\acute{\Omega}_{-,\tau}^{(\iota)}(s_q) + \grave{\Omega}_{+,\tau}^{(\iota)}(s_q) \right]^{-1} (1 + \mathrm{o_p}(1)) \\
& \cdot \left[\begin{pmatrix} \acute{\mu}_{0,\tau}^{(\iota)} \Omega_-(s_q) + \grave{\mu}_{0,\tau}^{(\iota)} \Omega_+(s_q) \\ \acute{\mu}_{1,\tau}^{(\iota)} \Omega_-(s_q) + \grave{\mu}_{1,\tau}^{(\iota)} \Omega_+(s_q) \end{pmatrix} a_-(s_q) + \begin{pmatrix} \grave{\mu}_{0,\tau}^{(\iota)} \Omega_+(s_q) \\ \grave{\mu}_{1,\tau}^{(\iota)} \Omega_+(s_q) \end{pmatrix} d_q + \mathrm{o_p}(1) \right] \\
=& \begin{pmatrix} I_p \\ 0_p \end{pmatrix} a_-(s_q) + \begin{pmatrix} \Xi_{0,\tau}^{(\iota)} \\ \Xi_{1,\tau}^{(\iota)} \end{pmatrix} d_q + \mathrm{o_p}(1),
\end{aligned}
\tag{35}
$$

where I_p is the $p \times p$ identity matrix, 0_p is the null matrix of size $p \times p$, and

$$\begin{pmatrix} \Xi_{0,\tau}^{(\iota)} \\ \Xi_{1,\tau}^{(\iota)} \end{pmatrix} = \left[\acute{\Omega}_{-,\tau}^{(\iota)}(s_q) + \grave{\Omega}_{+,\tau}^{(\iota)}(s_q) \right]^{-1} \begin{pmatrix} \acute{\mu}_{0,\tau}^{(\iota)} \Omega_+(s_q) \\ \acute{\mu}_{1,\tau}^{(\iota)} \Omega_+(s_q) \end{pmatrix}. \tag{36}$$

Note that, according to the definition of the right-sided kernel $K^{(r)}(\cdot)$ in (2), one has for $\tau \in (0, 1)$,

$$\acute{\mu}_{j,\tau}^{(r)} = \int_{-1}^{-\tau} u^j K^{(c)}(u) \mathbf{1} \{ u \geq 0 \} \, \mathrm{d}u = 0, \tag{37}$$

which implies that $\acute{\Omega}_{-,\tau}^{(r)}(s_q) = 0_{2p}$ and

$$\begin{pmatrix} \Xi_{0,\tau}^{(r)} \\ \Xi_{1,\tau}^{(r)} \end{pmatrix} = \grave{\Omega}_{+,\tau}^{(r)-1}(s_q) \begin{pmatrix} \grave{\mu}_{0,\tau}^{(r)} \Omega_+(s_q) \\ \grave{\mu}_{1,\tau}^{(r)} \Omega_+(s_q) \end{pmatrix} = \begin{pmatrix} I_p \\ 0_p \end{pmatrix} \tag{38}$$

due to Lemma 8(ii). Similarly, for $\tau \in (-1, 0)$ and the left-sided kernel $K^{(l)}(\cdot)$, we obtain

$$\grave{\mu}_{j,\tau}^{(l)} = 0, \quad \Xi_{0,\tau}^{(l)} = 0_p, \quad \text{and} \quad \Xi_{1,\tau}^{(l)} = 0_p. \tag{39}$$

Proof of Theorem 4.

In order to prove Theorem 4 for continuous conditional error variance function $\sigma^2(z)$ (Assumption A6), we analyze the limiting properties of each term of the decomposition of $N_n^{(\iota)}$ in (26). First, by Lemma 4(iv), $N_{n,1}^{(\iota)} = f_Z(s_q)\mu_0^{(\iota)}\sigma^2(s_q) + \mathrm{o}_p(1)$. Using equations

(33)–(35), one obtains

$$N_{n,2}^{(\iota)} = \frac{2}{n} \sum_{i=1}^{n} \left[a(Z_i) - \hat{a}_n^{(\iota)}(z) - h_n \hat{b}_n^{(\iota)}(z) \left(\frac{Z_i - z}{h_n} \right) \right]^{\top} X_i \varepsilon_i K_h^{(\iota)}(Z_i - z)$$

$$= \frac{2}{n} \sum_{i:Z_i < s_q} \left[\underbrace{a(Z_i) - a_-(s_q)}_{O(Z_i - s_q)} - \Xi_{0,\tau}^{(\iota)} d_q - \left(\frac{Z_i - z}{h_n} \right) \Xi_{1,\tau}^{(\iota)} d_q \right]^{\top} X_i \varepsilon_i K_h^{(\iota)}(Z_i - z)$$

$$+ \frac{2}{n} \sum_{i:Z_i \geq s_q} \left[\underbrace{a(Z_i) - a_-(s_q)}_{d_q + O(Z_i - s_q)} - \Xi_{0,\tau}^{(\iota)} d_q - \left(\frac{Z_i - z}{h_n} \right) \Xi_{1,\tau}^{(\iota)} d_q \right]^{\top} X_i \varepsilon_i K_h^{(\iota)}(Z_i - z)$$

$$+ o_p(1)$$

$$= -2[\Xi_{0,\tau}^{(\iota)} d_q]^{\top} \acute{F}_{n,0}^{(\iota)} - 2[\Xi_{1,\tau}^{(\iota)} d_q]^{\top} \acute{F}_{n,1}^{(\iota)} - 2[(\Xi_{0,\tau}^{(\iota)} - I_p) d_q]^{\top} \grave{F}_0^{(\iota)} - 2[\Xi_{1,\tau}^{(\iota)} d_q]^{\top} \grave{F}_{n,1}^{(\iota)}$$

$$+ O_p(h_n) + o_p(1).$$

44

Hence, $N_{n,2}^{(\iota)} = o_p(1)$ due to the consistency results for $\acute{F}_{n,j}^{(\iota)}$ and $\grave{F}_{n,j}^{(\iota)}$ in Lemma 4(ii). Again, it follows from (33)–(35) that

$$
\begin{aligned}
N_{n,3}^{(\iota)} =& \frac{1}{n} \sum_{i=1}^{n} \left[X_i^\top a(Z_i) - X_i^\top \left\{ \hat{a}_n^{(\iota)}(z) + h_n \hat{b}_n^{(\iota)}(z) \left(\frac{Z_i - z}{h_n} \right) \right\} \right]^2 K_h^{(\iota)}(Z_i - z) \\
=& \frac{1}{n} \sum_{i:Z_i < s_q} \left[X_i^\top \left\{ \underbrace{a(Z_i) - a_-(s_q)}_{O(Z_i - s_q)} - \Xi_{0,\tau}^{(\iota)} d_q - \left(\frac{Z_i - z}{h_n} \right) \Xi_{1,\tau}^{(\iota)} d_q \right\} \right]^2 K_h^{(\iota)}(Z_i - z) \\
&+ \frac{1}{n} \sum_{i:Z_i \geq s_q} \left[X_i^\top \left\{ \underbrace{a(Z_i) - a_-(s_q)}_{d_q + O(Z_i - s_q)} - \Xi_{0,\tau}^{(\iota)} d_q - \left(\frac{Z_i - z}{h_n} \right) \Xi_{1,\tau}^{(\iota)} d_q \right\} \right]^2 K_h^{(\iota)}(Z_i - z) \\
&+ o_p(1) \\
=& d_q^\top \Xi_{0,\tau}^{(\iota)\top} \acute{S}_{n,0}^{(\iota)} \Xi_{0,\tau}^{(\iota)} d_q + 2 d_q^\top \Xi_{0,\tau}^{(\iota)\top} \acute{S}_{n,1}^{(\iota)} \Xi_{1,\tau}^{(\iota)} d_q + d_q^\top \Xi_{1,\tau}^{(\iota)\top} \acute{S}_{n,2}^{(\iota)} \Xi_{1,\tau}^{(\iota)} d_q \\
&+ d_q^\top [\Xi_{0,\tau}^{(\iota)} - I_p]^\top \grave{S}_{n,0}^{(\iota)} [\Xi_{0,\tau}^{(\iota)} - I_p] d_q + 2 d_q^\top [\Xi_{0,\tau}^{(\iota)} - I_p]^\top \grave{S}_{n,1}^{(\iota)} \Xi_{1,\tau}^{(\iota)} d_q \\
&+ d_q^\top \Xi_{1,\tau}^{(\iota)\top} \grave{S}_{n,2}^{(\iota)} \Xi_{1,\tau}^{(\iota)} d_q + O_p(h_n) + o_p(1) \\
=& d_q^\top \begin{pmatrix} \Xi_{0,\tau}^{(\iota)} \\ \Xi_{1,\tau}^{(\iota)} \end{pmatrix}^\top \begin{bmatrix} \acute{S}_{n,0}^{(\iota)} & \acute{S}_{n,1}^{(\iota)} \\ \acute{S}_{n,1}^{(\iota)} & \acute{S}_{n,2}^{(\iota)} \end{bmatrix} \begin{pmatrix} \Xi_{0,\tau}^{(\iota)} \\ \Xi_{1,\tau}^{(\iota)} \end{pmatrix} d_q \\
&+ d_q^\top \begin{pmatrix} \Xi_{0,\tau}^{(\iota)} - I_p \\ \Xi_{1,\tau}^{(\iota)} \end{pmatrix}^\top \begin{bmatrix} \grave{S}_{n,0}^{(\iota)} & \grave{S}_{n,1}^{(\iota)} \\ \grave{S}_{n,1}^{(\iota)} & \grave{S}_{n,2}^{(\iota)} \end{bmatrix} \begin{pmatrix} \Xi_{0,\tau}^{(\iota)} - I_p \\ \Xi_{1,\tau}^{(\iota)} \end{pmatrix} d_q + O_p(h_n) + o_p(1).
\end{aligned}
$$

It follows from the consistency results for $\acute{S}_{n,j}^{(\iota)}$ and $\grave{S}_{n,j}^{(\iota)}$ in Lemma 4(i) that

$$
\begin{aligned}
N_{n,3}^{(\iota)} =& f_Z(s_q) d_q^\top \left\{ \begin{pmatrix} \Xi_{0,\tau}^{(\iota)} \\ \Xi_{1,\tau}^{(\iota)} \end{pmatrix}^\top \acute{\Omega}_{-,\tau}^{(\iota)}(s_q) \begin{pmatrix} \Xi_{0,\tau}^{(\iota)} \\ \Xi_{1,\tau}^{(\iota)} \end{pmatrix} \right. \\
&+ \left. \begin{pmatrix} \Xi_{0,\tau}^{(\iota)} - I_p \\ \Xi_{1,\tau}^{(\iota)} \end{pmatrix}^\top \grave{\Omega}_{+,\tau}^{(\iota)}(s_q) \begin{pmatrix} \Xi_{0,\tau}^{(\iota)} - I_p \\ \Xi_{1,\tau}^{(\iota)} \end{pmatrix} \right\} d_q + o_p(1) \\
=& f_Z(s_q) d_q^\top \mu_0^{(\iota)} C_\tau^{(\iota)} d_q + o_p(1),
\end{aligned}
$$

45

where

$$
\begin{aligned}
C_\tau^{(\iota)} = {} & \frac{1}{\mu_0^{(\iota)}} \begin{pmatrix} \Xi_{0,\tau}^{(\iota)} \\ \Xi_{1,\tau}^{(\iota)} \end{pmatrix}^{\top} \acute{\Omega}_{-,\tau}^{(\iota)}(s_q) \begin{pmatrix} \Xi_{0,\tau}^{(\iota)} \\ \Xi_{1,\tau}^{(\iota)} \end{pmatrix} \\
& + \frac{1}{\mu_0^{(\iota)}} \begin{pmatrix} \Xi_{0,\tau}^{(\iota)} - I_p \\ \Xi_{1,\tau}^{(\iota)} \end{pmatrix}^{\top} \grave{\Omega}_{+,\tau}^{(\iota)}(s_q) \begin{pmatrix} \Xi_{0,\tau}^{(\iota)} - I_p \\ \Xi_{1,\tau}^{(\iota)}. \end{pmatrix}.
\end{aligned} \tag{40}
$$

For $\tau \in (0,1)$ and $\iota = c$ or l, $\acute{\mu}_{0,\tau}^{(\iota)}$ and $\grave{\mu}_{0,\tau}^{(\iota)}$ are nonzero. According to Lemma 10, both matrices $\Xi_{0,\tau}^{(\iota)}$ and $\Xi_{0,\tau}^{(\iota)} - I_p$ have rank p. Hence, $[\Xi_{0,\tau}^{(\iota)} \ \Xi_{1,\tau}^{(\iota)}]^{\top}$ and $[\Xi_{0,\tau}^{(\iota)} - I_p \ \Xi_{1,\tau}^{(\iota)}]^{\top}$ have the same rank p, thus full column rank. By the result that $\acute{\Omega}_{-,\tau}^{(\iota)}(s_q)$ and $\grave{\Omega}_{+,\tau}^{(\iota)}(s_q)$ are positive definite for $\tau \in (0,1)$ and $\iota = c$ or l in Lemma 9, the property that $A + B \succ 0$ for any $A \succ 0$ and $B \succ 0$, and the fact that $A^{\top}BA \succ 0$ if $B \succ 0$ and A has full column rank, we conclude that matrices $C_\tau^{(c)}$ and $C_\tau^{(l)}$ are positive definite for $\tau \in (0,1)$.

For $\tau \in (0,1)$ and $\iota = r$, it follows from equation (37): $\acute{\mu}_{j,\tau}^{(r)} = 0$ that

$$
C_\tau^{(r)} = \frac{1}{\mu_0^{(r)}} \begin{pmatrix} \Xi_{0,\tau}^{(r)} - I_p \\ \Xi_{1,\tau}^{(r)} \end{pmatrix}^{\top} \begin{bmatrix} \grave{\mu}_{0,\tau}^{(r)}\Omega_+(s_q) & \grave{\mu}_{1,\tau}^{(r)}\Omega_+(s_q) \\ \grave{\mu}_{1,\tau}^{(r)}\Omega_+(s_q) & \grave{\mu}_{2,\tau}^{(r)}\Omega_+(s_q) \end{bmatrix} \begin{pmatrix} \Xi_{0,\tau}^{(r)} - I_p \\ \Xi_{1,\tau}^{(r)} \end{pmatrix}.
$$

Since $\Xi_{0,\tau}^{(r)} = I_p$ and $\Xi_{1,\tau}^{(r)} = 0_p$ (equation (38)), $C_\tau^{(r)}$ is a null matrix for $\tau \in (0,1)$.

Similarly, for $\tau \in (-1,0)$, we have positive definite matrices $C_\tau^{(c)} \succ 0$ and $C_\tau^{(r)} \succ 0$ and the null matrix $C_\tau^{(l)} = 0_p$. Combining the limiting results of $N_{n,1}^{(\iota)}$, $N_{n,2}^{(\iota)}$, $N_{n,3}^{(\iota)}$, and $K_n^{(\iota)}$ (due to Lemma 4(iii)) yields Theorem 4. $\qquad \square$

Proof of Theorem 5.

Following the proof of Theorem 3.2 in Gijbels et al. (2007), we write the jump-preserving estimator $\check{a}_n(z)$ as

$$\check{a}_n(z) = \hat{a}_n^{(c)}(z)\mathbf{1}\{A_n(z)\} + \hat{a}_n^{(l)}(z)\mathbf{1}\{B_n(z)\} + \hat{a}_n^{(r)}(z)\mathbf{1}\{C_n(z)\}$$
$$+ \frac{\hat{a}_n^{(l)}(z) + \hat{a}_n^{(r)}(z)}{2}\mathbf{1}\{BC_n(z)\},$$

in which $A_n(z)$, $B_n(z)$, $C_n(z)$, and $BC_n(z)$ correspond to the inequalities in (6) from top to bottom, respectively. Apparently, these sets are mutually exclusive, and for any $z \in D$,

$$\mathbf{1}\{A_n(z)\} + \mathbf{1}\{B_n(z)\} + \mathbf{1}\{C_n(z)\} + \mathbf{1}\{BC_n(z)\} = 1. \tag{41}$$

The rest of the proof is separated into three parts, which correspond to the regions D_{1n}, $D_{2n,\delta}$ for some $\delta \in (0, 1/2)$, and D_{2n} given in equation (9).

Part (i)

First, we consider z in the continuous region D_{1n}. According to Theorem 1, there exist a positive integer $n^{(\iota)}$ and a positive constant $C^{(\iota)} > 0$ such that for $n > n^{(\iota)}$,

$$\sup_{z \in D_{1n}} \sqrt{\frac{nh_n}{\ln n}} \left\| \hat{a}_n^{(\iota)}(z) - a(z) \right\| \le C^{(\iota)}, \quad \iota = c, l, r,$$

with probability approaching to 1. Take $\zeta = \max_{\iota=\{c,l,r\}} C^{(\iota)}$; for $n > \max_{\iota=\{c,l,r\}} n^{(\iota)}$, it

follows that

$$
\sup_{z \in D_{1n}} \sqrt{\frac{nh_n}{\ln n}} \left\| \breve{a}_n(z) - a(z) \right\| = \sup_{z \in D_{1n}} \sqrt{\frac{nh_n}{\ln n}} \left\| \hat{a}_n^{(c)}(z) - a(z) \right\| \mathbf{1}\left\{ A_n(z) \right\}
$$

$$
+ \sup_{z \in D_{1n}} \sqrt{\frac{nh_n}{\ln n}} \left\| \hat{a}_n^{(l)}(z) - a(z) \right\| \mathbf{1}\left\{ B_n(z) \right\}
$$

$$
+ \sup_{z \in D_{1n}} \sqrt{\frac{nh_n}{\ln n}} \left\| \hat{a}_n^{(r)}(z) - a(z) \right\| \mathbf{1}\left\{ C_n(z) \right\}
$$

$$
+ \sup_{z \in D_{1n}} \sqrt{\frac{nh_n}{\ln n}} \left\| \frac{\hat{a}_n^{(l)}(z) + \hat{a}_n^{(r)}(z)}{2} - a(z) \right\| \mathbf{1}\left\{ BC_n(z) \right\}
$$

$$
\leq \zeta
$$

with probability approaching to 1, which implies that

$$
\sup_{z \in D_{1n}} \sqrt{\frac{nh_n}{\ln n}} \left\| \breve{a}_n(z) - a(z) \right\| = \mathrm{O_p}(1).
$$

Part (ii)

Next, we prove the uniform consistency for $\breve{a}_n(z)$ in the region $D_{2n,\delta}$ for some $\delta \in (0, 1/2)$, which contains neighborhoods of discontinuities excluding any small regions around centers of s_q and around end points $s_q - h_n$ and $s_q + h_n$. For some $\delta \in (0, 1/2)$, the region $D_{2n,\delta}$ consists of two disjoint sets:

$$
\acute{D}_{2n,\delta} = D \cap \bigcup_{q=0}^{Q+1} [s_q - (1-\delta)h_n, s_q - \delta h_n]
$$

and

$$
\grave{D}_{2n,\delta} = D \cap \bigcup_{q=0}^{Q+1} [s_q + \delta h_n, s_q + (1-\delta)h_n].
$$

Consider the region $\acute{D}_{2n,\delta}$ and an arbitrarily small number $\epsilon > 0$. Any given point z

in $\acute{D}_{2n,\delta}$ satisfies $z = s_q + \tau h_n$ with $\tau \in [-1+\delta, -\delta]$ and s_q is one of $\{s_q\}_{q=0}^{Q+1}$. According to Theorem 1, for some $\zeta > 0$ and any $\epsilon > 0$, there exist a positive integer n_1 such that for $n > n_1$,

$$\sup_{z \in \acute{D}_{2n,\delta}} \sqrt{\frac{nh_n}{\ln n}} \left\| \hat{a}_n^{(l)}(z) - a(z) \right\| \le \zeta$$

with probability larger than $1 - \epsilon$. In the following, we show that for any $z \in \acute{D}_{2n,\delta}$, there exists another positive integer $n_3 > 0$ such that the difference of $\breve{a}_n(z)$ and $\hat{a}_n^{(l)}(z)$ is negligible in probability.

By Theorem 4, for any $\kappa > 0$ and $\epsilon > 0$, there exists an integer $n_\kappa(\kappa)$ such that for $n > n_\kappa(\kappa)$,

$$\Psi_n^{(c)}(z) > d_q^\top C_\tau^{(c)} d_q + \sigma^2(s_q) - \kappa,$$
$$\Psi_n^{(l)}(z) < \sigma^2(s_q) + \kappa,$$
$$\Psi_n^{(r)}(z) > d_q^\top C_\tau^{(r)} d_q + \sigma^2(s_q) - \kappa$$

with probability larger than $1 - \epsilon$. For $\tau \in [-1+\delta, -\delta]$, matrices $C_\tau^{(c)}$ and $C_\tau^{(r)}$ are positive definite (see the proof of Theorem 4). Additionally, the continuity of $C_\tau^{(\iota)}$ in τ follows from the continuity of $\acute{\mu}_{j,\tau}^{(\iota)}$ and $\grave{\mu}_{j,\tau}^{(\iota)}$ as functions of the limits of integration. Given the continuity of $C_\tau^{(\iota)}$ and thus of $d_q^\top C_\tau^{(\iota)} d_q$, we have for any $d_q \ne 0$,

$$a_\tau = \inf_{\tau \in [-1+\delta, -\delta]} \min\{d_q^\top C_\tau^{(c)} d_q, d_q^\top C_\tau^{(r)} d_q\} = \min_{\tau \in [-1+\delta, -\delta]} \min\{d_q^\top C_\tau^{(c)} d_q, d_q^\top C_\tau^{(r)} d_q\} > 0.$$

Set $\kappa = \frac{a_\tau}{4}$. For $n > n_2 = n_\kappa(\frac{a_\tau}{4})$, it follows that

$$\Psi_n^{(c)}(z) - \Psi_n^{(l)}(z) \geq \min\{\Psi_n^{(c)}(z), \Psi_n^{(r)}(z)\} - \Psi_n^{(l)}(z)$$
$$> a_\tau - 2\kappa = a_\tau - \frac{a_\tau}{2} = \frac{a_\tau}{2} > 0,$$

and hence,

$$\mathrm{diff}(z) = \Psi_n^{(c)}(z) - \min\{\Psi_n^{(l)}(z), \Psi_n^{(r)}(z)\}$$
$$= \Psi_n^{(c)}(z) - \Psi_n^{(l)}(z) > \frac{a_\tau}{2} > 0$$

with probability larger than $1 - \epsilon$. Moreover, since $u_n \to 0$, for any $\eta > 0$ there exists $n_\eta(\eta) > 0$ such that, for $n > n_\eta(\eta)$, we have $|u_n| < \eta$ with probability larger than $1 - \epsilon$. Setting $\eta = a_\tau/4$, it follows for $n > n_3 = \max\{n_\eta(\frac{a_\tau}{4}), n_2\}$,

$$\mathrm{diff}(z) - u_n > \frac{a_\tau}{2} - u_n > \frac{a_\tau}{2} - \frac{a_\tau}{4} = \frac{a_\tau}{4} > 0,$$

which implies that Conditions $A_n(z)$, $C_n(z)$, and $BC_n(z)$ do not hold, i.e., $\mathbf{1}\{A_n(z)\} + \mathbf{1}\{C_n(z)\} + \mathbf{1}\{BC_n(z)\} = 0$ with probability larger than $1 - 2\epsilon$. Moreover, by equation (41), we can claim with an arbitrarily high probability that only Condition $B_n(z)$ is satisfied, which means that $\hat{a}_n^{(l)}(z)$ is chosen for $n > n_3$ with probability larger than $1 - 2\epsilon$. Hence when $n > n_4 = \max\{n_1, n_3\}$,

$$\sup_{z \in \acute{D}_{2n,\delta}} \sqrt{\frac{nh_n}{\ln n}} \|\check{a}_n(z) - a(z)\| = \sup_{z \in \acute{D}_{2n,\delta}} \sqrt{\frac{nh_n}{\ln n}} \|\hat{a}_n^{(l)}(z) - a(z)\| \leq \zeta$$

with probability larger than $1 - 3\epsilon$, which implies

$$\sup_{z \in \acute{D}_{2n,\delta}} \sqrt{\frac{nh_n}{\ln n}} \|\check{a}_n(z) - a(z)\| = O_{\mathrm{p}}(1). \tag{42}$$

Similarly, for $z \in \dot{D}_{2n,\delta}$, one can also show that

$$\sup_{z \in \dot{D}_{2n,\delta}} \sqrt{\frac{nh_n}{\ln n}} \|\breve{a}_n(z) - a(z)\| = \sup_{z \in \dot{D}_{2n,\delta}} \sqrt{\frac{nh_n}{\ln n}} \|\hat{a}_n^{(r)}(z) - a(z)\| + o_p(1)$$

$$= O_p(1). \tag{43}$$

Combining (42) and (43) gives

$$\sup_{z \in D_{2n,\delta}} \sqrt{\frac{nh_n}{\ln n}} \|\breve{a}_n(z) - a(z)\| = O_p(1).$$

Part (iii)

For $z \in D_{2n} \setminus D_{2n,\delta}$, we can show the consistency of $\breve{a}_n(z)$ analogously to the proof of Part (ii). Since there is no unique strictly positive lower bound a_τ exists, the result is not uniform with respect to z on $D_{2n} \setminus D_{2n,\delta}$. $\qquad\square$

Proof of Theorem 6.

We showed in the proof of Theorem 5 that the jump-preserving estimator $\breve{a}_n(z)$ picks consistently the correct local estimator for $z \in D \setminus \{s_q\}_{q=0}^{Q+1}$. By Theorem 2, each local linear estimator is asymptotically normal in the regions, where it is selected. Consequently, $\breve{a}_n(z)$ is asymptotically normal for $z \in D \setminus \{s_q\}_{q=0}^{Q+1}$ with distribution given in Theorem 6. A detailed argument is given in the proof of Theorem 3.1 of Casas and Gijbels (2012). $\qquad\square$

Proof of Theorem 7.

Recall that the estimated residual used in Theorems 7–10 is $\tilde{\varepsilon}_{n,i}^{(\iota)} = Y_i - X_i^\top \hat{a}_n^{(\iota)}(z)$ and the kernel $\tilde{K}$ refers to the uniform kernel. Let us denote

$$\tilde{N}_n^{(\iota)} = \frac{1}{n} \sum_{i=1}^n \hat{e}_{n,i}^{(\iota)2} \tilde{K}_h^{(\iota)}(Z_i - z),$$

$$\tilde{T}_n^{(\iota)} = \frac{1}{n} \sum_{i=1}^{n} \rho \left(\frac{Z_i - z}{h_n} \right) \tilde{\varepsilon}_{n,i}^{(\iota)} \tilde{K}_h^{(\iota)}(Z_i - z),$$

$$\tilde{T}_{n,2}^{(\iota)} = \frac{1}{n} \sum_{i=1}^{n} \rho \left(\frac{Z_i - z}{h_n} \right) X_i^\top \{a(Z_i) - \hat{a}_n^{(\iota)}(z)\} \tilde{K}_h^{(\iota)}(Z_i - z),$$

$$\tilde{S}_n^{(\iota)} = \frac{1}{n} \sum_{i=1}^{n} \rho \left(\frac{Z_i - z}{h_n} \right) \rho^\top \left(\frac{Z_i - z}{h_n} \right) \tilde{K}_h^{(\iota)}(Z_i - z), \tag{44}$$

$$\tilde{W}_{n,1}^{(\iota)} = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 \tilde{K}_h^{(\iota)}(Z_i - z), \tag{45}$$

$$\tilde{W}_{n,2}^{(\iota)} = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^\top \tilde{K}_h^{(\iota)}(Z_i - z), \tag{46}$$

$$\tilde{W}_{n,3}^{(\iota)} = \frac{1}{n} \sum_{i=1}^{n} \rho \left(\frac{Z_i - z}{h_n} \right) X_i^\top \tilde{K}_h^{(\iota)}(Z_i - z), \tag{47}$$

$$\tilde{W}_{n,4}^{(\iota)} = \frac{1}{n} \sum_{i=1}^{n} X_i \varepsilon_i \tilde{K}_h^{(\iota)}(Z_i - z), \quad \text{and} \tag{48}$$

$$\tilde{W}_{n,5}^{(\iota)} = \frac{1}{n} \sum_{i=1}^{n} \rho \left(\frac{Z_i - z}{h_n} \right) \varepsilon_i \tilde{K}_h^{(\iota)}(Z_i - z), \tag{49}$$

where $\rho(u) = (1, u, \ldots, u^m)^\top$ and $\hat{e}_{n,i}^{(\iota)} = \tilde{\varepsilon}_{n,i}^{(\iota)} - \rho^\top((Z_i - z)/h_n)\hat{\gamma}_n^{(\iota)}$. Further, we define the population counterparts of some of the above kernel weighted averages:

$$\mho(z) = \mathrm{E}[X^\top | Z = z], \quad \tilde{\mu}_0^{(\iota)} = \int_{-1}^{1} \tilde{K}(u) \mathrm{d}u, \quad \tilde{m}^{(\iota)} = \int_{-1}^{1} \rho(u) \tilde{K}(u) \mathrm{d}u, \quad \text{and}$$

$$\tilde{M}^{(\iota)} = \int_{-1}^{1} \rho(u) \rho^\top(u) \tilde{K}(u) \mathrm{d}u. \tag{50}$$

With the help of the above notation, we write $\hat{\gamma}_n^{(\iota)}$ in (10) as

$$\hat{\gamma}_n^{(\iota)}(z) = \tilde{S}_n^{(\iota)^{-1}} \tilde{T}_n^{(\iota)} = \tilde{S}_n^{(\iota)^{-1}} (\tilde{W}_{n,5}^{(\iota)} + \tilde{T}_{n,2}^{(\iota)}). \tag{51}$$

By Lemma 5(vi), $\tilde{W}_{n,5}^{(\iota)} = \mathrm{o_p}(1)$. To show $\tilde{T}_{n,2}^{(\iota)} = \mathrm{o_p}(1)$ by the consistency results for $\hat{a}_n^{(\iota)}(z)$ and $\tilde{W}_{n,3}^{(\iota)}$ in Lemmas 1(vi) and 5(iv), respectively, the Taylor expansion of $a(Z_i)$ for $Z_i \in [z - h_n, z] : z \in D_{1n}^{(l)}$, $Z_i \in [z, z + h_n] : z \in D_{1n}^{(r)}$, or $Z_i \in [z - h_n, z + h_n] : z \in D_{1n}^{(c)}$

is used along with the boundedness of $a'(\cdot)$ (Assumption A5):

$$
\begin{aligned}
\tilde{T}_{n,2}^{(\iota)} &= \frac{1}{n} \sum_{i=1}^{n} \rho\left(\frac{Z_i - z}{h_n}\right) X_i^\top \{a(z) - \hat{a}_n^{(\iota)}(z) + \mathrm{O}(h_n)\} \tilde{K}_h^{(\iota)}(Z_i - z) \\
&= \tilde{W}_{n,3}^{(\iota)} [a(z) - \hat{a}_n^{(\iota)}(z) + \mathrm{O}(h_n)] \\
&= \{f_Z(z) \tilde{m}^{(\iota)} \mho(z) + \mathrm{o}_{\mathrm{p}}(1)\} \cdot \mathrm{o}_{\mathrm{p}}(1) = \mathrm{o}_{\mathrm{p}}(1). \qquad (h_n \to 0)
\end{aligned}
$$

As $\tilde{T}_n^{(\iota)} = \tilde{W}_{n,5}^{(\iota)} + \tilde{T}_{n,2}^{(\iota)} = \mathrm{o}_{\mathrm{p}}(1)$ by Lemma 5(vi), the consistency result for $\tilde{S}_n^{(\iota)}$ in Lemma 5(i) and the invertibility conditions for its population counterparts in Assumptions A2 and D1 imply

$$
\hat{\gamma}_n^{(\iota)} = \frac{\tilde{M}^{(\iota)-1}}{f_Z(z)} \mathrm{o}_{\mathrm{p}}(1) = \mathrm{o}_{\mathrm{p}}(1). \tag{52}
$$

Further, for $Z_i \in [z - h_n, z] : z \in D_{1n}^{(l)}$, $Z_i \in [z, z + h_n] : z \in D_{1n}^{(r)}$, or $Z_i \in [z - h_n, z + h_n] :$ $z \in D_{1n}^{(c)}$, the squared error from the local mth polynomial fitting of $\hat{\varepsilon}_{n,i}^{(\iota)}$ equals

$$
\begin{aligned}
\hat{e}_{n,i}^{(\iota)2} &= \varepsilon_i^2 + \left\{\hat{\gamma}_n^{(\iota)\top} \rho\left(\frac{Z_i - z}{h_n}\right)\right\}^2 + \left\{X_i^\top [a(Z_i) - \hat{a}_n^{(\iota)}(z)]\right\}^2 \\
&\quad - 2\hat{\gamma}_n^{(\iota)\top} \rho\left(\frac{Z_i - z}{h_n}\right) X_i^\top [a(Z_i) - \hat{a}_n^{(\iota)}(z)] + 2\varepsilon_i X_i^\top \{a(Z_i) - \hat{a}_n^{(\iota)}(z)\} \\
&\quad - 2\hat{\gamma}_n^{(\iota)\top} \rho\left(\frac{Z_i - z}{h_n}\right) \varepsilon_i \\
&= \varepsilon_i^2 + \left\{\hat{\gamma}_n^{(\iota)\top} \rho\left(\frac{Z_i - z}{h_n}\right)\right\}^2 + \left\{X_i^\top [a(z) + \mathrm{O}(h_n) - \hat{a}_n^{(\iota)}(z)]\right\}^2 \\
&\quad - 2\hat{\gamma}_n^{(\iota)\top} \rho\left(\frac{Z_i - z}{h_n}\right) X_i^\top [a(z) + \mathrm{O}(h_n) - \hat{a}_n^{(\iota)}(z)] \\
&\quad + 2\varepsilon_i X_i^\top \{a(z) + \mathrm{O}(h_n) - \hat{a}_n^{(\iota)}(z)\} - 2\hat{\gamma}_n^{(\iota)\top} \rho\left(\frac{Z_i - z}{h_n}\right) \varepsilon_i
\end{aligned}
$$

uniformly in $i \in \mathbb{N}$ by the Taylor expansion of $a(\cdot)$. To analyze each term of $\tilde{\Psi}_n^{(\iota)}(z)$ in

(11), let us first look at $\tilde{N}_n^{(\iota)}$. By Lemma 5 and (52), we have after substitution for $\hat{e}_{n,i}^{(\iota)2}$

$$
\begin{aligned}
\tilde{N}_n^{(\iota)} =&\frac{1}{n}\sum_{i=1}^n \hat{e}_{n,i}^{(\iota)2}\tilde{K}_h^{(\iota)}(Z_i - z)\\
=&\tilde{W}_{n,1}^{(\iota)} + \hat{\gamma}_n^{(\iota)\top}\tilde{S}_n^{(\iota)}\hat{\gamma}_n^{(\iota)} + [a(z) - \hat{a}_n^{(\iota)}(z)]^\top \tilde{W}_{n,2}^{(\iota)}[a(z) - \hat{a}_n^{(\iota)}(z)](1 + \mathrm{O}(h_n))\\
&+ \left(-2\hat{\gamma}_n^{(\iota)\top}\tilde{W}_{n,3}^{(\iota)} + 2\tilde{W}_{n,4}^{(\iota)\top}\right)[a(z) - \hat{a}_n^{(\iota)}(z)](1 + \mathrm{O}(h_n)) - 2\hat{\gamma}_n^{(\iota)\top}\tilde{W}_{n,5}\\
=&f_Z(z)\tilde{\mu}_0^{(\iota)}\sigma^2(z) + \mathrm{o_p}(1)f_Z(z)\tilde{M}^{(\iota)}\mathrm{o_p}(1) + \mathrm{o_p}(1)f_Z(z)\tilde{\mu}_0^{(\iota)}\Omega(z)\mathrm{o_p}(1)\\
&- 2\mathrm{o_p}(1)f_Z(z)\tilde{m}^{(\iota)}\mho(z)\mathrm{o_p}(1) + 2\mathrm{o_p}(1)\mathrm{o_p}(1) - 2\mathrm{o_p}(1)\mathrm{o_p}(1)\\
=&f_Z(z)\tilde{\mu}_0^{(\iota)}\sigma^2(z) + \mathrm{o_p}(1).
\end{aligned} \tag{53}
$$

Combining equations (52) and (53), Lemma 5(i), and Assumption D1 finally yields

$$
\begin{aligned}
\tilde{\Psi}_n^{(\iota)}(z) =&\mathrm{o_p}(1)\frac{\tilde{M}^{(\iota)-1}}{f_Z(z)}(1 + \mathrm{o_p}(1))\left(\frac{f_Z(z)\tilde{M}^{(\iota)} + \mathrm{o_p}(1)}{f_Z(z)\tilde{\mu}_0^{(\iota)}\sigma^2(z)}\right)\\
&\cdot \frac{\tilde{M}^{(\iota)-1}}{f_Z(z)}(1 + \mathrm{o_p}(1))\mathrm{o_p}(1)\\
=&\mathrm{o_p}(1).
\end{aligned}
$$

□

Proof of Theorem 8.

For any $z = s_q + \tau h_n$ with $\tau \in (-1, 1)$, let

$$
\acute{\hat{S}}_n^{(\iota)} = \frac{1}{n}\sum_{i:Z_i<s_q}\rho\left(\frac{Z_i - z}{h_n}\right)\rho^\top\left(\frac{Z_i - z}{h_n}\right)\tilde{K}_h^{(\iota)}(Z_i - z), \tag{54}
$$

$$
\grave{\hat{S}}_n^{(\iota)} = \frac{1}{n}\sum_{i:Z_i\geq s_q}\rho\left(\frac{Z_i - z}{h_n}\right)\rho^\top\left(\frac{Z_i - z}{h_n}\right)\tilde{K}_h^{(\iota)}(Z_i - z), \tag{55}
$$

$$
\acute{\hat{W}}_{n,1}^{(\iota)} = \frac{1}{n}\sum_{i:Z_i<s_q}\varepsilon_i^2\tilde{K}_h^{(\iota)}(Z_i - z), \quad \grave{\hat{W}}_{n,1}^{(\iota)} = \frac{1}{n}\sum_{i:Z_i\geq s_q}\varepsilon_i^2\tilde{K}_h^{(\iota)}(Z_i - z), \tag{56}
$$

54

$$\acute{\tilde{W}}_{n,2}^{(\iota)} = \frac{1}{n} \sum_{i:Z_i < s_q} X_i X_i^\top \tilde{K}_h^{(\iota)}(Z_i - z), \quad \grave{\tilde{W}}_{n,2}^{(\iota)} = \frac{1}{n} \sum_{i:Z_i \geq s_q} X_i X_i^\top \tilde{K}_h^{(\iota)}(Z_i - z), \tag{57}$$

$$\acute{\tilde{W}}_{n,3}^{(\iota)} = \frac{1}{n} \sum_{i:Z_i < s_q} \rho\left(\frac{Z_i - z}{h_n}\right) X_i^\top \tilde{K}_h^{(\iota)}(Z_i - z), \tag{58}$$

$$\grave{\tilde{W}}_{n,3}^{(\iota)} = \frac{1}{n} \sum_{i:Z_i \geq s_q} \rho\left(\frac{Z_i - z}{h_n}\right) X_i^\top \tilde{K}_h^{(\iota)}(Z_i - z), \tag{59}$$

$$\acute{\tilde{W}}_{n,4}^{(\iota)} = \frac{1}{n} \sum_{i:Z_i < s_q} X_i \varepsilon_i \tilde{K}_h^{(\iota)}(Z_i - z), \quad \grave{\tilde{W}}_{n,4}^{(\iota)} = \frac{1}{n} \sum_{i:Z_i \geq s_q} X_i \varepsilon_i \tilde{K}_h^{(\iota)}(Z_i - z), \tag{60}$$

$$\acute{\tilde{W}}_{n,5}^{(\iota)} = \frac{1}{n} \sum_{i:Z_i < s_q} \rho\left(\frac{Z_i - z}{h_n}\right) \varepsilon_i \tilde{K}_h^{(\iota)}(Z_i - z), \quad \text{and} \tag{61}$$

$$\grave{\tilde{W}}_{n,5}^{(\iota)} = \frac{1}{n} \sum_{i:Z_i \geq s_q} \rho\left(\frac{Z_i - z}{h_n}\right) \varepsilon_i \tilde{K}_h^{(\iota)}(Z_i - z). \tag{62}$$

Further, we define the population counterparts of the above kernel weighted averages:

$$\mho_-(s_q) = \lim_{z \uparrow s_q} \mathrm{E}[X^\top | Z = z], \quad \mho_+(s_q) = \lim_{z \downarrow s_q} \mathrm{E}[X^\top | Z = z] \tag{63}$$

$$\acute{\mu}_{0,\tau}^{(\iota)} = \int_{-1}^{-\tau} \tilde{K}(u)\mathrm{d}u, \quad \grave{\mu}_{0,\tau}^{(\iota)} = \int_{-\tau}^{1} \tilde{K}(u)\mathrm{d}u, \tag{64}$$

$$\acute{m}_\tau^{(\iota)} = \int_{-1}^{-\tau} \rho(u)\tilde{K}(u)\mathrm{d}u, \quad \grave{m}_\tau^{(\iota)} = \int_{-\tau}^{1} \rho(u)\tilde{K}(u)\mathrm{d}u, \tag{65}$$

$$\acute{M}_\tau^{(\iota)} = \int_{-1}^{-\tau} \rho(u)\rho^\top(u)\tilde{K}(u)\mathrm{d}u, \quad \text{and} \quad \grave{M}_\tau^{(\iota)} = \int_{-\tau}^{1} \rho(u)\rho^\top(u)\tilde{K}(u)\mathrm{d}u. \tag{66}$$

Again, we use decomposition $\tilde{T}_n^{(\iota)} = \tilde{W}_{n,5}^{(\iota)} + \tilde{T}_{n,2}^{(\iota)}$ as in (51). By the consistency results for $\acute{\tilde{W}}_{n,5}^{(\iota)}$ and $\grave{\tilde{W}}_{n,5}^{(\iota)}$ in Lemma 6(vi),

$$\tilde{W}_{n,5}^{(\iota)} = \acute{\tilde{W}}_{n,5}^{(\iota)} + \grave{\tilde{W}}_{n,5}^{(\iota)} = \mathrm{o_p}(1) + \mathrm{o_p}(1).$$

By (33)–(35) and the consistency results for $\acute{\tilde{W}}_{n,4}^{(\iota)}$ and $\grave{\tilde{W}}_{n,4}^{(\iota)}$ in Lemma 6(v), we obtain

$$
\begin{aligned}
\tilde{T}_{n,2}^{(\iota)}(z) =& \frac{1}{n} \sum_{i=1}^{n} \rho\left(\frac{Z_i - z}{h_n}\right) X_i^\top \{a(Z_i) - \hat{a}_n^{(\iota)}(z)\} \tilde{K}_h^{(\iota)}(Z_i - z) \\
=& \frac{1}{n} \sum_{i:Z_i < s_q} \rho\left(\frac{Z_i - z}{h_n}\right) X_i^\top [\underbrace{a(Z_i) - a_-(s_q)}_{\mathrm{O}(h_n)} - \Xi_{0,\tau}^{(\iota)} d_q] \tilde{K}_h^{(\iota)}(Z_i - z) \\
& + \frac{1}{n} \sum_{i:Z_i \geq s_q} \rho\left(\frac{Z_i - z}{h_n}\right) X_i^\top [\underbrace{a(Z_i) - a_-(s_q)}_{d_q + \mathrm{O}(h_n)} - \Xi_{0,\tau}^{(\iota)} d_q] \tilde{K}_h^{(\iota)}(Z_i - z) \\
=& - \acute{\tilde{W}}_{n,4}^{(\iota)}\left(\Xi_{0,\tau}^{(\iota)} d_q + \mathrm{O}(h_n)\right) - \grave{\tilde{W}}_{n,4}^{(\iota)}\left((\Xi_{0,\tau}^{(\iota)} - I_p)d_q + \mathrm{O}(h_n)\right) \\
=& - f_Z(s_q)\left(\acute{m}_\tau^{(\iota)} \mho_-(s_q)\Xi_{0,\tau}^{(\iota)} + \grave{m}_\tau^{(\iota)} \mho_+(s_q)(\Xi_{0,\tau}^{(\iota)} - I_p)\right) d_q + \mathrm{o}_\mathrm{p}(1).
\end{aligned}
$$

Hence, it follows from the consistency results for $\tilde{S}_n^{(\iota)} = \acute{\tilde{S}}_n^{(\iota)} + \grave{\tilde{S}}_n^{(\iota)}$ in Lemma 6(i) that

$$
\hat{\gamma}_n^{(\iota)} = \gamma^{(\iota)} + \mathrm{o}_\mathrm{p}(1), \tag{67}
$$

where

$$
\begin{aligned}
\gamma^{(\iota)} =& - \left(\acute{M}_\tau^{(\iota)} + \grave{M}_\tau^{(\iota)}\right)^{-1}\left(\acute{m}_\tau^{(\iota)} \mho_-(s_q)\Xi_{0,\tau}^{(\iota)} + \grave{m}_\tau^{(\iota)} \mho_+(s_q)(\Xi_{0,\tau}^{(\iota)} - I_p)\right) d_q \\
=& - \tilde{M}^{(\iota)-1}\left(\acute{m}_\tau^{(\iota)} \mho_-(s_q)\Xi_{0,\tau}^{(\iota)} + \grave{m}_\tau^{(\iota)} \mho_+(s_q)(\Xi_{0,\tau}^{(\iota)} - I_p)\right) d_q. \tag{68}
\end{aligned}
$$

Next, for $Z_i < s_q$ and $|Z_i - z| \leq h_n$, the squared error $\hat{e}_{n,i}^{(\iota)2}$ equals by the Taylor expansion of $a(\cdot)$ and the boundedness of its derivatives (Assumption A5)

$$
\begin{aligned}
\hat{e}_{n,i}^{(\iota)2} =& \varepsilon_i^2 + \left\{\hat{\gamma}_n^{(\iota)\top} \rho\left(\frac{Z_i - z}{h_n}\right)\right\}^2 + \left\{X_i^\top [a_-(s_q) + \mathrm{O}(h_n) - \hat{a}_n^{(\iota)}(z)]\right\}^2 \\
& - 2\hat{\gamma}_n^{(\iota)\top} \rho\left(\frac{Z_i - z}{h_n}\right) X_i^\top [a_-(z) + \mathrm{O}(h_n) - \hat{a}_n^{(\iota)}(z)] \\
& + 2\varepsilon_i X_i^\top \{a_-(s_q) + \mathrm{O}(h_n) - \hat{a}_n^{(\iota)}(z)\} - 2\hat{\gamma}_n^{(\iota)\top} \rho\left(\frac{Z_i - z}{h_n}\right) \varepsilon_i
\end{aligned}
$$

uniformly in $i \in \mathbb{N}$, and by the same argument for $Z_i \geq s_q$ and $|Z_i - z| \leq h_n$,

$$
\begin{aligned}
\hat{e}_{n,i}^{(\iota)2} = & \varepsilon_i^2 + \left\{ \hat{\gamma}_n^{(\iota)\top} \rho \left(\frac{Z_i - z}{h_n} \right) \right\}^2 + \left\{ X_i^\top [a_-(s_q) + d_q + \mathrm{O}(h_n) - \hat{a}_n^{(\iota)}(z)] \right\}^2 \\
& - 2\hat{\gamma}_n^{(\iota)\top} \rho \left(\frac{Z_i - z}{h_n} \right) X_i^\top [a_-(s_q) + d_q + \mathrm{O}(h_n) - \hat{a}_n^{(\iota)}(z)] \\
& + 2\varepsilon_i X_i^\top \left\{ a_-(s_q) + d_q + \mathrm{O}(h_n) - \hat{a}_n^{(\iota)}(z) \right\} - 2\hat{\gamma}_n^{(\iota)\top} \rho \left(\frac{Z_i - z}{h_n} \right) \varepsilon_i
\end{aligned}
$$

uniformly in $i \in \mathbb{N}$. For the term $\tilde{N}_n^{(\iota)}$ of $\tilde{\Psi}_n^{(\iota)}(z)$ in (11), it now follows after substituting the above expressions for $\hat{e}_{n,i}^{(\iota)2}$ and using equations (33)–(35) that

$$
\begin{aligned}
\tilde{N}_n^{(\iota)} = & \frac{1}{n} \sum_{i:Z_i < s_q} \hat{e}_{n,i}^{(\iota)2} \tilde{K}_h^{(\iota)}(Z_i - z) + \frac{1}{n} \sum_{i:Z_i \geq s_q} \hat{e}_{n,i}^{(\iota)2} \tilde{K}_h^{(\iota)}(Z_i - z) \\
= & \acute{W}_{n,1}^{(\iota)} + \grave{W}_{n,1}^{(\iota)} + \left[\hat{\gamma}_n^{(\iota)\top} \tilde{S}_n^{(\iota)} \hat{\gamma}_n^{(\iota)} + d_q^\top \Xi_{0,\tau}^{(\iota)\top} \acute{W}_{n,2}^{(\iota)} \Xi_{0,\tau}^{(\iota)} d_q \right. \\
& + d_q^\top (\Xi_{0,\tau}^{(\iota)} - I_p)^\top \grave{W}_{n,2}^{(\iota)} (\Xi_{0,\tau}^{(\iota)} - I_p) d_q + 2\hat{\gamma}_n^{(\iota)\top} \acute{W}_{n,3} \Xi_{0,\tau}^{(\iota)} d_q \\
& + 2\hat{\gamma}_n^{(\iota)\top} \grave{W}_{n,3} (\Xi_{0,\tau}^{(\iota)} - I_p) d_q - 2d_q^\top \Xi_{0,\tau}^{(\iota)\top} \acute{W}_{n,4}^{(\iota)} - 2d_q^\top (\Xi_{0,\tau}^{(\iota)} - I_p)^\top \grave{W}_{n,4}^{(\iota)} \right] \\
& \cdot (1 + \mathrm{O}(h_n)) - 2\hat{\gamma}_n^{(\iota)\top} \acute{W}_{n,5} - 2\hat{\gamma}_n^{(\iota)\top} \grave{W}_{n,5}.
\end{aligned}
$$

By (67) and Lemma 6, we thus have

$$
\tilde{N}_n^{(\iota)} = f_Z(s_q) \left\{ \acute{\tilde{\mu}}_{0,\tau}^{(\iota)} \sigma_-^2(s_q) + \grave{\tilde{\mu}}_{0,\tau}^{(\iota)} \sigma_+^2(s_q) + \sigma_{e,\tau}^{(\iota)2}(s_q) \right\} + \mathrm{o}_p(1), \tag{69}
$$

where

$$
\begin{aligned}
\sigma_{e,\tau}^{(\iota)2}(s_q) = & \gamma^{(\iota)\top} \tilde{M}^{(\iota)} \gamma^{(\iota)} + d_q^\top \Xi_{0,\tau}^{(\iota)\top} \acute{\tilde{\mu}}_{0,\tau}^{(\iota)} \Omega_-(s_q) \Xi_{0,\tau}^{(\iota)} d_q \\
& + d_q^\top (\Xi_{0,\tau}^{(\iota)} - I_p)^\top \grave{\tilde{\mu}}_{0,\tau}^{(\iota)} \Omega_+(s_q) (\Xi_{0,\tau}^{(\iota)} - I_p) d_q \\
& + 2\gamma^{(\iota)\top} \acute{m}_\tau^{(\iota)} \mho_-(s_q) \Xi_{0,\tau} d_q + 2\gamma^{(\iota)\top} \grave{m}_\tau^{(\iota)} \mho_+(s_q) (\Xi_{0,\tau}^{(\iota)} - I_p) d_q.
\end{aligned}
$$

Since the term above can be rewritten as

$$\sigma_{e,\tau}^{(\iota)2}(s_q) = \int_{-1}^{-\tau} \int (x^\top \Xi_{0,\tau} d_q + \gamma^{(\iota)\top} \rho(u))^2 \tilde{K}(u) \frac{f(x, s_q)}{f_Z(s_q)} \mathrm{d}x \mathrm{d}u$$
$$+ \int_{-\tau}^{1} \int (x^\top [\Xi_{0,\tau} - I_p] d_q + \gamma^{(\iota)\top} \rho(u))^2 \tilde{K}(u) \frac{f(x, s_q)}{f_Z(s_q)} \mathrm{d}x \mathrm{d}u, \tag{70}$$

it is clearly non-negative.

By equations (67)–(69) and Lemma 6(i), we conclude that

$$\tilde{\Psi}_n^{(\iota)}(z) = \gamma^{(\iota)\top} \tilde{C}_\tau^{(\iota)} \gamma^{(\iota)} + \mathrm{o}_\mathrm{p}(1),$$

where

$$\tilde{C}_\tau^{(\iota)} = \left(\frac{\tilde{M}^{(\iota)}}{\acute{\tilde{\mu}}_{0,\tau}^{(\iota)} \sigma_-^2(s_q) + \grave{\tilde{\mu}}_{0,\tau}^{(\iota)} \sigma_+^2(s_q) + \sigma_{e,\tau}^{(\iota)2}(s_q)} \right). \tag{71}$$

By the positive definiteness of $\tilde{M}^{(\iota)}$ (Assumption D1) and non-negative $\sigma_{e,\tau}^{(\iota)2}(s_q)$ from (70), we claim that $\tilde{C}_\tau^{(\iota)} \succ 0$ for any $\tau \in (-1, 1)$ and $\iota = c, l, r$. According to Assumption D2, some elements of $\gamma^{(\iota)}$, $\iota = c, l$, are non-zero for $\tau \in (0, 1)$. Hence, the limits of $\tilde{\Psi}_n^{(c)}(z)$ and $\tilde{\Psi}_n^{(l)}(z)$ are strictly positive, i.e., $\gamma^{(\iota)\top} \tilde{C}_\tau^{(\iota)} \gamma^{(\iota)} > 0$ for $\tau \in (0, 1)$ and $\iota = c, l$. For $\tau \in (0, 1)$ and $\iota = r$, we have $\acute{\tilde{\mu}}_{0,\tau}^{(\iota)} = 0$ and $\acute{m}_\tau^{(r)} = 0$. By the expressions of $\gamma^{(\iota)}$ in (68) and the fact (38), $\Xi_{0,\tau}^{(r)} = I_p$ for $\tau \in (0, 1)$, we conclude that $\gamma^{(r)} = 0$ and hence $\gamma^{(r)\top} \tilde{C}_\tau^{(r)} \gamma^{(r)} = 0$. Similarly for $\tau \in (-1, 0)$, we have $\gamma^{(c)\top} \tilde{C}_\tau^{(c)} \gamma^{(c)} \succ 0$, $\gamma^{(r)\top} \tilde{C}_\tau^{(r)} \gamma^{(r)} \succ 0$, and $\gamma^{(l)\top} \tilde{C}_\tau^{(l)} \gamma^{(l)} = 0$ due to equation (39), $\Xi_{0,\tau}^{(l)} = 0_p$. $\square$

Proof of Theorem 9.

Being based on the results of Theorems 7 and 8, it follows the same steps as in the proof of Theorem 5. $\square$

Proof of Theorem 10.

Being based on the results of Theorems 7 and 8, it follows the same steps as in the proof of Theorem 6. ☐

B   Some auxiliary lemmas

Lemma 1. *Suppose Assumptions A and B hold. For any $z \in D_{1n}^{(\iota)}$ and $\iota = c, l, r$, it holds as $n \to +\infty$ that*

(i) $S_{n,j}^{(\iota)} = \mu_j^{(\iota)} f_Z(z) \Omega(z) + o_p(1)$ *with* $j = 0, 1, 2, 3$,

(ii) $S_n^{(\iota)-1} = \dfrac{f_Z^{-1}(z)}{\mu_0^{(\iota)} \mu_2^{(\iota)} - \mu_1^{(\iota)2}} \begin{pmatrix} \mu_2^{(\iota)} & -\mu_1^{(\iota)} \\ -\mu_1^{(\iota)} & \mu_0^{(\iota)} \end{pmatrix} \otimes \Omega^{-1}(z)(1 + o_p(1))$,

(iii) $F_{n,j}^{(\iota)} = o_p(1)$ *with* $j = 0, 1$,

(iv) $K_n^{(\iota)} = \mu_0^{(\iota)} f_Z(z) + o_p(1)$,

(v) $N_{n,1}^{(\iota)} = \mu_0^{(\iota)} f_Z(z) \sigma^2(z) + o_p(1)$,

(vi) $\hat{a}_n^{(\iota)}(z) = a(z) + o_p(1)$,

(vii) $\hat{b}_n^{(\iota)}(z) = a'(z) + o_p(h_n^{-1})$,

where the above objects are defined in (15)–(17), (25), and (26).

Proof. By Assumptions A1–A3 and B1–B2, the conditions for the weak law of large number for kernel estimators in Hansen (2008) are satisfied. Applying Theorem 1 in Hansen (2008) leads to

$$S_{n,j}^{(\iota)} = \mathrm{E}[S_{n,j}^{(\iota)}] + o_p(1).$$

After a change of variable ($\dot{z} = z + vh_n$) and the Taylor expansion of the density f in which it partial derivatives with respect to Z are uniformly bounded due to Assumption A2, the expectation of $S_{n,j}^{(\iota)}$ equals

$$
\begin{aligned}
\mathrm{E}[S_{n,j}^{(\iota)}] =& \mathrm{E}\left[X_i X_i^\top \left(\frac{Z_i - z}{h_n}\right)^j K_h^{(\iota)}(Z_i - z)\right] \\
=& \frac{1}{h_n} \int\int \dot{x}\dot{x}^\top \left(\frac{\dot{z} - z}{h_n}\right)^j K^{(\iota)}\left(\frac{\dot{z} - z}{h_n}\right) f(\dot{x}, \dot{z}) \mathrm{d}\dot{z}\mathrm{d}\dot{x} \\
=& \int\int \dot{x}\dot{x}^\top v^j K^{(\iota)}(v) f(\dot{x}, z + vh_n)\mathrm{d}\dot{x}\mathrm{d}v \\
=& \int v^j K^{(\iota)}(v)\mathrm{d}v \cdot f_Z(z) \cdot \int \dot{x}\dot{x}^\top \frac{f(\dot{x}, z)}{f_Z(z)}\mathrm{d}\dot{x}\mathrm{d}v + \mathrm{O}(h_n) \\
=& \mu_j^{(\iota)} f_Z(z)\Omega(z) + \mathrm{O}(h_n),
\end{aligned}
$$

where $\Omega(z) = \mathrm{E}(XX^\top|Z = z)$. This concludes part (i). Part (ii) the follows trivially by part (i), Lemma 7(i): $\mu_0^{(\iota)}\mu_2^{(\iota)} - \mu_1^{(\iota)2} \neq 0$, the full rank conditions for $\Omega(z)$ in Assumption A4, and $f_Z(z) > 0$ in Assumption A2. Similarly to part (i), one can easily show (iii)–(v). Finally, using (21), parts (i)-(iii), and Assumption A5, we have

$$
\begin{aligned}
\left\|H_n(\hat{\beta}_n^{(\iota)} - \beta)\right\| \leq& \left\|S_n^{(\iota)-1}F_n^{(\iota)}\right\| + \left\|\frac{h_n^2}{2}S_n^{(\iota)-1}\begin{pmatrix}S_{n,2}^{(\iota)}\\S_{n,3}^{(\iota)}\end{pmatrix}a''(z)\right\| + \mathrm{o}(h_n^2) \\
=& \left\|\frac{\Omega^{-1}(z)}{f_Z(z)}\left(c_0^{(\iota)}F_{n,0}^{(\iota)} + c_1^{(\iota)}F_{n,1}^{(\iota)}\right)(1 + \mathrm{o}_\mathrm{p}(1))\right\| \\
&+ \left\|\frac{h_n^2}{2}\left(c_0^{(\iota)}\mu_2^{(\iota)} + c_1^{(\iota)}\mu_3^{(\iota)}\right)a''(z)(1 + \mathrm{o}_\mathrm{p}(1))\right\| + \mathrm{o}(h_n^2) \\
\leq& \mathrm{o}_\mathrm{p}(1) + \mathrm{O}_\mathrm{p}(h_n^2)\left\|a''(z)\right\| + \mathrm{o}(h_n^2) \\
=& \mathrm{o}_\mathrm{p}(1),
\end{aligned}
$$

where $c_0^{(\iota)}$ and $c_1^{(\iota)}$ are defined in (7). This completes the proofs of (vi) and (vii). $\qquad\square$

Lemma 2. *Under Assumptions A and B, it holds as* $n \to +\infty$ *that*

(i) $h_n \mathrm{var}(W_1^{(\iota)}) \to f_Z(z)\Theta(z)\left[c_0^{(\iota)2}\nu_0^{(\iota)} + 2c_0^{(\iota)}c_1^{(\iota)}\nu_1^{(\iota)} + c_1^{(\iota)2}\nu_2^{(\iota)}\right]$,

(ii) $h_n \sum_{j=1}^{n-1} |\mathrm{cov}(W_1^{(\iota)}, W_{j+1}^{(\iota)})| = o(1)$, *and*

(iii) $nh_n \mathrm{var}(U_n^{(\iota)}) \to f_Z(z)\Theta(z)\left[c_0^{(\iota)2}\nu_0^{(\iota)} + 2c_0^{(\iota)}c_1^{(\iota)}\nu_1^{(\iota)} + c_1^{(\iota)2}\nu_2^{(\iota)}\right]$,

where $U_n^{(\iota)}$ *and* $W_i^{(\iota)}$ *are given in* (23)–(24), $\Theta(z) = \mathrm{E}(XX^\top\sigma^2(X,Z)|Z = z)$, *and* $c_j^{(\iota)}$ *and* $\nu_j^{(\iota)}$ *are defined in equation* (7).

Proof. By conditioning on (X_1, Z_1), a change of variables, and the Taylor expansion,

$$
\begin{aligned}
h_n \mathrm{var}(W_1^{(\iota)}) =& h_n\mathrm{E}\left[X_1X_1^\top\sigma^2(X_1,Z_1)\left\{c_0^{(\iota)} + c_1^{(\iota)}\left(\frac{Z_1 - z}{h_n}\right)\right\}^2 K_h^{(\iota)2}(Z_1 - z)\right] \\
=& \int\int xx^\top\sigma^2(x, z + h_nu)\left(c_0^{(\iota)2} + 2c_0^{(\iota)}c_1^{(\iota)}u + c_1^{(\iota)2}u^2\right)K^{(\iota)2}(u) \\
& \times f(x, z + h_nu)\mathrm{d}u\mathrm{d}x \\
=& f_Z(z)\Theta(z)\left[c_0^{(\iota)2}\nu_0^{(\iota)} + 2c_0^{(\iota)}c_1^{(\iota)}\nu_1^{(\iota)} + c_1^{(\iota)2}\nu_2^{(\iota)}\right] + \mathrm{O}(h_n)
\end{aligned}
$$

due to Assumptions A2, A5, and A6. Since part (iii) follows trivially from (i) and (ii) by

$$
\begin{aligned}
nh_n\mathrm{var}(U_n^{(\iota)}) =& \frac{h_n}{n}\mathrm{var}\left(\sum_{i=1}^n W_i^{(\iota)}\right) \\
=& h_n\mathrm{var}(W_1^{(\iota)}) + 2h_n\sum_{j=1}^{n-1}\left(1 - \frac{j}{n}\right)\mathrm{cov}(W_1^{(\iota)}, W_{j+1}^{(\iota)}),
\end{aligned}
$$

it remains to prove (ii). To this end, let $c_n \to \infty$ be a sequence of positive integers such that $c_nh_n \to 0$. We write

$$
h_n\sum_{j=1}^{n-1}\left|\mathrm{cov}(W_1^{(\iota)}, W_{j+1}^{(\iota)})\right| = h_n\sum_{j=1}^{c_n}\left|\mathrm{cov}(W_1^{(\iota)}, W_{j+1}^{(\iota)})\right| + h_n\sum_{j=c_n}^{n-1}\left|\mathrm{cov}(W_1^{(\iota)}, W_{j+1}^{(\iota)})\right|
$$

$$
= J_{1,n} + J_{2,n}.
$$

61

We complete the proof by showing that $J_{1,n} = o(1)$ and $J_{2,n} = o(1)$.

First, for $j \leq c_n$, by conditioning on Z_1 and Z_{j+1} and Assumption A3(iii), we have,

$$|\text{cov}(W_1^{(\iota)}, W_{j+1}^{(\iota)})| \leq C_1 \text{E}\left(|X_1 X_{j+1}^\top \varepsilon_1 \varepsilon_{j+1}| K_h^{(\iota)}(Z_1 - z) K_h^{(\iota)}(Z_{j+1} - z)\right)$$

$$\leq C_2 \text{E}\left(|X_1 X_{j+1}^\top \varepsilon_1 \varepsilon_{j+1}| \big| Z_1 = z, Z_{j+1} = z\right) (f_{Z_1 Z_{j+1}}(z, z) + \text{O}(h_n))$$

$$\leq C_3,$$

for positive constants C_1, C_2, C_3, which implies that $J_{1,n} \leq h_n c_n C = o(1)$ by the choice of c_n. Next, let $W_{j,m}^{(\iota)}$ be the m-th element of $W_j^{(\iota)}$. Using Davydov's inequality (Fan and Yao, 2003, Proposition 2.5 with $p = q = \delta$), one has

$$|\text{cov}(W_{1,l}^{(\iota)}, W_{j+1,m}^{(\iota)})| \leq C\alpha^{1-2/\delta}(j) \left(\text{E}|W_{1,l}^{(\iota)}|^\delta\right)^{1/\delta} \left(\text{E}|W_{j+1,m}^{(\iota)}|^\delta\right)^{1/\delta}. \tag{72}$$

By conditioning on Z_1 and Assumptions A2 and A3(ii),

$$\text{E}|W_{1,l}^{(\iota)}|^\delta \leq C_1 \text{E}[|X_{1,l}\varepsilon_1|^\delta K_h^{(\iota)\delta}(Z_1 - z)]$$

$$\leq C_2 h_n^{1-\delta}\{\text{E}[|X_{1,l}\varepsilon_1|^\delta \big| Z_1 = z](f_Z(z) + \text{O}(h_n))\}$$

$$\leq C_3 h_n^{1-\delta}. \tag{73}$$

for some $C_1, C_2, C_3 > 0$. It follows from equations (72), (73), and Assumption A1 that

$$J_{2,n} = h_n \sum_{j=c_n+1}^{n-1} |\text{cov}(W_1^{(\iota)}, W_{j+1}^{(\iota)})|$$

$$\leq C_1 h_n h_n^{2(1-\delta)/\delta} \sum_{j=c_n+1}^{\infty} \alpha^{1-2/\delta}(j)$$

$$\leq C_2 h_n^{2/\delta-1} \sum_{j=c_n+1}^{\infty} j^{-(2-2/\delta)}$$

$$\leq C_3 h_n^{2/\delta-1} c_n^{2/\delta-1} = o(1),$$

where constants $C_1, C_2, C_3 > 0$ and the last inequality follows from the fact that

$$\sum_{j=k+1}^{\infty} j^{-\tau} \leq \int_k^{\infty} x^{-\tau} \mathrm{d}x = \frac{k^{1-\tau}}{\tau - 1}.$$

$\square$

Lemma 3. *Under Assumptions A, B, and C, we have for $n \to +\infty$ and $\iota = c, l, r$,*

$$\sup_{z \in D_{1n}^{(\iota)}} \left\| S_{n,j}^{(\iota)} - \mu_j^{(\iota)} f_Z(z) \Omega(z) \right\| = O_p\left(\sqrt{\frac{\ln n}{nh_n}} \right) + O(h_n) \text{ for } j = 0, 1, 2, 3,$$

$$\sup_{z \in D_{1n}^{(\iota)}} \left\| F_{n,j}^{(\iota)} \right\| = O_p\left(\sqrt{\frac{\ln n}{nh_n}} \right) \text{ for } j = 0, 1,$$

and

$$f_Z(z) S_n^{(\iota)-1} = \frac{\begin{pmatrix} \mu_2^{(\iota)} \Omega^{-1}(z) & -\mu_1^{(\iota)} \Omega^{-1}(z) \\ -\mu_1^{(\iota)} \Omega^{-1}(z) & \mu_0^{(\iota)} \Omega^{-1}(z) \end{pmatrix}}{\mu_0^{(\iota)} \mu_2^{(\iota)} - \mu_1^{(\iota)2}} \left\{ 1 + O_p\left(\sqrt{\frac{\ln n}{nh_n}} \right) + O(h_n) \right\}$$

uniformly for $z \in D_{1n}^{(\iota)}$.

Proof. By Assumptions A, B, and C, the conditions for weak uniform convergence result for kernel estimators over expanding sets in Hansen (2008) are satisfied. First, we consider case $\iota = c$, which uses both left and right neighborhoods. For the continuous region $D_{1n}^{(c)} = \bigcup_{q=0}^{Q}(s_q + h_n, s_{q+1} - h_n)$, we apply Theorem 2 in Hansen (2008) on each subregion $(s_q + h_n, s_{q+1} - h_n)$:

$$\sup_{z \in (s_q+h_n, s_{q+1}-h_n)} \left\| S_{n,j}^{(c)} - \mathrm{E}(S_{n,j}^{(c)}) \right\| = O_p\left(\sqrt{\frac{\ln n}{nh_n}} \right).$$

Notice the expanding sets considered in Hansen (2008) are allowed to grow to infinity

63

slowly, as $n \to \infty$, while the subregion $(s_q + h_n, s_{q+1} - h_n)$ expands to a bounded set (s_q, s_{q+1}). Taking the maximum over all subregions yields

$$\sup_{z \in D_{1n}^{(c)}} \left\| S_{n,j}^{(c)} - \mathrm{E}(S_{n,j}^{(c)}) \right\| \leq (Q+1) \cdot \max_q \sup_{z \in (s_q + h_n, s_{q+1} - h_n)} \left\| S_{n,j}^{(c)} - \mathrm{E}(S_{n,j}^{(c)}) \right\|$$
$$= \mathrm{O_p} \left(\sqrt{\frac{\ln n}{n h_n}} \right).$$

Since $\mathrm{E}(S_{n,j}^{(c)}) = \mu_j^{(c)} f_Z(z) \Omega(z) + \mathrm{O}(h_n)$, which is shown in the proof in Lemma 1, we have

$$\sup_{z \in D_{1n}^{(c)}} \left\| S_{n,j}^{(c)} - \mu_j^{(c)} f_Z(z) \Omega(z) \right\| = \mathrm{O_p} \left(\sqrt{\frac{\ln n}{n h_n}} \right) + \mathrm{O}(h_n).$$

Although Theorem 2 in Hansen (2008) originally excludes the case of one-sided kernel, his theorem is still applicable for one-sided kernel by taking 'one-sided' covering sets A_j, which boosts the size of covering by a constant multiplier 2^p, instead of 'two-sided' A_j in his proof. Then, by similar argument as for $S_{n,j}^{(c)}$, one can prove the uniform consistency results for $S_{n,j}^{(l)}$ and $S_{n,j}^{(r)}$.

Analogously, we can apply Theorem 2 in Hansen (2008) to $F_{n,j}^{(\iota)}$ with $\iota = c, l, r$, where the uniform convergence rates stays equal to $\mathrm{O_p}(\sqrt{\ln n / (n h_n)})$ since $\mathrm{E}(F_{n,j}^{(\iota)}) = 0$. $\square$

Lemma 4. *Suppose Assumptions A and B hold. For any $z = s_q + \tau h_n$ with $\tau \in (-1, 1)$ and $\iota = c, l, r$, we have as $n \to +\infty$,*

(i) *$\acute{S}_{n,j}^{(\iota)} = f_Z(s_q) \Omega_-(s_q) \acute{\mu}_{j,\tau}^{(\iota)} + \mathrm{o_p}(1)$ and $\grave{S}_{n,j}^{(\iota)} = f_Z(s_q) \Omega_+(s_q) \grave{\mu}_{j,\tau}^{(\iota)} + \mathrm{o_p}(1)$ for $j = 0, 1, 2$;*

(ii) *$\acute{F}_{n,j}^{(\iota)} = \grave{F}_{n,j}^{(\iota)} = \mathrm{o_p}(1)$ for $j = 0, 1$;*

(iii) *$K_n^{(\iota)} = f_Z(s_q) \mu_0^{(\iota)} + \mathrm{o_p}(1)$;*

64

(iv) *further, if the derivative of $\sigma^2(x,z)$ with respect to z is continuous and bounded on the complete D, $N_{n,1}^{(\iota)} = f_Z(s_q)\mu_0^{(\iota)}\sigma^2(s_q) + o_p(1)$,*

where the above terms are defined in (27)–(32).

Proof. After a change of variable and the Taylor expansion, we have

$$
\begin{aligned}
\mathrm{E}[\acute{S}_{n,j}^{(\iota)}] &= \mathrm{E}\left[X_i X_i^\top \left(\frac{Z_i - z}{h_n}\right)^j K_h^{(\iota)}(Z_i - z) \bigg| Z_i < s_q \right] \\
&= \int\int_{-1}^{-\tau} xx^\top u^j K^{(\iota)}(u) f(x, s_q + (\tau + u)h_n) \mathrm{d}u\mathrm{d}x \\
&= \int_{-1}^{-\tau} u^j K^{(\iota)}(u)\mathrm{d}u \cdot \lim_{z\uparrow s_q} f_Z(z) \int xx^\top \frac{f(x,z)}{f_Z(z)}\mathrm{d}x + \mathrm{O}(h_n) \\
&= \acute{\mu}_{j,\tau}^{(\iota)} f_Z(s_q)\Omega_-(s_q) + \mathrm{O}(h_n)
\end{aligned}
$$

due to Assumption A2. The convergence of $\acute{S}_{n,j}^{(\iota)}$ to its expectation follows again by applying Theorem 1 of Hansen (2008), which is allowed due to Assumptions A and B. The convergence results for $\grave{S}_{n,j}^{(\iota)}$ and (ii)–(iv) can be proven in a similar manner. $\qquad\square$

Lemma 5. *Suppose Assumptions A, B, and D1 hold. It holds for $n \to +\infty$ and $\iota = c, l, r$,*

(i) $\tilde{S}_n^{(\iota)} = f_Z(z)\tilde{M}^{(\iota)} \otimes \Omega(z) + o_p(1)$ *and* $\tilde{S}_n^{(\iota)-1} = \dfrac{\tilde{M}^{(\iota)-1}}{f_Z(z)}(1 + o_p(1))$;

(ii) $\tilde{W}_{n,1}^{(\iota)} = f_Z(z)\tilde{\mu}_0^{(\iota)}\sigma^2(z) + o_p(1)$;

(iii) $\tilde{W}_{n,2}^{(\iota)} = f_Z(z)\tilde{\mu}_0^{(\iota)}\Omega(z) + o_p(1)$;

(iv) $\tilde{W}_{n,3}^{(\iota)} = f_Z(z)\tilde{m}^{(\iota)} \otimes \Omega(z) + o_p(1)$;

(v) $\tilde{W}_{n,4}^{(\iota)} = o_p(1)$;

(vi) $\tilde{W}_{n,5}^{(\iota)} = o_p(1)$,

where the above terms are defined in (44)–(50).

65

Proof. This lemma is analogous to Lemma 1 and the results follow by direct applications of Theorem 1 in Hansen (2008). $\qquad\square$

Lemma 6. *Suppose Assumptions A, B, and D1 hold. For any $z = s_q + \tau h_n$ with $\tau \in (-1, 1)$ and $\iota = c, l, r$, we have as $n \to +\infty$,*

(i) $\acute{\tilde{S}}_n^{(\iota)} = f_Z(s_q)\acute{M}_\tau^{(\iota)} + \mathrm{o_p}(1)$ *and* $\grave{\tilde{S}}_n^{(\iota)} = f_Z(s_q)\grave{M}_\tau^{(\iota)} + \mathrm{o_p}(1)$;

(ii) $\acute{\tilde{W}}_{n,1}^{(\iota)} = f_Z(s_q)\acute{\tilde{\mu}}_{0,\tau}^{(\iota)}\sigma_-^2(s_q) + \mathrm{o_p}(1)$ *and* $\grave{\tilde{W}}_{n,1}^{(\iota)} = f_Z(s_q)\grave{\tilde{\mu}}_{0,\tau}^{(\iota)}\sigma_+^2(s_q) + \mathrm{o_p}(1)$;

(iii) $\acute{\tilde{W}}_{n,2}^{(\iota)} = f_Z(s_q)\acute{\tilde{\mu}}_{0,\tau}^{(\iota)}\Omega_-(s_q) + \mathrm{o_p}(1)$ *and* $\grave{\tilde{W}}_{n,2}^{(\iota)} = f_Z(s_q)\grave{\tilde{\mu}}_{0,\tau}^{(\iota)}\Omega_+(s_q) + \mathrm{o_p}(1)$;

(iv) $\acute{\tilde{W}}_{n,3}^{(\iota)} = f_Z(s_q)\acute{m}_\tau^{(\iota)}\mho_-(s_q) + \mathrm{o_p}(1)$ *and* $\grave{\tilde{W}}_{n,3}^{(\iota)} = f_Z(s_q)\grave{m}_\tau^{(\iota)}\mho_+(s_q) + \mathrm{o_p}(1)$;

(v) $\acute{\tilde{W}}_{n,4}^{(\iota)} = \grave{\tilde{W}}_{n,4}^{(\iota)} = \mathrm{o_p}(1)$;

(vi) $\acute{\tilde{W}}_{n,5}^{(\iota)} = \grave{\tilde{W}}_{n,5} = \mathrm{o_p}(1)$,

where the above terms are defined in (54)–(66).

Proof. This lemma is similar to Lemma 4. The results follow mainly by applying Theorem 1 in Hansen (2008). $\qquad\square$

Lemma 7. *Under Assumption B1, we have*

(i) $\mu_0^{(\iota)}\mu_2^{(\iota)} - \mu_1^{(\iota)2} > 0, \quad \iota = c, l, r$;

(ii)
$$\acute{\mu}_{0,\tau}^{(\iota)}\acute{\mu}_{2,\tau}^{(\iota)} - \acute{\mu}_{1,\tau}^{(\iota)2} \begin{cases} > 0, & \text{if } \iota = c \text{ and } \tau \in (-1,1), \\ > 0, & \text{if } \iota = l \text{ and } \tau \in (-1,1), \\ > 0, & \text{if } \iota = r \text{ and } \tau \in (-1,0), \\ = 0, & \text{if } \iota = r \text{ and } \tau \in [0,1); \end{cases}$$

(iii)

$$\acute{\mu}_{0,\tau}^{(\iota)}\acute{\mu}_{2,\tau}^{(\iota)} - \acute{\mu}_{1,\tau}^{(\iota)2} \quad \begin{cases} > 0, & \textit{if } \iota = c \textit{ and } \tau \in (-1,1), \\ = 0, & \textit{if } \iota = l \textit{ and } \tau \in (-1,0], \\ > 0, & \textit{if } \iota = l \textit{ and } \tau \in (0,1), \\ > 0, & \textit{if } \iota = r \textit{ and } \tau \in (-1,1), \end{cases}$$

Proof. Here, we prove part (ii) only and (i) and (iii) can be shown analogically. Suppose that U has a density $K^{(\iota)}(\cdot)$. We have

$$\begin{aligned} \mathrm{var}(U|U < -\tau) =& \mathrm{E}[\{U - \mathrm{E}(U|U < -\tau)\}^2 | U < -\tau] \\ =& \mathrm{E}(U^2|U < -\tau) - \{\mathrm{E}(U|U < -\tau)\}^2 \\ =& \int_{-1}^{-\tau} u^2 \frac{K^{(\iota)}(u)}{\int_{-1}^{-\tau} K^{(\iota)}(u)\mathrm{d}u}\mathrm{d}u - \left\{\int_{-1}^{-\tau} u \frac{K^{(\iota)}(u)}{\int_{-1}^{-\tau} K^{(\iota)}(u)\mathrm{d}u}\mathrm{d}u\right\}^2 \\ =& \frac{\int_{-1}^{-\tau} u^2 K^{(\iota)}(u)\mathrm{d}u}{\int_{-1}^{-\tau} K^{(\iota)}(u)\mathrm{d}u} - \left\{\frac{\int_{-1}^{-\tau} u K^{(\iota)}(u)\mathrm{d}u}{\int_{-1}^{-\tau} K^{(\iota)}(u)\mathrm{d}u}\right\}^2 \\ =& \frac{\acute{\mu}_{2,\tau}^{(\iota)}}{\acute{\mu}_{0,\tau}^{(\iota)}} - \frac{\acute{\mu}_{1,\tau}^{(\iota)2}}{\acute{\mu}_{0,\tau}^{(\iota)2}}. \end{aligned}$$

By Assumption B1 and definitions of $K^{(r)}(\cdot)$ and $K^{(l)}(\cdot)$ in (2),

$$\acute{\mu}_{0,\tau}^{(\iota)}\acute{\mu}_{2,\tau}^{(\iota)} - \acute{\mu}_{1,\tau}^{(\iota)2} = \acute{\mu}_{0,\tau}^{(\iota)2}\mathrm{var}(U|U < -\tau) \quad \begin{cases} > 0, & \text{if } \iota = c \text{ and } \tau \in (-1,1), \\ > 0, & \text{if } \iota = l \text{ and } \tau \in (-1,1), \\ > 0, & \text{if } \iota = r \text{ and } \tau \in (-1,0), \\ = 0, & \text{if } \iota = r \text{ and } \tau \in [0,1). \end{cases}$$

□

Lemma 8. *Let X be a symmetric matrix given by*

$$X = \begin{pmatrix} A & B^\top \\ B & C \end{pmatrix}.$$

Then

(i) *X is positive definite if and only if A and the Schur complement of A, $C - BA^{-1}B^\top$, are both positive definite.*

(ii)
$$X^{-1} \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} I_p \\ 0_p \end{pmatrix},$$

where I_p is the $p \times p$ identity matrix and 0_p is the null matrix of size $p \times p$, if A, B and C are $p \times p$ matrices.

Proof. Part (i) is one of the fundamental results of Schur complement, where the proof can be found in Zhang (2005, Theorem 1.12). For part (ii), since $X^{-1}X = I_{2p}$, we have

$$X^{-1}X \begin{pmatrix} I_p \\ 0_p \end{pmatrix} = I_{2p} \begin{pmatrix} I_p \\ 0_p \end{pmatrix}$$

$$\Leftrightarrow X^{-1} \begin{pmatrix} A & B^\top \\ B & C \end{pmatrix} \begin{pmatrix} I_p \\ 0_p \end{pmatrix} = \begin{pmatrix} I_p & 0_p \\ 0_p & I_p \end{pmatrix} \begin{pmatrix} I_p \\ 0_p \end{pmatrix}$$

$$\Leftrightarrow X^{-1} \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} I_p \\ 0_p \end{pmatrix}.$$

$\square$

Lemma 9. *Under Assumptions B1 and A4,*

(i) *the variance matrix*

$$\acute{\Omega}^{(\iota)}_{-,\tau}(s_q) = \begin{pmatrix} \acute{\mu}^{(\iota)}_{0,\tau}\Omega_-(s_q) & \acute{\mu}^{(\iota)}_{1,\tau}\Omega_-(s_q) \\ \acute{\mu}^{(\iota)}_{1,\tau}\Omega_-(s_q) & \acute{\mu}^{(\iota)}_{2,\tau}\Omega_-(s_q) \end{pmatrix}$$

is

$$\begin{cases} positive\ definite, & if\ \iota = c\ and\ \tau \in (-1,1), \\[2mm] positive\ definite, & if\ \iota = l\ and\ \tau \in (-1,1), \\[2mm] positive\ definite, & if\ \iota = r\ and\ \tau \in (-1,0), \\[2mm] a\ null\ matrix, & if\ \iota = r\ and\ \tau \in [0,1); \end{cases}$$

(ii) *the variance matrix*

$$\grave{\Omega}^{(\iota)}_{+,\tau}(s_q) = \begin{pmatrix} \grave{\mu}^{(\iota)}_{0,\tau}\Omega_+(s_q) & \grave{\mu}^{(\iota)}_{1,\tau}\Omega_+(s_q) \\ \grave{\mu}^{(\iota)}_{1,\tau}\Omega_+(s_q) & \grave{\mu}^{(\iota)}_{2,\tau}\Omega_+(s_q) \end{pmatrix}$$

is

$$\begin{cases} positive\ definite, & if\ \iota = c\ and\ \tau \in (-1,1), \\[2mm] a\ null\ matrix, & if\ \iota = l\ and\ \tau \in (-1,0], \\[2mm] positive\ definite, & if\ \iota = l\ and\ \tau \in (0,1), \\[2mm] positive\ definite, & if\ \iota = r\ and\ \tau \in (-1,1); \end{cases}$$

(iii) *for $\tau \in (-1,1)$ and $\iota = c,l,r$, the variance matrix $\acute{\Omega}^{(\iota)}_{-,\tau}(s_q) + \grave{\Omega}^{(\iota)}_{+,\tau}(s_q)$ is positive definite.*

Proof. By Assumptions B1 and A4, $\acute{\mu}^{(\iota)}_{0,\tau}\Omega_-(s_q)$ is positive definite except for $\iota = r$ and

$\tau \in [0, 1)$ when it equals the null matrix. Also, the Schur complement of $\acute{\mu}_{0,\tau}^{(\iota)} \Omega_-(s_q)$ is

$$\acute{\mu}_{2,\tau}^{(\iota)} \Omega_-(s_q) - \acute{\mu}_{1,\tau}^{(\iota)} \Omega_-(s_q) \acute{\mu}_{0,\tau}^{(\iota)-1} \Omega_-^{-1}(s_q) \acute{\mu}_{1,\tau}^{(\iota)} \Omega_-(s_q) = \left(\acute{\mu}_{2,\tau}^{(\iota)} - \frac{\acute{\mu}_{1,\tau}^{(\iota)2}}{\acute{\mu}_{0,\tau}^{(\iota)}} \right) \Omega_-(s_q),$$

which is also positive definite by Lemma 7 and Assumption A4 except for the case of $\iota = r$ and $\tau \in [0, 1)$ when it equals the null matrix. After applying Lemma 8(i), the proof of part (i) is complete. Similarly, one can prove (ii). The claim (iii) then follows immediately from (i) and (ii). □

Lemma 10. *Under Assumptions B1 and A4, for $\iota = l, r, c$,*

(i) rank $\left(\Xi_{0,\tau}^{(\iota)} \right) = p$, *if $\grave{\mu}_{0,\tau}^{(\iota)} > 0$;*

(ii) rank $\left(\Xi_{0,\tau}^{(\iota)} - I_p \right) = p$, *if $\acute{\mu}_{0,\tau}^{(\iota)} > 0$,*

where the matrix $\Xi_{0,\tau}^{(\iota)}$ is defined in (36).

Proof. Using Lemma 9(iii) and the properties of a positive definite matrix, matrix $\acute{\Omega}_{-,\tau}^{(\iota)}(s_q) + \grave{\Omega}_{+,\tau}^{(\iota)}(s_q)$ is non-singular and its inverse is also positive definite. By Lemma 9(ii) and the fact that $AB \succ 0$ if $A \succ 0$ and $B \succ 0$, the matrix

$$\Xi_\tau^{(\iota)} = \left[\acute{\Omega}_{-,\tau}^{(\iota)}(s_q) + \grave{\Omega}_{+,\tau}^{(\iota)}(s_q) \right]^{-1} \grave{\Omega}_{+,\tau}^{(\iota)}(s_q) \succ 0$$

if $\grave{\mu}_{0,\tau}^{(\iota)} > 0$. Since

$$
\begin{aligned}
\Xi_{0,\tau}^{(\iota)} &= [I_p \ \ 0_p] \begin{pmatrix} \Xi_{0,\tau}^{(\iota)} \\ \Xi_{1,\tau}^{(\iota)} \end{pmatrix} \\
&= [I_p \ \ 0_p] \left[\acute{\Omega}_{-,\tau}^{(\iota)}(s_q) + \grave{\Omega}_{+,\tau}^{(\iota)}(s_q) \right]^{-1} \begin{pmatrix} \grave{\mu}_{0,\tau}^{(\iota)} \Omega_+(s_q) \\ \grave{\mu}_{1,\tau}^{(\iota)} \Omega_+(s_q) \end{pmatrix} \\
&= [I_p \ \ 0_p] \left[\acute{\Omega}_{-,\tau}^{(\iota)}(s_q) + \grave{\Omega}_{+,\tau}^{(\iota)}(s_q) \right]^{-1} \grave{\Omega}_{+,\tau}^{(\iota)}(s_q) \begin{pmatrix} I_p \\ 0_p \end{pmatrix}
\end{aligned}
$$

and the property of positive definite matrix that $A^\top B A \succ 0$ if $B \succ 0$ and A has full column rank, we conclude that $\Xi_{0,\tau}^{(\iota)} \succ 0$. Hence $\Xi_{0,\tau}^{(\iota)}$ has full rank, i.e., $\text{rank}\left(\Xi_{0,\tau}^{(\iota)}\right) = p$, which completes the proof of (i).

To show (ii), we write

$$
\begin{aligned}
I_p - \Xi_{0,\tau}^{(\iota)} &= [I_p \ \ 0_p] \left\{ I_{2p} - \left[\acute{\Omega}_{-,\tau}^{(\iota)}(s_q) + \grave{\Omega}_{+,\tau}^{(\iota)}(s_q) \right]^{-1} \grave{\Omega}_{+,\tau}^{(\iota)}(s_q) \right\} \begin{pmatrix} I_p \\ 0_p \end{pmatrix} \\
&= [I_p \ \ 0_p] \left[\acute{\Omega}_{-,\tau}^{(\iota)}(s_q) + \grave{\Omega}_{+,\tau}^{(\iota)}(s_q) \right]^{-1} \left\{ \acute{\Omega}_{-,\tau}^{(\iota)}(s_q) + \grave{\Omega}_{+,\tau}^{(\iota)}(s_q) - \grave{\Omega}_{+,\tau}^{(\iota)}(s_q) \right\} \begin{pmatrix} I_p \\ 0_p \end{pmatrix} \\
&= [I_p \ \ 0_p] \left[\acute{\Omega}_{-,\tau}^{(\iota)}(s_q) + \grave{\Omega}_{+,\tau}^{(\iota)}(s_q) \right]^{-1} \acute{\Omega}_{-,\tau}^{(\iota)}(s_q) \begin{pmatrix} I_p \\ 0_p \end{pmatrix}.
\end{aligned}
$$

By similar arguments as in part (i) and Lemmas 9(i) and 9(iii), it follows that $I_p - \Xi_{0,\tau}^{(\iota)} \succ 0$. As a result, $I_p - \Xi_{0,\tau}^{(\iota)}$ has the full rank just as matrix $\Xi_{0,\tau}^{(\iota)} - I_p$. $\qquad\square$

References

Ahmad, I., S. Leelahanon, and Q. Li (2005). Efficient estimation of a semiparametric partially linear varying coefficient model. *The Annals of Statistics 33*, 258–283.

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics 20*(1), 105 – 134.

Cai, Z., J. Fan, and R. Li (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association 95*(451), 888–902.

Cai, Z., J. Fan, and Q. Yao (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association 95*(451), 941–956.

Cai, Z. and Q. Li (2008). Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory 24*(5), 1321–1342.

Casas, I. and I. Gijbels (2012). Unstable volatility: the break-preserving local linear estimator. *Journal of Nonparametric Statistics 24*(4), 883–904.

Fan, J. and T. Huang (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli 11*(6), 1031–1057.

Fan, J. and Q. Yao (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika 85*(3), pp. 645–660.

Fan, J. and Q. Yao (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer Series in Statistics. New York: Springer-Verlag.

Fan, J., Q. Yao, and Z. Cai (2003). Adaptive varying-coefficient linear models. *Journal of the Royal Statistical Society Series B 65*(1), 57–80.

Fan, J. and J. Zhang (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society Series B 62*(2), 303–322.

Gijbels, I. and A.-C. Goderniaux (2004). Bandwidth selection for change point estimation in nonparametric regression. *Technometrics 46*, 76–86.

Gijbels, I., A. Lambert, and P. Qiu (2007). Jump-preserving regression and smoothing using local linear fitting: A compromise. *Annals of the Institute of Statistical Mathematics 59*(2), 235–272.

Godtliebsen, F., E. Spjotvoll, and J. S. Marron (1997). A nonlinear gaussian filter applied to images with discontinuities. *Journal of Nonparametric Statistics 8*(1), 21–43.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory 24*(3), 726–748.

Hastie, T. J. and R. J. Tibshirani (1993). Varying-coefficient models. *Journal of the Royal Statistical Society Series B 55*, 757–796.

Hoover, D. R., J. A. Rice, C. O. Wu, and L.-P. Yang (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika 85*, 809–822.

Huang, J. Z. and H. Shen (2004). Functional coefficient regression models for non-linear time series: A polynomial spline approach. *Scandinavian Journal of Statistics 31*(4), 515–534.

Kang, K.-H., J.-Y. Koo, and C.-W. Park (2000). Kernel estimation of discontinuous regression functions. *Statistics & Probability Letters 47*, 277–285.

Lee, E. R. and E. Mammen (2016). Local linear smoothing for sparse high dimensional varying coefficient models. *Electronic Journal of Statistics 10*(1), 855–894.

Müller, H.-G. (1992). Change-points in nonparametric regression analysis. *The Annals of Statistics 20*, 737–761.

Patton, A., D. N. Politis, and H. White (2009). Correction to "automatic block-length selection for the dependent bootstrap". *Econometric Reviews 28*, 372–375.

Polzehl, J. and V. G. Spokoiny (2000). Adaptive weights smoothing with applications to image restoration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62*(2), 335–354.

Spokoiny, V. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *Ann. Stat. 26*(4), 1356–1378.

Sun, Y., R. J. Carroll, and D. Li (2009). Semiparametric estimation of fixed-effects panel data varying coefficient models. In Q. Li and J. S. Racine (Eds.), *Nonparametric Econometric Methods*, Volume 25 of *Advances in Econometrics*, pp. 101–129. Emerald Group Publishing Limited.

Čížek, P. and C. Koo (2015). Semiparametric transition models. Unpublished manuscript, Tilburg University.

Wu, C. O., C. T. Chiang, and D. R. Hoover (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American Statistical Association 93*, 1388–1402.

Wu, J. S. and C. K. Chu (1993). Kernel-type estimators of jump points and values of a regression function. *The Annals of Statistics 21*, 1545–1566.

Zhang, F. (2005). *The Schur Complement and Its Applications*, Volume 4 of *Numerical Methods and Algorithms*. Springer.

Zhang, W., S.-Y. Lee, and X. Song (2002). Local polynomial fitting in semivarying coefficient model. *Journal of Multivariate Analysis 82*(1), 166–188.

Zhao, Y.-Y., J.-G. Lin, X.-F. Huang, and H.-X. Wang (2016). Adaptive jump-preserving estimates in varying-coefficient models. *Journal of Multivariate Analysis 149*, 65–80.

Zhu, H., J. Fan, and L. Kong (2014). Spatially varying coefficient model for neuroimaging data with jump discontinuities. *Journal of the American Statistical Association 109*, 1084–1098.