Wavelet Estimators for LATE under Discontinuous (Kink) Incentive Assignment Mechanism*

Heng Chen†
Department of Economics
Vanderbilt University

July 2011

Abstract

In an effort to improve the asymptotic properties of the jump size estimators, a number of alternatives have been suggested. These include partial spline estimator, Speckman estimator, kernel estimator, partial linear estimator and the recent wavelet OLS estimators constructed by Chen and Fan (2011). We show that all these estimators share a common structure, being members of a class of local least squares wavelet estimators. We use this structure to compare their asymptotic bias and mean square error. We find that the local least squares wavelet estimator attains the optimal convergence rate for semiparametric estimation of the local average treatment effect (LATE) under a broader set of conditions. We also find that the local least squares wavelet estimator has the better finite sample performance and could jointly estimate other order derivative discontinuous jump sizes. These features are illustrated by a comprehensive Monte Carlo simulations. The scale selection, discontinuous order specification and estimation for the unknown discontinuous location are also discussed.

Keywords: Nonparametric Methods; Wavelet Estimation for Jump and Kink sizes; Incentive Assignment Mechanism; Local Average Treatment Effect;

JEL codes: C13; C14; C21

---

*I owe my gratitude to my advisor, Yanqin Fan, for her guidance, constant support and encouragement. I am grateful to Tong Li, Mototsugu Shintani and Bryan Shepherd for their insightful comments and suggestions. Any mistakes and errors are of course my own.

†Email: heng.chen@vanderbilt.edu. Comments are welcome.
1 Introduction

Estimating various treatment effects under the switching regime model has recently become an interest for empirical work in economics. Particularly for the local average treatment effect under the discontinuous incentive assignment mechanism, there are numerous semiparametric and nonparametric estimators available in the fields of econometrics and statistics. The main purpose of this paper is to propose a new and intuitive estimation approach, local least squares wavelet estimator, which is optimal and robust under a broader set of conditions. Moreover this local least squares wavelet estimator could be easily adapted to the kink incentive assignment mechanism.

Heuristically the reason of constructing the proposed estimator from the wavelet theory is that the wavelet coefficient of a deterministic function at a given location characterizes its pointwise smoothness, where a small (large) wavelet coefficient corresponds to a high (low) smoothness of the function at that point (See Appendix A for the rigorous statement). Because of this pointwise smoothness characterization, wavelet coefficients have been commonly used to detect the discontinuous location in (1995) and Raimondo (1998). For estimating the jump size Park and Kim (2006) presented a simple wavelet estimator using only one pair wavelet coefficient, while Chen and Fan (2011) contained all pairs of wavelet coefficients around the discontinuous location in their wavelet OLS estimators and showed the estimators had a better asymptotic property. However the wavelet OLS estimators failed to take any further potential discontinuous jumps into consideration, and this omittance turned out to be at loss of the optimal convergence rate under a broader set of conditions. The local least squares wavelet estimator in this paper combines the strength of using more wavelet coefficients with an explicit structure for potential higher order discontinuous jumps, so that the asymptotic property of the estimator is better in general.

Besides the wavelet estimators, other alternatives have been suggested in an effort to improve the asymptotic properties of the jump size estimators. These include partial spline estimator by Eubank and Whitney (1989) and Speckman estimator by Eubank and Speckman (1994); kernel estimators (Nadaraya-Watson and local polynomial kernel regression estimator) summarized by Porter (2003) and Robinson (1988)-based partial linear estimator proposed by Porter (2003). We show that all these estimators share a common structure, being members of a class of local least squares wavelet estimators. In specific we could reformulate aforementioned estimators either in terms of the equivalent wavelet estimators, or their objective functions minimizing the local least squares function based on the sample wavelet coefficients. We find that the local least squares wavelet estimator attains the optimal convergence rate for semiparametric estimation of the LATE.

In the finite sample performance, the local least squares wavelet estimator is superior to the current favorite local polynomial kernel regression approach due to the finite unconditional variance. It is well-known in the statistics literature that local polynomial kernel regression estimators suffer from a serious drawback: for compactly supported kernels, the unconditional variance of a local polynomial kernel estimator is infinite so that the MSE (mean square error) and the MSE optimal bandwidth are not defined, see Seifert and Gasser
In order to overcome such the defect, our estimator equips spaces (Hall, et al., 1998) the original data before running the regression. Not only such transformation could guarantee the finite unconditional variance for the local least squares wavelet estimator, but also would not degrade features of the function through prior smoothing, especially when there are jumps. Through the asymptotic derivations, we find the local least squares wavelet jump size estimator adapts to both random and fixed designs, to both highly clustered and nearly uniform design; thus this design-adaptiveness could induce an even better finite sample performance.

An interesting by-product of the local least squares wavelet estimator is that it could jointly estimate any potential order discontinuous jump sizes at the discontinuous location. In the case of the discontinuous incentive assignment mechanism, the proposed estimator could estimate the jump size (our primary interest), the kink size (the first derivative discontinuous jump size), the second derivative discontinuous jump size and up to the $p$-th derivative discontinuous jump size altogether. Similar results apply to the kink incentive assignment mechanism. Furthermore this joint estimates of all the order discontinuous jump sizes are informative in prior deciding whether we are falling into the discontinuous incentive assignment mechanism or the kink incentive assignment mechanism, or both.

Therefore without assuming the knowledge of the underlying incentive assignment mechanism, the success of the LATE identification hinges on the lowest discontinuous order at the discontinuous location. In Section 5, we would discuss this issue by LASSO (least absolute shrinkage and selection operator) under the wavelet-transformed linear model.

The outline of the paper is as follows. In Section 2 the local least squares wavelet estimator under the discontinuous incentive assignment mechanism is described. Section 3 summarize the wavelet-based jump size estimators and obtains other existing estimator as special cases of the local least squares wavelet estimator. Section 4 provides the asymptotic results under the kink incentive assignment mechanism. Section 5 discusses the scale selection, discontinuous order specification, and estimation for the unknown discontinuous location. The proposed methods are examined in Section 6 through Monte Carlo simulations. Section 7 concludes. Proofs of the results are given in Appendix.

## 2 Wavelet Estimators under the Discontinuous Incentive Assignment Mechanism

In this section, we first review the LATE identification results and place them into two auxiliary regressions for estimation. Next we provide our local least squares wavelet estimator and establish its asymptotic property.

Let $(\Omega, \mathcal{F}, P)$ denote a probability space. To simplify technical arguments, we assume the random variables $V \in \mathcal{V} \subset \mathcal{R}$, $U \in \mathcal{U} \subset \mathcal{R}$, and $W \in \mathcal{W} \subset \mathcal{R}^d$ are continuous random variables/vectors defined on $(\Omega, \mathcal{F}, P)$ and that the distributions of $W$, $V$, $U$ are absolutely continuous with respect to the Lebesgue measure with pdfs $f_W(w)$, $w \in \mathcal{W}$, $f_V(v)$, $v \in \mathcal{V}$, $f_U(u)$, $u \in \mathcal{U}$. Throughout the rest of this paper, we adopt the

---

1 In Dong (2010), she allowed for the co-existence of the discontinuous and kink incentive assignment mechanisms which led to a “weighted” identification result for LATE.
following notation: $\int dy = \int_H dy$, $\int dw = \int_W dw$, and $\int dv = \int_Y dv$. Let $[A_{i,j}]_{m \times n}$ denote a $(i,j)$-th element in a $m$ by $n$ matrix. And $x^n$ means that $x$ is to the power $a$, while $x^{(a)}$ is the $a$-th order derivative of $x$. At last the jump size is meant to be $\lim_{v \downarrow v_0} E(\cdot | V = v) - \lim_{v \uparrow v_0} E(\cdot | V = v)$; the first derivative jump size is meant to be $\lim_{v \downarrow v_0} dE(\cdot | V = v)/dv - \lim_{v \uparrow v_0} dE(\cdot | V = v)/dv$, which is also called the kink size interchangeably; and the $l$-th derivative jump size is defined as $\lim_{v \downarrow v_0} d^l E(\cdot | V = v)/dv^l - \lim_{v \uparrow v_0} d^l E(\cdot | V = v)/dv^l$ for any $l \geq 2$.

## 2.1 LATE Identification and Auxiliary Regressions

In Hahn, et al. (2001) seminal paper, they identified the LATE under RDD (regression discontinuity model). Built upon the switching regime model without self-selection into the treatment, Lee (2008) established the LATE identification result under the discontinuous incentive assignment mechanism, while Card, et al. (2009) obtained the LATE under the kink incentive assignment mechanism. Henceforth Chen and Fan (2011) incorporated both the incentive assignment mechanism and the individual treatment selection mechanism into the switching model, that is, allowing for heterogenous choices among agents assigned the same incentive, to identify the LATE.

Consider the switching regime model in Chen and Fan (2011):

\begin{align}
Y_1 &= g_1 (V, W), \quad Y_0 = g_0 (V, W), \quad \text{(1)}
\end{align}

\begin{align}
D &= I\{b(V) - U \geq 0\}, \quad \text{(2)}
\end{align}

where $U$ is the individual’s unobservable covariate affecting selection, $W$ is a vector of individual’s unobservable covariates affecting potential outcomes, and $g_1$, $g_0$ are unknown real-valued measurable functions. The agent’s observable covariate $V$ affects both the potential outcomes and selection through the incentive assignment mechanism $b(\cdot)$. The incentive assignment mechanism $b$ (could be unknown) is assumed to be either discontinuous at a known cut-off point $v_0$ (discontinuous incentive assignment mechanism) or continuous but non-differentiable at $v_0$ (kink incentive assignment mechanism). When assuming the treatment selection mechanism is the same as the incentive assignment mechanism, the model in Lee (2008) is a special case of (1) and (2) in which the binary treatment $D = b(V) = I\{V \geq v_0\}$, and Card, et al. (2009) considered a kink incentive assignment mechanism under the continuous treatment $D = b(V)$. Thus by allowing agent’s treatment selection to depend on her unobservable covariate $U$ in (2), Chen and Fan (2011) general framework is consistent with the observations that agents being assigned the same incentive often choose different treatments.

Assume the econometrician observes $(V, Y, D)$ and the individual’s realized outcome $Y$ follows $Y \equiv DY_1 + (1 - D)Y_0$. Let $P(v) \equiv \Pr(D = 1|V = v) = E(D|V = v)$. Under the discontinuous incentive assignment mechanism, the LATE, $\lim_{v \downarrow v_0} E(Y_1 - Y_0 | V = v, D(v_0 + e) - D(v_0 - e) = 1)$, would be identified as

\begin{align}
\frac{\lim_{v \downarrow v_0} E(Y | V = v) - \lim_{v \uparrow v_0} E(Y | V = v)}{\lim_{v \downarrow v_0} P(v) - \lim_{v \uparrow v_0} P(v)}, \quad \text{(3)}
\end{align}
which is the ratio of the jump size in $E(Y|V)$ to the jump size in $P(v)$ at $v_0$. We establish auxiliary regressions linking the LATE to jump sizes $\delta_0$ and $\zeta_0$ in (4) and (5) for the discontinuous incentive assignment mechanism. Thus estimating the LATE in is the same as estimating $\delta_0/\zeta_0$ in two auxiliary regressions, which are

\begin{align}
Y &= h(V) + \delta_0 I\{V \geq v_0\} + \varepsilon, \\
D &= g(V) + \zeta_0 I\{V \geq v_0\} + \varepsilon, 
\end{align}

(4) \tag{4}

(5) \tag{5}

where $E(\varepsilon|V) = 0$, $E(\varepsilon|V) = 0$, $\delta_0 \equiv \lim_{v \uparrow v_0} E(Y|V = v) - \lim_{v \downarrow v_0} E(Y|V = v)$ and $\zeta_0 \equiv \lim_{v \uparrow v_0} P(v) - \lim_{v \downarrow v_0} P(v)$.

Assumption Set D is as follows to make sure the auxiliary regressions (4) and (5) are equivalent to the LATE under the discontinuous incentive assignment mechanism.

**Assumption Set D:**

**Assumption D1.** Assume (i) $f_{V|W}(v|w)$ is continuous and strictly positive on $V$ for every $w \in W$; (ii) $f_{V}(v)$ is continuous and strictly positive on $V$; (iii) $f_{V|W,U}(v|w,u)$ is continuous and strictly positive on $V$ for $u \in U$ and $w \in W$.

**Assumption D2.** Assume $g_1(v, w)$ and $g_0(v, w)$ are continuous on $V$ for every $w \in W$.

**Assumption D3.** For $j = 1, 0$, assume (i) $E|Y_j| < \infty$; (ii) $\int_{V} \sup_{v \in V} |g_j(v, w) f_{W|V}(w|v)| dw < \infty$.

**Assumption D4.** $b(v)$ is continuous in $v \in V$ except at $v_0$ and is $\lim_{v \uparrow v_0} b(v) = b(v_0) \neq \lim_{v \downarrow v_0} b(v)^2$ at $v = v_0$.

**Assumption D5.** (i) Assume $F_{U|V}(u|v)$ is continuous in $u \in U$ and $v \in V$; (ii) Assume $F_{U|V,W}(u|v, w)$ is continuous\(^2\) in $u \in U$ and $v \in V$ for every $w \in W$.

Under Assumption Set D, the functions $h(\cdot)$ and $g(\cdot)$ in (4) and (5) are continuous on the support of $V$ (see Proposition 2.2 in Chen and Fan, 2011). And this is sufficient to introduce our local least squares wavelet estimator in (8). Notice that the above auxiliary regressions (4) and (5) model is a special partial liner model and raises a new identification and estimation strategy for $\delta_0$ and $\zeta_0$ due to the fact that the functions $h(\cdot)$ and $g(\cdot)$ are smoother than $I\{\cdot \geq v_0\}$ and thus their wavelet coefficients would have different magnitude.

### 2.2 Local Least Squares Wavelet Estimator with Single Scale

Under discontinuous incentive assignment mechanism, LATE is identified as $\theta = \delta_0/\zeta_0$, where $\delta_0$ and $\zeta_0$ are respectively the parameters in the auxiliary regressions (4) and (5). Since the idea underlying the estimation of $\delta_0$ and $\zeta_0$ is the same, we focus on the estimation of $\delta_0$.

Let us begin with the estimation for $\delta_0$. Let $F_V(\cdot)$ and $\tilde{F}_V(\cdot)$ denote the ture and empirical distribution functions of $V$ and $\tau \equiv F_V(v_0)$. Let $V_{i:n} \leq \cdots \leq V_{i:n}$ denote the order statistics of $\{V_i\}_{i=1}^{n}$ and $\{Y_{i:n}\}_{i=1}^{n}$.

\[^2\]A slightly weaker condition is that $b(v)$ is semi-continuous at $v = v_0$ as long as $\lim_{v \uparrow v_0} P(v) \neq \lim_{v \downarrow v_0} P(v)$.

\[^3\]The continuity assumptions in D1, D2 and D4 are required for the validity of the LATE identification. See Lee (2008) and McCrary (2008) for the formal mathematical arguments. Hartman et al. (2010) provided the economics interpretation of the validity about the strategic self-selection in the dynamic structural model.
the concomitants (induced order statistics) of \( \{V_{i:n}\}_{i=1}^n \). Further let \( t_i = i/n \) for \( 1 \leq i \leq n \). Then the induced order statistics \( \{Y_{i:n}\}_{i=1}^n \) satisfy:

\[
Y_{i:n} = h(V_{i:n}) + \delta_0 I\{V_{i:n} \geq v_0\} + \varepsilon_{i:n}
\]

\[
= h(\hat{F}_V^{-1}(t_i)) + \delta_0 I\{t_i \geq \hat{F}_V(v_0)\} + \varepsilon_{i:n}
\]

\[
= H(t_i) + \delta_0 I\{t_i \geq \tau\} + \varepsilon_i,
\]

where \( H(t) \equiv h(F_V^{-1}(t)) \), \( \hat{\tau} \equiv \hat{F}_V(v_0) \) and

\[
\varepsilon_i \equiv \delta_0 [I\{t_i \geq \hat{\tau}\} - I\{t_i \geq \tau\}] - \left[ h(F_V^{-1}(t_i)) - h(F_V^{-1}(t_i)) \right] + \varepsilon_{i:n}.
\]

**Assumption set A.**

**Assumption A1.** A random sample \( (V_i, Y_i, D_i), i = 1, ..., n \), is available.

**Assumption A2.**

(H). Let \( H(t) \equiv h(F_V^{-1}(t)) \). \( H(t) \) is \( p \) times continuously differentiable for \( t \in (0,1) \setminus \{\tau\} \), and \( H(\cdot) \) is continuous at \( \tau \) with finite right and left-hand derivatives up to the order \( p \geq m + 1 \).

(G). Let \( G(t) \equiv g(F_V^{-1}(t)) \). \( G(t) \) is \( q \) times continuously differentiable for \( t \in (0,1) \setminus \{\tau\} \), and \( G(\cdot) \) is continuous at \( \tau \) with finite right and left-hand derivatives up to the order \( q \geq m + 1 \).

**Assumption A3.**

(H) \( \sigma^2_{\varepsilon}(v) \equiv E(\varepsilon^2|V = v) \) is continuous at \( v \neq v_0 \) and its right and left-hand limits at \( v_0 \) exist; (b) For some \( \zeta > 0 \), \( E[|\varepsilon|^{2+\zeta}|v] \) is uniformly bounded on the support of \( V \).

(G) \( \sigma^2_{\varepsilon}(v) \equiv E(\varepsilon^2|V = v) \) is continuous at \( v \neq v_0 \) and its right and left-hand limits at \( v_0 \) exist; (b) For some \( \zeta > 0 \), \( E[|\varepsilon|^{2+\zeta}|v] \) is uniformly bounded on the support of \( V \).

(HG) \( \sigma_{\varepsilon}\varepsilon(v) \equiv E(\varepsilon\varepsilon_i|V_i = v) \) is continuous at \( v \neq v_0 \) and its right and left-hand limits at \( v_0 \) exist.

**Assumption A4.** (a) The real-valued (mother) wavelet function \( \psi(\cdot) \) is continuous with compact support \( [a, b] \), where \( a < 0 < b \) and \( m \) vanishing moments, i.e., \( \int_a^b u^j \psi(u) \, du = 0 \) for \( j = 0, 1, ..., m - 1 \); (b) \( \int_a^b u^m \psi(u) \, du \neq 0 \) and \( \int_a^b |u^m \psi(u)| \, du < \infty \); (c) \( \psi \) has a bounded derivative and an admissibility condition that \( \int |\hat{\psi}(\xi)|^2 / |\xi| \, d\xi < \infty \), where \( \hat{\psi}(\xi) \) is the Fourier transform of \( \psi(t) \).

**Assumption A5.** As \( n \to \infty \), \( j_0 \to \infty \), \( \frac{2j_0}{n} \to 0 \) and \( \left( \frac{1}{2^n} \right)^{2m-1} \sqrt{\frac{n}{2j_0}} \to C < \infty \).

**Assumption A6.** \( F_V^{-1}(v) \) is continuously differentiable on the support of \( V \), where \( F_V^{-1}(\cdot) \) is the quantile function for the random variable \( V \).

Assumption A1 may be relaxed to allow for non i.i.d. data following Chu and Jacho-Chavez (2010). Assumption A2 allows for the potential highest \( p(q) \)-th order derivative discontinuity at \( \tau \); and more importantly this smoothness assumption is more general, which contains both Assumption 2(a): \( H(\tau) \) is continuous but non-differentiable, and 2(b): \( H(\tau) \) is \( p \)-th order differentiable in Porter (2003). Assumption A3 imposes conditions on the possible heteroskedastic conditional variance function \( E[|\varepsilon|^{2+\zeta}|v] \) and \( E[|\varepsilon|^{2+\zeta}|v] \). Assumption A4 specifies the class of wavelet functions \( \psi \) and its vanishing moment. Notice that the wavelet function integrates to zero, \( \int_a^b \psi(t) \, dt = 0 \), and this leads great differences and implications from the kernel
function. See Section 3 for further discussion. Examples of $\psi$ include wavelet functions such as the class of Daubechies’ compactly supported wavelet functions $D(L)$ and the class of least asymmetric wavelet functions $LA(L)$, where $m = L$ and $[a, b] = [- (L - 1), L]$. Assumption A5 imposes conditions on the scale parameter $j_0$. In the end, Assumption A6 is technical so that we could apply the first order Taylor expansion.

When incorporating the potential highest $p$-th order derivative discontinuity at $\tau$, the equation (6) would be rewritten as

$$
Y_{i:n} = H^{*}(t_i) + \sum_{k=0}^{p} \delta_k \cdot [F_{V}^{-1}(t_i) - v_0]^{k} I\{t_i \geq \tau\} + \varepsilon_i, \quad (7)
$$

where

$$
H(t_i) \equiv H^{*}(t_i) + \sum_{k=1}^{p} \delta_k \cdot [F_{V}^{-1}(t_i) - v_0]^{k} I\{t_i \geq \tau\},
$$

and $\{\delta_k\}_{k=0}^{p}$ is the potential different order derivative jump sizes at $v_0(\tau)$, i.e., $\delta_0$ is the jump size, $\delta_1$ is the kink size, $\delta_2$ is the second order derivative discontinuous jump size and so on. Thanks to the reformation in equation (7), the function $H^{*}(\cdot)$ is $p$-th times differentiable on the whole support including the point $\tau$.

Let $\tilde{\Delta}^A_{j_0}(t)$ denote the wavelet coefficient of $\{A_i\}_{i=1}^{n}$ at the location parameter $t \in [0, 1]$ and scale parameter $j_0$:

$$
\tilde{\Delta}^A_{j_0}(t) = \frac{2^{j_0/2}}{n} \sum_{i=1}^{n} A_i \psi \left[ 2^{j_0}(t_i - t) \right].
$$

Applying the wavelet transformation on both sides of the equation (7), we obtain the wavelet-transformed partial linear model

$$
\tilde{\Delta}^{Y}_{j_0}(t) = \tilde{\Delta}^{H^{*}}_{j_0}(t) + \sum_{k=0}^{p} \delta_k \cdot \tilde{\Delta}^{d k}_{j_0}(t) + \tilde{\Delta}^{e}_{j_0}(t),
$$

where $d k(t_i) \equiv [F_{V}^{-1}(t_i) - v_0]^{k} I\{t_i \geq \tau\}$ and

$$
\tilde{\Delta}^{Y}_{j_0}(t) = \frac{2^{j_0/2}}{n} \sum_{i=1}^{n} Y_{i:n} \psi \left[ 2^{j_0}(t_i - t) \right],
$$

$$
\tilde{\Delta}^{d k}_{j_0}(t) = \frac{2^{j_0/2}}{n} \sum_{i=1}^{n} d k(t_i) \psi \left[ 2^{j_0}(t_i - t) \right],
$$

$$
\tilde{\Delta}^{e}_{j_0}(t) = \frac{2^{j_0/2}}{n} \sum_{i=1}^{n} e_i \psi \left[ 2^{j_0}(t_i - t) \right].
$$

It is known that the wavelet coefficient $\tilde{\Delta}^A_{j_0}(t)$ captures the variation of the sequence $\{A_i\}_{i=1}^{n}$ at the location parameter $t$ and scale parameter $j_0$. When the scale parameter $j_0$ is large enough, $\tilde{\Delta}^A_{j_0}(\tau)$ should be small unless there is a jump or isolated singularity in $\{A_i\}_{i=1}^{n}$ at $\tau$. Since the function $H^{*}(\cdot)$ is $p$-th times differentiable on the whole support $(0, 1)$, thus the wavelet coefficient $\tilde{\Delta}^{H^{*}}_{j_0}(t)$ should be close to zero on $t \in [0, 1]$. On the contrast the function $d k(\cdot)$ has the $k$-th order derivative discontinuity at the point $\tau$, whose wavelet coefficient $\tilde{\Delta}^{d k}_{j_0}(t)$ should be significantly different from zero in the neighborhood of the point $\tau$. Following these arguments, an approximate parametric linear regression at the neighborhood of $\tau$ is

$$
\tilde{\Delta}^{Y}_{j_0}(t) \approx \delta^T \cdot \tilde{\Delta}^{d}_{j_0}(t) + \tilde{\Delta}^{e}_{j_0}(t),
$$

5
where
\[
\delta_J^T = [\delta_0, \delta_1, \ldots, \delta_p],
\]
\[
\hat{\Delta}_{j0}^d (t) = \left[ \hat{\Delta}_{j0}^{d0} (t), \hat{\Delta}_{j0}^{d1} (t), \ldots, \hat{\Delta}_{j0}^{dp} (t) \right].
\]

Motivated by the ordinary least squares estimator, we propose the local least squares wavelet estimator \( \hat{\delta}_J \) in the equation(8). We emphasis local in the context, because the wavelet coefficients \( \hat{\Delta}_{j0}^d (t) \) would be small if \( t \) is far away from the point \( \tau \), in which case it becomes a smooth function. Notice that for the \( \hat{\delta}_J \), we replace \( F_V^{-1}(\cdot) \) and \( \tau \) with the empirical quantile function \( \hat{F}_V^{-1}(\cdot) \) and \( \hat{\tau} \),

\[
\hat{\delta}_J = \arg \min_{\delta_0, \ldots, \delta_p} \sum_{i=1}^{n} \left[ \hat{\Delta}_{j0}^Y (t_i) - \sum_{k=0}^{p} \delta_k \cdot \hat{\Delta}_{j0}^d (t_i) \right]^2 \hat{D}_{j0} (t_i), \tag{8}
\]

where \( \hat{\Delta}_0 (t_i) = \left[ \hat{F}_V^{-1}(t_i) - v_0 \right] I\{t_i \geq \hat{\tau} \} \) and \( \hat{D}_{j0} (t_i) = I\{a \leq \hat{\tau} - t_i \leq b \} \) which is the cone of influence defined in Mallat (2009)\(^5\).

The local least squares wavelet estimator \( \hat{\delta}_J \) has a closed-form expression

\[
\hat{\delta}_J = \left[ \hat{\Delta}_{j0}^\hat{\tau} \right]^T \cdot \hat{D}_{j0} \cdot \hat{\Delta}_{j0}^\hat{\tau}^{-1} \left[ \hat{\Delta}_{j0}^\hat{\tau} \right]^T \cdot \hat{D}_{j0} \cdot \hat{\Delta}_{j0}^\hat{\tau},
\]

where

\[
\hat{\Delta}_{j0}^\hat{\tau} = \left[ \hat{\Delta}_{j0}^{\hat{\tau}0}, \hat{\Delta}_{j0}^{\hat{\tau}1}, \ldots, \hat{\Delta}_{j0}^{\hat{\tau}p} \right],
\]

\[
\left( \hat{\Delta}_{j0}^\hat{\tau} \right)^T = \left[ \left( \hat{\Delta}_{j0}^{\hat{\tau}0} (t_1) \right)^T, \left( \hat{\Delta}_{j0}^{\hat{\tau}0} (t_2) \right)^T, \ldots, \left( \hat{\Delta}_{j0}^{\hat{\tau}0} (t_n) \right)^T \right],
\]

\[
\hat{\Delta}_{j0}^\hat{\tau} (t) = \left[ \hat{\Delta}_{j0}^{\hat{\tau}0} (t), \hat{\Delta}_{j0}^{\hat{\tau}1} (t), \ldots, \hat{\Delta}_{j0}^{\hat{\tau}p} (t) \right],
\]

\[
\hat{D}_{j0} = \text{diag} \left[ \hat{D}_{j0} (t_1), \hat{D}_{j0} (t_2), \ldots, \hat{D}_{j0} (t_n) \right],
\]

\[
\left( \hat{\Delta}_{j0}^\hat{\tau} \right)^T = \left[ \hat{\Delta}_{j0}^\hat{\tau} (t_1), \hat{\Delta}_{j0}^\hat{\tau} (t_2), \ldots, \hat{\Delta}_{j0}^\hat{\tau} (t_n) \right].
\]

THEOREM 2.1 Under Assumption Set D and A and \( p \geq 2m \),

1. The asymptotic bias of the local least squares wavelet estimator \( \hat{\delta}_J \) is

\[
\lim_{n \to \infty} \text{diag} \left[ \frac{2^{(2m-1)j_0} \cdots 2^{(2m-2)p} \cdots 2^{(2m-p-1)j_0}}{H^*(2m-1)(\tau) \cdot (M^*)^{-1}_{(0,0)} N^*_0}, \ldots, \frac{2^{(2m-1)j_0} \cdots 2^{(2m-p-1)j_0}}{H^*(2m-1)(\tau) \cdot (M^*)^{-1}_{(p,0)} N^*_0} \right],
\]

\[
E(\hat{\delta}_J) = \delta_J
\]

\[\frac{2^{(2m-1)j_0} \cdots 2^{(2m-p-1)j_0}}{H^*(2m-1)(\tau) \cdot (M^*)^{-1}_{(0,0)} N^*_0}, \ldots, \frac{2^{(2m-1)j_0} \cdots 2^{(2m-p-1)j_0}}{H^*(2m-1)(\tau) \cdot (M^*)^{-1}_{(p,0)} N^*_0} \].

\[^4\text{When substituting the quadratic loss function with the absolute (check) loss function, we would be able to estimate the median (quantile) jump regression model, respectively.}\]

\[^5\text{Other support-shrinking functions would also work, such as, Epanechnikov kernel and Gaussian kernel substituted for } \hat{D}_{j0} (t_i), \text{although we still do not know the effects from the additional smoothing parameter induced from these general kernel functions. When these kernels are used, we would either assume the bandwidth } h \text{ in the kernel equalling to } \frac{1}{2\lambda_0} \text{ or apply bivariate cross-validation for selecting } (h,j_0).\]

6
where
\[
\left[ M^*_{(i,j)} \right]_{(p+1)\times(p+1)}
= \frac{1}{f_{i+j}^*(v_0)} \int_a^b \int_a^b (w-t)^i(v-t)^j I\{w-t \geq 0\}I\{v-t \geq 0\}\psi(w)\psi(v)\,dw\,dv \, dt \text{ for } 0 \leq i,j \leq p
\] (9)
and
\[
N^*_0 = \frac{1}{m!(m-1)!} \int_a^b \psi(u)u^m \, du \cdot \int_a^b \int_a^b I\{w-t \geq 0\}(-t)^{m-1}\psi(w)\,dt\,dw.
\]

(2) The asymptotic variance of the local least square wavelet estimator \( \widehat{\delta}_J \) is
\[
\lim_{n \to \infty} n \cdot \Xi \cdot \text{Var}(\widehat{\delta}_J) = (M^*)^{-1}V^*(M^*)^{-1},
\]
where
\[
[\Xi_{(i,j)}]_{(p+1)\times(p+1)} = 2^{(1+2i)j_0} \text{ for } 0 \leq i \leq p \text{ and } 0 \text{ otherwise}
\]
and
\[
\left[ V^*_{(i,j)} \right]_{(p+1)\times(p+1)}
= \frac{\sigma^2_{i+j}(v_0)}{f_{i+j}^*(v_0)} \int_a^b \int_a^b \left( \int_a^b I\{w-t \geq 0\}(w-t)^i\psi(w)\psi(u+t)\,dw \right) \, du \\
+ \frac{\sigma^2_{i+j}(v_0)}{f_{i+j}^*(v_0)} \int_a^b \int_a^b \left( \int_a^b I\{w-t \geq 0\}(w-t)^i\psi(w)\psi(u+t)\,dw \right) \, du.
\]

(3) For the jump size estimator \( \widehat{\delta}_0^J \), we have
\[
\sqrt{\frac{n}{2j_0}} \left( \widehat{\delta}_0^J - \delta_0 \right) \overset{d}{\to} N(CB_0, V_0),
\]
where
\[
B_0 \equiv H^{*(2m-1)}(\tau) \cdot (M^*)^{-1}(0,0) N^*_0,
\]
\[
V_0 \equiv ((M^*)^{-1}V^*(M^*))^{-1}(0,0).
\]

From the above theorem the asymptotic bias order of \( \widehat{\delta}_0^J \) is \( O \left( \frac{1}{2^{2m-1+j_0}} \right) \) under the general Assumption A2(H), which only states the existence of right and left-hand derivatives up to the \( p \)-th order at \( \tau \) without
assuming any differentiability. This result presents a different picture: whereas the partial linear estimator in Porter (2003) and the wavelet OLS estimator in Chen and Fan (2011) have a fixed bias order $O(h^2)$ and $O\left(\frac{1}{2m-1}\right)$ when $H(\tau)$ is non-differentiable, our local least squares wavelet estimator $\hat{\delta}_{0j}^d$ has $O(\frac{1}{2m^2-1})$. Thus even without continuity in the derivatives of $H$ at $\tau$, the local least squares wavelet estimator improves on the bias behavior similar to the local polynomial kernel regression approach. This is not surprising, given that the local polynomial kernel regression approach is shown to be asymptotically equivalent to the local least squares wavelet estimator with a specific $\psi$ wavelet function. Section 3.2 provides a detailed explanation.

Remarks.

(1) The first order asymptotic bias of $\hat{\delta}_{0j}^d$ is independent of the underlying density $f_V(v_0)$, whose feature is usually called design-adaptive (Fan, 1992) to both random and fixed designs, to both highly clustered and nearly uniform design.

In the finite sample, our local least squares wavelet estimator $\hat{\delta}_{0j}^d$ has the finite unconditional variance through the equispacing method as in Hall, et al. (1998) Thus $\hat{\delta}_{0j}^d$ is robust with respect to the choice of the scales (around the optimal scale of MSE) and support-shrinking functions $\tilde{D}_{0j}^\gamma(\cdot)$. However for the local polynomial kernel regression approach, a finite-sample analysis in Seifert and Gasser (1996) showed that local polynomial kernel estimators are non-robust with respect to the choice of bandwidth and kernels, that is, kernel functions with compact support can lead to a breakdown at the unconditional variance which occurs sharply at some random bandwidth, depending on the realization of the design. Our local least squares wavelet estimator $\hat{\delta}_{0j}^d$ does not show this behavior and is thus a safe and fast approach.

(2) Besides the jump size $\delta_0$, our local least squares wavelet estimator $\hat{\delta}_{0j}^d$ also provides the estimates for other order derivative jump sizes. And their asymptotic behaviors could be seen from Theorem 2.1.

(3) The optimal convergence rate of $\delta_0$ under the equation(6) was proved by Porter (2003) and Sun (2005). And our local least squares wavelet estimator $\hat{\delta}_{0j}^d$ could attain the optimal convergence rate when $2m - 1 \geq p$. Notice that this condition does not contradict the condition $p \geq 2m$ in Theorem 2.1, since the condition $p \geq 2m$ in Theorem 2.1 is merely for obtaining an explicit asymptotic bias term so that a direct plug-in method would be possible for selecting the scale parameter $j_0$. When there is no explicit asymptotic bias term ($p < 2m$), we could still select the optimal scale $j_0^{opt}$ by the local cross-validation in Section 5.1.

(4) The local least squares wavelet estimator $\hat{\delta}_{0j}^d$ does not consider any correlation structure within the wavelet-transformed error $\tilde{\Delta}_{0j}^\epsilon(t)$ in the linear model $\tilde{\Delta}_{0j}^\gamma(t) \approx \delta_0 \cdot \tilde{\Delta}_{0j}^d(t) + \tilde{\Delta}_{0j}^\epsilon(t)$, hence a more efficient estimator for $\delta_0$ would be possible. However according to Ruckstuhl et al.(2000) and Lin and Carrol (2000), such more efficient estimator would be too difficult to construct due to the fact the correlation is a global property of the error structure which is not important to methods which act locally in the covariate space.

\footnote{The fixed bias order means that (1) the asymptotic bias order from the partial linear estimator is independent of the kernel vanishing moment; (2) the asymptotic bias order from the wavelet OLS estimator is independent of the wavelet vanishing moment.}

\footnote{Other ways of attaining the finite unconditional variance would be also available, such as, the local polynomial ridge regression by Seifert and Gasser (1996, 2000) and the shrinkage local linear regression by Hall and Marron (1997).}
An exception case for a specific stationary ARMA\( (p, q) \) error is proposed by Xiao, et al. (2003), where a pre-whitening transformation was applied in advance so that a modified more efficient local polynomial kernel estimator was achieved.

(5) The asymptotic variances of \( \hat{\delta}_0 \), which are required to construct Wald-type confidence sets, are rather complicated due to discontinuities in the conditional variance and multiple integration. Typically in order to estimate the asymptotic variances of \( \hat{\delta}_0 \), we need additional nonparametric regressions to estimate the left and right limits of the conditional variances and we also need nonparametric density estimation for the forcing variable. In this paper we implement the basic bootstrap method\(^8\) which does not require complicated asymptotic variance estimation.

Unlike the classical bootstrap for the parametric coefficients in the partial linear model (Cheng and Huang, 2011 and Hardle et al., 2004), the proposed bootstrap in this paper does not need to estimate the nonparametric function \( H^*(\cdot) \) at all because of the closed-form of our local least squares wavelet estimator. There are three variations of the bootstrap regimes: the first one is the classical \( n \) out of \( n \) nonparametric bootstrap; the second one is \( m \) out of \( m \) nonparametric bootstrap, which \( m \) is the effective sample from the cone of influence function in the local least squares wavelet estimation function; the last one is the parametric bootstrap based on the resampling of the wavelet coefficients for the linear model \( \hat{\Delta}_{j_0}^V (t) \approx \hat{\delta}_0 \hat{\Delta}_{j_0}^{\hat{V}} (t) + \hat{\delta}_0 \), and for this bootstrap being consistent we need to a fast decay rate \( j_0 \) to make the neglection of \( \hat{\Delta}_{j_0}^{H^*} (t) \) becomes asymptotically negligible.

We are now ready to estimate the LATE parameter \( \delta_0 \). Let \( \{D_{i:n}\}_{i=1}^n \) denote the concomitants of \( \{V_{i:n}\}_{i=1}^n \) corresponding to \( \{D_{i:n}\}_{i=1}^n \). Let a finer representation for the equation(5) \( D_{i:n} = G^*(t_i) + \sum_{k=0}^{p} \zeta_k \cdot [F_{V}^{-1}(t_i) - v_0]^k I\{t_i \geq \tau\} + e_i^D \),

\[ (10) \]

where

\[ e_i^D \equiv \zeta_0 \left[ I\{t_i \geq \tau\} - I\{t_i \geq \tau\} \right] - \left[ g(F_{V}^{-1}(t_i)) - g(\hat{F}_{V}^{-1}(t_i)) \right] + e_i^{D,v} \]

\[ G(t_i) \equiv G^*(t_i) + \sum_{k=1}^{p} \zeta_k \cdot [F_{V}^{-1}(t_i) - v_0]^k I\{t_i \geq \tau\}. \]

The local least squares wavelet estimator for the conditional probability model \( \Pr (D = 1|V = v) \) is

\[ \{\hat{\zeta}_0, \hat{\zeta}_1, \ldots, \hat{\zeta}_p \} = \arg \min_{\zeta_0, \zeta_1, \ldots, \zeta_p} \sum_{i=1}^{n} \left[ \hat{\Delta}_{j_0}^D (t_i) - \sum_{k=0}^{p} \zeta_k \cdot \hat{\Delta}_{j_0}^{\hat{V}} (t_i) \right]^2 \hat{D}_{j_0}(t_i) \]

For simplicity, we have used the same mother wavelet \( \psi (\cdot) \) and scale level \( j_0 \) to estimate \( \delta_0 \) and \( \zeta_0 \). This

\[ \text{8} \] Another practical method would be to use empirical likelihood for the confidence construction in Otsu and Xu (2011). They showed that by internalizing the the estimation of an asymptotic variance, the empirical likelihood confidence set has better higher-order coverage properties than Wald-type.

\[ \text{9} \] Since \( \zeta_0 \) is constrained to lie in the closed set \([-1, 1]\), we might propose the constrained local least square wavelet regression

\[ \min_{\zeta_0, \zeta_1, \ldots, \zeta_p} \sum_{i=1}^{n} \left[ \hat{\Delta}_{j_0}^D (t_i) - \sum_{k=0}^{p} \zeta_k \cdot \hat{\Delta}_{j_0}^{\hat{V}} (t_i) \right]^2 \hat{D}_{j_0}(t_i) \text{ subject to } \zeta_0 \in [-1, 1]. \]

The asymptotic distribution could be derived in similar to \( \hat{\delta}_0 \), while accommodating to the constraint by the epi-convergence in distribution in Wang (1996) and Knight (2002).
can be relaxed at the expense of more tedious derivations.

**THEOREM 2.2** Under Assumption Set D and A and $p \geq 2m$, $q \geq 2m$,

\[
(n/2^n)^{1/2} \begin{pmatrix}
\frac{\hat{\delta}_j}{\hat{\zeta}_j} - \frac{\delta_0}{\zeta_0}
\end{pmatrix}
\xrightarrow{d} N \left( \begin{pmatrix}
CB_0 \\
CB_0^D
\end{pmatrix}, \begin{pmatrix}
V_0^Y \\
V_0^Y D
\end{pmatrix} \right),
\]

and

\[
(n/2^n)^{1/2} \begin{pmatrix}
\frac{\hat{\delta}_j}{\hat{\zeta}_j} - \frac{\delta_0}{\zeta_0}
\end{pmatrix}
\xrightarrow{d} N \left( \frac{1}{\zeta_0} C \left( B_0 - \frac{\delta_0}{\zeta_0} D_0 \right), \frac{1}{\zeta_0} \left[ V_0 - \frac{2\delta_0}{\zeta_0} V^Y_0 + \frac{\delta_0^2}{\zeta_0} V^D_0 \right] \right),
\]

where

\[
B_0^D = G^* (2m-1)(\tau) \cdot (M^*)^{-1} (0,0) N^*_0,
\]

\[
V_0^D = ((M^*)^{-1} V^D (M^*))^{-1} (0,0),
\]

\[
V_0^Y D = ((M^*)^{-1} V^Y D (M^*))^{-1} (0,0),
\]

and $V^D$ and $V^Y D$ are similar to $V^*$, except $(\sigma_{2-}, \sigma_{2+})$ replaced by $(\sigma_{2-}, \sigma_{2^+})$ and $(\sigma_{2-}, \sigma_{2^+})$.

The proof follows that apply the Cramer-Wold Device to establish the joint limiting distribution of $(n/2^n)^{1/2}(\hat{\delta}_j - \delta_0)$ and $(n/2^n)^{1/2}(\hat{\zeta}_j - \zeta_0)$, then use Delta method to establish the asymptotic distribution for $\hat{\delta}_0/\hat{\zeta}_j$.

### 2.3 Local Least Squares Wavelet Estimator with Multiscale

In this section, we would briefly mention the local least squares wavelet estimator with multiscale parameters. The insight is gained from the fact: given the cut-off location $v_0$, the wavelet coefficients are large in many scales other than $j_0$. And this efficiency gain from using more information in multiscale estimators is confirmed by Theorem 4.3 and 4.5 in Chen and Fan (2011).

For the location parameter $a \leq 2^{j} (\tau - t) \leq b$ and the scale parameter $j_L \leq j \leq j_U$, the multiscale wavelet-transformed linear models are

\[
\hat{\Delta}_{jL}^Y (t) \approx \delta_j^* \cdot \hat{\Delta}_{jL}^d (t) + \hat{\Delta}_{jL}^c (t),
\]

\[
\hat{\Delta}_{jU}^Y (t) \approx \delta_j^* \cdot \hat{\Delta}_{jU}^d (t) + \hat{\Delta}_{jU}^c (t),
\]

Then the local least squares wavelet estimator with multiscale is

\[
\hat{\delta}_{JM} = \arg \min_{\delta_j} \sum_{j=L}^{U} \sum_{l=1}^{n} \left[ \hat{\Delta}_{jL}^Y (t_l) - \delta_{k_l} \cdot \hat{\Delta}_{jL}^d (t_l) \right]^2 D_j (t_l).
\]
where
\[ \mathbf{\hat{\delta}_{JM}} = \left[ \mathbf{\hat{\delta}_{0}}, \mathbf{\hat{\delta}_{1}}, \ldots, \mathbf{\hat{\delta}_{p}} \right]. \]

Notice that \( \mathbf{\hat{\delta}_{JM}} \) is actually a pooled OLS estimation, ignoring the error dependence structure in the model\( (11) \) through all the wavelet-coefficients data. Another possible approach, which we call component OLS estimation (Ruckstuhl et al., 2000), involve fitting separate linear models using the local least squares wavelet estimator with single scale and then combining these estimators to produce an overall estimator. For these two estimators we do not incorporate the information contained in the error covariance structure in the model\( (11) \), because Lin and Carroll (2000) showed that in typical random effects panel data models, it had better to estimate the regression by ignoring the correlation structure - the working independence\(^{10}\) approach.

### 3 Jump Size Estimators

In the first part of this section, we discuss the advancement of the wavelet estimators focusing on the asymptotic bias comparison. There are two crucial ingredients involved to improve the asymptotic properties of the wavelet estimators: one is to increase the effective sample size by including more wavelet coefficients at different location and scale parameters, and the other is to obtain smoother function \( H^*(\cdot) \) by considering the potential higher order discontinuous jumps in \( H(\tau) \). The second part of this section shows that most existing jump size estimators are asymptotically equivalent to the local least squares wavelet estimator in terms of specific wavelet functions and discontinuous order \( p \) in equation\( (8) \). We think such generalization would provide a new understanding for the discontinuous conditional mean model.

#### 3.1 Wavelet-based Jump Size Estimators

For a complete knowledge of the wavelet-based jump size estimators, we begin with the single scale with single location wavelet estimator \( \mathbf{\hat{\delta}_{0PK}} \)\(^{11}\) in Park and Kim (2006),

\[ \mathbf{\hat{\delta}_{0PK}} = \frac{\mathbf{\hat{\Delta}_{j_0}}(\mathbf{\bar{\tau}})}{\mathbf{\Delta}_{j_0}(\mathbf{\bar{\tau}})}. \]

The estimator \( \mathbf{\hat{\delta}_{0PK}} \) is motivated from estimating \( \delta_0 \) from only one pair of wavelet coefficients \( \left( \mathbf{\hat{\Delta}_{j_0}}(\mathbf{\bar{\tau}}), \mathbf{\hat{\Delta}_{j_0}}(\mathbf{\bar{\tau}}) \right) \) at the location \( \mathbf{\bar{\tau}} \) and scale \( j_0 \). The locally linear model becomes

\[ \mathbf{\hat{\Delta}_{j_0}}(\mathbf{\bar{\tau}}) \approx \delta_0 \cdot \mathbf{\hat{\Delta}_{j_0}}(\mathbf{\bar{\tau}}) + \mathbf{\hat{\Delta}_{j_0}}(\mathbf{\bar{\tau}}). \]

According to Lemma 1 in Wang\( (1995) \), \( \mathbf{\hat{\Delta}_{j_0}}(\mathbf{\bar{\tau}}) \) would be negligible compared to \( \mathbf{\hat{\Delta}_{j_0}}(\mathbf{\bar{\tau}}) \) when choosing the appropriate \( j_0 \). Therefore we solve \( \delta_{0PK}^{\mathbf{\hat{\delta}}} \) by \( \mathbf{\hat{\Delta}_{j_0}}(\mathbf{\bar{\tau}}) = \mathbf{\hat{\delta}_{0PK}} / \mathbf{\hat{\Delta}_{j_0}}(\mathbf{\bar{\tau}}). \)

\(^{10}\)Nevertheless under specific conditions, there exist more efficient nonparametric estimators by considering the error covariance structure, such as, Welsh and Yee (2006) and Martins-Filho and Yao (2008). But how to optimally incorporate the error covariance to attain the most efficient estimator (against oracle procedures for estimating \( \delta_0 \) based on knowledge of \( H(\tau) \)) is still an open question.

\(^{11}\)The original estimator \( \delta_{0PK}^{\mathbf{\hat{\delta}}} \) in Park and Kim (2006) is \( 2^{j_0/2} \mathbf{\hat{\Delta}_{j_0}}(\mathbf{\bar{\tau}}) / \int_{0^+}^b \psi(u)du \), but \( \sqrt{n} \left( \delta_{0PK}^{\mathbf{\hat{\delta}}} - \delta_0 \right) = o_p(1). \)
Under Assumption Set D and A the asymptotic bias of $\hat{d}_0^{PK}$ depends on the differentiability of $H(\tau)$, while the asymptotic variance are the same under both cases.

If $H(\tau)$ is continuous but non-differentiable, then

$$E\left(\hat{d}_0^{PK}\right) = \delta_0 + \left(\frac{1}{2^{3b}}\right) \left[H_1^{(1)}(\tau) - H_{\tau}^{(1)}(\tau)\right] \int_0^1 \psi(u)u \, du + o\left(\frac{1}{2^{3b}}\right) .$$

If $H(\tau)$ is $p(>m)$-th order differentiable, then

$$E\left(\hat{d}_0^{PK}\right) = \delta_0 + \left(\frac{1}{2^{3b}}\right) \left[H^{(m)}(\tau) \int_0^1 \psi(u)u^m \, du \right] + o\left(\frac{1}{2^{3b}}\right) .$$

The asymptotic variance of $\hat{d}_0^{PK}$ is

$$\left(\frac{\hat{d}_0}{n}\right) \frac{\sigma^2_{\hat{d}_0}(v_0) \int_0^1 \psi^2(u) \, du + \sigma^2_{\hat{\tau}_0}(v_0) \int_0^1 \psi^2(u) \, du}{\left(\int_0^1 \psi(u) \, du\right)^2} .$$

Although it has the simplest expression and asymptotically consistent, $\hat{d}_0^{PK}$ is shown to have the worst asymptotic property among all the wavelet-based jump size estimators: only achieving the optimal convergence rate when stronger smoothness assumption is assumed, and has the worst asymptotic bias order and MSE.

Observing the inadequacy of $\hat{d}_0^{PK}$, Chen and Fan (2011) constructed a wavelet OLS estimators $\hat{d}_0^{OLS}$ to improve upon $\hat{d}_0^{PK}$. The insight is gained from the fact: given the scale parameter, the wavelet coefficients are large for the location parameter close to the cut-off point $\tau$; and given the location parameter $\tau$, the wavelet coefficients are also large for multiple scale parameters. Such improvement is closely related to add more information (wavelet coefficients) into the effective sample, which is usually called integrating\textsuperscript{12} in the statistical literature. In particular the spectral density estimator in Hannan (1962) achieved Gauss-Markov efficiency by integrating all the sample spectral densities for a linear regression, which increased more effective sample from single spectral density estimator $b(k)$ to multiple spectral densities estimator $b$ in Hannan (1962). In the context of the nonparametric regression the integrating idea could be also found in He and Huang (2009), where they proposed a double-smoothing local linear estimator being constructed by internally combining all fitted values of local lines in its neighborhood with another round of smoothing.

The motivation for $\hat{d}_0^{OLS}$ using the integration to increase the effective sample is as follows. When allowing the location parameter $t \in [\hat{\tau} - \omega, \hat{\tau} + \omega]$ where $\omega$ is infinitesimal and the scale parameter $j \in [j_L, j_U]$, we have a collection wavelet coefficients \{\hat{\Delta}^Y_j(t), \hat{\Delta}^{d0}_j(t)\}_t=\hat{\tau}-\omega,j=j_L\text{ for the locally linear model } \hat{\Delta}^Y_j(t) \approx \delta_0 \cdot \hat{\Delta}^{d0}_j(t) + \hat{\Delta}^\tau_j(t) . \text{ Thus we have our local OLS } \hat{d}_0^{OLS} \text{ estimator based on the wavelet coefficients restricted to this range of the local and scale parameter. For example when } a \le 2^j(\hat{\tau} - t) \le b \text{ and } j \in [j_L, j_U]$$

\text{Another plausible way of improving the asymptotic property upon } \hat{d}_0^{PK} \text{ would be to form a finite (independent of } n) \text{ linear combination of a preliminary estimator evaluated at nearby cut-point } \tau \text{ with the specified weights. See Ming-Yen Cheng et.al. (2004) to reduce the asymptotic variance and Choi and Hall (1998) for the bias reduction under this methodology.}
where \( \lim_{n \to \infty} (j_U - j_L) = \infty \),

\[
\tilde{\sigma}^2_{\text{OLS}} = \frac{\sum_{j=j_L}^{j_U} \int_{0}^{1} \Delta_j^Y(t) \, \tilde{\Delta}_j^{\theta_0}(t) \, \tilde{D}_j(t) \, dt}{\sum_{j=j_L}^{j_U} \int_{0}^{1} \left[ \Delta_j^{\theta_0}(t) \right]^2 \, \tilde{D}_j(t) \, dt}.
\]  

(13)

Similar to \( \tilde{\sigma}^2_{\text{PK}} \), the asymptotic bias of \( \tilde{\sigma}^2_{\text{OLS}} \) also depends on the differentiability of \( H(\tau) \) while the asymptotic variance does not.

If \( H(\tau) \) is continuous but non-differentiable, then

\[
E \left( \tilde{\sigma}^2_{\text{OLS}} \right) = \delta_0 + \left( \frac{1}{2^{m_0}} \right) \left( \frac{6}{7} \right) \left[ H_+^{(1)}(\tau) - H_-^{(1)}(\tau) \right] \int_{a}^{b} \int_{a}^{b} L(t) \psi(s-t) I\{ s-t \geq 0 \} \, ds \, dt \, M(v) \, dv + o( 1 )
\]

in which

\[
L(t) = \int_{a}^{b} I\{ w \geq t \} \psi(w) \, dw \text{ and } M(v) = \int_{a}^{b} \int_{a}^{b} I\{ w \geq t + v \} \psi(w) \psi(t) \, dt \, dw.
\]

If \( H(\tau) \) is \( p(>2m) \)-th order differentiable, then

\[
E \left( \tilde{\sigma}^2_{\text{OLS}} \right) = \delta_0 + \left( \frac{1}{2^{m_0}} \right)^{2m-1} \frac{3}{4} \left[ 1 - \left( \frac{1}{2^{m_0}} \right)^{2m+1} \right] \left[ \frac{H^{(2m-1)}(\tau)}{m!(m-1)!} \int_{a}^{b} \int_{a}^{b} L(t) (-t)^{m-1} \, dt \, M(v) \, dv \right] + o( 1 )^{2m-1}.
\]

The asymptotic variance of \( \tilde{\sigma}^2_{\text{PK}} \) is

\[
\tilde{\sigma}^2_{\text{OLS}} = \left( \frac{2^{m_0}}{n} \right) \left( \frac{9}{14} \right) \frac{\sigma^2_+(v_0) \int_{a}^{b-a} M^2(v) \, dv + \sigma^2_-(v_0) \int_{a}^{b} M^2(v) \, dv}{\int_{a}^{b-a} M(v) \, dv}.
\]

The asymptotic results for \( \tilde{\sigma}^2_{\text{PK}} \) and \( \tilde{\sigma}^2_{\text{OLS}} \) reveal the role of the additional information by including more wavelet coefficients. The asymptotic bias order of \( \tilde{\sigma}^2_{\text{OLS}} \) under the assumption \( H(\tau) \) being \( p \)-th order differentiable reduces further to \( O \left( \left( \frac{1}{2^{m_0}} \right)^{2m-1} \right) \) from \( O \left( \left( \frac{1}{2^{m_0}} \right)^m \right) \) for \( \tilde{\sigma}^2_{\text{PK}} \). However under the assumption \( H(\tau) \) being non-differentiable, the order of the asymptotic bias of \( \tilde{\sigma}^2_{\text{OLS}} \) remains the same as that of \( \tilde{\sigma}^2_{\text{PK}} \).

Keep in mind that the above comparison is based on the fixed vanishing moment \( m \), however for a varying vanishing moment the \( \tilde{\sigma}^2_{\text{PK}} \) with \( (2m - 1) \) vanishing moment has the same asymptotic bias order as the \( \tilde{\sigma}^2_{\text{OLS}} \) with \( m \) vanishing moment.

When \( t = \hat{\tau} \) and \( j \in [j_L, j_U] \) where \( j_L < j_U \), the \( \tilde{\sigma}^2_{\text{OLS}} \) degenerates to the multiscale with single location estimator(\( \tilde{D}_j(\cdot) \) is a dirac function now), which reduces the asymptotic bias and variance only proportionally(not the order) regardless of the differentiability of \( H(\tau) \). When \( a \leq 2^j(\hat{\tau} - t) \leq b \) and \( j = j_L = j_U \), the \( \tilde{\sigma}^2_{\text{OLS}} \) degenerates to the single scale with many locations estimator, which only reduces the asymptotic bias order if \( H(\tau) \) is \( p \)-th order differentiable. In the end, the simulation results from Chen and Fan (2011) showed that the non-degenerated \( \tilde{\sigma}^2_{\text{OLS}} \) has smaller MSE than \( \tilde{\sigma}^2_{\text{PK}} \) under the optimal scale \( j^*_{\text{opt}} \).

It is seen that the \( \tilde{\sigma}^2_{\text{PK}} \) and \( \tilde{\sigma}^2_{\text{OLS}} \) do not attain the optimal convergence rate when \( H(\tau) \) is non-differentiable. The deeper reason for the no-optimality is because when deriving the asymptotic bias, the non-differentiable of \( H(\tau) \) keeps us from using two-sided Taylor expansion so that around the cut-off point \( \tau \), only the one-sided Taylor expansions are available and introduce \( H_+^{(1)}(\tau) - H_-^{(1)}(\tau) \) in the bias term (see the
proof in Appendix B). In an effort to get around the problem, we realize that we could model the potential higher order derivative discontinuities of \( H(\tau) \) by rewriting \( H(\cdot) = H^*(\cdot) + \sum_{k=1}^{p} \delta_k \cdot (F_{V_k}^{-1}(\cdot) - v_0)^k I\{ \cdot \geq \tau \} \), where \( H^*(\cdot) \) is \( p \)-th order differentiable on the whole support. After this reformulation, we expect a better asymptotic bias estimator would be possible and it results in our \( \hat{\Delta}_J \). Interestingly from the traditional semiparametric regression model, we know that the best asymptotic bias order of the parametric estimator depends on the smoothness of the unknown nonparametric function, thus the smoother the unknown function, the higher the best asymptotic bias order of the parametric estimator. Although here the \( \hat{\Delta}_0 \) has a nonparametric convergence rate, its asymptotic bias order still shares this general result.

Recall the asymptotic bias order of \( \hat{\Delta}_0 \) is \( O\left((\frac{1}{2\tau_0})^{2m-1}\right) \) no matter whether \( H(\tau) \) is differentiable or not. And when \( \{\delta_k\}_{k=1}^{p} = 0 \), the multiscale \( \hat{\Delta}_l \) would become \( \hat{\Delta}_0^{OLS} \). Thus \( \hat{\Delta}_0 \) attains the optimal convergence rate under a broader set of conditions and has the best asymptotic MSE among all the wavelet-based estimators.

### 3.2 Other Jump Size Estimators

This section exhibits various existing jump size estimators as special cases of the local least squares wavelet estimator. Let us begin with reviewing the relationship between the kernel function \( k(\cdot) \) and wavelet function \( \psi(\cdot) \). First and most importantly,

\[
\int_{\sup p(k)} k(t) dt = 1 \quad \text{but} \quad \int_{\sup p(\psi)} \psi(t) dt = 0.
\]

Thus most nonparametric statistics using the kernel function actually perform the \textit{averaging} operation, but the statistics using the (mother) wavelet function indeed perform the \textit{differencing} operation. Hence for the nonparametric sieve\(^{13}\) regression using the wavelet basis, we both apply the father wavelet (a special kernel) basis to obtain the average feature of the function, and use the mother wavelet basis to capture the finer details. Based on this, we point out our second observation: for two different kernel function \( k_1(\cdot) \) and \( k_2(\cdot) \),

\[
\int_{\sup p(\psi)} \psi(t) dt = 0 \\
= \int_{\sup p(k_1) \cup \sup p(k_2)} [k_1(t) - k_2(t)] dt \\
= \int_{\sup p(k_1)} k_1(t) dt - \int_{\sup p(k_2)} k_2(t) dt \\
= 1 - 1 = 0,
\]

Therefore we could generate a wavelet function \( \psi(\cdot) \) by simply taking the difference between two kernel functions \( k_1(\cdot) \) and \( k_2(\cdot) \). This insight is the basic to generalize most of existing jump size estimators. For the jump size estimator using the difference between two kernel estimates (Nadaraya-Watson or local polynomial

\(^{13}\)Almost all the jump size estimators are built upon the kernel method instead of the sieve method. This is because the technical difficulty associated to the sieve method: "there does not yet exist a general theory of pointwise limiting distribution for a sieve extremum estimator of an unknown function" (Chen, 2008), or only the asymptotic normality of the undersmoothing spline series LS estimator is available in Theorem 3.7 of Chen (2008). Those arguments exclude the analysis on the asymptotic bias term based on the sieve method.
kernel approaches), we expect that such kernel approach corresponds to a wavelet estimator with the wavelet function induced by these two kernels. Furthermore when estimating the $\delta_0$ from the perspective of profiling the unknown nonparametric part in the partial linear model, such as, partial spline estimator, Speckman estimator and partial linear estimator, there still exists an equivalent wavelet interpretation to be within our local least squares wavelet estimator framework.

**Nadaraya-Watson estimator $\widehat{\delta}_0^{NW}$ in Porter (2003):** This estimator $\widehat{\delta}_0^{NW}$ is asymptotically equivalent to $\widehat{\delta}_0^{PK}$ with the only one vanishing moment wavelet. To see this,

$$\widehat{\delta}_0^{NW} = \frac{\sum_j k_h(\tau - t_j)I\{t_j \geq \tau\}y_i}{\sum_j k_h(\tau - t_j)I\{t_j \geq \tau\}} - \frac{\sum_j k_h(\tau - t_i) [1 - I\{t_i \geq \tau\}] y_i}{\sum_j k_h(\tau - t_j) [1 - I\{t_j \geq \tau\}]}$$

$$= \sum_i \left[ \frac{\sum_j k_h(\tau - t_i)I\{t_i \geq \tau\}}{\sum_j k_h(\tau - t_j)I\{t_j \geq \tau\}} - \frac{\sum_j k_h(\tau - t_i) [1 - I\{t_i \geq \tau\}]}{\sum_j k_h(\tau - t_j) [1 - I\{t_j \geq \tau\}]} \right] y_i$$

$$= \sum_i \psi\left(\frac{\tau - t_i}{h}\right)y_i,$$

where $\psi(\cdot) = \frac{k(\cdot)I\{\cdot > 0\}}{\sum_j k_h(\tau - t_j)I\{t_j \geq \tau\}} - \frac{k(\cdot)I\{\cdot < 0\}}{\sum_j k_h(\tau - t_j) [1 - I\{t_j \geq \tau\}]]$. Notice the reciprocal relation between the bandwidth $h$ and the scale parameter $j$ in the nonparametric regression.

**Local polynomial kernel regression estimator $\widehat{\delta}_0^{LP}$ in Porter (2003):** This estimator $\widehat{\delta}_0^{LP}$ is asymptotically equivalent to $\widehat{\delta}_0$, thanks to the transformation of the object function. For example consider the local linear kernel regression for estimating the discontinuous jump size $\delta_0$,

$$\widehat{\delta}_0^{LP} = \arg \min_{\alpha, \beta, \delta_0, \delta_1} \frac{1}{n} \sum_{i=1}^n \left[ Y_i - \alpha - \beta(t_i - \tau) - \delta_0 I\{t_i \geq \tau\} - \delta_1 (F_i^{-1}(t_i) - v_0) I\{t_i \geq \tau\} \right]^2 \left(\frac{t_i - \tau}{h}\right)^P$$

$$= \arg \min_{\alpha, \beta, \delta_0, \delta_1} \frac{1}{n} \sum_{i=1}^n \left[ \Delta Y_{j_0}(t_i) - \Delta \alpha + \beta(t_i - \tau) \right]^2 \left(\frac{t_i - \tau}{h}\right)^P$$

$$= \arg \min_{\alpha, \beta, \delta_0, \delta_1} \frac{1}{n} \sum_{i=1}^n \left[ \Delta Y_{j_0}(t_i) - \delta_0 \Delta \alpha + \beta(t_i - \tau) \right]^2 \left(\frac{t_i - \tau}{h}\right)^P$$

where the second equality comes from applying wavelet transformation to the local linear kernel object function, and the last equality of $\Delta \alpha + \beta(t_i - \tau)$ being small order term are due to the smoothness of $\alpha + \beta(t_i - \tau)$ whose wavelet coefficients are close to zero. Hence for the local linear regression estimator $\widehat{\delta}_0^{LP}$, it behaves like $\widehat{\delta}_0$ with $p = 1$ and the $(2m - 1)$ vanishing moment wavelet. In addition, the relationship between $\widehat{\delta}_0^{LP}$ and $\widehat{\delta}_0^{PK}$ supports that the local least squares wavelet estimator $\widehat{\delta}_0^{PK}$ should have the optimal convergence rate from $\widehat{\delta}_0^{LP}$. Conversely, we think that any other jump size estimators which could not be written in terms of the local least squares wavelet estimator $\widehat{\delta}_0^{PK}$ would not have the optimal convergence rate under the broader set of conditions.

**Partial spline estimator $\widehat{\delta}_0^{PS}$ by Eubank and Whitney (1989):** This estimator $\widehat{\delta}_0^{PS}$ is asymptotically equivalent to $\widehat{\delta}_0^{PK}$ with the $(m-1)$ vanishing moment wavelet. To see this denote $S = \frac{1}{mh} \left[ k\left(\frac{t_i - t_j}{h}\right) \right]_{i,j=1,n}$
with the $m$ vanishing moment kernel $k(\cdot)$, then
\[
\delta_{0}^{\text{PS}} = \arg \min_{H(\cdot), \delta_0} \frac{1}{n} \sum_{i=1}^{n} [y_i - \delta_0 I\{t_i \geq \tau\} - H(t_i)]^2 + \lambda \int_{0}^{1} H'(t)^2 dt \\
= \arg \min_{\delta} [Y - \delta_0 I\{t \geq \tau\}]^T (I - S) [y - \delta_0 I\{t \geq \tau\}] \\
= \left[ I^T \{t \geq \tau\} (I - S) I\{t \geq \tau\} \right]^{-1} [I^T \{t \geq \tau\} (I - S) Y] \\
= \sum_{i=1}^{n} \psi\left(\frac{t_i - \tau}{\sqrt{h}}\right) y_i + \text{s.o.},
\]
where the second equality comes from the linear smoothing spline in Eubank (1994), and the expression for $\psi(\cdot)$ in the last equality could be found in Equation (6f) and (7g) in Cline, et al. (1995).

**Speckman estimator $\delta_{0}^{\text{SP}}$ by Eubank and Speckman (1994):** This estimator $\delta_{0}^{\text{SP}}$ is asymptotically equivalent to $\delta_{0}^{\text{PS}}$ with the $(2m - 1)$ vanishing moment of the wavelet. To see this
\[
\widehat{\delta}_{0}^{\text{SP}} = \left[ I^T \{t \geq \tau\} (I - S)^2 I\{t \geq \tau\} \right]^{-1} [I^T \{t \geq \tau\} (I - S)^2 Y] \\
= \sum_{i=1}^{n} \psi\left(\frac{t_i - \tau}{h}\right) y_i + \text{s.o.},
\]
where the expression for $\psi(\cdot)$ in the last equality could be found in Lemma 3 in Speckman (1994).

**Partial linear estimator $\delta_{0}^{\text{PL}}$ by Porter (2003):** Interestingly the Robinson (1988)-based partial linear estimator $\delta_{0}^{\text{PL}}$ by Porter (2003) also falls into the framework of $\delta_{0}^{\text{PS}}$, when $H(\tau)$ is $p$-th order differentiable. To see this define $W_i^j = \frac{k_h(t_i - t_j)}{\sum_{i=1}^{k_h(t_i - t_j)}}$ and $H'(t_i) = \sum_{j=1}^{n} W_i^j [y_j - \delta_0 I\{t_j \geq \tau\}]$, then
\[
\delta_{0}^{\text{PL}} = \arg \min_{\delta_0} \sum_{i=1}^{n} \left[ y_i - \delta_0 I\{t_i \geq \tau\} - \sum_{j=1}^{n} W_i^j [y_j - \delta_0 I\{t_j \geq \tau\}] \right]^2 \\
= \arg \min_{\delta_0} \sum_{i=1}^{n} \left[ \widehat{\Delta}_0^Y (t_i) - \delta_0 \widehat{\Delta}_0^0 (t_i) - \delta_0 H'(t_i) \right]^2 \widehat{D}_{0} (t_i) \\
= \arg \min_{\delta_0} \sum_{i=1}^{n} \left[ \widehat{\Delta}_0^Y (t_i) - \delta_0 \widehat{\Delta}_0^0 (t_i) - \text{s.o.} \right]^2 \widehat{D}_{0} (t_i),
\]
where the second equality comes from applying wavelet transformation to the least squares object function, and the last equality is because $H(\cdot) \approx H' (\cdot)$ when $h \to 0$ and $H(\cdot)$ is continuous. Notice that since $\delta_{0}^{\text{PL}}$ behave like $\delta_{0}^{\text{OLS}}$ which fails to attain the optimal convergence rate when $H(\tau)$ is non-differentiable, Yu (2010) proposed a partial polynomial kernel estimator to explicitly consider the potential higher order discontinuities and proved it to have the optimal convergence rate even under the non-differentiable case.

### 4 Wavelet Estimators under the Kink Incentive Assignment Mechanism

Similar to the discontinuous incentive assignment mechanism, we place the LATE identification results into two auxiliary regressions then provide the wavelet estimator for it.
4.1 LATE Identification and Auxiliary Regressions

Under the kink incentive assignment mechanism, the LATE would be identified as (see Theorem 2.3 in Chen and Fan, 2011)

\[
\lim_{v \downarrow v_0} \frac{dE(Y|V = v)/dv - \lim_{v \uparrow v_0} dE(Y|V = v)/dv}{\lim_{v \downarrow v_0} P'(v) - \lim_{v \uparrow v_0} P'(v)},
\]

which is the ratio of the kink size in \( E(Y|V) \) to the kink size in \( P(v) \) at \( v_0 \).

We establish auxiliary regressions linking the LATE to kink sizes \( \delta_1 \) and \( \zeta_1 \) in (15) and (16) for the kink incentive assignment mechanism. Thus estimating the LATE is the same as estimating \( \delta_1/\zeta_1 \) in two auxiliary regressions, which are

\[
Y = h_K(V) + \delta_1(V - v_0)I\{V \geq v_0\} + \varepsilon_K, \quad (15)
\]

\[
D = g_K(V) + \zeta_1(V - v_0)I\{V \geq v_0\} + \epsilon_K, \quad (16)
\]

where \( E(\varepsilon_K|V) = 0, E(\epsilon_K|V) = 0, \delta_1 \equiv \lim_{v \downarrow v_0} dE(Y|V = v)/dv - \lim_{v \uparrow v_0} dE(Y|V = v)/dv \) and \( \zeta_1 \equiv \lim_{v \downarrow v_0} P'(v) - \lim_{v \uparrow v_0} P'(v) \).

Assumption Set K is as follows to make sure the auxiliary regressions (15) and (16) are equivalent to the LATE under the kink incentive assignment mechanism.

**Assumption Set K.**

**Assumption K1.** Assume (i) \( f_{V|W}(v|w) \) is continuously differentiable on \( V \) for every \( w \in W \); (ii) \( f_{V}(v) \) is continuously differentiable on \( V \); (iii) \( f_{V|W,U}(v|w,u) \) is continuously differentiable on \( V \) for \( u \in U \) and \( w \in W \).

**Assumption K2.** Assume \( g_1(v, w) \) and \( g_0(v, w) \) are continuously differentiable on \( V \) for every \( w \in W \).

**Assumption K3.** For \( j = 1, 0, \) assume (i) \( E[Y_j] < \infty \);
(ii) \( \sup_v |\frac{\partial f_{V|W}(v|w)}{\partial w}| < \infty \) and \( \int \sup_v |\frac{\partial f_{V|W}(v|w)}{\partial v}| dv < \infty \);
(iii) \( \int \sup_v |\frac{\partial f_{W,U,V}(w,u,v)}{\partial v}| dv < \infty \) and \( \int \sup_v |\frac{\partial f_{W,U,V}(w,u,v)}{\partial u}| dv < \infty \).

**Assumption K4.** \( b(v) \) is continuously differentiable for \( v \in V \) except at \( v_0 \), and is \( \lim_{v \downarrow v_0} b'(v) \neq \lim_{v \uparrow v_0} b'(v) \) at \( v = v_0 \).

**Assumption K5.** (i) Assume \( F_{U|V}(u|v) \) is continuously differentiable in both \( u \in U \) and \( v \in V \); (ii) Assume \( F_{U|V,W}(u|v,w) \) is continuously differentiable in both \( u \in U \) and \( v \in V \) for every \( w \in W \).

Under Assumption Set K, the functions \( h_K(\cdot) \) and \( g_K(\cdot) \) in (15) and (16) are continuously differentiable on the support of \( V \) (see Proposition 2.4 in Chen and Fan, 2011). And this is sufficient to introduce our local least squares wavelet kink size estimator in (18). Notice that the above auxiliary regressions (15) and (16) model is a special partial liner model and raises a new identification and estimation strategy for \( \delta_1 \) and \( \zeta_1 \) due to the fact that the functions \( h_K(\cdot) \) and \( g_K(\cdot) \) are smoother than \( (\cdot - v_0)I\{\cdot \geq v_0\} \) and thus their wavelet coefficients would have different magnitude.
4.2 Local Least Squares Wavelet Estimator with Single Scale

Since the idea underlying the estimation of $\delta_1$ and $\zeta_1$ is the same, we focus on the estimation of $\delta_1$. The induced order statistics $\{Y_{i:n}\}_{i=1}^{n}$ satisfy:

\[ Y_{i:n} = h_K(V_{i:n}) + \delta_1(V_{i:n} - v_0)I\{V_{i:n} \geq v_0\} + \varepsilon_{i:n}^k \]
\[ = h_K(\tilde{F}_V^{-1}(t_i)) + \delta_1 \left( \tilde{F}_V^{-1}(t_i) - v_0 \right) I\{t_i \geq \tilde{F}_V(v_0)\} + \varepsilon_{i:n}^k \]
\[ = H_K(t_i) + \delta_1 (F_V^{-1}(t_i) - v_0) I\{t_i \geq \tau\} + e_i^k, \]

where $H_K(t) \equiv h_K(F_V^{-1}(t))$ and

\[ e_i^k = \delta_1 \left[ I\{t_i \geq \tau\} \left( \tilde{F}_V^{-1}(t_i) - v_0 \right) - I\{t_i \geq \tau\} \left( F_V^{-1}(t_i) - v_0 \right) \right] - \left[ h_K(F_V^{-1}(t_i)) - h_K(\tilde{F}_V^{-1}(t_i)) \right] + \varepsilon_{i:n}^k. \]

**Assumption set B:**

**Assumption B1.** A random sample $(V_i, Y_i, D_i), i = 1, \ldots, n,$ is available.

**Assumption B2.**

(H) Let $H_K(t) \equiv h_K(F_V^{-1}(t))$. $H_K(t)$ is $(p + 1)$ times continuously differentiable for $t \in (0, 1) \setminus \{\tau\}$, and $H_K(\cdot)$ is continuous differentiable at $\tau$ with finite right and left-hand derivatives up to the order $p + 1$ ($\geq m + 2$).

(G) Let $G_K(t) \equiv g_K(F_V^{-1}(t))$. $G_K(t)$ is $(q + 1)$ times continuously differentiable for $t \in (0, 1) \setminus \{\tau\}$, and $G_K(\cdot)$ is continuous differentiable at $\tau$ with finite right and left-hand derivatives up to the order $q + 1$ ($\geq m + 2$).

**Assumption B3.**

(H) $\sigma_{K}(v) \equiv E(\varepsilon_K^2|V = v)$ is continuous at $v \neq v_0$ and its right and left-hand limits at $v_0$ exist; (b) For some $\zeta > 0$, $E[|\varepsilon_K|^{2+\zeta}|v]$ is uniformly bounded on the support of $V$.

(G) $\sigma_{K}(v) \equiv E(\varepsilon_K^2|V = v)$ is continuous at $v \neq v_0$ and its right and left-hand limits at $v_0$ exist; (b) For some $\zeta > 0$, $E[|\varepsilon_K|^{2+\zeta}|v]$ is uniformly bounded on the support of $V$.

(HG) $\sigma_{\varepsilon\varepsilon}(v) \equiv E(\varepsilon_K^2\varepsilon_i^K|V_i = v)$ is continuous at $v \neq v_0$ and its right and left-hand limits at $v_0$ exist.

**Assumption B4.** (a) The wavelet function $\psi(\cdot)$ is continuous with compact support $[a, b]$, where $a < 0 < b$ and $(m + 1)$ vanishing moments, i.e., $\int_a^b u^j \psi(u) du = 0$ for $j = 0, 1, \ldots, m$; (b) $\int_a^b u^m \psi(u) du \neq 0$ and $\int_a^b |u^m \psi(u)| du < \infty$; (c) $\psi$ has a bounded derivative.

**Assumption B5.** As $n \rightarrow \infty$, $j_0 \rightarrow \infty$, $\frac{\sigma_0}{n} \rightarrow 0$ and $\left(\frac{1}{2\sigma_0}\right)^{2m-1} \frac{1}{\sqrt{2\sigma_0}} C^K < \infty$.

**Assumption B6.** $F_V^{-1}(v)$ is continuously differentiable on the support of $V$, where $F_V^{-1}(\cdot)$ is the quantile function for the random variable $V$.

Assumption B2 is assumed to continuous differentiable at $\tau$ from the identification requirement under the kink incentive assignment mechanism, instead of only continuity in Assumption A2. The scale parameter $j_0$ in Assumption B5 is changed for the convergence rate of the kink size estimator. The rest assumptions are accommodated to the kink incentive assignment mechanism.
The local least squares wavelet estimator $\hat{\delta}_K$ is

$$\hat{\delta}_K = \arg \min_{\delta_1, \ldots, \delta_{p+1}} \sum_{t=1}^{n} \left[ \hat{\Delta}_{j_0}(t) - \sum_{k=1}^{p+1} \delta_k \cdot \hat{\Delta}_{j_0}^{d_k}(t) \right]^2 \hat{D}_{j_0}(t),$$

where

$$\hat{\Delta}_{j_0}^d(t) = \begin{bmatrix} \hat{\Delta}_{j_0}^{d_1}(t), \hat{\Delta}_{j_0}^{d_2}(t), \ldots, \hat{\Delta}_{j_0}^{d_{p+1}}(t) \end{bmatrix},$$

and

$$\hat{D}_{j_0} = \text{diag}\left[ \hat{D}_{j_0}(t_1), \hat{D}_{j_0}(t_2), \ldots, \hat{D}_{j_0}(t_n) \right].$$

**THEOREM 4.1** Under Assumption Set K and B and $p \geq 2m$, then

$$\lim_{n \to \infty} \frac{n}{2^{3j_0}} \text{Var}(\hat{\delta}_K) = d < \infty,$$

where $c$ and $d$ are some generic constants.

Since the proof is very similar to Theorem 2.1, it is omitted. Following Section 2.5 in Tsybakov (2009), we could show $\hat{\delta}_K$ attains the optimal convergence rate for the kink size estimator$^{14}$.

And for $\psi_1$ the local least squares wavelet estimator $\hat{\psi}_K$ is

$$\hat{\psi}_K = \arg \min_{\psi_1, \ldots, \psi_{p+1}} \sum_{t=1}^{n} \left[ \hat{\Delta}_{j_0}^d(t) - \sum_{k=1}^{p+1} \psi_k \cdot \hat{\Delta}_{j_0}^{d_k}(t) \right]^2 \hat{D}_{j_0}(t),$$

where

$$\hat{\psi}_K = \begin{bmatrix} \hat{\psi}_1, \hat{\psi}_2, \ldots, \hat{\psi}_{p+1} \end{bmatrix}.$$
The local cross-validation for selecting the scale $j_0$ under the discontinuous incentive assignment mechanism is

$$
\hat{j}^{opt}_0 = \arg \min_{j_0} \frac{1}{n} \sum_{i=1}^{n} \left[ \hat{\Delta}Y_{j_0}(t_i) - \hat{\delta}_0(j_0, -i) \hat{\Delta} t_j(t_i) \right]^2 \hat{D}_{j_0}(t_i)
$$

where

$$
\hat{\Delta}Y_{j_0}(t_i) = \frac{2j^*/2}{n} \sum_{m=1}^{n} Y_{m,n} \psi \left[ 2j^*(t_m - t_i) \right],
$$

$$
\hat{\Delta} t_j(t_i) = \frac{2j^*/2}{n} \sum_{m=1}^{n} I \left( t_m \geq \tau \right) \psi \left[ 2j^*(t_m - t_i) \right],
$$

$$
\hat{D}_{j_0}(t_i) = I \left\{ a \leq 2j^* (\tilde{\tau} - t_i) \leq b \right\},
$$

$$
\hat{\delta}_0(j_0, -i) = \arg \min_{\delta_0, \ldots, \delta_p} \sum_{i=1}^{n} \left[ \hat{\Delta}Y_{j_0}(t_i) - \sum_{k=0}^{p} \delta_k \cdot \hat{\Delta} t_j(t_i) \right]^2 \hat{D}_{j_0}(t_i).
$$

Here $j^*$ is a pre-specified scale\(^{15}\), which is less important according to in Mielniczuk et al. (1989) and Vieu (1991). Notice that the $j_0$ only appears in the leave-one-out estimator $\hat{\delta}_0(j_0, -i)$. And we expect the selected scale $\hat{j}^{opt}_0$ to be

$$
\frac{\text{MSE} \left( \hat{\delta}_0(\hat{j}^{opt}_0) \right)}{\inf_j \text{MSE}(\hat{\delta}_0(j_0))} \xrightarrow{a.s.} 1.
$$

The local cross-validation selected $\hat{j}^{opt}_0$ is robust, but might not have the optimal convergence rate for the relative error of the scale selection.

\textbf{(2) Specify the discontinuous order $p$:}

In application, we need to specify the discontinuous order $p$ before applying the estimator $\hat{\delta}_j$ in equation (8). Suppose that the function $H(\cdot)$ is continuous at $\tau$ with the highest unequal finite right and left-hand derivatives to order $p$ then if we use the value $p_{\text{small}}$ which is strictly smaller $p$, then the estimator $\hat{\delta}_0(p_{\text{small}})$ might not have the smallest asymptotic bias order; on the other hand, if we use the value $p_{\text{big}}$ which is strictly bigger than $p$, then the estimator $\hat{\delta}_0(p_{\text{big}})$ might be more varied in sense of the asymptotic variance.

We suggest to use the LASSO\(^{16}\) (least absolute shrinkage and selection operator) method to achieve the goal of simultaneous estimation and variable selection. By introducing a penalty term on different order discontinuous jump sizes, we achieve the LASSO local least squares wavelet estimator as

$$
\hat{\delta}_j^{LASSO} = \arg \min_{\delta_0, \ldots, \delta_p} \sum_{i=1}^{n} \left[ \hat{\Delta}Y_{j_0}(t_i) - \sum_{k=0}^{p} \delta_k \cdot \hat{\Delta} t_j(t_i) \right]^2 \hat{D}_{j_0}(t_i) + \sum_{k=0}^{p} \lambda_k |\delta_k|,
$$

where $\{\lambda_k\}_{k=0}^{p}$ are another set smoothing parameters satisfying $\lambda_k \rightarrow 0$ as $n \rightarrow \infty$.\[^{15}\] A reasonable prior choice for $j^*$is to satisfy $j^* \rightarrow \infty$ and $2j^*/n \rightarrow 0$ as $n \rightarrow \infty$, but the sensitivity and the choice of these parameters certainly need further evaluation.\[^{16}\]

\(^{15}\)Mallat (2009) Chapter 6 and Kotlyarova and V. Zinde-Walsh (2006, 2008) suggested different ways of estimating the pointwise smoothness $p$ at the cut-off point $v_0$. Based on their estimates $\hat{p}$, we plug in our local least squares wavelet estimator $\hat{\delta}_0(\hat{p})$. But these $\hat{p}$ are only the estimates of the lowest discontinuous order and thus we still face the issue of using $p_{\text{small}}$.\[^{16}\]
Although full investigation is reserved for future work, we suggest the \( \hat{\delta}^{\text{LASSO}}_J \) can contain the exact zero coefficients for these overspecified higher order discontinuous jump sizes. See Xie and Huang (2009) and Ding, et al. (2009).

(3) Estimate the unknown discontinuous location \( v_0 \):

All the above jump size estimator are calculated at the known cut-off point \( v_0 \). When the cut-off point \( v_0 \) is unknown, we might first estimate the cut-off point location \( v_0 \) then substitute this location estimate into the jump size estimator. This is alike a plug-in method, or two-step estimation. See Theorem 1 in Eubank and Speckman (1994) for the asymptotic normality of two-step jump size estimator under this principal.

Contrast to two-step method, we propose a joint estimation for the jump size and cut-off location following a modified local least squares wavelet estimator:

\[
(\hat{\delta}_J, \hat{v}_0) = \arg \min_{\delta_0, \delta_1, \ldots, \delta_p, v_0 \in S} \sum_{l=1}^n \left[ \hat{\Delta} Y_{j_0}(t_l) - \sum_{k=0}^p \delta_k \cdot \hat{\Delta}^{dk}_{j_0}(t_l) \right]^2 \hat{D}_{j_0}(t_l),
\]

where

\[
\hat{\delta}_J = \left[ \hat{\delta}_0^J, \hat{\delta}_1^J, \ldots, \hat{\delta}_p^J \right],
\]

\[
\hat{\Delta}^{dk}_{j_0}(t_l) = \frac{1}{n} \sum_{i=1}^n (\hat{F}_V^{-1}(t_i) - v_0)^k I\{t_i \geq \hat{F}_V(v_0)\} 2^{j_0/2} \psi \left[ 2^{j_0}(t_i - t_l) \right],
\]

\[
\hat{D}_{j_0}(\cdot) = I\{a \leq 2^j(\hat{F}_V(v_0) - t_i) \leq b\},
\]

\[
S = [\min(V_i) + \omega, \max(V_i) - \omega] \text{ where } \omega \to 0.
\]

The set \( S \) excludes two endpoints of the support \( V \) and we are implicitly assuming one discontinuous point\(^{17} \) in the conditional mean model.

### 6 Numerical analysis

This section presents results from a Monte Carlo simulation study\(^{18} \). We focus on the finite sample performances of our local least squares wavelet estimators in the auxiliary regressions, although it is desirable to begin with the switching model.(1) and (2).

#### 6.1 Wavelet Jump Size Estimator under the Discontinuous Incentive Assignment Mechanism

The auxiliary model for the jump size is

\[
Y = \begin{cases} 
V + V^2 + W, & 0 \leq V < 0.5 \\
1 + 2V + 3V^2 + W, & 0.5 \leq V \leq 1
\end{cases}
\]

\(^{17}\)The more involved question would be to decide the number of the discontinuous points in the underlying curve (see Muller and Stadtmüller, 1999).

\(^{18}\)All the codes are programmed under R and available upon readers’ request.
where $V \sim U\left[0, 1\right]$ and $W|V \sim N(0, 0.01)$. And the above auxiliary (20) satisfies Assumption Set D and A
with the jump size $\delta_0$ being 2, the kink size $\delta_1$ being 3 and the second derivative jump size $\delta_2$ being 4 at the discontinuous location $V = 0.5$.

Let examine the finite sample performance of $\hat{\delta}_{0}^{(p)}$ under different sample size \{500, 2500, 5000\} with D(4)
nother wavelet and 250 times simulations. In addition, we also carry out four estimators $\hat{\delta}_{0}^{(p)}(p)$ with the
different discontinuous order $p$ in equation (8) These four estimators are: the zeta2, $\hat{\delta}_{0}^{(2)}(2)$, with the exact
highest discontinuous order $p = 2$; the zeta1, $\hat{\delta}_{0}^{(1)}(1)$, with the underspecification of the discontinuous order
$p = 1$; the zeta0, $\hat{\delta}_{0}^{(0)}(0)$, with the underspecification of the discontinuous order $p = 0$ and indeed this is $\hat{\delta}_{0}^{OLS}$;
at the last, the zeta3, $\hat{\delta}_{0}^{(3)}(3)$, with the overspecification of the discontinuous order $p = 3$.

Several observations are in order. First among all the different sample size \{500, 2500, 5000\}, the zeta0,
$\hat{\delta}_{0}^{(0)}(0)$, performs the worst because it does not consider any other derivative jump sizes, except the only jump
size. Notice that when the scale parameter $j_0$ is small, the finite sample MSE of $\hat{\delta}_{0}^{(0)}(0)$ is the largest due
to its poor asymptotic bias, while after some threshold for the scale parameter all these four estimators
perform similarly due to the undersmoothing. See Figure 1, 3, 5. Second we find that the zeta1, $\hat{\delta}_{0}^{(1)}(1)$, has
the best finite sample MSE even without considering the second order derivative jump. Although in theory
using $p = 2$ is optimal, the theoretical improvement for $p = 2$ is not generally noticeable in finite samples
compared to $p = 1$. See Marron and Wand (1992) for a similar argument. The improvement from $p = 0$ to
$p = 1$ is quite significant due to the reduced asymptotic bias order, while the improvement from $p = 1$ to
$p = 2$ is trivial and even negative because of the asymptotic variance increasing with $p$. See Figure 2, 4, 6.

To check the robustness of our local least squares wavelet estimator, we implement the following scenarios
and the results are encouraging.

(1) the distribution of $W|V$ follows a multivariate studentized t distribution (Heckman, Tobias and
Vytlaclil, RES, 2003) with parameters (0, 0, $I_{2 \times 2}$); see Figure 7.

(2) the conditional variance function $Var(W|V)$ is heteroskedastic with $W|V \sim N(0, 0.01 \times V^2)$; see
Figure 8.

(3) the the forcing variable $V$ could be either a random design or fixed design, such as, exponential
distribution with the parameter 2, norm distribution $N(0.5, 0.1^2)$ and the beta distribution with parameters
(1, 1, 0); see Figure 9, 10, 11.

(4) we allow for different signal to noise level, such as, $W|V \sim N(0, 0.1^2), W|V \sim N(0, 0.2^2), W|V \sim
N(0, 0.4^2)$ and $W|V \sim N(0, 0.6^2)$; see Figure 12.

(5) different vanishing moment wavelets $\psi$ are checked. For example, we use the Daubechies \{4, 6, 8\}
wavelet functions; see Figure 13.

(6) different support-shrinking functions $\hat{D}_{j_0}(\cdot)$ in the equation (8) are selected, such as, Epanechnkov
kernel with $h = 2^{-j_0}$ and Gaussian kernel $h = 2^{-j_0}$; see Figure 14.

(7) we also calculate the MSE on the interval $[0.1, 0.9]$ in order to see whether $\hat{\delta}_{0}$ would provide a
reasonable estimate for these continuou locations. Here we estimate these MSE under two scales $j_0 = 4.1
and \( j_0 = 5.1 \); see Figure 15.

(8) perturb (20) by adding an additive sine function \( \sin(10(v - 0.5)) \) such that the response function has a finer structure; see Figure 16.

### 6.2 Wavelet Kink Size Estimator under the Kink Incentive Assignment Mechanism

The auxiliary model for the kink size is

\[
Y = \begin{cases} 
V - 0.5 + W, & 0 \leq V < 0.5 \\
10(V - 0.5) + W, & 0.5 \leq V \leq 1
\end{cases}
\] (21)

where \( V \sim U[0, 1] \) and \( W|V \sim N(0, 0.02^2) \). And the above auxiliary (20) satisfies Assumption Set K and B with the kink size \( \delta_1 \) being 9.

Let examine the finite sample performance of \( \hat{\delta}_1^K \) under the sample size 500 with D(4) mother wavelet and 250 times simulations. Here we provide four estimators which are divided into two categories: one is to use the \( \hat{\delta}_J \) in equation (8) for estimating the kink size, and the other is to apply \( \hat{\delta}_K \) in equation (18). For the first category \( \hat{\delta}_J \), we have the kink_012 \( \hat{\delta}_J^1(2) \) with \( p = 2 \), and the kink_01 \( \hat{\delta}_J^1(1) \) with \( p = 1 \); while the kink_12 \( \hat{\delta}_K^1(2) \) with the overspecification discontinuous order \( p = 2 \), and the kink_12 \( \hat{\delta}_K^1(1) \) with the correct discontinuous order \( p = 1 \). Overall all these four estimators for the kink size perform well, although there are some variations towarding to the large scale \( j_0 \). See Figure 17.

### 7 Conclusion

This paper presents the local least squares wavelet estimator for the LATE, respectively under the discontinuous (kink) incentive assignment mechanism. The proposed estimator offers an unified framework to connect and compare various jump (kink) size estimators. And we show that the local least squares wavelet estimator possess the optimal convergence rate under a broader set of conditions and obtains other order derivative jump sizes at the same time.

In future, we are going to adopt the wavelet estimation strategy from the discontinuous conditional mean model to the discontinuous conditional density model (Chernozhukov and Han, 2004). This is important, because in general a discontinuous conditional density \( f_{Y|V}(y) \) is raised either from the discontinuous conditional mean function \( h(V) + \delta_0 I\{V \geq v_0\} \) or a discontinuous conditional density function \( f_{e|V}(e) \). At the moment, we are working on the local likelihood estimation (Loader, 1996) based on the wavelet coefficients. Moreover we would like to extend the univariate LATE identification and estimation to the multivariate scenario.
Appendix A: Important facts about the wavelet

This appendix serves to review some basic and useful facts used in this paper. For a comprehensive study, the readers should refer to Daubechies (1992) and Mallat (2009).

Wavelet vanishing moment:

\[ \int t^k \psi(t) dt = 0 \text{ for } 0 \leq k < m. \]

Hence a wavelet with \( m \) vanishing moments is orthogonal to polynomials of degree \( m-1 \): suppose the polynomial \( p \) has degree at most \( m-1 \), then

\[ Wp(t) = \int p(t) \frac{1}{\sqrt{s}} \psi\left( \frac{t-u}{s} \right) dt = 0. \]

**THEOREM 7.1** (Mallat, Theorem 6.4) If \( f \in L^2(R) \) is Lipschitz \( \alpha \leq n \) at \( v \), then there exists \( A \) such that

\[ \forall (u, s) \in R \times R^+, |Wf(u, s)| \leq As^{\alpha+1/2} \left( 1 + \frac{|u-v|}{s} \right)^\alpha. \]

Conversely, if \( \alpha < n \) is not an integer and there exist \( A \) and \( \alpha' < \alpha \) such that

\[ \forall (u, s) \in R \times R^+, |Wf(u, s)| \leq As^{\alpha+1/2} \left( 1 + \frac{|u-v|}{s} \right)^{\alpha'}, \]

then \( f \) is Lipschitz \( \alpha \leq n \) at \( v \).

Appendix B: Proof of Theorem 2.1

For notational compactness, we might use \( \psi_j[\cdot] \equiv 2^j \psi [2^j \cdot] \) throughout the paper.

The proofs rely heavily on Theorem 1 in Yang (1981). For completeness, we restate it in Lemma 7.1 below. Note that we need to extend Theorem 1 in Yang (1981) to allow the function \( J \) to depend on \( n \) as in Remark 2 in Yang (1981), and also extend to the vector-valued scenario. Let \( (X_i, Y_i) \ (i = 1, 2, \ldots, n) \) be independent and identically distributed as \( (X, Y) \). The \( r \)-th ordered \( X \) variate is denoted by \( X_{r:n} \) and the \( Y \) variate paired with it is denoted by \( Y_{r:n} \). Let

\[ S_n = n^{-1} \sum_{i=1}^n J(i/(n+1)) Y_{i:n}, \]

where \( J \) is some bounded smooth function and may depend on \( n \). Further let

\[ m(x) \equiv E(Y|X = x), \sigma^2(x) \equiv Var(Y|X = x), \]

\[ F^{-1}(u) \equiv \inf\{x|F(x) \geq u\}, \ m \circ F^{-1}(u) \equiv m(F^{-1}(u)). \]

**Lemma 7.1** Suppose the following conditions are satisfied: \( E(Y^2) < \infty; \ m(x) \) is a right continuous function of bounded variation in any finite interval; \( J \) is bounded and continuous ae \( m \circ F^{-1}; \) and the cdf of \( X \), \( (F(x)) \) is a continuous function. Let

\[ \sigma^2 = \int_{-\infty}^{+\infty} J^2(F(x)) \sigma^2(x) dF(x) \]

\[ + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [F(x \land y) - F(x)F(y)] \times J(F(x)) J(F(y)) dm(x) dm(y). \]
Then

\[
\lim_{n \to \infty} n\operatorname{Var}(S_n) = \sigma^2
\]

and

\[
\lim_{n \to \infty} E(S_n) = \int_{-\infty}^{+\infty} m(x)J(F(x))dF(x)
\]

Furthermore if \(\sigma^2 > 0\), then

\[
\frac{S_n - E(S_n)}{\sqrt{\operatorname{Var}(S_n)}} \xrightarrow{d} N(0,1).
\]

**Lemma 7.2** (Extension to the \(Q\) vector-valued estimators \(S_n\)) Suppose the above conditions are satisfied and let

\[
S_n = \left[ S_n^{[1]}, S_n^{[2]}, \ldots, S_n^{[Q]} \right]^T,
\]

where for each \(1 \leq q \leq Q\) and \(J^{[q]}(\cdot)\) may depend on \(n\):

\[
S_n^{[q]} = n^{-1} \sum_{i=1}^n J^{[q]}(i/(n+1)) Y_i.n.
\]

Define for any \(1 \leq q_1, q_2 \leq Q\),

\[
\sigma^2_{(q_1, q_2)} = \int_{-\infty}^{+\infty} J^{[q_1]}(F(x)) \cdot J^{[q_2]}(F(x)) \cdot \sigma^2(x)dF(x)
\]

\[
+ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [F(x \wedge y) - F(x)F(y)] \times J^{[q_1]}(F(x))J^{[q_2]}(F(x))dm(x)dm(y).
\]

then

\[
\lim_{n \to \infty} E(S_n)
\]

\[
= \left[ \int_{-\infty}^{+\infty} m(x)J^{[1]}(F(x))dF(x), \int_{-\infty}^{+\infty} m(x)J^{[2]}(F(x))dF(x), \ldots, \int_{-\infty}^{+\infty} m(x)J^{[Q]}(F(x))dF(x) \right]^T
\]

and

\[
\lim_{n \to \infty} n\operatorname{Var}(S_n) = \left[ \sigma^2_{(q_1, q_2)} \right]_{Q \times Q}.
\]

Furthermore if \(\left[ \sigma^2_{(q_1, q_2)} \right]_{Q \times Q}\) is positive definite, then

\[
[\operatorname{Var}(S_n)]^{-1/2} \left[ S_n - E(S_n) \right] \xrightarrow{d} N(0, I).
\]

**Proof:** For the asymptotic bias part, it is straightforward from the original theorem because of the closed-form expression. And we would apply Cramer-Wold device for deriving its asymptotic variance part.

**Lemma 7.3** Let

\[
M = \left[
\begin{array}{cccc}
\int_0^1 \hat{\Delta}_{j_0}^{d_0}(t) \Delta_{j_0}^{d_0}(t)D_{j_0}(t)dt & \int_0^1 \hat{\Delta}_{j_0}^{d_1}(t) \Delta_{j_0}^{d_1}(t)D_{j_0}(t)dt & \ldots & \int_0^1 \hat{\Delta}_{j_0}^{d_0}(t) \Delta_{j_0}^{d_p}(t)D_{j_0}(t)dt \\
\int_0^1 \hat{\Delta}_{j_0}^{d_0}(t) \Delta_{j_0}^{d_1}(t)D_{j_0}(t)dt & \int_0^1 \hat{\Delta}_{j_0}^{d_1}(t) \Delta_{j_0}^{d_1}(t)D_{j_0}(t)dt & \ldots & \int_0^1 \hat{\Delta}_{j_0}^{d_1}(t) \Delta_{j_0}^{d_p}(t)D_{j_0}(t)dt \\
\vdots & \vdots & \ddots & \vdots \\
\int_0^1 \hat{\Delta}_{j_0}^{d_0}(t) \Delta_{j_0}^{d_p}(t)D_{j_0}(t)dt & \int_0^1 \hat{\Delta}_{j_0}^{d_p}(t) \Delta_{j_0}^{d_1}(t)D_{j_0}(t)dt & \ldots & \int_0^1 \hat{\Delta}_{j_0}^{d_p}(t) \Delta_{j_0}^{d_p}(t)D_{j_0}(t)dt
\end{array}
\right]
\]
Then
\[ M = \text{diag} \left[ \frac{1}{2^{j_0}}, \frac{1}{2^{j_0}}, \ldots, \frac{1}{2^{(p+1)j_0}} \right] \cdot M^* \cdot \text{diag} \left[ \frac{1}{2^{j_0}}, \frac{1}{2^{j_0}}, \ldots, \frac{1}{2^{(p+1)j_0}} \right] + \text{s.o.}, \]
where for \( 0 \leq i, j \leq p, \)
\[ \left[ M_{(i,j)} \right]_{(p+1) \times (p+1)} = \frac{1}{f_{W(1)}(v_0)^{p+1}} \int_a^b \int_a^b (w - t)^i (v - t)^j \psi(w) \psi(v) dw dt. \]

**Proof:** First, from the term \( T_{W1D} \) in Chen&Fan (2011)'s Theorem 4.1:
\[
\int_0^1 \left[ \Delta_{j_0}^{d_0}(t) \right]^2 D_{j_0}(t) dt = \frac{1}{2^{2j_0}} \int_{a-b}^0 M(v) dv + \text{s.o.},
\]
where \( M(v) = \int_a^b \int_a^b I\{w \geq t + v\} \psi(w) \psi(t) dw dt. \)

Second, from the term \( T_{K1D} \) in Chen&Fan (2011)'s Theorem 4.7:
\[
\int_0^1 \left[ \Delta_{j_0}^{d_1}(t) \right]^2 D_{j_0}(t) dt = \frac{1}{2^{2j_0} f_V^2(v_0)} \int_{a-b}^0 \left[ 2^j(t) - t M_1(t) - t M_2(t) \right] dt + \text{s.o.},
\]
where
\[
M_1(s) = \int_a^b (-t) I\{w \geq t + s\} \psi(w) \psi(t) dt dw,
\]
\[
M_2(s) = \int_a^b w I\{w \geq t + s\} \psi(w) \psi(t) dtdw.
\]

For the general term \( \int_0^1 \Delta_{j_0}^{d_1}(t) \Delta_{j_0}^{d_1}(t) D_{j_0}(t) dt : \)
\[
\int_0^1 \Delta_{j_0}^{d_1}(t) \Delta_{j_0}^{d_1}(t) D_{j_0}(t) dt = \int \int D_{j_0}(t) \left[ F_V^{-1}(w) - v_0 \right]^i \psi(w) \psi(t) dt dw + \text{s.o.}
\]
\[
= \frac{1}{2^{2j_0}} \int \int \left[ F_V^{-1}(w - t) + v_0 \right]^i \psi(w - t) \psi(t) dt dw + \text{s.o.}
\]
\[
= \frac{1}{2^{2j_0}} \int \int \left[ (w - t) / 2^{j_0} \right] \psi(w - t) \psi(t) dt dw + \text{s.o.}
\]
\[
= \frac{1}{2^{(j+j+2)j_0}} \int \int \left[ (w - t) / 2^{j_0} \right] \psi(w - t) \psi(t) dt dw + \text{s.o.}
\]
where the s.o. term in first equality comes from replacing the finite double summations with integration, whose precision could be controlled by the Koksma–Hlawka inequality. Q.E.D.

**Lemma 7.4** Define
\[
\Delta_{j_0}^{H^*}(t) = \frac{1}{H^*(t_m)} \left[ 2^{j_0} \right] \psi(2^{j_0} (t_m - t)),
\]
\[ \cdot \]
and let

\[
N \equiv \left[ \int_0^1 \Delta_{j_0}^{d_0}(t)D_{j_0}(t)\Delta_{j_0}^{H^*}(t)dt \right]
\]

\[
\vdots
\]

\[
\int_0^1 \Delta_{j_0}^{dp}(t)D_{j_0}(t)\Delta_{j_0}^{H^*}(t)dt
\]

Then

\[
N = \frac{1}{2(2m+1)j_0}N^* + \text{s.o.},
\]

where

\[
\left[ N^* \right]_{(p+1)\times 1} \equiv \frac{1}{m!(m-1)!} H^{(2m-1)}(\tau) \int_a^b \psi(u)u^mdu \cdot \int_a^b I\{w - t \geq 0\}(-t)^{m-1}\psi(w)dtdw,
\]

and for \(1 \leq i \leq p\) and \(K_i(\frac{u}{2j_0}, \frac{w}{2j_0}, \frac{t}{2j_0}) \equiv H^*(\frac{u-t}{2j_0} + \tau) \cdot \left[ F_{V_i}^{-1}(\frac{u-t}{2j_0} + \tau) - v_0 \right]^i\)

\[
\left[ N^* \right]_{(i)\times 1} \equiv \int_a^b \sum_{\alpha_2 + \alpha_3 = m-1} \frac{1}{m!\alpha_2!\alpha_3!} \frac{\partial^{m-1}K_i(k_1,k_2,k_3)}{\partial^{m-1}k_1^{\alpha_2}k_2^{\alpha_3}k_3^{\alpha_3}} |k_1=k_2=k_3=0| u^m w^{\alpha_2}(-t)^{\alpha_3}dwdudt.
\]

**Proof:** First let us look at the \(\int_0^1 \Delta_{j_0}^{d_0}(t)D_{j_0}(t)\Delta_{j_0}^{H^*}(t)dt\) term:

\[
\int_0^1 \Delta_{j_0}^{d_0}(t)D_{j_0}(t)\Delta_{j_0}^{H^*}(t)dt
\]

\[
= \int_0^1 H^*(u)I\{w \geq \tau\}D_{j_0}(t)2^{j_0}\psi\left[2^{j_0}(w-t)\right] \psi\left[2^{j_0}(u-t)\right] dwdudt + \text{s.o.}
\]

\[
= \frac{1}{2^{2j_0}} \int_a^b H^*(\frac{u-t}{2j_0} + \tau)I\{w - t \geq 0\}\psi(w)\psi(u)dwdudt + \text{s.o.}
\]

\[
= \frac{1}{2^{2j_0}} \int_a^b \left[ H^*(\tau) + \sum_{k=1}^{2m-1} \frac{H^*(k)(\tau)}{k!} \left(\frac{u-t}{2j_0}\right)^k \right] I\{w - t \geq 0\}\psi(w)\psi(u)dwdudt + \text{s.o.}
\]

\[
= \frac{1}{2^{2j_0}} \int_a^b \frac{1}{(2m-1)!} H^{(2m-1)}(\tau) \left(\frac{u-t}{2j_0}\right)^{2m-1} I\{w - t \geq 0\}\psi(w)\psi(u)dwdudt + \text{s.o.}
\]

\[
= \frac{1}{2^{(2m+1)j_0} m!(m-1)!} H^{(2m-1)}(\tau) \int_a^b \psi(s)s^m ds \cdot \int_a^b I\{w - t \geq 0\}\psi(w)(-t)^{m-1}dwdt + \text{s.o.},
\]

where the s.o. term in first equality comes from replacing the finite double summations with integration, whose precision could be controlled by the Koksma–Hlawka inequality, and the second last equality is from employing the vanishing moment \(\int_a^b u^j\psi(u)du = 0\) for \(j = 0, 1, \ldots, m - 1\), and \(\int_a^b t^j\int_a^b I\{w \geq t\}\psi(w)dw = 0^{19}\) for \(j = 0, 1, \ldots, m - 2\).

---

\(^{19}\)See Heng and Fan (2011) Lemma C.1.
For then general term \( \int_{0}^{1} \Delta_{j_0}^{d_i}(t) D_{j_0}(t) \Delta_{j_0}^{H^*}(t) \) where \( 1 \leq i \leq p \):

\[
\int_{0}^{1} \Delta_{j_0}^{d_i}(t) D_{j_0}(t) \Delta_{j_0}^{H^*}(t) dt
= \int_{0}^{1} \left[ F_{V}^{-1}(w) - v_0 \right]^{i} D_{j_0}(t) 2^{j_0} \psi \left[ 2^{j_0}(w - t) \right] \psi \left[ 2^{j_0}(u - t) \right] dwdudt + s.o.
\]

\[
= \frac{1}{2^{2j_0}} \int_{a}^{b} \int_{0}^{1} H^{*}(\frac{w - t}{2^{j_0}} + \tau) \left[ F_{V}^{-1}(\frac{w - t}{2^{j_0}} + \tau) - v_0 \right]^{i} I\{w - t \geq 0\} \psi(w) \psi(u) dwdudt + s.o.
\]

\[
= \frac{1}{2^{(2m+1)j_0}} \int_{a}^{b} \sum_{\alpha_2 + \alpha_3 = m} \frac{1}{m! \alpha_2! \alpha_3!} \frac{\partial^{2m-1} K_i(k_1, k_2, k_3)}{\partial^{\alpha_2} k_1 \partial^{\alpha_3} k_2 \partial^{\alpha_3} k_3} \left|_{k_1 = k_2 = k_3 = 0} \right. u^{m} w^{\alpha_2} (-t)^{\alpha_3} dwdudt + s.o.,
\]

where the last equality comes from the trivariate Taylor expansion of \( H^{*}(\frac{w - t}{2^{j_0}} + \tau) \left[ F_{V}^{-1}(\frac{w - t}{2^{j_0}} + \tau) - v_0 \right]^{i} \equiv K_i(\frac{w}{2^{j_0}}, \frac{w - t}{2^{j_0}}, \frac{u - t}{2^{j_0}}) \) and then apply the vanishing moment of \( \psi(w) \) and \( L(t) \equiv \int_{a}^{b} I\{w \geq t\} \psi(w) dw). \ \ \ \ \textbf{Q.E.D}

Lemma 7.5 Define for \( 0 \leq i, j \leq p \)

\[
V_{(i,j)} = \int_{0}^{1} Z_i(u) Z_j(u) \sigma^{2} \left[ F_{V}^{-1}(u) \right] du,
\]

where

\[
Z_i(u) \equiv 2^{j_0} \int_{0}^{1} D_{j_0}(t) I\{w \geq \tau\} \psi \left[ 2^{j_0}(w - t) \right] \psi \left[ 2^{j_0}(u - t) \right] dwdt.
\]

Then

\[
V_{(i,j)} = \frac{1}{2^{(3+j+i)j_0}} V_{(i,j)}^{*} + s.o.,
\]

where

\[
\left[ V_{(i,j)}^{*} \right]_{(p+1) \times (p+1)} = \sigma^{2} (v_0) \int_{a}^{b} \left[ \int_{a}^{b} I\{w - t \geq 0\} (w - t)^{i} \psi(w) \psi(u + t) \right] dwdt
\]

\[
+ \sigma^{2} (v_0) \int_{a}^{b} \left[ \int_{a}^{b} I\{w - t \geq 0\} (w - t)^{i} \psi(w) \psi(u + t) \right] dwdt
\]

\[
\left[ \int_{a}^{b} I\{w - t \geq 0\} (w - t)^{i} \psi(w) \psi(u + t) \right] dwdt
\]

28
Proof: First let us look at the term \( V_{(0,0)} \), where \( S(u) \equiv \int_a^b I\{w - t \geq 0\} \psi(w) \psi[2^{j_0}u - 2^{j_0} \tau + t] dw dt \):

\[
V_{(0,0)} = \int_0^1 \left[ \int_a^b I\{w - t \geq 0\} \psi(w) \psi[2^{j_0}u - 2^{j_0} \tau + t] \frac{1}{2^{j_0}} dw dt \right]^2 \sigma^2 [F_{\psi}^{-1}(u)] du
\]

\[
= \frac{1}{2^{2j_0}} \int_0^1 [S(2^{j_0}u - 2^{j_0} \tau)]^2 \sigma^2 [F_{\psi}^{-1}(u)] du
\]

\[
= \frac{1}{2^{2j_0}} \int_a^b S^2(u) \sigma^2 [F_{\psi}^{-1}(\frac{u}{2^{j_0}} + \tau)] du
\]

\[
= \frac{1}{2^{2j_0}} \left[ \sigma_+^2(v_0) \int_0^b S^2(u) du + \sigma_-^2(v_0) \int_0^0 S^2(u) du + s.o. \right]
\]

\[
= \frac{1}{2^{2j_0}} \left[ \sigma_+^2(v_0) \int_0^b S^2(u) du \left( \int_a^b I\{w - t \geq 0\} \psi(w) \psi(u + t) dw dt \right) du \right] + s.o.,
\]

where the last equality comes from the fact that \( \int_a^b I\{w \geq t + v\} \psi(w) \psi(t) dw dt \) has compact support \([a - b, b - a]\).

By the same procedure, we could prove for other \( V_{(i,j)} \) for \( 0 \leq i, j \leq p \). Q.E.D

Lemma 7.6 (Asymptotic equivalence between the feasible estimator and its counterpart) Observe the local least square wavelets estimator \( \hat{\delta}_J \):

\[
\hat{\delta}_J = \arg \min_{\delta_0, \ldots, \delta_p} \sum_{l=1}^n \left[ \hat{\Delta}_{j_0}^Y(t_l) - \sum_{k=0}^p \delta_k \cdot \hat{\Delta}_{j_0}^d(t_l) \right]^2 \hat{D}_{j_0}(t_l)
\]

and its infeasible counterpart \( \overline{\delta}_J \):

\[
\overline{\delta}_J = \arg \min_{\delta_0, \ldots, \delta_p} \sum_{l=1}^n \left[ \hat{\Delta}_{j_0}^Y(t_l) - \sum_{k=0}^p \delta_k \cdot \hat{\Delta}_{j_0}^d(t_l) \right]^2 D_{j_0}(t_l)
\]

where

\[
D_{j_0}(t_l) = I\{a \leq 2^i(\tau - t_l) \leq b\},
\]

\[
(\hat{\Delta}_{j_0}^d)^T = \left[ (\hat{\Delta}_{j_0}^d(t_1))^T, (\hat{\Delta}_{j_0}^d(t_2))^T, \ldots, (\hat{\Delta}_{j_0}^d(t_n))^T \right]^T,
\]

\[
D_{j_0} = \text{diag}[D_{j_0}(t_1), D_{j_0}(t_2), \ldots, D_{j_0}(t_n)].
\]
When \( n^{-1/2} \left( \log \log n \right)^{1/2} / (2^{-j_0}) \to 0 \), then
\[
\sqrt{\text{diag} \left[ \frac{n}{2^{j_0}}, \frac{n}{2^{2j_0}}, \ldots, \frac{n}{2^{(p+1)j_0}} \right]} \left( \delta - \hat{\delta} \right) = o_p(I_{(p+1)\times 1}).
\]

**Proof:** First, let us look at the numerator terms \( \left( \hat{d}^{\Delta}_{j_0} \right)^T \cdot \hat{D}_{j_0} \cdot \hat{\Delta}^Y_{j_0} \) and \( \left( d^{\Delta}_{j_0} \right)^T \cdot D_{j_0} \cdot \Delta^Y_{j_0} \) in these two estimators. Then the corresponding element-wise difference between these two numerators is defined as \( \text{diff}_k \) for \( 0 \leq k \leq (p+1) \)
\[
\text{diff}_k = \sum_{i=1}^{n} \left( \frac{1}{n} \sum_{i=1}^{n} (F^{-1}(t_i) - v_0)^k I \{ t_i \geq \tau \} 2^{j_0/2} \psi \left[ 2^{j_0} (t_i - t_t) \right] \right) D_{j_0}(t_i) \hat{\Delta}^Y_{j_0}(t_i)
- \sum_{i=1}^{n} \left( \frac{1}{n} \sum_{i=1}^{n} (\hat{F}^{-1}(t_i) - v_0)^k I \{ t_i \geq \hat{\tau} \} 2^{j_0/2} \psi \left[ 2^{j_0} (t_i - t_t) \right] \right) \hat{D}_{j_0}(t_i) \hat{\Delta}^Y_{j_0}(t_i).
\]
Notice that
\[
\text{diff}_0 = \sum_{i=1}^{n} \left( \frac{1}{n} \sum_{i=1}^{n} I \{ t_i \geq \tau \} 2^{j_0/2} \psi \left[ 2^{j_0} (t_i - t_t) \right] \right) D_{j_0}(t_i) \hat{\Delta}^Y_{j_0}(t_i)
- \sum_{i=1}^{n} \left( \frac{1}{n} \sum_{i=1}^{n} I \{ t_i \geq \hat{\tau} \} 2^{j_0/2} \psi \left[ 2^{j_0} (t_i - t_t) \right] \right) \hat{D}_{j_0}(t_i) \hat{\Delta}^Y_{j_0}(t_i)
= \left\{ \begin{array}{ll}
\sum_{i=1}^{n} \left( \frac{1}{n} \sum_{i=1}^{n} I \{ t_i \geq \tau \} 2^{j_0/2} \psi \left[ 2^{j_0} (t_i - t_t) \right] \right) D_{j_0}(t_i) \Delta^Y_{j_0}(t_i)
- \sum_{i=1}^{n} \int_0^{2^{j_0}} I \{ w \geq \tau \} \psi_{j_0} \left[ w - t \right] \psi \left[ 2^{j_0} (t_i - t_t) \right] dtdw \cdot Y_{i:n}
\end{array} \right.
\]
\[
+ \left\{ \begin{array}{ll}
\sum_{i=1}^{n} \int_0^{2^{j_0}} I \{ w \geq \hat{\tau} \} \psi_{j_0} \left[ w - t \right] \psi \left[ 2^{j_0} (t_i - t_t) \right] dtdw \cdot Y_{i:n}
- \sum_{i=1}^{n} \int_0^{2^{j_0}} I \{ w \geq \hat{\tau} \} \psi_{j_0} \left[ w - t \right] \psi \left[ 2^{j_0} (t_i - t_t) \right] dtdw \cdot Y_{i:n}
\end{array} \right.
\]
\[
= O_p\left( \frac{1}{n2^{2j_0}} \right) + O_p\left( \frac{1}{n2^{2j_0}} \right).
\]
where the first two terms are \( O_p\left( \frac{1}{n2^{2j_0}} \right) \) in the last equality which is derived from the second step proof of Theorem 4.1 Heng & Fan (2011), and the third term of being \( O_p\left( \frac{1}{n2^{2j_0}} \right) \) in the last equality is from the change of variables.

Since
\[
\sup_{t_i \in D_{j_0}} \left| F^{-1}(t_i) - v_0 \right| = O\left( \frac{1}{2^{j_0}} \right)
\]
and
\[
\sup_{t_i \in D_{j_0}} \left| \hat{F}^{-1}(t_i) - v_0 \right| \leq \sup_{t_i \in D_{j_0}} \left| F^{-1}(t_i) - v_0 \right| + \sup_{t_i \in D_{j_0}} \left| \hat{F}^{-1}(t_i) - F^{-1}(t_i) \right| = O_p\left( \frac{1}{2^{2j_0}} \right)
\]
where the last equality comes from the equation (30) \( \left| \hat{F}^{-1}(t_i) - F^{-1}(t_i) \right| = O_p\left( n^{-1/2} \left( \log \log n \right)^{1/2} \right) \) in Wang and Cai (2010). Then \( \text{diff}_{k+1} = O_p\left( \frac{1}{2^{2j_0}} \text{diff}_k \right) \) for each \( 0 \leq k \leq p \).
Now we have

\[
\left[ (\hat{\Delta}_j^d)^T \cdot D_{j0} \cdot \hat{\Delta}_j^d \right]^{-1} \left[ (\hat{\Delta}_j^d)^T \cdot D_{j0} \cdot \hat{\Delta}_j^Y - (\hat{\Delta}_j^d)^T \cdot \hat{D}_{j0} \cdot \hat{\Delta}_j^Y \right]
\]

\[
= O_p \begin{bmatrix}
2^{2j_0} \text{diff} f_0 + 2^{3j_0} \text{diff} f_1 + \ldots + 2^{(p+2)j_0} \text{diff} f_p \\
2^{2j_0} \text{diff} f_0 + 2^{3j_0} \text{diff} f_1 + \ldots + 2^{(p+3)j_0} \text{diff} f_p \\
\vdots \\
2^{(p+2)j_0} \text{diff} f_0 + 2^{(p+2)j_0} \text{diff} f_1 + \ldots + 2^{(p+1)j_0} \text{diff} f_p 
\end{bmatrix}_{(p+1) \times 1}
\]

Therefore

\[
\sqrt{\text{diag} \left[ \frac{n}{2j_0}, \frac{n}{2j_0}, \ldots, \frac{n}{2(p+1)j_0} \right]}.
\]

\[
\left[ (\hat{\Delta}_j^d)^T \cdot D_{j0} \cdot \hat{\Delta}_j^d \right]^{-1} \left[ (\hat{\Delta}_j^d)^T \cdot D_{j0} \cdot \hat{\Delta}_j^Y - (\hat{\Delta}_j^d)^T \cdot \hat{D}_{j0} \cdot \hat{\Delta}_j^Y \right]
\]

\[
= O_p \left( \sqrt{\frac{2j_0}{n}} \right)_{(p+1) \times 1} = o_p(1)_{(p+1) \times 1}.
\]

In the end by Sluksty theorem and \( \left\{ (\hat{\Delta}_j^d)^T \cdot D_{j0} \cdot \hat{\Delta}_j^d - (\hat{\Delta}_j^d)^T \cdot \hat{D}_{j0} \cdot \hat{\Delta}_j^d \right\} = o_p(1)_{(p+1) \times (p+1)} \), we obtain

the asymptotic equivalence between the infeasible and feasible estimators. Q.E.D

**Proof of Theorem 2.1:** First let us look at the \((\hat{\Delta}_j^d)^T \cdot D_{j0} \cdot \hat{\Delta}_j^d\) term:

\[
(\hat{\Delta}_j^d)^T \cdot D_{j0} \cdot \hat{\Delta}_j^d
\]

\[
= \sum_{i=1}^{n} \hat{\Delta}_{j0}^{d0}(t_i) \cdot D_{j0}(t_i) \cdot \sum_{i=1}^{n} \hat{\Delta}_{j0}^{d0}(t_i) \hat{\Delta}_{j0}^{d1}(t_i) \cdot D_{j0}(t_i) \cdot \ldots \cdot \sum_{i=1}^{n} \hat{\Delta}_{j0}^{d0}(t_i) \hat{\Delta}_{j0}^{d1}(t_i) \hat{\Delta}_{j0}^{d2}(t_i) \cdot D_{j0}(t_i) \\
= \sum_{i=1}^{n} \hat{\Delta}_{j0}^{d0}(t_i) \hat{\Delta}_{j0}^{d1}(t_i) \hat{\Delta}_{j0}^{d2}(t_i) \cdot D_{j0}(t_i) \cdot \ldots \cdot \sum_{i=1}^{n} \hat{\Delta}_{j0}^{d0}(t_i) \hat{\Delta}_{j0}^{d1}(t_i) \hat{\Delta}_{j0}^{d2}(t_i) \hat{\Delta}_{j0}^{d3}(t_i) \cdot D_{j0}(t_i) \\
= \ldots
\]

\[
= \ldots
\]

\[
= \ldots
\]

\[
= M + \text{s.o.}
\]
Second, let us look at the \(\left(\hat{\Delta}^f_j\right)^T \cdot D_j \cdot \hat{\Delta}^Y_j\) term:

\[
\left(\hat{\Delta}^f_j\right)^T \cdot D_j \cdot \hat{\Delta}^Y_j = \begin{bmatrix}
\sum_{i=1}^n \hat{\Delta}^{d_0}_j(t_i)D_j(t_i)\hat{\Delta}^Y_j(t_i) \\
\vdots \\
\sum_{i=1}^n \hat{\Delta}^{d_p}_j(t_i)D_j(t_i)\hat{\Delta}^Y_j(t_i)
\end{bmatrix} = \begin{bmatrix}
\int_0^1 \hat{\Delta}^{d_0}_j(t)D_j(t)\hat{\Delta}^Y_j(t)dt \\
\vdots \\
\int_0^1 \hat{\Delta}^{d_p}_j(t)D_j(t)\hat{\Delta}^Y_j(t)dt
\end{bmatrix} + s.o.
\]

Now let us derive the asymptotic bias term using Lemma 7.3 and Lemma 7.4 and the asymptotic equivalence between the feasible and infeasible estimators from Lemma 7.6:

\[
\lim_{n \to \infty} \text{diag} \left[ 2^{(2m-1)j_0}, 2^{(2m-2)j_0}, \ldots, 2^{(2m-p-1)j_0} \right] \left[ E(\hat{\delta}_f) - \delta_f \right]
= \begin{bmatrix}
H^{(2m-1)}(\tau) \cdot (M^*)_{(0,0)}^{-1} N^*_0 \\
H^{(2m-1)}(\tau) \cdot (M^*)_{(1,0)}^{-1} N^*_0 \\
\vdots \\
H^{(2m-1)}(\tau) \cdot (M^*)_{(p,0)}^{-1} N^*_0
\end{bmatrix}_{(p+1) \times 1}
\]

For the asymptotic variance term, we employ Lemma 7.5 and the asymptotic equivalence between the feasible and infeasible estimators from Lemma 7.6 \(^{20}\):

\[
\lim_{n \to \infty} n \cdot \Xi \cdot \text{Var}(\hat{\delta}_f) = (M^*)^{-1}V^*(M^*)^{-1}
\]

In the end, the asymptotic normality of \(\hat{\delta}_0^f\) is established by following from Lemma 7.2. Q.E.D.

\(^{20}\)Notice that for the asymptotic variance term in Lemma 7.2, the second term is smaller order than the first one. The proof is shown in the Chen and Fan (2011)’s Theorem 4.1.
References


Figure 1: Four jump size estimators under 500 sample size

Figure 2: Three jump size estimators under 500 sample size
Figure 3: Four jump size estimators under 2500 sample size

Figure 4: Three jump size estimators under 2500 sample size
Figure 5: Four jump size estimators under 5000 sample size

Figure 6: Three jump size estimators under 5000 sample size
Figure 7: Jump size estimator under the multivariate student t distribution

Figure 8: Jump size estimator with heteroscedastic error
Figure 9: Jump size estimator under random design with exponential distribution

Figure 10: Jump size estimator under random design with normal distribution
Figure 11: Jump size estimator under random design with beta distribution

Figure 12: Jump size estimators under different noise levels
Figure 13: Jump size estimators with $D(4, 6, 8)$

Figure 14: Jump size estimators with different $\hat{D}_{j_0}(\cdot)$
Figure 15: Jump size estimator for the underlying curve within the interval $[0.1, 0.9]$.

Figure 16: Jump size estimator with the finer structure.
Figure 17: Four kink size estimators under 500 sample size