Identification and Wavelet Estimation of LATE in a Class of Switching Regime Models*

Heng Chen and Yanqin Fan
Department of Economics
Vanderbilt University
VU Station B #351819
2301 Vanderbilt Place
Nashville, TN 37235-1819

This version: April 2011

Abstract

This paper makes two main contributions to the econometrics literature on program evaluation. First, we show that under very mild conditions, a policy parameter, LATE, is identified in a class of switching regime models. The identification is achieved through a discontinuity or a kink in the incentive assignment mechanism depending on which the agent selects treatment according to a threshold-crossing model. In contrast to Lee (2008) and Card, Lee, and Pei (2009), we allow for not only (possibly) endogenous observable covariate but also (possibly) endogenous unobservable covariate to affect program participation. Second, we introduce several wavelet estimators of LATE for both discontinuous and kink incentive assignment mechanisms and establish their asymptotic properties. The finite sample performances of our wavelet estimators are examined through a simulation study.

Keywords: Incentive assignment mechanism; Regression discontinuity design; Regression kink design; Selection-on-unobservables; Wavelet transform

JEL codes: C13; C14; C35; C51

---

*We thank participants at the Triangle Econometrics seminar and Montreal Econometrics seminar for helpful comments.

1 Introduction

As described in Heckman (2008), "incorporating choice into the analysis of treatment effects is an essential and distinctive ingredient of the econometric approach to the evaluation of social programs," and "under a more comprehensive definition of treatment, agents are assigned incentives like taxes, subsidies, endowments and eligibility that affect their choices, but the agent chooses the treatment selected."

This paper studies a class of switching regime models to explicitly account for the role of an incentive assignment mechanism in an agent's selection of a binary treatment. Let $V \in \mathcal{V} \subset \mathcal{R}$ be a continuous random variable denoting the agent's observable covariate based on which incentives are assigned to the agent according to the incentive assignment mechanism $b : \mathcal{V} \mapsto \mathcal{R}$. Based on the incentive received $b\left(V\right)$ and her characteristic U, the agent chooses the treatment $D = 1$ or $D = 0$ with potential outcomes Y_1 (with treatment) or Y_0 (without treatment) respectively. Let

$$\begin{aligned} Y_1 &= g_1\left(V, W\right), Y_0 = g_0\left(V, W\right), && (1) \\ D &= I\{b(V) - U \geq 0\}, && (2) \end{aligned}$$

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.[1] The agent's observable covariate V affects both the potential outcomes and selection (through the incentive assignment mechanism b). The incentive assignment mechanism b is assumed to be either discontinuous at a known cut-off v_0 or differentiable with a discontinuous derivative at v_0. We refer to the latter class of incentive assignment mechanisms as kink incentive assignment mechanisms. Many incentive assignment mechanisms fall into one of these two categories. A well known example in the first category is $b\left(V\right) = I\left\{V \geq v_0\right\}$, which includes the allocation of merit awards, see Thistlethwaite and Campbell (1960), and many threshold rules often used by educational institutions to estimate the effect of financial aid and class size, respectively, on educational outcomes, see e.g., Van der Klaauw (2002) and Angrist and Lavy (1999). Lee and Lemieux (2009) provides many other such examples. Unemployment benefits assignment and the income tax system in most countries belong to the second category, see Card, Lee, and Pei (2009) for more examples.

The above switching regime model can be rewritten as a nonseparable simultaneous equations model using the individual's realized outcome: $Y \equiv DY_1 + (1 - D)\, Y_0$. The econometrician observes (V, Y, D). Let $Y = y(D, V, W)$, where $y(\cdot, \cdot, \cdot)$ is a real-valued measurable function. Then $g_1\left(V, W\right) = y(1, V, W)$ and $g_0\left(V, W\right) = y(0, V, W)$. In terms of the realized outcome Y, the poten-

---

[1] This set-up allows for the potential outcomes Y_1, Y_0 to depend on different components of W. Agent selects treatment based on the threshold-crossing model (2). As shown in Vytlacil (2006), there is a larger class of latent index models that will have a representation of this form.

tial outcomes model (1) and (2) can be written as

$$Y = y(D, V, W), \; D = I\{b(V) - U \geq 0\}. \tag{3}$$

(3) is a nonseparable structural model with an endogenous dummy variable D and a possibly endogenous continuous variable V. The endogeneity of D arises from the possible endogeneity of V and the dependence between the unobservable errors W and U. It is well known that in a general nonseparable structural model like (3) with possibly endogenous covariates V and D, it is difficult to identify the structural parameters in the model including g_1, g_0, and the conditional distribution of (W, U) given V. Often instruments and other conditions are required, see e.g., Chesher (2003, 2005) and Matzkin (2007) and references therein. However, as noted by Marschak (1953), I quote this from Heckman (2008) who refers to it as Marschak's Maxim: "For many specific questions of policy analysis, it is not necessary to identify fully specified models that are invariant to classes of policy modifications. All that may be required for any policy analysis are combinations of subsets of the structural parameters, corresponding to the parameters required to forecast particular policy modifications, which are often much easier to identify (i.e., require fewer and weaker assumptions)." Examples of important work following Marschak's Maxim include Heckman and Vytlacil (2005), Lee (2008), Florens, et al. (2008), Imbens and Newey (2009), Vytlacil and Yildiz (2007), Card, Lee, and Pei (2009), Chernozhukov and Hansen (2005), among others.

All the above-cited work except Lee (2008) and Card, Lee, and Pei (2009) make use of instruments or control variables to identify policy parameters of interest. The model in Lee (2008) is a special case of (1) and (2) in which $D = b(V) = I\{V \geq v_0\}$. Thus, the treatment selection mechanism is the same as the incentive assignment mechanism in Lee (2008) excluding the possibility of agent choosing the treatment. By allowing agent's selection of treatment to depend on her unobservable covariate U in (2), our general model is consistent with the observation that often agents assigned the same incentive choose different treatments. Card, Lee, and Pei (2009) considers the case of a known kink incentive assignment mechanism b and a continuous treatment $D = b(V)$, so the treatment assignment mechanism is the same as the known kink incentive assignment mechanism, again excluding self-selection of the agent.

The first contribution of this paper is to show that under mild conditions, a policy parameter: local average treatment effect (LATE), is identified in (3), where the source of identification is either the presence of a discontinuity or kink in the incentive assignment mechanism b. For discontinuous incentive assignment mechanisms, this result generalizes a similar result in Lee (2008) established for the case: $D = b(V) = I\{V \geq v_0\}$, by allowing for general incentive assignment mechanisms b and more importantly, for heterogenous choices among agents assigned the same incentive. For kink incentive assignment mechanisms, our result is similar to a result in Card, Lee, and Pei (2009) with several important differences: First and most important, Card, Lee, and Pei (2009) assumes that $D = b(V)$, thus excluding heterogenous choices among agents assigned the same incentive; Second, they assume the incentive assignment mechanism b is known; Third, they consider a continuous treatment instead of a binary treatment. Our identification result for discontinuous incentive assignment mechanisms is related to a similar result for regression discontinuity design (RDD) in Hahn, Todd, and van der Klaauw (2001) and our identification result for kink incentive assignment mechanisms is related to a similar result for regression kink design (RKD) in Dong (2010). Hahn, Todd, and van der Klaauw (2001) imposes smoothness conditions directly on the regression functions $E(Y_1|V = v)$ and $E(Y_0|V = v)$ and exploits certain local independence conditions to identify LATE, while Dong (2010) adopts a similar set-up. Instead, we impose smoothness conditions on the structural parameters in (3) and by exploiting the specific structure in (2), we are able to dispense with the local independence conditions.

The second contribution of this paper is to propose several nonparametric estimators of LATE using wavelets. First, we establish auxiliary regressions linking the policy parameter, LATE, to jump sizes δ and ζ in (4) and (5) for discontinuous incentive assignment mechanisms, or kink sizes δ_K and ζ_K in (6) and (7) for kink incentive assignment mechanisms. In particular, the policy parameter LATE in (3) is given by δ/ζ for discontinuous incentive assignment mechanisms and δ_K/ζ_K for kink incentive assignment mechanisms. Thus, estimating the policy parameter LATE in (3) with discontinuous/kink incentive assignment mechanisms is equivalent to estimating the jump/kink sizes of two auxiliary regressions. For discontinuous incentive assignment mechanisms, work in the recent econometrics literature on estimating LATE for RDD are applicable. These include estimators based on Nadaraya-Watson (NW) kernel regression (local constant regression) or local polynomial regression estimators of the jump sizes δ and ζ in which δ (ζ) is estimated by the difference between two kernel regression estimators or between two local polynomial regression estimators using respectively the observations to the right and to the left of the cut-off v_0, see Hahn, Todd, and van der Klaauw (2001), Porter (2003), Imbens and Kalyanaramang (2009), Ludwig and Miller (2007), and Sun (2007). Porter (2003) also proposes a partial linear estimator of LATE based on Robinson's (1998) partial linear estimators of δ and ζ and established asymptotic properties of all three estimators under general conditions allowing for conditionally heteroscedastic errors and the presence of jump discontinuities in the derivatives of the auxiliary regression functions. For kink incentive assignment mechanisms, Dong (2010) proposes to extend existing work on local linear (polynomial) estimators for RDD to RKD without establishing the corresponding asymptotic theory.

Existing work in the econometrics literature suggest that the local polynomial regression estimator appears to have the smallest asymptotic bias order among the alternative estimators and achieves the optimal rate established in Porter (2003), providing theoretical justification for the popularity of local polynomial, especially local linear estimators in applied research. In addition, for compactly supported kernels, Imbens and Kalyanaraman (2009) derive the optimal bandwidth for local linear estimators of LATE. However, it is known in the statistics literature that local polynomial estimators suffer from a serious drawback that for compactly supported kernels, the unconditional variance of a local polynomial estimator is infinite, and hence the asymptotic MSE and the asymptotic MSE optimal bandwidth are not defined, see Seifert and Gasser (1996). The afore-mentioned work on local polynomial estimators of LATE are based on asymptotic expansions of the conditional variance and conditional MSE of the local polynomial estimators. The infinite unconditional variance of local polynomial estimators may lead to their poor finite sample performance, see Seifert and Gasser (1996). Modifications have been proposed to rectify this problem, including local polynomial ridge regression (see Seifert and Gasser (1996, 2000)); local polynomial estimator using asymmetric kernels (see Chen (2002) and Cheng (2007)); and binning and transforming random design to regularly spaced fixed design (see Hall, Park, and Turlach (1998)). Hall, Park and Turlach (1998) demonstrates that in general their idea of bining and transforming the data is superior to other approaches especially when there are jumps in the regression function.

This paper proposes several wavelet estimators of jump and kink sizes or equivalently LATE in our model by combining the idea of binning and transformation in Hall, Park and Turlach (1998) and the method of wavelets. It is well known that wavelet transform coefficients of a function at a given location characterize its degree of local regularity (smoothness), so that large wavelet transform coefficients at large scales correspond to low regularity of the function at that point, see e.g., Daubechies (1992). Because of this special feature, wavelet transform coefficients have been used to detect the location of a jump point, see Wang (1995) and Antoniadis and Gijbels (1997) and more generally the location of any order of a cusp point,[2] see Abramovich and Samarov (2000), Li and Xie (2000), Raimondo (1998), and Park and Kim (2006) for random samples and Wang and Cai (2010) for long memory time series. In addition to detecting the location of a jump or cusp, Li and Xie (2000), Park and Kim (2006), and Wang and Cai (2010) present a simple estimator ($\overline{\delta}$ in our notation, see Section 3) of the jump size and establish its asymptotic distribution under conditions

---

[2] A cusp in a function g with domain $[0,1]$ is defined as follows. Consider a class of functions on $[0,1]$ with either a single jump point $\alpha = 0$ or a single cusp point $\alpha > 0$:

(a) $\mathcal{F}_{\prime}$ is a class of functions g on $[0,1]$ such that,

(i) pointwise Lipschitz irregularity at τ: $\liminf_{h\to 0} |g(\tau+h) - g(\tau-h)| > 0$ for a unique $\tau \in (0,1)$;

(ii) uniformly Lipschitz regularity except at τ: $\sup_{0<x<y<\tau} |g(x) - g(y)|/|x-y|^{\alpha'} < \infty$ and $\sup_{\tau<x<y<1} |g(x) - g(y)|/|x-y|^{\alpha'} < \infty$ for some α', $0 < \alpha' \leq 1$.

(b) $\mathcal{F}_\alpha$ $(0 < \alpha < 1)$ is a class of functions g on $[0,1]$ such that,

(i) $\liminf_{h\to 0} |g(\tau+h) - g(\tau-h)| / |h|^\alpha > 0$ for a unique $\tau \in (0,1)$;

(ii) g is differentiable on $(0,1)$ except at τ.

(c) $\mathcal{F}_\alpha$ $(\alpha \geq 1)$ is a class of functions g on $[0,1]$ such that,

(i) g is N times differentiable on $(0,1)$, where N is the largest integer part of α;

(ii) $g^{(N)} \in \mathcal{F}_{\alpha-\mathcal{N}}$.

In sum, for $g \in \mathcal{F}_\alpha$ $(\alpha \geq 0)$, a single jump point or a single cusp point τ satisfy: $\liminf_{h\to 0} \left| g^{(N)}(\tau+h) - g^{(N)}(\tau-h) \right| / |h|^{\alpha-N} > 0$.

For $\alpha = 0$, τ is a jump point of g; for $0 < \alpha < 1$; τ is a cusp of g; for $\alpha = 1$, τ corresponds to our definition of a kink point in g; for a general, interger α, τ is a jump point in the α-th derivative of g.

including homoscedastic errors and undersmoothing. To the best of the authors' knowledge, the method of wavelets has not been used to estimate kink size.[3] Given the close connection between the estimation of LATE in (3) and of jump/kink sizes in the corresponding auxiliary regressions, it seems natural to exploit this special feature of wavelets to estimate LATE. The second part of this paper accomplishes this objective.

For discontinuous incentive assignment mechanisms, the first wavelet estimator of LATE we propose makes use of wavelet estimators of the jump sizes δ and ζ similar to that of Park and Kim (2006). We motivate our estimator using the representation of the auxiliary regressions in the wavelet domain. In addition, we establish the asymptotic distribution of our estimator under more general conditions than Park and Kim (2006). First, we allow for conditionally heteroscedastic errors in the auxiliary regressions, and second *we allow for the presence of jump discontinuities in the derivatives of the auxiliary regression functions at the known cut-off point* v_0. Like the estimator of Park and Kim (2006), our first estimator makes use of only one wavelet transform coefficient corresponding to the location v_0 and a given scale. The representation of each auxiliary regression in the wavelet domain corresponding to different locations and scales suggests that the wavelet transform coefficients at locations close to v_0 and relatively large scales may also contain information on the jump size which motivate our subsequent wavelet estimators of the jump sizes and of the LATE parameter. Specifically, we propose three new wavelet estimators of the jump sizes δ and ζ using wavelet coefficients at locations close to v_0 and/or more than one scale: the single scale estimator making use of wavelet transform coefficients at a single scale and more than one location; the single location estimator making use of wavelet coefficients at one location v_0 and more than one scale; and the multiple scale and multiple location estimator. We call our new wavelet estimators wavelet OLS estimators. We establish their asymptotic properties and the asymptotic properties of estimators of the LATE parameter based on them. The asymptotic results confirm that indeed our wavelet OLS estimators using more than one wavelet coefficients have better asymptotic properties than the single coefficient wavelet estimator currently available in the literature. In particular, our wavelet OLS estimator using more than one location reduces the order of the asymptotic bias and the estimator using more than one scale reduces the asymptotic bias proportionally. A simulation study investigates the finite sample performance of the proposed wavelet estimators and confirms our theoretical findings. It reveals the best overall performance by the wavelet OLS estimator based on more than one scale and more than one location.

All the wavelet estimators of LATE proposed for discontinuous incentive assignment mechanisms have analogues for kink incentive assignment mechanisms and share similar properties. For space considerations, we only provide asymptotic properties of the wavelet estimator based on one wavelet coefficient only and the wavelet OLS estimator based on single scale and many locations.

---

[3] Zhou, et al. (2010) provide an estimator of the cusp size for $0 < \alpha < 1$.

The rest of this paper is organized as follows. In section 2, we establish conditions under which LATE is identified in (3) and conditions under which the auxiliary regressions hold for both discontinuous and kink assignment mechanisms. Section 3 presents our first wavelet estimators of LATE for both discontinuous and kink incentive assignment mechanisms. Under regularity conditions, we establish their asymptotic distributions allowing for conditional heteroscedasticity and for the presence of jump discontinuity in the derivatives of auxiliary regression functions at v_0 for discontinuous assignment incentive mechanisms and in the higher derivatives of auxiliary regression functions for kink incentive assignment mechanisms. Motivated by the wavelet representations of the auxiliary regressions, we propose three additional wavelet estimators of LATE for discontinuous assignment mechanisms and establish their asymptotic distributions in Section 4. For kink incentive assignment mechanisms, we propose and establish the asymptotic distribution of the single scale wavelet OLS estimator. Section 5 presents results from a Monte Carlo simulation study investigating the finite sample performance of our wavelet estimators. The final section concludes the paper and outlines some future research. Technical proofs are relegated to Appendices A, B, and C.

We close this section by briefly reviewing some work in the statistics literature on jump/kink detection and their size estimation. While nonparametric estimation of LATE in RDD or RKD is a relatively new topic in econometrics, nonparametric detection and estimation of the location and size of a jump/kink of a regression function has a long history in statistics. In fact, all three approaches (NW, partial linear and local polynomial estimators) in existing work on estimating LATE in RDD have been used to detect/estimate jump/kink locations and sizes in early work in statistics. One important difference is that most work in statistics focus on fixed, equally spaced design and homoscedastic errors (some on normal errors). We mention a few papers here and refer the interested reader to references therein. First, work using differences between two kernel estimators include Muller (1992) in which he constructs estimators of both jump and kink sizes and established their asymptotic distributions for random samples.[4] In fact, Muller (1992) employs boundary kernels to overcome the well-known boundary problem associated with standard kernel estimators. Second, Eubank and Whitney (1989) proposes a partial spline estimator of the kink size and establishes the lower bound for its rate of convergence. Similar partial spline idea can be found in Koo (1997) for detecting change point. Eubank and Speckman (1994) proposes a partial linear (kernel) estimator[5] of the kink size and established its asymptotic distribution, and Cline, et al. (1995) extends the partial linear (kernel) estimator to a more general framework including the presence of discontinuity in any order of derivatives of the regression function. Third, for the use of difference between two local polynomial estimators, we refer the reader to Loader (1996), Qiu and

---

[4] Delgado and Hidalgo (2000) extends estimators of Muller (1992) to time series model.

[5] Shiaua and Wahba (1988) and Eubank and Speckman (1994) contrast the partial spline and partial linear (kernel) methods under various smoothness conditions on the Fourier transform of the nonparametric function in a partial linear model.

Yandell (1998), and Bowman, et al. (2006) for detecting the jump point; Gijbels and Goderniauxa (2005) for detecting the kink point; and Gao, et al. (1998), Spokoiny (1998), Gijbels, et al. (1999, 2007), and Desmet and Gijbels (2009) for adaptively estimating the regression curve with a jump point.

2 Identification and Auxiliary Regressions

There are two parts in this section for discontinuous and kink incentive assignment mechanisms respectively. In each part, we first provide conditions under which LATE is identified in (3) and then establish auxiliary regressions that will be used to estimate the identified LATE in Sections 3 and 4.

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 following notation: $\int \cdot du = \int_{\mathcal{U}} \cdot du$, $\int \cdot dw = \int_{\mathcal{W}} \cdot dw$, and $\int \cdot dv = \int_{\mathcal{V}} \cdot dv$. In addition, $F_{A|B}(a|b)$ and $f_{A|B}(a|b)$ denote respectively the conditional distribution function and conditional density function of A given $B = b$; all the limits are taken as the sample size n goes to ∞ unless stated otherwise.

2.1 Discontinuous Incentive Assignment Mechanism

2.1.1 Identification

The following conditions will be used to prove identification of LATE.

Condition D1. Assume (i) $f_{V|W}(v|w)$ is continuous and strictly positive at $v = v_0$ for every $w \in \mathcal{W}$; (ii) $f_V(v)$ is continuous and strictly positive at $v = v_0$; (iii) $f_{V|W,U}(v|w,u)$ is continuous and strictly positive at $v = v_0$ for every $u \in \mathcal{U}$ and $w \in \mathcal{W}$.

Condition D2. Assume $g_1(v, w)$ and $g_0(v, w)$ are continuous at $v = v_0$ for every $w \in \mathcal{W}$.

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

Condition D4. (i) Assume $b(v)$ is an increasing and continuous function in a neighborhood of v_0 except at v_0 and is right continuous at $v = v_0$; (ii) Denote $b^+ \equiv \lim_{v \downarrow v_0} b(v) = b(v_0)$ and $b^- \equiv \lim_{v \uparrow v_0} b(v)$. We assume $[b^-, b^+] \cap \mathcal{U}$ is not empty.

Condition D5. (i) Assume $F_{U|V}(u|v)$ is continuous in $u \in \mathcal{U}$ and $v = v_0$; (ii) Assume $F_{U|V,W}(u|v, w)$ is continuous in $u \in U$ and $v = v_0$ for every $w \in \mathcal{W}$.

Condition D1 rules out complete manipulation at v_0 and imposes smoothness condition on the corresponding density functions. Tests for Condition D1 are available, see Otsu and Xu (2010) and the references therein. Condition D2 imposes continuity at v_0 of the potential outcome functions. Condition D3 is a regularity condition. Let $D(v) = I\{b(v) - U \geq 0\}$ for $v \in \mathcal{V}$. Then $D = D(V)$ and the propensity score is given by

$$P(v) \equiv \Pr(D = 1|V = v) = F_{U|V}(b(v)|v).$$

Condition D4 imposes conditions on the incentive assignment mechanism b. Without loss of generality, we assume in Condition D4 (i) that $b(v)$ is increasing and right continuous at $v = v_0$. Further we assume in Condition D4 (ii) that $[b^-, b^+]$ and the support of U are not mutually exclusive; otherwise, the propensity score $P(v)$ would be continuous at $v = v_0$ taking values 0 or 1. Obviously the incentive assignment mechanism $b(v) = I\{v \geq v_0\}$ satisfies Condition D4 as long as $[0,1] \cap \mathcal{U}$ is not empty. Condition D5 imposes smoothness conditions on the conditional distribution functions of U. Under Conditions D4 and D5 (i), the propensity score is discontinuous at v_0:

$$\begin{aligned} \lim_{v \downarrow v_0} P(v) &= F_{U|V}(\lim_{v \downarrow v_0} b(v)|v_0) = F_{U|V}(b^+|v_0) = F_{U|V}(b(v_0)|v_0), \\ \lim_{v \uparrow v_0} P(v) &= F_{U|V}(b^-|v_0). \end{aligned}$$

Conditions D1, D2, and D3 imply Assumptions (A1) and (A2) in Hahn, Todd, and van der Klaauw (2001) which assumes continuity of the regression functions $E(Y_0|V = v)$ and $E(Y_1|V = v)$ at v_0. However, compared with the identification results in Hahn, Todd, and van der Klaauw (2001), Theorem 2.1 below does not require any local independence assumption. Let $\Delta = Y_1 - Y_0$.

THEOREM 2.1 *Under Conditions D1-D5, we have*

$$\begin{aligned} &\frac{\lim_{v \downarrow v_0} E(Y|V = v) - \lim_{v \uparrow v_0} E(Y|V = v)}{\lim_{v \downarrow v_0} P(v) - \lim_{v \uparrow v_0} P(v)} \\ &= \lim_{e \downarrow 0} E(\Delta|V = v_0, D(v_0 + e) - D(v_0 - e) = 1) \\ &= \frac{1}{f_V(v_0) \int_{b^-}^{b^+} f_{U|V}(u|v_0) du} E_{W,U}\left[f_{V|W,U}(v_0|W,U) I\{b^- \leq U \leq b^+\}(g_1(v_0, W) - g_0(v_0, W))\right]. \end{aligned}$$

Theorem 2.1 implies that in model (1) and (2), under conditions D1-D5, we identify a weighted average treatment effect for the subpopulation of individuals whose treatment status will change if the value of V is changed from a value slightly smaller than v_0 to a value slightly larger than v_0, i.e., the LATE parameter introduced in Imbens and Angrist (1994). Those individuals who are more likely to obtain a draw of V near v_0 receive more weight than those who are unlikely to obtain such a draw. It is worth emphasizing that Conditions D1-D5 are not sufficient to identify the structural parameters g_1, g_0, and $f_{W,U|V}$, but sufficient to identify the policy parameter LATE.

When $D = I\{V \geq v_0\}$, Theorem 2.1 reduces to Proposition 3 in Lee (2008):

$$\begin{aligned} \lim_{v \downarrow v_0} E(Y|V &= v_0) - \lim_{v \uparrow v_0} E(Y|V = v) = E(\Delta|V = v_0) \\ &= \frac{1}{f_V(v_0)} E_W\left[f_{V|W}(v_0|W)(g_1(v_0, W) - g_0(v_0, W))\right]. \end{aligned}$$

In this case, we identify a weighted average treatment effect for the entire population and this weighted average treatment effect is identical to $\delta = E\left(\Delta|V = v_0\right)$.

Proof of Theorem 2.1. Take $v_+ \in \mathcal{V}$ and $v_- \in \mathcal{V}$ such that $v_+ > v_0 > v_-$. We will look at $E(Y|V = v_+)$ and $E(Y|V = v_-)$ separately. Under condition D4 (i), $b(v_+) \geq b(v_-)$.

First, we have

$$\begin{aligned}
E(Y|V &= v_+) \\
&= E(Y|V = v_+, D(v_+) = 1, D(v_-) = 0)\Pr(D(v_+) = 1, D(v_-) = 0|V = v_+) \\
+E(Y|V &= v_+, D(v_+) = 1, D(v_-) = 1)\Pr(D(v_+) = 1, D(v_-) = 1|V = v_+) \\
+E(Y|V &= v_+, D(v_+) = 0, D(v_-) = 0)\Pr(D(v_+) = 0, D(v_-) = 0|V = v_+) \\
+E(Y|V &= v_+, D(v_+) = 0, D(v_-) = 1)\Pr(D(v_+) = 0, D(v_-) = 1|V = v_+) \\
&= E(Y_1|V = v_+, D(v_+) = 1, D(v_-) = 0)\Pr(D(v_+) = 1, D(v_-) = 0|V = v_+) \\
+E(Y_1|V &= v_+, D(v_+) = 1, D(v_-) = 1)\Pr(D(v_+) = 1, D(v_-) = 1|V = v_+) \\
+E(Y_0|V &= v_+, D(v_+) = 0, D(v_-) = 0)\Pr(D(v_+) = 0, D(v_-) = 0|V = v_+) \\
+E(Y_0|V &= v_+, D(v_+) = 0, D(v_-) = 1)\Pr(D(v_+) = 0, D(v_-) = 1|V = v_+).
\end{aligned}$$

Now from Condition D5 (i), we obtain:

$$\begin{aligned}
\lim_{v_+, v_- \to v_0}\Pr(D_i(v_+) &= 1, D_i(v_-) = 0|V_i = v_+) \\
&= \lim_{v_+, v_- \to v_0}\Pr\left(b\left(v_-\right) < U_i \leq b\left(v_+\right)|V_i = v_+\right) \\
&= \Pr\left(b^- < U_i \leq b^+|V_i = v_0\right).
\end{aligned}$$

Similarly, we obtain:

$$\begin{aligned}
\lim_{v_+, v_- \to v_0}\Pr(D(v_+) &= 1, D(v_-) = 1|V = v_+) = \Pr\left(U \leq b^-|V = v_0\right), \\
\lim_{v_+, v_- \to v_0}\Pr(D(v_+) &= 0, D(v_-) = 0|V = v_+) = \Pr\left(U > b^+|V = v_0\right), \\
\lim_{v_+, v_- \to v_0}\Pr(D(v_+) &= 0, D(v_-) = 1|V = v_+) = \Pr\left(b^+ < U \leq b^-|V = v_0\right) = 0.
\end{aligned}$$

As a result,

$$\begin{aligned}
\lim_{v_+ \to v_0} E(Y|V &= v_+) \\
&= \left[\lim_{v_+, v_- \to v_0} E(Y_1|V = v_+, b\left(v_-\right) < U \leq b\left(v_+\right))\right]\Pr\left(b^- < U \leq b^+|V = v_0\right) \\
&\quad + \left[\lim_{v_+, v_- \to v_0} E(Y_1|V = v_+, U \leq b\left(v_-\right))\right]\Pr\left(U \leq b^-|V = v_0\right) \\
&\quad + \left[\lim_{v_+, v_- \to v_0} E(Y_0|V = v_+, U > b\left(v^+\right))\right]\Pr\left(U > b^+|V = v_0\right).
\end{aligned}$$

Similarly, we can show:

$$
\begin{aligned}
\lim_{v_-\to v_0} E(Y|V &= v_-) \\
&= \left[\lim_{v_+,v_-\to v_0} E(Y_0|V = v_-, b\left(v_-\right) < U \le b\left(v_+\right))\right] \Pr\left(b^- < U \le b^+|V = v_0\right) \\
&\quad + \left[\lim_{v_+,v_-\to v_0} E(Y_1|V = v_-, U \le b\left(v_-\right))\right] \Pr\left(U \le b^-|V = v_0\right) \\
&\quad + \left[\lim_{v_+,v_-\to v_0} E(Y_0|V = v_-, U > b\left(v^+\right))\right] \Pr\left(U > b^+|V = v_0\right).
\end{aligned}
$$

Lemma A.2 implies that for $j = 1, 0$:

$$
\begin{aligned}
\lim_{v_+,v_-\to v_0} E(Y_j|V &= v_+, b\left(v_-\right) < U \le b\left(v_+\right)) = \lim_{v_+,v_-\to v_0} E(Y_j|V = v_-, b\left(v_-\right) < U \le b\left(v_+\right)) \\
&= E(Y_j|V = v_0, b^- < U \le b^+),
\end{aligned}
$$

$$
\begin{aligned}
\lim_{v_+,v_-\to v_0} E(Y_j|V &= v_+, U \le b\left(v_-\right)) = \lim_{v_+,v_-\to v_0} E(Y_j|V = v_-, U \le b\left(v_-\right)) \\
&= E(Y_j|V = v_0, U \le b^-),
\end{aligned}
$$

and

$$
\begin{aligned}
\lim_{v_+,v_-\to v_0} E(Y_j|V &= v_+, U > b\left(v_+\right)) = \lim_{v_+,v_-\to v_0} E(Y_j|V = v_-, U > b\left(v_+\right)) \\
&= E(Y_j|V = v_0, U > b^+).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\frac{\lim_{v_+\to v_0} E(Y|V = v_+) - \lim_{v_-\to v_0} E(Y|V = v_-)}{\lim_{v\downarrow v_0} P(v) - \lim_{v\uparrow v_0} P(v)} \\
&= E(Y_1|V = v_0, b^- < U \le b^+) - E(Y_0|V = v_0, b^- < U \le b^+) \\
&= \lim_{v_+,v_-\to v_0} E(\Delta|V = v_0, D(v_+) - D(v_-) = 1).
\end{aligned}
$$

Finally let $A = \{b^- < U_i \le b^+\}$. It follows from Lemma A.1 that for $j = 1, 0$,

$$
\begin{aligned}
E(Y_j|V &= v_0, b^- < U \le b^+) \\
&= \int g_j(v_0, w) f_{W|V,U}(w|v_0, A) dw \\
&= \int g_j(v_0, w) \frac{\int_{b^-}^{b^+} f_{V|W,U}(v_0|w, u) f_{W,U}(w, u) du}{f_V(v_0) \int_{b^-}^{b^+} f_{U|V}\left(u|v_0\right) du} dw \\
&= E_{W,U}\left[\frac{f_{V|W,U}(v|W, U)}{f_V(v_0) \int_{b^-}^{b^+} f_{U|V}\left(u|v_0\right) du} I\left\{b^- \le U \le b^+\right\} g_j(v_0, W)\right].
\end{aligned}
$$

Q.E.D.

2.1.2 Auxiliary Regressions

In this subsection, we present conditions on the structural parameters in (1) and (2) to justify the auxiliary regressions below:

$$\begin{aligned} Y &= g(V) + \delta I\{V \geq v_0\} + \varepsilon, & (4) \\ D &= h(V) + \zeta I\{V \geq v_0\} + \epsilon, & (5) \end{aligned}$$

where $E(\varepsilon|V) = 0$, $E(\epsilon|V) = 0$, and

$$\delta = \lim_{v \downarrow v_0} E(Y|V = v) - \lim_{v \uparrow v_0} E(Y|V = v),\ \zeta = \lim_{v \downarrow v_0} P(v) - \lim_{v \uparrow v_0} P(v).$$

Unlike Porter (2003) and Imbens and Kalyanaramang (2009) who directly assume continuity of g and h, we impose sufficient conditions on the structural parameters in (1) and (2) to ensure that g and h are continuous on the support of V.

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

Condition D2(A). Assume $g_1(v,w)$ and $g_0(v,w)$ are continuous on $\mathcal{V}$ for every $w \in \mathcal{W}$.

Condition D4(A). $b(v)$ is continuous in $v \in \mathcal{V}$ except at v_0.

Condition D5(A). (i) Assume $F_{U|V}(u|v)$ is continuous in $u \in \mathcal{U}$ and $v \in \mathcal{V}$; (ii) Assume $F_{U|V,W}(u|v,w)$ is continuous in $u \in \mathcal{U}$ and $v \in \mathcal{V}$ for every $w \in \mathcal{W}$.

Proposition 2.2 *Under Conditions D1(A), D2(A), D3, D4, D4(A), and D5(A), the functions $g(\cdot)$ and $h(\cdot)$ are continuous on the support of V.*

Remark 2.1. It is clear from the proof of Proposition 2.2 that under Conditions D1-D5, the functions $g(\cdot)$ and $h(\cdot)$ are continuous at v_0. For the jump wavelet estimator introduced in Section 3, it is sufficient that $g(\cdot)$ and $h(\cdot)$ are continuous at v_0.

2.2 Kink Incentive Assignment Mechanism

2.2.1 Identification

Many policy assignment mechanisms including allocation of unemployment benefits and income tax systems violate Condition D4. Instead they satisfy Condition K4 below.

Condition K1. Assume (i) $f_{V|W}(v|w)$ is continuously differentiable in a neighborhood of v_0 and $f_{V|W}(v_0|w) > 0$ for every $w \in \mathcal{W}$; (ii) $f_V(v)$ is continuously differentiable in a neighborhood of v_0 and $f_V(v_0) > 0$; (iii) $f_{V|W,U}(v|w,u)$ is continuously differentiable in a neighborhood of v_0 and $f_{V|W,U}(v_0|w,u) > 0$ for $u \in \mathcal{U}$ and $w \in \mathcal{W}$.

Condition K2. Assume $g_1(v,w)$ and $g_0(v,w)$ are continuously differentiable in a neighborhood of v_0 for every $w \in \mathcal{W}$.

Condition K3. For $j = 1, 0$, assume (i) $E|Y_j| < \infty$;
(ii) For all $u \in \mathcal{U}$, $\sup_v |\frac{\partial f_{U|V}(u|v)}{\partial v}| < \infty$ and $\int \sup_v |\frac{\partial f_{U|V}(u|v)}{\partial v}| du < \infty$;
(iii) For all $w \in \mathcal{W}$, $\int \sup_v |\frac{\partial f_{W,U|V}(w,u|v)}{\partial v}| du < \infty$ and $\int \int \sup_v |\frac{\partial f_{W,U|V}(w,u|v)}{\partial v}| dudw < \infty$.

Condition K4. **(i)** Assume $b(v)$ is increasing and continuously differentiable in a neighborhood of v_0 except at v_0, where its derivative is right continuous at $v = v_0$; (ii) $b(v_0) \in \mathcal{U}$.

Condition K5. (i) Assume $F_{U|V}(u|v)$ is continuously differentiable in $u \in \mathcal{U}$, and continuously differentiable in a neighborhood of v_0 as well; (ii) Assume $F_{U|V,W}(u|v, w)$ is continuously differentiable in $u \in \mathcal{U}$ and continuously differentiable in a neighborhood of v_0 for every $w \in \mathcal{W}$.

We note that Condition K4 is also used in Card, Lee, and Pei (2009) which assumes $D = b(V)$ implying a continuous treatment D under Condition K4. Instead we focus on a binary treatment D and allow for the unobserved covariate U to affect the agent's selection of treatment status.

Denote $b'^{+} \equiv \lim_{v\downarrow v_0} b'(v) = b'(v_0) < \infty$ and $b'^{-} \equiv \lim_{v\uparrow v_0} b'(v) < \infty$. Condition K4 (i) implies: $b'^{-} \neq b'^{+}$. Under Conditions K4 and K5 (i), the propensity score is discontinuous in its first derivative at v_0. To see this, we note that

$$P'(v) = f_{U|V}(b(v)|v)b'(v) + \frac{\partial F_{U|V}(b(v)|v)}{\partial v}.$$

So under conditions K4 and K5, we obtain:

$$\begin{aligned}
\lim_{v\downarrow v_0} P'(v) &= f_{U|V}(b(v_0)|v_0)b'^{+} + \int_{-\infty}^{b(v_0)} \frac{\partial f_{U|V}(u|v_0)}{\partial v} du, \\
\lim_{v\uparrow v_0} P'(v) &= f_{U|V}(b(v_0)|v_0)b'^{-} + \int_{-\infty}^{b(v_0)} \frac{\partial f_{U|V}(u|v_0)}{\partial v} du,
\end{aligned}$$

and

$$\lim_{v\downarrow v_0} P'(v) - \lim_{v\uparrow v_0} P'(v) = f_{U|V}(b(v_0)|v_0)\left[b'^{+} - b'^{-}\right] \neq 0.$$

THEOREM 2.3 *Under Conditions K1-K5, we have*

$$\begin{aligned}
&\frac{\lim_{v\downarrow v_0} dE(Y|V = v)/dv - \lim_{v\uparrow v_0} dE(Y|V = v)/dv}{\lim_{v\downarrow v_0} P'(v) - \lim_{v\uparrow v_0} P'(v)} \\
&= \lim_{e\downarrow 0} E(\Delta|V = v_0, D(v_0 + e) - D(v_0 - e) = 1) \\
&= E_W\left[[g_1(v_0, W) - g_0(v_0, W)]\frac{f_{W|U,V}(W|b(v_0), v_0)}{f_W(W)}\right].
\end{aligned}$$

2.2.2 Auxiliary Regressions

For kink incentive assignment mechanisms, we establish the following auxiliary regressions:

$$Y = g_K(V) + \delta_K(V - v_0)I\{V \geq v_0\} + \varepsilon_K, \tag{6}$$

$$D = h_K(V) + \zeta_K(V - v_0)I\{V \geq v_0\} + \epsilon_K. \tag{7}$$

where $E(\varepsilon_K|V) = 0$, $E(\epsilon_K|V) = 0$, and

$$\begin{aligned}
\delta_K &= \lim_{v \downarrow v_0} dE(Y|V=v)/dv - \lim_{v \uparrow v_0} dE(Y|V=v)/dv, \\
\zeta_K &= \lim_{v \downarrow v_0} P'(v) - \lim_{v \uparrow v_0} P'(v).
\end{aligned}$$

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

Condition K2(A). Assume $g_1(v,w)$ and $g_0(v,w)$ are continuously differentiable on $\mathcal{V}$ for every $w \in \mathcal{W}$.

Condition K4(A). $b(v)$ is continuously differentiable for $v \in \mathcal{V}$ except at v_0, where it is only continuous.

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

Proposition 2.4 *Under Conditions K1(A), K2(A), K3, K4, K4(A), and K5(A), the functions* $g_K(\cdot)$ *and* $h_K(\cdot)$ *are continuously differentiable on* $\mathcal{V}$*.*

3 The First Wavelet Estimator

Let θ denote the identified LATE parameter. In this and the next sections, we propose wavelet estimators of LATE for both discontinuous and kink incentive assignment mechanisms. Throughout this and the next sections, we assume the conditions of Propositions 2.2 and 2.4 hold respectively for discontinuous and kink incentive assignment mechanisms and a random sample (V_i, Y_i, D_i), $i = 1, ..., n$, is available.

3.1 Discontinuous Incentive Assignment Mechanism

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

The wavelet estimator of δ we propose in this section is closely related to the estimator of the jump size of the regression function $E(Y_i|V_i = v)$ studied in Park and Kim (2006). Let $F_V(\cdot)$ denote the distribution function of V_i and $\tau \equiv F_V(v_0)$. Let $V_{1:n} \leq \cdots \leq V_{n:n}$ denote the order statistics of $\{V_i\}_{i=1}^n$ and $\left\{Y_{[i:n]}\right\}_{i=1}^n$ the concomitants of $\{V_{i:n}\}_{i=1}^n$ or induced order statistics. Further let $t_i = i/n$ for $1 \leq i \leq n$.

To motivate our first wavelet estimator and wavelet OLS estimators in Section 4, we consider the auxiliary regression in the wavelet domain. Let $\widehat{F}_V(\cdot)$ denote the empirical distribution function[6] of $\{V_i\}_{i=1}^n$. Then the induced order statistics $\left\{Y_{[i:n]}\right\}_{i=1}^n$ satisfy:

$$\begin{aligned} Y_{[i:n]} &= g(V_{i:n}) + \delta I\{V_{i:n} \geq v_0\} + \varepsilon_{[i:n]} \\ &= g(\widehat{F}_V^{-1}(t_i)) + \delta I\{t_i \geq \widehat{F}_V(v_0)\} + \varepsilon_{[i:n]} \\ &\equiv G(t_i) + \delta I\{t_i \geq \widehat{\tau}\} + e_i, \end{aligned}$$

where $G(t) \equiv g(F_V^{-1}(t))$, $\widehat{\tau} = \widehat{F}_V(v_0)$, and

$$e_i = \left[g(\widehat{F}_V^{-1}(t_i)) - g(F_V^{-1}(t_i))\right] + \varepsilon_{[i:n]}.$$

Suppose $\psi(t)$ is a real-valued (mother) wavelet function on the interval $[a, b]$ with $-\infty < a < 0 < b < \infty$, i.e., it satisfies:

$$\int_a^b \psi(t)\, dt = 0, \int_a^b \psi^2(t)\, dt = 1,$$

and an admissibility condition that $\int \left|\widehat{\psi}(\xi)\right|^2 / |\xi|\, d\xi < \infty$, where $\widehat{\psi}(\xi)$ is the Fourier transform of $\psi(t)$. Let $\widehat{\Delta}_{j_0}^A(\tau)$ denote the wavelet coefficient of $\{A_i\}_{i=1}^n$ at cut-off point τ and scale j_0:

$$\widehat{\Delta}_{j_0}^A(\tau) = \frac{2^{j_0/2}}{n} \sum_{i=1}^n A_i \psi(2^{j_0}(t_i - \tau)).$$

Then we have:

$$\widehat{\Delta}_{j_0}^Y(\tau) = \widehat{\Delta}_{j_0}^G(\tau) + \delta \widehat{\Delta}_{j_0}^d(\tau) + \widehat{\Delta}_{j_0}^e(\tau) \tag{8}$$

where $d_i = I\{t_i \geq \widehat{\tau}\}$ and

$$\widehat{\Delta}_{j_0}^Y(\tau) = \frac{2^{j_0/2}}{n} \sum_{i=1}^n Y_{[i:n]} \psi(2^{j_0}(t_i - \tau)).$$

It is well known that the wavelet coefficient $\widehat{\Delta}_{j_0}^A(\tau)$ captures the variation of the sequence $\{A_i\}_{i=1}^n$ at cut-off point τ and scale j_0. When the scale is large enough, $\widehat{\Delta}_{j_0}^A(\tau)$ is small unless there is a jump or isolated singularity in $\{A_i\}_{i=1}^n$ at τ. Since $G(t)$ is continuous at τ, we expect $\widehat{\Delta}_{j_0}^G(\tau)$ to be small at large enough j_0 motivating our first wavelet estimator $\widehat{\delta}$:

$$\widehat{\delta} = \frac{\widehat{\Delta}_{j_0}^Y(\widehat{\tau})}{\widehat{\Delta}_{j_0}^d(\widehat{\tau})}.$$

The estimator studied in Park and Kim (2006) is:

$$\overline{\delta} = \frac{2^{j_0/2} \widehat{\Delta}_{j_0}^Y(\widehat{\tau})}{\int_0^b \psi(u) du}.$$

---

[6] Park and Kim (2006) chooses a piecewise linear version of the empirical distribution function. All the results in this paper carry over to this case.

To establish asymptotic properties of $\widehat{\delta}$, we adopt the following assumptions. We note that the ψ function needs to satisfy assumption A4 only. All wavelet functions satisfy A4, but functions satisfying A4 may not be wavelet functions. For the ease of exposition, we will refer to any function satisfying A4 as a 'wavelet' function, the corresponding transform coefficients as 'wavelet' coefficients, and our estimators as wavelet estimators.

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

Assumption A2.

(G). Let $G(t) \equiv g(F_V^{-1}(t))$. (a) $G(t)$ is l_G times continuously differentiable for $t \in (0,1)\backslash\{\tau\}$, and $G(\cdot)$ is continuous at τ with finite right and left-hand derivatives to order $l_G \geq m+1$; (b) Right and left hand derivatives of $G(t)$ up to order $l_G \geq m+1$ are equal at τ, where m is defined in Assumption A4.

(H). Let $H(t) \equiv h(F_V^{-1}(t))$. (a) $H\left(t\right)$ is l_H times continuously differentiable for $t \in (0,1)\backslash\{\tau\}$, and $H\left(\cdot\right)$ is continuous at τ with finite right and left-hand derivatives to order $l_H \geq m+1$; (b) Right and left hand derivatives of $H\left(t\right)$ to order $l_H \geq m+1$ are equal at τ, where m is defined in Assumption A4.

Assumption A3. **(G).** (a) $\sigma_\varepsilon^2(v) \equiv E(\varepsilon^2|V = v)$ 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.

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

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

Assumption A4. (a) The function $\psi\left(\cdot\right)$ is continuous with compact support $[a, b]$, where $a < 0 < b$ and m vanishing moments, i.e., $\int_a^b u^j \psi\left(u\right) du = 0$ for $j = 0, 1, ..., m-1$; (b) $\int_0^b \psi(u) du \neq 0$, $\int_a^b u^m \psi\left(u\right) du \neq 0$, and $\int_a^b \left|u^m \psi\left(u\right)\right| du < \infty$; (c) ψ has a bounded derivative.

Assumption A5. (a) As $n \to \infty$, $j_0 \to \infty$, $\frac{2^{j_0}}{n} \to 0$, and $\frac{1}{2^{j_0}}\sqrt{\frac{n}{2^{j_0}}} \to C_a < \infty$; (b) As $n \to \infty$, $j_0 \to \infty$, $\frac{2^{j_0}}{n} \to 0$, and $(\frac{1}{2^{j_0}})^m \sqrt{\frac{n}{2^{j_0}}} \to C_b < \infty$.

Assumption A1 may be relaxed to allow for non i.i.d. data by using the extension of Theorem 1 in Yang (1981) presented in Chu and Jacho-Chavez (2010). Assumption A2(G) (a) allows for jumps in the derivatives of G at τ up to order l_G. Work in the statistics literature on detection of jumps such as Wang (1995) and Park and Kim (2006) assume away the presence of jumps in the derivatives of G at τ so that Assumption A2(G) (b) holds. Assumption A3(G) imposes conditions on the conditional variance function and $E[|\varepsilon|^{2+\zeta}|v]$. Park and Kim (2006) assumes a constant conditional variance function. Assumption A4 specifies the class of functions ψ. In contrast to a kernel function which integrates to one, the function ψ integrates to zero and shares the properties of an m-th order kernel otherwise. Examples of ψ 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 $[a, b] = [-(L-1), L]$. In addition, the second derivative functions of kernel functions constructed in Cheng and Raimondo (2008) for which $[a, b] = [-1, 1]$ and $m = s - 1$ and the differences between the two kernel functions used in Wu and Chu (1993) also satisfy Assumption A4[7]. Assumption A5 imposes conditions on the scale j_0.

THEOREM 3.1 *Suppose A1, A3(G), and A4 hold.*

(i) When A2(G) (a) and A5 (a) hold, we obtain: $\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta} - \delta)$ *and* $\sqrt{\frac{n}{2^{j_0}}}(\overline{\delta} - \delta)$ *have the same asymptotic distribution and*

$$\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta} - \delta) \xrightarrow{d} N\left(C_a B_a, V\right),$$

where

$$\begin{aligned} B_a &= \frac{\left[G_+^{(1)}(\tau) - G_-^{(1)}(\tau)\right] \int_0^b \psi(u) u du}{\int_0^b \psi(u) du}, \\ V &= \frac{\sigma_{\varepsilon+}^2(v_0) \int_0^b \psi^2(u) du + \sigma_{\varepsilon-}^2(v_0) \int_a^0 \psi^2(u) du}{\left(\int_0^b \psi(u) du\right)^2}. \end{aligned}$$

(ii) When A2(G) (b) and A5 (b) hold, we obtain: $\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta} - \delta)$ *and* $\sqrt{\frac{n}{2^{j_0}}}(\overline{\delta} - \delta)$ *have the same asymptotic distribution and*

$$\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta} - \delta) \xrightarrow{d} N\left(C_b B_b, V\right),$$

where

$$B_b = \frac{G^{(m)}(\tau) \int_a^b u^m \psi(u) du}{\int_0^b \psi(u) du}.$$

When $\sigma_\varepsilon^2(v)$ is a constant, the asymptotic distribution of $\overline{\delta}$ given in Theorem 3.1 (ii) reduces to that in Park and Kim (2006). Theorem 3.1 (i) reveals a similar asymptotic behavior of $\widehat{\delta}$ to the Nadaraya-Watson kernel estimator in Porter (2003) under A2(G) (a). However, as revealed in Theorem 3.1 (ii), although A2(G) (b) does not affect the asymptotic distribution of the Nadaraya-Watson kernel estimator in Porter (2003), it does affect the asymptotic distribution of our wavelet estimator $\widehat{\delta}$. In particular, it reduces the order of the asymptotic bias of $\widehat{\delta}$ from 2^{-j_0} to 2^{-mj_0}. Thus, in terms of asymptotic bias, $\widehat{\delta}$ behaves more like the partial linear estimator of Porter (2003). This is not surprising, given that we transformed our covariates $\{V_i\}_{i=1}^n$ to equally spaced design $\{t_i\}_{i=1}^n$ and the observation in Eubank and Speckman (1994) that for equally spaced design, their partial linear estimator of the jump size is asymptotically equivalent to an estimator similar to $\overline{\delta}$ with a specific ψ function. In general, transforming a random design to an equally spaced fixed

---

[7] See Assumption $C_{1,s}$ in Cheng and Raimondo (2008).

design before applying nonparametric method leads to better finite sample performance, see Hall, et al. (1998). In our context, this also leads to estimators that are design-adaptive: the asymptotic bias and variance of our estimators do not depend on the density $f_V(v)$. In addition, the discrete version of our estimator can be computed by applying the Cascades Algorithm by Mallat (2009).

Remark 3.1. It is interesting to observe that τ is the location of the jump in the regression model with regularly spaced design points, so we can estimate τ by any existing estimator of the location of the jump proposed in the literature, such as Raimondo (1998). Under standard regularity conditions, all the existing estimators converge at rates faster than $n^{-1/2}$ so the conclusions in Theorem 3.1 and in all other theorems in this paper for the discontinuous assignment mechanism remain valid.

We are now ready to estimate the LATE parameter θ. Let $\left\{D_{[i:n]}\right\}_{i=1}^{n}$ denote the concomitants of $\{V_{i:n}\}_{i=1}^{n}$ corresponding to $\{D_i\}_{i=1}^{n}$. Our first wavelet estimator of $\theta \equiv \delta/\zeta$ is defined as $\widehat{\theta} = \widehat{\delta}/\widehat{\zeta}$, where

$$\widehat{\zeta} = \frac{\widehat{\Delta}_{j_0}^{D}(\widehat{\tau})}{\widehat{\Delta}_{j_0}^{d}(\widehat{\tau})}.$$

For simplicity, we have used the same mother wavelet $\psi(\cdot)$ and scale j_0 to estimate δ and ζ. This can be relaxed at the expense of more lengthy derivations.

THEOREM 3.2 *Suppose A1, A3, and A4 hold.*

(i) When A2 (a) and A5 (a) hold, we obtain:

$$(n/2^{j_0})^{1/2}\begin{pmatrix}\widehat{\delta}-\delta\\ \widehat{\zeta}-\zeta\end{pmatrix}\xrightarrow{d} N\left(\begin{pmatrix}C_aB_a\\ C_aB_a^D\end{pmatrix},\begin{pmatrix}V & V^{YD}\\ V^{YD} & V^D\end{pmatrix}\right),$$

and

$$(n/2^{j_0})^{1/2}\left(\frac{\widehat{\delta}}{\widehat{\zeta}}-\frac{\delta}{\zeta}\right)\xrightarrow{d} N\left(\frac{1}{\zeta}C_a\left[B_a-\frac{\delta}{\zeta}B_a^D\right],\frac{1}{\zeta^2}\left[V-\frac{2\delta}{\zeta}V^{YD}+\frac{\delta^2}{\zeta^2}V^D\right]\right),$$

where

$$\begin{aligned}
B_a^D &= \frac{\left[H_+^{(1)}(\tau)-H_-^{(1)}(\tau)\right]\int_0^b\psi(u)u\,du}{\int_0^b\psi(u)du},\\
V^D &= \frac{\sigma_{\epsilon+}^2(v_0)\int_0^b\psi^2(u)du+\sigma_{\epsilon-}^2(v_0)\int_a^0\psi^2(u)du}{\left(\int_0^b\psi(u)du\right)^2},\\
V^{YD} &= \frac{\sigma_{\varepsilon\epsilon+}^2(v_0)\int_0^b\psi^2(u)du+\sigma_{\varepsilon\epsilon-}^2(v_0)\int_a^0\psi^2(u)du}{\left(\int_0^b\psi(u)du\right)^2}.
\end{aligned}$$

(ii) When A2 (b) and A5 (b) hold, we obtain:

$$(n/2^{j_0})^{1/2}\begin{pmatrix}\widehat{\delta}-\delta\\ \widehat{\zeta}-\zeta\end{pmatrix}\xrightarrow{d} N\left(\begin{pmatrix}C_bB_b\\ C_bB_b^D\end{pmatrix},\begin{pmatrix}V & V^{YD}\\ V^{YD} & V^D\end{pmatrix}\right),$$

and

$$(n/2^{j_0})^{1/2}\left(\frac{\widehat{\delta}}{\widehat{\zeta}}-\frac{\delta}{\zeta}\right)\xrightarrow{d}N\left(\frac{1}{\zeta}C_b\left[B_b-\frac{\delta}{\zeta}B_b^D\right],\frac{1}{\zeta^2}\left[V-\frac{2\delta}{\zeta}V^{YD}+\frac{\delta^2}{\zeta^2}V^D\right]\right),$$

where

$$B_b^D=\frac{H^{(m)}(\tau)\int_a^b u^m\psi(u)du}{\int_0^b\psi(u)du}.$$

3.2 Kink Incentive Assignment Mechanism

For a kink incentive assignment mechanism, the auxiliary regressions are given in (6) and (7). Again we focus on the estimation of δ_K. First we note that the induced order statistics $\left\{Y_{[i:n]}\right\}_{i=1}^n$ satisfy:

$$\begin{aligned}Y_{[i:n]} &= g_K\left(V_{i:n}\right)+\delta_K(V_{i:n}-v_0)I\{V_{i:n}\geq v_0\}+\varepsilon_{K,[i:n]}\\ &= g_K\left(\widehat{F}_V^{-1}\left(t_i\right)\right)+\delta_K(\widehat{F}_V^{-1}\left(t_i\right)-v_0)I\{t_i\geq\widehat{\tau}\}+\varepsilon_{K,[i:n]}.\end{aligned}$$

Similar to the discontinuous incentive assignment mechanism case, we propose the following estimator of δ_K:

$$\widehat{\delta}_K=\frac{\widehat{\Delta}_{j_0}^Y\left(\widehat{\tau}\right)}{\widehat{\Delta}_{j_0}^k\left(\widehat{\tau}\right)},$$

where $k_i=\left[\widehat{F}_V^{-1}(t_i)-v_0\right]I\{t_i\geq\widehat{\tau}\}$. We will show that under conditions stated below, $\widehat{\delta}_K$ has the same asymptotic distribution as

$$\overline{\delta}_K=\frac{1}{n}\sum_{i=1}^n\frac{2^{j_0}\psi\left[2^{j_0}(\frac{i}{n}-\tau)\right]}{\int_0^b\left[F_V^{-1}(\frac{u}{2^{j_0}}+\tau)-v_0\right]\psi(u)du}Y_{[i:n]}.$$

Like the discontinuous incentive assignment mechanism case, $\widehat{\tau}$ could be replaced with any existing estimator of the kink location in the first derivative of a nonparametric regression function including Raimondo (1998).

Assumption A2K.

(G). Let $G_K(t)\equiv g_K(F_V^{-1}(t))$. (a) $G_K(t)$ is l_G+1 times continuously differentiable for $t\in(0,1)\backslash\{\tau\}$, and $G_K(\cdot)$ is continuously differentiable at τ with finite right and left-hand derivatives to order $l_G+1\geq m+2$; (b) Right and left hand derivatives of $G_K(t)$ to order $l_G+1\geq m+2$ are equal at τ, where m is defined in Assumption A4K.

(H). Let $H_K(t)\equiv h_K(F_V^{-1}(t))$. (a) $H_K\left(t\right)$ is l_H+1 times continuously differentiable for $t\in(0,1)\backslash\{\tau\}$, and $H_K\left(\cdot\right)$ is continuously differentiable at τ with finite right and left-hand derivatives to order $l_H+1\geq m+2$; (b) Right and left hand derivatives of $H_K\left(t\right)$ to order $l_H+1\geq m+2$ are equal at τ, where m is defined in Assumption A4K.

Assumption A4K. (a) The 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 $j=0,1,...,m$; (b) $\int_0^b u\psi(u)du \neq 0$, $\int_a^b u^{m+1}\psi(u)\,du \neq 0$, and $\int_a^b \left|u^{m+1}\psi(u)\right| du < \infty$; (c) ψ has a bounded derivative.

Assumption A5K. (a) As $n\to\infty$, $j_0\to\infty$, $\frac{2^{3j_0}}{n}\to 0$, and $\frac{1}{2^{j_0}}\sqrt{\frac{n}{2^{3j_0}}}\to C_{Ka}<\infty$; (b) As $n\to\infty$, $j_0\to\infty$, $\frac{2^{3j_0}}{n}\to 0$, and $(\frac{1}{2^{j_0}})^m\sqrt{\frac{n}{2^{3j_0}}}\to C_{Kb}<\infty$.

Assumption A6K. (a) $F_V^{-1}(v)$ is continuously differentiable on the support of V; (b) $F_V^{-1}(v)$ is m times continuously differentiable on the support of V.

THEOREM 3.3 *Suppose A1, A3(G) for ε_K, and A4K hold.*

(i) When A2K(G) (a), A5K (a) and A6K (a) hold, we obtain: $\sqrt{\frac{n}{2^{3j_0}}}(\widehat{\delta}_K-\delta_K)$ and $\sqrt{\frac{n}{2^{3j_0}}}(\overline{\delta}_K-\delta_K)$ have the same asymptotic distribution and

$$\sqrt{\frac{n}{2^{3j_0}}}(\widehat{\delta}_K-\delta_K)\xrightarrow{d} N\left(C_{Ka}B_{Ka},V_K\right),$$

where

$$\begin{aligned} B_{Ka} &= \frac{\left[G_{K+}^{(2)}(\tau)-G_{K-}^{(2)}(\tau)\right]\int_0^b u^2\psi(u)du}{2\int_0^b u\psi(u)du} f_V(v_0), \\ V_K &= \frac{f_V^2(v_0)\left[\sigma_{\varepsilon+}^2(v_0)\int_0^b \psi^2(u)du+\sigma_{\varepsilon-}^2(v_0)\int_a^0 \psi^2(u)du\right]}{\left(\int_0^b u\psi(u)du\right)^2}. \end{aligned}$$

(ii) When A2K(G) (b), A5K (b) and A6K (b) hold, we obtain: $\sqrt{\frac{n}{2^{3j_0}}}(\widehat{\delta}_K-\delta_K)$ and $\sqrt{\frac{n}{2^{3j_0}}}(\overline{\delta}_K-\delta_K)$ have the same asymptotic distribution and

$$\sqrt{\frac{n}{2^{3j_0}}}(\widehat{\delta}_K-\delta_K)\xrightarrow{d} N\left(C_{Kb}B_{Kb},V_K\right),$$

where

$$B_{Kb}=\frac{G_K^{(m+1)}(\tau)\int_a^b u^{m+1}\psi(u)du}{(m+1)!\int_0^b u\psi(u)du}f_V(v_0).$$

Comparing Theorems 3.1 and 3.3, we observe the same qualitative behavior of $\widehat{\delta}$ and $\widehat{\delta}_K$ in terms of the order of their asymptotic bias: the order of the asymptotic bias of $\widehat{\delta}_K$ depends on whether there are jump discontinuities in the second and higher order derivatives of G_K.

Finally our first LATE estimator for kink incentive assignment mechanisms is defined as $\widehat{\delta}_K/\widehat{\zeta}_K$, where $\widehat{\zeta}_K=\widehat{\Delta}_{j_0}^D(\widehat{\tau})/\widehat{\Delta}_{j_0}^k(\widehat{\tau})$.

THEOREM 3.4 *Suppose A1, A3 for ε_K and ϵ_K, and A4K hold.*

(i) When A2K (a), A5K (a) and A6K(a) hold, we obtain:

$$\sqrt{\frac{n}{2^{3j_0}}}\left(\frac{\widehat{\delta}_K}{\widehat{\zeta}_K}-\frac{\delta_K}{\zeta_K}\right)\xrightarrow{d} N\left(\frac{1}{\zeta}C_{Ka}\left[B_{Ka}-\frac{\delta}{\zeta}B_{Ka}^D\right],\frac{1}{\zeta^2}\left[V_K-\frac{2\delta}{\zeta}V_K^{YD}+\frac{\delta^2}{\zeta^2}V_K^D\right]\right),$$

where

$$\begin{aligned}
B_{Ka}^{D} &= \frac{\left[H_{K+}^{(2)}(\tau)-H_{K-}^{(2)}(\tau)\right]\int_0^b u^2\psi(u)du}{2\int_0^b u\psi(u)du} f_V(v_0), \\
V_K^D &= \frac{f_V^2(v_0)\left[\sigma_{\epsilon+}^2(v_0)\int_0^b \psi^2(u)du+\sigma_{\epsilon-}^2(v_0)\int_a^0 \psi^2(u)du\right]}{\left(\int_0^b u\psi(u)du\right)^2}, \\
V_K^{YD} &= \frac{f_V^2(v_0)\left[\sigma_{\varepsilon\epsilon+}^2(v_0)\int_0^b \psi^2(u)du+\sigma_{\varepsilon\epsilon-}^2(v_0)\int_a^0 \psi^2(u)du\right]}{\left(\int_0^b u\psi(u)du\right)^2}.
\end{aligned}$$

(*ii) When A2K (b), A5K (b) and A6K(b) hold, we obtain:*

$$\sqrt{\frac{n}{2^{3j_0}}}\left(\frac{\widehat{\delta}_K}{\widehat{\zeta}_K}-\frac{\delta_K}{\zeta_K}\right)\xrightarrow{d} N\left(\frac{1}{\zeta}C_{Kb}\left[B_{Kb}-\frac{\delta}{\zeta}B_{Kb}^D\right],\frac{1}{\zeta^2}\left[V_K-\frac{2\delta}{\zeta}V_K^{YD}+\frac{\delta^2}{\zeta^2}V_K^D\right]\right),$$

where

$$B_{Kb}^D=\frac{H_K^{(m+1)}(\tau)\int_a^b u^{m+1}\psi(u)du}{(m+1)!\int_0^b u\psi(u)du}f_V(v_0).$$

4 The Wavelet OLS Estimators

4.1 Discontinuous Incentive Assignment Mechanism

Note that the wavelet estimators $\widehat{\delta}$ and $\overline{\delta}$ make use of one wavelet coefficient of $\left\{Y_{[i:n]}\right\}_{i=1}^n$ only, the one at location τ and scale j_0. Heuristically wavelet coefficients of $\left\{Y_{[i:n]}\right\}_{i=1}^n$ at locations near τ and other fine scales contain information about δ as well. Formally, it follows from

$$Y_{[i:n]}=G\left(t_i\right)+\delta I\{t_i\geq \widehat{F}_V\left(v_0\right)\}+\left\{\left[g(\widehat{F}_V^{-1}\left(t_i\right))-G\left(t_i\right)\right]+\varepsilon_{[i:n]}\right\},$$

that

$$\widehat{\Delta}_j^Y\left(t\right)=\widehat{\Delta}_j^G\left(t\right)+\delta\widehat{\Delta}_j^d\left(t\right)+\widehat{\Delta}_j^e\left(t\right),\text{ for all } j \text{ and } t\in[0,1], \tag{9}$$

where $\widehat{\Delta}_j^A\left(t\right)$ denotes the wavelet coefficient of $\{A_i\}_{i=1}^n$ at location t and scale j, i.e.,

$$\widehat{\Delta}_j^A\left(t\right)=\frac{2^{j/2}}{n}\sum_{i=1}^n A_i\psi(2^j(t_i-t)).$$

The wavelet regression (9) suggests that provided $G\left(t\right)$ is continuous, all the wavelet coefficients $\widehat{\Delta}_j^Y\left(t\right)$ at large enough scales and locations t near τ should contain information on δ. This motivates us to propose the following general class of wavelet OLS estimators of δ:

$$\widehat{\delta}_W=\frac{\sum_{j=j_L}^{j_U}\int_0^1\widehat{\Delta}_j^Y\left(t\right)\widehat{\Delta}_j^d\left(t\right)\ \ D_j(t)dt}{\sum_{j=j_L}^{j_U}\int_0^1\left[\widehat{\Delta}_j^d\left(t\right)\right]^2\ \ D_j(t)dt}, \tag{10}$$

where $D_j(t) \equiv I\{a \leq 2^j(\widehat{\tau} - t) \leq b\}$ is the 'cone of influence', see p. 215 in Mallet (2009), $j_L \leq j_U$, and $j_L \equiv j_{Ln} \to \infty$ as $n \to \infty$.

The class of estimators in (10) include estimators using wavelet coefficients at a single scale and multiple locations, at multiscale and a single location, and at multiscale and multiple locations. We'll establish the asymptotic properties of these estimators in the rest of this section and then extend them to the corresponding results for estimators of LATE.

4.1.1 One scale with many locations

In this part, we consider the asymptotic properties of a subclass of wavelet OLS estimators for which only one scale is used. The fixed scale j_0 wavelet OLS estimator of δ is defined as:

$$\widehat{\delta}_{W1} = \frac{\int_0^1 \widehat{\Delta}^Y_{j_0}(t)\widehat{\Delta}^d_{j_0}(t) D_{j_0}(t)dt}{\int_0^1 \left[\widehat{\Delta}^d_{j_0}(t)\right]^2 D_{j_0}(t)dt}.$$

Assumption A5. (b)' As $n \to \infty$, $j_0 \to \infty$, $\frac{2^{j_0}}{n} \to 0$, and $\left(\frac{1}{2^{j_0}}\right)^{2m-1} \sqrt{\frac{n}{2^{j_0}}} \to C^b_{W1} < \infty$.

THEOREM 4.1 *Suppose A1, A3(G), and A4 hold. In addition,* $\int_{a-b}^0 M(v)dv \neq 0$, *where* $M(\cdot)$ *is defined below.*

(i) When A2(G) (a) and A5 (a) hold, we obtain:

$$\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_{W1} - \delta) \xrightarrow{d} N\left(C_a B^a_{W1}, V_{W1}\right),$$

where

$$\begin{aligned} B^a_{W1} &= \frac{\left[G^{(1)}_+(\tau) - G^{(1)}_-(\tau)\right] \int_a^b \int_a^b L(t)\psi(s)(s-t)I\{s-t \geq 0\}dsdt}{\int_{a-b}^0 M(v)dv}, \\ V_{W1} &= \frac{\sigma^2_+(v_0)\int_0^{b-a} M^2(v)dv + \sigma^2_-(v_0)\int_{a-b}^0 M^2(v)dv}{\left[\int_{a-b}^0 M(v)dv\right]^2}, \end{aligned}$$

in which

$$L(t) = \int_a^b I\{w \geq t\}\psi(w)dw \text{ and } M(v) = \int_a^b \int_a^b I\{w \geq t+v\}\psi(w)\psi(t)dtdw.$$

(ii) When A2(G) (b) with $l_G \geq 2m$ *and A5 (b)' hold, we obtain*

$$\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_{W1} - \delta) \xrightarrow{d} N\left(C^b_{W1} B^b_{W1}, V_{W1}\right),$$

where

$$B^b_{W1} = \frac{G^{(2m-1)}(\tau)\int_a^b \psi(s)s^m ds \int_a^b L(t)(-t)^{m-1}dt}{m!(m-1)!\int_{a-b}^0 M(v)dv}.$$

Theorems 3.1 and 4.1 reveal the role of the additional information in wavelet coefficients at locations other than τ. When A2(G)(b) holds with $l_G \geq 2m$, the use of additional wavelet coefficients ($\widehat{\delta}_{W1}$) reduces the order of the asymptotic bias of the wavelet estimator further to $O\left((2^{-j_0})^{2m-1}\right)$ instead of $O\left((2^{-j_0})^{m}\right)$ for $\widehat{\delta}$. This feature resembles that of the local polynomial estimator, see Porter (2003). The additional assumption that $l_G \geq 2m$ allows us to obtain an explicit expression for the asymptotic bias of $\widehat{\delta}_{W1}$. However, when only A2(G)(a) holds, the order of the asymptotic bias of $\widehat{\delta}_{W1}$ remains the same as that of $\widehat{\delta}$.

To estimate LATE, we estimate ζ by $\widehat{\zeta}_{W1}$ below:

$$\widehat{\zeta}_{W1} = \frac{\int_0^1 \widehat{\Delta}_{j_0}^D(t)\widehat{\Delta}_{j_0}^d(t)D_{j_0}(t)dt}{\int_0^1 \left[\widehat{\Delta}_{j_0}^d(t)\right]^2 D_{j_0}(t)dt}.$$

In the end, LATE is estimated by $\widehat{\theta}_{W1} = \widehat{\delta}_{W1}/\widehat{\zeta}_{W1}$.

THEOREM 4.2 *Suppose A1, A3, and A4 hold. In addition,* $\int_{a-b}^0 M(v)dv \neq 0$.

(i) When A2 (a) and A5 (a) hold, we obtain:

$$(n/2^{j_0})^{1/2}\left(\frac{\widehat{\delta}_{W1}}{\widehat{\zeta}_{W1}} - \frac{\delta}{\zeta}\right) \xrightarrow{d} N\left(\frac{1}{\zeta}C_a\left[B_{W1}^a - \frac{\delta}{\zeta}B_{aW1}^D\right], \frac{1}{\zeta^2}\left[V_{W1} - 2\frac{\delta}{\zeta}V_{W1}^{YD} + \frac{\delta^2}{\zeta^2}V_{W1}^D\right]\right),$$

where

$$\begin{aligned} B_{aW1}^D &= \frac{\left[H_+^{(1)}(\tau) - H_-^{(1)}(\tau)\right]\int_a^b\int_a^b L(t)\psi(s)(s-t)I\{s-t\geq 0\}dsdt}{\int_{a-b}^0 M(v)dv}, \\ V_{W1}^D &= \frac{\sigma_{\epsilon+}^2(v_0)\int_0^{b-a}M^2(v)dv + \sigma_{\epsilon-}^2(v_0)\int_{a-b}^0 M^2(v)dv}{\left[\int_{a-b}^0 M(v)dv\right]^2}, \\ V_{W1}^{YD} &= \frac{\sigma_{\varepsilon\epsilon+}^2(v_0)\int_0^{b-a}M^2(v)dv + \sigma_{\varepsilon\epsilon-}^2(v_0)\int_{a-b}^0 M^2(v)dv}{\left[\int_{a-b}^0 M(v)dv\right]^2}. \end{aligned}$$

(ii) When A2 (b) with $\min\{l_G, l_H\} \geq 2m$ *and A5 (b)' hold, we obtain***:**

$$\begin{aligned} &(n/2^{j_0})^{1/2}\left(\frac{\widehat{\delta}_{W1}}{\widehat{\zeta}_{W1}} - \frac{\delta}{\zeta}\right) \\ &\xrightarrow{d} N\left(\frac{1}{\zeta}\left[C_{W1}^b B_{W1}^b - \frac{\delta}{\zeta}C_{W1}^b B_{bW1}^D\right], \frac{1}{\zeta^2}\left[V_{W1} - \frac{2\delta}{\zeta}V_{W1}^{YD} + \frac{\delta^2}{\zeta^2}V_{W1}^D\right]\right), \end{aligned}$$

where

$$B_{bW1}^D = \frac{H^{(2m-1)}(\tau)\int_a^b \psi(s)s^m ds \cdot \int_a^b L(t)(-t)^{m-1}dt}{m!(m-1)!\int_{a-b}^0 M(v)dv}.$$

4.1.2 Multiscale with a single location

We now investigate the role of using more than one scale in estimating the LATE parameter.[8] First we consider the estimator of δ:

$$\widehat{\delta}_{W2} = \frac{\sum_{j=j_L}^{j_U} \widehat{\Delta}_j^Y(\widehat{\tau})\widehat{\Delta}_j^d(\widehat{\tau})}{\sum_{j=j_L}^{j_U} \left[\widehat{\Delta}_j^d(\widehat{\tau})\right]^2},$$

where $j_L < j_U$. Let $(j_U - j_L) = K_n$.

THEOREM 4.3 *Suppose A1, A3(G), and A4 hold. In addition, suppose $\sigma_{\varepsilon+}^2(v_0) = \sigma_{\varepsilon-}^2(v_0)$.*[9]

(i) When A2(G) (a) and A5 (a) hold for j_L and j_U, we obtain:

if $\lim_{n\to\infty} K_n < \infty$, then

$$\sqrt{\frac{n}{2^{j_L}}}(\widehat{\delta}_{W2} - \delta) \xrightarrow{d} N\left(C_a B_{W2}^a, \frac{V}{2\left[1 - (\frac{1}{2})^{\lim K_n + 1}\right]}\right),$$

where

$$B_{W2}^a = \left[1 + (\frac{1}{2})^{\lim K_n + 1}\right] \frac{2B_a}{3};$$

if $\lim_{n\to\infty} K_n = \infty$, then

$$\sqrt{\frac{n}{2^{j_L}}}(\widehat{\delta}_{W2} - \delta) \xrightarrow{d} N\left(\frac{2}{3}C_a B_a, \frac{V}{2}\right),$$

(ii) When A2(G) (b) and A5 (b) hold for j_L and j_U, we obtain:

if $\lim_{n\to\infty} K_n < \infty$, then

$$\sqrt{\frac{n}{2^{j_L}}}(\widehat{\delta}_{W2} - \delta) \xrightarrow{d} N\left(C_b B_{W2}^b, \frac{V}{2\left[1 - (\frac{1}{2})^{\lim K_n + 1}\right]}\right),$$

where

$$B_{W2}^b = \frac{\left[1 - (\frac{1}{2})^{(m+1)(\lim K_n + 1)}\right]}{2\left[1 - (\frac{1}{2})^{\lim K_n + 1}\right]\left[1 - (\frac{1}{2})^{m+1}\right]} B_b;$$

if $\lim_{n\to\infty} K_n = \infty$, then

$$\sqrt{\frac{n}{2^{j_L}}}(\widehat{\delta}_{W2} - \delta) \xrightarrow{d} N\left(C_b \frac{1}{2\left[1 - (\frac{1}{2})^{m+1}\right]} B_b, \frac{V}{2}\right).$$

---

[8] This has the flavor of Kotlyarova and Zinde-Walsh (2006, 2008) which average kernel density estimators using different bandwidths.

[9] For notational compactness, we only report results when $\sigma_{\varepsilon+}^2(v_0) = \sigma_{\varepsilon-}^2(v_0)$ in the main text. General results without this assumption can be found in the proof of this theorem in Appendix C.

Theorems 3.1, 4.1, and 4.3 reveal the interesting effects of using additional information in wavelet coefficients at multiple scales and multiple locations. First, the multiple locations estimator $\widehat{\delta}_{W1}$ reduces the order of the asymptotic bias of $\widehat{\delta}$ under A2(G)(b); Second, the multiple scales estimator $\widehat{\delta}_{W2}$ reduces the asymptotic bias and variance of the wavelet estimator $\widehat{\delta}$ only proportionally, but under both A2(G)(a) and A2(G)(b).

To estimate the LATE parameter, we let

$$\widehat{\zeta}_{W2} = \frac{\sum_{j=j_L}^{j_U} \widehat{\Delta}_j^D(\tau)\widehat{\Delta}_j^d(\tau)}{\sum_{j=j_L}^{j_U} \left[\widehat{\Delta}_j^d(\tau)\right]^2}.$$

THEOREM 4.4 *Suppose A1, A3, and A4 hold. In addition, suppose* $\sigma^2_{\varepsilon+}(v_0) = \sigma^2_{\varepsilon-}(v_0)$, $\sigma^2_{\epsilon+}(v_0) = \sigma^2_{\epsilon-}(v_0)$, *and* $\sigma^2_{\varepsilon\epsilon+}(v_0) = \sigma^2_{\varepsilon\epsilon-}(v_0)$.

(a) When A2 (a) and A5 (a) hold for j_L *and* j_U*, we obtain:*

if $\lim_{n\to\infty} K_n < \infty$*, then*

$$(n/2^{j_L})^{1/2}\left(\frac{\widehat{\delta}_{W2}}{\widehat{\zeta}_{W2}} - \frac{\delta}{\zeta}\right) \xrightarrow{d} N\left(\frac{1}{\zeta}C_a\left[B^a_{W2} - \frac{\delta}{\zeta}B^{Da}_{W2}\right], \frac{V - \frac{2\delta}{\zeta}V^{YD} + \frac{\delta^2}{\zeta^2}V^D}{2\left[1-(\frac{1}{2})^{\lim K_n+1}\right]\zeta^2}\right),$$

where

$$B^{Da}_{W2} = \left[1 + (\frac{1}{2})^{\lim K_n+1}\right]\frac{2B^D_a}{3};$$

if $\lim_{n\to\infty} K_n = \infty$*, then*

$$(n/2^{j_L})^{1/2}\left(\frac{\widehat{\delta}_{W2}}{\widehat{\zeta}_{W2}} - \frac{\delta}{\zeta}\right) \xrightarrow{d} N\left(\frac{2}{3\zeta}C_a\left[B_a - \frac{\delta}{\zeta}B^D_a\right], \frac{V - \frac{2\delta}{\zeta}V^{YD} + \frac{\delta^2}{\zeta^2}V^D}{2\zeta^2}\right).$$

(b) When A2 (b) and A5 (b) hold for j_L *and* j_U*, we obtain:*

if $\lim_{n\to\infty} K_n < \infty$*, then*

$$(n/2^{j_L})^{1/2}\left(\frac{\widehat{\delta}_{W2}}{\widehat{\zeta}_{W2}} - \frac{\delta}{\zeta}\right) \xrightarrow{d} N\left(\frac{1}{\zeta}C_b\left[B^b_{W2} - \frac{\delta}{\zeta}B^{Db}_{W2}\right], \frac{V - \frac{2\delta}{\zeta}V^{YD} + \frac{\delta^2}{\zeta^2}V^D}{2\left[1-(\frac{1}{2})^{\lim K_n+1}\right]\zeta^2}\right),$$

where

$$B^{Db}_{W2} = \frac{\left[1-(\frac{1}{2})^{(m+1)(\lim K_n+1)}\right]}{2\left[1-(\frac{1}{2})^{\lim K_n+1}\right]\left[1-(\frac{1}{2})^{m+1}\right]}B^D_b;$$

if $\lim_{n\to\infty} K_n = \infty$, *then*

$$(n/2^{j_L})^{1/2}\left(\frac{\widehat{\delta}_{W2}}{\widehat{\zeta}_{W2}} - \frac{\delta}{\zeta}\right)$$

$$\xrightarrow{d} N\left(C_b \frac{1}{2\zeta\left[1-(\frac{1}{2})^{m+1}\right]}\left[B_b - \frac{\delta}{\zeta}B_b^D\right], \frac{V - \frac{2\delta}{\zeta}V^{YD} + \frac{\delta^2}{\zeta^2}V^D}{2\zeta^2}\right).$$

4.1.3 Multiscale with many locations

We now establish the asymptotic distribution of the general estimator $\widehat{\delta}_W$ in (10):

$$\widehat{\delta}_W = \frac{\sum_{j=j_L}^{j_U}\int_0^1 \widehat{\Delta}_j^Y(t)\,\widehat{\Delta}_j^d(t)\,D_j(t)dt}{\sum_{j=j_L}^{j_U}\int_0^1\left[\widehat{\Delta}_j^d(t)\right]^2 D_j(t)dt}.$$

Again for notational compactness, we only establish the asymptotic distribution of $\widehat{\delta}_W$ under the condition that $\sigma^2_{\varepsilon+}(v_0) = \sigma^2_{\varepsilon-}(v_0)$.

THEOREM 4.5 *Suppose A1, A3, and A4 hold. In addition, suppose* $\sigma^2_{\varepsilon+}(v_0) = \sigma^2_{\varepsilon-}(v_0)$.

(i) When A2(G) (a) and A5(a) hold for j_L *and* j_U*, we obtain:*

if $\lim_{n\to\infty} K_n < \infty$, *then*

$$\sqrt{\frac{n}{2^{j_L}}}(\widehat{\delta}_W - \delta) \xrightarrow{d} N\left(C_a B_W^a, V_W\right),$$

where

$$\begin{aligned} B_W^a &= \frac{6}{7}\frac{\left[1-(\frac{1}{8})^{\lim K_n+1}\right]}{\left[1-(\frac{1}{4})^{\lim K_n+1}\right]}B_{W1}^a, \\ V_W &= \frac{9}{14}\frac{1-(\frac{1}{8})^{\lim K_n+1}}{\left[1-(\frac{1}{4})^{\lim K_n+1}\right]^2}\ V_{W1}; \end{aligned}$$

if $\lim_{n\to\infty} K_n = \infty$, *then*

$$\sqrt{\frac{n}{2^{j_L}}}(\widehat{\delta}_W - \delta) \xrightarrow{d} N\left(C_a B_W^{a*}, V_W^*\right),$$

where

$$B_W^{a*} = \frac{6}{7}B_{W1}^a, V_W^* = \frac{9}{14}\ V_{W1}.$$

(ii) When A2(G) (b) with $l_G \geq 2m$ *and (A5) (b)' hold for* j_L *and* j_U*, we obtain:*

if $\lim_{n\to\infty} K_n < \infty$, *then*

$$\sqrt{\frac{n}{2^{j_L}}}(\widehat{\delta}_W - \delta) \xrightarrow{d} N\left(C_{W1}^b B_W^b, V_W\right),$$

where

$$B_W^b = \frac{3}{4\left[1-(\frac{1}{2})^{2m+1}\right]}\frac{1-(\frac{1}{2})^{(2m+1)(\lim K_n+1)}}{1-(\frac{1}{4})^{\lim K_n+1}}\frac{G^{(2m-1)}(\tau)\int_a^b\psi(s)s^m ds\int_a^b L(t)(-t)^{m-1}dt}{m!(m-1)!\int_{a-b}^0 M(v)dv},$$

if $\lim_{n\to\infty} K_n = \infty$*, then*

$$\sqrt{\frac{n}{2^{j_L}}}(\widehat{\delta}_W - \delta) \xrightarrow{d} N\left(C_{W1}^{b} B_W^{b*}, V_W^{*}\right),$$

where

$$B_W^{b*} = \frac{3}{4\left[1-(\frac{1}{2})^{2m+1}\right]} \frac{G^{(2m-1)}(\tau)\int_a^b \psi(s)s^m ds \int_a^b L(t)(-t)^{m-1}dt}{m!(m-1)!\int_{a-b}^0 M(v)dv}.$$

As expected, $\widehat{\delta}_W$ inherit the properties of $\widehat{\delta}_{W1}$and $\widehat{\delta}_{W2}$: when A2(G)(b) holds, it reduces the order of the asymptotic bias of $\widehat{\delta}$ or $\widehat{\delta}_{W2}$ reflecting the additional information in the multiple locations used in $\widehat{\delta}_W$ and under both A2(G)(a) and A2(G)(b), the asymptotic bias and variance of $\widehat{\delta}_W$ are proportionally smaller than those of $\widehat{\delta}_{W1}$ reflecting the additional information in the multiple scales used in $\widehat{\delta}_W$.

Define

$$\widehat{\zeta}_W = \frac{\sum_{j=j_L}^{j_U}\int_0^1 \widehat{\Delta}_j^D(t)\,\widehat{\Delta}_j^d(t)\,D_j(t)dt}{\sum_{j=j_L}^{j_U}\int_0^1\left[\widehat{\Delta}_j^d(t)\right]^2 D_j(t)dt}.$$

Our estimator of LATE is given by $\widehat{\delta}_W/\ \widehat{\zeta}_W$.

THEOREM 4.6 *Suppose A1, A3, and A4 hold. In addition, suppose* $\sigma_{\varepsilon+}^2(v_0) = \sigma_{\varepsilon-}^2(v_0)$, $\sigma_{\epsilon+}^2(v_0) = \sigma_{\epsilon-}^2(v_0)$*, and* $\sigma_{\varepsilon\epsilon+}^2(v_0) = \sigma_{\varepsilon\epsilon-}^2(v_0)$*.*

(i) When A2 (a) and A5 (a) hold for j_L *and* j_U*, we obtain:*

if $\lim_{n\to\infty} K_n < \infty$*, then*

$$(n/2^{j_L})^{1/2}\left(\frac{\widehat{\delta}_W}{\widehat{\zeta}_W} - \frac{\delta}{\zeta}\right) \xrightarrow{d} N\left(\frac{1}{\zeta}C_a\left[B_W^a - \frac{\delta}{\zeta}B_W^{Da}\right], \frac{1}{\zeta^2}\left[V_W - \frac{2\delta}{\zeta}V_W^{YD} + \frac{\delta^2}{\zeta^2}V_W^D\right]\right),$$

where

$$\begin{aligned} B_W^{Da} &= \frac{6}{7}\frac{\left[1-(\frac{1}{8})^{\lim K_n+1}\right]}{\left[1-(\frac{1}{4})^{\lim K_n+1}\right]}B_{aW1}^D, \\ V_W^D &= \frac{9}{14}\frac{1-(\frac{1}{8})^{\lim K_n+1}}{\left[1-(\frac{1}{4})^{\lim K_n+1}\right]^2}V_{W1}^D, \end{aligned}$$

if $\lim_{n\to\infty} K_n = \infty$*, then*

$$(n/2^{j_L})^{1/2}\left(\frac{\widehat{\delta}_W}{\widehat{\zeta}_W} - \frac{\delta}{\zeta}\right) \xrightarrow{d} N\left(\frac{1}{\zeta}C_a\left[B_W^{a*} - \frac{\delta}{\zeta}B_{aW}^{D*}\right], \frac{1}{\zeta^2}\left[V_W^* - 2\frac{\delta}{\zeta}V_W^{YD*} + \frac{\delta^2}{\zeta^2}V_W^{D*}\right]\right),$$

where

$$\begin{aligned} B_{aW}^{D*} &= \frac{6}{7}B_{aW1}^D, \\ V_W^{YD} &= \frac{9}{14}\frac{1-(\frac{1}{8})^{K_n+1}}{\left[1-(\frac{1}{4})^{K_n+1}\right]^2}V_{W1}^{YD}, \\ V_W^{D*} &= \frac{9}{14}V_{W1}^D,\ V_{W3}^{YD*} = \frac{9}{14}V_{W1}^{YD}. \end{aligned}$$

(ii) When A2 (b) with $\min\{l_G, l_H\} \geq 2m$ *and (A5)(b)' hold for* j_L *and* j_U*, we obtain:*

if $\lim_{n\to\infty} K_n < \infty$*, then*

$$(n/2^{j_L})^{1/2}\left(\frac{\widehat{\delta}_W}{\widehat{\zeta}_W} - \frac{\delta}{\zeta}\right) \xrightarrow{d} N\left(\frac{1}{\zeta}C_a\left[B_W^b - \frac{\delta}{\zeta}B_W^{Db}\right], \frac{1}{\zeta^2}\left[V_W - \frac{2\delta}{\zeta}V_W^{YD} + \frac{\delta^2}{\zeta^2}V_W^D\right]\right),$$

where

$$B_W^{Db} = \frac{3}{4\left[1-(\frac{1}{2})^{2m+1}\right]}\frac{1-(\frac{1}{2})^{(2m+1)(\lim K_n+1)}}{1-(\frac{1}{4})^{\lim K_n+1}}\frac{H^{(2m-1)}(\tau)\int_a^b \psi(s)s^m ds\int_a^b L(t)(-t)^{m-1}dt}{m!(m-1)!\int_{a-b}^0 M(v)dv};$$

if $\lim_{n\to\infty} K_n = \infty$*, then*

$$(n/2^{j_L})^{1/2}\left(\frac{\widehat{\delta}_W}{\widehat{\zeta}_W} - \frac{\delta}{\zeta}\right) \xrightarrow{d} N\left(\frac{1}{\zeta}C_{W1}^b\left[B_W^{b*} - \frac{\delta}{\zeta}B_{bW}^{D*}\right], \frac{1}{\zeta^2}\left[V_W^* - 2\frac{\delta}{\zeta}V_W^{YD*} + \frac{\delta^2}{\zeta 2}V_W^{D*}\right]\right),$$

where

$$B_{bW}^{D*} = \frac{3}{4\left[1-(\frac{1}{2})^{2m+1}\right]}\frac{H^{(2m-1)}(\tau)\int_a^b \psi(s)s^m ds\int_a^b L(t)(-t)^{m-1}dt}{m!(m-1)!\int_{a-b}^0 M(v)dv}.$$

4.2 Kink Incentive Assignment Mechanism

All three wavelet OLS estimators for discontinuous incentive assignment mechanisms proposed in Section 4.1 can be extended to kink incentive assignment mechanisms and they share the same qualitative properties. To illustrate, we present a detailed analysis of the wavelet OLS estimator based on wavelet coefficients at a single scale and many locations and report results on the other estimators in a separate paper.

Let

$$\widehat{\delta}_{KW1} = \frac{\int_0^1 \widehat{\Delta}_{j_0}^Y(t)\widehat{\Delta}_{j_0}^k(t)D_{j_0}(t)dt}{\int_0^1\left[\widehat{\Delta}_{j_0}^k(t)\right]^2 D_{j_0}(t)dt}$$

and

$$\overline{\delta}_{KW1} = \frac{1}{n}\sum_{i=1}^n J_{KW1}(\frac{i}{n})Y_{i:n}$$

where

$$J_{KW1}(\frac{i}{n}) = \frac{\int_0^1\int_0^1 D_{j_0}(t)\left(F_V^{-1}(w)-v_0\right)I\{w\geq\tau\}2^{j_0}\psi\left[2^{j_0}(w-t)\right]\psi\left[2^{j_0}(\frac{i}{n}-t)\right]dtdw}{\int_0^1\int_0^1\int_0^1\left\{\begin{array}{c} D_{j_0}(t)\left(F_V^{-1}(w)-v_0\right)I\{w\geq\tau\}\left(F_V^{-1}(v)-v_0\right) \\ \times I\{v\geq\tau\}2^{j_0}\psi\left[2^{j_0}(w-t)\right]\psi\left[2^{j_0}(v-t)\right]\end{array}\right\}dwdvdt}.$$

We first show that $\widehat{\delta}_{KW1}$ has the same asymptotic distribution as $\overline{\delta}_{KW1}$ and then establish the asymptotic distribution of $\overline{\delta}_{KW1}$.

Assumption A5K. (b)' As $n\to\infty$, $j_0\to\infty$, $\frac{2^{3j_0}}{n}\to 0$, and $\left(\frac{1}{2^{j_0}}\right)^{2m-1}\sqrt{\frac{n}{2^{3j_0}}}\to C_{K'W1}^b<\infty$.

Assumption A6K. (c) $F_V^{-1}(v)$ is $2m$ times continuously differentiable on the support of V.

THEOREM 4.7 *Suppose A1, A3(G), and A4K hold. Let* $M_{12}(s) = M_1(s) + M_2(s) - sM(s)$, *where*

$$\begin{aligned} M(s) &= \int_a^b \int_a^b I\{w \geq t+s\}\psi(w)\psi(t)dtdw, \\ M_1(s) &= \int_a^b \int_a^b (-t)I\{w \geq t+s\}\psi(w)\psi(t)dtdw, \\ M_2(s) &= \int_a^b \int_a^b wI\{w \geq t+s\}\psi(w)\psi(t)dtdw. \end{aligned}$$

Assume $\int_{a-b}^0 M_{12}(t)\,dt \neq 0$.

(i) When A2K(G) (a), A5K (a) and A6K(a) hold, we obtain:

$$\sqrt{\frac{n}{2^{3j_0}}}(\widehat{\delta}_{KW1} - \delta_K) \xrightarrow{d} N\left(C_{Ka}B^a_{KW1}, V_{KW1}\right),$$

where

$$\begin{aligned} & B^a_{KW1} \\ = & -\frac{\left[G^{(2)}_{K+}(\tau) - G^{(2)}_{K-}(\tau)\right] f_V(v_0) \int_a^b \int_a^b (s-w)^2\psi(s)I\{s-w \geq 0\}\left[L_1(w) - wL_0(w)\right] dsdw}{2\int_{a-b}^0 M_{12}(t)\,dt}, \end{aligned}$$

$$V_{KW1} = \frac{f^2_V(v_0)\left[\sigma^2_{\varepsilon+}(v_0)\int_{a-b}^0 M^2_{12}(t)\,dt + \sigma^2_{\varepsilon-}(v_0)\int_0^{b-a} M^2_{12}(t)\,dt\right]}{\left[\int_{a-b}^0 M_{12}(t)\,dt\right]^2},$$

in which, for $i = 0, 1, \cdots, m-1$, L_i *defined below has* $(m-i)$ *vanishing moments:*

$$L_i(2^{j_0}(\tau - t)) = \int_a^b w^i I\{w \geq 2^{j_0}(\tau - t)\}\psi(w)dw.$$

(ii) When A2K(G) (b) with $l_G \geq 2m$, *A5K (b)' and A6K(c) hold, we obtain:*

$$\sqrt{\frac{n}{2^{3j_0}}}(\widehat{\delta}_{kW1} - \delta_k) \xrightarrow{d} N\left(C^b_{K'W1}B^b_{KW1}, V_{KW1}\right),$$

where

$$B^b_{KW1} = -f^2_V(v_0)\frac{\int_a^b \psi(s)s^{m+1}ds\left[\Gamma_0 + \sum_{i=1}^m \Gamma_i\right]}{\int_{a-b}^0 M_{12}(t)\,dt},$$

in which

$$\begin{aligned} \Gamma_0 &= \int_a^b L_0(t)(-t)^m dt\left[\sum_{i=1}^m [F_V^{-1}(\tau)]^{(i)}\frac{G_K^{(2m+1-i)}(\tau)}{i!(m+1)!(m-i)!}\right], \\ \Gamma_i &= \frac{1}{i!}\int_a^b L_i(t)(-t)^{m-i}dt\left[\sum_{l=i}^m [F_V^{-1}(\tau)]^{(l)}\frac{G_K^{(2m-l+1)}(\tau)}{(m-i)!(m+1)!(m-l)!}\right] \text{ for } i \geq 1. \end{aligned}$$

Now we provide the LATE estimator: $\widehat{\delta}_{KW1}/\widehat{\zeta}_{KW1}$, where

$$\widehat{\zeta}_{KW1} = \frac{\int_0^1 \widehat{\Delta}^D_{j_0}(t)\widehat{\Delta}^k_{j_0}(t)D_{j_0}(t)dt}{\int_0^1 \left[\widehat{\Delta}^k_{j_0}(t)\right]^2 D_{j_0}(t)dt}.$$

THEOREM 4.8 *Suppose A1, A3, and A4K hold.*

(a) When A2K (a), A5K (a) and A6K(a) hold, we obtain:

$$\sqrt{\frac{n}{2^{3j_0}}}\left(\frac{\widehat{\delta}_{KW1}}{\widehat{\zeta}_{KW1}}-\frac{\delta_K}{\zeta_K}\right)$$

$$\xrightarrow{d} N\left(\frac{1}{\zeta}C_{Ka}\left[B^{a}_{KW1}-\frac{\delta}{\zeta}B^{Da}_{KW1}\right],\frac{1}{\zeta^2}\left[V_{KW1}-\frac{2\delta}{\zeta}V^{YD}_{KW1}+\frac{\delta^2}{\zeta^2}V^{D}_{KW1}\right]\right),$$

where

$$\begin{aligned}
B^{Da}_{KW1} &= -\frac{\left[H^{(2)}_{K+}(\tau)-H^{(2)}_{K-}(\tau)\right]f_V(v_0)\int_a^b\int_a^b(s-w)^2\psi(s)I\{s-w\geq 0\}\left[L_1(w)-wL_0(w)\right]dsdw}{2\int_{a-b}^{0}M_{12}\left(t\right)dt},\\
V^{D}_{KW1} &= \frac{f_V^2(v_0)\left[\sigma^2_{\epsilon+}(v_0)\int_{a-b}^{0}M_{12}^2\left(t\right)dt+\sigma^2_{\epsilon-}(v_0)\int_0^{b-a}M_{12}^2\left(t\right)dt\right]}{\left[\int_{a-b}^{0}M_{12}\left(t\right)dt\right]^2},\\
V^{YD}_{KW1} &= \frac{f_V^2(v_0)\left[\sigma^2_{\epsilon\varepsilon+}(v_0)\int_{a-b}^{0}M_{12}^2\left(t\right)dt+\sigma^2_{\epsilon\varepsilon-}(v_0)\int_0^{b-a}M_{12}^2\left(t\right)dt\right]}{\left[\int_{a-b}^{0}M_{12}\left(t\right)dt\right]^2}.
\end{aligned}$$

(b) When A2K (b) with $\min\{l_G,l_H\}\geq 2m$*, A5K(b)' and A6K(c) hold, we obtain:*

$$\sqrt{\frac{n}{2^{3j_0}}}\left(\frac{\widehat{\delta}_{KW1}}{\widehat{\zeta}_{KW1}}-\frac{\delta_K}{\zeta_K}\right)$$

$$\xrightarrow{d} N\left(\frac{1}{\zeta}C^{b}_{K'W1}\left[B^{b}_{KW1}-\frac{\delta}{\zeta}B^{Db}_{KW1}\right],\frac{1}{\zeta^2}\left[V_{KW1}-\frac{2\delta}{\zeta}V^{YD}_{KW1}+\frac{\delta^2}{\zeta^2}V^{D}_{KW1}\right]\right),$$

where

$$B^{Db}_{KW1}=-f_V^2(v_0)\frac{\int_a^b\psi(s)s^{m+1}ds\left[\Lambda_0+\sum_{i=1}^{m}\Lambda_i\right]}{\int_{a-b}^{0}M_{12}\left(t\right)dt},$$

in which

$$\begin{aligned}
\Lambda_0 &= \int_a^b L_0(t)(-t)^m dt\left[\sum_{i=1}^{m}[F_V^{-1}(\tau)]^{(i)}\frac{H_K^{(2m+1-i)}(\tau)}{i!(m+1)!(m-i)!}\right],\\
\Lambda_i &= \frac{1}{i!}\int_a^b L_i(t)(-t)^{m-i}dt\left[\sum_{l=i}^{m}[F_V^{-1}(\tau)]^{(l)}\frac{H_K^{(2m-l+1)}(\tau)}{(m-i)!(m+1)!(m-l)!}\right]\ \textit{for } i\geq 1.
\end{aligned}$$

5 Monte-Carlo Simulation

This section presents results from a small simulation study. We focus on the finite sample performances of our wavelet estimators of the jump size in two classes of models. The first class includes switching regime models and the second class reduced form auxiliary regression models.

The two switching regime models are:

Model 1.

$$\begin{aligned} Y_1 &= 1.25 + V + V^2 + W_1,\ Y_0 = V + 2V^2 + W_2, \\ D &= I\{V \geq 0.5\}. \end{aligned}$$

Model 2.

$$\begin{aligned} Y_1 &= 1 + V^7 + W_1,\ Y_0 = V^7 + W_2, \\ D &= I\{V \geq 0.5\}. \end{aligned}$$

In both models, $V \sim U\ [0,1]$, $(W_1, W_2) \sim N\left[(0,0), \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}\right]$, and V is independent of (W_1, W_2). Lengthy algebras show that the corresponding auxiliary regression models are:

$$\text{Model 1} : Y = \begin{cases} V + 2V^2 + W,\ 0 \leq V < 0.5 \\ 1.25 + V + V^2 + W,\ 0.5 \leq V \leq 1 \end{cases},$$

and

$$\text{Model 2} : Y = \begin{cases} V^7 + W,\ 0 \leq V < 0.5 \\ 1 + V^7 + W,\ 0.5 \leq V \leq 1 \end{cases},$$

where $W \sim N(0, 0.01)$. The auxiliary regression functions indicate that Model 1 satisfies A2(G)(a) and Model 2 satisfies A2(G)(b).

We also generate data directly from the two auxiliary regression models below:

Model 3.

$$Y = \begin{cases} V + W,\ 0 \leq V < 0.5 \\ 0.5 + 2V + W,\ 0.5 \leq V \leq 1 \end{cases}.$$

Model 4.

$$Y = \begin{cases} V + W,\ 0 \leq V < 0.5 \\ 1 + V + W,\ 0.5 \leq V \leq 1 \end{cases}.$$

In both models, $V \sim U\ [0,1]$, $W \sim N(0, 0.01)$, and V is independent of W. Obviously Model 3 satisfies A2(G)(a) and Model 4 satisfies A2(G)(b).

From each model, we generate random samples of sizes 500, 2,500, 5,000 respectively and computed our wavelet estimates using Daub4. Daub4 has 4 vanishing moments supported on $[-3, 4]$. All four estimators depend on the choice of the scale. For single scale estimators, we chose six scale levels, $1, 2, 3, 4, 5, 6$, while for many scales estimators, we chose $j_L = 1, 2, 3, 4, 5, 6$ and $K_n = 2$. We repeated this for 5,000 times and computed the bias, standard deviation, and MSE of each estimator. To save space, results for samples of sizes 500 and 2,500 are reported in Tables 1-4.

Insert tables here.

Tables 1-4 reveal the same qualitative behavior of each estimator for all models. First, as the sample size increases, the MSEs of all estimators decrease; Second, overall many locations estimator $\widehat{\delta}_{W1}$ performs much better than the single location estimator $\widehat{\delta}$ in terms of bias, standard error, and MSE; Third, many scales estimator $\widehat{\delta}_{W2}$ performs better than the single scale estimator $\widehat{\delta}$, but the reduction in MSE is not as much as that of the many locations estimator $\widehat{\delta}_{W1}$ in comparison with the single scale and single location estimator $\widehat{\delta}$; Fourth, for all estimators and all models, as the scale level increases, the MSE decreases initially and then begins to increase. It seems that for most cases considered, the optimal scale level is either 3 or 4.[10] Overall, the numerical results confirm our theoretical findings that it is advantageous to use more locations and more scales in estimating the jump size compared with single scale and single location estimator currently available in the literature and that our many scales and many locations estimator $\widehat{\delta}_W$ performs the best whether A2(G)(a) or A2(G)(b) holds.

6 Conclusion

In this paper, we have studied the identification of LATE in two classes of switching regime models. Both allow for individuals to make decision based on not only incentives assigned to them but also their unobserved characteristics. The first class of switching regime models accounts for discontinuous incentive assignment mechanisms and the second accounts for kink incentive assignment mechanisms. For each class of switching regime models, we established auxiliary regressions for estimating LATE based on which we have presented a systematic treatment of wavelet estimation of LATE. In addition to making use of the existing wavelet estimator of the jump size or kink size, we have developed new wavelet OLS estimators improving upon the existing wavelet estimator by employing more wavelet transform coefficients. The asymptotic properties of all the estimators are established and their finite sample properties are investigated via a simulation study.

This paper has focused on incentive assignment mechanisms depending on one forcing variable and having a single known cut-off. In some empirical applications, the cut-off point may be unknown to the econometrician and there may be more than one forcing variables. For example, Hoekstra (2009) applied RDD to studying the effect of attending the flagship state university on earnings. For the university and data set he used, the admission's cut-off depends on both SAT score and high school GPA. Hoekstra (2009) constructed an adjusted SAT score for a given GPA and estimated a parametric model with the adjusted SAT score as the forcing variable. Since the university didn't keep records of the exact admission rules used, the cut-off point is unknown and estimated. It would be interesting to extend the wavelet OLS estimators proposed in this paper to allow for unknown cut-off and/or more than one forcing variables.

---

[10] We have limited results on the selection of the optimal scale. For space considerations, we will report details on this in a separate paper.

This paper also suppresses other covariates X (say) in the potential outcomes equations and the selection equation. An extension of the model (1) and (2) accounting for the presence of other covariates is:

$$\begin{aligned} Y_1 &= g_1(X, V, W), \; Y_0 = g_0(X, V, W), \quad (11) \\ D &= I\{b(V) + g_3(X) - U \geq 0\}, \quad (12) \end{aligned}$$

where g_3 is an unknown function of X. Under appropriate conditions, the auxiliary regressions established earlier for estimating LATE in both discontinuous and kink incentive assignment mechanisms still hold and our wavelet estimators still apply. Alternatively, one may take into account the observable covariates X in estimating LATE. This may be done by making use of the alternative auxiliary regressions:

$$\begin{aligned} Y &= g(X, V) + \delta(X) I\{V \geq v_0\} + \varepsilon, \\ D &= h(X, V) + \zeta(X) I\{V \geq v_0\} + \epsilon, \end{aligned}$$

where $E[\varepsilon|X, V] = 0$ and $E[\epsilon|X, V] = 0$. Frölich (2007) proposes a local linear estimator taking into account the covariate X and compares it with the local linear estimator without using X. The asymptotic analysis in Frölich (2007) seems to suggest that using the covariate X may not always improve the performance of the LATE estimator. It would be interesting to extend our wavelet estimators to take into account the covariate X. We'll leave this to future research.

Appendix A: Technical Proofs and Lemmas Used in Section 2

Lemma A.1 *For any a, b satisfying: $-\infty \le a < b \le \infty$ and $[a,b] \subseteq \mathcal{U}$, we have:*

$$f_{W|V,A}(w|v,A) = \frac{\int_a^b f_{V|W,U}(v|w,u) f_{W,U}(w,u)du}{f_V(v)\int_a^b f_{U|V}(u|v)\, du} \ ,$$

where $A = \{\ a < U \le b\}$.

Proof. By definition, we have

$$\begin{aligned} & f_{W|V,A}(w|v,A) \\ = \ & \frac{1}{\Pr(a<U\le b|V=v)} \frac{\partial \{\Pr(W\le w|V=v, a<U\le b)\Pr(a<U\le b|V=v)\}}{\partial w} \\ = \ & \frac{1}{\Pr(a<U\le b|V=v)} \frac{\partial \{\Pr(W\le w, a<U\le b|V=v)\}}{\partial w} \\ = \ & \frac{1}{\Pr(a<U\le b|V=v)} \frac{\partial}{\partial w}\left\{\int_{-\infty}^{w}\int_a^b f_{W,U|V}(w',u|v)dw'du\right\} \\ = \ & \frac{1}{\Pr(a<U\le b|V=v)} \int_a^b f_{W,U|V}(w,u|v)du. \end{aligned}$$

Q.E.D.

Lemma A.2 *Under the conditions of Theorem 2.1, we get: for $j = 0, 1$,*

$$\lim_{v_+, v_- \to v_0} E(Y_j|V=v_+, A(v_+,v_-)) = E(Y_j|V=v_0, A),$$

where $\{A(v_+,v_-), A\} = \{\{b(v_-) < U \le b(v_+)\}, \{b^- < U \le b^+\}\}$, or $\{\{U \le b(v_-)\}, \{U \le b^-\}\}$, or $\{\{U > b(v^+)\}, \{U > b^+\}\}$.

Proof. Without loss of generality, we provide the proof for $j = 0$ and

$$\{A(v_+,v_-), A\} = \{\{b(v_-) < U \le b(v_+)\}, \{b^- < U \le b^+\}\}.$$

By definition,

$$E(Y_0|V=v_+, A(v_+,v_-)) = \int g_0(v_+,w) f_{W|V,A(v_+,v_-)}(w|v_+, A(v_+,v_-))dw.$$

Lemma A.1 implies:

$$f_{W|V,A(v_+,v_-)}(w|v_+, A(v_+,v_-)) = \frac{\int_{b(v_-)}^{b(v_+)} f_{V|W,U}(v_+|w,u) f_{W,U}(w,u)du}{f_V(v_+)\int_{b(v_-)}^{b(v_+)} f_{U|V}(u|v_+)\, du}.$$

Then

$$\begin{aligned} & f_{W|V,A(v_+,v_-)}(w|v_+, A(v_+,v_-)) \\ = \ & \frac{\int_{b(v_-)}^{b(v_+)} f_{U|V,W}(u|v_+,w)du f_{V,W}(v_+,w)}{f_V(v_+)\left[F_{U|V}(b(v_+)|v_+) - F_{U|V}(b(v_-)|v_-)\right]} \\ = \ & \frac{\left[F_{U|V,W}(b(v_+)|v_+,w) - F_{U|V,W}(b(v_-)|v_+,w)\right] f_{V|W}(v_+|w) f_W(w)}{f_V(v_+)\left[F_{U|V}(b(v_+)|v_+) - F_{U|V}(b(v_-)|v_-)\right]}. \end{aligned}$$

It follows from Conditions D1, D4 and D5 that

$$
\begin{aligned}
& \lim_{v_+,v_-\to v_0} f_{W|V,A(v_+,v_-)}(w|v_+, A\left(v_+,v_-\right)) \\
= \quad & \frac{\lim_{v_+,v_-\to v_0}\left\{\left[F_{U|V,W}(b\left(v_+\right)|v_+,w) - F_{U|V,W}(b\left(v_-\right)|v_+,w)\right] f_{V|W}\left(v_+|w\right) f_W\left(w\right)\right\}}{\lim_{v_+,v_-\to v_0}\left\{f_V(v_+)\left[F_{U|V}\left(b\left(v_+\right)|v_+\right) - F_{U|V}\left(b\left(v_-\right)|v_-\right)\right]\right\}} \\
= \quad & \frac{[F_{U|V,W}(b^+|v_0,w) - F_{U|V,W}(b^-|v_0,w)]\cdot f_{V|W}\left(v_0|w\right)\cdot f_W\left(w\right)}{\left[F_{U|V}\left(b^+\right) - F_{U|V}\left(b^-\right)\right]\cdot f_V(v_0)} \\
= \quad & \frac{\int_{b^-}^{b^+} f_{V|W,U}(v_0|w,u) f_{W,U}(w,u)du}{f_V(v_0)\int_{b^-}^{b^+} f_{U|V}\left(u|v_0\right)du}.
\end{aligned}
$$

Thus by Condition D2, Condition D3, and the dominated convergence theorem, we get

$$
\begin{aligned}
\lim_{v_+,v_-\to v_0} E(Y_0|V \quad &= \quad v_+, A\left(v_+,v_-\right)) \\
&= \quad \lim_{v_+,v_-\to v_0}\int g_0(v_+,w) f_{W|V,A(v_+,v_-)}(w|v_+, A\left(v_+,v_-\right))dw \\
&= \quad \int \lim_{v_+,v_-\to v_0} g_0(v_+,w) \lim_{v_+,v_-\to v_0} f_{W|V,A(v_+,v_-)}(w|v_+, A\left(v_+,v_-\right))dw \\
&= \quad \int g_0(v_0,w)\frac{\int_{b^-}^{b^+} f_{V|W,U}(v_0|w,u) f_{W,U}(w,u)du}{f_V(v_0)\int_{b^-}^{b^+} f_{U|V}\left(u|v_0\right)du}dw \\
&= \quad E(Y_0|V = v_0, A).
\end{aligned}
$$

Q.E.D.

Proof of Proposition 2.2. The proofs for g and h are similar so we provide a proof for g only and complete it in three steps:

Step 1. We prove continuity of $g(\cdot)$ at v_0;

Step 2. We prove continuity of $g(v)$ at any $v^* < v_0$;

Step 3. We prove continuity of $g(v)$ at any $v^* > v_0$.

Proof of Step 1. Note that

$$
\lim_{v\downarrow v_0} g(v) = \lim_{v\downarrow v_0} E(Y|V=v) - \delta = \lim_{v\uparrow v_0} E(Y|V=v) = \lim_{v\uparrow v_0} g(v).
$$

By definition, we have

$$
\begin{aligned}
& \quad g(v_0) \\
&= \quad E(Y|V=v_0) - \left[\lim_{v\downarrow v_0} E(Y|V=v) - \lim_{v\uparrow v_0} E(Y|V=v)\right] \\
&= \quad E(Y|V=v_0, b^- < U \le b^+)\Pr(b^- < U \le b^+|V=v_0) \\
+E(Y|V &= \quad v_0, U \le b^-)\Pr(U \le b^-|V=v_0) \\
+E(Y|V &= \quad v_0, U > b^+)\Pr(U > b^+|V=v_0) \\
+E(Y|V &= \quad v_0, b^+ < U \le b^-)\Pr(b^+ < U \le b^-|V=v_0) - (Y^+ - Y^-) \\
&= \quad \lim_{v_+\to v_0} E(Y|V=v_+) - \left[\lim_{v\downarrow v_0} E(Y|V=v) - \lim_{v\uparrow v_0} E(Y|V=v)\right] \\
&= \quad \lim_{v\uparrow v_0} E(Y|V=v).
\end{aligned}
$$

Then $g(v_0) = \lim_{v\downarrow v_0} g(v) = \lim_{v\uparrow v_0} g(v)$.

Proof of Step 2. For $v^* < v_0$, we know: $g(v^*) = E(Y|V = v^*)$. Then,

$$\begin{aligned}
\lim_{v\to v^*} E(Y|V &= v) \\
&= \lim_{v\to v^*}[E(Y|V = v, D(v) = 1)\Pr(D(v) = 1|V = v)] \\
+\lim_{v\to v^*}[E(Y|V &= v, D(v) = 0)\Pr(D(v) = 0|V = v)] \\
&= \lim_{v\to v^*}[E(Y_1|V = v, D(v) = 1)\Pr(D(v) = 1|V = v)] \\
+\lim_{v\to v^*}[E(Y|V &= v, D(v) = 0)\Pr(D(v) = 0|V = v)] \\
&= \lim_{v\to v^*}[E(Y_1|V = v, U \le b(v))\Pr(U \le b(v)|V = v)] \\
+\lim_{v\to v^*}[E(Y_0|V &= v, U > b(v))\Pr(U > b(v)|V = v)] \\
&= \lim_{v\to v^*}[E(Y_1|V = v, U \le b(v))]\Pr(U \le \lim_{v\to v^*} b(v)|V = v^*) \\
+\lim_{v\to v^*}[E(Y_0|V &= v, U > b(v))]\Pr(U > \lim_{v\to v^*} b(v)|V = v^*) \\
&= E(Y_1|V = v^*, U \le \lim_{v\to v^*} b(v))]\Pr(U \le \lim_{v\to v^*} b(v)|V = v^*) \\
+E(Y_0|V &= v^*, U > \lim_{v\to v^*} b(v))\Pr(U > \lim_{v\to v^*} b(v)|V = v^*) \\
&= E(Y|V = v^*),
\end{aligned}$$

where we have used:

$$\begin{aligned}
\lim_{v\to v^*}[E(Y_1|V &= v, U \le b(v))] = \lim_{v\to v^*}\int g_1(v,w) f_{W|V,U\le b(v)}(w|v, U \le b(v))dw \\
&= \int g_1(v^*,w)\frac{\int_{-\infty}^{\lim_{v\to v^*} b(v)} f_{V|W,U}(v^*|w,u) f_{W,U}(w,u)du}{f_V(v^*)\int_{-\infty}^{b(v^*)} f_{U|V}(u|v)du}dw, \\
&= E(Y_1|V = v^*, U \le \lim_{v\to v^*} b(v)].
\end{aligned}$$

A similar argument leads to: $\lim_{v\to v^*}[E(Y_0|V = v, U > b(v))] = E(Y_0|V = v^*, U > \lim_{v\to v^*} b(v))$.

Proof of Step 3. It is similar to that of Step 2 and thus omitted. **Q.E.D.**

Proof of Theorem 2.3. We complete the proof in two steps:

Step 1. We show:

$$\begin{aligned}
\lim_{e\downarrow 0} E(\Delta|V &= v_0, D(v_0 + e) - D(v_0 - e) = 1) \\
&= E_W\left[[g_1(v_0, W) - g_0(v_0, W)]\frac{f_{W|U,V}(W|b(v_0), v_0)}{f_W(W)}\right].
\end{aligned}$$

Step 2. We show:

$$\begin{aligned}
&\frac{\lim_{v\downarrow v_0} dE(Y|V = v)/dv - \lim_{v\uparrow v_0} dE(Y|V = v)/dv}{\lim_{v\downarrow v_0} P'(v) - \lim_{v\uparrow v_0} P'(v)} \\
= \; &E_W\left[[g_1(v_0, W) - g_0(v_0, W)]\frac{f_{W|U,V}(W|b(v_0), v_0)}{f_W(W)}\right].
\end{aligned}$$

Proof of Step 1. It follows from Condition K4(i):

$$\begin{aligned}
\lim_{e\downarrow 0} E(\Delta|V &= v_0, D(v_0 + e) - D(v_0 - e) = 1) \\
&= \lim_{e\downarrow 0} E(\Delta|V = v_0, D(v_0 + e) = 1, D(v_0 - e) = 0).
\end{aligned}$$

The right hand side expression is:

$$\begin{aligned}
&= \lim_{e\downarrow 0} E(g_1(v_0,W)-g_0(v_0,W)|V=v_0, b(v_0+e)\geq U > b(v_0-e)) \\
&= \lim_{e\downarrow 0} \int [g_1(v_0,w)-g_0(v_0,w)]\, f_{W|V,U}(w|v_0, b(v_0+e)\geq U > b(v_0-e))dw \\
&= \lim_{e\downarrow 0} \int [g_1(v_0,w)-g_0(v_0,w)] \left[\frac{\int_{b(v_0-e)}^{b(v_0+e)} f_{W,U|V}(w,u|v_0)du}{\int_{b(v_0-e)}^{b(v_0+e)} f_{U|V}(u|v_0)du}\right] dw \\
&= \int [g_1(v_0,w)-g_0(v_0,w)] \lim_{e\downarrow 0} \left[\frac{\int_{b(v_0-e)}^{b(v_0+e)} f_{W,U|V}(w,u|v_0)du}{\int_{b(v_0-e)}^{b(v_0+e)} f_{U|V}(u|v_0)du}\right] dw \\
&= \int [g_1(v_0,w)-g_0(v_0,w)] \frac{f_{W,U|V}(w,b(v_0)|v_0)}{f_{U|V}(b(v_0)|v_0)} dw \\
&= \int [g_1(v_0,w)-g_0(v_0,w)]\, f_{W|U,V}(w|b(v_0),v_0)dw \\
&= E_W\left[[g_1(v_0,W)-g_0(v_0,W)] \frac{f_{W|U,V}(W|b(v_0),v_0)}{f_W(W)}\right],
\end{aligned}$$

where we have used the following result:

$$\begin{aligned}
&\lim_{e\downarrow 0} \left[\frac{\int_{b(v_0-e)}^{b(v_0+e)} f_{W,U|V}(w,u|v_0)du}{\int_{b(v_0-e)}^{b(v_0+e)} f_{U|V}(u|v_0)du}\right] \\
&= \lim_{e\downarrow 0} \left[\frac{e^{-1}\{P(b(v_0+e))-P(b(v_0))-(P(b(v_0-e))-P(b(v_0)))\}}{e^{-1}\{Q(b(v_0+e))-Q(b(v_0))-(Q(b(v_0-e))-Q(b(v_0)))\}}\right] \\
&= \frac{f_{W,U|V}(w,b(v_0)|v_0)\,[b'^{+}(v)-b'^{-}(v)]}{f_{U|V}(b(v_0)|v_0)\,[b'^{+}(v)-b'^{-}(v)]} \\
&= \frac{f_{W,U|V}(w,b(v_0)|v_0)}{f_{U|V}(b(v_0)|v_0)},
\end{aligned}$$

in which $P(x)-P(a)=\int_a^x f_{W,U|V}(w,u|v_0)du$ and $Q(x)-Q(b)=\int_b^x f_{U|V}(u|v_0)du$.

Proof of Step 2. Consider $E(Y|V=v)$:

$$\begin{aligned}
E(Y|V &= v) \\
&= E(Y_1|V=v,D(v)=1)\Pr(D(v)=1|V=v)+E(Y_0|V=v,D(v)=0)\Pr(D(v)=0|V=v) \\
&= \left[\begin{array}{c} \int g_1(v,w) f_{W|V,U}(w|v,b(v)\geq u)dw \int_{-\infty}^{b(v)} f_{U|V}(u|v)du \\ +\int g_0(v,w) f_{W|V,U}(w|v,b(v)<u)dw \int_{b(v)}^{\infty} f_{U|V}(u|v)du \end{array}\right] \\
&= \int g_1(v,w)\frac{\int_{-\infty}^{b(v)} f_{W,U|V}(w,u|v)du}{\int_{-\infty}^{b(v)} f_{U|V}(u|v)du}dw \int_{-\infty}^{b(v)} f_{U|V}(u|v)du \\
&\quad + \int g_0(v,w)\frac{\int_{b(v)}^{\infty} f_{W,U|V}(w,u|v)du}{\int_{b(v)}^{\infty} f_{U|V}(u|v)du}dw \int_{b(v)}^{\infty} f_{U|V}(u|v)du \\
&= \int g_1(v,w)\int_{-\infty}^{b(v)} f_{W,U|V}(w,u|v)dudw + \int g_0(v,w)\int_{b(v)}^{\infty} f_{W,U|V}(w,u|v)dudw.
\end{aligned}$$

Taking derivatives on both sides of the last equality above, we get:

$$\begin{aligned}
& \frac{dE(Y|V=v)}{dv} \\
= & \int \frac{\partial}{\partial v}\left(g_1(v,w)\int_{-\infty}^{b(v)} f_{W,U|V}(w,u|v)du\right)dw + \int \frac{\partial}{\partial v}\left(g_0(v,w)\int_{b(v)}^{\infty} f_{W,U|V}(w,u|v)du\right)dw \\
= & \int g_1'(v,w)\int_{-\infty}^{b(v)} f_{W,U|V}(w,u|v)dudw + \int g_1(v,w)[b'(v)f_{W,U|V}(w,b(v)|v) \\
& + \int g_0'(v,w)\int_{b(v)}^{\infty} f_{W,U|V}(w,u|v)dudw + \int g_0(v,w)[-b'(v)f_{W,U|V}(w,b(v)|v) \\
& + \int_{-\infty}^{b(v)} \frac{\partial}{\partial v}(f_{W,U|V}(w,u|v))du]dw + \int_{b(v)}^{\infty} \frac{\partial}{\partial v}(f_{W,U|V}(w,u|v))du]dw.
\end{aligned}$$

Now taking limits leads to

$$\begin{aligned}
& \lim_{v\downarrow v_0} \frac{dE(Y|V=v)}{dv} \\
= & \int g_1'(v_0,w)\int_{-\infty}^{b(v_0)} f_{W,U|V}(w,u|v_0)dudw \\
& + \int g_1(v_0,w)[b'^{+} f_{W,U|V}(w,b(v_0)|v_0) + \int_{-\infty}^{b(v_0)} \frac{\partial}{\partial v}(f_{W,U|V}(w,u|v_0))du]dw \\
& + \int g_0'(v_0,w)\int_{b(v_0)}^{\infty} f_{W,U|V}(w,u|v_0)dudw \\
& + \int g_0(v_0,w)[-b'^{+} f_{W,U|V}(w,b(v_0)|v_0) + \int_{b(v_0)}^{\infty} \frac{\partial}{\partial v}(f_{W,U|V}(w,u|v_0))du]dw
\end{aligned}$$

and

$$\begin{aligned}
& \lim_{v\uparrow v_0} \frac{dE(Y|V=v)}{dv} \\
= & \int g_1'(v_0,w)\int_{-\infty}^{b(v_0)} f_{W,U|V}(w,u|v_0)dudw \\
& + \int g_1(v_0,w)[b'^{-} f_{W,U|V}(w,b(v_0)|v_0) + \int_{-\infty}^{b(v_0)} \frac{\partial}{\partial v}(f_{W,U|V}(w,u|v_0))du]dw \\
& + \int g_0'(v_0,w)\int_{b(v_0)}^{\infty} f_{W,U|V}(w,u|v_0)dudw \\
& + \int g_0(v_0,w)[-b'^{-} f_{W,U|V}(w,b(v_0)|v_0) + \int_{b(v_0)}^{\infty} \frac{\partial}{\partial v}(f_{W,U|V}(w,u|v_0))du]dw.
\end{aligned}$$

As a result, we have:

$$\begin{aligned}
& \lim_{v\downarrow v_0} \frac{dE(Y|V=v)}{dv} - \lim_{v\uparrow v_0} \frac{dE(Y|V=v)}{dv} \\
= & \int \left[g_1(v_0,w) - g_0(v_0,w)\right] f_{W,U|V}(w,b(v_0)|v_0)dw \left[b'^{+} - b'^{-}\right]
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\lim_{v\downarrow v_0} dE(Y|V=v)/dv - \lim_{v\uparrow v_0} dE(Y|V=v)/dv}{\lim_{v\downarrow v_0} P'(v) - \lim_{v\uparrow v_0} P'(v)} \\
= \;& \frac{\int [g_1(v_0,w)-g_0(v_0,w)]\, f_{W,U|V}(w,b(v_0)|v_0)dw\, [b'^{+}(v)-b'^{-}(v)]}{f_{U|V}(b(v_0)|v_0)\, [b'^{+}(v)-b'^{-}(v)]} \\
= \;& \int [g_1(v_0,w)-g_0(v_0,w)]\, \frac{f_{W,U|V}(w,b(v_0)|v_0)}{f_{U|V}(b(v_0)|v_0)} dw.
\end{aligned}$$

Q.E.D.

Proof of Proposition 2.4. We provide a proof for g_K only. This will be done in two steps:

Step 1. We show g_K is continuous;

Step 2. We show g_K is continuously differentiable.

Proof of Step 1. By definition, $g_K(V) = E(Y|V) - \delta_K(V-v_0)I\{V \geq v_0\}$. We only need to show that it is continuous at v_0. Under Condition K4(A), we know: $\lim_{v\downarrow v_0} E(Y|V=v) - \lim_{v\downarrow v_0} E(Y|V=v) = 0 = \delta$, thus $E(Y|V) = g(v)$ from Proposition 2.2, which is continuous on the support of V. Since $\delta_K(V-v_0)I\{V \geq v_0\}$ is continuous on the support of V, $g_K(V)$ is continuous.

Proof of Step 2. When $v = v_0$,

$$\lim_{v\downarrow v_0} \frac{dg_K(v)}{dv} = \lim_{v\downarrow v_0} \frac{dE(Y|V=v)}{dv} - \delta_K = \lim_{v\uparrow v_0} \frac{dE(Y|V=v)}{dv}$$

and $\lim_{v\uparrow v_0} \frac{dg_K(v)}{dv} = \lim_{v\uparrow v_0} \frac{dE(Y|V=v)}{dv}$. Thus

$$\lim_{v\downarrow v_0} \frac{dg_K(v)}{dv} = \lim_{v\uparrow v_0} \frac{dg_K(v)}{dv} = \lim_{v\uparrow v_0} \frac{dE(Y|V=v)}{dv} = \lim_{v\to v_0} \frac{dg(v)}{dv}.$$

Now let us consider $v^* < v_0$. Then following the proof of Theorem 2.3 except that we are looking at v^* instead of v_0, we obtain: $\lim_{v\downarrow v^*} dE(Y|V=v)/dv - \lim_{v\uparrow v^*} dE(Y|V=v)/dv = 0$. A similar proof applies to $v^* > v_0$. **Q.E.D.**

Appendix B: Technical Proofs for Results in Section 3

We will make extensive use of the following Taylor expansions in the proofs in Appendices B and C. Under A2(G)(a), we have:

$$G(\tau \pm h) = G(\tau) + \sum_{k=1}^{l_G - 1} \frac{G_{\pm}^{(k)}(\tau)}{k!} (\pm h)^k + R_G^{\pm},$$

where $G_{\pm}^{(k)}(\tau)$ denote the right and left k-th order derivatives of $G(t)$ at τ,

$$|R_G^{\pm}| \leq K|h|^{l_G} \sup_{t \in (0,1)} |G_{\pm}^{(l_G)}(t)| < \infty$$

with K a large positive number. Under A2(G)(b), we have:

$$G(\tau + h) = G(\tau) + \sum_{k=1}^{l_G - 1} \frac{G^{(k)}(\tau)}{k!} (h)^k + R_G, \tag{B.1}$$

where for a large positive number K, $|R_G| \leq K|h|^{l_G} \sup_{t \in (0,1)} |G^{(l_G)}(t)| < \infty$.

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

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

where J is some bounded smooth function and may depend on n. Further, let

$$\begin{aligned} m(x) &= E(Y|X = x),\ \sigma^2(x) = Var(Y|X = x), \\ F^{-1}(u) &= \inf\{x|F(x) \geq u\},\ m \circ F^{-1}(u) = m(F^{-1}(u)). \end{aligned}$$

Lemma B.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*

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{+\infty} J^2(F(x))\sigma^2(x) dF(x) \\ &\quad + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [F(x \wedge y) - F(x)F(y)] J(F(x)) J(F(y)) dm(x) dm(y). \end{aligned}$$

Then $\lim_{n\to\infty} nVar(S_n) = \sigma^2$ *and* $\lim_{n\to\infty} E(S_n) = \int_{-\infty}^{+\infty} m(x)J(F(x))dF(x)$*. Furthermore, if* $\sigma^2 > 0$*, then*

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

We note that all the proposed wavelet estimators including the wavelet OLS estimators converge at rates slower than $n^{-1/2}$. Since $\widehat{\tau}$ converges at rate $n^{-1/2}$, under regularity conditions, the asymptotic distributions of all our estimators are not affected by estimating τ by $\widehat{\tau}$. This is universal to all nonparametric estimators constructed after a first step estimation of the location of a jump or kink, see e.g., the work in statistics cited in Section 1 including Wang and Cai (2010). Because of this, we will work with the infeasible versions of our estimators with $\widehat{\tau}$ replaced by τ in Appendices B and C. With slight abuse of notation, we will use the same notations to denote the corresponding infeasible estimators. For notational compactness, we let $\psi_j\left[\cdot\right] = 2^j\psi\left[2^j\cdot\right]$ in both Appendix B and Appendix C.

Proof of Theorem 3.1. We will complete the proof in two steps:

Step 1. We show that $\overline{\delta}$ has the asymptotic distributions stated in the theorem;

Step 2. We show:$\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}-\overline{\delta}) = o_p\left(1\right)$.

Proof of Step 1: Note that we can write $\overline{\delta}$ as: $\overline{\delta} = \frac{1}{n}\sum_{i=1}^{n} J(\frac{i}{n})Y_{[i:n]}$, where

$$J(\frac{i}{n}) = \frac{\psi_{j_0}\left[\frac{i}{n}-\tau\right]}{\int_0^b \psi(u)du}.$$

We will use Lemma B.1 to show that $\sqrt{\frac{n}{2^{j_0}}}(\overline{\delta}-\delta)$ has the limiting distribution stated in Theorem 3.1. The conditions in Lemma B.1 are satisfied: $E(Y|V=v) = g(v)+\delta I\{v \geq v_0\}$ is right continuous by Proposition 2.2 and of bounded variation in any finite interval; $J(\frac{i}{n})$ is bounded and continuous by Assumption A4.

First, let us calculate $\lim_{n\to\infty} E(\overline{\delta})$:

$$\begin{aligned}
&\int_{-\infty}^{+\infty}\left[g(v)+\delta I\{v \geq v_0\}\right]\frac{\psi_{j_0}(F_V(v)-\tau)}{\int_0^b \psi(u)du}dF_V(v)\\
= \quad & \frac{\int_{-\infty}^{+\infty} g(v)\psi_{j_0}(F_V(v)-\tau)dF_V(v)}{\int_0^b \psi(u)du} + \frac{\int_{-\infty}^{+\infty} \delta I\{v \geq v_0\}\psi_{j_0}(F_V(v)-\tau)dF_V(v)}{\int_0^b \psi(u)du}\\
= \quad & \frac{\int_a^b g\left[F_V^{-1}(\frac{u}{2^{j_0}}+\tau)\right]\psi(u)du}{\int_0^b \psi(u)du} + \delta\\
= \quad & \left\{\begin{array}{l}
\dfrac{\frac{1}{2^{j_0}}\left[G_+^{(1)}(\tau)-G_-^{(1)}(\tau)\right]\int_0^b \psi(u)udu}{\int_0^b \psi(u)du} + \delta + s.o., \text{ under A2(G)(a)}\\
\dfrac{(\frac{1}{2^{j_0}})^m G^{(m)}(\tau)\int_a^b u^m\psi(u)du}{\int_0^b \psi(u)du} + \delta + s.o, \text{ under A2(G)(b)}
\end{array}\right. .
\end{aligned}$$

Then,

$$\begin{aligned}
\lim_{n\to\infty} 2^{j_0}\left[E(\overline{\delta}_1)-\delta\right] &= \frac{\left[G_+^{(1)}(\tau)-G_-^{(1)}(\tau)\right]\int_0^b \psi(u)udu}{\int_0^b \psi(u)du}, \text{ under A2(G)(a)},\\
\lim_{n\to\infty} 2^{mj_0}\left[E(\overline{\delta}_1)-\delta\right] &= \frac{G^{(m)}(\tau)\int_a^b u^m\psi(u)du}{\int_0^b \psi(u)du}, \text{ under A2(G)(b)}.
\end{aligned}$$

Second, let us calculate $\lim_{n\to\infty} nVar(\overline{\delta})$:

$$\begin{aligned}
&\int_{-\infty}^{+\infty}\left[\frac{\psi_{j_0}(F_V(v)-\tau)}{\int_0^b \psi(u)du}\right]^2 \sigma^2(v)dF_V(v)\\
&+\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[F_V(v_1\wedge v_2)-F_V(v_1)F_V(v_2)\right]\frac{\psi_{j_0}(F_V(v_1)-\tau)\psi_{j_0}(F_V(v_2)-\tau)}{\left(\int_0^b \psi(u)du\right)^2}dm(v_1)dm(v_2)\\
= \quad & A_1 + A_2,
\end{aligned}$$

where

$$\begin{aligned}
m(v) &= g(v) + \delta I\{v \geq v_0\}, \\
A_1 &= \int_{-\infty}^{+\infty} \left[\frac{\psi_{j_0}(F_V(v) - \tau)}{\int_0^b \psi(u) du} \right]^2 \sigma^2(v) dF_V(v), \\
A_2 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [F_V(v_1 \wedge v_2) - F_V(v_1)F_V(v_2)] \frac{\psi_{j_0}(F_V(v_1) - \tau)\psi_{j_0}(F_V(v_2) - \tau)}{\left(\int_0^b \psi(u) du\right)^2} dm(v_1) dm(v_2).
\end{aligned}$$

Then

$$\begin{aligned}
A_1 = \int_{-\infty}^{+\infty} \left[\frac{2^{j_0}\psi(u)}{\int_0^b \psi(u) du} \right]^2 \sigma^2(F_V^{-1}(\frac{u}{2^{j_0}} + \tau)) \frac{1}{2^{j_0}} du \\
&= \frac{2^{j_0} \int_{-\infty}^{+\infty} \psi^2(u) \sigma^2(F_V^{-1}(\frac{u}{2^{j_0}} + \tau)) du}{\left(\int_0^b \psi(u) du\right)^2} \\
&= \frac{2^{j_0} \left[\int_0^b \psi^2(u) \sigma^2(F_V^{-1}(\frac{u}{2^{j_0}} + \tau)) du + \int_a^0 \psi^2(u) \sigma^2(F_V^{-1}(\frac{u}{2^{j_0}} + \tau)) du \right]}{\left(\int_0^b \psi(u) du\right)^2}.
\end{aligned}$$

So $A_1 = O(2^{j_0})$ and $A_2 = O(1)$. Thus we get:

$$\lim_{n \to \infty} \frac{n}{2^{j_0}} Var(\overline{\delta}) = \frac{\left[\sigma_+^2(v_0) \int_0^b \psi^2(u) du + \sigma_-^2(v_0) \int_a^0 \psi^2(u) du \right]}{\left(\int_0^b \psi(u) du\right)^2}.$$

By Lemma B.1, we obtain:

$$\sqrt{\frac{n}{2^{j_0}}} (\overline{\delta}_1 - \delta) \xrightarrow{d} \begin{cases} N(C_a B_a, V), \text{ under A2(G)(a)} \\ N(C_b B_b, V), \text{ under A2(G)(b)} \end{cases}.$$

Proof of Step 2: Note that

$$\begin{aligned}
& \sqrt{\frac{n}{2^{j_0}}} (\widehat{\delta} - \overline{\delta}) \\
= & \sqrt{\frac{n}{2^{j_0}}} \overline{\delta} \left(\frac{\int_0^b \psi(u) du - \frac{1}{n} \sum_{j=1}^n I\{t_j \geq \tau\} \psi_{j_0}[t_j - \tau]}{\frac{1}{n} \sum_{j=1}^n I\{t_j \geq \tau\} \psi_{j_0}[t_j - \tau]} \right),
\end{aligned}$$

where the first term satisfies: $\sqrt{\frac{n}{2^{j_0}}} \overline{\delta} = O_p(\sqrt{\frac{n}{2^{j_0}}})$. For large enough j_0, the numerator of the second term satisfies:

$$\begin{aligned}
& |\frac{1}{n} \sum_{j=1}^n I\{t_j \geq \tau\} \psi_{j_0}[t_j - \tau] - \int_0^b \psi(u) du| \\
= & |\frac{1}{n} \sum_{j=1}^n I\{t_j \geq \tau\} \psi_{j_0}[t_j - \tau] - \int_0^1 I\{t \geq \tau\} \psi_{j_0}[t_j - \tau] dt| \\
\leq & \frac{1}{n+1} V_0^1(f),
\end{aligned}$$

where the last inequality above is obtained from the Koksma–Hlawka inequality in which

$$f(t) = I\{t \geq \tau\} \psi_{j_0}[t_j - \tau]$$

and $V_0^1(f)$ is the bounded variation of f on $[0,1]$.

Note that for large enough j_0,

$$\begin{aligned} V_0^1(f) &= V_\tau^1(\psi_{j_0}\left[\cdot-\tau\right]) = 2^{j_0}\int_\tau^1 |2^{j_0}\psi^{(1)}\left[2^{j_0}(t-\tau)\right]|dt \\ &= 2^{j_0}\int_0^b |\psi^{(1)}\left[t\right]|dt = O\left(2^{j_0}\right) \end{aligned}$$

since $\int_0^b |\psi^{(1)}\left[t\right]|dt < \infty$. Hence,

$$\begin{aligned} \sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}-\overline{\delta}) &= \sqrt{\frac{n}{2^{j_0}}}\overline{\delta}\left(\frac{\int_0^b \psi(u)du}{\frac{1}{n}\sum_{j=1}^n I\left\{t_j \geq \tau\right\}\psi_{j_0}\left[t_j-\tau\right]}-1\right) \\ &= O_p(\sqrt{\frac{n}{2^{j_0}}})O_p(\frac{2^{j_0}}{n}) = o_p(1). \end{aligned}$$

Q.E.D.

Proof of Theorem 3.2. First, we derive the asymptotic covariance between $\widehat{\delta}$ and $\widehat{\zeta}$:

$$\begin{aligned} &\lim_{n\to\infty} Cov\left[(n/2^{j_0})^{1/2}(\widehat{\delta}-\delta),(n/2^{j_0})^{1/2}(\widehat{\zeta}-\zeta)\right] = \lim_{n\to\infty}\frac{n}{2^{j_0}}Cov\left[\widehat{\delta},\widehat{\zeta}\right] \\ =\ & \lim_{n\to\infty}\frac{n}{2^{j_0}}Cov\left[\frac{1}{n}\sum_{i=1}^n J(\frac{i}{n})Y_{[i:n]},\frac{1}{n}\sum_{i=1}^n J(\frac{i}{n})D_{[i:n]}\right] \\ =\ & \frac{1}{2^{j_0}}\int_{-\infty}^{+\infty} J^2(F_V(x))\sigma^2_{\varepsilon\epsilon}(x)dF_V(x) \\ & +\frac{1}{2^{j_0}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[F_V(x\wedge y)-F_V(x)F_V(y)\right]J(F_V(x))J(F_V(y))dm(x)dm^D(y) \\ =\ & \frac{\left[\sigma^2_{\varepsilon\epsilon+}(v_0)\int_0^b\psi^2(u)du+\sigma^2_{\varepsilon\epsilon-}(v_0)\int_a^0\psi^2(u)du\right]}{\left(\int_0^b\psi(u)du\right)^2}, \end{aligned}$$

where $m^D(x) = E(D|V=x)$ and the second last equality follows from the proof of equation (8) in Yang (1981).

Next, we apply the Cramer-Wold Device to establish the joint limiting distribution of $(n/2^{j_0})^{1/2}(\widehat{\delta}-\delta)$ and $(n/2^{j_0})^{1/2}(\widehat{\zeta}-\zeta)$. In the end, we use Delta method to establish the asymptotic distribution for $\widehat{\delta}/\widehat{\zeta}$. **Q.E.D.**

Proof of Theorem 3.3. First we work with the bias term for $\overline{\delta}_K$:

$$\begin{aligned} &\int_{-\infty}^{+\infty}\left[g_K(v)+\delta_K(V-v_0)I\{v\geq v_0\}\right]\frac{\psi_{j_0}(F_V(v)-\tau)}{\int_0^b\left[F_V^{-1}(\frac{u}{2^{j_0}}+\tau)-v_0\right]\psi(u)du}dF_V(v) \\ =\ & \frac{\int_{-\infty}^{+\infty}g_K(v)\psi_{j_0}(F_V(v)-\tau)dF_V(v)}{\int_0^b\left[F_V^{-1}(\frac{u}{2^{j_0}}+\tau)-v_0\right]\psi(u)du}+\frac{\int_{-\infty}^{+\infty}\delta_K vI\{v\geq v_0\}\psi_{j_0}(F_V(v)-\tau)dF_V(v)}{\int_0^b\left[F_V^{-1}(\frac{u}{2^{j_0}}+\tau)-v_0\right]\psi(u)du} \\ =\ & \frac{\int_a^b g_K\left[F_V^{-1}(\frac{u}{2^{j_0}}+\tau)\right]\psi(u)du}{\int_0^b\left[F_V^{-1}(\frac{u}{2^{j_0}}+\tau)-v_0\right]\psi(u)du}+\delta_K. \end{aligned}$$

Under A2K(G)(a), we have

$$\int_a^b g_K\left[F_V^{-1}(\frac{u}{2^{j_0}}+\tau)\right]\psi(u)du = \frac{\left[G^{(2)}_{K+}(\tau)-G^{(2)}_{K-}(\tau)\right]\int_0^b u^2\psi(u)du}{2^{2j_0+1}}+s.o.$$

Under A2K(G)(b):

$$\int_a^b g_K \left[F_V^{-1}(\frac{u}{2^{j_0}}+\tau)\right]\psi(u)du = \frac{G_K^{(m+1)}(\tau)\int_a^b u^{m+1}\psi(u)du}{(m+1)!2^{(m+1)j_0}} + s.o.$$

and

$$\int_0^b \left[F_V^{-1}(\frac{u}{2^{j_0}}+\tau)-v_0\right]\psi(u)du = \frac{\int_0^b u\psi(u)du}{2^{j_0}f_V(v_0)} + s.o.$$

Then,

$$\begin{aligned}
\lim_{n\to\infty} 2^{j_0}\left[E(\overline{\delta}_K)-\delta_K\right] &= \frac{\left[G_{K+}^{(2)}(\tau)-G_{K-}^{(2)}(\tau)\right] f_V(v_0)\int_0^b u^2\psi(u)du}{2\int_0^b u\psi(u)du}, \text{ under A2K(G)(a)},\\
\lim_{n\to\infty} 2^{mj_0}\left[E(\overline{\delta}_K)-\delta_K\right] &= \frac{G_K^{(m+1)}(\tau)}{(m+1)!}\frac{f_V(v_0)\int_a^b u^{m+1}\psi(u)du}{\int_0^b u\psi(u)du}, \text{ under A2K(G)(b)}.
\end{aligned}$$

Now we work on the variance term:

$$\begin{aligned}
&\int_{-\infty}^{+\infty}\left[\frac{2^{j_0}\psi(u)}{\int_0^b\left[F_V^{-1}(\frac{u}{2^{j_0}}+\tau)-v_0\right]\psi(u)du}\right]^2\sigma^2(F_V^{-1}(\frac{u}{2^{j_0}}+\tau))\frac{1}{2^{j_0}}du\\
=\;& \frac{2^{j_0}\int_{-\infty}^{+\infty}\psi^2(u)\sigma^2(F_V^{-1}(\frac{u}{2^{j_0}}+\tau))du}{\left(\frac{1}{2^{j_0}}\frac{1}{f_V(v_0)}\int_0^b u\psi(u)udu\right)^2}\\
=\;& \frac{2^{j_0}}{\left(\frac{1}{2^{j_0}}\frac{1}{f_V(v_0)}\int_0^b u\psi(u)udu\right)^2}\left[\int_0^b \psi^2(u)\sigma^2(F_V^{-1}(\frac{u}{2^{j_0}}+\tau))du+\int_a^0\psi^2(u)\sigma^2(F_V^{-1}(\frac{u}{2^{j_0}}+\tau))du\right].
\end{aligned}$$

Thus we get

$$\lim_{n\to\infty}\frac{n}{2^{3j_0}}Var(\overline{\delta_K}) = \frac{f_V^2(v_0)\left[\sigma_{\varepsilon+}^2(v_0)\int_0^b\psi^2(u)du+\sigma_{\varepsilon-}^2(v_0)\int_a^0\psi^2(u)du\right]}{\left(\int_0^b u\psi(u)du\right)^2}.$$

Finally the asymptotic normality of $\widehat{\delta}_K$ is established by following a similar proof to that of Theorem 3.1. **Q.E.D.**

Appendix C: Technical Proofs for Results in Section 4

We recall $\psi_j\left[\cdot\right] = 2^j \psi\left[2^j \cdot\right]$ and the following expressions for functions $L\left(t\right)$ and $M\left(v\right)$:

$$L(t) = \int_a^b I\{w \geq t\}\psi(w)dw \text{ and } M(v) = \int_a^b \int_a^b I\{w \geq t+v\}\psi(w)\psi(t)dtdw.$$

Lemma C.1 *Under Assumption (A4), (i) $L(t)$ has $(m-1)$ vanishing moments and compact support $[a,b]$; (ii) $M(t)$ has compact support $[a-b, b-a]$.*

Proof of Lemma C.1. By induction. **Q.E.D.**

Proof of Theorem 4.1. Let

$$\overline{\delta}_{W1} = n^{-1}\sum_{i=1}^{n} J_{W1}\left(\frac{i}{n}\right)Y_{[i:n]},$$

where

$$J_{W1}(\frac{i}{n}) = \frac{\int_0^1\int_0^1 D_{j_0}(t)I\{w\geq \tau\}\psi_{j_0}\left[w-t\right]\psi_{j_0}\left[\frac{i}{n}-t\right]dtdw}{\int_0^1\int_0^1\int_0^1 D_{j_0}(t)I\{w\geq\tau\}I\{v\geq\tau\}\psi_{j_0}\left[w-t\right]\psi_{j_0}\left[v-t\right]dwdvdt}.$$

We will complete the proof in two steps:

Step 1. We show $\overline{\delta}_{W1}$ has the asymptotic distributions stated in the theorem;

Step 2. We show: $\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_{W1} - \overline{\delta}_{W1}) = o_p\left(1\right)$.

Proof of Step 1: First, let us calculate $\lim_{n\to\infty} E(\overline{\delta}_{W1})$:

$$\int_{-\infty}^{\infty}\left[g(v) + \delta I\{v\geq v_0\}\right]J_{W1}(F_V(v))dF_V(v) = \int_{-\infty}^{\infty} g(v)J_{W1}(F_V(v))dF_V(v) + \delta.$$

For the first term on the right hand side of the above equation,

$$\begin{aligned}
&\int_{-\infty}^{\infty} g(v)J_{W1}(F_V(v))dF_V(v)\\
=\;& \frac{2^{-j_0}\int_{-\infty}^{\infty} g(v)\left[\int_0^1\int_0^1 D_{j_0}(t)I\{w\geq\tau\}\psi_{j_0}\left[w-t\right]\psi_{j_0}\left[F_V(v)-t\right]dtdw\right]dF_V(v)}{2^{-j_0}\int_0^1\int_0^1\int_0^1 D_{j_0}(t)I\{w\geq\tau\}I\{v\geq\tau\}\psi_{j_0}\left[w-t\right]\psi_{j_0}\left[v-t\right]dwdvdt}\\
=\;& \frac{T_{W1}}{T_{W1D}},
\end{aligned}$$

where

$$\begin{aligned}
T_{W1} &= 2^{-j_0}\int_{-\infty}^{\infty} g(v)\left[\int_0^1\int_0^1 D_{j_0}(t)I\{w\geq\tau\}\psi_{j_0}\left[w-t\right]\psi_{j_0}\left[F_V(v)-t\right]dtdw\right]dF_V(v),\\
T_{W1D} &= 2^{-j_0}\int_0^1\int_0^1\int_0^1 D_{j_0}(t)I\{w\geq\tau\}I\{v\geq\tau\}\psi_{j_0}\left[w-t\right]\psi_{j_0}\left[v-t\right]dwdvdt.
\end{aligned}$$

For large enough j_0, we obtain:

$$\begin{aligned}
T_{W1} &= \int_0^1\int_0^1\left[\int_a^b g(F_V^{-1}(\frac{s}{2^{j_0}}+t))\psi(s)ds\right]D_{j_0}(t)I\{w\geq\tau\}\psi\left[2^{j_0}(w-t)\right]dtdw\\
&= \int_0^1\int_0^1 W(t)D_{j_0}(t)I\{w\geq\tau\}\psi\left[2^{j_0}(w-t)\right]dtdw\\
&= \frac{1}{2^{j_0}}\int_0^1\left[\int_a^b I\{w\geq 2^{j_0}(\tau-t)\}\psi\left[w\right]dw\right]D_{j_0}(t)W(t)dt\\
&= \frac{1}{2^{j_0}}\int_0^1 L(2^{j_0}(\tau-t))D_{j_0}(t)W(t)dt\\
&= (\frac{1}{2^{j_0}})^2\int_a^b L(t)W(\tau-\frac{t}{2^{j_0}})dt,
\end{aligned}$$

where

$$W(t) \equiv \int_a^b G\left(\frac{s}{2^{j_0}} + t\right)\psi(s)ds.$$

Ignoring higher order terms, we obtain: under A2(G)(a):

$$\begin{aligned}
W(\tau - \frac{t}{2^{j_0}}) &= \int_a^b G\left(\tau + \frac{s-t}{2^{j_0}}\right)\psi(s)ds \\
&= \int_a^b G\left(\tau + \frac{s-t}{2^{j_0}}\right)\psi(s)I\{\frac{s-t}{2^{j_0}} \geq 0\}ds + \int_a^b G\left(\tau + \frac{s-t}{2^{j_0}}\right)\psi(s)I\{\frac{s-t}{2^{j_0}} < 0\}ds \\
&= \sum_{k=1}^{l_G-1} \frac{G_+^{(k)}(\tau)}{k!}\int_a^b (\frac{s-t}{2^{j_0}})^k\psi(s)I\{\frac{s-t}{2^{j_0}} \geq 0\}ds + \sum_{k=1}^{l_G-1} \frac{G_-^{(k)}(\tau)}{k!}\int_a^b (\frac{s-t}{2^{j_0}})^k\psi(s)I\{\frac{s-t}{2^{j_0}} < 0\}ds.
\end{aligned}$$

Then,

$$T_{W1} = (\frac{1}{2^{j_0}})^3 \left[G_+^{(1)}(\tau) - G_-^{(1)}(\tau)\right]\int_a^b\int_a^b L(t)\psi(s)(s-t)I\{s-t \geq 0\}dsdt.$$

Under A2(G)(b):

$$W(\tau - \frac{t}{2^{j_0}}) = \sum_{k=1}^{l_G-1} \frac{G^{(k)}(\tau)}{k!}\int_a^b (\frac{s-t}{2^{j_0}})^k\psi(s)ds + s.o.$$

Then,

$$\begin{aligned}
T_{W1} &= (\frac{1}{2^{j_0}})^2 \sum_{k=1}^{l_G-1} \frac{G^{(k)}(\tau)}{k!}\int_a^b\int_a^b L(t)\psi(s)(\frac{s-t}{2^{j_0}})^k dsdt + s.o. \\
&= \begin{cases} (\frac{1}{2^{j_0}})^{2m+1}\frac{G^{(2m-1)}(\tau)}{m!(m-1)!}\int_a^b \psi(s)s^m ds \int_a^b L(t)(-t)^{m-1}dt + s.o. & \text{if } \ l_g \geq 2m; \\ O((\frac{1}{2^{j_0}})^{l_G+2}) & \text{if } \ l_g < 2m. \end{cases}
\end{aligned}$$

For large enough j_0,

$$\begin{aligned}
T_{W1D} &= \int_0^1\int_0^1 \left[\int_a^b I\{w \geq 2^{j_0}(\tau - t)\}\psi\left[w\right]dw\right] D_{j_0}(t)I\{v \geq \tau\}\psi\left[2^{j_0}(v-t)\right]dtdv \\
&= \int_0^1\int_0^1 L(2^{j_0}(\tau - t))I\{a \leq 2^{j_0}(\tau - t) \leq b\}I\{v \geq \tau\}\psi\left[2^{j_0}(v-t)\right]dtdv \\
&= \frac{1}{2^{j_0}}\int_0^1 \left[\int_a^b L(2^{j_0}(\tau - v) + t)\psi(t)dt\right] I\{v \geq \tau\}dv \\
&= \frac{1}{2^{j_0}}\int_0^1 M(2^{j_0}(\tau - v))I\{v \geq \tau\}dv \\
&= \frac{1}{2^{2j_0}}\int_{a-b}^0 M(v)dv.
\end{aligned}$$

Therefore, under A2(G)(a):

$$\frac{T_{W1}}{T_{W1D}} = \frac{(\frac{1}{2^{j_0}})\left[G_+^{(1)}(\tau) - G_-^{(1)}(\tau)\right]\int_a^b\int_a^b L(t)\psi(s)(s-t)I\{s-t \geq 0\}dsdt}{\int_{a-b}^0 M(v)dv} + s.o.;$$

under A2(G)(b):

$$\frac{T_{W1}}{T_{W1D}}=\begin{cases}\frac{(\frac{1}{2^{j_0}})^{2m-1}\cdot G^{(2m-1)}(\tau)\int_a^b\psi(s)s^m ds\cdot\int_a^b L(t)(-t)^{m-1}dt}{m!(m-1)!\int_{a-b}^0 M(v)dv}+s.o. & \text{if } l_G\geq 2m\\ O((\frac{1}{2^{j_0}})^{l_G}) & \text{if } l_G<2m\end{cases}.$$

Thus, under A2(G)(a), we obtain:

$$\lim_{n\to\infty}2^{j_0}\left[E(\overline{\delta}_{W1})-\delta\right]=\frac{\left[G_+^{(1)}(\tau)-G_-^{(1)}(\tau)\right]\int_a^b\int_a^b L(t)\psi(s)(s-t)I\{s-t\geq 0\}dsdt}{\int_{a-b}^0 M(v)dv};$$

under A2(G)(b), we obtain:

$$\lim\nolimits_{n\to\infty}(2^{j_0})^{2m-1}\left[E(\overline{\delta}_{W1})-\delta\right]=\frac{G^{(2m-1)}(\tau)\int_a^b\psi(s)s^m ds\cdot\int_a^b L(t)(-t)^{m-1}dt}{m!(m-1)!\int_{a-b}^0 M(v)dv},\text{ if } l_G\geq 2m;$$
$$\lim\nolimits_{n\to\infty}(2^{j_0})^{l_G}\left[E(\overline{\delta}_{W1})-\delta\right]=O(1),\text{ if } l_G<2m.$$

Second, let us calculate the asymptotic variance of $\overline{\delta}_{W1}$:

$$\begin{aligned}&\frac{\int_{-\infty}^{\infty}\left[\int_0^1\int_0^1 D_{j_0}(t)I\{w\geq\tau\}\psi_{j_0}[w-t]\,\psi\left[2^{j_0}(F_V(v)-t)\right]dtdw\right]^2\sigma^2(v)dF_V(v)}{T_D^2}\\&+\frac{\int\int\left\{\begin{array}{c}[F_V(v_1\wedge v_2)-F_V(v_1)F_V(v_2)]\cdot\\ \cdot\left[\int_0^1\int_0^1 D_{j_0}(t)I\{w\geq\tau\}\psi_{j_0}[w-t]\,\psi\left[2^{j_0}(F_V(v_1)-t)\right]dtdw\right]\\ \cdot\left[\int_0^1\int_0^1 D_{j_0}(t)I\{w\geq\tau\}\psi_{j_0}[w-t]\,\psi\left[2^{j_0}(F_V(v_2)-t)\right]dtdw\right]\end{array}\right\}dm(v_1)dm(v_2)}{T_D^2}\\=&\;\frac{T_{W12}}{T_{W1D}^2}+\frac{T_{W13}}{T_{W1D}^2},\end{aligned}$$

where

$$\begin{aligned}T_{W12}&=\int_{-\infty}^{\infty}\left[\int_0^1\int_0^1 D_{j_0}(t)I\{w\geq\tau\}\psi_{j_0}[w-t]\,\psi\left[2^{j_0}(F_V(v)-t)\right]dtdw\right]^2\sigma^2(v)dF_V(v),\\ T_{W13}&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left\{\begin{array}{c}[F_V(v_1\wedge v_2)-F_V(v_1)F_V(v_2)]\\ \left[\int_0^1\int_0^1 D_{j_0}(t)I\{w\geq\tau\}\psi_{j_0}[w-t]\,\psi\left[2^{j_0}(F_V(v_1)-t)\right]dtdw\right]\\ \left[\int_0^1\int_0^1 D_{j_0}(t)I\{w\geq\tau\}\psi_{j_0}[w-t]\,\psi\left[2^{j_0}(F_V(v_2)-t)\right]dtdw\right]\end{array}\right\}dm(v_1)dm(v_2).\end{aligned}$$

For the T_{W12} term

$$T_{W12}=\int_0^1 P_{W1}^2(u)\,\sigma^2(F_V^{-1}(u))du,$$

where

$$P_{W1}(u)=\int_0^1\int_0^1 D_{j_0}(t)I\{w\geq\tau\}\psi_{j_0}[w-t]\,\psi\left[2^{j_0}(u-t)\right]dtdw.$$

Notice that for large enough j_0,

$$\begin{aligned}P_{W1}(u)&=\int_0^1\left[\int_a^b I\{w\geq 2^{j_0}(\tau-t)\}\psi(w)dw\right]D_{j_0}(t)\psi\left[2^{j_0}(u-t)\right]dt\\&=\int_0^1 L(2^{j_0}(\tau-t))D_{j_0}(t)\psi\left[2^{j_0}(u-t)\right]dt\\&=\frac{1}{2^{j_0}}M(2^{j_0}(\tau-u)).\end{aligned}$$

Therefore,

$$T_{W12} = \frac{1}{2^{3j_0}} \int_{a-b}^{b-a} M^2(u)\sigma^2(F^{-1}(\tau - \frac{u}{2^{j_0}}))du.$$

Notice that when $n \to \infty$, $\frac{T_{W12}}{T_{W1D}^2} = O(2^{j_0})$, while $\frac{T_{W13}}{T_{W1D}^2} = O(1)$.

In the end,

$$\begin{aligned} \lim_{n\to\infty} \frac{n}{2^{j_0}} Var(\overline{\delta}_{W1}) &= \frac{\int_{a-b}^{b-a} M^2(u)\sigma^2(F^{-1}(\tau - \frac{u}{2^{j_0}}))du}{\left[\int_{a-b}^{b-a} M(v)I\{v \le 0\}dv\right]^2} \\ &= \frac{\sigma_+^2(v_0)\int_0^{b-a} M^2(v)dv + \sigma_-^2(v_0)\int_{a-b}^0 M^2(v)dv}{\left[\int_{a-b}^0 M(v)dv\right]^2}. \end{aligned}$$

Proof of Step 2: Note that

$$\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_{W1} - \overline{\delta}_{W1}) = \sqrt{\frac{n}{2^{j_0}}}\left[\frac{1}{n}\frac{\sum_{i=1}^n (A_i - C_i)\, Y_{i:n}}{B}\right] + \sqrt{\frac{n}{2^{j_0}}}\overline{\delta}_{W1}\left[\frac{D}{B} - 1\right],$$

where

$$\begin{aligned} A_i &= \int_0^1 \frac{1}{n}\sum_{j=1}^n I\{t_j \ge \tau\}\psi_{j_0}\left[t_j - t\right]\psi\left[2^{j_0}(t_i - t)\right] D_{j_0}(t)dt, \\ B &= \int_0^1 \left[\frac{1}{n}\sum_{j=1}^n 2^{j_0/2} I\{t_j \ge \tau\}\psi\left[2^{j_0}(t_j - t)\right]\right]^2 D_{j_0}(t)dt, \\ C_i &= \int_0^1\int_0^1 D_{j_0}(t)I\{w \ge \tau\}\psi_{j_0}\left[w - t\right]\psi\left[2^{j_0}(t_i - t)\right] dtdw, \\ D &= \int_0^1\int_0^1\int_0^1 D_{j_0}(t)I\{w \ge \tau\}I\{v \ge \tau\}\psi_{j_0}\left[w - t\right]\psi\left[2^{j_0}(v - t)\right] dwdvdt. \end{aligned}$$

For the term $\sqrt{\frac{n}{2^{j_0}}}\overline{\delta}_{W1}\left[\frac{D}{B} - 1\right]$, note that

$$\begin{aligned} & \lim_{n\to\infty} |B - D| \\ \le\ & \lim_{n\to\infty} \int_0^1 \left| \frac{1}{n}\sum_{j=1}^n 2^{j_0/2} I\{t_j \ge \tau\}\psi\left[2^{j_0}(t_j - t)\right] - \int_0^1 2^{j_0/2} I\{w \ge \tau\}\psi\left[2^{j_0}(w - t)\right] dw \right| \\ & \cdot \left| \frac{1}{n}\sum_{j=1}^n 2^{j_0/2} I\{t_j \ge \tau\}\psi\left[2^{j_0}(t_j - t)\right] + \int_0^1 2^{j_0/2} I\{v \ge \tau\}\psi\left[2^{j_0}(v - t)\right] dv \right| D_{j_0}(t)dt \\ \le\ & \lim_{n\to\infty} \sup_{t\in D(t)} 2^{j_0/2} \left| \frac{1}{n}\sum_{j=1}^n I\{t_j \ge \tau\}\psi\left[2^{j_0}(t_j - t)\right] - \int_0^1 I\{w \ge \tau\}\psi\left[2^{j_0}(w - t)\right] dw \right| \\ & \cdot \sup_{t\in D(t)} 2^{j_0/2} \left| \frac{1}{n}\sum_{j=1}^n I\{t_j \ge \tau\}\psi\left[2^{j_0}(t_j - t)\right] + \int_0^1 I\{v \ge \tau\}\psi\left[2^{j_0}(v - t)\right] dv \right| \int_0^1 D_{j_0}(t)dt \\ =\ & O(\frac{2^{j_0/2}}{n}) \cdot O(\frac{1}{2^{j_0/2}}) \cdot O(\frac{1}{2^{j_0}}) \\ =\ & O(\frac{1}{n2^{j_0}}). \end{aligned}$$

Thus,

$$\sqrt{\frac{n}{2^{j_0}}}\bar{\delta}_{W1}\left[\frac{D}{B}-1\right]=o_p(1).$$

Since $D=O(\frac{1}{2^{2j_0}})$ from T_{W1D} term, thus $B=O(\frac{1}{2^{2j_0}})$. And note that

$$\begin{aligned}
& \sqrt{\frac{n}{2^{j_0}}}\left[\frac{1}{n}\frac{\sum_{i=1}^{n}\left(A_i-C_i\right)Y_{[i:n]}}{B}\right] \\
\leq\ & \sqrt{\frac{n}{2^{j_0}}}\frac{1}{B}\sup_{t\in D(t)}\left|\begin{array}{c}\frac{1}{n}\sum_{j=1}^{n}I\{t_j\geq\tau\}\psi_{j_0}\left[t_j-t\right] \\ -\int_0^1 I\{w\geq\tau\}\psi_{j_0}\left[w-t\right]dw\end{array}\right|\frac{1}{n}\sum_{i=1}^{n}\int_0^1 D_{j_0}(t)dt\sup_{t\in D(t)}\left|\psi\left[2^{j_0}(t_i-t)\right]\right|Y_{[i:n]} \\
=\ & \sqrt{\frac{n}{2^{j_0}}}\frac{1}{B}\sup_{t\in D(t)}\left|\begin{array}{c}\frac{1}{n}\sum_{j=1}^{n}I\{t_j\geq\tau\}\psi_{j_0}\left[t_j-t\right] \\ -\int_0^1 I\{w\geq\tau\}\psi_{j_0}\left[w-t\right]dw\end{array}\right|\frac{1}{n}\sum_{i=1}^{n}\sup_{t\in D(t)}\left|\psi_{j_0}\left[t_i-t\right]\right|Y_{[i:n]}\frac{\int_0^1 D_{j_0}(t)dt}{2^{j_0}} \\
=\ & O(\sqrt{\frac{n}{2^{j_0}}})\cdot O(2^{2j_0})\cdot O(\frac{2^{j_0}}{n})\cdot O_p(1)\cdot O(\frac{1}{2^{2j_0}}) \\
=\ & O_p(\sqrt{\frac{2^{j_0}}{n}})=o_p(1).
\end{aligned}$$

In the end, we have: $\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_{W1}-\bar{\delta}_{W1})=o_p\left(1\right)$. **Q.E.D.**

Proof of Theorem 4.3: We begin with the simplest case where $j_L=j_0$ and $j_U=j_0+1$. Then by induction we prove the general case.

When $j_L=j_0$, and $j_U=j_0+1$, we have:

$$\begin{aligned}
\widehat{\delta}_{W2} &= \frac{\widehat{\Delta}_{j_0}^{Y}(\tau)\widehat{\Delta}_{j_0}^{d}(\tau)+\widehat{\Delta}_{j_0+1}^{Y}(\tau)\widehat{\Delta}_{j_0+1}^{d}(\tau)}{\left[\widehat{\Delta}_{j_0}^{d}(\tau)\right]^2+\left[\widehat{\Delta}_{j_0+1}^{d}(\tau)\right]^2} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n}\left\{\begin{array}{c}\frac{1}{n}\sum_{l=1}^{n}I\{t_l\geq\tau\}\psi_{j_0}[t_l-\tau]\psi[2^{j_0}(t_i-\tau)] \\ +\frac{2}{n}\sum_{l=1}^{n}I\{t_l\geq\tau\}\psi_{j_0}[2(t_l-\tau)]\psi[2^{j_0+1}(t_i-\tau)]\end{array}\right\}Y_{[i:n]}}{\left[\frac{2^{j_0/2}}{n}\sum_{l=1}^{n}I\{t_l\geq\tau\}\psi[2^{j_0}(t_i-\tau)]\right]^2+\left[\frac{2^{(j_0+1)/2}}{n}\sum_{l=1}^{n}I\{t_l\geq\tau\}\psi[2^{j_0+1}(t_i-\tau)]\right]^2}.
\end{aligned}$$

Let

$$\bar{\delta}_{W2}=\frac{1}{n}\sum_{i=1}^{n}J_{W2}(\frac{i}{n})Y_{[i:n]},$$

where

$$J_{W2}(\frac{i}{n})=\frac{2^{j_0}\left\{\psi[2^{j_0}(\frac{i}{n}-\tau)]+\psi[2^{j_0+1}(\frac{i}{n}-\tau)]\right\}}{(1+\frac{1}{2})\int_0^b\psi(t)dt}.$$

We will complete the proof in two steps:

Step 1. We show $\bar{\delta}_{W2}$ has the asymptotic distributions stated in the theorem;

Step 2. We show $\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_{W2}-\bar{\delta}_{W2})=o_p\left(1\right)$.

Proof of Step 1: First

$$\begin{aligned}
& \int_{-\infty}^{\infty}\left[g(v)+\delta I\{v\geq v_0\}\right]J_{W2}(F_V(v))dF_V(v)=\int_{-\infty}^{\infty}g(v)J_{W2}(F_V(v))dF_V(v)+\delta \\
=\ & \frac{2}{3}\int_{-\infty}^{\infty}g(v)\frac{\psi_{j_0}[F_V(v)-\tau]}{\int_0^b\psi(t)dt}dF_V(v)+\frac{1}{3}\int_{-\infty}^{\infty}g(v)\frac{2\psi_{j_0}[2(F_V(v)-\tau)]}{\int_0^b\psi(t)dt}dF_V(v)+\delta \\
=\ & \frac{2}{3}P_1+\frac{1}{3}P_2+\delta,
\end{aligned}$$

where

$$\begin{aligned}P_1 &= \int_{-\infty}^{\infty} g(v)\frac{\psi_{j_0}[F_V(v)-\tau]}{\int_0^b \psi(t)dt}dF_V(v),\\ P_2 &= \int_{-\infty}^{\infty} g(v)\frac{2\psi_{j_0}[2(F_V(v)-\tau)]}{\int_0^b \psi(t)dt}dF_V(v).\end{aligned}$$

From the proof of Theorem 3.1, we know: under A2(G)(a):

$$\begin{aligned}P_1 &= \frac{\frac{1}{2^{j_0}}\left[G_{+}^{(1)}(\tau)-G_{-}^{(1)}(\tau)\right]\int_0^b \psi(u)udu}{\int_0^b \psi(u)du}+s.o.,\\ P_2 &= \frac{\frac{1}{2^{j_0+1}}\left[G_{+}^{(1)}(\tau)-G_{-}^{(1)}(\tau)\right]\int_0^b \psi(u)udu}{\int_0^b \psi(u)du}+s.o.\end{aligned}$$

Then,

$$\frac{2}{3}P_1+\frac{1}{3}P_2=\frac{\frac{5}{2^{j_0}}\left[G_{+}^{(1)}(\tau)-G_{-}^{(1)}(\tau)\right]\int_0^b \psi(u)udu}{6\int_0^b \psi(u)du}+s.o.$$

Again from the proof of Theorem 3.1, we know: under A2(G)(b):

$$\begin{aligned}P_1 &= \frac{(\frac{1}{2^{j_0}})^m G^{(m)}(\tau)\int_a^b u^m\psi(u)du}{\int_0^b \psi(u)du}+s.o.,\\ P_2 &= \frac{(\frac{1}{2^{j_0+1}})^m G^{(m)}(\tau)\int_a^b u^m\psi(u)du}{\int_0^b \psi(u)du}+s.o.\end{aligned}$$

Then,

$$\frac{2}{3}P_1+\frac{1}{3}P_2=\frac{(\frac{1}{2^{j_0}})^m(2+\frac{1}{2^m})G^{(m)}(\tau)\int_a^b u^m\psi(u)du}{3\int_0^b \psi(u)du}+s.o.$$

Thus when $K_n=1$,

$$\begin{aligned}\lim_{n\to\infty}2^{j_0}\left[E(\overline{\delta}_{W2})-\delta\right] &= \frac{5\left[G_{+}^{(1)}(\tau)-G_{-}^{(1)}(\tau)\right]\int_0^b \psi(u)udu}{6\int_0^b \psi(u)du}, \text{ under A2(G)(a);}\\ \lim_{n\to\infty}2^{mj_0}\left[E(\overline{\delta}_{W2})-\delta\right] &= \frac{(2+\frac{1}{2^m})G^{(m)}(\tau)\int_a^b u^m\psi(u)du}{3\int_0^b \psi(u)du}, \text{ under A2(G)(b).}\end{aligned}$$

Second, let us calculate $\lim_{n\to\infty} nVar(\overline{\delta}_{W2})$ with $K_n=1$:

$$\begin{aligned}&\int_{-\infty}^{\infty}\left[\frac{2^{j_0}\left\{\psi[2^{j_0}(F_V(v)-\tau)]+\psi[2^{j_0+1}(F_V(v)-\tau)]\right\}}{(1+\frac{1}{2})\int_0^b \psi(t)dt}\right]^2\sigma^2(v)dF_V(v)\\ &+\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left\{\frac{\begin{array}{c}[F_V(v_1\wedge v_2)-F_V(v_1)F_V(v_2)]\\ \left\{\begin{array}{c}\psi_{j_0}[F_V(v_1)-\tau]\\ +\psi_{j_0}[2(F_V(v_1)-\tau)]\end{array}\right\}\left\{\begin{array}{c}\psi_{j_0}[F_V(v_2)-\tau]\\ +\psi_{j_0}[2(F_V(v_2)-\tau)]\end{array}\right\}\end{array}}{\left[\frac{3}{2}\int_0^b \psi(t)dt\right]^2}\right\}dm(v_1)dm(v_2)\\ =\ & A_{W21}+A_{W22},\end{aligned}$$

where

$$\begin{aligned}
A_{W21} &= \int_{-\infty}^{\infty}\left[\frac{2^{j_0}\left\{\psi[2^{j_0}(F_V(v)-\tau)]+\psi[2^{j_0+1}(F_V(v)-\tau)]\right\}}{\left(1+\frac{1}{2}\right)\int_0^b\psi(t)dt}\right]^2\sigma^2(v)dF_V(v),\\
A_{W22} &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left\{\begin{array}{c}[F_V(v_1\wedge v_2)-F_V(v_1)F_V(v_2)]\\ \dfrac{\left\{\begin{array}{c}\psi_{j_0}[F_V(v_1)-\tau]\\ +\psi_{j_0}[2(F_V(v_1)-\tau)]\end{array}\right\}\left\{\begin{array}{c}\psi_{j_0}[F_V(v_2)-\tau]\\ +\psi_{j_0}[2(F_V(v_2)-\tau)]\end{array}\right\}}{\left[\left(1+\frac{1}{2}\right)\int_0^b\psi(t)dt\right]^2}\end{array}\right\}dm(v_1)dm(v_2).
\end{aligned}$$

Since,

$$\begin{aligned}
A_{W21} &= \int_{-\infty}^{\infty}\left[\frac{2^{j_0}\left\{\psi[2^{j_0}(F_V(v)-\tau)]+\psi[2^{j_0+1}(F_V(v)-\tau)]\right\}}{\left(1+\frac{1}{2}\right)\int_0^b\psi(t)dt}\right]^2\sigma^2(v)dF_V(v)\\
&= \frac{4}{9}\int_{-\infty}^{\infty}\left[\frac{\psi_{j_0}[F_V(v)-\tau]}{\int_0^b\psi(t)dt}\right]^2\sigma^2(v)dF_V(v)+\frac{1}{9}\int_{-\infty}^{\infty}\left[\frac{2\psi_{j_0}[2(F_V(v)-\tau)]}{\int_0^b\psi(t)dt}\right]^2\sigma^2(v)dF_V(v)\\
&\quad +\frac{4}{9}\int_{-\infty}^{\infty}\frac{\psi_{j_0}[F_V(v)-\tau]2^{j_0+1}\psi[2^{j_0+1}(F_V(v)-\tau)]}{\left[\int_0^b\psi(t)dt\right]^2}\sigma^2(v)dF_V(v)\\
&= \frac{2^{j_0+1}}{3}A_1+\frac{4}{9}\int_{-\infty}^{\infty}\frac{2^{j_0}\psi[2^{j_0}(F_V(v)-\tau)]2^{j_0+1}\psi[2^{j_0+1}(F_V(v)-\tau)]}{\left[\int_0^b\psi(t)dt\right]^2}\sigma^2(v)dF_V(v)+s.o.\\
&= \frac{2^{j_0+1}}{3}A_1+\frac{8}{9}\frac{2^{j_0}}{\left[\int_0^b\psi(t)dt\right]^2}\left[\sigma_+^2(v_0)-\sigma_-^2(v_0)\right]\int_0^b\psi(u)\psi(2u)du+s.o.,
\end{aligned}$$

when $n\to\infty$, $A_{W21}=O(2^{j_0})$, while $A_{W22}=O(1)$. Therefore when $K_n=1$,

$$\lim_{n\to\infty}\frac{n}{2^{j_0}}Var(\overline{\delta}_{W2})=\frac{2}{3}V+\frac{2^{j_0+3\left[\sigma_+^2(v_0)-\sigma_-^2(v_0)\right]\int_0^b\psi(u)\psi(2u)du}}{9\left[\int_0^b\psi(t)dt\right]^2}.$$

Proof of Step 2. It is similar to that of Theorem 3.1. **Q.E.D.**

Proof of Theorem 4.5: We begin with the simplest case with $j_L=j_0$ and $j_U=j_0+1$. Then by induction, we prove the general case.

When $j_L=j_0$, and $j_U=j_0+1$, we have:

$$\begin{aligned}
\widehat{\delta}_{W3} &= \frac{\sum_{j=j_L}^{j_U}\int_0^1\widehat{\Delta}_j^Y(t)\,\widehat{\Delta}_j^d(t)\,D_j(t)dt}{\sum_{j=j_L}^{j_U}\int_0^1\left[\widehat{\Delta}_j^d(t)\right]^2D_j(t)dt}\\
&= \frac{\int_0^1\widehat{\Delta}_{j_0}^Y(t)\,\widehat{\Delta}_{j_0}^d(t)\,D_{j_0}(t)dt+\int_0^1\widehat{\Delta}_{j_0+1}^Y(t)\,\widehat{\Delta}_{j_0+1}^d(t)\,D_{j_0+1}(t)dt}{\int_0^1\left[\widehat{\Delta}_{j_0}^d(t)\right]^2D_{j_0}(t)dt+\int_0^1\left[\widehat{\Delta}_{j_0+1}^d(t)\right]^2D_{j_0+1}(t)dt}.
\end{aligned}$$

Let

$$\overline{\delta}_{W3}=\frac{1}{n}\sum_{i=1}^nJ_{W3}\left(\frac{i}{n}\right)Y_{[i:n]},$$

where

$$\begin{aligned}
J_{W3}(\frac{i}{n}) &= \frac{Z_{W3}^{1}(\frac{i}{n}) + Z_{W3}^{2}(\frac{i}{n})}{Q_1 + Q_2}, \\
Z_{W3}^{1}(\frac{i}{n}) &= \int_0^1 \int_0^1 I\{w \geq \tau\} \psi_{j_0}\left[w - t\right] \psi\left[2^{j_0}(\frac{i}{n} - t)\right] D_{j_0}(t) dt dw, \\
Z_{W3}^{2}(\frac{i}{n}) &= \int_0^1 \int_0^1 I\{w \geq \tau\} 2\psi_{j_0}\left[2(w - t)\right] \psi\left[2^{j_0+1}(\frac{i}{n} - t)\right] D_{j_0+1}(t) dt dw, \\
Q_1 &= \int_0^1 \int_0^1 \int_0^1 I\{u \geq \tau\} I\{v \geq \tau\} \psi_{j_0}\left[u - t\right] \psi\left[2^{j_0}(v - t)\right] D_{j_0}(t) du dv dt, \\
Q_2 &= \int_0^1 \int_0^1 \int_0^1 I\{u \geq \tau\} I\{v \geq \tau\} 2\psi_{j_0}\left[2(u - t)\right] \psi\left[2^{j_0+1}(v - t)\right] D_{j_0+1}(t) du dv dt.
\end{aligned}$$

We will complete the proof in two steps:

Step 1. We show $\overline{\delta}_{W3}$ has the asymptotic distributions stated in the theorem;

Step 2. We show: $\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_{W3} - \overline{\delta}_{W3}) = o_p(1)$.

Proof of Step 1: First, let us calculate the $\lim_{n \to \infty} E(\overline{\delta}_{W3})$:

$$\int_{-\infty}^{\infty} \left[g(v) + \delta I\{v \geq v_0\}\right] J_{W3}(F_V(v)) dF_V(v) = \int_{-\infty}^{\infty} g(v) J_{W3}(F_V(v)) dF_V(v) + \delta.$$

From the proof of Theorem 4.1, we know

$$Q_1 + Q_2 = \frac{5 \int_{-\infty}^{0} M(v) dv}{2^{2j_0+2}}.$$

Under A2(G)(a):

$$\begin{aligned}
&\int_{-\infty}^{\infty} g(v) \left[Z_{W3}^{1}(F_V(v)) + Z_{W3}^{2}(F_V(v))\right] dF_V(v) \\
= \;& 9(\frac{1}{2^{j_0+1}})^3 \left[G_{+}^{(1)}(\tau) - G_{-}^{(1)}(\tau)\right] \int_a^b \int_a^b L(t) \psi(s)(s - t) I\{s - t \geq 0\} ds dt + s.o..
\end{aligned}$$

Then when $K_n = 1$, ignoring higher order terms, we have

$$\begin{aligned}
&\int_{-\infty}^{\infty} g(v) J_{W3}(F_V(v)) dF_V(v) \\
= \;& \frac{9(\frac{1}{2^{j_0}}) \left[G_{+}^{(1)}(\tau) - G_{-}^{(1)}(\tau)\right] \int_a^b \int_a^b L(t) \psi(s)(s - t) I\{s - t \geq 0\} ds dt}{10 \int_{a-b}^{0} M(v) dv}.
\end{aligned}$$

Under A2(G)(b) and $l_g \geq 2m$:

$$\begin{aligned}
&\int_{-\infty}^{\infty} g(v) \left[Z_{W3}^{1}(F_V(v)) + Z_{W3}^{2}(F_V(v))\right] dF_V(v) \\
= \;& [1 + (\frac{1}{2})^{2m+1}] \frac{(\frac{1}{2^{j_0}})^{2m+1} G^{(2m-1)}(\tau) \int_a^b \psi(s) s^m ds \int_a^b L(t)(-t)^{m-1} dt}{m!(m-1)!} + s.o..
\end{aligned}$$

Then when $K_n = 1$,

$$\begin{aligned}
&\int_{-\infty}^{\infty} g(v) J_{W3}(F_V(v)) dF_V(v) \\
= \;& \frac{\left[4 + (\frac{1}{2})^{2m-1}\right] (\frac{1}{2^{j_0}})^{2m-1} G^{(2m-1)}(\tau) \int_a^b \psi(s) s^m ds \int_a^b L(t)(-t)^{m-1} dt}{5m!(m-1)! \int_{a-b}^{0} M(v) dv} + s.o..
\end{aligned}$$

Therefore, under A2(G)(a),

$$\begin{aligned}&\lim_{n\to\infty} 2^{j_0}\left[E(\overline{\delta}_{W3})-\delta\right]\\ =\quad&\frac{9\left[G_+^{(1)}(\tau)-G_-^{(1)}(\tau)\right]\int_a^b\int_a^b L(t)\psi(s)(s-t)I\{s-t\geq 0\}dsdt}{10\int_{a-b}^0 M(v)dv};\end{aligned}$$

under A2(G)(b),

$$\begin{aligned}&\lim_{n\to\infty} 2^{(2m-1)j_0}\left[E(\overline{\delta}_{W3})-\delta\right]\\ =\quad&\frac{\left[4+(\frac{1}{2})^{2m-1}\right]G^{(2m-1)}(\tau)\int_a^b\psi(s)s^m ds\int_a^b L(t)(-t)^{m-1}dt}{5m!(m-1)!\int_{a-b}^0 M(v)dv}.\end{aligned}$$

Second, let us calculate $\lim_{n\to\infty} nVar(\overline{\delta}_{W3})$ with $K_n=1$,

$$\begin{aligned}&\lim_{n\to\infty}\frac{n}{2^{j_0}}Var(\overline{\delta}_{W3})\\ =\quad&\frac{\int_{a-b}^{b-a}M^2(u)\sigma^2(F_V^{-1}(\tau-\frac{u}{2^{j_0}}))du}{2^{4j_0}(Q_1+Q_2)^2}+\frac{\int_{a-b}^{b-a}M^2(u)\sigma^2(F_V^{-1}(\tau-\frac{u}{2^{j_0}}))du}{2^{3j_0+3}(Q_1+Q_2)^2}\\ &+\frac{2}{(Q_1+Q_2)^2}\int_{-\infty}^{+\infty}Z_{W3}^1(F_V(v))Z_{W3}^2(F_V(v))\sigma^2(v)dF_V(v)\\ =\quad&\frac{18}{25}V_{W1}+\frac{Cross_1}{(Q_1+Q_2)^2},\end{aligned}$$

where

$$\begin{aligned}&Cross_1=2\int_{-\infty}^{+\infty}Z_{W3}^1(F_V(v))Z_{W3}^2(F_V(v))\sigma^2(v)dF_V(v)\\ =\quad&\frac{1}{2^{3j_0}}\int_{-\infty}^{+\infty}M(u)M(2u)\sigma^2(F_V^{-1}(\tau-\frac{u}{2^{j_0}}))du.\end{aligned}$$

Therefore when $K_n=1$,

$$\lim_{n\to\infty}\frac{n}{2^{j_0}}Var(\overline{\delta}_{W3})=\frac{18}{25}V_{W1}+\frac{2^{j_0+3}\left[\sigma_+^2(v_0)-\sigma_-^2(v_0)\right]\int_0^b M(u)M(2u)du}{9\left[\int_0^b\psi(t)dt\right]^2}.$$

The proof of second step could be obtained analogous to Theorem 4.1. **Q.E.D.**

Proof of Theorem 4.7: We will complete the proof in two steps:

Step 1. We show $\overline{\delta}_{KW1}$ has the asymptotic distributions stated in the theorem;

Step 2. We show: $\sqrt{\frac{n}{2^{3j_0}}}(\widehat{\delta}_{KW1}-\overline{\delta}_{KW1})=o_p(1)$.

Proof of Step 1. Note that

$$\begin{aligned}&\int_{-\infty}^{\infty}\left[g_K(v)+\delta_K(V-v_0)I\{v\geq v_0\}\right]J_{KW1}(F_V(v))dF_V(v)\\ =\quad&\int_{-\infty}^{\infty}g_K(v)J_{KW1}(F_V(v))dF_V(v)+\delta_K..\end{aligned}$$

For the first term,

$$\int_{-\infty}^{\infty}g_K(v)J_{KW1}(F_V(v))dF_V(v)=\frac{T_{KW1}}{T_{KW1D}},$$

where

$$\begin{aligned}
T_{KW1} &= \int_{-\infty}^{\infty} g_K(v) \begin{bmatrix} \int_0^1 \int_0^1 D_{j_0}(t)\left(F_V^{-1}(w) - v_0\right) I\{w \geq \tau\} \\ \cdot\psi_{j_0}\left[w - t\right] \psi\left[2^{j_0}(F_V(v) - t)\right] dtdw \end{bmatrix} dF_V(v), \\
T_{KW1D} &= \int_0^1 \int_0^1 \int_0^1 \begin{bmatrix} D_{j_0}(t)\left(F_V^{-1}(w) - v_0\right) I\{w \geq \tau\}\left(F_V^{-1}(v) - v_0\right) \\ \cdot I\{v \geq \tau\}\psi_{j_0}\left[w - t\right] \psi\left[2^{j_0}(v - t)\right] \end{bmatrix} dwdvdt.
\end{aligned}$$

Then for large enough j_0,

$$\begin{aligned}
& T_{KW1} \\
= {} & \int_0^1 \int_0^1 \left[\int_0^1 g_K(F_V^{-1}(u))\psi_{j_0}\left[u - t\right] du\right] D_{j_0}(t)\left(F_V^{-1}(w) - v_0\right) I\{w \geq \tau\}\psi\left[2^{j_0}(w - t)\right] dtdw \\
= {} & \int_0^1 \int_0^1 \left[\int_a^b g_K(F_V^{-1}(\frac{s}{2^{j_0}} + t))\psi(s)ds\right] D_{j_0}(t)\left(F_V^{-1}(w) - v_0\right) I\{w \geq \tau\}\psi\left[2^{j_0}(w - t)\right] dtdw \\
= {} & \int_0^1 \int_0^1 W_K(t)D_{j_0}(t)\left(F_V^{-1}(w) - v_0\right) I\{w \geq \tau\}\psi\left[2^{j_0}(w - t)\right] dtdw \\
= {} & \int_0^1 \left[\int_0^1 \left(F_V^{-1}(w) - v_0\right) I\{w \geq \tau\}\psi\left[2^{j_0}(w - t)\right] dw\right] D_{j_0}(t)W_K(t)dt \\
= {} & \frac{1}{2^{j_0}} \int_0^1 \left[\int_a^b \left[F_V^{-1}(\frac{w}{2^{j_0}} + t) - v_0\right] I\{w \geq 2^{j_0}(\tau - t)\}\psi\left[w\right] dw\right] D_{j_0}(t)W_K(t)dt,
\end{aligned}$$

where

$$W_K(t) = \int_a^b g_K(F_V^{-1}(\frac{s}{2^{j_0}} + t))\psi(s)ds.$$

Under the conditions of part (i) and ignoring higher order terms, we have:

$$\begin{aligned}
& T_{KW1} \\
= {} & \frac{1}{2^{2j_0}} \int_a^b L_0(W)\left[F_V^{-1}(\tau)\right]^{(1)}(-\frac{w}{2^{j_0}})W_K(\tau - \frac{w}{2^{j_0}})dw \\
& + \frac{1}{2^{2j_0}} \int_a^b L_1(w)\left[F_V^{-1}(\tau - \frac{w}{2^{j_0}})\right]^{(1)} W_K(\tau - \frac{w}{2^{j_0}})dw \\
= {} & \frac{1}{2^{5j_0} f_V(v_0)} \frac{1}{2}\left[G_{K+}^{(2)}(\tau) - G_{K-}^{(2)}(\tau)\right] \int_a^b \int_a^b (s - w)^2 \psi(s) I\{s - w \geq 0\}\left[L_1(w) - wL_0(w)\right] dsdw.
\end{aligned}$$

Under the conditions of part (ii) and ignoring higher order terms, we obtain:

$$\begin{aligned}
& T_{KW1} \\
= {} & \frac{1}{2^{j_0}} \int_0^1 \left[\int_a^b \left[\sum_{i=0}^{2m-1} \left[F_V^{-1}(t)\right]^{(i)} \left(\frac{w}{2^{j_0}}\right)^i \frac{1}{i!} - v_0)\right] I\{w \geq 2^{j_0}(\tau - t)\}\psi\left[w\right] dw\right] D_{j_0}(t)W_K(t)dt \\
= {} & \frac{1}{2^{j_0}} \left\{ \begin{matrix} \sum_{i=0}^{2m-1} \frac{1}{2^{(i+1)j_0}} \frac{1}{i!} \int_a^b L_i(w)\left[F_V^{-1}(\tau - \frac{w}{2^{j_0}})\right]^{(i)} W_K(\tau - \frac{w}{2^{j_0}})dw \\ - \frac{v_0}{2^{j_0}} \int_a^b L_0(w)W_K(\tau - \frac{w}{2^{j_0}})dw \end{matrix} \right\} \\
= {} & \frac{1}{2^{j_0}} \left\{ \sum_{p=1}^{2m-1} A_{1p}^{-} + \sum_{k=2}^{2m-1} A_k \right\},
\end{aligned}$$

where

$$\begin{aligned}
A_{1p}^{-} &= \frac{1}{2^{j_0}}\frac{1}{p!}\int_a^b L_0(W)\left[F_V^{-1}(\tau)\right]^{(p)}(-\frac{w}{2^{j_0}})^p W_K(\tau-\frac{w}{2^{j_0}})dw \text{ for } 2m-1\geq p\geq 1,\\
A_k &= \frac{1}{2^{kj_0}}\frac{1}{(k-1)!}\int_a^b L_{k-1}(w)\left[F_V^{-1}(\tau-\frac{w}{2^{j_0}})\right]^{(k-1)} W_K(\tau-\frac{w}{2^{j_0}})dw \text{ for } 2m-1\geq k\geq 2.
\end{aligned}$$

Notice that

$$\begin{aligned}
A_{11}^{-} &= \frac{1}{2^{(2+2m)j_0}}[F_V^{-1}(\tau)]^{(1)}\frac{G_K^{(2m)}(\tau)}{(m+1)!(m-1)!}\int_a^b L_0(t)(-t)^m dt\int_a^b \psi(s)s^{m+1}ds+s.o.,\\
A_{12}^{-} &= \frac{1}{2!}\frac{1}{2^{(2+2m)j_0}}[F_V^{-1}(\tau)]^{(2)}\frac{G_K^{(2m-1)}(\tau)}{(m+1)!(m-2)!}\int_a^b L_0(t)(-t)^m dt\int_a^b \psi(s)s^{m+1}ds+s.o.,\\
&\cdots\\
A_{1m}^{-} &= \frac{1}{m!}\frac{1}{2^{(2+2m)j_0}}[F_V^{-1}(\tau)]^{(m)}\frac{G_K^{(m+1)}(\tau)}{(m+1)!}\int_a^b L_0(t)(-t)^m dt\int_a^b \psi(s)s^{m+1}ds+s.o.,\\
A_{1p}^{-} &= o(\frac{1}{2^{(2+2m)j_0}}) \text{ for } p>m.
\end{aligned}$$

Thus,

$$\sum_{p=1}^{2m-1}A_{1p}^{-}=\frac{1}{2^{(2+2m)j_0}}\int_a^b L_0(t)(-t)^m dt\int_a^b \psi(s)s^{m+1}ds\sum_{i=1}^{m}[F_V^{-1}(\tau)]^{(i)}\frac{G_K^{(2m+1-i)}(\tau)}{i!(m+1)!(m-i)!}+s.o..$$

Ignoring higher order terms, we get: for A_2 term,

$$A_2=\frac{1}{2^{(2+2m)j_0}}\int_a^b L_1(t)(-t)^{m-1}dt\int_a^b \psi(s)s^{m+1}ds\sum_{i=1}^{m}[F_V^{-1}(\tau)]^{(i)}\frac{G_K^{(2m+1-i)}(\tau)}{(i-1)!(m+1)!(m-i)!};$$

for A_3 term,

$$\begin{aligned}
&A_3\\
=\ &\frac{1}{2!}\frac{1}{2^{(2+2m)j_0}}\int_a^b L_2(t)(-t)^{m-2}dt\int_a^b \psi(s)s^{m+1}ds\sum_{i=1}^{m-1}[F_V^{-1}(\tau)]^{(i+1)}\frac{G_K^{(2m-i)}(\tau)}{(i-1)!(m+1)!(m-i-1)!}.
\end{aligned}$$

Apply the similar procedure till A_{m+1} term,

$$A_{m+1}=\frac{1}{m!}\frac{1}{2^{(2+2m)j_0}}\int_a^b L_m(t)dt\int_a^b \psi(s)s^{m+1}ds\cdot[F_V^{-1}(\tau)]^{(m)}\frac{G_K^{(m+1)}(\tau)}{(m+1)!}+s.o..$$

And

$$A_q=o(\frac{1}{2^{(2+2m)j_0}}) \text{ when } q>m+1.$$

Therefore,

$$T_{KW1}=\frac{1}{2^{(3+2m)j_0}}\int_a^b \psi(s)s^{m+1}ds\left[\Gamma_0+\sum_{i=1}^{m}\Gamma_i\right].$$

For T_{KW1D} term,

$$\begin{aligned}
&\int_0^1\int_0^1\int_0^1 \left\{ \begin{array}{c} D_{j_0}(t)\left(F_V^{-1}(w)-v_0\right)I\{w\geq\tau\}\left(F_V^{-1}(v)-v_0\right) \\ \cdot I\{v\geq\tau\}\psi_{j_0}\left[w-t\right]\psi\left[2^{j_0}(v-t)\right] \end{array} \right\} dwdvdt \\
=\ & \int_0^1\int_0^1 \left\{ \begin{array}{c} \left\{\int_a^b \left[F_v^{-1}(t)-v_0+\sum_{i=1}^{\infty}\frac{1}{i!}[F_v^{-1}(t)]^{(i)}(\frac{w}{2^{j_0}})i\right] I\{w\geq 2^{j_0}(\tau-t)\}\psi(w)dw\right\} \\ \cdot D_{j_0}(t)\left(F_V^{-1}(v)-v_0\right)I\{v\geq\tau\}\psi\left[2^{j_0}(v-t)\right] \end{array} \right\} dtdv \\
=\ & C_1{+}C_2{+}RE,
\end{aligned}$$

where

$$\begin{aligned}
C_1 &= \int_0^1\int_0^1 \left[F_v^{-1}(t)-v_0\right] L_0(2^{j_0}(\tau-t))D_{j_0}(t)\left(F_V^{-1}(v)-v_0\right)I\{v\geq\tau\}\psi\left[2^{j_0}(v-t)\right]dtdv, \\
C_2 &= \int_0^1\int_0^1 \frac{\left[F_v^{-1}(t)\right]^{(1)}}{2^{j_0}} L_1(2^{j_0}(\tau-t))D_{j_0}(t)\left(F_V^{-1}(v)-v_0\right)I\{v\geq\tau\}\psi\left[2^{j_0}(v-t)\right]dtdv, \\
RE &= \sum_{i=2}^{\infty}\int_0^1\int_0^1 \frac{\left[F_v^{-1}(t)\right]^{(i)}}{i!2^{ij_0}} L_i(2^{j_0}(\tau-t))D_{j_0}(t)\left(F_V^{-1}(v)-v_0\right)I\{v\geq\tau\}\psi\left[2^{j_0}(v-t)\right]dtdv.
\end{aligned}$$

For C_1 term,

$$C_1{=}\frac{1}{2^{4j_0}f_V^2(v_0)}\int_{a-b}^0\left[t^2M(t)-tM_1(t)\right]dt+s.o.,$$

For C_2 term,

$$C2={}\frac{1}{2^{4j_0}f_V^2(v_0)}\int_{a-b}^0\left[-tM_2(t)\right]dt+s.o.,$$

For RE term, $RE=o(\frac{1}{2^{4j_0}})$. Hence,

$$T_{KW1D}{=}\frac{1}{2^{4j_0}f_V^2(v_0)}\int_{a-b}^0\left[t^2M(t)-tM_1(t)-tM_2(t)\right]dt+s.o..$$

Then under the conditions of part (i):

$$\begin{aligned}
&\lim_{n\to\infty}2^{j_0}\left[E(\overline{\delta}_K)-\delta_K\right] \\
=\ & \frac{\left[G_{K+}^{(2)}(\tau)-G_{K-}^{(2)}(\tau)\right]f_V(v_0)\int_a^b\int_a^b(s-w)^2\psi(s)I\{s-w\geq0\}\left[L_1(w)-wL_0(w)\right]dsdw}{2\int_{a-b}^0\left[t^2M(t)-tM_1(t)-tM_2(t)\right]dt}.
\end{aligned}$$

under the conditions of part (ii):

$$\lim_{n\to\infty}2^{mj_0}\left[E(\overline{\delta}_K)-\delta_K\right]{=}\frac{f_V^2(v_0)\int_a^b\psi(s)s^{m+1}ds\left[\Gamma_0+\sum_{i=1}^m\Gamma_i\right]}{\int_{a-b}^0\left[t^2M(t)-tM_1(t)-tM_2(t)\right]dt}.$$

For the asymptotic variance,

$$\begin{aligned}
&\frac{\int_{-\infty}^{\infty}\left[\int_0^1\int_0^1 D_{j_0}(t)(F_V^{-1}(w)-v_0)I\{w\geq\tau\}\psi_{j_0}\left[w-t\right]\psi\left[2^{j_0}(F_V(v)-t)\right]dtdw\right]^2\sigma^2(v)dF_V(v)}{T_{KW1D}^2} \\
=\ & \frac{T_{KW12}}{T_{KW1D}^2},
\end{aligned}$$

where

$$
\begin{aligned}
& T_{KW12} \\
= & \int_{-\infty}^{\infty} \left[\int_0^1 \int_0^1 D_{j_0}(t)(F_V^{-1}(w) - v_0) I\{w \geq \tau\} \psi_{j_0}[w-t] \, \psi \left[2^{j_0}(F_V(v) - t) \right] dtdw \right]^2 \sigma^2(v) dF_V(v) \\
= & \int_0^1 P_{KW1}^2 \sigma^2(F_V^{-1}(u)) du,
\end{aligned}
$$

in which

$$
\begin{aligned}
P_{KW1} = & \int_0^1 \int_0^1 D_{j_0}(t)(F_V^{-1}(w) - v_0) I\{w \geq \tau\} \psi_{j_0}[w-t] \, \psi \left[2^{j_0}(u - t) \right] dtdw \\
= & \int_0^1 \left\{ \int_a^b \left[F_V^{-1}(t) - v_0 + [F_V^{-1}(t)]^{(1)} (\frac{w}{2^{j_0}}) + s.o. \right] I\{w \geq 2^{j_0}(\tau - t)\} \psi(w) dw \right\} \\
& D_{j_0}(t) \psi \left[2^{j_0}(u - t) \right] dt \\
= & \int_0^1 \left[F_V^{-1}(t) - v_0 \right] L_0 \left[2^{j_0}(\tau - t) \right] D_{j_0}(t) \psi \left[2^{j_0}(u - t) \right] dt \\
& + \frac{1}{2^{j_0}} \int_0^1 \left[F_V^{-1}(t) \right]^{(1)} L_1 \left[2^{j_0}(\tau - t) \right] D_{j_0}(t) \psi \left[2^{j_0}(u - t) \right] dt + s.o. \\
= & \frac{1}{2^{j_0} f_V(v_0)} (u - \tau) M \left[2^{j_0}(\tau - u) \right] + \frac{1}{2^{2j_0} f_V(v_0)} M_1 \left[2^{j_0}(\tau - u) \right] \\
& + \frac{1}{2^{2j_0} f_V(v_0)} M_2 \left[2^{j_0}(\tau - u) \right] + s.o..
\end{aligned}
$$

Thus

$$
T_{KW12} = \frac{1}{2^{5j_0} f_V^2(v_0)} \left[\begin{array}{c} \sigma_{\varepsilon+}^2(v_0) \int_{a-b}^0 \left[M_1(t) + M_2(t) - tM(t) \right]^2 dt \\ + \sigma_{\varepsilon-}^2(v_0) \int_0^{b-a} \left[M_1(t) + M_2(t) - tM(t) \right]^2 dt \end{array} \right] + s.o.
$$

and

$$
\begin{aligned}
& \lim_{n \to \infty} \frac{n}{2^{3j_0}} Var(\overline{\delta}_{KW1}) \\
= & \frac{f_V^2(v_0) \left[\sigma_{\varepsilon+}^2(v_0) \int_{a-b}^0 \left[M_1(t) + M_2(t) - tM(t) \right]^2 dt + \sigma_{\varepsilon-}^2(v_0) \int_0^{b-a} \left[M_1(t) + M_2(t) - tM(t) \right]^2 dt \right]}{\left[\int_{a-b}^0 \left[t^2 M(t) - tM_1(t) - tM_2(t) \right] dt \right]^2}.
\end{aligned}
$$

The proof of the second step could be obtained analogous to Theorem 4.1. **Q.E.D.**

References

[1] Abramovich, F. and Samarov, A. (2000), "On One-Sided Estimation Of A Sharp Cusp Using Wavelets," *Unpublished manuscript.*

[2] Altonji, J.G. and R.L. Matzkin (2005), "Cross Section and Panel Data Estimators for Nonseparable Models With Endogenous Regressors," *Econometrica*, Vol. 73, 1053-1102.

[3] Antoniadis, A. and Gijbels, I. (1997), "Detecting Abrupt Changes by Wavelet Methods," *Journal of Nonparametric Statistics*, Vol 14, 7-29.

[4] Bowman, A. W., Pope, A. and Ismail, B. (2006), "Detecting discontinuities in nonparametric regression curves and surfaces," *Statistics and Computing,* Vol 16, 377-390.

[5] Card, D.E., Lee, D.S. and Pei, Z. (2009) "Quasi-Experimental Identification and Estimation in the Regression Kink Design," Working paper 1206, Princeton University.

[6] Cheng, M.Y. (2007), "On Design Sparseness Problem in Local Polynomial Regression," *International Congress of Chinese Mathematicians*, Vol. III, 538-545.

[7] Cheng, M.Y. and Raimondo, M. (2008), "Kernel Methods for Optimal Change-Points Estimation in Derivatives," *Journal of Computational and Graphical Statistics*, Vol 17, 56-75.

[8] Chen, S. (2002), "Local Linear Smoothers Using Asymmetric Kernels," *Annals of the Institute of Statistical Mathematics*, Vol 54, 312-323.

[9] Chesher, A. (2003), "Identification in Nonseparable Models," *Econometrica*, Vol. 71, 1405-1441.

[10] Chesher, A. (2005), "Nonparametric Identification Under Discrete Variation," *Econometrica*, Vol. 73, 1525-1550.

[11] Cline, D.B.H., Eubank, R.L. and Speckman, P.L. (1995), "Nonparametric Estimation of Regression Curves with Discontinuous Derivatives," *Journal of Statistical Research*, 17-30.

[12] Chu, B. and D. T. Jacho-Chavez (2010), "k-Nearest Neighbor Estimation of Inverse-Density-Weighted Expectations with Dependent Data," Working Paper.

[13] Delgado, MA. and Hidalgo, Javier (2000), "Nonparametric Inference on Structural Breaks," *Journal of Econometrics*, Vol. 96, 113-144.

[14] Desmet, L. and Gijbels, I. (2009), "Local Linear Fitting and Improved Estimation near Peaks," *Canadian Journal of Statistics*, Vol. 37, 453-475.

[15] Dong, Y. (2010), "Jumpy or Kinky? Regression Discontinuity without the Discontinuity," manuscript, California State University Fullerton.

[16] Eubank, R.L. and Speckman, P.L. (1994), "Nonparametric Estimation of Functions with Jump Discontinuities," *Lecture Notes-Monograph Series*, Vol 23, 130-144.

[17] Eubank,R. L. and Whitney, P. (1989), "Convergence Rates for Estimation in Certain Partially Linear Models," *Journal of Statistical Planning and Inference*, Vol. 23, 33-43.

[18] Florens, J.P., J.J. Heckman, C. Meghir, and E. Vytlacil (2008), "Identification of Treatment Effects Using Control Functions in Models With Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica*, Vol. 76, 1191-1206.

[19] Frölich, M. (2007), "Regression discontinuity design with covariates," IZA Discussion Papers 3024: 1-25.

[20] Gao, J., Pettitt, P. and Rodney, C.L.W. (1998), "Local Linear Kernel Estimation for Discontinuous Nonparametric Regression Functions," *Communications in Statistics-Theory and Methods*, Vol. 27, 2871-2894.

[21] Gijbels, I. and Goderniauxa, A.C. (2005), “Data-Driven Discontinuity Detection in Derivatives of a Regression Function,” *Communications in Statistics - Theory and Methods*, Vol. 33, 851-871.

[22] Gijbels, I., Hall, P. and Kneip, A. (1999), “On the Estimation of Jump Points in Smooth Curves,” *Annals of the Institute of Statistical Mathematics*, Vol. 51, 231-251.

[23] Gijbels, I., Lambert, A. and Peihua, Q. (2007), “Jump-Preserving Regression and Smoothing using Local Linear Fitting: A Compromise,” *Annals of the Institute of Statistical Mathematics*, Vol. 59, 235-272.

[24] Hahn, J., P. Todd, and W. Van der Klaauw (1999), “Evaluating The Effect of an Antidiscrimination Law using a Regression–Discontinuity Design,” NBER Working paper 7131: 1–40.

[25] Hahn, J., P. Todd, and W. Van der Klaauw (2001), “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design,” *Econometrica*, Vol. 69, 201-209.

[26] Hall, P., B.U. Park, and B.A. Turlach (1998), “A note on design transformation and binning in nonparametric curve estimation,” *Biometrika*, Vol. 85, 469–476.

[27] Heckman, J.J. (2008), “Econometric Causality,” *International Economic Review*, Vol. 76, 1-27.

[28] Heckman, J.J. and E. Vytlacil (2005), “Structural Equations, Treatment Effects, and Econometric Policy Evaluation,” *Econometrica*, Vol. 73, 669-738.

[29] Hoekstra, M. (2009), “The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach,” *The Review of Economics and Statistics*, Vol. 91, 717-724.

[30] Imbens, G. and K. Kalyanaramang (2009), “Optimal Bandwidth Choice for the Regression Discontinuity Estimator,” NBER Working paper 14726.

[31] Imbens, G. and W.K. Newey (2009), “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity,” *Econometrica*, Vol. 77, 1481-1512.

[32] Jones, M.C., Davies, S.J. and Park, B.U. (1994), “Versions of Kernel-Type Regression Estimators,” *Journal of the American Statistical Association*, Vol. 89, 823-832.

[33] Koo, J. (1997), “Spline Estimation of Discontinuous Regression Functions,” *Journal of Computational and Graphical Statistics*, Vol. 6, 266-284.

[34] Kotlyarova and V. Zinde-Walsh (2006), “Non and Semiparametric Estimation in Models with Unknown Smoothness,” *Economic Letters*, Vol. 93, 369-386.

[35] Kotlyarova and V. Zinde-Walsh (2008), “Robust Kernel Density Estimation in Models with Unknown Smoothness,” *Journal of Nonparametric Statistics*, Vol. 19, 89-101.

[36] Lee, D.S. (2008), “Randomized Experiments from Non-random Selection in U.S. House Elections,” *Journal of Econometrics*, Vol. 142, 675-697.

[37] Lee, D.S. and Card, D. (2008), ”Regression Discontinuity Inference with Specification Error,” *Journal of Econometrics*, Vol. 142, 655-674.

[38] Lee, D.S. and T. Lemieux (2009), “Regression Discontinuity Design in Economics,” NBER Working Paper 14723.

[39] Li, Y. and Xie, Z.J. (2000), “The Wavelet Detection of the Jump and Cusp Points of a Regression Function,” *Acta Mathematicae Applicatae Sinica*, Vol. 16, 283-291.

[40] Loader, C.R. (1996), “Change Point Estimation Using Nonparametric Regression,” *Annals of Statistics*, Vol. 24, 1667-1678.

[41] Ludwig, J. and D. Miller (2007), “Does Head Start Improve Children’s Life Chances? Evidence from a Regression Discontinuity Design,” *Quarterly Journal of Economics*, Vol. 122, 159-208.

[42] Mallat, Stephane (2009), *A Wavelet Tour of Signal Processing*, Third edition, Academic Press.

[43] Marschak, J. (1953), “Economic Measurements for Policy and Prediction,” *Studies in Econometric Method*, Eds. W. Hood & T. Koopmans, pp. 1-26. New York: Wiley.

[44] Muller, H. G. (1992), “Change-points in Nonparametric Regression Analysis,” *Annals of Statistics*, Vol. 20, 737-761.

[45] Nason, G. P. (2008), *Wavelet Methods in Statistics with R*, Springer

[46] Otsu, T. and K.-L. Xu (2010), “Estimation and Inference of Discontinuity in Density,” Manuscript.

[47] Park, C. and W. Kim (2004), “Estimation of a Regression Function with a Sharp Change Point using Boundary Wavelets,” *Statistics and Probability Letters*, Vol. 66, 435-448.

[48] Park, C. and W. Kim (2006), “Wavelet Estimation of a Regression Function with a Sharp Change Point in a Random Design,” *Journal of Statistical Planning and Inference*, Vol. 136, 2381-2394.

[49] Porter, J. (2003), “Estimation in the Regression Discontinuity Model,” manuscript, Harvard University.

[50] Qiu, P. and Yandell, B. (1998), “A Local Polynomial Jump-detection Algorithm in Nonparametric Regression,” *Technometrics*, Vol. 40, 141-152.

[51] Raimondo, M. (1998), “Minimax Estimation of Sharp Change Points,” *Annals of Statistics*, Vol. 26, 1379-1397.

[52] Rioul, O. (1992), “Simple Regularity Criteria for Subdivision Schemes,” *SIAM Journal on Mathematical Analysis*, Vol. 23, 1544-76.

[53] Seifert, B. and T. Gasser (1996), “Finite-Sample Variance of Local Polynomials: Analysis and Solutions,” *Journal of the American Statistical Association*, Vol. 91, 267-275.

[54] Seifert, B. and T. Gasser (2000), “Data Adaptive Ridging in Local Polynomial Regression,” *Journal of Computational and Graphical Statistics*, Vol. 9, 338–360.

[55] Shiaua, J.J. and Wahba, G. (1988), “Rates of Convergence of Some Estimators for a Semiparametric Model,” *Communications in Statistics - Simulation and Computation*, Vol. 17, 1117-1133.

[56] Spokoiny, V.G. (1998), “Estimation of a Function with Discontinuities via Local Polynomial Fit with an Adaptive Window Choice,” *Annals of Statistics*, Vol. 26, 1356-1378.

[57] Sun, Y. (2005), “Adaptive Estimation of the Regression Discontinuity Model,” Working Paper, UCSD.

[58] Vytlacil, E. (2006), “A Note on Additive Separability and Latent Index Models of Binary Choice: Representation Results,” *Oxford Bulletin of Economics and Statistics*, Vol. 68, 515-518.

[59] Vytlacil, E. and N. Yildiz (2007), “Dummy Endogenous Variables in Weakly Separable Models, ” *Econometrica*, Vol. 75, 757-779

[60] Wang L.H. and Cai, H.Y. (2010), “Wavelet Change-point Estimation for Long Memory Nonparametric Random Design Models,” *Journal of Time Series Analysis*, Vol. 31, 86-97.

[61] Wang, Y. (1995), “Jump and Sharp Cusp Detection by Wavelets,” *Biometrika*, Vol. 82, 385-397.

[62] Wishart, J. (2009), “Kink Estimation with Correlated Noise,” *Journal of the Korean Statistical Society*, Vol. 38, 131-143.

[63] Wu, J. S. and Chu, C. K. (1993), “Kernel-Type Estimators of Jump Points and Values of a Regression Function,” *Annals of Statistics*, Vol. 21, 1545-1566.

[64] Yang, S.-S. (1981), “Linear Functions of Concomitants of Order Statistics With Application to Nonparameteric Estimation of a Regression Function,” *Journal of the American Statistical Association*, Vol. 76, 658-662.

[65] Yang, Y. (2000), “Combining Different Procedures for Adaptive Regression,” *Journal of Multivariate Analysis*, 135-161.

[66] Zhou, Y., A.T.K. Wan, S. Xie, and X. Wang (2010), “Wavelet Analysis of Change-Points in a Nonparametric Regression with Heteroscedastic Variance,” *Journal of Econometrics*, forthcoming.

Table 1: Model 1

$n = 500$						
Scale	1	2	3	4	5	6
Single scale with single location $\widehat{\delta}$						
Bias	1.8899	0.51305	0.32227	0.23813	0.27612	0.21523
Std	0.02045	0.17181	0.24607	0.24607	0.35501	0.52015
MSE	3.5723	0.27465	0.13338	0.11725	0.20228	0.31688
Single scale with many locations $\widehat{\delta}_{W1}$						
Bias	1.1810	1.14143	0.60293	0.23328	0.25288	0.25542
Std	0.01060	0.02246	0.03693	0.05894	0.14328	0.14719
MSE	1.3950	1.30336	0.36489	0.05789	0.08448	0.08690
Many scales with single location $\widehat{\delta}_{W2}$ ($K_n = 2$)						
Bias	1.8091	0.43207	0.29186	0.24589	0.25263	0.21355
Std	0.01999	0.08601	0.12924	0.18799	0.27649	0.42293
MSE	3.2735	0.19408	0.10189	0.09581	0.14027	0.22447
Many scales with many locations $\widehat{\delta}_{W3}$ ($K_n = 2$)						
Bias	1.1609	1.02915	0.52829	0.23803	0.25363	0.25503
Std	0.00970	0.02070	0.03448	0.05750	0.10538	0.11776
MSE	1.3479	1.05958	0.28028	0.05996	0.07543	0.07891
$n = 2500$						
Scale	1	2	3	4	5	6
Single scale with single location $\widehat{\delta}$						
Bias	1.8256	0.52631	0.28373	0.24094	0.22803	0.25524
Std	0.00947	0.04580	0.07432	0.10881	0.15872	0.21760
MSE	3.3330	0.27910	0.08603	0.06989	0.07719	0.11250
Single scale with many locations $\widehat{\delta}_{W1}$						
Bias	1.1813	1.16796	0.48132	0.24961	0.25284	0.25431
Std	0.00495	0.00974	0.01878	0.02619	0.03831	0.05950
MSE	1.3957	1.36423	0.23203	0.06299	0.06539	0.06821
Many scales with single location $\widehat{\delta}_{W2}$ ($K_n = 2$)						
Bias	1.7437	0.43296	0.26356	0.23928	0.23944	0.25369
Std	0.00947	0.03819	0.05702	0.08025	0.11818	0.16321
MSE	3.0406	0.18892	0.07271	0.06369	0.07129	0.09099
Many scales with many locations $\widehat{\delta}_{W3}$ ($K_n = 2$)						
Bias	1.1621	1.04032	0.43238	0.25035	0.25310	0.25411
Std	0.00443	0.00923	0.01716	0.02357	0.03574	0.05456
MSE	1.3505	1.08235	0.18725	0.06323	0.06533	0.06755

Table 2: Model 2

$n = 500$						
Scale	1	2	3	4	5	6
Single scale with single location $\widehat{\delta}$						
Bias	0.56924	0.03767	-0.00589	-0.00827	0.00133	-0.00110
Std	0.02066	0.10921	0.17378	0.24238	0.35301	0.51970
MSE	0.32446	0.01334	0.03023	0.05881	0.12462	0.27009
Single scale with many locations $\widehat{\delta}_{W1}$						
Bias	0.16051	0.39249	0.19918	-0.00040	-0.00648	0.00037
Std	0.01073	0.02222	0.04216	0.05938	0.12250	0.14753
MSE	0.02588	0.15454	0.04145	0.00352	0.01504	0.02176
Many scales with single location $\widehat{\delta}_{W2}$ $(K_n = 2)$						
Bias	0.53853	0.02124	-0.00556	-0.00458	0.00166	0.00026
Std	0.02010	0.08693	0.13176	0.18840	0.27419	0.42692
MSE	0.29042	0.00800	0.01739	0.03551	0.07518	0.18226
Many scales with many locations $\widehat{\delta}_{W3}$ $(K_n = 2)$						
Bias	0.19271	0.35004	0.15825	-0.00143	-0.00400	0.00066
Std	0.00972	0.02115	0.03774	0.05535	0.09410	0.11833
MSE	0.03723	0.12297	0.02646	0.00306	0.00887	0.01400

$n = 2500$						
Scale	1	2	3	4	5	6
Single scale with single location $\widehat{\delta}$						
Bias	0.50762	0.09212	0.00137	-0.01301	-0.01088	-0.00316
Std	0.00954	0.04811	0.07736	0.10636	0.15041	0.22983
MSE	0.25777	0.01080	0.00598	0.01148	0.02274	0.05283
Single scale with many locations $\widehat{\delta}_{W1}$						
Bias	0.15464	0.42798	0.11587	-0.00024	-0.00138	-0.00730
Std	0.00518	0.01062	0.01712	0.02644	0.04047	0.05966
MSE	0.02394	0.18328	0.01372	0.00069	0.00164	0.00361
Many scales with single location $\widehat{\delta}_{W2}$ $(K_n = 2)$						
Bias	0.48121	0.05740	-0.00448	-0.0110	-0.00707	-0.00238
Std	0.00952	0.03820	0.05744	0.07872	0.11614	0.17130
MSE	0.23165	0.00475	0.00332	0.00631	0.01354	0.02935
Many scales with many locations $\widehat{\delta}_{W3}$ $(K_n = 2)$						
Bias	0.19268	0.36962	0.09124	-0.00075	-0.00285	-0.00795
Std	0.00472	0.00981	0.01567	0.02481	0.03719	0.05572
MSE	0.03715	0.13671	0.00857	0.00061	0.00139	0.00316

Table 3: Model 3

$n = 500$

Scale	1	2	3	4	5	6
Single scale with single location $\widehat{\delta}$						
Bias	1.3147	0.40031	0.17508	0.03774	0.03576	-0.00841
Std	0.02072	0.10775	0.17249	0.24849	0.34953	0.52059
MSE	1.7290	0.17186	0.06040	0.06317	0.12345	0.27108
Single scale with many locations $\widehat{\delta}_{W1}$						
Bias	0.72221	0.67465	0.26152	-0.01437	0.00140	0.00901
Std	0.01081	0.02238	0.03728	0.06064	0.12690	0.14753
MSE	0.52170	0.45565	0.06978	0.00388	0.01610	0.02184
Many scales with single location $\widehat{\delta}_{W2}$ $(K_n = 2)$						
Bias	1.2595	0.30005	0.11654	0.03105	0.01737	-0.01309
Std	0.02025	0.08542	0.13085	0.18742	0.27159	0.42165
MSE	1.5869	0.09732	0.03070	0.03609	0.07406	0.17796
Many scales with many locations $\widehat{\delta}_{W3}$ $(K_n = 2)$						
Bias	0.70405	0.58876	0.20586	-0.01022	0.00351	0.00749
Std	0.00980	0.02070	0.03460	0.05743	0.09661	0.11894
MSE	0.49579	0.34707	0.04357	0.00340	0.00934	0.01420

$n = 2500$

Scale	1	2	3	4	5	6
Single scale with single location $\widehat{\delta}$						
Bias	1.2660	0.40237	0.13346	0.05457	0.01173	0.01949
Std	0.00919	0.04578	0.07474	0.10968	0.16162	0.21138
MSE	1.6028	0.16400	0.02339	0.01500	0.02625	0.04506
Single scale with many locations $\widehat{\delta}_{W1}$						
Bias	0.72128	0.69221	0.17167	-0.00142	0.00070	0.00087
Std	0.00478	0.00975	0.01713	0.02711	0.03861	0.05751
MSE	0.52027	0.47926	0.02976	0.00073	0.00149	0.00330
Many scales with single location $\widehat{\delta}_{W2}$ $(K_n = 2)$						
Bias	1.2097	0.29508	0.09356	0.03735	0.01369	0.01499
Std	0.00893	0.03746	0.05788	0.08500	0.12059	0.16217
MSE	1.4636	0.08847	0.01210	0.00862	0.01473	0.02652
Many scales with many locations $\widehat{\delta}_{W3}$ $(K_n = 2)$						
Bias	0.70351	0.59553	0.13510	-0.00097	0.00065	0.00077
Std	0.00438	0.00910	0.01586	0.02480	0.03550	0.05481
MSE	0.49494	0.35474	0.01850	0.0006	0.00126	0.00300

Table 4: Model 4

			$n = 500$			
Scale	1	2	3	4	5	6
Single scale with single location $\widehat{\delta}$						
Bias	0.76726	0.13214	0.04122	-0.00451	0.01112	-0.01362
Std	0.02089	0.10759	0.17274	0.24254	0.35835	0.52279
MSE	0.58913	0.02904	0.03154	0.05884	0.12853	0.27350
Single scale with many locations $\widehat{\delta}_{W1}$						
Bias	0.46991	0.44709	0.17340	-0.00769	-0.00031	0.00466
Std	0.01096	0.02238	0.03677	0.05917	0.11750	0.14873
MSE	0.22094	0.20039	0.03142	0.00356	0.01380	0.02214
Many scales with single location $\widehat{\delta}_{W2}$ $(K_n = 2)$						
Bias	0.72995	0.09286	0.02400	-0.00127	0.00174	-0.01317
Std	0.02038	0.08484	0.13091	0.18938	0.28215	0.43292
MSE	0.53325	0.01582	0.01771	0.03586	0.07961	0.18759
Many scales with many locations $\widehat{\delta}_{W3}$ $(K_n = 2)$						
Bias	0.45927	0.39024	0.13675	-0.00567	0.00116	0.00395
Std	0.010004	0.02078	0.03406	0.05531	0.09130	0.12020
MSE	0.21103	0.15272	0.01986	0.00309	0.0083	0.01446

			$n = 2500$			
Scale	1	2	3	4	5	6
Single scale with single location $\widehat{\delta}$						
Bias	0.74038	0.13072	0.01623	0.00481	-0.00829	-0.00970
Std	0.00872	0.04857	0.07493	0.09976	0.14984	0.21606
MSE	0.54824	0.01944	0.00587	0.00997	0.02252	0.04677
Single scale with many locations $\widehat{\delta}_{W1}$						
Bias	0.47196	0.45551	0.10940	0.00092	0.00017	-0.00286
Std	0.00472	0.01058	0.01650	0.02731	0.03946	0.05732
MSE	0.22277	0.20760	0.01224	0.00074	0.00155	0.00329
Many scales with single location $\widehat{\delta}_{W2}$ $(K_n = 2)$						
Bias	0.70192	0.08772	0.00947	-0.00099	-0.00900	-0.01215
Std	0.00868	0.03806	0.05706	0.07918	0.11714	0.16592
MSE	0.49277	0.00914	0.00334	0.00627	0.01380	0.02767
Many scales with many locations $\widehat{\delta}_{W3}$ $(K_n = 2)$						
Bias	0.46062	0.39143	0.08640	0.00062	-0.00052	-0.00235
Std	0.00432	0.00970	0.01519	0.02541	0.03620	0.05276
MSE	0.21219	0.15331	0.00769	0.00064	0.00131	0.00278