ESSAYS ON WAVELET ESTIMATION OF THE LOCAL AVERAGE TREATMENT EFFECT

By

Heng Chen

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Economics

August, 2012

Nashville, Tennessee

Approved:

Professor Yanqin Fan

Professor Tong Li

Professor Mototsugu Shintani

Professor Bryan Shepherd

To my parents, Baoqi Chen and Xiaomei Wang.

ACKNOWLEDGMENTS

First and most deeply, I would thank my advisor, Professor Yanqin Fan. I'm very fortunate to be her student and learn a great deal from her profound knowledge. Her guidance and support help me get through many difficulties in my research and beyond. In particular her encouragements have inspired and gave me strengths in my most stressful periods. The first chapter is based on the joint work with her and the rest two chapters could not have been possible without her help and advices.

I am also very thankful to my committee members, Professor Tong Li, Professor Mototsugu Shintani and Professor Bryan Shepherd for their valuable comments and suggestions that have greatly improved the quality of the dissertation. In particular I am very grateful for Professor Tong Li and Professor Mototsugu Shintani to many simulating discussions and really appreciating Professor Bryan Shepherd for his thought-provoking explanations on the biostatistics examples.

Other faculty members, staff and classmates in the Department of Economics also gave me tremendous help over the years of my PhD study. I would like to thank them all.

My final and most heartfelt acknowledgment must go to my parents Baoqi Chen and Xiaomei Wang. Without their love, support and encouragement, I could not have been able to complete the dissertation on time.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

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

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 chapter 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 $b(V)$ 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

$$Y_1 = g_1(V, W), \; Y_0 = g_0(V, W), \tag{I.1}$$

$$D = I\{b(V) - U \geq 0\}, \tag{I.2}$$

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

[1]This set-up allows for the potential outcomes Y_1, Y_0 to depend on different components of W. Agent

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 but 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 $b(V) = I\{V \geq v_0\}$, 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 $g_1(V, W) = y(1, V, W)$ and $g_0(V, W) = y(0, V, W)$. In terms of the realized outcome Y, the potential outcomes model (I.1) and (I.2) can be written as

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

(I.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 (I.3) with possibly endogenous

selects treatment based on the threshold-crossing model (I.2). As shown in Vytlacil (2006), there is a larger class of latent index models that will have a representation of this form.

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 (I.1) and (I.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 selected. By allowing agent's selection of treatment to depend on her unobservable covariate U in (I.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 chapter is to show that under mild conditions, a policy parameter: the local average treatment effect (LATE), is identified in (I.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 the LATE, while Dong (2010) adopts a similar set-up. Instead, we impose smoothness conditions on the structural parameters in (I.1) and by exploiting the specific structure in (I.2), we are able to dispense with the local independence conditions.

The second contribution of this chapter is to propose several nonparametric estimators of the LATE using wavelets. First, we establish auxiliary regressions linking the policy parameter, the LATE, to jump sizes δ_0 and ζ_0 in (I.4) and (I.5) for discontinuous incentive assignment mechanisms, or kink sizes δ_1 and ζ_1 in (I.6) and (I.7) for kink incentive assignment mechanisms. In particular, the policy parameter LATE in (I.3) is given by δ_0/ζ_0 for discontinuous incentive assignment mechanisms and δ_1/ζ_1 for kink incentive assignment mechanisms. Thus, estimating the policy parameter LATE in (I.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 the LATE for RDD are applicable. These include estimators based on Nadaraya-Watson (NW) kernel regression (local constant kernel regression) or local polynomial kernel regression estimators of the jump sizes δ_0 and ζ_0 in which δ_0 (ζ_0) is estimated by the difference between two kernel 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 proposed a partial linear estimator of the LATE based on Robinson's (1998) partial linear estimators of δ_0 and ζ_0 and established asymptotic properties of these 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) proposed to extend existing work on local linear (polynomial) estimators from RDD to RKD without establishing the corresponding asymptotic theory.

Existing work in the econometrics literature suggest that the local polynomial kernel regression estimator appears to have the smallest asymptotic bias among the alternative estimators and achieves the optimal rate established in Porter (2003), which provide theoretical justifications for the popularity of local polynomial, especially local linear kernel estimators in applied research. In addition, for compactly supported kernels, Imbens and

Kalyanaraman (2009) derived the optimal bandwidth for local linear kernel estimators of the LATE. However, it is known in the statistics literature that local polynomial kernel estimators suffer from a serious drawback that for compactly supported kernels, the unconditional variance of a local polynomial kernel estimator is infinite and the MSE and the MSE optimal bandwidth are not defined, see Seifert and Gasser (1996). The afore-mentioned work on local polynomial kernel estimators of the LATE are based on expansions of the conditional variance and conditional MSE of the local polynomial kernel estimators. The infinite unconditional variance of local polynomial kernel 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 the random design to the regularly spaced fixed design (see Hall, Park, and Turlach (1998)). Hall, Park and Turlach (1998) demonstrate that in general their idea of binnng and transforming the data is superior to other approaches especially when there are jumps in the regression function.

This chapter proposes several local constant wavelet estimators of jump and kink sizes or equivalently the 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 coefficients of a function at a given location characterize its degree of local regularity (smoothness), so that large wavelet 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 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 i.i.d. 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 ($\bar{\delta}_0^{LC-SS}$ in our notation, see Section 3 (page 20)) of the jump size and establish its asymptotic distribution under the homoscedastic errors' condition. To the best of the authors' knowledge, the method of wavelets has not been used to estimate the kink size. Given the close connection between the estimation of the LATE in (I.3) and of jump/kink sizes in the corresponding auxiliary regressions, it seems natural to exploit this special feature of wavelet coefficients to estimate the LATE. The second part of this chapter accomplishes this objective.

For discontinuous incentive assignment mechanisms, the first wavelet estimator of the LATE we propose makes use of wavelet estimators of the jump sizes δ_0 and ζ_0 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). Firstly, 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*

[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) $\mathcal{F}_0$ 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 $\sup_{0<\tau<x<y} |g(x) - g(y)|/|x-y|^{\alpha'} < \infty$ 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 integer part of α;
(ii) $g^{(N)} \in \mathcal{F}_{\alpha-N}$.
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.

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 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 coefficients at locations close to v_0 and relatively large scales may also contain information on the jump size motivating our subsequent local constant wavelet estimators of the jump sizes and of the LATE parameter. Specifically, we propose three new wavelet estimators of the jump sizes δ_0 and ζ_0 using wavelet coefficients at locations close to v_0 and/or more than one scale: $\widehat{\delta}_0^{LC-SM}$, the single-scale estimator making use of wavelet coefficients at a single scale and more than one location; $\widehat{\delta}_0^{LC-MS}$, the single location estimator making use of wavelet coefficients at one location v_0 and more than one scale; and $\widehat{\delta}_0^{LC-MM}$, the multiple scale and multiple location estimator. We call our new wavelet estimators: local constant wavelet 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 local constant wavelet 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 local constant wavelet 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 local constant wavelet estimator based on more than one scale and more than one location.

All the local constant wavelet estimators of the 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 either one wavelet coefficient $\widehat{\delta}_1^{LC-SS}$, or wavelet coefficients from a single scale and more than one location $\widehat{\delta}_1^{LC-SM}$.

The rest of this chapter is organized as follows. In Section 2 (page 10), we establish conditions under which the LATE is identified in (I.3) and conditions under which the auxiliary regressions hold for both discontinuous and kink assignment mechanisms. Section 3 (page 17) presents our first wavelet estimators of the 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 local constant wavelet estimators of the LATE for discontinuous assignment mechanisms and establish their asymptotic distributions in Section 4 (page 27). For kink incentive assignment mechanisms, we propose and establish the asymptotic distribution for two types of the single-scale local constant wavelet estimator. Section 5 (page 38) presents results from a Monte Carlo simulation study investigating the finite sample performance of our wavelet estimators. The final section (page 41) concludes the chapter and outlines some future research. Technical proofs are relegated to Appendix A.

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 the 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 have a long history in statistics. In fact, all three approaches (NW, partial linear and local polynomial kernel estimators) in existing work on estimating the 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 constructed estimators of both jump and kink sizes and established their asymptotic distributions for random design samples.[3] In fact, Muller (1992) employed boundary kernels to overcome the well-known boundary problem associated with standard kernel estimators. Second, for the partial linear estimators, Eubank and Whitney (1989) proposed a partial spline estimator of the kink size and established 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) proposed a partial linear (kernel) estimator[4] of the kink size and established its asymptotic distribution, and Cline, et al. (1995) extended 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 kernel estimators, we refer the reader to Loader (1996), Qiu and 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.

Identification and Auxiliary Regressions

[3] Delgado and Hidalgo (2000) extended estimators of Muller (1992) to time series models.

[4] Shiaua and Wahba (1988) and Eubank and Speckman (1994) contrasted the partial spline and partial linear (kernel) methods under various smoothness conditions.

There are two parts in this section for discontinuous and kink incentive assignment mechanisms respectively. In each part, we first provide conditions under which the LATE is identified in (I.3) and then establish auxiliary regressions that will be used to estimate the identified LATE in Sections 3 (page 17) and 4 (page 27).

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 chapter, 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 $B = b$.

Discontinuous Incentive Assignment Mechanism

Identification

The following conditions will be used to prove identification of the 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 the continuity of the regression functions $E(Y_0|V = v)$ and $E(Y_1|V = v)$ at v_0. In addition, compared with the identification results in Hahn, Todd, and van der Klaauw (2001), Theorem 1 below does not require any local independence assumption. Let $\Delta = Y_1 - Y_0$.

Theorem 1 *Under Conditions D1-D5, we have*

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

Theorem 1 implies that in models (I.1) and (I.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 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 $\lim_{v\downarrow v_0} E(Y|V=v) - \lim_{v\uparrow v_0} E(Y|V=v)$.

Auxiliary Regressions

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

$$Y = g(V) + \delta_0 I\{V \geq v_0\} + \varepsilon, \tag{I.4}$$

$$D = h(V) + \zeta_0 I\{V \geq v_0\} + \epsilon, \tag{I.5}$$

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

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

Unlike Porter (2003) and Imbens and Kalyanaramang (2009) who directly assume the continuity of g and h, we impose sufficient conditions on the structural parameters in (I.1) and (I.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 1 *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 1 that under Conditions D1-D5, the functions $g(\cdot)$ and $h(\cdot)$ are only point-wise continuous at v_0. Thus for the LATE estimator $\widehat{\theta}^{LC-SS}$ introduced in Section 3 (page 17), it is still valid that $g(\cdot)$ and $h(\cdot)$ are only point-wise continuous at v_0.

Kink Incentive Assignment Mechanism

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) $\sup_v |\frac{\partial f_{U|V}(u|v)}{\partial v}| < \infty$ and $\int \sup_v |\frac{\partial f_{U|V}(u|v)}{\partial v}| du < \infty$;

(iii) $\int \sup_v |\frac{\partial f_{W,U|V}(w,u|v)}{\partial v}| du < \infty$ 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. Also notice that when U is degenerated in $D = I\{b(V) \geq 0\}$, but under Condition K4 we end up with the discontinuous incentive assignment mechanism instead of the kink incentive assignment mechanism.

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 *Under Conditions K1-K5, we have*

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

Auxiliary Regressions

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

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

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

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

$$\delta_1 = \lim_{v \downarrow v_0} dE(Y|V = v)/dv - \lim_{v \uparrow v_0} dE(Y|V = v)/dv,$$

$$\zeta_1 = \lim_{v \downarrow v_0} P'(v) - \lim_{v \uparrow v_0} P'(v).$$

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 *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 $\mathcal{V}$.*

The First Wavelet Estimator

Let θ denote the identified the LATE parameter. In this and the next sections, we propose local constant wavelet estimators of the LATE for both discontinuous and kink incentive assignment mechanisms. Throughout this and the next sections, we assume the conditions of Propositions 1 and 2 hold respectively for discontinuous and kink incentive assignment mechanisms and a random sample (V_i, Y_i, D_i), $i = 1, ..., n$, is available.

Discontinuous Incentive Assignment Mechanism

Under discontinuous incentive assignment mechanism, the LATE is identified as $\theta = \delta_0/\zeta_0$, where δ_0 and ζ_0 are respectively the parameters in the auxiliary regressions (I.4) and (I.5). Since the idea underlying the estimation of δ_0 and ζ_0 is the same, we focus on the estimation of δ_0.

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 $\widehat{\delta}_0^{LC-SS}$, we consider the auxiliary regression in the wavelet domain. Let $\widehat{F}_V(\cdot)$ denote the empirical distribution function[5] of $\{V_i\}_{i=1}^n$. Then the induced order statistics $\left\{Y_{[i:n]}\right\}_{i=1}^n$ satisfy:

$$\begin{aligned} Y_{[i:n]} &= g(V_{i:n}) + \delta_0 I\{V_{i:n} \geq v_0\} + \varepsilon_{[i:n]} \\ &= g(\widehat{F}_V^{-1}(t_i)) + \delta_0 I\{t_i \geq \widehat{F}_V(v_0)\} + \varepsilon_{[i:n]} \\ &\equiv G(t_i) + \delta_0 I\{t_i \geq \widehat{\tau}\} + e_i, \end{aligned}$$

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

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

[5] 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.

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 resolution level 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:

$$\begin{aligned} \widehat{\Delta}_{j_0}^{Y}(\tau) &= \widehat{\Delta}_{j_0}^{G}(\tau) + \delta_0 \widehat{\Delta}_{j_0}^{\widehat{D_0}}(\tau) + \widehat{\Delta}_{j_0}^{e}(\tau) & \text{(I.8)} \\ &\approx \delta_0 \cdot \widehat{\Delta}_{j_0}^{\widehat{D_0}}(\tau), & \text{(I.9)} \end{aligned}$$

where $\widehat{D_0}(t_i) = I\{t_i \geq \widehat{\tau}\}$.

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 resolution level j_0. When the resolution level 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 some large j_0 motivating our first wavelet estimator $\widehat{\delta}_0^{LC-SS}$:

$$\widehat{\delta}_0^{LC-SS} = \frac{\widehat{\Delta}_{j_0}^{Y}(\widehat{\tau})}{\widehat{\Delta}_{j_0}^{\widehat{D_0}}(\widehat{\tau})}.$$

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

$$\bar{\delta}_0^{LC-SS} = \frac{2^{j_0/2} \widehat{\Delta}_{j_0}^{Y}(\widehat{\tau})}{\int_0^b \psi(u) du}.$$

To establish asymptotic properties of $\widehat{\delta}_0^{LC-SS}$, we adopt the following assumptions.

We note here that the ψ function needs to satisfy assumption A4 only. 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) $H(t)$ is l_H times continuously differentiable for $t \in (0,1)\backslash\{\tau\}$, and $H(\cdot)$ 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 $H(t)$ 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). $\sigma_{\varepsilon\epsilon}(v) \equiv E(\varepsilon_i\epsilon_i|V_i = v)$ is continuous at $v \neq v_0$ and its right and left-hand limits at v_0 exist.

Assumption A4. (a) The function $\psi(\cdot)$ is continuous with compact support $[a, b]$, where $a < 0 < b$ and m vanishing moments, i.e., $\int_a^b u^j \psi(u)\, du = 0$ for $j = 0, 1, ..., m-1$; (b) $\int_0^b \psi(u)du \neq 0$, $\int_a^b u^m \psi(u)\, du \neq 0$, and $\int_a^b |u^m \psi(u)|\, du < \infty$; (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) assume 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 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[6]. Assumption A5 imposes conditions on the scale level j_0.

Theorem 3 *Suppose A1, A3(G), and A4 hold.*

(i) When A2(G) (a) and A5 (a) hold, we obtain: $\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_0^{LC-SS} - \delta_0)$ *and* $\sqrt{\frac{n}{2^{j_0}}}(\bar{\delta}_0^{LC-SS} - \delta_0)$ *have the same asymptotic distribution and*

$$\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_0^{LC-SS} - \delta_0) \xrightarrow{d} N\left(C_a B_a, V\right),$$

[6]See Assumption $C_{1,s}$ in Cheng and Raimondo (2008). And both of them could be treated as 'equivalent wavelets', as opposed to 'equivalent kernels'.

where

$$B_a = \frac{\left[G_+^{(1)}(\tau) - G_-^{(1)}(\tau)\right] \int_0^b \psi(u) u du}{\int_0^b \psi(u) du},$$

$$V = \frac{\sigma_{\varepsilon+}^2(v_0) \int_0^b \psi^2(u) du + \sigma_{\varepsilon-}^2(v_0) \int_a^0 \psi^2(u) du}{\left(\int_0^b \psi(u) du\right)^2}.$$

(ii) When A2(G) (b) and A5 (b) hold, we obtain: $\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_0^{LC-SS} - \delta_0)$ *and* $\sqrt{\frac{n}{2^{j_0}}}(\overline{\delta}_0^{LC-SS} - \delta_0)$ *have the same asymptotic distribution and*

$$\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_0^{LC-SS} - \delta_0) \xrightarrow{d} N\left(C_b B_b, V\right),$$

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 $\overline{\delta}_0^{LC-SS}$ given in Theorem 3 (ii) reduces to that in Park and Kim (2006). Theorem 3 (i) reveals a similar asymptotic behavior of $\widehat{\delta}_0^{LC-SS}$ to the Nadaraya-Watson kernel estimator in Porter (2003) under A2(G) (a). However, as revealed in Theorem 3 (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 $\widehat{\delta}_0^{LC-SS}$. In particular, it reduces the order of the asymptotic bias of $\widehat{\delta}_0^{LC-SS}$ from 2^{-j_0} to 2^{-mj_0}. Thus in terms of asymptotic bias, $\widehat{\delta}_0^{LC-SS}$ behaves more like the partial linear estimator in Porter (2003). This is not surprising, given their partial linear estimator of the jump size is asymptotically equivalent to $\overline{\delta}_0^{LC-SS}$ with a specific ψ function (more details in Chapter 2). Since transforming a random design to an equally spaced design before applying nonparametric method leads to better finite sample performance (Hall, et al., 1998), $\widehat{\delta}_0^{LC-SS}$ inherits such nice property. In addition this also leads to the estimator $\widehat{\delta}_0^{LC-SS}$ to be design-adaptive: the asymptotic bias and variance of our estimators do not depend on the density $f_V(v)$.

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 Raimondo (1998). Under standard regularity conditions, the estimators in Raimondo (1998) coverage at rates faster than $n^{-1/2}$ so the conclusions in Theorem 3 and in all other theorems in this chapter 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 $\theta \equiv \delta_0/\zeta_0$ is defined as $\widehat{\theta}^{LC-SS} = \widehat{\delta}_0^{LC-SS}/\widehat{\zeta}_0^{LC-SS}$, where

$$\widehat{\zeta}_0^{LC-SS} = \frac{\widehat{\Delta}_{j_0}^{D}\left(\widehat{\tau}\right)}{\widehat{\Delta}_{j_0}^{\widehat{D_0}}\left(\widehat{\tau}\right)}.$$

For simplicity, we have used the same mother wavelet $\psi(\cdot)$ and scale level j_0 to estimate δ_0 and ζ_0. This can be relaxed at the expense of more tedious derivations.

Theorem 4 *Suppose A1, A3, and A4 hold.*

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

$$(n/2^{j_0})^{1/2}\left(\frac{\widehat{\delta}_0^{LC-SS}}{\widehat{\zeta}_0^{LC-SS}} - \frac{\delta_0}{\zeta_0}\right) \xrightarrow{d} N\left(\frac{1}{\zeta}C_a\left[B_a - \frac{\delta}{\zeta}B_a^D\right], \frac{1}{\zeta^2}\left[V - \frac{2\delta}{\zeta}V^{YD} + \frac{\delta^2}{\zeta^2}V^D\right]\right),$$

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:

$$(n/2^{j_0})^{1/2}\left(\frac{\widehat{\delta}_0^{LC-SS}}{\widehat{\zeta}_0^{LC-SS}} - \frac{\delta_0}{\zeta_0}\right) \xrightarrow{d} N\left(\frac{1}{\zeta}C_b\left[B_b - \frac{\delta}{\zeta}B_b^D\right], \frac{1}{\zeta^2}\left[V - \frac{2\delta}{\zeta}V^{YD} + \frac{\delta^2}{\zeta^2}V^D\right]\right),$$

where

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

Kink Incentive Assignment Mechanism

For a kink incentive assignment mechanism, the auxiliary regressions are given in (I.6) and (I.7). Again we focus on the estimation of δ_1. First we note that the induced order statistics $\{Y_{[i:n]}\}_{i=1}^n$ satisfy:

$$\begin{aligned} Y_{[i:n]} &= g_K\left(V_{i:n}\right) + \delta_1(V_{i:n} - v_0)I\{V_{i:n} \geq v_0\} + \varepsilon_{K,[i:n]} \\ &= g_K\left(\widehat{F}_V^{-1}\left(t_i\right)\right) + \delta_1(\widehat{F}_V^{-1}\left(t_i\right) - v_0)I\{t_i \geq \widehat{\tau}\} + \varepsilon_{K,[i:n]}. \end{aligned}$$

Similar to the discontinuous incentive assignment mechanism case, we propose the following estimator of δ_1:

$$\widehat{\delta}_1^{LC-SS} = \frac{\widehat{\Delta}_{j_0}^{Y}\left(\widehat{\tau}\right)}{\widehat{\Delta}_{j_0}^{\widehat{D_1}}\left(\widehat{\tau}\right)},$$

where $\widehat{D_1}(t_i) = \left[\widehat{F}_V^{-1}(t_i) - v_0\right] I\{t_i \geq \widehat{\tau}\}$. We will show that under conditions stated below, $\widehat{\delta}_1^{LC-SS}$ has the same asymptotic distribution as

$$\overline{\delta}_1^{LC-SS} = \frac{1}{n}\sum_{i=1}^{n} \frac{2^{j_0}\psi\left[2^{j_0}(\frac{i}{n} - \tau)\right]}{\int_0^b \left[F_V^{-1}(\frac{u}{2^{j_0}} + \tau) - v_0\right]\psi(u)du} Y_{[i:n]}.$$

Like the discontinuous incentive assignment mechanism case, $\widehat{\tau}$, the estimate of the kink location, could follow 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) $H_K(t)$ is $l_H + 1$ times continuously differentiable for $t \in (0,1) \backslash \{\tau\}$, and $H_K(\cdot)$ 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 $H_K(t)$ 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 5 *Suppose A1, A3(G) for ε_K, and A4K hold.*

(i) When A2K(G) (a), A5K (a) and A6K (a) hold, we obtain: $\sqrt{\frac{n}{2^{3j_0}}}(\widehat{\delta}_1^{LC-SS} - \delta_1)$ *and* $\sqrt{\frac{n}{2^{3j_0}}}(\overline{\delta}_1^{LC-SS} - \delta_1)$ *have the same asymptotic distribution and*

$$\sqrt{\frac{n}{2^{3j_0}}}(\widehat{\delta}_1^{LC-SS} - \delta_1) \xrightarrow{d} N\left(C_{Ka}B_{Ka}, V_K\right),$$

where

$$B_{Ka} = \frac{\left[G_{K+}^{(2)}(\tau) - G_{K-}^{(2)}(\tau)\right]\int_0^b u^2\psi(u)du}{2\int_0^b u\psi(u)du} f_V(v_0),$$

$$V_K = \frac{f_V^2(v_0)\left[\sigma_{\varepsilon+}^2(v_0)\int_0^b \psi^2(u)du + \sigma_{\varepsilon-}^2(v_0)\int_a^0 \psi^2(u)du\right]}{\left(\int_0^b u\psi(u)du\right)^2}.$$

(ii) When A2K(G) (b), A5K (b) and A6K (b) hold, we obtain: $\sqrt{\frac{n}{2^{3j_0}}}(\widehat{\delta}_1^{LC-SS} - \delta_1)$ *and* $\sqrt{\frac{n}{2^{3j_0}}}(\overline{\delta}_1^{LC-SS} - \delta_1)$ *have the same asymptotic distribution and*

$$\sqrt{\frac{n}{2^{3j_0}}}(\widehat{\delta}_1^{LC-SS} - \delta_1) \xrightarrow{d} N\left(C_{Kb}B_{Kb}, V_K\right),$$

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 and 5, we observe the same qualitative behavior of $\widehat{\delta}_0^{LC-SS}$ and $\widehat{\delta}_1^{LC-SS}$ in terms of the order of their asymptotic bias: the order of the asymptotic bias of $\widehat{\delta}_1^{LC-SS}$ 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 $\widehat{\delta}_1^{LC-SS}/\widehat{\zeta}_1^{LC-SS}$, where $\widehat{\zeta}_1^{LC-SS}=\widehat{\Delta}_{j_0}^{D}(\widehat{\tau})/\widehat{\Delta}_{j_0}^{\widehat{D_1}}(\widehat{\tau})$.

Theorem 6 *Suppose A1, A3 for ε_K and ϵ_K, and A4K hold.*
(i) When A2K (a), A5K (a) and A6K(a) hold, we obtain:

$$\sqrt{\frac{n}{2^{3j_0}}}\left(\frac{\widehat{\delta}_1^{LC-SS}}{\widehat{\zeta}_1^{LC-SS}}-\frac{\delta_1}{\zeta_1}\right)\xrightarrow{d}N\left(\frac{1}{\zeta}C_{Ka}\left[B_{Ka}-\frac{\delta}{\zeta}B_{Ka}^{D}\right],\frac{1}{\zeta^2}\left[V_K-\frac{2\delta}{\zeta}V_K^{YD}+\frac{\delta^2}{\zeta^2}V_K^{D}\right]\right),$$

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:

$$\sqrt{\frac{n}{2^{3j_0}}}\left(\frac{\widehat{\delta}_1^{LC-SS}}{\widehat{\zeta}_1^{LC-SS}}-\frac{\delta_1}{\zeta_1}\right)\xrightarrow{d}N\left(\frac{1}{\zeta}C_{Kb}\left[B_{Kb}-\frac{\delta}{\zeta}B_{Kb}^{D}\right],\frac{1}{\zeta^2}\left[V_K-\frac{2\delta}{\zeta}V_K^{YD}+\frac{\delta^2}{\zeta^2}V_K^{D}\right]\right),$$

where

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

Local Constant Wavelet Estimators

Discontinuous Incentive Assignment Mechanism

Note that the wavelet estimators $\widehat{\delta}_0^{LC-SS}$ and $\overline{\delta}_0^{LC-SS}$ make use of one wavelet coefficient of $\left\{Y_{[i:n]}\right\}_{i=1}^n$ only, the one at location τ and resolution level 2^{-j_0}. Heuristically wavelet coefficients of $\left\{Y_{[i:n]}\right\}_{i=1}^n$ at locations near τ and other fine resolution levels contain information about δ as well. Formally, it follows from

$$Y_{[i:n]} = G\left(t_i\right) + \delta_0 I\{t_i \geq \widehat{F}_V\left(v_0\right)\} + \left\{\left[g(\widehat{F}_V^{-1}\left(t_i\right)) - G\left(t_i\right)\right] + \varepsilon_{[i:n]}\right\},$$

that

$$\begin{aligned} \widehat{\Delta}_j^Y\left(t\right) &= \widehat{\Delta}_j^G\left(t\right) + \delta_0 \widehat{\Delta}_j^{\widehat{D_0}}\left(t\right) + \widehat{\Delta}_j^e\left(t\right) && \text{(I.10)} \\ &\approx \delta_0 \cdot \widehat{\Delta}_j^{\widehat{D_0}}\left(t\right) + \widehat{\Delta}_j^e\left(t\right), \text{ for all } j \text{ and } t \in [0,1], && \text{(I.11)} \end{aligned}$$

where $\widehat{\Delta}_j^A\left(t\right)$ denotes the wavelet coefficient of $\{A_i\}_{i=1}^n$ at location t and scale level 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 approximately linear regression (I.10) in the wavelet domain suggests that provided $G\left(t\right)$ is continuous, all the wavelet coefficients $\widehat{\Delta}_j^Y\left(t\right)$ at large enough resolution levels j and locations t near τ should contain information on δ_0. This motivates us to propose the following general class of local constant wavelet estimators of δ_0:

$$\widehat{\delta}_0^{LC-MM} = \frac{\sum_{j=j_L}^{j_U}\int_0^1 \widehat{\Delta}_j^Y\left(t\right)\widehat{\Delta}_j^{\widehat{D_0}}\left(t\right)\ \widehat{I}_j(t)dt}{\sum_{j=j_L}^{j_U}\int_0^1 \left[\widehat{\Delta}_j^{\widehat{D_0}}\left(t\right)\right]^2\ \widehat{I}_j(t)dt}, \tag{I.12}$$

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

The class of estimators in (I.12) 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 the LATE.

Single-scale with many locations

In this part, we consider the asymptotic properties of a subclass of local constant wavelet estimators for which only one resolution level is used. The fixed level j_0 local constant wavelet estimator of δ_0 is defined as:

$$\widehat{\delta}_0^{LC-SM} = \frac{\int_0^1 \widehat{\Delta}_{j_0}^{Y}(t)\widehat{\Delta}_{j_0}^{\widehat{D_0}}(t)\widehat{I}_{j_0}(t)dt}{\int_0^1 \left[\widehat{\Delta}_{j_0}^{\widehat{D_0}}(t)\right]^2 \widehat{I}_{j_0}(t)dt}.$$

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

Theorem 7 *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:

$$\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_0^{LC-SM} - \delta_0) \xrightarrow{d} N\left(C_a B_{W1}^a, V_{W1}\right),$$

where

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

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

$$\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_0^{LC-SM} - \delta_0) \xrightarrow{d} N\left(C_{W1}^b B_{W1}^b, V_{W1}\right),$$

where

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

Theorems 3 and 7 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 ($\widehat{\delta}_0^{LC-SM}$) reduces the order of the asymptotic bias of the wavelet estimator further to $O\left(\left(2^{-j_0}\right)^{2m-1}\right)$ from $O\left(\left(2^{-j_0}\right)^m\right)$ for $\widehat{\delta}_0^{LC-SS}$. However, when only A2(G)(a) holds, the order of the asymptotic bias of $\widehat{\delta}_0^{LC-SM}$ remains the same as that of $\widehat{\delta}_0^{LC-SS}$.

To estimate the LATE, we estimate ζ_0 by:

$$\widehat{\zeta}_0^{LC-SM} = \frac{\int_0^1 \widehat{\Delta}_{j_0}^D(t)\widehat{\Delta}_{j_0}^{\widehat{D_0}}(t)\widehat{I}_{j_0}(t)dt}{\int_0^1 \left[\widehat{\Delta}_{j_0}^{\widehat{D_0}}(t)\right]^2 \widehat{I}_{j_0}(t)dt}.$$

The LATE is estimated by $\widehat{\theta}^{LC-SM} = \widehat{\delta}_0^{LC-SM}/\widehat{\zeta}_0^{LC-SM}$.

Theorem 8 *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:

$$(n/2^{j_0})^{1/2}\left(\frac{\widehat{\delta}_0^{LC-SM}}{\widehat{\zeta}_0^{LC-SM}} - \frac{\delta_0}{\zeta_0}\right) \xrightarrow{d} N\left(\frac{1}{\zeta}C_a\left[B_{W1}^a - \frac{\delta}{\zeta}B_{aW1}^D\right], \frac{1}{\zeta^2}\left[V_{W1} - 2\frac{\delta}{\zeta}V_{W1}^{YD} + \frac{\delta^2}{\zeta^2}V_{W1}^D\right]\right),$$

where

$$B_{aW1}^D = \frac{\left[H_+^{(1)}(\tau) - H_-^{(1)}(\tau)\right]\int_a^b\int_a^b L(t)\psi(s)(s-t)I\{s-t \geq 0\}dsdt}{\int_{a-b}^0 M(v)dv},$$

$$V_{W1}^D = \frac{\sigma_{\epsilon+}^2(v_0)\int_0^{b-a} M^2(v)dv + \sigma_{\epsilon-}^2(v_0)\int_{a-b}^0 M^2(v)dv}{\left[\int_{a-b}^0 M(v)dv\right]^2},$$

$$V_{W1}^{YD} = \frac{\sigma_{\varepsilon\epsilon+}^2(v_0)\int_0^{b-a} M^2(v)dv + \sigma_{\varepsilon\epsilon-}^2(v_0)\int_{a-b}^0 M^2(v)dv}{\left[\int_{a-b}^0 M(v)dv\right]^2}.$$

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

$$(n/2^{j_0})^{1/2}\left(\frac{\widehat{\delta}_0^{LC-SM}}{\widehat{\zeta}_0^{LC-SM}} - \frac{\delta_0}{\zeta_0}\right)$$
$$\xrightarrow{d} N\left(\frac{1}{\zeta}\left[C_{W1}^b B_{W1}^b - \frac{\delta}{\zeta}C_{W1}^b B_{bW1}^D\right], \frac{1}{\zeta^2}\left[V_{W1} - \frac{2\delta}{\zeta}V_{W1}^{YD} + \frac{\delta^2}{\zeta^2}V_{W1}^D\right]\right),$$

where

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

Multiscale with a single location

We now investigate the role of using more than one scale in estimating the LATE parameter.[7] First we consider the estimator of δ_0:

$$\widehat{\delta}_0^{LC-MS} = \frac{\sum_{j=j_L}^{j_U} \widehat{\Delta}_j^Y(\widehat{\tau})\widehat{\Delta}_j^{\widehat{D_0}}(\widehat{\tau})}{\sum_{j=j_L}^{j_U}\left[\widehat{\Delta}_j^{\widehat{D_0}}(\widehat{\tau})\right]^2},$$

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

Theorem 9 *Suppose A1, A3(G), and A4 hold. In addition, suppose* $\sigma_{\varepsilon+}^2(v_0) = \sigma_{\varepsilon-}^2(v_0)$.[8]
(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*

$$\sqrt{\frac{n}{2^{j_L}}}(\widehat{\delta}_0^{LC-MS} - \delta_0) \xrightarrow{d} N\left(C_a B_{W2}^a, \frac{V}{2\left[1-(\frac{1}{2})^{\lim K_n+1}\right]}\right),$$

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*

$$\sqrt{\frac{n}{2^{j_L}}}(\widehat{\delta}_0^{LC-MS} - \delta_0) \xrightarrow{d} N\left(\frac{2}{3}C_a B_a, \frac{V}{2}\right),$$

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

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

[8] 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 A.

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

$$\sqrt{\frac{n}{2^{j_L}}}(\widehat{\delta}_0^{LC-MS} - \delta_0) \xrightarrow{d} N\left(C_b B_{W2}^b, \frac{V}{2\left[1-(\frac{1}{2})^{\lim K_n+1}\right]}\right),$$

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*

$$\sqrt{\frac{n}{2^{j_L}}}(\widehat{\delta}_0^{LC-MS} - \delta_0) \xrightarrow{d} N\left(C_b \frac{1}{2\left[1-(\frac{1}{2})^{m+1}\right]} B_b, \frac{V}{2}\right).$$

Theorems 3, 7, and 9 reveal the interesting effects of using additional information in wavelet coefficients at multiple scales and multiple locations. First, the multiple locations estimator $\widehat{\delta}_0^{LC-SM}$ reduces the order of the asymptotic bias of $\widehat{\delta}_0^{LC-SS}$ under A2(G)(b) with $l_G \geq 2m$; Second, the multiple scales estimator $\widehat{\delta}_0^{LC-MS}$ reduces the asymptotic bias and variance of the wavelet estimator $\widehat{\delta}_0^{LC-SS}$ only proportionally, but under both A2(G)(a) and A2(G)(b).

To estimate the LATE parameter, we let

$$\widehat{\zeta}_0^{LC-MS} = \frac{\sum_{j=j_L}^{j_U} \widehat{\Delta}_j^D(\tau)\widehat{\Delta}_j^{\widehat{D_0}}(\tau)}{\sum_{j=j_L}^{j_U}\left[\widehat{\Delta}_j^{\widehat{D_0}}(\tau)\right]^2}.$$

Theorem 10 *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* $\sigma_{\varepsilon\epsilon+}^2(v_0) = \sigma_{\varepsilon\epsilon-}^2(v_0)$*.*
(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*

$$(n/2^{j_L})^{1/2}\left(\frac{\widehat{\delta}_0^{LC-MS}}{\widehat{\zeta}_0^{LC-MS}} - \frac{\delta_0}{\zeta_0}\right) \xrightarrow{d} N\left(\frac{1}{\zeta}C_a\left[B_{W2}^a - \frac{\delta}{\zeta}B_{W2}^{Da}\right], \frac{V - \frac{2\delta}{\zeta}V^{YD} + \frac{\delta^2}{\zeta^2}V^D}{2\left[1-(\frac{1}{2})^{\lim K_n+1}\right]\zeta^2}\right),$$

where

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

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

$$(n/2^{j_L})^{1/2}\left(\frac{\widehat{\delta}_0^{LC-MS}}{\widehat{\zeta}_0^{LC-MS}} - \frac{\delta_0}{\zeta_0}\right) \xrightarrow{d} N\left(\frac{2}{3\zeta}C_a\left[B_a - \frac{\delta}{\zeta}B_a^D\right], \frac{V - \frac{2\delta}{\zeta}V^{YD} + \frac{\delta^2}{\zeta^2}V^D}{2\zeta^2}\right).$$

(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*

$$(n/2^{j_L})^{1/2}\left(\frac{\widehat{\delta}_0^{LC-MS}}{\widehat{\zeta}_0^{LC-MS}} - \frac{\delta_0}{\zeta_0}\right)$$
$$\xrightarrow{d} N\left(\frac{1}{\zeta}C_b\left[B_{W2}^b - \frac{\delta}{\zeta}B_{W2}^{Db}\right], \frac{V - \frac{2\delta}{\zeta}V^{YD} + \frac{\delta^2}{\zeta^2}V^D}{2\left[1-(\frac{1}{2})^{\lim K_n+1}\right]\zeta^2}\right),$$

where

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

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

$$(n/2^{j_L})^{1/2}\left(\frac{\widehat{\delta}_0^{LC-MS}}{\widehat{\zeta}_0^{LC-MS}} - \frac{\delta_0}{\zeta_0}\right)$$
$$\xrightarrow{d} N\left(C_b\frac{1}{2\zeta\left[1-(\frac{1}{2})^{m+1}\right]}\left[B_b - \frac{\delta}{\zeta}B_b^D\right], \frac{V - \frac{2\delta}{\zeta}V^{YD} + \frac{\delta^2}{\zeta^2}V^D}{2\zeta^2}\right).$$

Multiscale with many locations

We now establish the asymptotic distribution of the general estimator $\widehat{\delta}_0^{LC-MM}$:

$$\widehat{\delta}_0^{LC-MM} = \frac{\sum_{j=j_L}^{j_U}\int_0^1 \widehat{\Delta}_j^Y(t)\,\widehat{\Delta}_j^{\widehat{D_0}}(t)\,\widehat{I}_{j0}(t)dt}{\sum_{j=j_L}^{j_U}\int_0^1\left[\widehat{\Delta}_j^{\widehat{D_0}}(t)\right]^2\widehat{I}_{j0}(t)dt}.$$

Again for notational compactness, we only establish the asymptotic distribution of $\widehat{\delta}_0^{LC-MM}$ under the condition that $\sigma_{\varepsilon+}^2(v_0) = \sigma_{\varepsilon-}^2(v_0)$.

Theorem 11 *Suppose A1, A3, and A4 hold. In addition, suppose $\sigma_{\varepsilon+}^2(v_0) = \sigma_{\varepsilon-}^2(v_0)$.*
(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*

$$\sqrt{\frac{n}{2^{j_L}}}(\widehat{\delta}_0^{LC-MM} - \delta_0) \xrightarrow{d} N\left(C_aB_W^a, V_W\right),$$

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*

$$
\sqrt{\frac{n}{2^{j_L}}}(\widehat{\delta}_0^{LC-MM} - \delta_0) \xrightarrow{d} N\left(C_a B_W^{a*}, V_W^*\right),
$$

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*

$$
\sqrt{\frac{n}{2^{j_L}}}(\widehat{\delta}_0^{LC-MM} - \delta_0) \xrightarrow{d} N\left(C_{W1}^b B_W^b, V_W\right),
$$

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*

$$
\sqrt{\frac{n}{2^{j_L}}}(\widehat{\delta}_0^{LC-MM} - \delta_0) \xrightarrow{d} N\left(C_{W1}^b B_W^{b*}, V_W^*\right),
$$

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, $\widehat{\delta}_0^{LC-MM}$ inherit the properties of both $\widehat{\delta}_0^{LC-SM}$ and $\widehat{\delta}_0^{LC-MS}$: when A2(G)(b) holds with $l_G \geq 2m$, it reduces the order of the asymptotic bias of $\widehat{\delta}_0^{LC-SS}$ or $\widehat{\delta}_0^{LC-MS}$ reflecting the additional information in the multiple locations used in $\widehat{\delta}_0^{LC-MM}$; and under both A2(G)(a) and A2(G)(b), the asymptotic bias and variance of $\widehat{\delta}_0^{LC-MM}$ are proportionally smaller than those of $\widehat{\delta}_0^{LC-SM}$ reflecting the additional information in the multiple scales used in $\widehat{\delta}_0^{LC-MM}$.

Define

$$\widehat{\zeta}_0^{LC-MM} = \frac{\sum_{j=j_L}^{j_U} \int_0^1 \widehat{\Delta}_j^D(t)\, \widehat{\Delta}_j^{\widehat{D_0}}(t)\, \widehat{I}_{j_0}(t)dt}{\sum_{j=j_L}^{j_U} \int_0^1 \left[\widehat{\Delta}_j^{\widehat{D_0}}(t)\right]^2 \widehat{I}_{j_0}(t)dt}.$$

Our estimator of the LATE is given by $\widehat{\delta}_0^{LC-MM} / \widehat{\zeta}_0^{LC-MM}$.

Theorem 12 *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 $\sigma^2_{\varepsilon\epsilon+}(v_0) = \sigma^2_{\varepsilon\epsilon-}(v_0)$.*
(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

$$(n/2^{j_L})^{1/2}\left(\frac{\widehat{\delta}_0^{LC-MM}}{\widehat{\zeta}_0^{LC-MM}} - \frac{\delta_0}{\zeta_0}\right) \xrightarrow{d} N\left(\frac{1}{\zeta}C_a\left[B_W^a - \frac{\delta}{\zeta}B_W^{Da}\right], \frac{1}{\zeta^2}\left[V_W - \frac{2\delta}{\zeta}V_W^{YD} + \frac{\delta^2}{\zeta^2}V_W^D\right]\right),$$

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

$$(n/2^{j_L})^{1/2}\left(\frac{\widehat{\delta}_0^{LC-MM}}{\widehat{\zeta}_0^{LC-MM}} - \frac{\delta_0}{\zeta_0}\right) \xrightarrow{d} N\left(\frac{1}{\zeta}C_a\left[B_W^{a*} - \frac{\delta}{\zeta}B_{aW}^{D*}\right], \frac{1}{\zeta^2}\left[V_W^* - 2\frac{\delta}{\zeta}V_W^{YD*} + \frac{\delta^2}{\zeta^2}V_W^{D*}\right]\right),$$

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

$$(n/2^{j_L})^{1/2}\left(\frac{\widehat{\delta}_0^{LC-MM}}{\widehat{\zeta}_0^{LC-MM}} - \frac{\delta_0}{\zeta_0}\right) \xrightarrow{d} N\left(\frac{1}{\zeta}C_a\left[B_W^b - \frac{\delta}{\zeta}B_W^{Db}\right], \frac{1}{\zeta^2}\left[V_W - \frac{2\delta}{\zeta}V_W^{YD} + \frac{\delta^2}{\zeta^2}V_W^D\right]\right),$$

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*

$$(n/2^{j_L})^{1/2}\left(\frac{\widehat{\delta}_0^{LC-MM}}{\widehat{\zeta}_0^{LC-MM}} - \frac{\delta_0}{\zeta_0}\right) \xrightarrow{d} N\left(\frac{1}{\zeta}C_{W1}^{b}\left[B_W^{b*} - \frac{\delta}{\zeta}B_{bW}^{D*}\right], \frac{1}{\zeta^2}\left[V_W^* - 2\frac{\delta}{\zeta}V_W^{YD*} + \frac{\delta^2}{\zeta 2}V_W^{D*}\right]\right),$$

where

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

Kink Incentive Assignment Mechanism

All three local constant wavelet estimators for discontinuous incentive assignment mechanisms proposed in Section 4 (page 27) can be extended to kink incentive assignment mechanisms and they share the same qualitative properties. To illustrate, we present a detailed analysis of the local constant wavelet estimator based on wavelet coefficients at a single scale and many locations and report results on the other estimators in a separate paper.

Let

$$\widehat{\delta}_1^{LC-SM} = \frac{\int_0^1 \widehat{\Delta}_{j_0}^{Y}(t)\widehat{\Delta}_{j_0}^{\widehat{D_1}}(t)\widehat{I}_{j_0}(t)dt}{\int_0^1\left[\widehat{\Delta}_{j_0}^{\widehat{D_1}}(t)\right]^2 \widehat{I}_{j_0}(t)dt}$$

and

$$\overline{\delta}_1^{LC-SM} = \frac{1}{n}\sum_{i=1}^{n} J_{KW1}(\frac{i}{n})Y_{i:n}$$

where

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

We first show that $\widehat{\delta}_1^{LC-SM}$ has the same asymptotic distribution as $\overline{\delta}_1^{LC-SM}$ and then establish the asymptotic distribution of $\overline{\delta}_1^{LC-SM}$.

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 13 *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:

$$\sqrt{\frac{n}{2^{3j_0}}}(\widehat{\delta}_1^{LC-SM} - \delta_1) \xrightarrow{d} N\left(C_{Ka}B_{KW1}^a, V_{KW1}\right),$$

where

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

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

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:

$$\sqrt{\frac{n}{2^{3j_0}}}(\widehat{\delta}_1^{LC-SM} - \delta_1) \xrightarrow{d} N\left(C_{K'W1}^b B_{KW1}^b, V_{KW1}\right),$$

where

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

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: $\widehat{\delta}_1^{LC-SM}/\widehat{\zeta}_1^{LC-SM}$, where

$$\widehat{\zeta}_1^{LC-SM} = \frac{\int_0^1 \widehat{\Delta}^D_{j_0}(t)\widehat{\Delta}^{\widehat{D_1}}_{j_0}(t)\widehat{I}_{j_0}(t)dt}{\int_0^1\left[\widehat{\Delta}^{\widehat{D_1}}_{j_0}(t)\right]^2\widehat{I}_{j_0}(t)dt}.$$

Theorem 14 *Suppose A1, A3, and A4K hold.*
(a) When A2K (a), A5K (a) and A6K(a) hold, we obtain:

$$\begin{aligned}&\sqrt{\frac{n}{2^{3j_0}}}\left(\frac{\widehat{\delta}_1^{LC-SM}}{\widehat{\zeta}_1^{LC-SM}}-\frac{\delta_1}{\zeta_1}\right)\\ &\xrightarrow{d} N\left(\frac{1}{\zeta}C_{Ka}\left[B^a_{KW1}-\frac{\delta}{\zeta}B^{Da}_{KW1}\right],\frac{1}{\zeta^2}\left[V_{KW1}-\frac{2\delta}{\zeta}V^{YD}_{KW1}+\frac{\delta^2}{\zeta^2}V^D_{KW1}\right]\right),\end{aligned}$$

where

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

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

$$\begin{aligned}&\sqrt{\frac{n}{2^{3j_0}}}\left(\frac{\widehat{\delta}_1^{LC-SM}}{\widehat{\zeta}_1^{LC-SM}}-\frac{\delta_1}{\zeta_1}\right)\\ &\xrightarrow{d} N\left(\frac{1}{\zeta}C^b_{K'W1}\left[B^b_{KW1}-\frac{\delta}{\zeta}B^{Db}_{KW1}\right],\frac{1}{\zeta^2}\left[V_{KW1}-\frac{2\delta}{\zeta}V^{YD}_{KW1}+\frac{\delta^2}{\zeta^2}V^D_{KW1}\right]\right),\end{aligned}$$

where

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

in which

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

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 is based on 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). Tedious 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 generated 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 generated 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 a 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.

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 $\widehat{\delta}_0^{LC-SM}$ performs much better than the single location estimator $\widehat{\delta}_0^{LC-SS}$ in terms of bias, standard error, and MSE; Third, many scales estimator $\widehat{\delta}_0^{LC-MS}$ performs better than the single scale estimator $\widehat{\delta}_0^{LC-SS}$, but the reduction in MSE is not as much as that of the many locations estimator $\widehat{\delta}_0^{LC-SM}$ in comparison with the single scale and single location estimator $\widehat{\delta}_0^{LC-SS}$; 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.[9] 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 $\widehat{\delta}_0^{LC-MM}$ performs the best whether A2(G)(a) or A2(G)(b) holds.

Conclusion

In this chapter, we have studied the identification of the LATE in two classes of switching regime models. Both allow for individuals to make decisions 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 the LATE based on which

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

we have presented a systematic treatment of wavelet estimation of the LATE. In addition to making use of the existing wavelet estimator of the jump size or kink size, we have developed local constant wavelet estimators improving upon the existing wavelet estimator by employing more wavelet coefficients. The asymptotic properties of all the estimators are established and their finite sample properties are investigated via a simulation study.

This chapter 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 local constant wavelet estimators proposed in this chapter to allow for unknown cut-off and/or more than one forcing variables.

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

$$Y_1 = g_1(X, V, W), \; Y_0 = g_0(X, V, W), \tag{I.13}$$

$$D = I\{b(V) + g_3(X) - U \geq 0\}. \tag{I.14}$$

Under appropriate conditions, the auxiliary regressions established earlier for estimating the 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 the LATE. This may be done by making use of the alternative auxiliary regressions:

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

where $E[\varepsilon|X,V] = 0$ and $E[\epsilon|X,V] = 0$. Frölich (2007) proposed a local linear kernel estimator taking into account the covariates X and compared it with the local linear estimator without using X. The asymptotic analysis in Frölich (2007) seems to suggest that using the covariates 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 covariates X. We'll leave this to future research.

CHAPTER II

Local Polynomial Wavelet Estimation of the LATE

Introduction

Estimating treatment effect parameters has become routine in empirical work in economics. One such parameter that has recently gained considerable attention is the local average treatment effect (LATE). Numerous semiparametric and nonparametric estimators for the LATE have been proposed in econometrics literature. The main purpose of this chapter is to propose a new and intuitive approach—local polynomial wavelet approach—for estimating the LATE under discontinuous incentive assignment mechanisms. We show that, under a broad set of conditions, the local polynomial wavelet approach is optimal, and could be easily adapted to estimating the LATE under kink incentive assignment mechanisms.

The motivation of our wavelet approach is the characterization of the point-wise smoothness of a deterministic function by its wavelet coefficient; a small (large) wavelet coefficient corresponds to a high (low) smoothness of the function (see Appendix B for a detailed statement). As a result of this point-wise smoothness characterization, wavelet coefficients have been widely used to detect discontinuous locations—see Wang (1995) and Raimondo (1998), and estimate jump size (the difference between right- and left-hand limits)—see Park and Kim (2006) in statistics literature and local constant wavelet estimators in Chapter 1. While Park and Kim (2006) presented a simple estimator using only one pair of the wavelet coefficient at the discontinuous location, local constant wavelet estimators integrated all pairs of wavelet coefficients around the discontinuous location and the resulting estimators had better asymptotic properties. Furthermore, local constant wavelet estimators are applied to the LATE estimations in a class of switching regime models. However, these wavelet estimators failed to take any potential higher-order derivative discontinuities into consideration, and this omission resulted in a sub-optimal convergence rate under the presence of slope or higher-order derivative discontinuities. The local polynomial wavelet approach in this chapter explicitly accounts for potential higher-order derivative discontinuities and fully explores wavelet coefficients generated from both time and frequency domains. In an application for discontinuous (kink) incentive assignment mechanisms, we find that local polynomial wavelet estimators attain the optimal convergence rate of the LATE.

Compared to local polynomial kernel estimators for jump size, local polynomial wavelet estimators have finite unconditional variances in finite samples. It is well-known in the statistics literature that local polynomial kernel estimators suffer from a serious drawback: for compactly supported kernels, the unconditional variance of local polynomial kernel estimators is infinite, so the mean squared error (MSE) and MSE optimal bandwidth are not defined; see Seifert and Gasser (1996). In order to overcome this defect, our estimators equispaced (Hall, et al., 1998) the original data in the first step. Not only could such a transformation guarantee finite unconditional variances for local polynomial wavelet estimators while maintaining the first order property, it also would not degrade features of the function through prior smoothing, especially when there are jumps. Moreover, local polynomial wavelet estimators of jump size adapt to both random and fixed designs, and to both highly clustered and nearly uniform designs in the large sample.

An interesting by-product of the local polynomial wavelet approach is that it could jointly and optimally estimate jump sizes in any order derivatives, such as jump size, kink size (the difference between right- and left-hand first derivative limits), and up to jump sizes in higher-order derivatives. Such joint estimates of jump and kink sizes are informative about the LATE regimes we identified—either discontinuous incentive assignment mechanisms or kink incentive assignment mechanisms.

Next we show that all existing jump size estimators (based on the equispaced data) share a common structure, being members of a class of local polynomial wavelet estimators. Local polynomial kernel estimators are in the framework of local polynomial wavelet estimators, which incorporate potential higher-order derivative discontinuities into the criterion function. On the other hand, the class of local constant wavelet estimators consists of time-frequency domain wavelet estimators by Park and Kim (2006). Also, in the context of the time domain approach, Nadaraya-Watson estimator, partial smoothing kernel estimator by Eubank and Speckman (1994), and profiled partial linear estimator by Porter (2003) could also be asymptotically expressed as local constant wavelet estimators. In general, jump size estimates from local constant wavelet estimators have asymptotic bias orders being inferior to the ones from local polynomial wavelet estimators, unless the slope or higher-order derivatives are continuous.

The outline of the chapter follows. In Section 2 (page 47), local polynomial wavelet estimators under discontinuous incentive assignment mechanisms are described. Section 3 (page 58) summarizes existing jump size estimators as being members of local polynomial wavelet estimators. Section 4 (page 61) provides asymptotic results under kink incentive assignment mechanisms. The proposed methods are examined in Section 5 (page 64), using Monte Carlo simulations. Section 6 (page 67) suggests possible directions for the future research. Proofs of the results are given in Appendix B.

Wavelet Estimators under Discontinuous Incentive Assignment Mechanism

In this section, we first review the LATE identifications by Chapter 1, and then translate the LATE estimate as the ratio of two jump sizes from two auxiliary regressions. Next, we introduce two types of local polynomial wavelet estimators: single-scale versus multiscale, each of which are motivated from the time–frequency representation of wavelet transformations. The asymptotic property of the LATE under single-scale local polynomial wavelet estimator is provided, and we discuss the better asymptotic MSE through multiscale local polynomial wavelet estimators.

Throughout the rest of this chapter, we adopt the following notation: x^α means that x is to the power α, while $x^{(\alpha)}$ is the α-th derivative of x.

The LATE Identification and Auxiliary Regressions

In their seminal 2001 paper, Hahn, et al. identified the LATE under the regression discontinuity design (RDD) with the local independence conditions. On the other hand, Chapter 1 showed that without imposing the local independence conditions, the LATE under the switching regime model could be identified under the presence of a discontinuity in the incentive assignment. Thee results in Chapter 1 generalized Lee (2008) by allowing for general incentive assignment mechanisms, and more importantly, for heterogenous choices among agents being assigned the same incentive.

Let $V \in \mathcal{V} \subset \mathcal{R}$ be a continuous random variable denoting the agent's observable covariate based on which the incentive assignment $b : \mathcal{V} \mapsto \mathcal{R}$ is assigned. Based on the incentive received $b(V)$ and one's 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.

We have

$$Y_1 = g_1(V,W) \text{ and } Y_0 = g_0(V,W), \tag{II.1}$$

$$D = I\{b(V) - U \geq 0\}, \tag{II.2}$$

where U is the individual's unobservable covariate affecting selection, W is a vector of the individual's unobservable covariates affecting potential outcomes, and g_1 and g_0 are unknown real-valued measurable functions. The agent's observable covariate V affects both potential outcomes and selection through the incentive assignment $b(\cdot)$. The function $b(\cdot)$ (could be unknown) is assumed to be either discontinuous at a known cutoff point[1] v_0 (discontinuous incentive assignment mechanisms) or continuous but non-differentiable at v_0 (kink incentive assignment mechanisms). Assume the econometrician observes (V, Y, D), where the individual's realized outcome $Y = DY_1 + (1-D)Y_0$. Let $P(v) = \Pr(D=1|V=v) = E(D|V=v)$. Under discontinuous incentive assignment mechanisms, the LATE, $\lim_{e\downarrow 0} E(Y_1 - Y_0|V = v_0, D(v_0+e) - D(v_0-e) = 1)$, is identified as

$$\frac{\lim_{v\downarrow v_0} E(Y|V=v) - \lim_{v\uparrow v_0} E(Y|V=v)}{\lim_{v\downarrow v_0} P(v) - \lim_{v\uparrow v_0} P(v)}, \tag{II.3}$$

which is the ratio of the jump size in $E(Y|V)$ to the jump size in $P(V)$ at $V = v_0$. Chapter 1 established auxiliary regressions linking the LATE[2] under discontinuous incentive assignment mechanisms to jump sizes δ_0 and ζ_0 in Equation (II.4) and (II.5), which are

$$Y = g(V) + \delta_0 I\{V \geq v_0\} + \varepsilon, \tag{II.4}$$

$$D = h(V) + \zeta_0 I\{V \geq v_0\} + \epsilon, \tag{II.5}$$

[1]For the unknown cutoff point, we could apply Wang (1995) and Raimondo's (1998) wavelet methods to detect discontinuous locations. Under standard regularity conditions, the estimated cutoff points converge at rates faster than $n^{-1/2}$, so that the asymptotic results of local polynomial wavelet estimators in this paper remain valid even under the estimated cutoff point.

[2]See Proposition 1 in Chapter 1.

where $E(\varepsilon|V) = 0,\ E(\epsilon|V) = 0,\ \delta_0 = \lim_{v\downarrow v_0} E(Y|V = v) - \lim_{v\uparrow v_0} E(Y|V = v),\ \zeta_0 = \lim_{v\downarrow v_0} P(v) - \lim_{v\uparrow v_0} P(v)$, and both h and g are continuous on the support of V.

Single-scale Local Polynomial Wavelet Estimator

Under discontinuous incentive assignment mechanisms, we could estimate the LATE by using the estimation of δ_0/ζ_0. Since the idea for estimating δ_0 and ζ_0 is the same, we focus on the estimation of δ_0.

Let $F_V(\cdot)$ and $\widehat{F}_V(\cdot)$ denote the true and empirical distribution functions of V and $\tau = F_V(v_0)$. Denote $V_{1:n} \leq \cdots \leq V_{n:n}$ as the order statistics of $\{V_i\}_{i=1}^n$. Further, let $t_i = i/n$ for $1 \leq i \leq n$. Then, with the induced order statistics $\{Y_{i:n}\}_{i=1}^n$ satisfy

$$\begin{aligned} Y_{i:n} &= g(V_{i:n}) + \delta_0 I\{V_{i:n} \geq v_0\} + \varepsilon_{i:n} \\ &= g(\widehat{F}_V^{-1}(t_i)) + \delta_0 I\{t_i \geq \widehat{F}_V(v_0)\} + \varepsilon_{i:n} \\ &= G(t_i) + \delta_0 I\{t_i \geq \tau\} + e_i, \end{aligned} \tag{II.6}$$

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

$$e_i = \delta_0 \left[I\{t_i \geq \widehat{\tau}\} - I\{t_i \geq \tau\}\right] - \left[g(F_V^{-1}(t_i)) - g(\widehat{F}_V^{-1}(t_i))\right] + \varepsilon_{i:n}.$$

Assumption Set B.

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

Assumption B2.

(G). Let $G(t) = g(F_V^{-1}(t))$. $G(t)$ is p-th continuously differentiable at $t \in (0,1)\backslash\{\tau\}$, and is continuous at $t = \tau$ with finite right- and left-hand derivatives up to the order p.

(H). Let $H(t) = h(F_V^{-1}(t))$. $H(t)$ is q-th continuously differentiable at $t \in (0,1)\backslash\{\tau\}$, and is continuous at $t = \tau$ with finite right- and left-hand derivatives up to the order q.

Assumption B3.

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

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

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

Assumption B4. (a) The real-valued wavelet function $\psi(\cdot)$ is continuous with compact support $[a, b]$, where $a < 0 < b$ and m vanishing moments, i.e., $\int_a^b u^j \psi(u)\, du = 0$ for $j = 0, 1, ..., m-1$; (b) $\int_a^b u^m \psi(u)\, du \neq 0$ and $\int_a^b |u^m \psi(u)|\, du < \infty$; (c) ψ has a bounded derivative and satisfies the admissibility condition that $\int \left|\widehat{\psi}(\xi)\right|^2 / |\xi|\, d\xi < \infty$, where $\widehat{\psi}(\xi)$ is the Fourier transform of $\psi(t)$.

Assumption B5. As $n \to \infty$, $j_0 \to \infty$, $2^{j_0}/n \to 0$ and $\left(1/2^{j_0}\right)^{2m-1} \sqrt{n/2^{j_0}} \to C < \infty$.

Assumption B6. The function $F_V^{-1}(v)$ is continuously differentiable on the support of V.

Assumption B2 allows functions G and H could have the potential slope changes or derivative discontinuities at $t = \tau$ up to p-th and q-th orders, respectively. Assumption B3 allows for the possible heteroskedasticity. Assumptions B4 (a), (b), and (c) specify the class of wavelet functions ψ. Compared to the kernel function, the wavelet function ψ integrates to zero. Examples of wavelet functions ψ include classes of Daubechies wavelet functions and least asymmetric wavelet functions. Assumption B5 provides conditions on the scale parameter j_0, whose role is analogous to the reciprocal of the bandwidth parameter in the kernel estimation.

Based on Assumption B2 (G), we model potential derivative discontinuities up to p-th order for the function G at $t = \tau$

$$G(t_i) = G^*(t_i) + \sum_{k=1}^{p} \delta_k \cdot \left[F_V^{-1}(t_i) - v_0\right]^k I\{t_i \geq \tau\}, \tag{II.7}$$

where $\{\delta_k\}_{k=1}^p$ includes the kink size δ_1, the jump size in the second derivative ($\delta_2/2!$) till the jump size in the p-th derivative ($\delta_p/p!$). Notice that $G^*(\cdot)$ is p-th differentiable on the whole support. Substitute Equation (II.7) into Equation (II.6) and we have

$$Y_{i:n} = G^*(t_i) + \sum_{k=0}^{p} \delta_k \cdot \left[F_V^{-1}(t_i) - v_0\right]^k I\{t_i \geq \tau\} + e_i. \tag{II.8}$$

Let $\widehat{\Delta}_{j_0}^{A}(t)$ denote the wavelet coefficient of $\{A_i\}_{i=1}^n$ at the location $t \in [0,1]$ and a fixed scale[3] j_0, where t and j_0 represent the time and frequency parameters, respectively:

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

Applying the wavelet transformation to both sides of Equation (II.8), we have

$$\widehat{\Delta}_{j_0}^{Y}(t_i) = \widehat{\Delta}_{j_0}^{G^*}(t_i) + \sum_{k=0}^{p} \delta_k \cdot \widehat{\Delta}_{j_0}^{D_k}(t_i) + \widehat{\Delta}_{j_0}^{e}(t_i), \tag{II.9}$$

[3]In the next subsection, we would vary the scale parameter to construct multiscale local polynomial wavelet estimators.

where $D_k(t) = \left[F_V^{-1}(t) - v_0\right]^k I\{t \geq \tau\}$ for $k = 0, 1, \cdots, p$ and

$$\begin{aligned}
\widehat{\Delta}_{j_0}^{Y}(t) &= \frac{2^{j_0/2}}{n} \sum_{i=1}^{n} Y_{i:n} \psi\left[2^{j_0}(t_i - t)\right], \\
\widehat{\Delta}_{j_0}^{D_k}(t) &= \frac{2^{j_0/2}}{n} \sum_{i=1}^{n} D_k(t_i) \psi\left[2^{j_0}(t_i - t)\right], \\
\widehat{\Delta}_{j_0}^{e}(t) &= \frac{2^{j_0/2}}{n} \sum_{i=1}^{n} e_i \psi\left[2^{j_0}(t_i - t)\right].
\end{aligned}$$

It is well-known that the wavelet coefficient $\widehat{\Delta}_{j_0}^{A}(t)$ captures the variation of the sequence $\{A_i\}_{i=1}^n$ at the location parameter t. Heuristically, when the scale j_0 is large, $\widehat{\Delta}_{j_0}^{A}(t)$ is small unless there is a discontinuity in $\{A_i\}_{i=1}^n$ at $i = t$. Since G^* is p-th differentiable on the whole support, its wavelet coefficient is expected to be small. On the other hand, D_k has the k-th derivative discontinuity at $t = \tau$, so wavelet coefficients $\widehat{\Delta}_{j_0}^{D_k}(t)$ would be large in the neighborhood of τ. Combining these arguments[4], Equation (II.9) reduces to

$$\widehat{\Delta}_{j_0}^{Y}(t_i) \approx \sum_{k=0}^{p} \delta_k \cdot \widehat{\Delta}_{j_0}^{D_k}(t_i) + \widehat{\Delta}_{j_0}^{e}(t_i).$$

This motivates single-scale local polynomial wavelet estimator:

$$\widehat{\delta}^{LP-SM} = \arg \min_{\{\delta_k\}_{k=0}^p} \sum_{l=1}^{n} \left[\widehat{\Delta}_{j_0}^{Y}(t_l) - \sum_{k=0}^{p} \delta_k \cdot \widehat{\Delta}_{j_0}^{\widehat{D}_k}(t_l)\right]^2 \widehat{I}_{j_0}(t_l), \tag{II.10}$$

where

$$\begin{aligned}
\widehat{\delta}^{LP-SM} &= \left[\widehat{\delta}_0^{LP-SM}, \widehat{\delta}_1^{LP-SM}, \cdots, \widehat{\delta}_p^{LP-SM}\right], \\
\widehat{D}_k(t) &= \left[\widehat{F}_V^{-1}(t) - v_0\right]^k I\{t \geq \widehat{\tau}\} \text{ for } k = 0, 1, \cdots, p
\end{aligned}$$

and $\widehat{I}_{j_0}(t) = I\{a \leq 2^{j_0}(\widehat{\tau} - t) \leq b\}$, which is the cone of influence defined in Mallat (2009). The qualifier *single-scale* comes from the fact that we are only using wavelet coefficients

[4] A formal explanation would be: within the interval $\{t : a \leq 2^{j_0}(\tau - t) \leq b\}$, $\widehat{\Delta}_{j_0}^{G^*}(t)$ and $\widehat{\Delta}_{j_0}^{D}(t)$ are of orders $2^{(-p-1/2)j_0}$ and $\left\{2^{(-s-1/2)j_0}\right\}_{s=0}^{p-1}$ as j_0 tends to infinity. If we pick up a scale j_0 such that the orders of $\widehat{\Delta}_{j_0}^{G^*}(t)$ and $\widehat{\Delta}_{j_0}^{e}(t)$ are balanced, the $\widehat{\Delta}_{j_0}^{Y}(t)$ would be dominated by $\widehat{\Delta}_{j_0}^{D}(t)$.

$\left\{\widehat{\Delta}_{j_0}^{Y}(t), \widehat{\Delta}_{j_0}^{D}(t) : t \in [0,1]\right\}$ for a given scale j_0, instead of $\left\{\widehat{\Delta}_{j}^{Y}(t), \widehat{\Delta}_{j}^{D}(t) : j \in \mathbb{Z}^{+}, t \in [0,1]\right\}$ in multiscale local polynomial wavelet estimators. The qualifier *polynomial* is to indicate the wavelet transformation of $\left[F_V^{-1}(t) - v_0\right]^k I\{t \geq \tau\}$ for $k = 0, 1, \cdots, p$, and notice that there is no intercept term in Equation (II.10). Single-scale local polynomial wavelet estimator $\widehat{\delta}^{LP-SM}$ has a closed-form expression

$$\widehat{\delta}^{LP-SM} = \left[\left(\widehat{\Delta}_{j_0}^{\widehat{DP}}\right)^T \cdot \widehat{I}_{j_0} \cdot \widehat{\Delta}_{j_0}^{\widehat{DP}}\right]^{-1} \left[\left(\widehat{\Delta}_{j_0}^{\widehat{DP}}\right)^T \cdot \widehat{I}_{j_0} \cdot \widehat{\Delta}_{j_0}^{Y}\right],$$

where

$$\begin{aligned}
\left(\widehat{\Delta}_{j_0}^{\widehat{DP}}\right)^T &= \left[\left(\widehat{\Delta}_{j_0}^{\widehat{DP}}(t_1)\right), \left(\widehat{\Delta}_{j_0}^{\widehat{DP}}(t_2)\right), \cdots, \left(\widehat{\Delta}_{j_0}^{\widehat{DP}}(t_n)\right))\right], \\
\widehat{\Delta}_{j_0}^{\widehat{DP}}(t) &= \left[\widehat{\Delta}_{j_0}^{\widehat{D}_0}(t), \widehat{\Delta}_{j_0}^{\widehat{D}_1}(t), \cdots, \widehat{\Delta}_{j_0}^{\widehat{D}_p}(t)\right]^T, \\
\widehat{I}_{j_0} &= diag\left[\widehat{I}_{j_0}(t_1), \widehat{I}_{j_0}(t_2), \cdots, \widehat{I}_{j_0}(t_n)\right], \\
\left(\widehat{\Delta}_{j_0}^{Y}\right)^T &= \left[\widehat{\Delta}_{j_0}^{Y}(t_1), \widehat{\Delta}_{j_0}^{Y}(t_2), \cdots, \widehat{\Delta}_{j_0}^{Y}(t_n)\right].
\end{aligned}$$

Theorem 15 *Under Assumption Set B and $p \geq 2m$:*

(1) the asymptotic bias of single-scale local polynomial wavelet estimator $\widehat{\delta}^{LP-SM}$ is

$$\begin{aligned}
&\lim_{n\to\infty} diag\left[2^{(2m-1)j_0}, 2^{(2m-2)j_0}, ..., 2^{(2m-p-1)j_0}\right]\left[E(\widehat{\delta}^{LP-SM}) - \delta\right] \\
= &\left[\begin{array}{c} G^{*(2m-1)}(\tau) \cdot (M^*)^{-1}_{(0,0)} N^*_{(0)} \\ G^{*(2m-1)}(\tau) \cdot (M^*)^{-1}_{(1,0)} N^*_{(0)} \\ \cdots \\ G^{*(2m-1)}(\tau) \cdot (M^*)^{-1}_{(p,0)} N^*_{(0)} \end{array}\right],
\end{aligned}$$

where

$$\begin{aligned}
&M^*_{(i,j)} \\
= &\frac{1}{f_V^{i+j}(v_0)} \iiint_a^b (w-t)^i (v-t)^j I\{w - t \geq 0\} I\{v - t \geq 0\} \psi(w)\psi(v) dw dv dt \text{ for } 0 \leq i, j \leq p
\end{aligned}$$

and

$$N^*_{(0)} = \frac{1}{m!(m-1)!} \int_a^b \psi(u) u^m du \cdot \iint_a^b I\{w - t \geq 0\}(-t)^{m-1} \psi(w) dt dw;$$

(2) the asymptotic variance of single-scale local polynomial wavelet estimator $\widehat{\delta}^{LP-SM}$ *is*

$$\lim_{n\to\infty} n \cdot \Xi \cdot Var(\widehat{\delta}^{LP-SM}) = (M^*)^{-1}V^*(M^*)^{-1},$$

where

$$\Xi^{-1}_{(i,j)} = \begin{cases} 2^{(1+2i)j_0}, \text{ when } 0 \le i = j \le p \\ 0, \text{ otherwise} \end{cases}$$

and for $0 \le i, j \le p$,

$$\begin{aligned} & V^*_{(i,j)} \\ = \; & \frac{\sigma^2_{\varepsilon-}(v_0)}{f_V^{i+j}(v_0)} \int_{a-b}^{0} \left[\iint_a^b I\{w-t \ge 0\}(w-t)^i \psi(w)\psi(u+t)dwdt \right] \\ & \left[\iint_a^b I\{w-t \ge 0\}(w-t)^j \psi(w)\psi(u+t)dwdt \right] du \\ & + \\ & \frac{\sigma^2_{\varepsilon+}(v_0)}{f_V^{i+j}(v_0)} \int_{0}^{b-a} \left[\iint_a^b I\{w-t \ge 0\}(w-t)^i \psi(w)\psi(u+t)dwdt \right] \\ & \left[\iint_a^b I\{w-t \ge 0\}(w-t)^j \psi(w)\psi(u+t)dwdt \right] du; \end{aligned}$$

(3)

$$\sqrt{\frac{n}{2^{j_0}}} \left(\widehat{\delta}_0^{LP-SM} - \delta_0 \right) \xrightarrow{d} N(CB_0, V_0),$$

where

$$\begin{aligned} B_0 &= G^{*(2m-1)}(\tau) \cdot (M^*)^{-1}_{(0,0)} N^*_{(0)}, \\ V_0 &= \left[(M^*)^{-1} V^* (M^*) \right]^{-1}_{(0,0)} . \end{aligned}$$

Remarks

(1) In the finite sample, single-scale local polynomial wavelet estimator has the finite unconditional variance via equispacing[5] as in Hall, et al. (1998). However, for local

[5] Other ways of achieving the finite unconditional variance are also available, such as, local polynomial ridge regression by Seifert and Gasser (1996, 2000) and shrinkage local linear regression by Hall and Marron (1997). However those methods require nontrivial restrictions on additional tuning parameters to maintain the first order asymptotic properties.

polynomial kernel approach, a finite-sample analysis by Seifert and Gasser (1996) showed that local polynomial kernel estimators with a compactly supported kernel[6] would have the infinite unconditional variance. Therefore, local polynomial kernel estimators are sensitive to the choices of bandwidths and kernels.

The asymptotic bias term of $\widehat{\delta}_0^{LP-SM}$ is independent of the underlying density f_V. This feature is usually called design-adaptive (Fan, 1992) to both random and fixed designs, and to both highly clustered and nearly uniform designs. In contrast, local polynomial kernel estimators with symmetric kernels are only design-adaptive when the polynomial order is even.

(2) For the known smoothness p, the single-scale local polynomial wavelet estimator $\widehat{\delta}^{LP-SM}$ in Equation (II.10) achieves the optimal convergence rate[7] under Assumption B2 (H), where only the existence of right- and left-hand derivatives are assumed: for the jump size estimate $\widehat{\delta}_0^{LP-SM}$, it could have the optimal convergence rate $n^{-p/(2p+1)}$; for jump sizes in higher-order derivatives, single-scale local polynomial wavelet estimator also achieves their optimal convergence rates, for example, the optimal rate of convergence for kink size estimate is $n^{-p/(2p+3)}$.[8]

(3) In order to select the first order optimal[9] scale j_0, we suggest a local cross-validation approach, since δ_0 is locally defined at $t = \tau$ and cross-validation is consistent

[6] The optimal kernel for jump size from the local polynomial kernel estimation is the compactly supported Bartlett kernel. See Remark 4 in Sun (2005).

[7] When the order of local polynomial wavelet estimators is chosen to be smaller than p, resulting estimators do not have the optimal convergence rate under Assumption B2 (H); on the other hand, using a larger polynomial order than p will only inflate the asymptotic variance without the benefit of bias reduction.

[8] The proof is essentially following Section 2.5 in Tsybakov (2009) and is available upon request.

[9] The first order optimal selected scale $\widehat{j}_0^{opt}$ is defined to satisfy $MSE\left[\widehat{\delta}_0^{LP-SM}(\widehat{j}_0^{opt})\right] \big/ \inf_j MSE\left[\widehat{\delta}_0^{LP-SM}(j)\right] \xrightarrow{a.s.} 1$, although it might not be higher-order optimal where the relative error of the selected scale has the optimal convergence rate.

under weak conditions (Stone, 1984). The selection criterion function is

$$\arg\min_{j_0} \frac{1}{n} \sum_{l=1}^{n} \left[\widehat{\Delta}_{j^*}^{Y}(t_l) - \left(\widehat{\delta}_{0,-l}^{J}(j_0)\right)^T \cdot \widehat{\Delta}_{j^*}^{\widehat{D}_0}(t_l) \right]^2 \widehat{I}_{j^*}(t_l),$$

where

$$\widehat{\delta}_{0,-l}^{LP-SM}(j_0) \text{ is the first entry of } \widehat{\delta}_{-l}^{LP-SM}(j_0)$$

and

$$\widehat{\delta}_{-l}^{LP-SM}(j_0) = \arg \min_{\{\delta_k\}_{k=0}^{p}} \sum_{i=1, l\neq i}^{n} \left[\widehat{\Delta}_{j_0}^{Y}(t_i) - \sum_{k=0}^{p} \delta_k \cdot \widehat{\Delta}_{j_0}^{\widehat{D}_k}(t_i) \right]^2 \widehat{I}_{j_0}(t_i).$$

Here the preliminary scale j^* is chosen to satisfy $j^* \to \infty$ and $2^{j^*}/n \to 0$ as $n \to \infty$, which is less sensitive, according to Mielniczuk, et al. (1989) and Vieu (1991), in the nonparametric curve estimation. Notice that the selected scale j_0 only appears in the leave-one-out estimators $\widehat{\delta}_{-l}^{LP-SM}(j_0)$.

We are now ready to estimate the LATE. Let $\{D_{i:n}\}_{i=1}^{n}$ denote the induced order statistics from $\{V_{i:n}\}_{i=1}^{n}$, and also we have single-scale local polynomial wavelet estimator,

$$\widehat{\zeta}^{LP-SM} = \arg\min_{\{\zeta_k\}_{k=0}^{q}} \sum_{l=1}^{n} \left[\widehat{\Delta}_{j_0}^{D}(t_l) - \sum_{k=0}^{q} \zeta_k \cdot \widehat{\Delta}_{j_0}^{\widehat{D}_k}(t_l) \right]^2 \widehat{I}_{j_0}(t_l)$$

where

$$\begin{aligned} \widehat{\zeta}^{LP-SM} &= \left[\widehat{\zeta}_0^{LP-SM}, \widehat{\zeta}_1^{LP-SM}, ..., \widehat{\zeta}_q^{LP-SM} \right], \\ \widehat{\Delta}_{j_0}^{D}(t) &= \frac{2^{j_0/2}}{n} \sum_{i=1}^{n} D_{i:n} \psi\left[2^{j_0}(t_i - t)\right]. \end{aligned}$$

For simplicity, we use the same wavelet function ψ and scale j_0 to estimate ζ_0.[10]

[10] Since ζ_0 is constrained to $[-1, 1]$, the constrained single-scale local polynomial wavelet estimator is

$$\min_{\{\zeta_k\}_{k=0}^{q}} \sum_{l=1}^{n} \left[\widehat{\Delta}_{j_0}^{D}(t_l) - \sum_{k=0}^{q} \zeta_k \cdot \widehat{\Delta}_{j_0}^{\widehat{D}_k}(t_l) \right]^2 \widehat{I}_{j_0}(t_l) \text{ subject to } \zeta_0 \in [-1, 1].$$

This can be relaxed at the expense of more tedious derivations.

Theorem 16 *Under Assumption Set B and $p \geq 2m$, $q \geq 2m$,*

$$(n/2^{j_0})^{1/2}\left(\frac{\widehat{\delta}_0^{LP-SM}}{\widehat{\zeta}_0^{LP-SM}}-\frac{\delta_0}{\zeta_0}\right)\xrightarrow{d}N\left(\frac{1}{\zeta_0}C\left[B_0-\frac{\delta_0}{\zeta_0}B_0^D\right],\frac{1}{\zeta_0^2}\left[V_0-\frac{2\delta_0}{\zeta_0}V_0^{YD}+\frac{\delta_0^2}{\zeta_0^2}V_0^D\right]\right),$$

where

$$\begin{aligned} B_0^D &= H^{*(2m-1)}(\tau)\cdot(M^*)^{-1}_{(0,0)}N^*_{(0)}, \\ V_0^D &= \left((M^*)^{-1}V^{D*}(M^*)\right)^{-1}_{(0,0)}, \\ V_0^{YD} &= \left((M^*)^{-1}V^{YD*}(M^*)\right)^{-1}_{(0,0)}, \end{aligned}$$

and V^{D*} and V^{YD*} are defined similar to V^*, except for replacing the entity $(\sigma^2_{\varepsilon-},\sigma^2_{\varepsilon+})$ with $(\sigma^2_{\epsilon-},\sigma^2_{\epsilon+})$ and $(\sigma^2_{\varepsilon\epsilon-},\sigma^2_{\varepsilon\epsilon+})$.

Remarks Single-scale local polynomial wavelet approach could test the validity of the LATE identification under discontinuous incentive assignment mechanisms (Proposition 2 in Lee, 2008). If we suppose that there is a pre-determined variable $X \in \mathcal{X} \subset \mathcal{R}$ (one whose value has already been determined prior to treatment assignment), then the argument that $\Pr[X \leq x | V = v_0]$ is continuous for every $x \in \mathcal{X}$ would be necessary for the LATE identification. Using single-scale local polynomial wavelet approach, we are able to examine whether $E(X|V)$ changes discontinuously around $V = v_0$, and then we could justify sufficient conditions for the identification of the LATE.

Multiscale Local Polynomial Wavelet Estimator

In this section, we briefly discuss multiscale local polynomial wavelet estimators. The insight is gained from the fact that, given the cutoff location $V = v_0$, wavelet coefficients are large in many scales other than only j_0. Thus, for multiscale local polynomial wavelet estimators, we would use wavelet coefficients $\left\{\widehat{\Delta}_j^Y(t), \widehat{\Delta}_j^D(t) : j \in \mathbb{Z}^+, t \in [0,1]\right\}$ from both time and frequency domains. In the context of nonparametric kernel density estimations, this has similarities to that of Kotlyarova and Zinde-Walsh (2006, 2008), whose estimator averaged different bandwidths and kernels. Analogously, our multiscale local polynomial wavelet estimators are averaging estimators from different scales. A pooled multiscale local polynomial wavelet estimator[11] is

$$\begin{aligned}&\widehat{\delta}^{LP-MM}\\ = \;& \arg\min_{\{\delta_k\}_{k=0}^{p}} \sum_{j=j_L}^{j_U} \sum_{l=1}^{n} \left[\widehat{\Delta}_j^{Y}(t_l) - \sum_{k=0}^{p} \delta_k \cdot \widehat{\Delta}_j^{\widehat{D}_k}(t_l) \right]^2 \widehat{I}_j(t_l),\end{aligned}$$

where

$$\widehat{\delta}^{LP-MM} = \left[\widehat{\delta}_0^{LP-MM}, \widehat{\delta}_1^{LP-MM}, \cdots, \widehat{\delta}_p^{LP-MM} \right].$$

Another possible approach, which we call a component multiscale local polynomial wavelet estimator, involves estimating different single-scale local polynomial wavelet estimators separately, and then combining individual estimators with optimal weights. The asymptotic mean squared errors (AMSE) of the combined estimator will not be larger than the smallest AMSE of the individual estimator that is included in the combination.

A Synthesis of Existing Jump Size Estimators

In this section, we group existing jump size estimators into either *local constant wavelet estimators* or *local polynomial wavelet estimators*. This includes the wavelet approach from both time and frequency domains, such as in Park and Kim (2006); or estimators from only the time domain, such as in Eubank and Speckman (1994) and Porter

[11] There is an efficiency gain from using more information in multiscale wavelet estimators confirmed by Theorem 7 and 11 in Chapter 1. Heuristically this is because when $\sigma^2_{\varepsilon+}(v_0) = \sigma^2_{\varepsilon-}(v_0)$ and $\{\psi(2^j t - w) : j \in \mathbb{Z}^+, w \in \mathbb{N}\}$ constitute an orthonormal basis for square integrable functions, each of single-scale wavelet estimators is asymptotically independent so that their combination would reduce the variance.

(2003).

Wavelet Approach

The wavelet estimator $\widehat{\delta}_0^{LC-SS}$ in Chapter 1 used only one pair of the wavelet coefficient at the discontinuous location, and had the expression $\widehat{\Delta}_{j_0}^{Y}(\widehat{\tau}) \diagup \widehat{\Delta}_{j_0}^{\widehat{D}_0}(\widehat{\tau})$; on the other hand, $\widehat{\delta}_0^{LC-MM}$ in Chapter 1 integrated all pairs of wavelet coefficients around the discontinuous location, and had the expression as

$$\widehat{\delta}_0^{LC-MM} = \frac{\sum_{j=j_L}^{j_U} \sum_{l=1}^{n} \widehat{\Delta}_j^{Y}(t_l)\, \widehat{\Delta}_j^{\widehat{D}_0}(t_l)\, \widehat{I}_j(t_l)}{\sum_{j=j_L}^{j_U} \sum_{l=1}^{n} \left[\widehat{\Delta}_j^{\widehat{D}_0}(t_l)\right]^2 \widehat{I}_j(t_l)}.$$

The difference between $\widehat{\delta}_0^{LC-SS}$ and $\widehat{\delta}_0^{LC-MM}$ is that the latter uses more sample information by including more wavelet coefficients at different locations and different scales, so that it improved the AMSE. This idea is similar to the nonparametric curve estimations where He and Huang (2009) established an integral estimator with respect to locations, while Choi and Hall (1998) and Cheng, et al. (2007) formed a linear combination of estimators based on different locations.

We define the local constant wavelet estimator as

$$\widehat{\delta}_0^{LC-MM} = \arg\min_{\delta_0} \sum_{j=j_L}^{j_U} \sum_{l=1}^{n} \left[\widehat{\Delta}_j^{Y}(t_l) - \delta_0 \cdot \widehat{\Delta}_j^{\widehat{D}_0}(t_l)\right]^2 \widehat{I}_j(t_l)$$

which only considers jump size δ_0, compared to $\{\delta_k\}_{k=0}^{p}$ in local polynomial wavelet estimators.

Corollary 1 *(1) $\widehat{\delta}_0^{LC-SS}$ is a special case of local constant wavelet estimators $\widehat{\delta}_0^{LC-MM}$, where $j_L = j_U = j_0$ and $\{t_l\}_{l=1}^{n} = \{\widehat{\tau}\}$;*

(3) both $\widehat{\delta}_0^{LC-SS}$ and $\widehat{\delta}_0^{LC-MM}$ only have the optimal convergence rate under the assumption of the function G being p-th differentiable at $t = \tau$.

Time-Domain Approach

This section shows that all time-domain jump size estimators could be asymptotically expressed as either local constant wavelet estimators or local polynomial wavelet estimators. In the literature, there are two types of time-domain estimators: the first one is to estimate jump size by differencing two nonparametric estimators, and the other one is to estimate jump size in the context of a partial linear model. For the purpose of comparison, all these estimators are equispaced. Given a symmetric kernel function K with m vanishing moment, Nadaraya-Watson estimator is

$$\widehat{\delta}_0^{NW} = \frac{\sum_{i=1}^{n} K(\frac{t_i-\widehat{\tau}}{h}) I\{t_i \geq \widehat{\tau}\} Y_{i:n}}{\sum_{i=1}^{n} K(\frac{t_i-\widehat{\tau}}{h}) I\{t_i \geq \widehat{\tau}\}} - \frac{\sum_{i=1}^{n} K(\frac{t_i-\widehat{\tau}}{h}) \left[1 - I\{t_i \geq \widehat{\tau}\}\right] Y_{i:n}}{\sum_{i=1}^{n} K(\frac{t_i-\widehat{\tau}}{h}) \left[1 - I\{t_i \geq \widehat{\tau}\}\right]},$$

partial smoothing kernel estimator (Eubank and Speckman, 1994) is

$$\widehat{\delta}_0^{ES} = \left[I^T\{\mathbf{t} \geq \widehat{\tau}\}(I-S)^2 I\{\mathbf{t} \geq \widehat{\tau}\}\right]^{-1} \left[I^T\{\mathbf{t} \geq \widehat{\tau}\}(I-S)^2 Y_{:n}\right],$$

profiled partial linear estimator (Porter, 2003) is

$$\widehat{\delta}_0^{PO} = \arg\min_{\delta_0} \sum_{i=1}^{n} \left[Y_{i:n} - \delta_0 I\{t_i \geq \widehat{\tau}\} - \sum_{j=1}^{n} \frac{K(\frac{t_i-t_j}{h})}{\sum_{l=1}^{n} K(\frac{t_i-t_l}{h})} \left[Y_{j:n} - \delta_0 I\{t_j \geq \widehat{\tau}\}\right] \right]^2,$$

local polynomial kernel estimator is

$$\widehat{\delta}_0^{LP} = \arg \min_{\{\beta_r\}_{r=0}^{p}, \{\delta_k\}_{k=0}^{p}} \frac{1}{n} \sum_{i=1}^{n} \left[Y_{i:n} - \sum_{r=0}^{p} \beta_j (t_i - \widehat{\tau})^r - \sum_{k=0}^{p} \delta_k (\widehat{F}_V^{-1}(t_i) - v_0)^k I\{t_i \geq \widehat{\tau}\} \right]^2 K(\frac{t_i - \widehat{\tau}}{h}),$$

where

$$\begin{aligned} I^T\{\mathbf{t} &\geq \widehat{\tau}\} = \left[I\{t_1 \geq \widehat{\tau}\}, \cdots, I\{t_n \geq \widehat{\tau}\}\right], \\ S_{(i,j)} &= \frac{1}{nh}\left[K(\frac{t_i - t_j}{h})\right] \text{ for } 1 \leq i, j \leq n, \\ Y_{:n} &= \left[Y_{1:n}, \cdots, Y_{n:n}\right]^T. \end{aligned}$$

Theorem 17 *(1) $\widehat{\delta}_0^{NW}$, $\widehat{\delta}_0^{ES}$ and $\widehat{\delta}_0^{PO}$ could be asymptotically expressed as local constant wavelet estimators $\widehat{\delta}_0^{LC-MM}$, and only have the optimal convergence rate under the assump-*

tion of the function G being p-th differentiable at $t = \tau$;

(2) $\widehat{\delta}_0^{LP}$ could be asymptotically expressed as local polynomial wavelet estimator $\widehat{\delta}_0^{LP-MM}$, and has the optimal convergence rate under Assumption B2 (G).

Wavelet Estimators under Kink Incentive Assignment Mechanism

The LATE Identification and Auxiliary Regressions

Under kink incentive assignment mechanisms, Chapter 1 identified the LATE when there is a kink in the incentive assignment b. Their result is similar to Card, et al. (2009) with several important differences: first and most important, Card, et al. (2009) assumed that $D = b(V)$, thus excluding heterogenous choices among agents being assigned the same incentive; second, they assumed the incentive assignment mechanism b is known; third, they considered a continuous treatment instead of a binary treatment. In addition, the identification of the LATE under kink incentive assignment mechanisms in Chapter 1 is related to a similar result for regression kink design (RKD) in Dong (2010), except her result is not derived under a switching regime model.[12] Under kink incentive assignment mechanisms, the LATE is identified as

$$\frac{\lim_{v\downarrow v_0} dE(Y|V=v)/dv - \lim_{v\uparrow v_0} dE(Y|V=v)/dv}{\lim_{v\downarrow v_0} P'(v) - \lim_{v\uparrow v_0} P'(v)}, \tag{II.11}$$

which is the ratio of the kink size in $E(Y|V)$ to the kink size in $P(v)$ at $V = v_0$. Chapter 1 established auxiliary regressions linking the LATE[13] under kink incentive assignment

[12] In general, a switching regime model is easy to rationalize the observed and counterfactual data (Vytlacil, 2002).

[13] See Proposition 2 in Chapter 1.

mechanisms to kink sizes δ_1 and ζ_1 in (II.12) and (II.13)

$$Y = g_K(V) + \delta_1(V - v_0)I\{V \geq v_0\} + \varepsilon_K, \tag{II.12}$$

$$D = h_K(V) + \zeta_1(V - v_0)I\{V \geq v_0\} + \epsilon_K, \tag{II.13}$$

where $E(\varepsilon_K|V) = 0$, $E(\epsilon_K|V) = 0$, $\delta_1 = \lim_{v\downarrow v_0} dE(Y|V = v)/dv - \lim_{v\uparrow v_0} dE(Y|V = v)/dv$, $\zeta_1 = \lim_{v\downarrow v_0} P'(v) - \lim_{v\uparrow v_0} P'(v)$, and both g_K and h_K are continuously differentiable on the support of V.

Single-scale Local Polynomial Wavelet Estimator

Since the idea of estimating δ_1 and ζ_1 is the same, we focus on the estimation of δ_1. The induced order statistics $\{Y_{i:n}\}_{i=1}^n$ satisfy

$$\begin{aligned} Y_{i:n} &= g_K(V_{i:n}) + \delta_1(V_{i:n} - v_0)I\{V_{i:n} \geq v_0\} + \varepsilon_{i:n}^k \\ &= g_K(\widehat{F}_V^{-1}(t_i)) + \delta_1 \left(\widehat{F}_V^{-1}(t_i) - v_0\right) I\{t_i \geq \widehat{F}_V(v_0)\} + \varepsilon_{i:n}^k \qquad \text{(II.14)} \\ &= G_K(t_i) + \delta_1 \left(F_V^{-1}(t_i) - v_0\right) I\{t_i \geq \tau\} + e_i^k, \end{aligned}$$

where $G_K(t) = g_K(F_V^{-1}(t))$ and

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

Assumption Set BK.

Assumption B1K. A random sample (V_i, Y_i, D_i), $i = 1, ..., n$, is available.

Assumption B2K.

(H). Let $G_K(t) = g_K(F_V^{-1}(t))$. $H_K(t)$ is $(p + 1)$-th continuously differentiable at $t \in (0, 1)\backslash\{\tau\}$, and is continuously differentiable at $t = \tau$ with finite right- and left-hand derivatives up to the order $p + 1$.

(G). Let $H_K(t) = h_K(F_V^{-1}(t))$. $G_K(t)$ is $(q+1)$-th continuously differentiable at $t \in (0,1)\backslash\{\tau\}$, and is continuously differentiable at $t = \tau$ with finite right- and left-hand derivatives up to the order $q+1$.

Assumption B3K.

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

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

(HG) $\sigma_{\varepsilon\epsilon}(v) \equiv E(\varepsilon_K\epsilon_K|V = v)$ is continuous at $v \neq v_0$ and its right and left-hand limits at v_0 exist.

Assumption B4K. (a) The wavelet function ψ 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) $\int_a^b u^m\psi(u)\,du \neq 0$ and $\int_a^b |u^m\psi(u)|\,du < \infty$; (c) ψ has a bounded derivative 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 $\psi(t)$.

Assumption B5K. As $n \to \infty$, $j_k \to \infty$ and $2^{3j_k}/n \to 0$.

Assumption B6K. The function $F_V^{-1}(v)$ is continuously differentiable on the support of V.

Assumption B2K assumes continuously differentiable of G_K and H_K at $t = \tau$ under kink incentive assignment mechanisms. The scale j_k in Assumption B5K is accommodating the convergence rate of the kink size estimate. Single-scale local polynomial wavelet estimator $\widehat{\delta}_K^{LP-SM}$ is

$$\widehat{\delta}_K^{LP-SM} = \arg \min_{\{\delta_k\}_{k=1}^{p+1}} \sum_{l=1}^{n} \left[\widehat{\Delta}_{j_k}^{Y}(t_l) - \sum_{k=1}^{p+1} \delta_k \cdot \widehat{\Delta}_{j_k}^{\widehat{D}_k}(t_l) \right]^2 \widehat{I}_{j_k}(t_l), \tag{II.15}$$

where

$$\widehat{\delta}_K^{LP-SM} = \left[\widehat{\delta}_{K,1}^{LP-SM}, \widehat{\delta}_{K,2}^{LP-SM}, \cdots, \widehat{\delta}_{K,p+1}^{LP-SM} \right].$$

Under kink incentive assignment mechanisms, single-scale local polynomial wavelet estimator $\widehat{\delta}_K^{LP-SM}$ consists of $\left\{ \widehat{\Delta}_{j_k}^{\widehat{D}_k}(t) \right\}_{k=1}^{p+1}$, instead of $\left\{ \widehat{\Delta}_{j_0}^{\widehat{D}_k}(t) \right\}_{k=0}^{p}$ under discontinuous incentive assignment mechanisms.

Theorem 18 *Under Assumption Set BK and $p \geq 2m$,*

$$\begin{aligned} \lim_{n \to \infty} 2^{(2m-1)j_k} \left[E(\widehat{\delta}_{K,1}^{LP-SM}) - \delta_1 \right] &= c < \infty, \\ \lim_{n \to \to \infty} \frac{n}{2^{3j_k}} Var(\widehat{\delta}_{K,1}^{LP-SM}) &= d < \infty, \end{aligned}$$

where c and d are some generic constants.

Since the proof is very similar to Theorem 15, it is omitted. In order to estimate ζ_1, the unconstrained single-scale local polynomial wavelet estimator $\widehat{\zeta}_{KJ}$ is

$$\widehat{\zeta}_K^{LP-SM} = \arg \min_{\{\zeta_k\}_{k=1}^{p+1}} \sum_{l=1}^{n} \left[\widehat{\Delta}_{j_k}^{D}(t_l) - \sum_{k=1}^{p+1} \zeta_k \cdot \widehat{\Delta}_{j_k}^{\widehat{D}_k}(t_l) \right]^2 \widehat{I}_{j_k}(t_l),$$

where

$$\widehat{\zeta}_K^{LP-SM} = \left[\widehat{\zeta}_{K,1}^{LP-SM}, \widehat{\zeta}_{K,2}^{LP-SM}, \cdots, \widehat{\zeta}_{K,p+1}^{LP-SM} \right].$$

In the end, the LATE under kink incentive assignment mechanisms is calculated as $\widehat{\delta}_{K,1}^{LP-SM} / \widehat{\zeta}_{K,1}^{LP-SM}$, which asymptotically converges to a normal distribution at the rate $\left(n/2^{3j_k}\right)^{1/2}$.

Numerical Analysis

This section presents results from Monte Carlo simulation studies. We focus on finite sample performances of local polynomial wavelet estimators. Notice that the unconditional MSE under the finite sample is calculated from $E\left(\widehat{\delta}-\delta\right)^2 = E\left[Var\left(\widehat{\delta} \mid \mathcal{V}\right)\right] + E\left[E\left(\widehat{\delta} \mid \mathcal{V}\right) - \delta\right]^2$ for $\mathcal{V} \equiv \{V_i\}_{i=1}^n$.

Local Polynomial Wavelet Estimator of the Jump Size

The jump size model is

$$Y = \begin{cases} V + V^2 + W,\ 0 \leq V < 0.5 \\ 1 + 2V + 3V^2 + W,\ 0.5 \leq V \leq 1 \end{cases}, \tag{II.16}$$

where $V \sim U\ [0,1]$ and $W \sim N(0, 0.01)$ are independent. The model (II.16) satisfies Assumption Set A, with jump size ζ_0 being 2, kink size ζ_1 being 3, and the second derivative jump size ζ_2 being 4 at the discontinuous location $V = 0.5$. We examine the finite sample performance of $\widehat{\delta}_0^{LP-SM}$ under different sample sizes $\{500, 2500, 5000\}$ with a Daubechies-4 wavelet and 250 simulations within 100 realizations of the design V. We carry out four estimators with different polynomial orders in Equation (II.10). These four estimators are: Zeta2, the single-scale local quadratic wavelet estimator; Zeta1, the single-scale local linear wavelet estimator; Zeta0, the single-scale local constant wavelet estimator; and Zeta3, the single-scale local cubic wavelet estimator.

Several observations are in order from Figure 1 to 3. First, for the sample size $\{500, 2500, 5000\}$, Zeta0 performs the worst because it does not consider any slope change or higher-order derivative discontinuities. Notice that when the scale j_0 is small, the MSE from Zeta0 is the largest due to its worst bias order reduction; however, for larger scales, all four estimators perform similarly due to dominant variances. Second, we find that Zeta1 has the best finite sample MSE even without considering the second order derivative discontinuity. Although according to Theorem 15, Zeta2 should be optimal in the large sample, the asymptotic improvement from Zeta2 is not generally noticeable in finite samples compared with Zeta1. See Marron and Wand (1992) for a similar argument about higher-order kernels. The MSE improvement from Zeta0 to Zeta1 is quite significant due to the bias order reduction, while the improvement from Zeta1 to Zeta2 is trivial and even negative because of the increased variance. Nevertheless, Zeta1 is not that robust to small variations around the optimal scale, so that we still recommend Zeta2 in practice when lacking the reliable scale selection criterion.

To check the robustness of local polynomial wavelet estimators, we implemented the following scenarios and the results in Figure 4 and 5 are encouraging: (1) $W|V$ following a multivariate studentized t distribution (Heckman, Tobias and Vytlacil, 2003); (2) $Var(W|V)$ being heteroskedastic; (3) the marginal distribution of V being different; (4) different signal to noise level; (5) different vanishing moment wavelet functions; (6) different kernel functions to replace $\widehat{I}_{j_0}(t)$ in Equation (II.10); and (7) perturbing Equation (II.16) with an added sine function.

Local Polynomial Wavelet Estimator of the Kink Size

The kink size model is

$$Y = \begin{cases} V - 0.5 + W,\ 0 \leq V < 0.5 \\ 10(V - 0.5) + W,\ 0.5 \leq V \leq 1 \end{cases}, \tag{II.17}$$

where $V \sim U\ [0, 1]$ and $W \sim N(0, 0.02^2)$ are independent. The model (II.17) satisfies Assumption Set B, with kink size δ_1 being 9 at the discontinuous location $V = 0.5$. We examine the finite sample performance of $\widehat{\delta}_{K,1}^{LP-SM}$ under the sample size 500, with a Daubechies-4 wavelet and 250 simulations within 100 realizations of the design V. Four estimators are carried out based on different polynomial orders: Kink_012 is from the single-scale local quadratic wavelet estimator, Kink_01 is from the single-scale local linear wavelet estimator, Kink_12 is from the single-scale local quadratic wavelet estimator without considering the jump size, and finally Kink_1 is from the single-scale local linear wavelet estimator without considering the jump size. All four estimators perform well by Figure 6, although there are some variations at small scales.

Future Research

Local polynomial wavelet estimators could achieve the LATE's optimal convergence rate when the smoothness parameters p and q in Assumption A2 or Assumption B2 are known. However, in practice, this is impossible. We would need to make our estimators adaptive, where the modified versions could have the optimal convergence rate[14] (or up to a logarithmic factor) without knowing the smoothness. In the context of local polynomial kernel estimators of the LATE, Sun (2005) proposed an adaptive estimator based on the estimated smoothness.[15] In order to construct our adaptive local polynomial wavelet estimators, we suggest using the least absolute shrinkage and selection operator (LASSO) to achieve the goal of simultaneous estimation of $\{\delta_k\}_{k=0}^{p}$ and polynomial order selection for p. By introducing a penalty on the absolute values of $\{\delta_k\}_{k=0}^{p^*}$, LASSO local polynomial wavelet estimator is

$$\widehat{\delta}^{LASSO} = \arg \min_{\{\delta_k\}_{k=0}^{p^*}} \sum_{l=1}^{n} \left[\widehat{\Delta}_{j_0}^{Y}(t_l) - \sum_{k=0}^{p^*} \delta_k \cdot \widehat{\Delta}_{j_0}^{\widehat{D}_k}(t_l) \right]^2 \widehat{I}_{j_0}(t_l) + \sum_{k=0}^{p^*} \lambda_k \left| \delta_k \right|,$$

where $\{\lambda_k\}_{k=0}^{p^*}$ are penalty parameters satisfying $\lambda_k \longrightarrow 0$ as $n \to \infty$.

[14] Wavelet curve estimators have this adaptation property (Section 11.3 in Härdle, et al., 2000) even when the underlying function has discontinuities (Park and Kim, 2006), so a natural adaptive estimator for jump sizes could be constructed from differencing two wavelet curve estimates.

[15] Alternatively, we could use maxima propagation of wavelet coefficients by Mallat (2009) to estimate the degree of the smoothness.

The identification and estimation of the LATE in this chapter have focused on incentive assignment mechanisms $b(V)$ depending on only one forcing variable $V \in \mathcal{V} \subset \mathcal{R}$. In some empirical applications, the V may be more than one dimension. For example, Hoekstra (2009) applied RDD to study the effect of attending the flagship state university on earnings. For the data used in the paper, the admissions rule depended on both the student's 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 single forcing variable. It would be interesting to extend local polynomial wavelet estimators to allow for more than one forcing variable. Since multidimensional wavelet functions are powerful tools for edge detection in image processing, wavelet estimators for jump size along the discontinuous curve would be more direction-oriented (Wang, 1998) and more sparsely represented (Mallat, 2009).

In the end, local polynomial wavelet estimators under weak dependent data are left for the future, where special interest is in derivation of the asymptotic normality from linear functions of concomitants of order statistics, with an application to nonparametric jump size estimations.

CHAPTER III

Wavelet Estimators for the Discontinuous Quantile Model

Introduction

This chapter provides a new and intuitive two-step procedure for estimating sizes of the discontinuities in the nonparametric quantile model. For example, we are interested in estimating the jump size δ_0 in the discontinuous median model

$$Y = g(x) + \delta_0 I\left(x \geq x_0\right) + \xi, \text{ where } median(\xi) = 0. \tag{III.1}$$

Such discontinuous sizes from nonparametric *quantile* models are essential to construct credible policy parameters under the framework of the quantile regression discontinuity design (Frandsen, Frolich and Melly, 2011), which is very different from the classical regression discontinuity design (Hahn, Klaauw and Todd, 2001) of only requiring the discontinuous sizes from nonparametric *mean* models. Toward addressing distributional treatment effects, general nonparametric quantile models with potential discontinuities would have more broad applications than nonparametric mean models, such as in Oka(2011).

Our first step is to approximate the above discontinuous nonparametric median model[1] with the discontinuous nonparametric mean model by local medians transformation (Zhou, 2006). Such local medians transformation is turning the problem of nonparametric regression with zero median errors into the one with zero mean errors, where the approximation errors are negligible for our later wavelet estimators. Then the second step is

[1]Without loss of generality, we are focusing on the discontinuous nonparametric median model. For any other discontinuous k-th quantile model ($0 < k < 1$), the local k-th quantile transformation is applied in the first step and the following wavelet estimator is built upon the transformed data.

to carry out local polynomial wavelet estimators (Chapter 2) for the resulting discontinuous nonparametric mean model. Our wavelet estimators here are obtained via wavelet transformation, which approximates the discontinuous nonparametric mean model with the discontinuous parametric mean model.

Our proposed two-step method enjoys several desirable properties: first it is computationally efficient, because the estimated jump size is directly solved from the least squares loss function so that many standard econometrics/statistics software are applicable. Thus our two-step method is in contrast to the standard approach (Oka, 2011), that is taking the difference between the right- and left-hand limits from their check loss functions[2] instead.

Second the consistency and asymptotic normality of our two-step estimator are easily established without involving Bahadur-type techniques. After the local medians transformation in the first step and the wavelet transformation in the second step, our estimated jump size is exactly written as sums of i.i.d. transformed data so that the standard techniques of asymptotic theory are capable. On the other hand when one uses the standard approach from the check loss functions, its jump size estimates could no longer be explicit forms of sums of independent random variables then we have to use the Bahadur-type techniques (cf Section 2.5 in Serfling (1980)).

Third our two-step estimator has the optimal rate of convergence to a wide class of underlying regression functions, which is a powerful consequence from *asymptotic equivalence*[3] between discontinuous median and mean models. Heuristically speaking the main

[2]Yu and Jones(1998) provided a full-fledged asymptotic properties for the estimated quantile function at the limit (boundary) point based on the check loss function.

[3]The asymptotic equivalence here is reserved for two different statistical models/experiments, instead of two different statistics under single statistical model/experiment.

goal of the asymptotic equivalence theory[4] is to approximate a general statistical model with the simpler one: if a complex model is asymptotically equivalent to a simple model, then all asymptotically optimal procedures can be carried over from the simple model to the complex one and the study of the complex model is then essentially simplified. Since local polynomial wavelet estimators are optimal for discontinuous mean models (Chapter 2), our two-step estimator yields an analogous optimal result for discontinuous median models by the asymptotic equivalence.

This chapter is organized as follows. In Section 2 (page 71), we first discuss how to apply local medians transformation in order to approximate the discontinuous nonparametric median model with the discontinuous nonparametric mean model. Later wavelet estimators are used for estimating the jump size δ_0. Section 3 (page 78) would show the asymptotic normality of our two-step approach jointly with the asymptotic equivalence between discontinuous nonparametric median and mean models. Section 4 (page 82) would point out directions for the future work and other potential applications.

The Two-step Estimator of the Jump Size

Our two-step method is built upon two important transformations: the local medians transformation in the first step and the wavelet transformation in the second step, where the local medians transformation is approximating the discontinuous nonparametric median models with discontinuous nonparametric mean models, while the wavelet transformation is approximating the discontinuous nonparametric mean models with discontinuous parametric mean models. Hence our proposed estimator is making use of the transformed

[4] Nussbaum (1996) established the asymptotic equivalence of density estimation and Gaussian white noise under Holder smoothness condition. Brown, Carter, Low and Zhang (2004) extended the result of Nussbaum (1996) under a sharp Besov smoothness constraint condition.

data; and by appropriately controlling for the approximation errors, our two-step estimator could be treated as if we were working on the discontinuous parametric mean model, which is easy to handle and fast to compute.

Local medians transformation is described as binning the original sample into many subintervals then picking local medians within each subinterval. Zhou (2006) provided a tight bound between such local medians and normal random variables, in which we could treat local medians as if they were normal random variables. With the number of bins being chosen in a suitable range, the approximation error between the discontinuous nonparametric median and mean models is small, so that the discontinuous nonparametric mean model based on the local medians transformation data is our new data situation for estimating the jump size δ_0.

Wavelet transformation is generating the wavelet coefficients (the wavelet transformed data) which could characterize the point-wise smoothness of the function: a small (large) wavelet coefficient corresponds to a high (low) smoothness of the function. For our discontinuous nonparametric mean model, it is consisting of both an unknown nonparametric continuous function and a parametric discontinuous indicator function. Hence after the wavelet transformation, we could expect the wavelet coefficients from the unknown nonparametric continuous function are small compared to the ones from the parametric indicator function, that is, the new discontinuous parametric mean model based on the wavelet transformed data is reducing the dimensionality of the discontinuous nonparametric mean model from infinite to finite. Therefore for the discontinuous nonparametric mean model through the local medians transformation, we apply wavelet transformation in order to further approximate with the discontinuous parametric mean model. Notice that in the sequential discontinuous parametric mean model, our ultimate transformed data are the wavelet coefficients of the local medians. Moreover either local constant wavelet estimators (Chapter 1) or local polynomial wavelet estimators (Chapter 2) could be used under these ultimate transformed data.

To illustrate the key ideas and main procedures, we divide this section into two parts and they are:

Part 1: Use local medians transformation to approximate the discontinuous nonparametric median model with the discontinuous nonparametric mean model.

Part 2: Use wavelet transformation to approximate the discontinuous nonparametric mean model with the discontinuous parametric mean model.

Local Medians Transformation

Let work on equation (III.1) where the design points $\{x_i\}_{i=1}^n$ are equally spaced on the interval $[0,1]$. Let the sample $\{Y_i\}_{i=1}^n$ be given as

$$Y_i = g(x_i) + \delta_0 I(x_i \geq x_0) + \xi_i, \text{ where } median(\xi_i) = 0$$

where $x_i = \frac{i}{n}$ and ξ_i are i.i.d. with an unknown density f_ξ. Set the number of subintervals T and the number of observations within each subinterval $m = n/T$. We divide the interval $[0,1]$ into T equal-length subintervals. For $1 \leq j \leq T$, let $I_j = \left\{Y_i : x_i \in (\frac{j-1}{T}, \frac{j}{T}]\right\}$ be the j-th subinterval. Thus for our local medians transformation, the local median Y_j^{med} is the median of the observations in I_j and let $x_{[j]}$ be the induced order statistics from Y_j^{med}. According to Theorem 19, we could treat the local median Y_j^{med} as if it were almost a normal random variable with mean $g(x_{[j]}) + \delta_0 I\left(x_{[j]} \geq x_0\right)$ and variance $1/\left[4mf_\xi^2(0)\right]$.

Assumption 1. Let $\xi_1, \cdots, \xi_n$ be i.i.d. random variables with density function f_ξ, where $\int_{-\infty}^0 f_\xi(u)du = \frac{1}{2}$, $f_\xi(0) > 0$ and $f_\xi(u)$ is Lipschitz[5] at $u = 0$.

[5]Assumption 1 is the standard assumption as in the literature to guarantee the uniqueness of the median,

Assumption 2. The number of subintervals $T = O(n^{3/4})$.

Theorem 19 *Assumptions 1,2 hold. Then* Y_j^{med} *can be written as*

$$\sqrt{m}Y_j^{med} = \sqrt{m}\left[g(x_{[j]}) + \delta_0 I(x_{[j]} \geq x_0)\right] + \frac{Z_j}{2} + \zeta_j \tag{III.2}$$

where

(i) Z_j *is i.i.d.* $N(0, \frac{1}{f_\xi^2(0)})$*;*

(ii) ζ_j *are independent and "stochastically small" random variables satisfying, for any* $l > 0$

$$E\left|\zeta_j\right|^l \leq C_l m^{-l/2}$$

and for any $a > 0$

$$P(\left|\zeta_j\right| > a) \leq C_l(a^2 m)^{-l/2}$$

where $C_l > 0$ *is a constant depending on* l *only.*

Remarks:

(a) In the following sections, we shall assume without loss of generality that $\frac{1}{f_\xi^2(0)}$ is known, since it can be estimated accurately in the sense that the asymptotic properties of our two-step estimator does not change when replacing $\frac{1}{f_\xi^2(0)}$ with an accurate estimate $\frac{1}{\widehat{f_\xi^2}(0)}$. A candidate suggested in Brown, Cai and Zhou (2008) is

$$\frac{1}{\widehat{f_\xi^2}(0)} = \frac{8m}{T}\sum\left(Y_{2j-1}^{med} - Y_{2j}^{med}\right)^2.$$

(b) When the design points $\{x_i\}_{i=1}^n$ are not equally spaced, one can bin the sample so that each bin contains the same number of observations and then make the median of each subinterval. This method produces unequally spaced medians that are homoscedastic since the number of observations in the intervals are the same. An alternative method is to group the sample data using equal-length subintervals and then make the median of each subinterval. This method produces the local median with the heteroskedastic variance depending on the number of observations in the subintervals.

see Chaudhuri (1991).

(c) The stochastic errors ζ_j has the negligible contribution to the asymptotic mean squared errors(MSE) for our two-step estimator, due to its tail bound $P\left(\left|\zeta_j\right| > a\right)$ decaying faster than any polynomial[6] of n. In addition although we might lose the sample size from n to T from binning, the volatility of the new error terms Z_j and ζ_j are also reduced by m $(= \frac{n}{T})$, thus the convergence rate of our two-step estimator is still the same as the standard approach based on the original data size n.

(d) When setting $\sqrt{m}\left[g(\cdot) + \delta_0 I(\cdot)\right]$ to be zero, we have $\sqrt{m}Y_j^{med} = \frac{Z_j}{2} + \zeta_j$ where $E\left(\zeta_j^2\right) = O(m^{-1})$. This expression is very alike to the Bahadur representation for local medians in Bhattacharya and Gangopadhyay (1990), where for $F_\xi'(0) = f_\xi(0) > 0$ with probability 1,

$$\sqrt{m}Y_j^{med} = \frac{\sqrt{m}\left[0.5 - F_{\xi n}(0)\right]}{f_\xi(0)} + R_m,$$

where $E\left(R_m^2\right) = O(m^{-1/2})$ (Duttweiler, 1973). Apparently in terms of approximating local medians with normal random variables, our theorem 19 provides a sharper bound than the Bahadur representation.

Wavelet Transformation

After we reformulate the jump size δ_0 in eq (III.2) by an approximately discontinuous mean model, both local constant and local polynomial wavelet estimators are applicable. They are from minimizing the sum of squared wavelet residuals in the approximately discontinuous parametric mean model[7]. In particular, local polynomial wavelet estimator in Chapter 2 explicitly accounted for potential higher-order derivative discontinuities and fully

[6]Cai and Zhou (2010) showed the tail bound $P\left(\left|\zeta_j\right| > a\right)$ has indeed the exponential decay rate for natural exponential families, so that one direct implication of this strengthed result is that we might choose the bin size m at smaller order. I expect this also holds for more general cases and am leaving it for future research.

[7]Such model is obtained after the wavelet transformation and hence the new dependent and independent variables of the model are wavelet coefficients.

explored wavelet coefficients generated from both time and frequency domains. In an application for discontinuous (kink) incentive assignment mechanisms, local polynomial wavelet estimators also attained the optimal convergence rate. Another desirable property of local polynomial wavelet estimators is that it could jointly and optimally estimate jump sizes in any order derivatives, such as jump size, kink size (the difference between right- and left-hand first derivative limits), and up to jump sizes in higher-order derivatives. In the end, Chapter 2 showed that all existing jump size estimators (based on the equispaced data) share a common structure: being members of a class of local polynomial wavelet estimators.

Following eq (III.2), let take the wavelet transformation on both sides in order to reduce the dimensionality from infinite to finite.

Recall for $1 \leq j \leq T$,

$$\begin{aligned} Y_j^{med} &= g(x_{[j]}) + \delta_0 I(x_{[j]} \geq x_0) + \frac{Z_j}{2\sqrt{m}} + \frac{\zeta_j}{\sqrt{m}} && \text{(III.3)} \\ &\approx g(x_{[j]}) + \delta_0 I(x_{[j]} \geq x_0) + \frac{Z_j}{2\sqrt{m}} && \text{(III.4)} \end{aligned}$$

where the approximately equality comes from the smaller ζ_j than Z_j, because from the previous section we have $Var(Z_j) = \frac{1}{4f_\xi^2(0)}$, while $Var(\zeta_j) = O(\frac{1}{m})$.

Equispace the design point $x_{[j]}$ into $t = \frac{j}{T}$ for $1 \leq j \leq T$. Let $\widehat{\Delta}_{j_0}^A(t)$ denote the wavelet coefficient of $\{A_i\}_{i=1}^n$ at the location $t \in [0,1]$ and a scale j_0, where t and j_0 represent the time and frequency parameters, respectively:

$$\widehat{\Delta}_{j_0}^A(t) = \frac{2^{j_0/2}}{T} \sum_{i=1}^{T} A_i \psi\left[2^{j_0}(t_i - t)\right].$$

After the wavelet transformation on both sides of eq (III.3), we obtain

$$\begin{aligned}
\widehat{\Delta}_{j_0}^{Y^{med}}(t) &\approx \widehat{\Delta}_{j_0}^{g}(t) + \delta_0 \cdot \widehat{\Delta}_{j_0}^{D_0}(t) + \frac{\widehat{\Delta}_{j_0}^{Z}(t)}{2\sqrt{m}} \\
&\approx \delta_0 \cdot \widehat{\Delta}_{j_0}^{D_0}(t) + \frac{\widehat{\Delta}_{j_0}^{Z}(t)}{2\sqrt{m}}
\end{aligned}$$

where the approximately equality comes from the wavelet coefficients $\widehat{\Delta}_{j_0}^{g}(t)$ has the smaller magnitude than $\widehat{\Delta}_{j_0}^{D_0}(t)$, since the continuous function $g(\cdot)$ is smoother than the indicator function $I(\cdot \geq x_0)$. And the corresponding wavelet coefficients are

$$\begin{aligned}
\widehat{\Delta}_{j_0}^{Y^{med}}(t) &= \frac{2^{j_0/2}}{T}\sum_{i=1}^{T} Y_i^{med}\psi\left[2^{j_0}(t_i - t)\right] \\
\widehat{\Delta}_{j_0}^{D_0}(t) &= \frac{2^{j_0/2}}{T}\sum_{i=1}^{T} I(t_i \geq x_0)\psi\left[2^{j_0}(t_i - t)\right] \\
\widehat{\Delta}_{j_0}^{Z}(t) &= \frac{2^{j_0/2}}{T}\sum_{i=1}^{T} Z_i\psi\left[2^{j_0}(t_i - t)\right].
\end{aligned}$$

Thus if the function g is p-th continuously differentiable at x_0, the local constant wavelet estimator for δ_0 (Chapter 1) is

$$\widehat{\delta}_0^{LC-med} = \arg\min_{\delta_0}\sum_{l=1}^{T}\left[\widehat{\Delta}_{j_0}^{Y^{med}}(t_l) - \delta_0 \cdot \widehat{\Delta}_{j_0}^{D_0}(t_l)\right]^2 \widehat{I}_{j_0}(t_l). \tag{III.5}$$

Otherwise if the function g is continuous but with finite right- and left-hand derivatives up to the order p at x_0, the local polynomial wavelet estimator for δ_0 (Chapter 2) is motivated by adding the terms $\sum_{k=1}^{p}\delta_k \cdot \widehat{\Delta}_{j_0}^{D_k}(t)$ in order to capture the potential discontinuities in higher-order derivatives in $\widehat{\Delta}_{j_0}^{g}(t)$. Thus the refined discontinuous parametric mean model is becoming

$$\begin{aligned}
\widehat{\Delta}_{j_0}^{Y^{med}}(t) &\approx \delta_0 \cdot \widehat{\Delta}_{j_0}^{D_0}(t) + \sum_{k=1}^{p}\delta_k \cdot \widehat{\Delta}_{j_0}^{D_k}(t) + \frac{\widehat{\Delta}_{j_0}^{Z}(t)}{2\sqrt{m}} \\
&\approx \sum_{k=0}^{p}\delta_k \cdot \widehat{\Delta}_{j_0}^{D_k}(t) + \frac{\widehat{\Delta}_{j_0}^{Z}(t)}{2\sqrt{m}}.
\end{aligned}$$

Therefore our the local polynomial wavelet estimator $\widehat{\delta}_0^{LP-med}$ is computed from

$$\widehat{\delta}_0^{med} = e_1 \cdot \widehat{\delta}^{LP-med}$$

where the selection vector $e_1 \equiv (1, \underbrace{0, \cdots, 0}_{p})$ and

$$\widehat{\delta}^{LP-med} = \arg \min_{\{\delta_k\}_{k=0}^{p}} \sum_{l=1}^{T} \left[\widehat{\Delta}_{j_0}^{Y^{med}}(t_l) - \sum_{k=0}^{p} \delta_k \cdot \widehat{\Delta}_{j_0}^{D_k}(t_l) \right]^2 \widehat{I}_{j_0}(t_l), \tag{III.6}$$

and

$$\widehat{\Delta}_{j_0}^{D_k}(t) = \frac{2^{j_0/2}}{T} \sum_{i=1}^{T} I(t_i \geq x_0)^k \left[t_i - x_0\right]^k \psi \left[2^{j_0}(t_i - t)\right].$$

Remarks: Notice that the difference between the current two-step wavelet estimators and wavelet estimators proposed in Chapter 1 and Chapter 2: here our two-step estimators are based on the local medians instead of the original data.

Asymptotic Properties of the Two-step Estimator

Since the local polynomial wavelet estimator $\widehat{\delta}_0^{LP-med}$ in eq (III.6) is optimal and preferred to the local constant wavelet estimator $\widehat{\delta}_0^{LC-med}$ in eq (III.5), we are focusing on the asymptotic properties of $\widehat{\delta}_0^{LP-med}$ in terms of the asymptotic normality and asymptotic optimality, respectively.

Asymptotic Normality

Assumption 3. $g(\cdot)$ is p-th continuously differentiable at $(0,1)\backslash\{x_0\}$, and is continuous at x_0 with finite right- and left-hand derivatives up to the order p. When allowing for the potential higher-order derivative discontinuities, we could decompose $g(\cdot) \equiv G(\cdot) + \sum_{k=1}^{p} \delta_k \cdot I(\cdot \geq x_0)^k [\cdot - x_0]$ where $G(\cdot)$ is p-th continuously differentiable at $(0,1)$.

Assumption 4. (a) The real-valued wavelet function $\psi(\cdot)$ is continuous with compact support $[a, b]$, where $a < 0 < b$ and m vanishing moments, i.e., $\int_a^b u^j \psi(u)\, du = 0$ for $j = 0, 1, ..., m-1$; (b) $\int_a^b u^m \psi(u)\, du \neq 0$ and $\int_a^b |u^m \psi(u)|\, du < \infty$; (c) ψ has a bounded derivative and satisfies the admissibility condition that $\int \left|\widehat{\psi}(\xi)\right|^2 / |\xi|\, d\xi < \infty$, where $\widehat{\psi}(\xi)$ is the Fourier transform of $\psi(t)$.

Assumption 5. As $n \to \infty$, $j_0 \to \infty$, $2^{j_0}/n \to 0$ and $\left(1/2^{j_0}\right)^{2m-1} \sqrt{n/2^{j_0}} \to C < \infty$.

Theorem 20 *Under Assumptions 1-5 and $p \geq 2m$:*

(1) the asymptotic bias of $\widehat{\delta}^{LP-med}$ is

$$\lim_{n\to\infty} diag\left[2^{(2m-1)j_0}, 2^{(2m-2)j_0}, ..., 2^{(2m-p-1)j_0}\right]\left[E(\widehat{\delta}^{LP-med}) - \delta\right]$$

$$= \begin{bmatrix} G^{(2m-1)}(x_0) \cdot (M)^{-1}_{(0,0)} N^*_{(0)} \\ G^{(2m-1)}(x_0) \cdot (M)^{-1}_{(1,0)} N^*_{(0)} \\ \cdots \\ G^{(2m-1)}(x_0) \cdot (M)^{-1}_{(p,0)} N^*_{(0)} \end{bmatrix},$$

where

$$M_{(i,j)}$$

$$= \iiint_a^b (w-t)^i (v-t)^j I\{w-t \geq 0\} I\{v-t \geq 0\} \psi(w)\psi(v) dwdvdt \text{ for } 0 \leq i, j \leq p$$

and

$$N^*_{(0)} = \frac{1}{m!(m-1)!} \int_a^b \psi(u) u^m du \cdot \iint_a^b I\{w-t \geq 0\}(-t)^{m-1}\psi(w) dtdw;$$

(2) the asymptotic variance of $\widehat{\delta}^{LP-med}$ is

$$\lim_{n\to\infty} n \cdot \Xi \cdot Var(\widehat{\delta}^{LP-med}) = (M)^{-1} V (M)^{-1},$$

where

$$\Xi^{-1}_{(i,j)} = \begin{cases} 2^{(1+2i)j_0}, \text{ when } 0 \leq i = j \leq p \\ 0, \text{ otherwise} \end{cases}$$

and for $0 \leq i, j \leq p$,

$$\begin{aligned} & V_{(i,j)} \\ = \; & \frac{1}{f_\xi^2(0)} \int_{a-b}^{b-a} \left[\iint_a^b I\{w-t \geq 0\}(w-t)^i \psi(w)\psi(u+t) dwdt \right] \\ & \left[\iint_a^b I\{w-t \geq 0\}(w-t)^j \psi(w)\psi(u+t) dwdt \right] du; \end{aligned}$$

(3)

$$\sqrt{\frac{n}{2^{j_0}}} \left(\widehat{\delta}_0^{LP-med} - \delta_0 \right) \xrightarrow{d} N(CB_{med}, V_{med}),$$

where

$$\begin{aligned} B_{med} &= G^{(2m-1)}(x_0) \cdot (M)^{-1}_{(0,0)} N^*_{(0)}, \\ V_{med} &= \left[(M)^{-1} V (M) \right]^{-1}_{(0,0)}. \end{aligned}$$

Remark:

(a) In order to calculate the asymptotic variance of $\widehat{\delta}_0^{LP-med}$ easily, we could directly use the robust variance results[8] from the standard OLS packages based on the wavelet coefficients of the local medians. For example when $p = 0$ in eq (III.6),

$$\widehat{\delta}_0^{LP-med} = \frac{\sum_{l=1}^T \left(\widehat{\Delta}_{j_0}^{Y^{med}}(t_l) \widehat{I}_{j_0}(t_l) \right) \left(\widehat{\Delta}_{j_0}^{D_0}(t_l) \widehat{I}_{j_0}(t_l) \right)}{\sum_{l=1}^T \left[\widehat{\Delta}_{j_0}^{D_0}(t_l) \widehat{I}_{j_0}(t_l) \right]^2}$$

which is resembling an OLS estimator with the dependent variable $\left(\widehat{\Delta}_{j_0}^{Y^{med}}(t_l) \widehat{I}_{j_0}(t_l) \right)_{l=1}^T$ and the independent variable $\left(\widehat{\Delta}_{j_0}^{D_0}(t_l) \widehat{I}_{j_0}(t_l) \right)_{l=1}^T$.

(b) In order to select the scale (smoothing parameter) j_0 in our two-step estimator, we are suggesting the mean cross-validation based on the approximately discontinuous mean model. This is another theoretical advantage of our two-step estimator through local medians transformation, since the smoothing parameters selection under the nonparametric mean model is far more well-studied than the one under the nonparametric median model.

[8]The proof for the consistency of the estimated variance is upon request.

Otherwise if we use the standard approach from the check loss functions, we have to use the complicated median cross-validations (Zheng and Yang, 1998) for the original discontinuous median model.

Asymptotic Optimality

In this subsection we would establish the asymptotic equivalence between the discontinuous median and mean models, so that our two-step estimator has the optimal rate of convergence to a wide class of underlying regression functions. Henceforth we follow Cai and Zhou (2009)'s notations to define the asymptotic equivalence for unbounded loss functions[9].

Definition 1 *(Cai and Zhou, 2009) Two sequences of experiments E_n and F_n are called asymptotically equivalent with respect to the sets of procedures Λ_{E_n} and Λ_{F_n} and set of loss functions Γ_n if*

$$\Delta(E_n, F_n; \Gamma_n, \Lambda_{E_n}, \Lambda_{F_n}) \longrightarrow 0 \text{ as } n \to \infty$$

where

$$\Delta(E_n, F_n; \Gamma_n, \Lambda_{E_n}, \Lambda_{F_n}) = \max\{\delta(E_n, F_n; \Gamma_n, \Lambda_{E_n}, \Lambda_{F_n}), \delta(F_n, E_n; \Gamma_n, \Lambda_{F_n}, \Lambda_{E_n})\}$$

and $\delta(E_n, F_n; \Gamma_n, \Lambda_{E_n}, \Lambda_{F_n}) \equiv \inf\{\epsilon \geq 0;$ for every procedure $\tau_n \in \Lambda_{F_n}$ there exists a procedure $\xi_n \in \Lambda_{E_n}$ such that $R(\theta; \xi_n) \leq R(\theta; \tau_n) + 2\varepsilon$ for every $\theta \in \Theta$ for any loss function $L \in \Gamma_n$ and its associated risk function $R\}$.

Let the discontinuous median model E_n and the discontinuous mean model F_n to be

$$E_n \; : \; Y_i = g(x_i) + \delta_0 I(x_i \geq x_0) + \xi_i, \text{ where } median(\xi_i) = 0 \text{ and } i = 1, ..., n,$$

$$F_n \; : \; Y_j^{med} = g(x_j) + \delta_0 I(x_j \geq x_0) + \frac{Z_j}{2\sqrt{m}}, \text{ where } Z_j \text{ is i.i.d. } N(0, \frac{1}{f_\xi^2(0)}) \text{ and } j = 1, ..., T.$$

[9]Although the definition of the asymptotic equivalence is abstract and lengthy, we could actually reduce the general comparison between two procedures from two different experiments to simply a comparison of Le Cam deficiency between two experiments, which is independent of procedures.

Assumption 6. Assume that

$$\begin{aligned} \mu(a) &\leq Ca^2, \\ E\exp\left[t\left(r_a(\xi) - \mu(a)\right)\right] &\leq \exp(Ct^2a^2), \end{aligned}$$

for $0 \leq |a| < \varepsilon$ and $0 \leq |ta| < \varepsilon$ for some $\varepsilon > 0$, where $r_a(\xi) = \log \frac{h(\xi-a)}{h(\xi)}$ and $\mu(a) = Er(\xi)$.

Theorem 21 *Under Assumptions 1-6, the two experiments F_n and E_n are asymptotically equivalent with respect to the set of procedures Λ_n and set of loss functions Γ_n, where Λ_n are meant to be the estimates of the jump size δ_0.*

Thus our two-step estimator $\widehat{\delta}_0^{LP-med}$ has the optimal rate of convergence under the discontinuous nonparametric median model E_n.

Remarks: Notice that in Cai and Zhou (2009) they established the similar asymptotic equivalence between the nonparametric median and mean models under the certain smoothness conditions with the focus of estimating of the underlying function. However our primary interest here is the estimation of jump size δ_0 instead of the function itself, thus we will not need stringent smoothness conditions as in their paper.

Future research

Besides introducing the new approach for discontinuous quantile models, our chapter is largely expository to spotlight some important "local transformation" inequality which might have many empirical implications in econometrics. For example, we could use the local means transformation[10] (bin the original sample into many subintervals, then compute the mean within each subinterval) under the switching regime model so that Heckman Two-Step procedure would be robust to misspecification of outcome error distributions, such as,

[10]Zhou (2006) also provided a tight bound between such local means and normal random variables, in which we could treat local means as if they were normal random variables.

Hall, et al. (1998) considered a similar transformation to address the irregular design of X_i under the conditional mean model, however they did not derive the bound of the approximation error.

non-normal/normal, asymmetric/symmetric outcome errors (Chen, Fan and Wu, 2012). It is hoped that this chapter helps to make the local transformation tool known to a wider econometrics audience.

Another interesting topic would research on the local transformation for the non i.i.d. random observations. Lemma 4 in Brown, Cai and Zhou (2010) provided a specific local medians transformation result for the natural exponential distributions' family with independent but not identically distributed observations. However for the general dependent and non-identically distributed observations, there are no results available in the literature.

In the end, we would like to extend the current discontinuous median model under the univariate fixed design to the general one under the multivariate random design.

Proofs of Chapter 1

Proof of Theorem 1. Take $v_+ \in V$ and $v_- \in V$ 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:

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

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

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

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

and

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

The same results hold for $\lim_{v_+,v_-\to v_0} E(Y_j|V=v_-,\cdot)$. 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 $j = 1, 0$,

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

Q.E.D.

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 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

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

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

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

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

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

Q.E.D.

Proof of Proposition 1. 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

$$\lim_{v\downarrow v_0}g(v)=\lim_{v\downarrow v_0}E(Y|V=v)-\delta_0=\lim_{v\uparrow v_0}E(Y|V=v)=\lim_{v\uparrow v_0}g(v).$$

By definition, we have

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

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. 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:

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

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:

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

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. 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, $g_K(V) = E(Y|V) - \delta_1(V - v_0)I\{V \geq v_0\}$. We only need to show that it is continuous at v_0. Under Condition K4(A), we know: $\lim_{v\downarrow v_0} E(Y|V = v) - \lim_{v\downarrow v_0} E(Y|V = v) = 0 = \delta_0$, thus $E(Y|V) = g(v)$ from Proposition 2.2, which is continuous on the support of V. Since $\delta_1(V - v_0)I\{V \geq v_0\}$ is continuous on the support of V, $g_K(V)$ is continuous.

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

$$\lim_{v\downarrow v_0} \frac{dg_K(v)}{dv} = \lim_{v\downarrow v_0} \frac{dE(Y|V = v)}{dv} - \delta_1 = \lim_{v\uparrow v_0} \frac{dE(Y|V = v)}{dv}$$

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 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.

Next we will make extensive use of the following Taylor expansions. 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 local constant wavelet 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 of Wang and Cai (2010). Because of this, we will work with the infeasible versions of our estimators with $\widehat{\tau}$ replaced by τ. With slight abuse of notation, we will use the same notations to denote the corresponding infeasible estimators. For notational compactness, we let $\psi_j[\cdot] = 2^j\psi\left[2^j\cdot\right]$.

Proof of Theorem 3. We will complete the proof in two steps:

Step 1. We show that $\overline{\delta}_0^{LC-SS}$ has the asymptotic distributions stated in the theorem;

Step 2. We show: $\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_0^{LC-SS} - \overline{\delta}_0^{LC-SS}) = o_p(1)$.

Proof of Step 1: Note that we can write $\overline{\delta}_0^{LC-SS}$ as: $\overline{\delta}_0^{LC-SS} = \frac{1}{n}\sum_{i=1}^n J(\frac{i}{n})Y_{[i:n]}$, 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 $\sqrt{\frac{n}{2^{j_0}}}(\overline{\delta}_0^{LC-SS} - \delta_0)$ has the limiting distribution stated in Theorem 3. The conditions in Lemma B.1 are satisfied: $E(Y|V=v) = g(v) + \delta_0 I\{v \geq v_0\}$ is right continuous by Proposition 1 and of bounded variation in any finite interval; $J(\frac{i}{n})$ is bounded and continuous by Assumption A4.

First, let us calculate $\lim_{n\to\infty} E(\overline{\delta}_0^{LC-SS})$:

$$\begin{aligned}
& \int_{-\infty}^{+\infty} [g(v) + \delta_0 I\{v \geq v_0\}] \frac{\psi_{j_0}(F_V(v)-\tau)}{\int_0^b \psi(u)du} dF_V(v) \\
= \quad & \frac{\int_{-\infty}^{+\infty} g(v)\psi_{j_0}(F_V(v)-\tau)dF_V(v)}{\int_0^b \psi(u)du} + \frac{\int_{-\infty}^{+\infty} \delta_0 I\{v \geq v_0\}\psi_{j_0}(F_V(v)-\tau)dF_V(v)}{\int_0^b \psi(u)du} \\
= \quad & \frac{\int_a^b g\left[F_V^{-1}(\frac{u}{2^{j_0}}+\tau)\right]\psi(u)du}{\int_0^b \psi(u)du} + \delta_0 \\
= \quad & \begin{cases} \frac{\frac{1}{2^{j_0}}\left[G_+^{(1)}(\tau) - G_-^{(1)}(\tau)\right]\int_0^b \psi(u)udu}{\int_0^b \psi(u)du} + \delta_0 + s.o., \text{ under A2(G)(a)} \\ \frac{(\frac{1}{2^{j_0}})^m G^{(m)}(\tau)\int_a^b u^m\psi(u)du}{\int_0^b \psi(u)du} + \delta_0 + s.o, \text{ under A2(G)(b)} \end{cases}.
\end{aligned}$$

Then,

$$\begin{aligned}
\lim_{n\to\infty} 2^{j_0}\left[E(\overline{\delta}_0^{LC-SS}) - \delta_0\right] &= \frac{\left[G_+^{(1)}(\tau) - G_-^{(1)}(\tau)\right]\int_0^b \psi(u)udu}{\int_0^b \psi(u)du}, \text{ under A2(G)(a)}, \\
\lim_{n\to\infty} 2^{mj_0}\left[E(\overline{\delta}_0^{LC-SS}) - \delta_0\right] &= \frac{G^{(m)}(\tau)\int_a^b u^m\psi(u)du}{\int_0^b \psi(u)du}, \text{ under A2(G)(b)}.
\end{aligned}$$

Second, let us calculate $\lim_{n\to\infty} nVar(\bar{\delta}_0^{LC-SS})$:

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

where

$$\begin{aligned}
m(v) &= g(v)+\delta_0 I\{v\geq v_0\}, \\
A_1 &= \int_{-\infty}^{+\infty}\left[\frac{\psi_{j_0}(F_V(v)-\tau)}{\int_0^b \psi(u)du}\right]^2 \sigma^2(v)dF_V(v), \\
A_2 &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[F_V(v_1\wedge v_2)-F_V(v_1)F_V(v_2)]\frac{\psi_{j_0}(F_V(v_1)-\tau)\psi_{j_0}(F_V(v_2)-\tau)}{\left(\int_0^b \psi(u)du\right)^2}dm(v_1)dm(v_2).
\end{aligned}$$

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:

$$\lim_{n\to\infty}\frac{n}{2^{j_0}}Var(\bar{\delta}_0^{LC-SS}) = \frac{\left[\sigma_+^2(v_0)\int_0^b \psi^2(u)du+\sigma_-^2(v_0)\int_a^0 \psi^2(u)du\right]}{\left(\int_0^b \psi(u)du\right)^2}.$$

By Lemma B.1, we obtain:

$$\sqrt{\frac{n}{2^{j_0}}}(\bar{\delta}_0^{LC-SS}-\delta_0)\xrightarrow{d}\begin{cases} N(C_aB_a,V), \text{ under A2(G)(a)} \\ N(C_bB_b,V), \text{ under A2(G)(b)}\end{cases}.$$

Proof of Step 2: Note that

$$\begin{aligned}
& \sqrt{\frac{n}{2^{j_0}}}(\hat{\delta}_0^{LC-SS}-\bar{\delta}_0^{LC-SS}) \\
= \;& \sqrt{\frac{n}{2^{j_0}}}\bar{\delta}_0^{LC-SS}\left(\frac{\int_0^b \psi(u)du-\frac{1}{n}\sum_{j=1}^n I\{t_j\geq\tau\}\psi_{j_0}[t_j-\tau]}{\frac{1}{n}\sum_{j=1}^n I\{t_j\geq\tau\}\psi_{j_0}[t_j-\tau]}\right),
\end{aligned}$$

where the first term satisfies: $\sqrt{\frac{n}{2^{j_0}}\overline{\delta}_0^{LC-SS}} = O_p(\sqrt{\frac{n}{2^{j_0}}})$. 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,

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

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

$$
\begin{aligned}
\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_0^{LC-SS} - \overline{\delta}_0^{LC-SS}) &= \sqrt{\frac{n}{2^{j_0}}\overline{\delta}_0^{LC-SS}}\left(\frac{\int_0^b \psi(u)du}{\frac{1}{n}\sum_{j=1}^{n} I\{t_j \geq \tau\}\psi_{j_0}[t_j - \tau]} - 1\right) \\
&= O_p(\sqrt{\frac{n}{2^{j_0}}})O_p(\frac{2^{j_0}}{n}) = o_p(1).
\end{aligned}
$$

Q.E.D.

Proof of Theorem 4. First, we derive the asymptotic covariance between $\widehat{\delta}_0^{LC-SS}$ and $\widehat{\zeta}_0^{LC-SS}$:

$$
\begin{aligned}
& \lim_{n\to\infty} Cov\left[(n/2^{j_0})^{1/2}(\widehat{\delta}_0^{LC-SS} - \delta_0), (n/2^{j_0})^{1/2}(\widehat{\zeta}_0^{LC-SS} - \zeta_0)\right] = \lim_{n\to\infty}\frac{n}{2^{j_0}}Cov\left[\widehat{\delta}_0^{LC-SS}, \widehat{\zeta}_0^{LC-SS}\right] \\
= & \lim_{n\to\infty}\frac{n}{2^{j_0}}Cov\left[\frac{1}{n}\sum_{i=1}^{n} J(\frac{i}{n})Y_{[i:n]}, \frac{1}{n}\sum_{i=1}^{n} J(\frac{i}{n})D_{[i:n]}\right] \\
= & \frac{1}{2^{j_0}}\int_{-\infty}^{+\infty} J^2(F_V(x))\sigma_{\varepsilon\epsilon}^2(x)dF_V(x) \\
& +\frac{1}{2^{j_0}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[F_V(x\wedge y) - F_V(x)F_V(y)]J(F_V(x))J(F_V(y))dm(x)dm^D(y) \\
= & \frac{\left[\sigma_{\varepsilon\epsilon+}^2(v_0)\int_0^b \psi^2(u)du + \sigma_{\varepsilon\epsilon-}^2(v_0)\int_a^0 \psi^2(u)du\right]}{\left(\int_0^b \psi(u)du\right)^2},
\end{aligned}
$$

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 $(n/2^{j_0})^{1/2}(\widehat{\delta}_0^{LC-SS} - \delta_0)$ and $(n/2^{j_0})^{1/2}(\widehat{\zeta}_0^{LC-SS} - \zeta_0)$. In the end, we use Delta method to establish the asymptotic distribution for $\widehat{\delta}_0^{LC-SS}/\widehat{\zeta}_0^{LC-SS}$.

Q.E.D.

Proof of Theorem 5. First we work with the bias term for $\widehat{\delta}_1^{LC-SS}$:

$$\begin{aligned}
&\int_{-\infty}^{+\infty} \left[g_K(v) + \delta_1(V - v_0)I\{v \geq v_0\}\right] \frac{\psi_{j_0}(F_V(v) - \tau)}{\int_0^b \left[F_V^{-1}(\frac{u}{2^{j_0}} + \tau) - v_0\right] \psi(u)du} dF_V(v) \\
= \quad & \frac{\int_{-\infty}^{+\infty} g_K(v)\psi_{j_0}(F_V(v) - \tau)dF_V(v)}{\int_0^b \left[F_V^{-1}(\frac{u}{2^{j_0}} + \tau) - v_0\right] \psi(u)du} + \frac{\int_{-\infty}^{+\infty} \delta_1 vI\{v \geq v_0\}\psi_{j_0}(F_V(v) - \tau)dF_V(v)}{\int_0^b \left[F_V^{-1}(\frac{u}{2^{j_0}} + \tau) - v_0\right] \psi(u)du} \\
= \quad & \frac{\int_a^b g_K\left[F_V^{-1}(\frac{u}{2^{j_0}} + \tau)\right] \psi(u)du}{\int_0^b \left[F_V^{-1}(\frac{u}{2^{j_0}} + \tau) - v_0\right] \psi(u)du} + \delta_1.
\end{aligned}$$

Under A2K(G)(a), we have

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

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,

$$\begin{aligned}
\lim_{n\to\infty} 2^{j_0}\left[E(\widehat{\delta}_1^{LC-SS}) - \delta_1\right] &= \frac{\left[G_{K+}^{(2)}(\tau) - G_{K-}^{(2)}(\tau)\right] f_V(v_0) \int_0^b u^2\psi(u)du}{2\int_0^b u\psi(u)du}, \text{ under A2K(G)(a),} \\
\lim_{n\to\infty} 2^{mj_0}\left[E(\widehat{\delta}_1^{LC-SS}) - \delta_1\right] &= \frac{G_K^{(m+1)}(\tau)}{(m+1)!} \frac{f_V(v_0) \int_a^b u^{m+1}\psi(u)du}{\int_0^b u\psi(u)du}, \text{ under A2K(G)(b).}
\end{aligned}$$

Now we work on the variance term:

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

Thus we get

$$
\lim_{n\to\infty}\frac{n}{2^{3j_0}}Var(\widehat{\delta}_1^{LC-SS})=\frac{f_V^2(v_0)\left[\sigma_{\varepsilon+}^2(v_0)\int_0^b\psi^2(u)du+\sigma_{\varepsilon-}^2(v_0)\int_a^0\psi^2(u)du\right]}{\left(\int_0^b u\psi(u)du\right)^2}.
$$

Finally the asymptotic normality of $\widehat{\delta}_1^{LC-SS}$ is established by following a similar proof to that of Theorem 3.

Q.E.D.

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

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

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 Theorem 7. Let

$$\overline{\delta}_0^{LC-SM} = n^{-1} \sum_{i=1}^{n} J_{W1}\left(\frac{i}{n}\right) Y_{[i:n]},$$

where

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

We will complete the proof in two steps:

Step 1. We show $\overline{\delta}_0^{LC-SM}$ has the asymptotic distributions stated in the theorem;

Step 2. We show: $\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_0^{LC-SM} - \overline{\delta}_0^{LC-SM}) = o_p(1)$.

Proof of Step 1: First, let us calculate $\lim_{n\to\infty} E(\overline{\delta}_0^{LC-SM})$:

$$\int_{-\infty}^{\infty} [g(v) + \delta_0 I\{v \geq v_0\}]\, J_{W1}(F_V(v)) dF_V(v) = \int_{-\infty}^{\infty} g(v) J_{W1}(F_V(v)) dF_V(v) + \delta_0.$$

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

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

where

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

For large enough j_0, we obtain:

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

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,

$$
\begin{aligned}
T_{W1} &= (\frac{1}{2^{j_0}})^2 \sum_{k=1}^{l_G-1}\frac{G^{(k)}(\tau)}{k!}\int_a^b\int_a^b L(t)\psi(s)(\frac{s-t}{2^{j_0}})^k dsdt + s.o. \\
&= \{
\end{aligned}
$$

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

For large enough j_0,

$$T_{W1D} = \int_0^1 \int_0^1 \left[\int_a^b I\{w \geq 2^{j_0}(\tau - t)\}\psi\left[w\right]dw\right] I_{j_0}(t)I\{v \geq \tau\}\psi\left[2^{j_0}(v-t)\right]dtdv$$

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

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:

$$\lim_{n\to\infty} 2^{j_0}\left[E(\bar{\delta}_0^{LC-SM}) - \delta_0\right] = \frac{\left[G_+^{(1)}(\tau) - G_-^{(1)}(\tau)\right]\int_a^b\int_a^b L(t)\psi(s)(s-t)I\{s-t \geq 0\}dsdt}{\int_{a-b}^0 M(v)dv};$$

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

Second, let us calculate the asymptotic variance of $\bar{\delta}_0^{LC-SM}$:

$$\frac{\int_{-\infty}^{\infty}\left[\int_0^1\int_0^1 I_{j_0}(t)I\{w\geq\tau\}\psi_{j_0}\left[w-t\right]\psi\left[2^{j_0}(F_V(v)-t)\right]dtdw\right]^2\sigma^2(v)dF_V(v)}{T_D^2}$$

$$+\frac{\int\int\left\{\begin{array}{c}[F_V(v_1\wedge v_2)-F_V(v_1)F_V(v_2)]\cdot\\ \cdot\left[\int_0^1\int_0^1 I_{j_0}(t)I\{w\geq\tau\}\psi_{j_0}\left[w-t\right]\psi\left[2^{j_0}(F_V(v_1)-t)\right]dtdw\right]\\ \cdot\left[\int_0^1\int_0^1 I_{j_0}(t)I\{w\geq\tau\}\psi_{j_0}\left[w-t\right]\psi\left[2^{j_0}(F_V(v_2)-t)\right]dtdw\right]\end{array}\right\}dm(v_1)dm(v_2)}{T_D^2}$$

$$= \frac{T_{W12}}{T_{W1D}^2}+\frac{T_{W13}}{T_{W1D}^2},$$

where

$$T_{W12} = \int_{-\infty}^{\infty}\left[\int_0^1\int_0^1 I_{j_0}(t)I\{w\geq\tau\}\psi_{j_0}\left[w-t\right]\psi\left[2^{j_0}(F_V(v)-t)\right]dtdw\right]^2\sigma^2(v)dF_V(v),$$

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

For the T_{W12} term

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

where

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

Notice that for large enough j_0,

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

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

In the end,

$$\begin{aligned}
\lim_{n\to\infty} \frac{n}{2^{j_0}} Var(\bar{\delta}_0^{LC-SM}) &= \frac{\int_{a-b}^{b-a} M^2(u)\sigma^2(F^{-1}(\tau - \frac{u}{2^{j_0}}))du}{\left[\int_{a-b}^{b-a} M(v)I\{v \le 0\}dv\right]^2} \\
= \quad & \frac{\sigma_+^2(v_0)\int_0^{b-a} M^2(v)dv + \sigma_-^2(v_0)\int_{a-b}^0 M^2(v)dv}{\left[\int_{a-b}^0 M(v)dv\right]^2}.
\end{aligned}$$

Proof of Step 2: Note that

$$\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_0^{LC-SM} - \bar{\delta}_0^{LC-SM}) = \sqrt{\frac{n}{2^{j_0}}}\left[\frac{1}{n}\frac{\sum_{i=1}^n (A_i - C_i)\, Y_{i:n}}{B}\right] + \sqrt{\frac{n}{2^{j_0}}}\bar{\delta}_0^{LC-SM}\left[\frac{D}{B} - 1\right],$$

where

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

For the term $\sqrt{\frac{n}{2^{j_0}}}\bar{\delta}_0^{LC-SM}\left[\frac{D}{B} - 1\right]$, note that

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

Thus,

$$\sqrt{\frac{n}{2^{j_0}}}\bar{\delta}_0^{LC-SM}\left[\frac{D}{B}-1\right]=o_p(1).$$

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

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

In the end, we have: $\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_0^{LC-SM}-\bar{\delta}_0^{LC-SM})=o_p\left(1\right).$

Q.E.D.

Proof of Theorem 9: 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:

$$\begin{aligned}
\widehat{\delta}_0^{LC-MS} &= \frac{\widehat{\Delta}_{j_0}^{Y}(\tau)\widehat{\Delta}_{j_0}^{\widehat{D_0}}(\tau)+\widehat{\Delta}_{j_0+1}^{Y}(\tau)\widehat{\Delta}_{j_0+1}^{\widehat{D_0}}(\tau)}{\left[\widehat{\Delta}_{j_0}^{\widehat{D_0}}(\tau)\right]^2+\left[\widehat{\Delta}_{j_0+1}^{\widehat{D_0}}(\tau)\right]^2}\\
&= \frac{\frac{1}{n}\sum_{i=1}^{n}\left\{\begin{array}{c}\frac{1}{n}\sum_{l=1}^{n}I\{t_l\geq\tau\}\psi_{j_0}[t_l-\tau]\psi[2^{j_0}(t_i-\tau)]\\ +\frac{2}{n}\sum_{l=1}^{n}I\{t_l\geq\tau\}\psi_{j_0}[2(t_l-\tau)]\psi[2^{j_0+1}(t_i-\tau)]\end{array}\right\}Y_{[i:n]}}{\left[\frac{2^{j_0/2}}{n}\sum_{l=1}^{n}I\{t_l\geq\tau\}\psi[2^{j_0}(t_i-\tau)]\right]^2+\left[\frac{2^{(j_0+1)/2}}{n}\sum_{l=1}^{n}I\{t_l\geq\tau\}\psi[2^{j_0+1}(t_i-\tau)]\right]^2}.
\end{aligned}$$

Let

$$\bar{\delta}_0^{LC-MS}=\frac{1}{n}\sum_{i=1}^{n}J_{W2}(\frac{i}{n})Y_{[i:n]},$$

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 $\overline{\delta}_0^{LC-MS}$ has the asymptotic distributions stated in the theorem;

Step 2. We show $\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_0^{LC-MS} - \overline{\delta}_0^{LC-MS}) = o_p(1)$.

Proof of Step 1: First

$$\begin{aligned}
\int_{-\infty}^{\infty} [g(v) + \delta_0 I\{v \geq v_0\}]\, J_{W2}(F_V(v)) dF_V(v) &= \int_{-\infty}^{\infty} g(v) J_{W2}(F_V(v)) dF_V(v) + \delta_0 \\
= \frac{2}{3}\int_{-\infty}^{\infty} g(v)\frac{\psi_{j_0}[F_V(v)-\tau]}{\int_0^b \psi(t)dt} dF_V(v) &+ \frac{1}{3}\int_{-\infty}^{\infty} g(v)\frac{2\psi_{j_0}[2(F_V(v)-\tau)]}{\int_0^b \psi(t)dt} dF_V(v) + \delta_0 \\
= \frac{2}{3}P_1 + \frac{1}{3}P_2 + \delta_0,
\end{aligned}$$

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, 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, 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 $K_n = 1$,

$$\begin{aligned}
\lim_{n\to\infty} 2^{j_0}\left[E(\bar{\delta}_0^{LC-MS}) - \delta_0\right] &= \frac{5\left[G_+^{(1)}(\tau) - G_-^{(1)}(\tau)\right]\int_0^b \psi(u)udu}{6\int_0^b \psi(u)du}, \text{ under A2(G)(a);} \\
\lim_{n\to\infty} 2^{mj_0}\left[E(\bar{\delta}_0^{LC-MS}) - \delta_0\right] &= \frac{(2+\frac{1}{2^m})G^{(m)}(\tau)\int_a^b u^m\psi(u)du}{3\int_0^b \psi(u)du}, \text{ under A2(G)(b).}
\end{aligned}$$

Second, let us calculate $\lim_{n\to\infty} nVar(\bar{\delta}_0^{LC-MS})$ 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

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

Since,

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

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

$$\lim_{n\to\infty}\frac{n}{2^{j_0}}Var(\bar{\delta}_0^{LC-MS})=\frac{2}{3}V+\frac{2^{j_0+3}\left[\sigma_+^2(v_0)-\sigma_-^2(v_0)\right]\int_0^b\psi(u)\psi(2u)du}{9\left[\int_0^b\psi(t)dt\right]^2}.$$

Proof of Step 2. It is similar to that of Theorem 3.

Q.E.D.

Proof of Theorem 11: 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:

$$\begin{aligned}
\widehat{\delta}_0^{LC-MM} &= \frac{\sum_{j=j_L}^{j_U}\int_0^1\widehat{\Delta}_j^Y(t)\,\widehat{\Delta}_j^{\widehat{D_0}}(t)\,\widehat{I_{j_0}}(t)dt}{\sum_{j=j_L}^{j_U}\int_0^1\left[\widehat{\Delta}_j^{\widehat{D_0}}(t)\right]^2\widehat{I_{j_0}}(t)dt} \\
&= \frac{\int_0^1\widehat{\Delta}_{j_0}^Y(t)\,\widehat{\Delta}_{j_0}^{\widehat{D_0}}(t)\,\widehat{I_{j_0}}(t)dt+\int_0^1\widehat{\Delta}_{j_0+1}^Y(t)\,\widehat{\Delta}_{j_0+1}^{\widehat{D_0}}(t)\,\widehat{I_{j_0+1}}(t)dt}{\int_0^1\left[\widehat{\Delta}_{j_0}^{\widehat{D_0}}(t)\right]^2\widehat{I_{j_0}}(t)dt+\int_0^1\left[\widehat{\Delta}_{j_0+1}^{\widehat{D_0}}(t)\right]^2\widehat{I_{j_0+1}}(t)dt}.
\end{aligned}$$

Let

$$\bar{\delta}_0^{LC-MM}=\frac{1}{n}\sum_{i=1}^{n}J_{W3}(\frac{i}{n})Y_{[i:n]},$$

where

$$
\begin{aligned}
J_{W3}(\frac{i}{n}) &= \frac{Z^1_{W3}(\frac{i}{n}) + Z^2_{W3}(\frac{i}{n})}{Q_1 + Q_2}, \\
Z^1_{W3}(\frac{i}{n}) &= \int_0^1 \int_0^1 I\{w \geq \tau\}\psi_{j_0}[w-t]\,\psi\left[2^{j_0}(\frac{i}{n} - t)\right] I_{j_0}(t)dtdw, \\
Z^2_{W3}(\frac{i}{n}) &= \int_0^1 \int_0^1 I\{w \geq \tau\}2\psi_{j_0}[2(w-t)]\,\psi\left[2^{j_0+1}(\frac{i}{n} - t)\right] I_{j_0+1}(t)dtdw, \\
Q_1 &= \int_0^1 \int_0^1 \int_0^1 I\{u \geq \tau\}I\{v \geq \tau\}\psi_{j_0}[u-t]\,\psi\left[2^{j_0}(v-t)\right] I_{j_0}(t)dudvdt, \\
Q_2 &= \int_0^1 \int_0^1 \int_0^1 I\{u \geq \tau\}I\{v \geq \tau\}2\psi_{j_0}[2(u-t)]\,\psi\left[2^{j_0+1}(v-t)\right] I_{j_0+1}(t)dudvdt.
\end{aligned}
$$

We will complete the proof in two steps:

Step 1. We show $\overline{\delta}_0^{LC-MM}$ has the asymptotic distributions stated in the theorem;

Step 2. We show: $\sqrt{\frac{n}{2^{j_0}}}(\widehat{\delta}_0^{LC-MM} - \overline{\delta}_0^{LC-MM}) = o_p(1)$.

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

$$
\int_{-\infty}^{\infty} [g(v) + \delta_0 I\{v \geq v_0\}]\, J_{W3}(F_V(v))dF_V(v) = \int_{-\infty}^{\infty} g(v)J_{W3}(F_V(v))dF_V(v) + \delta_0.
$$

From the proof of Theorem 7, we know

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

Under A2(G)(a):

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

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

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

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

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

Then when $K_n = 1$,

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

Therefore, under A2(G)(a),

$$\begin{aligned}
&\lim_{n\to\infty} 2^{j_0} \left[E(\bar{\delta}_0^{LC-MM}) - \delta_0\right] \\
= \quad & \frac{9\left[G_+^{(1)}(\tau) - G_-^{(1)}(\tau)\right] \int_a^b \int_a^b L(t)\psi(s)(s-t) I\{s-t \geq 0\} ds dt}{10 \int_{a-b}^0 M(v) dv};
\end{aligned}$$

under A2(G)(b),

$$\begin{aligned}
&\lim_{n\to\infty} 2^{(2m-1)j_0} \left[E(\bar{\delta}_0^{LC-MM}) - \delta_0\right] \\
= \quad & \frac{\left[4 + (\frac{1}{2})^{2m-1}\right] G^{(2m-1)}(\tau) \int_a^b \psi(s) s^m ds \int_a^b L(t)(-t)^{m-1} dt}{5m!(m-1)! \int_{a-b}^0 M(v) dv}.
\end{aligned}$$

Second, let us calculate $\lim_{n\to\infty} nVar(\bar{\delta}_0^{LC-MM})$ with $K_n = 1$,

$$\begin{aligned}
&\lim_{n\to\infty} \frac{n}{2^{j_0}} Var(\bar{\delta}_0^{LC-MM}) \\
= \quad & \frac{\int_{a-b}^{b-a} M^2(u)\sigma^2(F_V^{-1}(\tau - \frac{u}{2^{j_0}}))du}{2^{4j_0}(Q_1+Q_2)^2} + \frac{\int_{a-b}^{b-a} M^2(u)\sigma^2(F_V^{-1}(\tau - \frac{u}{2^{j_0}}))du}{2^{3j_0+3}(Q_1+Q_2)^2} \\
& + \frac{2}{(Q_1+Q_2)^2} \int_{-\infty}^{+\infty} Z_{W3}^1(F_V(v)) Z_{W3}^2(F_V(v)) \sigma^2(v) dF_V(v) \\
= \quad & \frac{18}{25} V_{W1} + \frac{Cross_1}{(Q_1+Q_2)^2},
\end{aligned}$$

where

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

Therefore when $K_n = 1$,

$$\lim_{n\to\infty} \frac{n}{2^{j_0}} Var(\bar{\delta}_0^{LC-MM}) = \frac{18}{25} V_{W1} + \frac{2^{j_0+3}\left[\sigma_+^2(v_0) - \sigma_-^2(v_0)\right] \int_0^b M(u)M(2u)du}{9\left[\int_0^b \psi(t)dt\right]^2}.$$

The proof of second step could be obtained analogous to Theorem 7.

Q.E.D.

Proof of Theorem 13: We will complete the proof in two steps:

Step 1. We show $\bar{\delta}_1^{LC-SM}$ has the asymptotic distributions stated in the theorem;
Step 2. We show: $\sqrt{\frac{n}{2^{3j_0}}}(\widehat{\delta}_1^{LC-SM} - \bar{\delta}_1^{LC-SM}) = o_p(1)$.
Proof of Step 1. Note that

$$\begin{aligned}
& \int_{-\infty}^{\infty} \left[g_K(v) + \delta_1(V - v_0)I\{v \geq v_0\}\right] J_{KW1}(F_V(v))dF_V(v) \\
= & \int_{-\infty}^{\infty} g_K(v)J_{KW1}(F_V(v))dF_V(v) + \delta_1..
\end{aligned}$$

For the first term,

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

where

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

Then for large enough j_0,

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

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:

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

where

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

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,

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

where

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

For C_1 term,

$$C_1 = \frac{1}{2^{4j_0} f_V^2(v_0)} \int_{a-b}^{0} \left[t^2 M(t) - t M_1(t)\right] dt + s.o.,$$

For C_2 term,

$$C2 = \frac{1}{2^{4j_0} f_V^2(v_0)} \int_{a-b}^{0} \left[-t M_2(t)\right] dt + s.o.,$$

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

$$T_{KW1D} = \frac{1}{2^{4j_0} f_V^2(v_0)} \int_{a-b}^{0} \left[t^2 M(t) - t M_1(t) - t M_2(t)\right] dt + s.o..$$

Then under the conditions of part (i):

$$\begin{aligned} & \lim_{n\to\infty} 2^{j_0} \left[E(\bar{\delta}_1^{LC-SM}) - \delta_1\right] \\ = & \frac{\left[G_{K+}^{(2)}(\tau) - G_{K-}^{(2)}(\tau)\right] f_V(v_0) \int_a^b \int_a^b (s-w)^2 \psi(s) I\{s-w \geq 0\} \left[L_1(w) - w L_0(w)\right] ds dw}{2 \int_{a-b}^{0} \left[t^2 M(t) - t M_1(t) - t M_2(t)\right] dt}. \end{aligned}$$

under the conditions of part (ii):

$$\lim_{n\to\infty} 2^{m j_0} \left[E(\bar{\delta}_1^{LC-SM}) - \delta_1\right] = \frac{f_V^2(v_0) \int_a^b \psi(s) s^{m+1} ds \left[\Gamma_0 + \sum_{i=1}^m \Gamma_i\right]}{\int_{a-b}^{0} \left[t^2 M(t) - t M_1(t) - t M_2(t)\right] dt}.$$

For the asymptotic variance,

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

where

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

in which

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

Thus

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

and

$$
\begin{aligned}
& \lim_{n\to\infty} \frac{n}{2^{3j_0}} Var(\overline{\delta}_1^{LC-SM}) \\
= & \frac{f_V^2(v_0) \left[\sigma_{\varepsilon+}^2(v_0) \int_{a-b}^0 \left[M_1(t) + M_2(t) - tM(t)\right]^2 dt + \sigma_{\varepsilon-}^2(v_0) \int_0^{b-a} \left[M_1(t) + M_2(t) - tM(t)\right]^2 dt\right]}{\left[\int_{a-b}^0 \left[t^2 M(t) - tM_1(t) - tM_2(t)\right] dt\right]^2}.
\end{aligned}
$$

The proof of the second step could be obtained analogous to Theorem 7.

Q.E.D.

Proofs of Chapter 2

Important Facts about the Wavelet

For a comprehensive wavelet study, readers should refer to Daubechies (1992) and Mallat (2009).

Theorem 22 *(Mallat, Theorem 6.4) If $f \in L^2(R)$ is Lipschitz α at v, then there exists A such that*

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

Conversely, if α is not an integer and there exists A and $\alpha' < \alpha$ such that

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

then f is Lipschitz α at v.

For notational compactness, we might use $\psi_j[\cdot] \equiv 2^j \psi\left[2^j \cdot\right]$ throughout this appendix.

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

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

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

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

Lemma C.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*

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

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*

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

Lemma C.2 *(Extension to the* Q *vector-valued estimators* S_n*) Suppose the above conditions are satisfied and let*

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

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

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

Define for any $1 \leq q_1, q_2 \leq Q$*,*

$$\begin{aligned}\sigma^2_{(q_1,q_2)} \quad &= \quad \int_{-\infty}^{+\infty} J^{[q_1]}(F(x)) \cdot J^{[q_2]}(F(x)) \cdot \sigma^2(x)dF(x) \\ &\quad + \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} [F(x \wedge y) - F(x)F(y)] \times J^{[q_1]}(F(x))J^{[q_2]}(F(x))dm(x)dm(y),\end{aligned}$$

then

$$\begin{aligned}&\lim_{n\to\infty} E(S_n) \\ = \quad &\left[\int_{-\infty}^{+\infty} m(x)J^{[1]}(F(x))dF(x), \int_{-\infty}^{+\infty} m(x)J^{[2]}(F(x))dF(x), \cdots, \int_{-\infty}^{+\infty} m(x)J^{[Q]}(F(x))dF(x)\right]^T,\end{aligned}$$

and

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

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

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

Proof: For the asymptotic bias part, it is straightforward from Equation (12) and (13) of Theorem 2 in Yang (1981) because of the closed-form expression, and we would apply the Cramer-Wold device for deriving its asymptotic variance similar to Theorem 1 in Yang (1981). The asymptotic normality follows Theorem 6 in Yang (1977).

Q.E.D.

Lemma C.3 *Let*

$$M = \begin{bmatrix} \int_0^1 \left[\widehat{\Delta}_{j_0}^{D_0}(t)\right]^2 I_{j_0}(t)dt & \int_0^1 \widehat{\Delta}_{j_0}^{D_0}(t)\widehat{\Delta}_{j_0}^{D_1}(t)I_{j_0}(t)dt & \dots & \int_0^1 \widehat{\Delta}_{j_0}^{D_0}(t)\widehat{\Delta}_{j_0}^{D_p}(t)I_{j_0}(t)dt \\ \int_0^1 \widehat{\Delta}_{j_0}^{D_0}(t)\widehat{\Delta}_{j_0}^{D_1}(t)I_{j_0}(t)dt & \int_0^1 \left[\widehat{\Delta}_{j_0}^{D_1}(t)\right]^2 I_{j_0}(t)dt & \dots & \int_0^1 \widehat{\Delta}_{j_0}^{D_1}(t)\widehat{\Delta}_{j_0}^{D_p}(t)I_{j_0}(t)dt \\ \vdots & \vdots & \ddots & \vdots \\ \int_0^1 \widehat{\Delta}_{j_0}^{D_0}(t)\widehat{\Delta}_{j_0}^{D_p}(t)I_{j_0}(t)dt & \int_0^1 \widehat{\Delta}_{j_0}^{Dp}(t)\widehat{\Delta}_{j_0}^{D_1}(t)I_{j_0}(t)dt & \dots & \int_0^1 \left[\widehat{\Delta}_{j_0}^{D_p}(t)\right]^2 I_{j_0}(t)dt \end{bmatrix}.$$

Then

$$M = diag\left[\frac{1}{2^{j_0}}, \frac{1}{2^{2j_0}}, ..., \frac{1}{2^{(p+1)j_0}}\right] \cdot M^* \cdot diag\left[\frac{1}{2^{j_0}}, \frac{1}{2^{2j_0}}, ..., \frac{1}{2^{(p+1)j_0}}\right] + s.o.,$$

where for $0 \leq i, j \leq p$,

$$M^*_{(i,j)} = \frac{1}{f_V^{i+j}(v_0)} \iiint_a^b (w-t)^i (v-t)^j I\{w-t \geq 0\} I\{v-t \geq 0\} \psi(w)\psi(v) dwdvdt.$$

Proof: First, from the term T_{W1D} in Theorem 7 of Chapter 1:

$$\int_0^1 \left[\widehat{\Delta}_{j_0}^{D_0}(t)\right]^2 I_{j_0}(t)dt = \frac{1}{2^{2j_0}} \int_{a-b}^0 M(v)dv + s.o.,$$

where $M(v) \equiv \int_a^b \int_a^b I\{w \geq t+v\}\psi(w)\psi(t)dtdw$.

Second, from the term T_{KW1D} in Theorem 9 of Chapter 1:

$$\int_0^1 \left[\widehat{\Delta}_{j_0}^{D_1}(t)\right]^2 I_{j_0}(t)dt = \frac{1}{2^{4j_0} f_V^2(v_0)} \int_{a-b}^0 \left[t^2 M(t) - tM_1(t) - tM_2(t)\right] dt + s.o.,$$

where

$$M_1(s) = \iint_a^b (-t) I\{w \geq t+s\}\psi(w)\psi(t)dtdw,$$

$$M_2(s) = \iint_a^b w I\{w \geq t+s\}\psi(w)\psi(t)dtdw.$$

For the general term $\int_0^1 \widehat{\Delta}_{j_0}^{D_i}(t)\widehat{\Delta}_{j_0}^{D_j}(t)I_{j_0}(t)dt$ where $0 \le i, j \le p$:

$$
\begin{aligned}
&\int_0^1 \widehat{\Delta}_{j_0}^{D_i}(t)\widehat{\Delta}_{j_0}^{D_j}(t)I_{j_0}(t)dt \\
= &\iiint_0^1 \begin{bmatrix} I_{j_0}(t) \\ \\ \left[F_V^{-1}(w) - v_0\right]^i \end{bmatrix} \begin{bmatrix} I\{w \ge \tau\} \\ \\ \left[F_V^{-1}(v) - v_0\right]^j \end{bmatrix} I\{v \ge \tau\} \begin{Bmatrix} 2^{j_0}\psi\left[2^{j_0}(w-t)\right] \\ \\ \psi\left[2^{j_0}(v-t)\right] \end{Bmatrix} dwdvdt + s.o. \\
= &\frac{1}{2^{2j_0}} \iiint_a^b \begin{bmatrix} F_V^{-1}(\frac{w-t}{2^{j_0}} + \tau) \\ \\ -v_0 \end{bmatrix}^i I\{w - t \ge 0\} \begin{bmatrix} F_V^{-1}(\frac{v-t}{2^{j_0}} + \tau) \\ \\ -v_0 \end{bmatrix}^j I\{v - t \ge 0\}\psi(w)\psi(v)dwdvdt + s.o. \\
= &\frac{1}{2^{2j_0}} \iiint_a^b \begin{bmatrix} (\frac{w-t}{2^{j_0}}) \\ \\ \left(F_V^{-1}(\tau)\right)^{(1)} \end{bmatrix}^i I\{w - t \ge 0\} \begin{bmatrix} (\frac{v-t}{2^{j_0}}) \\ \\ \left(F_V^{-1}(\tau)\right)^{(1)} \end{bmatrix}^j I\{v - t \ge 0\}\psi(w)\psi(v)dwdvdt + s.o. \\
= &\frac{1}{2^{(i+j+2)j_0}\left[f_v(v_0)\right]^{i+j}} \iiint_a^b (w-t)^i(v-t)^j I\{w-t \ge 0\}I\{v-t \ge 0\}\psi(w)\psi(v)dwdvdt + s.o.
\end{aligned}
$$

where the $s.o.$ term in first equality comes from replacing the finite double summations with integration, whose precision could be controlled by the Koksma–Hlawka inequality.

Q.E.D.

Lemma C.4 *Define*

$$
\widehat{\Delta}_{j_0}^{G^*}(t) = \frac{1}{n}\sum_{m=1}^{n} G^*(t_m)2^{j_0/2}\psi\left[2^{j_0}(t_m - t)\right],
$$

and let

$$
N = \begin{bmatrix} \int_0^1 \widehat{\Delta}_{j_0}^{D_0}(t)I_{j_0}(t)\widehat{\Delta}_{j_0}^{G^*}(t)dt \\ \int_0^1 \widehat{\Delta}_{j_0}^{D_1}(t)I_{j_0}(t)\widehat{\Delta}_{j_0}^{G^*}(t)dt \\ \vdots \\ \int_0^1 \widehat{\Delta}_{j_0}^{D_p}(t)I_{j_0}(t)\widehat{\Delta}_{j_0}^{G^*}(t)dt \end{bmatrix}.
$$

Then

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

where

$$
N^*_{(0)} = \frac{1}{m!(m-1)!}G^{*(2m-1)}(\tau)\int_a^b \psi(u)u^m du \cdot \iint_a^b I\{w - t \ge 0\}(-t)^{m-1}\psi(w)dtdw,
$$

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

$$N_{(i)}^* = \iiint_a^b \sum_{\alpha_2+\alpha_3=m-1} \frac{1}{m!\alpha_2!\alpha_3!} \frac{\partial^{2m-1} K_i(k_1,k_2,k_3)}{\partial^m k_1 \partial^{\alpha_2} k_2 \partial^{\alpha_3} k_3}_{|k_1=k_2=k_3=0} u^m w^{\alpha_2} (-t)^{\alpha_3} dwdudt.$$

Proof: First let us look at the $\int_0^1 \widehat{\Delta}_{j_0}^{D_0}(t) I_{j_0}(t) \widehat{\Delta}_{j_0}^{G^*}(t)dt$ term:

$$\begin{aligned}
& \int_0^1 \widehat{\Delta}_{j_0}^{D_0}(t) I_{j_0}(t) \widehat{\Delta}_{j_0}^{G^*}(t)dt \\
= \; & \iiint_0^1 G^*(u) I\{w \geq \tau\} I_{j_0}(t) 2^{j_0} \psi\left[2^{j_0}(w-t)\right] \psi\left[2^{j_0}(u-t)\right] dwdudt + s.o. \\
= \; & \frac{1}{2^{2j_0}} \iiint_a^b G^*(\frac{u-t}{2^{j_0}} + \tau) I\{w - t \geq 0\} \psi(w)\psi(u) dwdudt + s.o. \\
= \; & \frac{1}{2^{2j_0}} \iiint_a^b \left[G^*(\tau) + \sum_{k=1}^{2m-1} \frac{G^{*(k)}(\tau)}{k!} \left(\frac{u-t}{2^{j_0}}\right)^k + s.o.\right] I\{w - t \geq 0\} \psi(w)\psi(u) dwdudt + s.o. \\
= \; & \frac{1}{2^{2j_0}} \iiint_a^b \frac{1}{(2m-1)!} G^{*(2m-1)}(\tau) \left(\frac{u-t}{2^{j_0}}\right)^{2m-1} I\{w - t \geq 0\} \psi(w)\psi(u) dwdudt + s.o. \\
= \; & \frac{1}{2^{(2m+1)j_0}} \frac{1}{m!(m-1)!} G^{*(2m-1)}(\tau) \int_a^b \psi(s) s^m ds \cdot \iint_a^b I\{w - t \geq 0\} \psi(w) (-t)^{m-1} dwdt + s.o.,
\end{aligned}$$

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

For the general term $\int_0^1 \widehat{\Delta}_{j_0}^{D_i}(t) I_{j_0}(t) \widehat{\Delta}_{j_0}^{G^*}(t)dt$ where $1 \leq i \leq p$:

$$\begin{aligned}
& \int_0^1 \widehat{\Delta}_{j_0}^{D_i}(t) I_{j_0}(t) \widehat{\Delta}_{j_0}^{G^*}(t)dt \\
= \; & \iiint_0^1 G^*(u) I\{w \geq \tau\} \left[F_V^{-1}(w) - v_0\right]^i I_{j_0}(t) 2^{j_0} \psi\left[2^{j_0}(w-t)\right] \psi\left[2^{j_0}(u-t)\right] dwdudt + s.o. \\
= \; & \frac{1}{2^{2j_0}} \iiint_a^b G^*(\frac{u-t}{2^{j_0}} + \tau) \left[F_V^{-1}(\frac{w-t}{2^{j_0}} + \tau) - v_0\right]^i I\{w - t \geq 0\} \psi(w)\psi(u) dwdudt + s.o. \\
= \; & \frac{1}{2^{(2m+1)j_0}} \iiint_a^b \sum_{\alpha_2+\alpha_3=m-1} \frac{1}{m!\alpha_2!\alpha_3!} \frac{\partial^{2m-1} K_i(k_1,k_2,k_3)}{\partial^m k_1 \partial^{\alpha_2} k_2 \partial^{\alpha_3} k_3}_{|k_1=k_2=k_3=0} u^m w^{\alpha_2} (-t)^{\alpha_3} dwdudt + s.o.,
\end{aligned}$$

where the last equality comes from the trivariate Taylor expansion of $G^*(\frac{u-t}{2^{j_0}} + \tau) \left[F_V^{-1}(\frac{w-t}{2^{j_0}} + \tau) - v_0\right]^i \equiv K_i(\frac{u}{2^{j_0}}, \frac{w}{2^{j_0}}, \frac{-t}{2^{j_0}})$, and then apply the vanishing moment from $\psi(w)$ and $L(t) \equiv \int_a^b I\{w \geq t\} \psi(w) dw$.

Q.E.D.

Lemma C.5 *Define for* $0 \le i, j \le p$

$$V_{(i,j)} = \int_0^1 Z_i(u) Z_j(u) \sigma^2 \left[F_V^{-1}(u) \right] du,$$

where

$$Z_i(u) = 2^{j_0} \iint\limits_0^1 I_{j_0}(t) \left[F_V^{-1}(w) - v_0 \right]^i I\{w \ge \tau\} \psi \left[2^{j_0}(w-t) \right] \psi \left[2^{j_0}(u-t) \right] dwdt.$$

Then

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

where for $0 \le i, j \le p$

$$\begin{aligned}
& V^*_{(i,j)} \\
= \quad & \sigma_-^2(v_0) \int_{a-b}^0 \left[\iint\limits_a^b \frac{I\{w-t \ge 0\}(w-t)^i \psi(w) \psi(u+t)}{f_V^i(v_0)} dwdt \right] \\
& \left[\iint\limits_a^b \frac{I\{w-t \ge 0\}(w-t)^j \psi(w) \psi(u+t)}{f_V^j(v_0)} dwdt \right] du \\
& + \\
& \sigma_+^2(v_0) \int_0^{b-a} \left[\iint\limits_a^b \frac{I\{w-t \ge 0\}(w-t)^i \psi(w) \psi(u+t)}{f_V^i(v_0)} dwdt \right] \\
& \left[\iint\limits_a^b \frac{I\{w-t \ge 0\}(w-t)^j \psi(w) \psi(u+t)}{f_V^j(v_0)} dwdt \right] du.
\end{aligned}$$

Proof: First let us look at the term $V_{(0,0)}$, where $S(u) = \iint_a^b \begin{bmatrix} I\{w-t\geq 0\} \\ \psi(w)\psi\left[2^{j_0}u - 2^{j_0}\tau + t\right] \end{bmatrix} dwdt$:

$$\begin{aligned}
& V_{(0,0)} \\
= & \int_0^1 \left[\iint_a^b I\{w-t\geq 0\}\psi(w)\psi\left[2^{j_0}u - 2^{j_0}\tau + t\right]\frac{1}{2^{j_0}}dwdt\right]^2 \sigma^2\left[F_V^{-1}(u)\right]du \\
= & \frac{1}{2^{2j_0}}\int_0^1 \left[S(2^{j_0}u - 2^{j_0}\tau)\right]^2 \sigma^2\left[F_V^{-1}(u)\right]du \\
= & \frac{1}{2^{3j_0}}\int_a^b S^2(u)\sigma^2\left[F_V^{-1}(\frac{u}{2^{j_0}}+\tau)\right]du \\
= & \frac{1}{2^{3j_0}}\left[\sigma_+^2(v_0)\int_0^b S^2(u)du + \sigma_-^2(v_0)\int_a^0 S^2(u)du + s.o.\right] \\
= & \frac{1}{2^{3j_0}}\begin{bmatrix} \sigma_+^2(v_0)\int_0^{b-a}\left(\iint_a^b I\{w-t\geq 0\}\psi(w)\psi(u+t)dwdt\right)^2 du \\ \sigma_-^2(v_0)\int_{a-b}^0\left(\iint_a^b I\{w-t\geq 0\}\psi(w)\psi(u+t)dwdt\right)^2 du \end{bmatrix} + s.o.,
\end{aligned}$$

where the last equality comes from the fact that $\iint_a^b I\{w \geq t+v\}\psi(w)\psi(t)dtdw$ has compact support $[a-b, b-a]$. By the same procedure, we could prove for other $V_{(i,j)}$ for $0 \leq i, j \leq p$.

Q.E.D.

Lemma C.6 *(Asymptotic equivalence between the feasible estimator and its counterpart) Observe the feasible local polynomial wavelets estimator* $\widehat{\delta}^{LP-SM}$:

$$\begin{aligned}
& \widehat{\delta}^{LP-SM} \\
= & \arg\min_{\{\delta_k\}_{k=0}^p} \sum_{l=1}^n \left[\widehat{\Delta}_{j_0}^Y(t_l) - \sum_{k=0}^p \delta_k \cdot \widehat{\Delta}_{j_0}^{\widehat{D}_k}(t_l)\right]^2 \widehat{I}_{j_0}(t_l) \\
= & \left[\left(\widehat{\Delta}_{j_0}^{\widehat{DP}}\right)^T \cdot \widehat{I}_{j_0} \cdot \widehat{\Delta}_{j_0}^{\widehat{DP}}\right]^{-1}\left[\left(\widehat{\Delta}_{j_0}^{\widehat{DP}}\right)^T \cdot \widehat{I}_{j_0} \cdot \widehat{\Delta}_{j_0}^Y\right],
\end{aligned}$$

where

$$\begin{aligned}
\widehat{I}_{j_0}(t_l) &= I\{a \leq 2^j(\widehat{\tau} - t_l) \leq b\}, \\
\widehat{I}_{j_0} &= diag\left[\widehat{I}_{j_0}(t_1), \widehat{I}_{j_0}(t_2), ..., \widehat{I}_{j_0}(t_n)\right], \\
\widehat{\Delta}_{j_0}^{\widehat{DP}}(t) &= \left[\widehat{\Delta}_{j_0}^{\widehat{D}_0}(t), \widehat{\Delta}_{j_0}^{\widehat{D}_1}(t), ..., \widehat{\Delta}_{j_0}^{\widehat{D}_p}(t)\right]^T \\
\left(\widehat{\Delta}_{j_0}^{\widehat{DP}}\right)^T &= \left[\left(\widehat{\Delta}_{j_0}^{\widehat{DP}}(t_1)\right)^T, \left(\widehat{\Delta}_{j_0}^{\widehat{DP}}(t_2)\right)^T, ..., \left(\widehat{\Delta}_{j_0}^{\widehat{DP}}(t_n)\right)^T)\right].
\end{aligned}$$

and its infeasible counterpart $\bar{\delta}^{LP-SM}$:

$$\begin{aligned}
& \bar{\delta}^{LP-SM} \\
= \ & \arg \min_{\{\delta_k\}_{k=0}^{p}} \sum_{l=1}^{n} \left[\widehat{\Delta}_{j_0}^{Y}(t_l) - \sum_{k=0}^{p} \delta_k \cdot \widehat{\Delta}_{j_0}^{D_k}(t_l) \right]^2 I_{j_0}(t_l) \\
= \ & \left[\left(\widehat{\Delta}_{j_0}^{DP} \right)^T \cdot I_{j_0} \cdot \widehat{\Delta}_{j_0}^{DP} \right]^{-1} \left[\left(\widehat{\Delta}_{j_0}^{DP} \right)^T \cdot I_{j_0} \cdot \widehat{\Delta}_{j_0}^{Y} \right],
\end{aligned}$$

where

$$\begin{aligned}
I_{j_0}(t_l) &= I\{a \leq 2^j(\tau - t_l) \leq b\}, \\
I_{j_0} &= diag\left[I_{j_0}(t_1), I_{j_0}(t_2), ..., I_{j_0}(t_n) \right], \\
\widehat{\Delta}_{j_0}^{DP}(t) &= \left[\widehat{\Delta}_{j_0}^{D_0}(t), \widehat{\Delta}_{j_0}^{D_1}(t), ..., \widehat{\Delta}_{j_0}^{D_p}(t) \right]^T \\
\left(\widehat{\Delta}_{j_0}^{DP} \right)^T &= \left[\left(\widehat{\Delta}_{j_0}^{DP}(t_1) \right)^T, \left(\widehat{\Delta}_{j_0}^{DP}(t_2) \right)^T, ..., \left(\widehat{\Delta}_{j_0}^{DP}(t_n) \right)^T) \right].
\end{aligned}$$

When $\left(n^{-1/2} (\log \log n)^{1/2} \right) / (2^{-j_0}) \to 0$*, then*

$$diag\left[\sqrt{\frac{n}{2^{j_0}}}, \sqrt{\frac{n}{2^{3j_0}}}, \cdots, \sqrt{\frac{n}{2^{(2p+1)j_0}}} \right] \left(\bar{\delta}^{LP-SM} - \widehat{\delta}^{LP-SM} \right) = o_p(\mathbf{1}).$$

Proof: First, let us look at the numerator terms $\left(\widehat{\Delta}_{j_0}^{\widehat{DP}} \right)^T \cdot \widehat{I}_{j_0} \cdot \widehat{\Delta}_{j_0}^{Y}$ and $\left(\widehat{\Delta}_{j_0}^{DP} \right)^T \cdot I_{j_0} \cdot \widehat{\Delta}_{j_0}^{Y}$ in these two estimators. Then the corresponding element-wise difference between these two numerators is defined as $diff_k$ for $0 \leq k \leq (p+1)$,

$$\begin{aligned}
diff_k = & \sum_{l=1}^{n} \left(\frac{1}{n} \sum_{i=1}^{n} (F_V^{-1}(t_i) - v_0)^k I\{t_i \geq \tau\} 2^{j_0/2} \psi \left[2^{j_0}(t_i - t_l) \right] \right) I_{j_0}(t_l) \widehat{\Delta}_{j_0}^{Y}(t_l) \\
& - \sum_{l=1}^{n} \left(\frac{1}{n} \sum_{i=1}^{n} (\widehat{F}_V^{-1}(t_i) - v_0)^k I\{t_i \geq \widehat{\tau}\} 2^{j_0/2} \psi \left[2^{j_0}(t_i - t_l) \right] \right) \widehat{I}_{j_0}(t_l) \widehat{\Delta}_{j_0}^{Y}(t_l).
\end{aligned}$$

Notice that

$$
\begin{aligned}
diff_0 &= \sum_{l=1}^{n}\left(\frac{1}{n}\sum_{i=1}^{n} I\{t_i \geq \tau\}2^{j_0/2}\psi\left[2^{j_0}(t_i-t_l)\right]\right) I_{j_0}(t_l)\widehat{\Delta}_{j_0}^{Y}(t_l) \\
&\quad -\sum_{l=1}^{n}\left(\frac{1}{n}\sum_{i=1}^{n} I\{t_i \geq \widehat{\tau}\}2^{j_0/2}\psi\left[2^{j_0}(t_i-t_l)\right]\right) \widehat{I}_{j_0}(t_l)\widehat{\Delta}_{j_0}^{Y}(t_l) \\
&= \left\{\begin{array}{c} \sum_{l=1}^{n}\left(\frac{1}{n}\sum_{i=1}^{n} I\{t_i \geq \tau\}2^{j_0/2}\psi\left[2^{j_0}(t_i-t_l)\right]\right) I_{j_0}(t_l)\widehat{\Delta}_{j_0}^{Y}(t_l) \\ -\sum_{i=1}^{n}\iint_0^1 I_{j_0}(t)I\{w\geq\tau\}\psi_{j_0}\left[w-t\right]\psi\left[2^{j_0}(t_i-t)\right]dtdw\cdot Y_{i:n} \end{array}\right\} \\
&\quad -\left\{\begin{array}{c} \sum_{l=1}^{n}\left(\frac{1}{n}\sum_{i=1}^{n} I\{t_i \geq \widehat{\tau}\}2^{j_0/2}\psi\left[2^{j_0}(t_i-t_l)\right]\right) \widehat{I}_{j_0}(t_l)\widehat{\Delta}_{j_0}^{Y}(t_l) \\ -\sum_{i=1}^{n}\iint_0^1 \widehat{I}_{j_0}(t)I\{w\geq\widehat{\tau}\}\psi_{j_0}\left[w-t\right]\psi\left[2^{j_0}(t_i-t)\right]dtdw\cdot Y_{i:n} \end{array}\right\} \\
&\quad +\left\{\begin{array}{c} \sum_{i=1}^{n}\iint_0^1 I_{j_0}(t)I\{w\geq\tau\}\psi_{j_0}\left[w-t\right]\psi\left[2^{j_0}(t_i-t)\right]dtdw\cdot Y_{i:n} \\ -\sum_{i=1}^{n}\iint_0^1 \widehat{I}_{j_0}(t)I\{w\geq\widehat{\tau}\}\psi_{j_0}\left[w-t\right]\psi\left[2^{j_0}(t_i-t)\right]dtdw\cdot Y_{i:n} \end{array}\right\} \\
&= O_p\left(\frac{1}{n2^{j_0}}\right)+O_p\left(\frac{1}{n2^{j_0}}\right),
\end{aligned}
$$

where the first two terms in the last equality are $O_p(\frac{1}{n2^{j_0}})$, which is derived from the proof of Theorem 7 in Chapter 1, and the third term $O_p(\frac{1}{n2^{j_0}})$ is from the change of variables.

Let $\Lambda = \left\{t : a \leq 2^{j_0}(\tau - t) \leq b\right\}$ and since

$$
\sup_{t_i\in\Lambda}\left|F_V^{-1}(t_i)-v_0\right| = O\left(\frac{1}{2^{j_0}}\right),
$$

and

$$
\sup_{t_i\in\Lambda}\left|\widehat{F_V^{-1}}(t_i)-v_0\right| \leq \sup_{t_i\in\Lambda}\left|F_V^{-1}(t_i)-v_0\right| + \sup_{t_i\in\Lambda}\left|\widehat{F_V^{-1}}(t_i)-F_V^{-1}(t_i)\right| = O_p\left(\frac{1}{2^{j_0}}\right),
$$

where the last equality comes from Equation (30), which is $\left|\widehat{F_V^{-1}}(t_i)-F_V^{-1}(t_i)\right| = O_p\left(n^{-1/2}\left(\log\log n\right)^{1/2}\right)$ in Wang and Cai (2010). Then $diff_{k+1} = O_p\left(\frac{1}{2^{j_0}}diff_k\right)$for each $0 \leq k \leq p$.

Now we have

$$
\begin{aligned}
& \left[\left(\widehat{\Delta}_{j_0}^{DP}\right)^T \cdot I_{j_0} \cdot \widehat{\Delta}_{j_0}^{DP}\right]^{-1}\left[\left(\widehat{\Delta}_{j_0}^{DP}\right)^T \cdot I_{j_0} \cdot \widehat{\Delta}_{j_0}^{Y} - \left(\widehat{\Delta}_{j_0}^{\widehat{DP}}\right)^T \cdot \widehat{I}_{j_0} \cdot \widehat{\Delta}_{j_0}^{Y}\right] \\
= \; & O_p \begin{bmatrix} 2^{2j_0} diff_0 + 2^{3j_0} diff_1 + \cdots + 2^{(p+2)j_0} diff_p \\ 2^{3j_0} diff_0 + 2^{4j_0} diff_1 + \cdots + 2^{(p+3)j_0} diff_p \\ \vdots \\ 2^{(p+2)j_0} diff_0 + 2^{(p+3)j_0} diff_1 + \cdots + 2^{2(p+1)j_0} diff_p \end{bmatrix} \\
= \; & O_p \begin{bmatrix} 2^{2j_0} diff_0 \\ 2^{3j_0} diff_0 \\ \vdots \\ 2^{(p+2)j_0} diff_0 \end{bmatrix}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& diag\left[\sqrt{\frac{n}{2^{j_0}}}, \sqrt{\frac{n}{2^{3j_0}}}, \cdots, \sqrt{\frac{n}{2^{(2p+1)j_0}}}\right] \cdot \\
& \left[\left(\widehat{\Delta}_{j_0}^{DP}\right)^T \cdot I_{j_0} \cdot \widehat{\Delta}_{j_0}^{DP}\right]^{-1}\left[\left(\widehat{\Delta}_{j_0}^{DP}\right)^T \cdot I_{j_0} \cdot \widehat{\Delta}_{j_0}^{Y} - \left(\widehat{\Delta}_{j_0}^{\widehat{DP}}\right)^T \cdot \widehat{I}_{j_0} \cdot \widehat{\Delta}_{j_0}^{Y}\right] \\
= \; & O_p\left(\sqrt{\frac{2^{j_0}}{n}}\right) = o_p\left(\mathbf{1}\right).
\end{aligned}
$$

In the end, using Sluskty theorem and $\left\{\left(\widehat{\Delta}_{j_0}^{DP}\right)^T \cdot I_{j_0} \cdot \widehat{\Delta}_{j_0}^{DP} - \left(\widehat{\Delta}_{j_0}^{\widehat{DP}}\right)^T \cdot \widehat{I}_{j_0} \cdot \widehat{\Delta}_{j_0}^{\widehat{DP}}\right\} = o_p\left(\mathbf{1}\right)$, we obtain the asymptotic equivalence between the infeasible and feasible estimators.

Q.E.D.

Proof of Theorem 15: First let us look at the $\left(\widehat{\Delta}_{j_0}^{DP}\right)^T \cdot I_{j_0} \cdot \widehat{\Delta}_{j_0}^{DP}$ term:

$$
\begin{aligned}
&\left(\widehat{\Delta}_{j_0}^{DP}\right)^T \cdot I_{j_0} \cdot \widehat{\Delta}_{j_0}^{DP} \\
= &\begin{bmatrix}
\sum_{i=1}^n \left[\widehat{\Delta}_{j_0}^{D_0}(t_i)\right]^2 I_{j_0}(t_i) & \sum_{i=1}^n \widehat{\Delta}_{j_0}^{D_0}(t_i)\widehat{\Delta}_{j_0}^{D_1}(t_i) I_{j_0}(t_i) & \dots & \sum_{i=1}^n \widehat{\Delta}_{j_0}^{D_0}(t_i)\widehat{\Delta}_{j_0}^{D_p}(t_i) I_{j_0}(t_i) \\
\sum_{i=1}^n \widehat{\Delta}_{j_0}^{D_0}(t_i)\widehat{\Delta}_{j_0}^{D_1}(t_i) I_{j_0}(t_i) & \sum_{i=1}^n \left[\widehat{\Delta}_{j_0}^{D_1}(t_i)\right]^2 I_{j_0}(t_i) & \dots & \sum_{i=1}^n \widehat{\Delta}_{j_0}^{D_1}(t_i)\widehat{\Delta}_{j_0}^{D_p}(t_i) I_{j_0}(t_i) \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^n \widehat{\Delta}_{j_0}^{D_0}(t_i)\widehat{\Delta}_{j_0}^{D_p}(t_i) I_{j_0}(t_i) & \sum_{i=1}^n \widehat{\Delta}_{j_0}^{D_p}(t_i)\widehat{\Delta}_{j_0}^{D_1}(t_i) I_{j_0}(t_i) & \dots & \sum_{i=1}^n \left[\widehat{\Delta}_{j_0}^{D_p}(t_i)\right]^2 I_{j_0}(t_i)
\end{bmatrix} \\
= &\begin{bmatrix}
\int_0^1 \left[\widehat{\Delta}_{j_0}^{D_0}(t)\right]^2 I_{j_0}(t)dt & \int_0^1 \widehat{\Delta}_{j_0}^{D_0}(t)\widehat{\Delta}_{j_0}^{D_1}(t) I_{j_0}(t)dt & \dots & \int_0^1 \widehat{\Delta}_{j_0}^{D_0}(t)\widehat{\Delta}_{j_0}^{D_p}(t) I_{j_0}(t)dt \\
\int_0^1 \widehat{\Delta}_{j_0}^{D_0}(t)\widehat{\Delta}_{j_0}^{D_1}(t) I_{j_0}(t)dt & \int_0^1 \left[\widehat{\Delta}_{j_0}^{D_1}(t)\right]^2 I_{j_0}(t)dt & \dots & \int_0^1 \widehat{\Delta}_{j_0}^{D_1}(t)\widehat{\Delta}_{j_0}^{D_p}(t) I_{j_0}(t)dt \\
\vdots & \vdots & \ddots & \vdots \\
\int_0^1 \widehat{\Delta}_{j_0}^{D_0}(t)\widehat{\Delta}_{j_0}^{D_p}(t) I_{j_0}(t)dt & \int_0^1 \widehat{\Delta}_{j_0}^{D_p}(t)\widehat{\Delta}_{j_0}^{D_1}(t) I_{j_0}(t)dt & \dots & \int_0^1 \left[\widehat{\Delta}_{j_0}^{D_p}(t)\right]^2 I_{j_0}(t)dt
\end{bmatrix} + s.o. \\
= &\ M + s.o.
\end{aligned}
$$

Second, let us look at the $\left(\widehat{\Delta}_{j_0}^{DP}\right)^T \cdot I_{j_0} \cdot \widehat{\Delta}_{j_0}^{G(\cdot)+\delta_0 I\{\cdot\geq\tau\}}$ term:

$$\left(\widehat{\Delta}_{j_0}^{DP}\right)^T \cdot I_{j_0} \cdot \widehat{\Delta}_{j_0}^{G(\cdot)+\delta_0 I\{\cdot\geq\tau\}}$$

$$= \begin{bmatrix} \sum_{i=1}^n \widehat{\Delta}_{j_0}^{D_0}(t_i) I_{j_0}(t_i) \widehat{\Delta}_{j_0}^{G(\cdot)+\delta_0 I\{\cdot\geq\tau\}}(t_i) \\ \sum_{i=1}^n \widehat{\Delta}_{j_0}^{D_1}(t_i) I_{j_0}(t_i) \widehat{\Delta}_{j_0}^{G(\cdot)+\delta_0 I\{\cdot\geq\tau\}}(t_i) \\ \vdots \\ \sum_{i=1}^n \widehat{\Delta}_{j_0}^{D_p}(t_i) I_{j_0}(t_i) \widehat{\Delta}_{j_0}^{G(\cdot)+\delta_0 I\{\cdot\geq\tau\}}(t_i) \end{bmatrix} = \begin{bmatrix} \int_0^1 \widehat{\Delta}_{j_0}^{D_0}(t) I_{j_0}(t) \widehat{\Delta}_{j_0}^{G(\cdot)+\delta_0 I\{\cdot\geq\tau\}}(t)dt \\ \int_0^1 \widehat{\Delta}_{j_0}^{D_1}(t) I_{j_0}(t) \widehat{\Delta}_{j_0}^{G(\cdot)+\delta_0 I\{\cdot\geq\tau\}}(t)dt \\ \vdots \\ \int_0^1 \widehat{\Delta}_{j_0}^{D_p}(t) I_{j_0}(t) \widehat{\Delta}_{j_0}^{G(\cdot)+\delta_0 I\{\cdot\geq\tau\}}(t)dt \end{bmatrix} + s.o.$$

$$= \begin{bmatrix} \int_0^1 \widehat{\Delta}_{j_0}^{D_0}(t) I_{j_0}(t) \widehat{\Delta}_{j_0}^{G^*}(t)dt + \delta_0 \int_0^1 \left[\widehat{\Delta}_{j_0}^{D_0}(t)\right]^2 I_{j_0}(t)dt + ... + \delta_p \int_0^1 \widehat{\Delta}_{j_0}^{D_0}(t) \widehat{\Delta}_{j_0}^{D_p}(t) I_{j_0}(t)dt \\ \int_0^1 \widehat{\Delta}_{j_0}^{D_1}(t) I_{j_0}(t) \widehat{\Delta}_{j_0}^{G^*}(t)dt + \delta_0 \int_0^1 \widehat{\Delta}_{j_0}^{D_0}(t) \widehat{\Delta}_{j_0}^{D_1}(t) I_{j_0}(t)dt + ... + \delta_p \int_0^1 \widehat{\Delta}_{j_0}^{D_1}(t) \widehat{\Delta}_{j_0}^{D_p}(t) I_{j_0}(t)dt \\ ... \\ \int_0^1 \widehat{\Delta}_{j_0}^{D_p}(t) I_{j_0}(t) \widehat{\Delta}_{j_0}^{G^*}(t)dt + \delta_0 \int_0^1 \widehat{\Delta}_{j_0}^{D_0}(t) \widehat{\Delta}_{j_0}^{D_p}(t) I_{j_0}(t)dt + ... + \delta_p \int_0^1 \left[\widehat{\Delta}_{j_0}^{D_p}(t)\right]^2 I_{j_0}(t)dt \end{bmatrix} + s.o.$$

$$= \begin{bmatrix} \int_0^1 \widehat{\Delta}_{j_0}^{D_0}(t) I_{j_0}(t) \widehat{\Delta}_{j_0}^{G^*}(t)dt \\ \int_0^1 \widehat{\Delta}_{j_0}^{D_1}(t) I_{j_0}(t) \widehat{\Delta}_{j_0}^{G^*}(t)dt \\ ... \\ \int_0^1 \widehat{\Delta}_{j_0}^{D_p}(t) I_{j_0}(t) \widehat{\Delta}_{j_0}^{G^*}(t)dt \end{bmatrix} + M \cdot \delta + s.o.$$

Now let us derive the asymptotic bias term using Lemma C.3 and Lemma C.4 and the asymptotic equivalence between the feasible and infeasible estimators from Lemma C.6:

$$\lim_{n\to\infty} diag\left[2^{(2m-1)j_0}, 2^{(2m-2)j_0}, ..., 2^{(2m-p-1)j_0}\right]\left[E(\widehat{\delta}^{LP-SM}) - \delta\right]$$

$$= \begin{bmatrix} G^{*(2m-1)}(\tau) \cdot (M^*)^{-1}_{(0,0)} N^*_{(0)} \\ G^{*(2m-1)}(\tau) \cdot (M^*)^{-1}_{(1,0)} N^*_{(0)} \\ ... \\ G^{*(2m-1)}(\tau) \cdot (M^*)^{-1}_{(p,0)} N^*_{(0)} \end{bmatrix}.$$

For the asymptotic variance term, we employ Lemma C.5, and the asymptotic equivalence between the feasible and infeasible estimators from Lemma C.6 [11]:

[11] Notice that, for the asymptotic variance term in Lemma C.5, the second term is smaller order than the first one. The proof is shown in Theorem 3 in Chapter 1.

$$\lim_{n\to\to\infty} n\cdot\Xi\cdot Var(\widehat{\delta}^{LP-SM}) = (M^*)^{-1}V^*(M^*)^{-1}.$$

In the end, the asymptotic normality of $\widehat{\delta}_0^{LP-SM}$ is established by following from Lemma C.2.

Q.E.D.

Proof of Theorem 16:

The proof follows the Cramer-Wold device to establish the joint limiting distribution of $(n/2^{j_0})^{1/2}(\widehat{\delta}_0^{LP-SM} - \delta_0)$ and $(n/2^{j_0})^{1/2}(\widehat{\zeta}_0^{LP-SM} - \zeta_0)$, then applies the Delta method to establish the asymptotic distribution for $\widehat{\delta}_0^{LP-SM}/\widehat{\zeta}_0^{LP-SM}$.

Q.E.D.

Proof of Corollary 1:

(1) and (2) are easy to obtain by substitution.

(3) The asymptotic properties of $\widehat{\delta}_0^{LC-SS}$ and $\widehat{\delta}_0^{LC-MM}$ are derived in Chapter 1. Here we list their results: if G is continuous but non-differentiable at τ, then

$$\lim_{n\to\infty} E\left(\widehat{\delta}_0^{LC-SS}\right) = \delta_0 + \left(\frac{1}{2^{j_0}}\right)\frac{\left[G_+^{(1)}(\tau) - G_-^{(1)}(\tau)\right]\int_0^b \psi(u)udu}{\int_0^b \psi(u)du} + o(\frac{1}{2^{j_0}}),$$

and

$$\begin{aligned}&\lim_{n\to\infty} E\left(\widehat{\delta}_0^{LC-MM}\right)\\ = \quad & \delta_0 + \left(\frac{1}{2^{j_0}}\right)\left(\frac{6}{7}\right)\frac{\left[G_+^{(1)}(\tau) - G_-^{(1)}(\tau)\right]\int_a^b\int_a^b L(t)\psi(s)(s-t)I\{s-t\geq 0\}dsdt}{\int_{a-b}^0 M(v)dv} + o(\frac{1}{2^{j_0}}),\end{aligned}$$

in which

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

It is seen that the $\widehat{\delta}_0^{LC-SS}$ and $\widehat{\delta}_0^{LC-MM}$ do not attain the optimal convergence rate when G is non-differentiable at τ. The reason for the non-optimality is because the non-differentiable of G at τ keeps us from using a two-sided Taylor expansion, so that around the cutoff point τ, only the one-sided Taylor expansions are available and introduce $G_+^{(1)}(\tau) - G_-^{(1)}(\tau)$ in the bias term. This fact leads us to model the potential higher-order derivative discontinuities of G at τ and results in local polynomial wavelet estimators.

However, if G is p-th differentiable at τ, then

$$\lim_{n\to\infty} E\left(\widehat{\delta}_0^{LC-SS}\right) = \delta_0 + (\frac{1}{2^{j_0}})^m\frac{G^{(m)}(\tau)\int_a^b u^m\psi(u)du}{\int_0^b\psi(u)du} + o\left((\frac{1}{2^{j_0}})^m\right),$$

and

$$\begin{aligned}&\lim_{n\to\infty} E\left(\widehat{\delta}_0^{LC-MM}\right)\\ = \quad & \delta_0 + (\frac{1}{2^{j_0}})^{2m-1}\frac{3}{4\left[1-(\frac{1}{2})^{2m+1}\right]}\frac{G^{(2m-1)}(\tau)\int_a^b\psi(s)s^m ds\int_a^b L(t)(-t)^{m-1}dt}{m!(m-1)!\int_{a-b}^0 M(v)dv} + o((\frac{1}{2^{j_0}})^{2m-1}).\end{aligned}$$

The asymptotic variance of $\widehat{\delta}_0^{LC-SS}$ is:

$$\left(\frac{2^{j_0}}{n}\right)\frac{\sigma_{\varepsilon+}^2(v_0)\int_0^b\psi^2(u)du + \sigma_{\varepsilon-}^2(v_0)\int_a^0\psi^2(u)du}{\left(\int_0^b\psi(u)du\right)^2},$$

and the asymptotic variance of $\widehat{\delta}_0^{LC-MM}$ is:

$$\left(\frac{2^{j_0}}{n}\right)\left(\frac{9}{14}\right)\frac{\sigma^2_{\varepsilon+}(v_0)\int_0^{b-a}M^2(v)dv+\sigma^2_{\varepsilon-}(v_0)\int_{a-b}^0 M^2(v)dv}{\left[\int_{a-b}^0 M(v)dv\right]^2}.$$

Q.E.D

Proof of Theorem 17:

(1) The equispaced Nadaraya-Watson estimator $\widehat{\delta}_0^{NW}$ could be written as

$$\widehat{\delta}_0^{NW}=\frac{1}{nh}\sum_{i=1}^{n}\widetilde{\psi}^{NW}(\frac{t_i-\tau}{h})Y_{i:n}+s.o.$$

for $t_i=i/n$ where $1\leq i\leq n$. To see this, consider

$$\begin{aligned}\widehat{\delta}_0^{NW} &\equiv \frac{\sum_{i=1}^n K(\frac{t_i-\widehat{\tau}}{h})I\{t_i\geq\widehat{\tau}\}Y_{i:n}}{\sum_{j=1}^n K(\frac{t_j-\widehat{\tau}}{h})I\{t_j\geq\widehat{\tau}\}}-\frac{\sum_{i=1}^n K(\tau-t_i)\left[1-I\{t_i\geq\widehat{\tau}\}\right]Y_{i:n}}{\sum_{j=1}^n K(\frac{t_j-\widehat{\tau}}{h})\left[1-I\{t_j\geq\widehat{\tau}\}\right]}\\ &= \frac{1}{nh}\sum_{i=1}^{n}\widetilde{\psi}^{NW}(\frac{t_i-\tau}{h})Y_{i:n}+s.o.,\end{aligned}$$

where

$$\widetilde{\psi}^{NW}(t)\equiv\frac{K(t)I\{t\geq 0\}}{\int_0^a K(t)dt}-\frac{K(t)\left[1-I\{t\geq 0\}\right]}{\int_{-a}^0 K(t)dt}.$$

The equivalent wavelet function $\widetilde{\psi}^{NW}$ satisfies $\int_{-a}^a\widetilde{\psi}^{NW}(u)du=0$ and $\int_{-a}^a u\widetilde{\psi}^{NW}(u)du\neq 0$, which only has one vanishing moment.

The equispaced partial smoothing kernel estimator[12] $\widehat{\delta}_0^{ES}$ could be written as $\widehat{\delta}_0^{ES}=\frac{1}{nh}\sum_{i=1}^n\widetilde{\psi}^{ES}(\frac{t_i-\widehat{\tau}}{h})Y_{i:n}+s.o.$,for $t_i=i/n$ where $1\leq i\leq n$ and

$$\begin{aligned}\widetilde{\psi}^{ES}(t) &\equiv \left\{I\{t\geq\widehat{\tau}\}-\int_{-4a}^{4a}\widetilde{K}(t-v)I\{v\geq\widehat{\tau}\}dv\right\}/B^{ES},\\ \widetilde{K}(t) &\equiv 2K(t)-\int K(t-v)K(v)dv,\\ B^{ES} &\equiv \int_{-2a}^{2a}\left[\widetilde{\psi}^{ES}(t)\right]^2dt.\end{aligned}$$

$\widetilde{\psi}^{ES}$ has the $(2m-1)$ vanishing moment[13]. Hence the better asymptotic bias performance of $\widehat{\delta}_0^{ES}$ is expected when H is smoother.

[12]The similar results apply to Eubank and Whitney (1989), where the equivalent wavelet function has m vanishing moment.

[13]These arguments could be found in Lemma 2, 3, and 6 in Speckman (1994), though he did not consider the implications of $\widetilde{\psi}^{ES}$ integrating to zero or the relationship to wavelet estimators.

For the equispaced profiled partial linear estimator $\widehat{\delta}_0^{PO}$, we see:

$$\begin{aligned}
\widehat{\delta}_0^{PO} &\equiv \arg\min_{\delta_0} \sum_{i=1}^{n} \left[Y_{i:n} - \delta_0 I\{t_i \geq \widehat{\tau}\} - \sum_{j=1}^{n} \frac{K(\frac{t_i - t_j}{h})}{\sum_{l=1}^{n} K(\frac{t_i - t_l}{h})} \left[Y_{j:n} - \delta_0 I\{t_j \geq \widehat{\tau}\} \right] \right]^2 \\
&= \arg\min_{\delta_0} \sum_{i=1}^{n} \left[\widehat{\Delta}_{j_0}^{Y}(t_i) - \delta_0 \widehat{\Delta}_{j_0}^{D_0}(t_i) - \widehat{\Delta}_{j_0}^{\widehat{G}}(t_i) \right]^2 \\
&= \arg\min_{\delta_0} \sum_{\substack{i=1, \\ t_i \in \left\{ t : a \leq 2^{j_0}(\widehat{\tau} - t) \leq b \right\}}}^{n} \left[\widehat{\Delta}_{j_0}^{Y}(t_i) - \delta_0 \widehat{\Delta}_{j_0}^{D_0}(t_i) - \widehat{\Delta}_{j_0}^{\widehat{G}}(t_i) \right]^2 \\
&\quad + \sum_{\substack{i=1, \\ t_i \notin \left\{ t : a \leq 2^{j_0}(\widehat{\tau} - t) \leq b \right\}}}^{n} \left[\widehat{\Delta}_{j_0}^{Y}(t_i) - \delta_0 \widehat{\Delta}_{j_0}^{D_0}(t_i) - \widehat{\Delta}_{j_0}^{\widehat{G}}(t_i) \right]^2 \\
&= \arg\min_{\delta_0} \sum_{i=1}^{n} \left[\widehat{\Delta}_{j}^{Y}(t_i) - \delta_0 \widehat{\Delta}_{j_0}^{D_0}(t_i) - s.o. \right]^2 \widehat{I}_{j_0}(t_i) + s.o.,
\end{aligned}$$

where

$$\widehat{G}(t) \equiv \sum_{j=1}^{n} \frac{K(\frac{t - t_j}{h})}{\sum_{l=1}^{n} K(\frac{t - t_l}{h})} \left[Y_j - \delta_0 I\{t_j \geq \widehat{\tau}\} \right].$$

The second equality comes from the property $W^T W = I/n + s.o.$, where

$$W \equiv \frac{2^{j_0/2}}{n} \left[\psi(2^{j_0} t_i - 2^{j_0} t_j) \right]_{i,j=1,n}$$

and $\left\{ \psi\left(2^{j_0} t - w\right); w \in \mathbb{Z} \right\}$ constitute an orthonormal basis (Chapter 5 in Daubechies,1992). The first term in the last equality is because G is close to $\widehat{G}$ when the bandwidth h goes to zero, so that $\widehat{\Delta}_{j_0}^{\widehat{G}}(t_i)$ is small because G is continuous. For the second term in the last equality when $t_i \notin \left\{ t : a \leq 2^{j_0}(\widehat{\tau} - t) \leq b \right\}$, the wavelet coefficient $\widehat{\Delta}_{j_0}^{D_0}(t_i)$ is also small due to D_0 being a constant, so that the second term no longer has the argument δ_0. Therefore, the profiled partial linear estimator $\widehat{\delta}_0^{PO}$ (approximately) shares the same objective function as local constant wavelet estimators.[14]

[14] Yu (2010) proposed a partial polynomial kernel estimator that could be asymptotically expressed as local polynomial wavelet estimator.

(2) The equispaced local polynomial kernel estimator[15] $\widehat{\delta}_0^{LP}$ is estimated as

$$\begin{aligned}
\widehat{\delta}_0^{LP} &= \arg \min_{\{\beta_j\}_{j=0}^p, \{\delta_k\}_{k=0}^p} \frac{1}{n}\sum_{i=1}^n \left[Y_{i:n} - \sum_{j=0}^p \beta_j (t_i - \widehat{\tau})^j - \sum_{k=0}^p \delta_k (\widehat{F}_V^{-1}(t_i) - v_0)^k I\{t_i \geq \widehat{\tau}\} \right]^2 K(\frac{t_i - \widehat{\tau}}{h}) \\
&= \arg \min_{\{\beta_j\}_{j=0}^p, \{\delta_k\}_{k=0}^p} \frac{1}{n}\sum_{i=1}^n \left[\widehat{\Delta}_{j_0}^{Y}(t_i) - \widehat{\Delta}_{j_0}^{G^P}(t_i) - \sum_{k=0}^p \delta_k \cdot \widehat{\Delta}_{j_0}^{\widehat{D}_k}(t_i) \right]^2 K(\frac{t_i - \widehat{\tau}}{h}) \\
&= \arg \min_{\{\beta_j\}_{j=0}^p, \{\delta_k\}_{k=0}^p} \sum_{\substack{i=1, \\ t_i \in \{t: a \leq 2^{j_0}(\widehat{\tau} - t) \leq b\}}}^n \left[\widehat{\Delta}_{j_0}^{Y}(t_i) - \widehat{\Delta}_{j_0}^{G^P}(t_i) - \sum_{k=0}^p \delta_k \cdot \widehat{\Delta}_{j_0}^{\widehat{D}_k}(t_i) \right]^2 K(\frac{t_i - \widehat{\tau}}{h}) \\
&\quad + \sum_{\substack{i=1, \\ t_i \notin \{t: a \leq 2^{j_0}(\widehat{\tau} - t) \leq b\}}}^n \left[\widehat{\Delta}_{j_0}^{Y}(t_i) - \widehat{\Delta}_{j_0}^{G^P}(t_i) - \sum_{k=0}^p \delta_k \cdot \widehat{\Delta}_{j_0}^{\widehat{D}_k}(t_i) \right]^2 K(\frac{t_i - \widehat{\tau}}{h}) \\
&= \arg \min_{\{\delta_k\}_{k=0}^p} \sum_{i=1}^n \left[\widehat{\Delta}_{j_0}^{Y}(t_i) - \sum_{k=0}^p \delta_k \cdot \widehat{\Delta}_{j_0}^{\widehat{D}_k}(t_i) - s.o. \right]^2 \widehat{I}_{j_0,h}(t_i) + s.o.,
\end{aligned}$$

where

$$\begin{aligned}
G^p(t) &\equiv \sum_{j=0}^p \beta_j (t - \widehat{\tau})^j, \\
\widehat{I}_{j_0,h}(t) &\equiv I\{a \leq 2^{j_0}(\widehat{\tau} - t) \leq b\} K(\frac{t - \widehat{\tau}}{h}).
\end{aligned}$$

The second equality comes from the property $W^T W = I/n + s.o.$, where $W \equiv \frac{2^{j_0/2}}{n}\left[\psi(2^{j_0}t_i - 2^{j_0}t_j)\right]_{i,j=}$ and $\left\{\psi\left(2^{j_0}t - w\right); w \in \mathbb{Z}\right\}$ constitute an orthonormal basis (Chapter 5 in Daubechies,1992). The first term in the last equality is because G^p is the p-th polynomial, so that $\widehat{\Delta}_{j_0}^{G^p}(t_i)$ is close to 0. For the second term in the last equality, when $t_i \notin \left\{t : a \leq 2^{j_0}(\widehat{\tau} - t) \leq b\right\}$, the wavelet coefficients $\left\{\widehat{\Delta}_{j_0}^{\widehat{D}_k}(t_i)\right\}_{k=0}^p$ go to zero because $\{D_k\}_{k=0}^p$ is continuous. Therefore the second term is independent of the arguments $\{\delta_k\}_{k=0}^p$. Notice that $\widehat{I}_{j_0,h}(t)$ is a more general weighting function than $\widehat{I}_{j_0}(t)$.

Q.E.D.

[15] The local polynomial kernel curve estimator, based on the equispaced data still, has the automatic boundary corrections discussed on page 473 of Hall, et al. (1998).

Proofs of Chapter 3

Proof of Theorem 19: Let $\eta_j = median\left(\{\xi_i : (j-1)m+1 \leq i \leq jm\}\right)$. We define $Z_j = \frac{1}{f_\xi(0)}\Phi^{-1}\left(G(\eta_j)\right)$ where G is the distribution of η_j. It follows from the local median coupling theorem that $\sqrt{4m}\eta_j$ is well approximated by Z_j whose distribution is $N(0, \frac{1}{f_\xi^2(0)})$.

Set

$$\zeta_j = \sqrt{m}\eta_j - \frac{Z_j}{2}.$$

Then ζ_j is the error of approximating the median by the Gaussian variable. According to Lemma C.1 and C.2, a bound for the approximation error ζ_j is given by

$$\left|\zeta_j\right| \leq \frac{C}{m^{1/2}}\left(1+\left|Z_j\right|^2\right) \text{ when } |Z_j| \leq \varepsilon\sqrt{m}$$

for some $\varepsilon > 0$, and the probability of $|Z_j| > \varepsilon\sqrt{m}$ is exponentially small. Hence for any finite integer $l \geq 1$ (here l is fixed and $m = n^\gamma \to \infty$),

$$\begin{aligned} E\left|\zeta_j\right|^l &= E\left|\zeta_j\right|^l \left\{|Z_j| \leq \varepsilon\sqrt{m}\right\} + E\left|\zeta_j\right|^l \left\{|Z_j| > \varepsilon\sqrt{m}\right\} \\ &\leq C_l m^{-l/2} + \left(E\left|\zeta_j\right|^{2l}\right)^{1/2}\left[P\left\{|Z_j| > \varepsilon\sqrt{m}\right\}\right]^{1/2} \end{aligned}$$

for some constant $C_l > 0$, where

$$P\left\{|Z_j| > \varepsilon\sqrt{m}\right\} \leq \frac{1}{2}\exp(-\frac{\varepsilon^2}{2}m).$$

By Mill's ratio inequality

$$\frac{\varphi(x)}{1-\Phi(x)} > \max\left\{x, \frac{2}{\sqrt{2\pi}}\right\} \geq \frac{1}{2}\left(x + \frac{2}{\sqrt{2\pi}}\right) \text{ for } x > 0$$

and

$$E\left|\sqrt{m}\eta_j\right|^{2l} \leq m^l E\left|\eta_j\right|^{2l} \leq D_l m^l$$

for some constant $D_l > 0$, so we have

$$E\left|\zeta_j\right|^l \leq C_l m^{-l/2}.$$

Assumption 1 implies

$$P(|\xi_i| \geq |x|) \leq \frac{C}{|x|^{\epsilon_3}}.$$

For $m = 2v + 1$ i.i.d. ξ_i, the density of the sample median is

$$\begin{aligned} g(x) &= \frac{\sqrt{8v}}{\sqrt{2\pi}} \left[4F_\xi(x)(1 - F_\xi(x))\right]^v f_\xi(x) \exp\left(O\left(\frac{1}{v}\right)\right) \\ &\leq \frac{\sqrt{8v}}{\sqrt{2\pi}} \left[\frac{4C}{|x|^{\epsilon_3}}\right]^v f_\xi(x) \exp\left(O\left(\frac{1}{v}\right)\right) \\ &= \frac{\sqrt{8v}}{\sqrt{2\pi}} \left[\frac{4C}{|x|^{\epsilon_3/2}}\right]^v \frac{1}{|x|^{v\epsilon_3/2}} f_\xi(x) \exp\left(O\left(\frac{1}{v}\right)\right). \end{aligned}$$

When $|x|^{\epsilon_3/2} \geq 8C$, we have

$$\frac{\sqrt{8v}}{\sqrt{2\pi}} \left[\frac{4C}{|x|^{\epsilon_3/2}}\right]^v \leq \frac{\sqrt{8v}}{\sqrt{2\pi}2^v}$$

which is bounded for all v. This implies as $v \to \infty$ ($m \sim n^\gamma$ in our procedure) the median has any finite moments.

Thus we have

$$E\left|\zeta_j\right|^l \leq 2^{l-1}\left(E\left|\zeta_{2j}\right|^l\right) \leq C_l m^{-l/2}.$$

The inequality $P(\left|\zeta_j\right| > a) \leq C_l(a^2 m)^{-l/2}$ then follows from Chebyshev's inequality.
Q.E.D.

Lemma C.1 *(Zhou, 2006) Let Z be a standard normal random variable. Let S_n be a random variable with a distributed function $G(x) = P(S_n \leq x)$. Assume that there is a positive ε such that for all n,*

$$\begin{aligned} P(S_n &< -x) = \Phi(-x)\exp\left(O(n^{-1}x^4 + n^{-1})\right), \\ 1 - P(S_n &< x) = \overline{\Phi}(x)\exp\left(O(n^{-1}x^4 + n^{-1})\right), \end{aligned}$$

where $\overline{G}(x) = 1 - G(x)$, and $\overline{\Phi}(x) = 1 - \Phi(x)$, and $O(n^{-1}x^4 + n^{-1})$ is uniform on the interval $x \in [0, \varepsilon\sqrt{n}]$. And the expression above holds when " $<$ " is replaced by " $\leq$ ". Then for every n, there is a random variable $\widetilde{S_n}$ with $L(\widetilde{S_n}) = L(S_n)$ such that

$$\left|\widetilde{S_n} - Z\right| \leq \frac{C_1}{n} + \frac{C_1}{n}\left|\widetilde{S_n}\right|^3$$

for $\left|\widetilde{S_n}\right| \leq \varepsilon_1\sqrt{n}$, where $C_1, \varepsilon_1 > 0$ do not depend on n. Or equivalently

$$\left|\widetilde{S_n} - Z\right| \leq \frac{C}{n}(1 + |Z|)^3$$

for $|Z| \leq \varepsilon\sqrt{n}$, where $C, \varepsilon > 0$ do not depend on n.

Theorem 23 ***Lemma C.2*** *(Brown, Cai and Zhou, 2008) Let Z be a standard normal variable and let $Y_1, ..., Y_n$ be i.i.d. with density function h where $n = 2k+1$ for some integer $k \geq 1$. Let Assumptions 1 and 2 hold. Then for every n there is a mapping $\widetilde{Y}_{med} : R \to R$*

such that $L(\widetilde{Y}_{med}(Z)) = L(Y_{med})$ *and*

$$\begin{aligned}\left|\sqrt{4nh(0)}\widetilde{Y}_{med} - Z\right| &\leq \frac{C}{\sqrt{n}} + \frac{C}{\sqrt{n}}\left|\sqrt{4nf_{\xi}(0)}\widetilde{Y}_{med}\right|^2 \\ \text{when } \left|\widetilde{Y}_{med}\right| &\leq \varepsilon\end{aligned}$$

where $C, \varepsilon > 0$ *depend on* f_{ξ} *but not on* n.

Proof of Theorem 17: First given δ_0, we define two experiments

$$\begin{aligned} E_n^* &: Y_i - \delta_0 I(x_i \geq x_0) = f(x_i) + \xi_i, \text{ where } median(\xi_i) = 0 \text{ and } i = 1, ..., n, \\ F_n^* &: X_j - \delta_0 I(x_j \geq x_0) = f(x_j) + \frac{1}{2h(0)\sqrt{m}} Z_j, \text{ where } Z_j \overset{i.i.d.}{\sim} (0,1) \text{ and } j = 1, ..., T. \end{aligned}$$

Then from Theorem 1 and 2 in Cai and Zhou(2009), we know two experiments F_n^* and E_n^* are asymptotically equivalent with respect to the set of procedures Λ_n^* and set of loss functions Γ_n, where Λ_n^* are meant to be the estimates of the function f.

Next since we could consistently estimate the jump size δ_0 from F_n from local polynomial wavelet estimators, we define the experiment

$$F_n^{**} : X_j = f(x_j) + \widehat{\delta}_0 I(x_j \geq x_0) + \frac{1}{2h(0)\sqrt{m}} Z_j, \text{ where } Z_j \overset{i.i.d.}{\sim} (0,1) \text{ and } j = 1, ..., T.$$

Therefore we have two experiments F_n^{**} and F_n^* are asymptotically equivalent, so that two experiments E_n^* and F_n^{**} are asymptotically equivalent and two experiments E_n^* and F_n are asymptotically equivalent.

In the end the two experiments E_n and E_n^* are actually defined for the same probability same for ξ_1, so that they are trivially asymptotically equivalent.

According to Lemma C.1 and the asymptotic equivalence, our two-step estimator $\widehat{\delta}_0^{LP-med}$ has the optimal rate of convergence under the discontinuous nonparametric median model E_n.

Q.E.D.

Lemma C.1 *(Porter, 2003) Under Assumptions 1-6. Then for some positive constant* D *and a small* $\epsilon > 0$:

$$\lim \inf_{n\to\infty} \inf_{\widetilde{\delta}_0} \sup_{\theta \in \Theta} P_\theta \left[n^{\frac{p}{2p+1}} \left| \widetilde{\delta}_0 - \delta_0(\theta) \right| \geq \epsilon \right] \geq D.$$

In other words, the optimal convergence rate for the jump size estimation under discontinuous mean model is $n^{-\frac{p}{2p+1}}$.

BIBLIOGRAPHY

[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] Bhattacharya, P. K. and Gangopadhyay, A. K. (1990), "Kernel and Nearest-neighbor Estimation of A Conditional Quantile," *Annals of Statistics*, p1400-1415.

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

[6] Brown, L. D. and Low, M. G. (1996), "Asymptotic Equivalence of Nonparametric Regression and White Noise," *Annals of Statistics*, p2384-2398.

[7] Brown, L. D., Cai, T. T., Low, M. G. and Zhang, C. H. (2002), "Asymptotic Equivalence Theory for Nonparametric Regression with Random Design," *Annals of Statistics*, p688-707.

[8] Brown, L. D., Cai, T. T. and Zhou, H. H. (2008), "Robust Nonparametric Estimation via Wavelet Median Regressio,". *Annals of Statistics*, p2055–2084.

[9] Brown, L. D., Cai, T. T. and Zhou, H. H. (2010), "Nonparametric Regression in Exponential Families," *Annals of Statistics*, p2005-2046.

[10] Cai, T. T. and Zhou, H. H. (2009), "Asymptotic Equivalence and Adaptive Estimation for Robust Nonparametric Regression," *Annals of Statistics*, p3204-3235.

[11] 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.

[12] Chaudhuri, P. (1991), "Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation," *Annals of Statistics*, p760-777.

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

[14] Cheng, M.Y., Peng, L. and Wu, J.S. (2007), "Reducing Variance in Univariate Smoothing," *Annals of Statistics*, Vol. 35, 522-542.

[15] 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.

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

[17] Chen, H. (2012), "Local Polynomial Wavelet Estimation of the Local Average Treatment Effect," Working Paper, Vanderbilt University.

[18] Chen, H. and Fan, Y. (2011), "Identification and Wavelet Estimation of the LATE in a Class of Switching Regime Models," Working Paper, Vanderbilt University.

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

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

[21] Choi, E. and Hall, P. (1998), "On Bias Reduction in local constant Smoothing," *Biometrika*, Vol. 85, 333-345.

[22] 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.

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

[24] Daubechies, I. (1992), "Ten Lectures on Wavelets," CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM.

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

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

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

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

[29] 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.

[30] Eubank, R.L. (1999), *Nonparametric Regression and Spline Smoothing*, Second Edition, Marcel Dekker, Inc.

[31] Fan, J. (1992), “Design-adaptive Nonparametric Regression,” *Journal of the American Statistical Association*, Vol. 87, 998-1004.

[32] 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.

[33] Frandsen, B. Frolich, M. and Melly, B. (2011), "Quantile Treatments Effects in the Regression Discontinuity Design," Working Paper, Brown University.

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

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

[42] Hall, P and Marron, J.S. (1997), “On the Role of the Shrinkage Parameter in local constant Smoothing,” *Probability Theory and Related Fields,* Vol. 108, 495-516.

[43] Härdle, W., Kerkyacharian, G., Picard, D., Tsybakov, A. (2000) “Wavelets, Approximation, and Statistical Applications,” Springer.

[44] He, H. and Huang, L. (2009), “Double-smoothing for Bias Reduction in local constant Regression,” *Journal of Statistical Planning and Inference,* Vol. 139, 1056-1072.

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

[46] Heckman, J., Tobias, J.L. and Vytlacil, E. (2003), “Simple Estimators for Treatment Parameters in a Latent-variable Framework,” *Review of Economics and Statistics,* Vol. 85, 748-755.

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

[48] 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.

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

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

[51] 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.

[52] Jung, S. H. (1996), "Quasi-Likelihood for Median Regression Models," *Journal of the American Statistical Association* 91, p251-257.

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

[54] Komlos, J., Major, P. and Tusnady, G. (1975), "An Approximation of Partial Sums of Independent rv’s and the Sample," *Probability Theory and Related Fields*, p111-131.

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

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

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

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

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

[60] 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.

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

[62] 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.

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

[64] Marron, J.S. and Wand, M.P. (1992), "Exact Mean Integrated Squared Error," *Annals of Statistics,* Vol. 20, 712-736.

[65] Mason, D. and Zhou, H. H. (2012), "Quantile Coupling Inequalities and Their Applications," *Probability Surveys.*

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

[67] Mielniczuk, J., Sarda P. and Vieu. P. (1989), "Local Data-driven Bandwidth Choice for Density Estimation," *Journal of Statistical Planning and Inference,* Vol. 23, 53-69.

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

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

[70] Nussbaum, M. (1996), "Asymptotic Equivalence of Density Estimation and Gaussian White Noise," *Annals of Statistics*, p2399-2430.

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

[72] 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.

[73] 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.

[74] Porter, J. (2003), "Estimation in the Regression Discontinuity Model," manuscript, Harvard University.

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

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

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

[78] Sakhanenko, A. I. (1996), "Estimates for the Accuracy of Coupling in the Central Limit Theorem," *Siberian Mathematical Journal* 37, p811-823.

[79] 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.

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

[81] Speckman, P. (1994), "Detection of Change-Points in Nonparametric Regression," Working Paper, University of Missouri at Columbia.

[82] 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.

[83] 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.

[84] Stone, C. (1984), "An Asymptotically Optimal Window Selection Rule for Kernel Density Estimates,", *Annals of Statistics,* Vol. 12, 1285-1297.

[85] Sun, Y. (2005), "Adaptive Estimation of the Regression Discontinuity Model," Working Paper, UCSD.

[86] Truong, K. Y. (1989), "Asymptotic Properties of Kernel Estimation Based on Local Medians," *Annals of Statistics*, p606-617.

[87] Tsybakov, A.B. (2010) *Introduction to Nonparametric Estimation,* Springer Press.

[88] Vieu, P. (1991), "Nonparametric Regression: Optimal Local Bandwidth Choice," *Journal of the Royal Statistical Society. Series B (Methodological),* Vol. 53, 453-464.

[89] Vytlacil, E. (2002), "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica,* Vol. 70, 331-341.

[90] 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.

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

[92] 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.

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

[94] Wang, Y. (1998), “Change Curve Estimation via Wavelets,” *Journal of the American Statistical Association*, Vol. 93, 163-172.

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

[96] 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.

[97] Yang, S.S. (1977), “Linear Functions of Concomitants of Order Statistics,” Technical Report No.7, MIT, Department of Mathematics.

[98] 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.

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

[100] Yu, K. and Jones, M.C. (1998), "Local Linear Quantile Regression," *Journal of the American Statistical Association* 93, p228-237.

[101] Yu, P. (2010), “Partially Polynomial Estimation in Regression Discontinuity,” Working Paper, University of Auckland.

[102] 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.

[103] Zheng, Z. G. and Yang, Y. (1998), "Cross-validation and Median Criterion," *Statistica Sinica*, p907-921.

[104] Zhou, H. H. (2006), "A note on quantile coupling inequalities and their applications," Submitted. Available at www.stat.yale.edu/~hz68.

Table 1. Switching Regime Model 1 under Assumption A2 (G) (a)

$n = 500$						
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. Switching Regime Model 2 under Assumption A2 (G) (b)

$n = 500$						
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. Auxiliary Regression Model 3 under Assumption A2 (G) (a)

$n = 500$						
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. Auxiliary Regression Model 4 under Assumption A2 (G) (b)

$n = 500$						
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

Table 5. Jump Size Model under the Sample Size 500

$n = 500$						
Scale	1	2	3	4	5	6
			Zeta0			
Bias	1.8269	1.7808	0.7111	-0.0356	0.0093	0.0084
Std	0.0106	0.0223	0.0359	0.0617	0.0859	0.1221
MSE	3.3376	3.1718	0.5069	0.0050	0.0074	0.0149
			Zeta1			
Bias	0.1201	0.0103	0.0270	-0.0172	0.0077	0.0105
Std	0.0166	0.0296	0.0408	0.0622	0.0855	0.1218
MSE	0.0147	0.0009	0.002	0.0041	0.0073	0.0149
			Zeta2			
Bias	0.5757	0.0577	0.0280	-0.0129	0.0091	0.0117
Std	0.0273	0.0308	0.0408	0.0708	0.0848	0.1233
MSE	0.3322	0.0042	0.0024	0.0051	0.0072	0.0153
			Zeta3			
Bias	2.2004	0.1498	0.0391	-0.0161	0.0095	0.0120
Std	0.0557	0.0381	0.0471	0.0846	0.0892	0.1231
MSE	4.8449	0.0239	0.0037	0.0074	0.0080	0.0153

Table 6. Jump Size Model under the Sample Size 2500

$n = 2500$						
Scale	1	2	3	4	5	6
			Zeta0			
Bias	1.8132	1.8547	0.4343	-0.0038	0.0023	-0.0116
Std	0.0041	0.0112	0.0173	0.0236	0.0337	0.0513
MSE	3.2878	3.4401	0.1889	0.0005	0.0011	0.0027
			Zeta1			
Bias	0.1181	-0.0028	0.0267	0.0021	0.0007	-0.0187
Std	0.0079	0.0134	0.0184	0.0238	0.0336	0.0567
MSE	0.0140	0.0001	0.0010	0.0005	0.0011	0.0035
			Zeta2			
Bias	0.5819	0.0387	0.0262	0.0065	0.0032	-0.0167
Std	0.0123	0.0139	0.0184	0.0254	0.0337	0.0577
MSE	0.3388	0.0016	0.0010	0.0006	0.0011	0.0036
			Zeta3			
Bias	2.0359	0.1315	0.0239	0.0095	0.0047	-0.0162
Std	0.0241	0.0177	0.0207	0.0300	0.0380	0.0574
MSE	4.1456	0.0176	0.0010	0.0009	0.0014	0.0035

Table 7. Jump Size Model under the Sample Size 5000

			$n = 5000$		
Scale	1	2	3	4	5
			Zeta0		
Bias	1.8611	1.7946	0.7579	-0.0111	0.0041
Std	0.0031	0.0071	0.0122	0.0196	0.0250
MSE	3.4639	3.2207	0.5746	0.0005	0.0006
			Zeta1		
Bias	0.1203	-0.0203	0.0110	0.0050	0.0001
Std	0.0050	0.0096	0.0141	0.0195	0.0249
MSE	0.0145	0.0005	0.0003	0.0003	0.0006
			Zeta2		
Bias	0.5936	0.0683	0.0193	-0.0023	0.0027
Std	0.0095	0.0100	0.0142	0.0243	0.0259
MSE	0.3525	0.0047	0.0005	0.0005	0.0006
			Zeta3		
Bias	2.1118	0.1671	0.0274	-0.0032	0.0028
Std	0.0175	0.0134	0.0170	0.0260	0.0282
MSE	4.4601	0.0281	0.0010	0.0006	0.0008

Table 8. Kink Size Model under the Sample Size 500

			$n = 500$			
Scale	1	2	3	4	5	6
			Kink_012			
Bias	-2.0714	0.6654	0.2618	0.2930	0.2010	0.5706
Std	0.0458	0.0423	0.1797	0.5078	1.7162	2.4197
MSE	4.2931	0.4446	0.1008	0.3437	2.9858	6.1807
			Kink_01			
Bias	1.5550	0.6157	0.2781	0.3420	0.2727	0.7417
Std	0.0213	0.0355	0.1771	0.5171	1.7496	2.1388
MSE	2.4185	0.3804	0.1087	0.3843	3.1357	5.1246
			Kink_12			
Bias	0.5013	1.0613	1.0118	0.3031	0.7470	1.0703
Std	0.0094	0.0119	0.0297	0.4998	1.2915	2.0182
MSE	0.2514	1.1266	1.0248	0.3417	2.2261	5.2190
			Kink_1			
Bias	0.8913	1.0223	1.0268	0.3959	0.7647	1.0462
Std	0.0060	0.0086	0.0255	0.5048	1.3004	2.0244
MSE	0.7945	1.0453	1.0550	0.4116	2.2760	5.1933

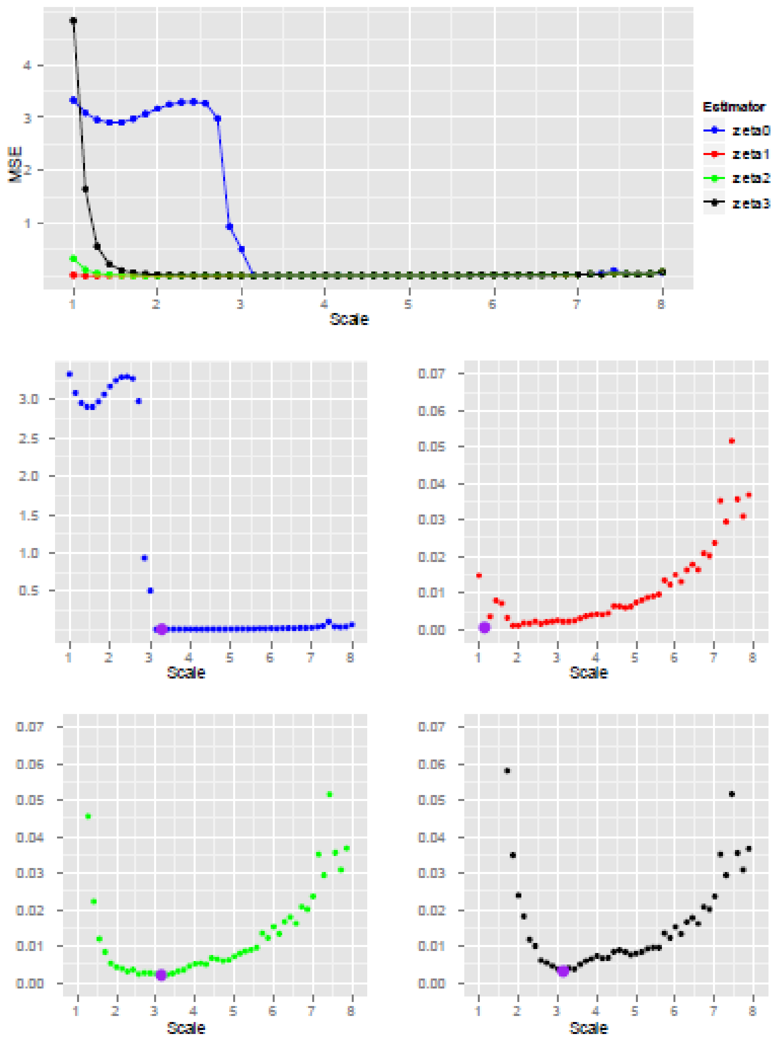

Figure 1. Local polynomial wavelet estimators under the sample size 500 for the jump size across different scales j_0. (Top) MSE comparisons among different polynomial orders; (Middle Left) Zeta0, the single-scale local constant wavelet estimator; (Middle Right) Zeta1, the single-scale local linear wavelet estimator; (Bottom Left) Zeta2, the single-scale local quadratic wavelet estimator; (Bottom Right) Zeta3, the single-scale local cubic wavelet estimator.

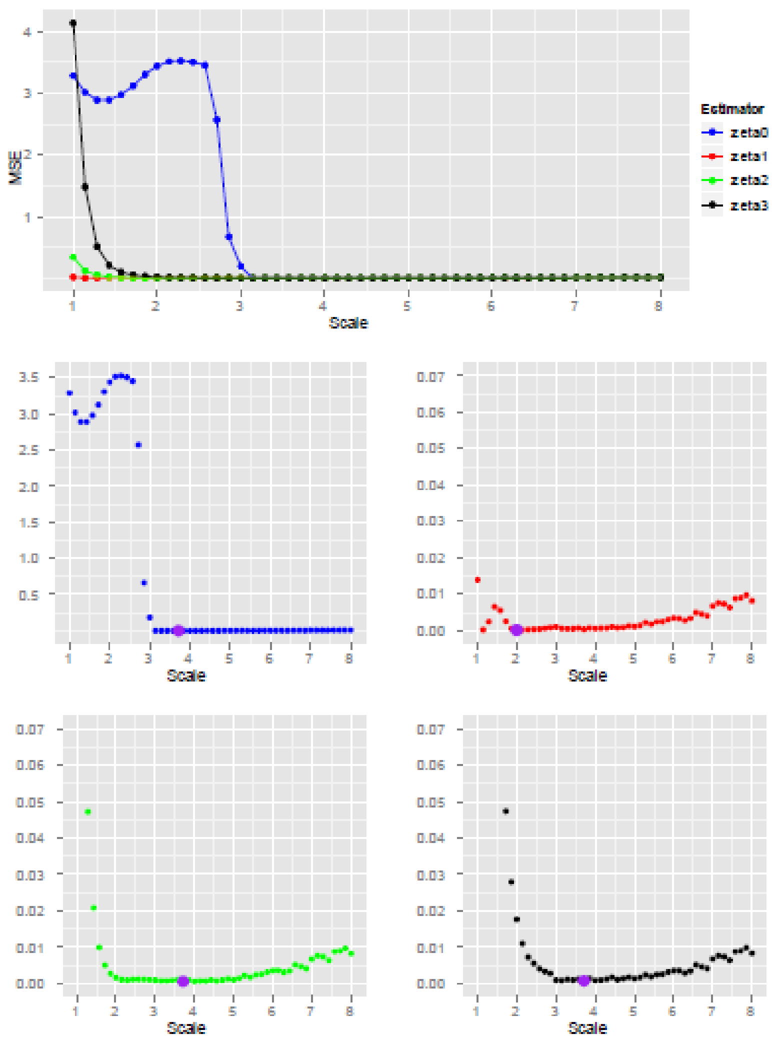

Figure 2. Local polynomial wavelet estimators under the sample size 2500 for the jump size across different scales j_0. (Top) MSE comparisons among different polynomial orders; (Middle Left) Zeta0, the single-scale local constant wavelet estimator; (Middle Right) Zeta1, the single-scale local linear wavelet estimator; (Bottom Left) Zeta2, the single-scale local quadratic wavelet estimator; (Bottom Right) Zeta3, the single-scale local cubic wavelet estimator.

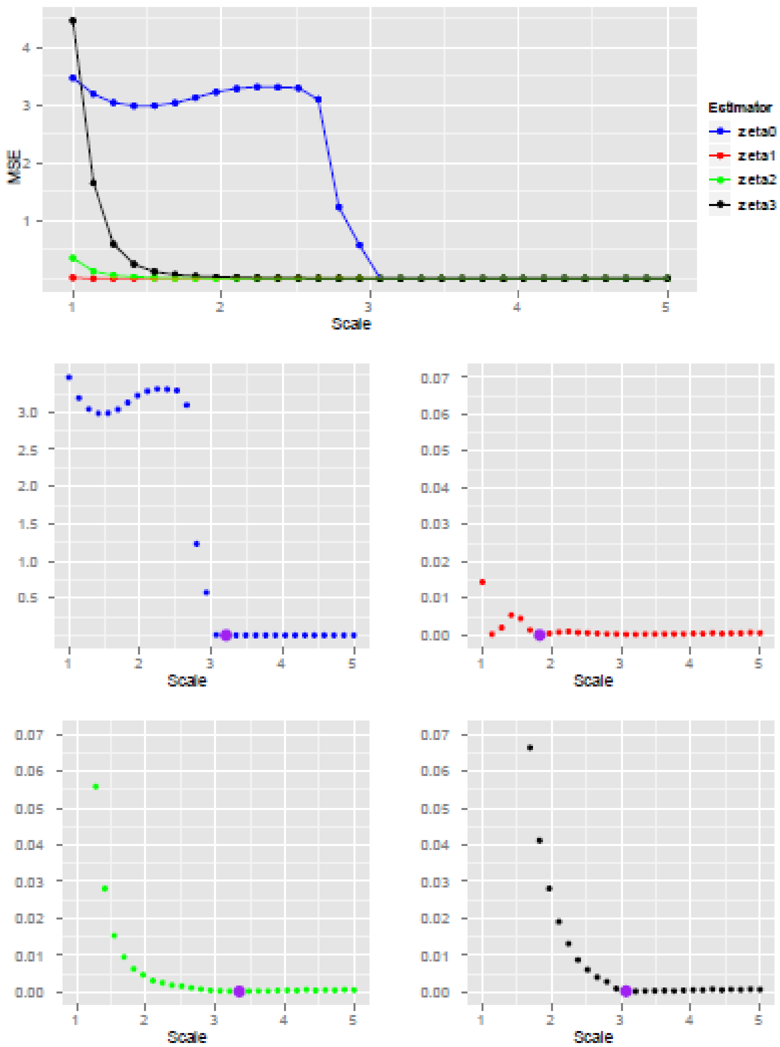

Figure 3. Local polynomial wavelet estimators under the sample size 5000 for the jump size across different scales j_0. (Top) MSE comparisons among different polynomial orders; (Middle Left) Zeta0, the single-scale local constant wavelet estimator; (Middle Right) Zeta1, the single-scale local linear wavelet estimator; (Bottom Left) Zeta2, the single-scale local quadratic wavelet estimator; (Bottom Right) Zeta3, the single-scale local cubic wavelet estimator.

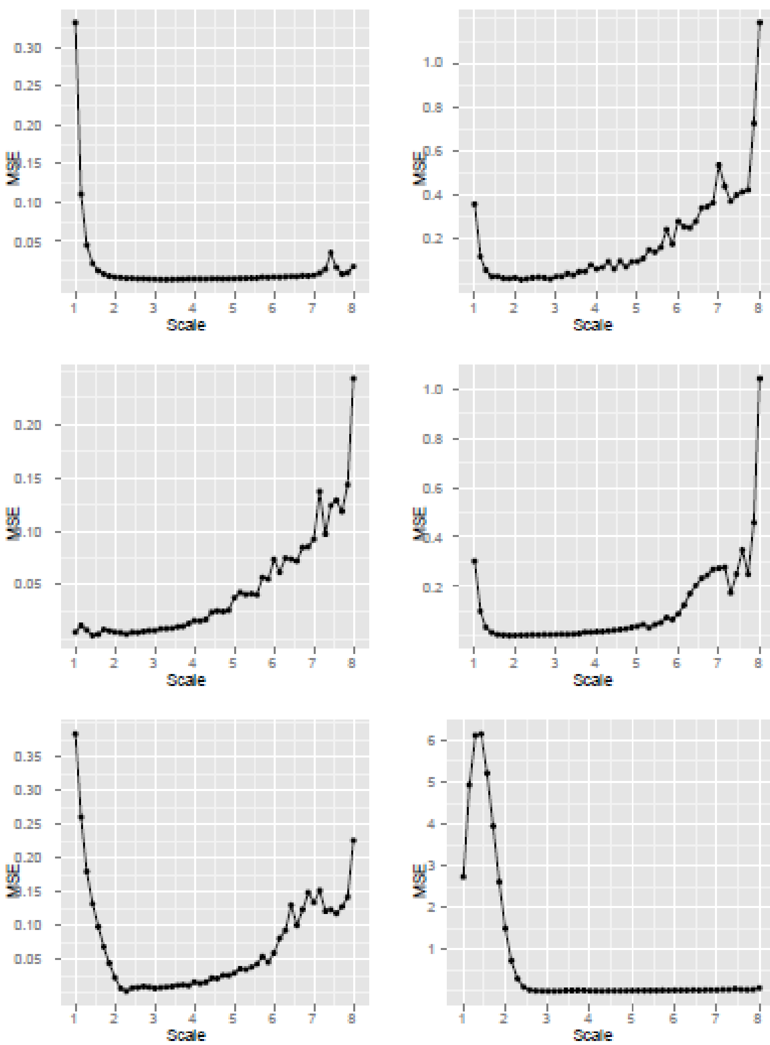

Figure 4. Robust check for the single-scale local quadratic wavelet estimator, Zeta2, when the sample size is 500: (Top Left) $W|V$ follows a multivariate studentized t distribution with parameters $(0, 0, I_{2\times2})$; (Top Right) $Var(W|V)$ is heteroskedastic with $W|V \sim N(0, 0.01\times V^2)$; (Middle Left) V follows the norm distribution $N(0.5, 0.1^2)$; (Middle Right) V follows the exponential distribution with the parameter 2; (Bottom Left) V follows the beta distribution with parameters $(1, 1, 0)$; (Bottom Right) Model-(II.16) is perturbed by adding an additive sine function $\sin[10(v - 0.5)]$.

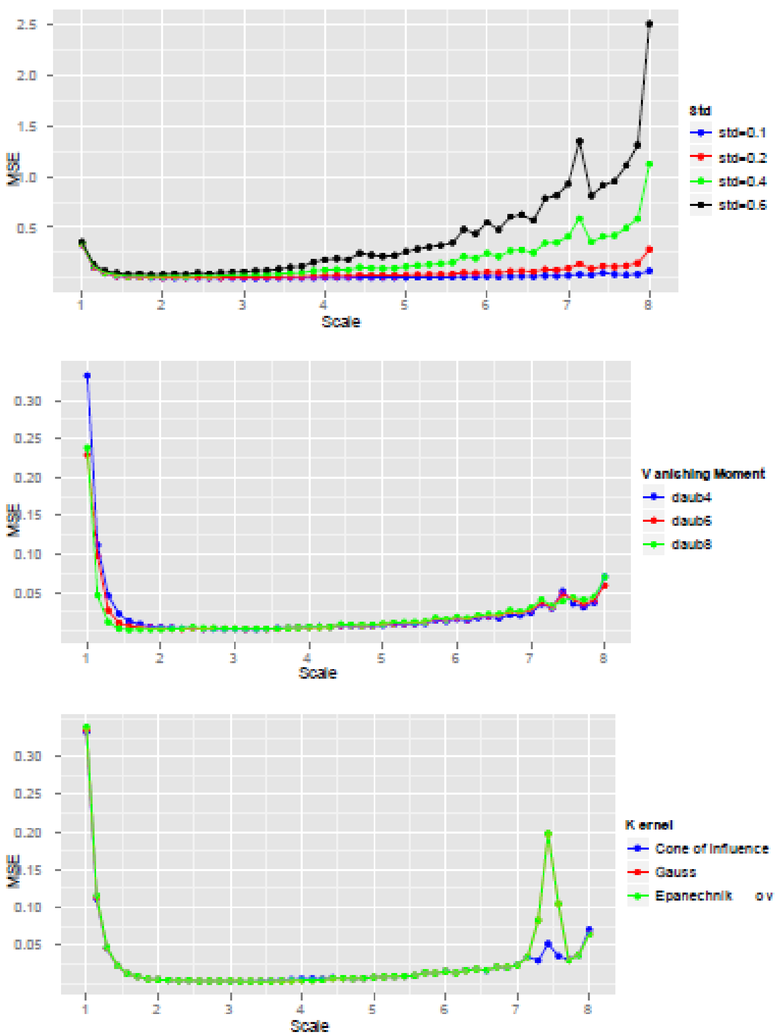

Figure 5. Robust check for the single-scale local quadratic wavelet estimator, Zeta2, when the sample size is 500: (Top) different signal to noise levels, such as, $W|V \sim N(0,0.1^2), W|V \sim N(0,0.2^2), W|V \sim N(0,0.4^2)$ and $W|V \sim N(0,0.6^2)$; (Middle) different vanishing moment wavelets ψ: Daubechies $\{4,6,8\}$ wavelet functions; (Bottom) different kernel functions to replace $\widehat{I}_{j_0}(\cdot)$ in Equation (II.10): Epanechnikov kernel with $h = 2^{-j_0}$ and Gaussian kernel $h = 2^{-j_0}$.

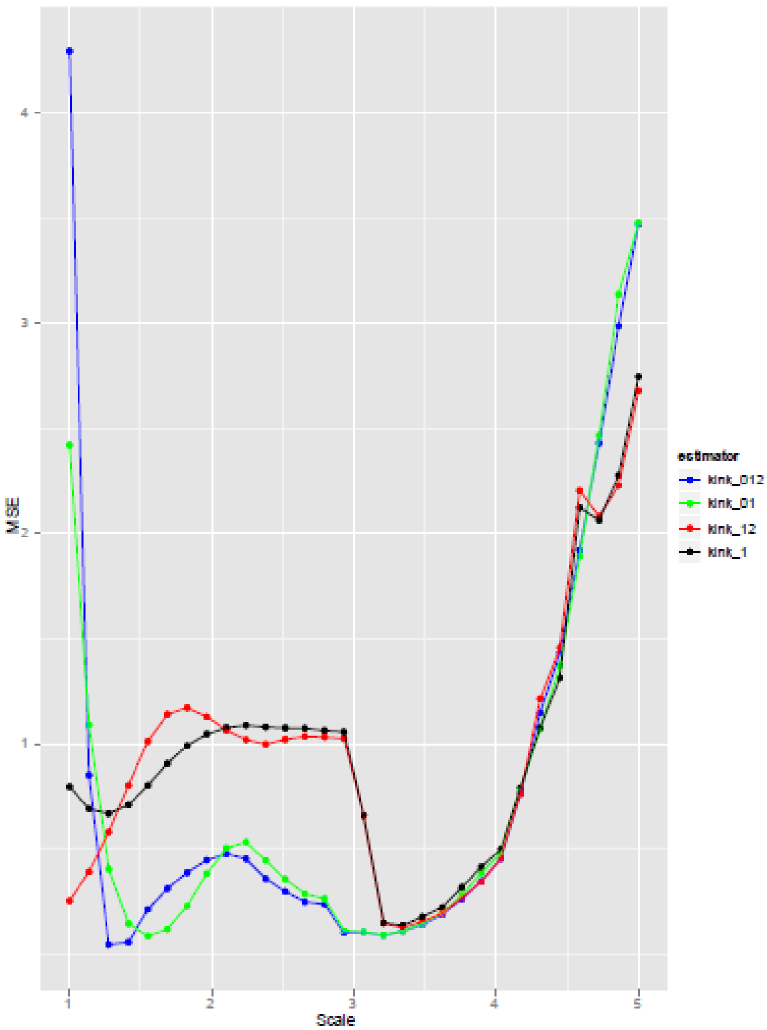

Figure 6. Local polynomial wavelet estimators under the sample size 500 for different polynomial orders. There are four different kink size wavelet estimatots: Kink_012 is the single-scale local quadratic wavelet estimator, Kink_01 is the single-scale local linear wavelet estimator, Kink_12 is the single-scale local quadratic wavelet estimator without considering the jump size, and finally Kink_1 is the single-scale local linear wavelet estimator without considering the jump size.