Supplement to "Simultaneous selection of optimal bandwidths for the sharp regression discontinuity estimator"

(*Quantitative Economics*, Vol. 9, No. 1, March 2018, 441–482)

Yoichi Arai
School of Social Sciences, Waseda University

Hidehiko Ichimura
Department of Economics, University of Tokyo

A. Introduction

We present a detailed procedure to obtain RMSE* provided in Table 1 in Section B. A detailed algorithm to implement the proposed method is described in Section C. The property of the minimizer of MMSE_n^p, $\hat{h}$, is discussed in Section D. Finally, Section E provides proofs for Lemmas 1 and 3 and Theorems 2 and 3.

B. A procedure to obtain RMSE*

We describe how RMSE* is computed for the LLR estimators based on the MMSE bandwidths, the IND bandwidths, the IK bandwidth, and the CCT bandwidth. We also show how θ_{IK} in page 12 of the main text is obtained.

Once the sample size, the form of a kernel function, the functional forms of $m_1(c)$, $m_0(c)$, $f(c)$, $\sigma_1^2(c)$, and $\sigma_0^2(c)$ are given, the AMSE can be computed using the formula of the AMSE in (2) for each of the bandwidths. We use the triangular kernel function. For other parameters, we use true values for each design.

The MMSE bandwidths can be obtained by minimizing $\text{MMSE}_n(h)$ (not $\text{MMSE}_n^p(h)$) provided on page 16 of the main text. The IND bandwidths can be obtained based on the formulae provided in the footnote of page 12.

The IK bandwidth can be obtained analogously except the regularization terms, $r_+ + r_-$. Note that

$$r_+ = \frac{2160\sigma_1^2(c)}{N_{2,+}h_{2,+}^4} \quad \text{and} \quad r_- = \frac{2160\sigma_0^2(c)}{N_{2,-}h_{2,-}^4},$$

where

$$h_{2,+} = 3.56\left(\frac{\sigma_1^2(c)}{f(c)\left[m_1^{(3)}(c)\right]^2}\right)^{1/7} N_+^{-1/7} \quad \text{and} \quad h_{2,-} = 3.56\left(\frac{\sigma_0^2(c)}{f(c)\left[m_0^{(3)}(c)\right]^2}\right)^{1/7} N_-^{-1/7}.$$

Yoichi Arai: yarai@waseda.jp
Hidehiko Ichimura: ichimura@e.u-tokyo.ac.jp

https://doi.org/10.3982/QE590

Hence the computation of the regularization term requires N_+, N_-, $N_{2,+}$, and $N_{2,-}$. Since N_+ and N_- are the number of observations to the left and right of the threshold, respectively (see page 942 of IK), their population analogues are computed by

$$N_+ = n\cdot\int_{-\infty}^{c} f(x)\,dx \quad \text{and} \quad N_- = n\cdot\int_{c}^{\infty} f(x)\,dx.$$

Similarly, since $N_{2,+}$ and $N_{2,-}$ are the numbers of observations with $c \le X_i \le c + h_{2,+}$ and $c - h_{2,-} \le X_i < c$, respectively, their population analogues are computed by

$$N_{2,+} = n\cdot\int_{c}^{c+h_{2,+}} f(x)\,dx \quad \text{and} \quad N_{2,-} = n\cdot\int_{c-h_{2,-}}^{c} f(x)\,dx.$$

The same procedure is used to obtain θ_{IK} on page 12 in the main text.

The CCT bandwidths differ from the IK bandwidth only by the regularization term given that homoskedasticity is imposed. The regularization term, r_n is given by $3V_{2,2}/nb_n^5$ where b_n is the pilot bandwidth to estimate the second-order derivatives and $V_{2,2}$ given in page 38 of Calonico, Cattaneo, and Titiunik (2014b) is a function of the kernel function, $f(c)$, $\sigma_1^2(c)$, and $\sigma_0^2(c)$. The pilot bandwidths b_n can be obtained in two steps. First, the pilot bandwidth to estimate the third derivative, c_n, is given by

$$c_n = C_{3,3}\left(\frac{\sigma_1^2(c)+\sigma_0^2(c)}{f(c)\big[m_1^{(4)}(c)-m_0^{(4)}(c)\big]^2}\right)^{1/9} n^{-1/9},$$

where $C_{3,3}$ is given in Section 3.2.3 of Fan and Gijbels (1996). Then the pilot bandwidth b_n is given by

$$b_n = C_{2,2}\left(\frac{\sigma_1^2(c)+\sigma_0^2(c)}{f(c)\big\{\big[m_1^{(3)}(c)-m_0^{(3)}(c)\big]^2 + 3\cdot(3!)^2 V_{3,3}/nc_n^7\big\}}\right)^{1/9} n^{-1/9}.$$

C. Implementation

To obtain the proposed bandwidths in the case of the sharp RD design, we need pilot estimates of the density, its first derivative, the second, and third derivatives of the conditional expectation functions, and the conditional variances at the cut-off point. For the sharp RK design, we need pilot estimates of the third and fourth derivatives in stead of the second and third derivatives. We obtain these pilot estimates in a number of steps.

C.1 *Sharp RD design*

Step 1: Obtain pilot estimates for the density $f(c)$ and its first derivative $f^{(1)}(c)$. We calculate the density of the assignment variable at the cut-off point, $f(c)$, which is estimated using the kernel density estimator with an Epanechnikov kernel.[1] A pilot bandwidth for

[1] IK estimated the density in a simpler manner (see Section 4.2 of IK). We used the kernel density estimator to be consistent with the estimation method used for the first derivative. Our unreported simulation experiments produced similar results for both methods.

kernel density estimation is chosen using the normal scale rule with Epanechnikov kernel, given by $2.34\hat{\sigma}n^{-1/5}$, where $\hat{\sigma}$ is the square root of the sample variance of X_i (see Silverman (1986) and Wand and Jones (1994) for the normal scale rules). The first derivative of the density is estimated using the method proposed by Jones (1994). The kernel first derivative density estimator is given by $\sum_{i=1}^{n} L((c - X_i)/h)/(nh^2)$, where L is the kernel function proposed by Jones (1994), $L(u) = -15u(1-u^2)1_{\{|u|<1\}}/4$. Again, a pilot bandwidth is obtained using the normal scale rule, given by $\hat{\sigma} \cdot (112\sqrt{\pi}/n)^{1/7}$.

Step 2: Obtain pilot bandwidths for estimating the second and third derivatives $m_j^{(2)}(c)$ and $m_j^{(3)}(c)$ for $j = 0, 1$. We next estimate the second and third derivatives of the conditional mean functions using the third-order LPR.

We obtain pilot bandwidths for the LPR based on the estimated fourth derivatives of $m_1^{(4)}(c) = \lim_{x \to c+} m_1^{(4)}(x)$ and $m_0^{(4)}(c) = \lim_{x \to c-} m_0^{(4)}(x)$. Following Fan and Gijbels (1996), Imbens and Kalyanaraman (2012), and Calonico, Cattaneo, and Titiunik (2014a), we use estimates of $m_1^{(4)}(c)$ that are not necessarily consistent by fitting global polynomial regressions. First, using observations for which $X_i \geq c$, we regress Y_i on 1, $(X_i - c)$, $(X_i - c)^2$, $(X_i - c)^3$, and $(X_i - c)^4$ to obtain the OLS coefficients $\hat{\gamma}_1$ and the variance estimate $\hat{s}_1^2$. Using the data with $X_i < c$, we repeat the same procedure to obtain $\hat{\gamma}_0$ and $\hat{s}_0^2$. The pilot estimates for fourth derivatives are $\hat{m}_1^{(4)}(c) = 24 \cdot \hat{\gamma}_1(5)$ and $\hat{m}_0^{(4)}(c) = 24 \cdot \hat{\gamma}_0(5)$, where $\hat{\gamma}_1(5)$ and $\hat{\gamma}_0(5)$ are the fifth elements of $\hat{\gamma}_1$ and $\hat{\gamma}_0$, respectively. The plug-in bandwidths for the third-order LPR used to estimate the second and third derivatives are calculated by

$$h_{\nu,j} = C_{\nu,3}(K)\left(\frac{\hat{s}_j^2}{\hat{f}(c) \cdot \hat{m}_j^{(4)}(c)^2 \cdot n_j}\right)^{1/9}$$

for $j = 0, 1$ where n_0 and n_1 are the numbers of observations on the left and right of the cut-off point, respectively (see Fan and Gijbels (1996, Section 3.2.3) for information on plug-in bandwidths and the definition of $C_{\nu,3}$).[2] We use $\nu = 2$ and $\nu = 3$ for estimating the second and third derivatives, respectively.

Step 3: Estimation of the second and third derivatives $m_j^{(2)}(c)$ and $m_j^{(3)}(c)$ as well as the conditional variances $\hat{\sigma}_j^2(c)$ for $j = 0, 1$. We estimate the second and third derivatives at the cut-off point using the third-order LPR with the pilot bandwidths obtained in Step 2. Following CCT, we use the triangular kernel, which yields $C_{2,3} = 5.7851$ and $C_{3,3} = 5.2774$. To estimate $\hat{m}_1^{(2)}(c)$, we construct a vector $Y_a = (Y_1, \ldots, Y_{n_a})'$ and an $n_a \times 4$ matrix, X_a, whose ith row is given by $(1, (X_i - c), (X_i - c)^2, (X_i - c)^3)$ for observations with $c \leq X_i \leq c + h_{2,1}$, where n_a is the number of observations with $c \leq X_i \leq c + h_{2,1}$. We also construct the weighting matrix W_a where $W_a = \text{diag}\{K((X_i - c)/h_{2,1})\}$ and K is the triangular kernel. The estimated second derivative is given by $\hat{m}_1^{(2)}(c) = 2 \cdot \hat{\beta}_{2,1}(3)$, where $\hat{\beta}_{2,1}(3)$ is the third element of $\hat{\beta}_{2,1}$ and $\hat{\beta}_{2,1} = (X_a{}' W_a X_a)^{-1} X_a W_a Y_a$. We estimate $\hat{m}_0^{(2)}(c)$ in the same manner. Replacing $h_{2,1}$ with $h_{3,1}$ leads to an estimated third derivative of $\hat{m}_1^{(3)}(c) = 6 \cdot \hat{\beta}_{3,1}(4)$, where $\hat{\beta}_{3,1}(4)$ is the fourth element of

[2] The bandwidth we use for estimating the third derivatives are not rate optimal when the underlying function has higher-order derivative. However, we use this bandwidth to avoid estimating higher-order derivatives.

$\hat{\beta}_{3,1}$, $\hat{\beta}_{3,1} = (X_b{}' W_b X_b)^{-1} X_b W_b Y_b$, $Y_b = (Y_1, \ldots, Y_{n_b})'$, X_b is an $n_b \times 4$ matrix whose ith row is given by $(1, (X_i - c), (X_i - c)^2, (X_i - c)^3)$ for observations with $c \leq X_i \leq c + h_{3,1}$, $W_b = \text{diag}\{K((X_i - c)/h_{3,1})\}$ and n_b is the number of observations with $c \leq X_i \leq c + h_{3,1}$. The conditional variance at the cut-off point $\sigma_1^2(c)$ is calculated as $\hat{\sigma}_1^2(c) = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 K((X_i - c)/h_{2,1}) / \text{trace}(W_a - W_a X_a (X_a' W_a X_a)^{-1} X_a' W_a)$, where $\hat{Y}_i$ denotes the fitted values from the regression used to estimate the second derivative. $\hat{\beta}_{2,0}$, $\hat{\beta}_{3,0}$, and $\hat{\sigma}_0^2(c)$ can be obtained analogously.

Step 4: Numerical optimization. The final step is to plug the pilot estimates into the MMSE^p given by equation (6) in the main text and to use numerical minimization over the compact region to obtain $\hat{h}_1$ and $\hat{h}_0$. Unlike $\text{AMSE}_{1n}(h)$ and $\text{AMSE}_{2n}(h)$ subject to the restriction given in Definition 1, the MMSE is not necessarily strictly convex, particularly when the sign of the product is positive. In minimizing the objective function, it is important to try optimization with several initial values, in order to avoid finding only a local minimum.

C.2 *Sharp RK design*

Step 1 is the same as the one for the sharp RD design. We modify Steps 2, 3, and 4 but these steps are analogous to those for the sharp RD design.

Step 2: Obtain pilot bandwidths for estimating the third and fourth derivatives $m_j^{(3)}(c)$ and $m_j^{(4)}(c)$ for $j = 0, 1$. We obtain pilot bandwidths for the LPR based on the estimated fifth derivatives. First, using observations for which $X_i \geq c$, we regress Y_i on 1, $(X_i - c)$, $(X_i - c)^2$, $(X_i - c)^3$, $(X_i - c)^4$, and $(X_i - c)^5$ to obtain the OLS coefficients $\hat{\gamma}_1$ and the variance estimate $\hat{s}_1^2$. Using the data with $X_i < c$, we repeat the same procedure to obtain $\hat{\gamma}_0$ and $\hat{s}_0^2$. The pilot estimates for fifth derivatives are $\hat{m}_1^{(5)}(c) = 120 \cdot \hat{\gamma}_1(6)$ and $\hat{m}_0^{(5)}(c) = 120 \cdot \hat{\gamma}_0(6)$, where $\hat{\gamma}_1(6)$ and $\hat{\gamma}_0(6)$ are the fifth elements of $\hat{\gamma}_1$ and $\hat{\gamma}_0$, respectively. The plug-in bandwidths for the third-order LPR used to estimate the third and fourth derivatives are calculated by

$$h_{\nu,j} = C_{\nu,4}(K) \left(\frac{\hat{s}_j^2}{\hat{f}(c) \cdot \hat{m}_j^{(5)}(c)^2 \cdot n_j} \right)^{1/11}$$

for $j = 0, 1$ where n_0 and n_1 are the numbers of observations on the left and right of the cut-off point, respectively (see Fan and Gijbels (1996, Section 3.2.3) for information on plug-in bandwidths and the definition of $C_{\nu,3}$). We use $\nu = 3$ and $\nu = 4$ for estimating the second and third derivatives, respectively.

Step 3: Estimation of the third and fourth derivatives $m_j^{(3)}(c)$ and $m_j^{(4)}(c)$ as well as the conditional variances $\hat{\sigma}_j^2(c)$ for $j = 0, 1$. We estimate the third and fourth derivatives at the cut-off point using the fourth-order LPR with the pilot bandwidths obtained in Step 2. Following CCT, we use the triangular kernel, which yields $C_{3,4} = 6.5261$ and $C_{4,4} = 6.1275$. To estimate $\hat{m}_1^{(3)}(c)$, we construct a vector $Y_a = (Y_1, \ldots, Y_{n_a})'$ and an $n_a \times 5$ matrix, X_a, whose ith row is given by $(1, (X_i - c), (X_i - c)^2, (X_i - c)^3, (X_i - c)^4)$ for observations with $c \leq X_i \leq c + h_{3,1}$, and $W_a = \text{diag}\{K((X_i - c)/h_{3,1})\}$ where n_a is the number of observations with $c \le X_i \le c + h_{3,1}$ and K is the kernel function. The estimated third derivative is given by $\hat{m}_1^{(3)}(c) = 6 \cdot \hat{\beta}_{3,1}(4)$, where $\hat{\beta}_{3,1}(4)$ is the fourth element of $\hat{\beta}_{3,1}$ and $\hat{\beta}_{3,1} = (X_a{}'W_a X_a)^{-1} X_a W_a Y_a$. We estimate $\hat{m}_0^{(3)}(c)$ in the same manner. Replacing $h_{3,1}$ with $h_{4,1}$ leads to an estimated fourth derivative of $\hat{m}_1^{(4)}(c) = 12 \cdot \hat{\beta}_{4,1}(5)$, where $\hat{\beta}_{4,1}(5)$ is the fifth element of $\hat{\beta}_{4,1}$, $\hat{\beta}_{4,1} = (X_b{}'W_b X_b)^{-1} X_b Y_b$, $Y_b = (Y_1, \ldots, Y_{n_b})'$, X_b is an $n_b \times 5$ matrix whose ith row is given by $(1, (X_i - c), (X_i - c)^2, (X_i - c)^3, (X_i - c)^4)$ for observations with $c \le X_i \le c + h_{4,1}$, $W_b = \text{diag}\{K((X_i - c)/h_{4,1})\}$ and n_b is the number of observations with $c \le X_i \le c + h_{4,1}$. The conditional variance at the cut-off point $\sigma_1^2(c)$ is calculated as $\hat{\sigma}_1^2(c) = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 K((X_i - c)/h_{3,1}) / \text{trace}(W_a - W_a X_a (X_a' W_a X_a)^{-1} X_a' W_a)$, where $\hat{Y}_i$ denotes the fitted values from the regression used to estimate the second derivative. $\hat{\beta}_{3,0}$, $\hat{\beta}_{4,0}$, and $\hat{\sigma}_0^2(c)$ can be obtained analogously.

Step 4: Numerical optimization. The final step is to plug the pilot estimates into the MMSE^p given by equation (7) in the main text and to use numerical minimization over the compact region to obtain $\hat{h}_1$ and $\hat{h}_0$.

C.3 *Confidence interval*

The expression of the objects necessary to construct the confidence interval provided in Appendix B of the main text simply depend on the data except the bandwidth, kernel function, and the conditional variance estimate. For the bandwidths, our recommendation is to use $(\hat{h}, \hat{h}, \hat{h})$ following the recommendation of CCT. For the kernel function and the conditional variance estimate, we follow the procedures described in Sections C.1 and C.2.

D. The property of $\hat{h}$

In this section, we describe the property of the minimizer of MMSE_n^p, $\hat{h}$. First, we show the uniqueness of $\hat{h}$ around $m_1^{(2)}(c) m_0^{(2)}(c) = 0$. Without loss of generality, we focus on the case where $m_1^{(2)}(c)$ is close to zero.

Remember that the objective function is written, with simplified notation, as

$$L = \left(\alpha_1 h_1^2 - \alpha_0 h_0^2\right)^2 + \left(\beta_1 h_1^3 - \beta_0 h_0^3\right)^2 + \frac{\gamma_1}{n h_1} + \frac{\gamma_0}{n h_0}. \tag{D.1}$$

Let $\lambda = h_1/h_0$, $a = \alpha_0/\alpha_1$, $b = \beta_0/\beta_1$ and $r = \gamma_1/\gamma_0$. Note that $\lambda > 0$ and $r > 0$. With this, the objective function is reparametrized as

$$L = \alpha_1^2 h_0^4 \left(\lambda^2 - a\right)^2 + \beta_1^2 h_0^6 \left(\lambda^3 - b\right)^2 + \frac{\gamma_0}{n h_0}\left(\frac{r}{\lambda} + 1\right).$$

Then the first-order conditions are given by

$$\frac{\partial L}{\partial h_0} = 4\alpha_1^2 \left(\lambda^2 - a\right)^2 h_0^3 + 6\beta_1^2 \left(\lambda^3 - b\right)^2 h_0^5 - \frac{\gamma_0}{n h_0^2}\left(\frac{r}{\lambda} + 1\right) = 0$$

and

$$\frac{\partial L}{\partial \lambda} = 4\alpha_1^2 \lambda \left(\lambda^2 - a\right) h_0^4 + 6\beta_1^2 \lambda^2 \left(\lambda^3 - b\right) h_0^6 - \frac{\gamma_0 r}{n \lambda^2 h_0} = 0. \tag{D.2}$$

Solving the first-order conditions with respect to h_0 produces

$$h_0 = \left(-\frac{2\alpha_1^2(\lambda^2 - a)(\lambda^3 + ar)}{3\beta_1^2(\lambda^3 - b)(\lambda^4 + br)} \right)^{1/2}. \tag{D.3}$$

We consider if these equations have a unique positive solution in h_0 and λ.

We need to consider four cases (i) $a \geq 0$ and $b \geq 0$, (ii) $a \geq 0$ and $b \leq 0$, (iii) $a \leq 0$ and $b \geq 0$, and (iv) $a \leq 0$ and $b \leq 0$. In the following, we exclude the case where $a = b = 0$.

D.1 *(i) $a \geq 0$ and $b \geq 0$*

We consider two cases of (i-1) $a^{1/2} < b^{1/3}$ and (i-2) $b^{1/3} < a^{1/2}$ separately.

D.1.1 *(i-1) $a \geq 0$, $b \geq 0$ and $a^{1/2} < b^{1/3}$* Since h_0 is positive, we can restrict our attention to the values of λ which satisfy $a^{1/2} < \lambda < b^{1/3}$ by equation (D.3). The first-order condition (D.2) can be expressed as

$$4\alpha_1^2\lambda(\lambda^2 - a)h_0^5 + 6\beta_1^2\lambda^2(\lambda^3 - b)h_0^7 - \frac{\gamma_0 r}{n\lambda^2} = 0.$$

Substituting the expression of h_0 given by equation (D.3) into this yields

$$\begin{aligned} &4\alpha_1^2\lambda^3(\lambda^2 - a)\left(-\frac{2\alpha_1^2(\lambda^2 - a)(\lambda^3 + ar)}{3\beta_1^2(\lambda^3 - b)(\lambda^4 + br)} \right)^{5/2} \\ &\quad + 6\beta_1^2\lambda^4(\lambda^3 - b)\left(-\frac{2\alpha_1^2(\lambda^2 - a)(\lambda^3 + ar)}{3\beta_1^2(\lambda^3 - b)(\lambda^4 + br)} \right)^{7/2} \\ &= \frac{\gamma_0 r}{n} \end{aligned} \tag{D.4}$$

or

$$4 \cdot \left(\frac{2}{3}\right)^{5/2} \cdot \frac{|\alpha_1|^7}{|\beta_1|^5} \cdot \frac{\lambda^3(\lambda^2 - a)^{7/2}(\lambda^3 + ar)^{5/2}(b - a\lambda)}{(b - \lambda^3)^{5/2}(\lambda^4 + br)^{7/2}} = \frac{\gamma_0}{n}.$$

Denote

$$f_{11}(\lambda) \equiv \frac{\lambda^3(\lambda^2 - a)^{7/2}(\lambda^3 + ar)^{5/2}(b - a\lambda)}{(b - \lambda^3)^{5/2}(\lambda^4 + br)^{7/2}}.$$

Since $\lim_{\lambda \downarrow a^{1/2}} f_{11}(\lambda) = 0$ and $\lim_{\lambda \uparrow b^{1/3}} f_{11}(\lambda) = \infty$, the uniqueness of λ that satisfy the first-order condition follows if $f_{11}(\lambda)$ is monotonically increasing. The derivative of $f_{11}(\lambda)$ is given by

$$\begin{aligned} \frac{df_{11}(\lambda)}{d\lambda} = &\frac{\lambda^2(\lambda^2 - a)^{5/2}(\lambda^3 + ar)^{3/2}}{(b - \lambda^3)^{7/2}(\lambda^4 + br)^{9/2}} \\ &\times \Big\{ 10(\lambda^2 - a)(\lambda^3 + ar)(b - a\lambda)(b - \lambda^3)(\lambda^4 + br) \end{aligned}$$

$$
\begin{aligned}
&+ 7a\big(\lambda^3 + ar\big)(b - a\lambda)\big(b - \lambda^3\big)\big(\lambda^4 + br\big) \\
&+ \frac{15}{2}\lambda^3\big(\lambda^2 - a\big)(b - a\lambda)\big(b - \lambda^3\big)\big(\lambda^4 + br\big) \\
&- a\lambda\big(\lambda^2 - a\big)\big(\lambda^3 + ar\big)\big(b - \lambda^3\big)\big(\lambda^4 + br\big) \\
&+ \frac{15}{2}\lambda^3\big(\lambda^2 - a\big)\big(\lambda^3 + ar\big)(b - a\lambda)\big(\lambda^4 + br\big) \\
&- 14\lambda^4\big(\lambda^2 - a\big)\big(\lambda^3 + ar\big)(b - a\lambda)\big(b - \lambda^3\big)\Big\}.
\end{aligned}
$$

Observe that, for λ which satisfies $a^{1/2} < \lambda < b^{1/3}$, we have $(\lambda^2 - a) > 0$, $(\lambda^3 + ar) > 0$, $(b - a\lambda) > 0$, $(b - \lambda^3) > 0$, and $(\lambda^4 + br) > 0$. Since the fraction outside of the curly bracket is positive, it is enough to show that the sum in the curly bracket is positive. It can be shown that the sum is equal to

$$
\begin{aligned}
&10br\big(\lambda^2 - a\big)\big(\lambda^3 + ar\big)(b - a\lambda)\big(b - \lambda^3\big) + 6a\big(\lambda^3 + ar\big)(b - a\lambda)\big(b - \lambda^3\big)\big(\lambda^4 + br\big) \\
&\quad + \frac{7}{2}\lambda^3\big(\lambda^2 - a\big)(b - a\lambda)\big(b - \lambda^3\big)\big(\lambda^4 + br\big) + a\big(\lambda^3 + ar\big)\big(b - \lambda^3\big)^2\big(\lambda^4 + br\big) \\
&\quad + \frac{15}{2}\lambda^3\big(\lambda^2 - a\big)\big(\lambda^3 + ar\big)(b - a\lambda)\big(\lambda^4 + br\big) + 4r\lambda^3\big(\lambda^2 - a\big)(b - a\lambda)^2\big(b - \lambda^3\big) \\
&\quad > 0,
\end{aligned}
$$

proving the required result.

D.1.2 *(i-2) $a \geq 0$, $b \geq 0$, and $b^{1/3} < a^{1/2}$* For this case, we can restrict our attention to the values of λ which satisfy $b^{1/3} < \lambda < a^{1/a}$ as in the previous case. The first-order condition (D.4) can be expressed as

$$
4 \cdot \left(\frac{2}{3}\right)^{5/2} \cdot \frac{|\alpha_1|^7}{|\beta_1|^5} \cdot \frac{\lambda^3\big(a - \lambda^2\big)^{7/2}\big(\lambda^3 + ar\big)^{5/2}(a\lambda - b)}{\big(\lambda^3 - b\big)^{5/2}\big(\lambda^4 + br\big)^{7/2}} = \frac{\gamma_0}{n}.
$$

Denote

$$
f_{12}(\lambda) \equiv \frac{\lambda^3\big(a - \lambda^2\big)^{7/2}\big(\lambda^3 + ar\big)^{5/2}(a\lambda - b)}{\big(\lambda^3 - b\big)^{5/2}\big(\lambda^4 + br\big)^{7/2}}.
$$

Since $\lim_{\lambda \downarrow b^{1/3}} f_{12}(\lambda) = \infty$ and $\lim_{\lambda \uparrow a^{1/2}} f_{12}(\lambda) = 0$, the uniqueness of λ that satisfy the first-order condition follows if $f_{12}(\lambda)$ is monotonically decreasing. The derivative of $f_{12}(\lambda)$ is given by

$$
\begin{aligned}
\frac{df_{12}(\lambda)}{d\lambda} = &\frac{\lambda^2\big(a - \lambda^2\big)^{5/2}\big(\lambda^3 + ar\big)^{3/2}}{\big(\lambda^3 - b\big)^{7/2}\big(\lambda^4 + br\big)^{9/2}} \\
&\times \Big\{10\big(a - \lambda^2\big)\big(\lambda^3 + ar\big)(a\lambda - b)\big(\lambda^3 - b\big)\big(\lambda^4 + br\big) \\
&- 7a\big(\lambda^3 + ar\big)(a\lambda - b)\big(\lambda^3 - b\big)\big(\lambda^4 + br\big)
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{15}{2}\lambda^3\big(a-\lambda^2\big)(a\lambda-b)\big(\lambda^3-b\big)\big(\lambda^4+br\big)\\
&+a\lambda\big(a-\lambda^2\big)\big(\lambda^3+ar\big)\big(\lambda^3-b\big)\big(\lambda^4+br\big)\\
&-\frac{15}{2}\lambda^3\big(a-\lambda^2\big)\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^4+br\big)\\
&-14\lambda^4\big(a-\lambda^2\big)\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^3-b\big)\Big\}.
\end{aligned}
$$

Observe that, for λ which satisfies $b^{1/3} < \lambda < a^{1/2}$, we have $(a-\lambda^2) > 0$, $(\lambda^3+ar) > 0$, $(a\lambda - b) > 0$, $(\lambda^3 - b) > 0$, and $(\lambda^4 + br) > 0$. Since the fraction outside of the curly bracket is positive, it is enough to show that the sum in the curly bracket is negative. It can be shown that the sum is equal to

$$
\begin{aligned}
&-\frac{5}{2}\lambda^2\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^3-b\big)\big(\lambda^4+br\big)-\frac{7}{2}a\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^3-b\big)\big(\lambda^4+br\big)\\
&\quad-\frac{15}{2}r\lambda^3\big(a-\lambda^2\big)(a\lambda-b)^2\big(\lambda^3-b\big)-a\big(\lambda^3+ar\big)\big(\lambda^3-b\big)^2\big(\lambda^4+br\big)\\
&\quad-\frac{13}{2}\lambda^4\big(a-\lambda^2\big)\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^3-b\big)-\frac{15}{2}b\big(a-\lambda^2\big)\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^4+br\big)\\
&\quad<0.
\end{aligned}
$$

This leads to the required result.

D.2 *(ii) $a \geq 0$ and $b \leq 0$*

We consider two cases of (i-a) $a^{1/2} < (-br)^{1/4}$ and $(-br)^{1/4} < a^{1/2}$ separately.

D.2.1 *(ii-1) $a \geq 0$, $b \leq 0$ and $a^{1/2} < (-br)^{1/4}$* Since h_0 is positive, we can restrict our attention to the values of λ which satisfy $a^{1/2} < \lambda < (-br)^{1/4}$ by equation (D.3). The first-order condition (D.4) can be expressed as

$$
4\cdot\left(\frac{2}{3}\right)^{5/2}\cdot\frac{|\alpha_1|^7}{|\beta_1|^5}\cdot\frac{\lambda^3\big(\lambda^2-a\big)^{7/2}\big(\lambda^3+ar\big)^{5/2}(a\lambda-b)}{\big(\lambda^3-b\big)^{5/2}\big(-\lambda^4-br\big)^{7/2}}=\frac{\gamma_0}{n}.
$$

Denote

$$
f_{21}(\lambda)\equiv\frac{\lambda^3\big(\lambda^2-a\big)^{7/2}\big(\lambda^3+ar\big)^{5/2}(a\lambda-b)}{\big(\lambda^3-b\big)^{5/2}\big(-\lambda^4-br\big)^{7/2}}.
$$

Since $\lim_{\lambda\downarrow a^{1/2}} f_{21}(\lambda) = 0$ and $\lim_{\lambda\uparrow(-br)^{1/4}} f_{21}(\lambda) = \infty$, the uniqueness of λ that satisfy the first-order condition follows if $f_{21}(\lambda)$ is monotonically increasing. The derivative of $f_{21}(\lambda)$ is given by

$$
\begin{aligned}
\frac{df_{12}(\lambda)}{d\lambda}=&\frac{\lambda^2\big(\lambda^2-a\big)^{5/2}\big(\lambda^3+ar\big)^{3/2}}{\big(\lambda^3-b\big)^{7/2}\big(-\lambda^4-br\big)^{9/2}}\\
&\times\Big\{10\big(\lambda^2-a\big)\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^3-b\big)\big(-\lambda^4-br\big)
\end{aligned}
$$

$$
\begin{aligned}
&+7a\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^3-b\big)\big(-\lambda^4-br\big)\\
&+\frac{15}{2}\lambda^3\big(\lambda^2-a\big)(a\lambda-b)\big(\lambda^3-b\big)\big(-\lambda^4-br\big)\\
&+a\lambda\big(\lambda^2-a\big)\big(\lambda^3+ar\big)\big(\lambda^3-b\big)\big(-\lambda^4-br\big)\\
&-\frac{15}{2}\lambda^3\big(\lambda^2-a\big)\big(\lambda^3+ar\big)(a\lambda-b)\big(-\lambda^4-br\big)\\
&+14\lambda^4\big(\lambda^2-a\big)\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^3-b\big)\bigg\}.
\end{aligned}
$$

Observe that, for λ which satisfies $a^{1/2} < \lambda < (-br)^{1/4}$, we have $(\lambda^2-a) > 0$, $(\lambda^3+ar) > 0$, $(a\lambda-b) > 0$, $(\lambda^3-b) > 0$, and $(-\lambda^4-br) > 0$. Since the fraction outside of the curly bracket is positive, it is enough to show that the sum in the curly bracket is positive. It can be shown that the sum is equal to

$$
\begin{aligned}
&\frac{5}{2}\big(\lambda^2-a\big)\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^3-b\big)\big(-\lambda^4-br\big)+7a\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^3-b\big)\big(-\lambda^4-br\big)\\
&\quad+\frac{15}{2}\lambda^3\big(\lambda^2-a\big)(a\lambda-b)\big(\lambda^3-b\big)\big(-\lambda^4-br\big)+a\lambda\big(\lambda^2-a\big)\big(\lambda^3+ar\big)\big(\lambda^3-b\big)\big(-\lambda^4-br\big)\\
&\quad-\frac{15}{2}b\big(\lambda^2-a\big)\big(\lambda^3+ar\big)(a\lambda-b)\big(-\lambda^4-br\big)+14\lambda^4\big(\lambda^2-a\big)\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^3-b\big)\\
&\quad>0.
\end{aligned}
$$

This leads to the required result.

D.2.2 *(ii-2)* $a \geq 0$, $b \leq 0$, *and* $(-br)^{1/4} < a^{1/2}$ Since h_0 is positive, we can restrict our attention to the values of λ which satisfy $(-br)^{1/4} < \lambda < a^{1/2}$ by equation (D.3). The first-order condition (D.4) can be expressed as

$$
4\cdot\left(\frac{2}{3}\right)^{5/2}\cdot\frac{|\alpha_1|^7}{|\beta_1|^5}\cdot\frac{\lambda^3\big(a-\lambda^2\big)^{7/2}\big(\lambda^3+ar\big)^{5/2}(a\lambda-b)}{\big(\lambda^3-b\big)^{5/2}\big(\lambda^4+br\big)^{7/2}}=\frac{\gamma_0}{n}.
$$

Denote

$$
f_{22}(\lambda)\equiv\frac{\lambda^3\big(a-\lambda^2\big)^{7/2}\big(\lambda^3+ar\big)^{5/2}(a\lambda-b)}{\big(\lambda^3-b\big)^{5/2}\big(\lambda^4+br\big)^{7/2}}.
$$

Since $\lim_{\lambda\downarrow(-br)^{1/4}} f_{22}(\lambda)=\infty$ and $\lim_{\lambda\uparrow a^{1/2}} f_{22}(\lambda)=0$, the uniqueness of λ that satisfy the first-order condition follows if $f_{22}(\lambda)$ is monotonically decreasing. The derivative of $f_{22}(\lambda)$ is given, as in (i-2), by

$$
\begin{aligned}
\frac{df_{22}(\lambda)}{d\lambda}=&\frac{\lambda^2\big(a-\lambda^2\big)^{5/2}\big(\lambda^3+ar\big)^{3/2}}{\big(\lambda^3-b\big)^{7/2}\big(\lambda^4+br\big)^{9/2}}\\
&\times\bigg\{10\big(a-\lambda^2\big)\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^3-b\big)\big(\lambda^4+br\big)\\
&-7a\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^3-b\big)\big(\lambda^4+br\big)
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{15}{2}\lambda^3\big(a-\lambda^2\big)(a\lambda-b)\big(\lambda^3-b\big)\big(\lambda^4+br\big)\\
&+a\lambda\big(a-\lambda^2\big)\big(\lambda^3+ar\big)\big(\lambda^3-b\big)\big(\lambda^4+br\big)\\
&-\frac{15}{2}\lambda^3\big(a-\lambda^2\big)\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^4+br\big)\\
&-14\lambda^4\big(a-\lambda^2\big)\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^3-b\big)\Big\}.
\end{aligned}
$$

Observe that, for λ which satisfies $(-br)^{1/4} < \lambda < a^{1/2}$, we have $(a-\lambda^2) > 0$, $(\lambda^3+ar) > 0$, $(a\lambda-b) > 0$, $(\lambda^3-b) > 0$, and $(\lambda^4+br) > 0$. Since the fraction outside of the curly bracket is positive, it is enough to show that the sum in the curly bracket is negative. It can be shown that the sum is equal to

$$
\begin{aligned}
&4br\big(a-\lambda^2\big)\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^3-b\big)-6\lambda^2\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^3-b\big)\big(\lambda^4+br\big)\\
&\quad-\frac{15}{2}r\lambda^3\big(a-\lambda^2\big)(a\lambda-b)^2\big(\lambda^3-b\big)-a\big(\lambda^3+ar\big)\big(\lambda^3-b\big)^2\big(\lambda^4+br\big)\\
&\quad-\frac{15}{2}\lambda^3\big(a-\lambda^2\big)\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^4+br\big)-\frac{5}{2}\lambda^4\big(a-\lambda^2\big)\big(\lambda^3+ar\big)(a\lambda-b)\big(\lambda^3-b\big)\\
&\quad<0.
\end{aligned}
$$

This leads to the required result.

D.3 *(iii) $a \le 0$ and $b \ge 0$*

We consider two cases of (i-a) $(-ar)^{1/3} < b^{1/3}$ and $b^{1/3} < (-ar)^{1/3}$ separately.

D.3.1 *(iii-1) $a \le 0$, $b \ge 0$, and $(-ar)^{1/3} < b^{1/3}$* Since h_0 is positive, we can restrict our attention to the values of λ which satisfy $(-ar)^{1/3} < \lambda < b^{1/3}$ by equation (D.3). The first-order condition (D.2) can be expressed as

$$
4\cdot\left(\frac{2}{3}\right)^{5/2}\cdot\frac{|\alpha_1|^7}{|\beta_1|^5}\cdot\frac{\lambda^3\big(\lambda^2-a\big)^{7/2}\big(\lambda^3+ar\big)^{5/2}(b-a\lambda)}{\big(b-\lambda^3\big)^{5/2}\big(\lambda^4+br\big)^{7/2}}=\frac{\gamma_0}{n}.
$$

Denote

$$
f_{31}(\lambda)\equiv\frac{\lambda^3\big(\lambda^2-a\big)^{7/2}\big(\lambda^3+ar\big)^{5/2}(b-a\lambda)}{\big(b-\lambda^3\big)^{5/2}\big(\lambda^4+br\big)^{7/2}}.
$$

Since $\lim_{\lambda\downarrow(-ar)^{1/3}} f_{31}(\lambda)=0$ and $\lim_{\lambda\uparrow b^{1/3}} f_{31}(\lambda)=\infty$, the uniqueness of λ that satisfy the first-order condition follows if $f_{31}(\lambda)$ is monotonically increasing. The derivative of $f_{31}(\lambda)$ is given by

$$
\begin{aligned}
\frac{df_{31}(\lambda)}{d\lambda}=&\frac{\lambda^2\big(\lambda^2-a\big)^{5/2}\big(\lambda^3+ar\big)^{3/2}}{\big(b-\lambda^3\big)^{7/2}\big(\lambda^4+br\big)^{9/2}}\\
&\times\Big\{3\big(\lambda^2-a\big)\big(\lambda^3+ar\big)(b-a\lambda)\big(b-\lambda^3\big)\big(\lambda^4+br\big)
\end{aligned}
$$

$$
\begin{aligned}
&+7\lambda^2(\lambda^3+ar)(b-a\lambda)(b-\lambda^3)(\lambda^4+br)\\
&+\frac{15}{2}\lambda^3(\lambda^2-a)(b-a\lambda)(b-\lambda^3)(\lambda^4+br)\\
&-a\lambda(\lambda^2-a)(\lambda^3+ar)(b-\lambda^3)(\lambda^4+br)\\
&+\frac{15}{2}\lambda^3(\lambda^2-a)(\lambda^3+ar)(b-a\lambda)(\lambda^4+br)\\
&-14\lambda^4(\lambda^2-a)(\lambda^3+ar)(b-a\lambda)(b-\lambda^3)\Big\}.
\end{aligned}
$$

Observe that, for λ which satisfies $(-ar)^{1/3} < \lambda < b^{1/3}$, we have $(\lambda^2 - a) > 0$, $(\lambda^3 + ar) > 0$, $(b - a\lambda) > 0$, $(b - \lambda^3) > 0$, and $(\lambda^4 + br) > 0$. Since the fraction outside of the curly bracket is positive, it is enough to show that the sum in the curly bracket is positive. It can be shown that the sum is equal to

$$
\begin{aligned}
&3br(\lambda^2-a)(\lambda^3+ar)(b-a\lambda)(b-\lambda^3)+\frac{15}{2}r\lambda^3(\lambda^2-a)(b-a\lambda)^2(b-\lambda^3)\\
&\quad-a\lambda(\lambda^2-a)(\lambda^3+ar)(b-a\lambda)^2(b-\lambda^3)\\
&\quad+\frac{1}{2}\lambda^3(\lambda^3+ar)(b-a\lambda)\{8\lambda^7-22a\lambda^5+7b\lambda^4+br\lambda^3+7ab\lambda^2-15abr\lambda+14b^2r\}.
\end{aligned}
$$

The first three terms are positive and the last term is also positive when a is close to zero. This leads to the required result.

D.3.2 *(iii-2)* $a \le 0$, $b \ge 0$, *and* $b^{1/3} < (-ar)^{1/3}$ Since h_0 is positive, we can restrict our attention to the values of λ which satisfy $b^{1/3} < \lambda < (-ar)^{1/3}$ by equation (D.3). The first-order condition (D.2) can be expressed as

$$
4\cdot\left(\frac{2}{3}\right)^{5/2}\cdot\frac{|\alpha_1|^7}{|\beta_1|^5}\cdot\frac{\lambda^3(\lambda^2-a)^{7/2}(-\lambda^3-ar)^{5/2}(b-a\lambda)}{(\lambda^3-b)^{5/2}(\lambda^4+br)^{7/2}}=\frac{\gamma_0}{n}.
$$

Denote

$$
f_{32}(\lambda)\equiv\frac{\lambda^3(\lambda^2-a)^{7/2}(-\lambda^3-ar)^{5/2}(b-a\lambda)}{(\lambda^3-b)^{5/2}(\lambda^4+br)^{7/2}}.
$$

Since $\lim_{\lambda\downarrow b^{1/3}} f_{32}(\lambda) = \infty$ and $\lim_{\lambda\uparrow(-ar)^{1/3}} f_{32}(\lambda) = 0$, the uniqueness of λ that satisfy the first-order condition follows if $f_{32}(\lambda)$ is monotonically decreasing. The derivative of $f_{32}(\lambda)$ is given by

$$
\begin{aligned}
\frac{df_{31}(\lambda)}{d\lambda}&=\frac{\lambda^2(\lambda^2-a)^{5/2}(-\lambda^3-ar)^{3/2}}{(\lambda^3-b)^{7/2}(\lambda^4+br)^{9/2}}\\
&\quad\times\Big\{3(\lambda^2-a)(-\lambda^3-ar)(b-a\lambda)(\lambda^3-b)(\lambda^4+br)\\
&\quad+7\lambda^2(-\lambda^3-ar)(b-a\lambda)(\lambda^3-b)(\lambda^4+br)
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{15}{2}\lambda^3\big(\lambda^2-a\big)(b-a\lambda)\big(\lambda^3-b\big)\big(\lambda^4+br\big)\\
&-a\lambda\big(\lambda^2-a\big)\big(-\lambda^3-ar\big)\big(\lambda^3-b\big)\big(\lambda^4+br\big)\\
&-\frac{15}{2}\lambda^3\big(\lambda^2-a\big)\big(-\lambda^3-ar\big)(b-a\lambda)\big(\lambda^4+br\big)\\
&-14\lambda^4\big(\lambda^2-a\big)\big(-\lambda^3-ar\big)(b-a\lambda)\big(\lambda^3-b\big)\bigg\}.
\end{aligned}
$$

Observe that, for λ which satisfies $b^{1/3} < \lambda < (-ar)^{1/3}$, we have $(\lambda^2-a) > 0$, $(-\lambda^3-ar) > 0$, $(b-a\lambda) > 0$, $(\lambda^3-b) > 0$, and $(\lambda^4+br) > 0$. Since the fraction outside of the curly bracket is positive, it is enough to show that the sum in the curly bracket is negative. It can be shown that the sum is equal to

$$
\begin{aligned}
&-3b\big(\lambda^2-a\big)\big(-\lambda^3-ar\big)(b-a\lambda)\big(\lambda^4+br\big)-7\lambda^2\big(-\lambda^3-ar\big)(b-a\lambda)^2\big(\lambda^4+br\big)\\
&\quad-b\lambda\big(\lambda^2-a\big)^2\big(-\lambda^3-ar\big)\big(\lambda^4+br\big)-\frac{1}{2}\lambda^3\big(\lambda^2-a\big)\big(-\lambda^3-ar\big)\\
&\quad\times\left\{6\lambda^4\big(-\lambda^3-ar\big)+15r\big(\lambda^3-b\big)(b-a\lambda)+13b\lambda^4+7br\lambda^3+13abr\lambda+7abr^2\right\}.
\end{aligned}
$$

The first three terms are negative and the last term is also negative when a is close to zero. This leads to the required result.

D.4 *(iv) $a \le 0$ and $b \le 0$*

For this case, we consider the original parametrization (D.1). Denote

$$
\begin{aligned}
L &= \big(\alpha_1 h_1^2-\alpha_0 h_0^2\big)^2+\big(\beta_1 h_1^3-\beta_0 h_0^3\big)^2+\frac{\gamma_1}{nh_1}+\frac{\gamma_0}{nh_0}\\
&\equiv g_1+g_2+g_3.
\end{aligned}
$$

First, we show that g_1 and g_2 are convex. For this purpose, it suffices to show their Hessian matrices are nonnegative definite. Observe that

$$
\begin{aligned}
\frac{\partial^2 g_1}{\partial h_1^2} &= 4\big(3\alpha_1^2 h_1^2-\alpha_1\alpha_0 h_0^2\big)\ge 0,\\
\frac{\partial^2 g_1}{\partial h_1^2}\cdot\frac{\partial^2 g_1}{\partial h_0^2}-\left(\frac{\partial^2 g_1}{\partial h_1\,\partial h_0}\right)^2 &= -48\alpha_1\alpha_0\big(\alpha_1 h_1^2-\alpha_0 h_0^2\big)^2\ge 0,\\
\frac{\partial^2 g_2}{\partial h_1^2} &= 30\beta_1^2 h_1^4-12\beta_1\beta_0 h_1 h_0^3\ge 0,\\
\frac{\partial^2 g_2}{\partial h_1^2}\cdot\frac{\partial^2 g_2}{\partial h_0^2}-\left(\frac{\partial^2 g_2}{\partial h_1\,\partial h_0}\right)^2 &= -360\beta_1\beta_0 h_1 h_0\big(\beta_1 h_1^3-\beta_0 h_0^3\big)^2\ge 0.
\end{aligned}
$$

It follows that g_1 and g_2 are convex. In addition, the calculation above show that g_1+g_2 is strictly convex unless $a=b=0$. In the same manner, we can show that g_3 is strictly convex since

$$\frac{\partial^2 g_3}{\partial h_1^2} = \frac{2\gamma_1}{nh_1^3} > 0,$$

$$\frac{\partial^2 g_3}{\partial h_1^2} \cdot \frac{\partial^2 g_3}{\partial h_0^2} - \left(\frac{\partial^2 g_1}{\partial h_1 \, \partial h_0}\right)^2 = \frac{4\gamma_1\gamma_0}{n^2 h_1^3 h_3^3} > 0.$$

These imply that L is strictly convex, leading to the required result.

Given the uniqueness of $\hat{h}$, the continuity of $\hat{h}$ with respect to a immediately follows from the implicit function theorem.

E. Proofs

E.1 *Proof of Lemmas 1 and 3*

For both lemmas, a contribution to the MSE from a variance component is standard. See Fan and Gijbels (1996, Section 3.2) for the details. Here, we consider the contribution made by the bias component. For the bias component, we provide a proof for the general results which encompass both lemmas.

Let $\hat{\beta}_{1,p}(h_1)$ and $\hat{\beta}_{0,p}(h_0)$ be the LPR estimators of order p ($p \in \mathbb{N}$) on the right and left of the cut-off point with bandwidth h_1 and h_0, respectively. Then the LPR estimator can be expressed as $\hat{\beta}_{j,p}(h_j) = (X_p(c)'W_{j,h_j}(c)X_p(c))^{-1}X_p(c)'W_{j,h_j}(c)Y$ for $j = 0,1$, where $X_p(c)$ is an $n \times (p+1)$ matrix whose ith row is given by $(1, X_i - c, \ldots, (X_i - c)^p)$, $Y = (Y_1, \ldots, Y_n)'$, $W_{j,h_j}(c) = \mathrm{diag}(K_{j,h_j}(X_i - c))$, $K_{1,h}(\cdot) = K(\cdot/h)\mathbb{I}\{\cdot \geq 0\}/h$ and $K_{0,h}(\cdot) = K(\cdot/h)\mathbb{I}\{\cdot < 0\}/h$. The LPR estimator of $m_j^{(\nu)}(c)$ based on the LPR of order p can be written as $\hat{m}_{j,p}^{(\nu)}(c) = \nu! e_\nu \hat{\beta}_{j,p}(h_j)$ for $j = 0,1$, where e_ν is the conformable unit vector having one in the $(\nu+1)$th entry and zero in the other entry.

Lemmas 1 and 3 can be obtained by applying the next lemma to construct the bias of $\hat{\tau}_{\nu,p} = \hat{m}_{1,p}^{(\nu)}(c) - \hat{m}_{0,p}^{(\nu)}(c)$ by choosing $(\nu, p) = (0,1)$ for the SRD and $(\nu, p) = (1,2)$ for the SRK, respectively.

Lemma E.1. *Suppose Assumptions* 1–4 *and* 5 *with* $\kappa = p+2$ *hold. Then it follows for* $j = 0,1$ *and* $p \in \mathbb{N}$, *that*

$$\begin{aligned}
\mathrm{Bias}\big(\hat{\beta}_{j,p}(h_1)|X\big) &= h^{p+1} H_p^{-1}(h_j) \frac{m_j^{(p+1)}(c)}{(p+1)!} S_{j,0,p}^{-1} c_{j,p+1,p} \\
&\quad + h^{p+2} H_p^{-1}(h_j) \bigg\{ \bigg(\frac{m_j^{(p+1)}(c)}{(p+1)!} \frac{f^{(1)}(c)}{f(c)} + \frac{m_j^{(p+2)}(c)}{(p+2)!} \bigg) S_{j,0,p}^{-1} c_{j,p+2,p} \\
&\quad - \frac{m_j^{(p+1)}(c)}{(p+1)!} \frac{f^{(1)}(c)}{f(c)} S_{j,0,p}^{-1} S_{j,1,p} S_{j,0,p}^{-1} c_{j,p+1,p} \bigg\} + H_p^{-1}(h_j) r_{j,n},
\end{aligned}$$

where $H_p(h_j) = \mathrm{diag}(1, h_j, \ldots, h_j^p)$, $S_{j,k,p} = (\mu_{j,k+\ell_1+\ell_2})_{0 \leq \ell_1, \ell_2 \leq p}$, $c_{j,k,p} = (\mu_{j,k+\ell})_{0 \leq \ell \leq p}$, $\mu_{1,s} = \int_0^\infty u^s K(s)\, ds$, $\mu_{0,s} = \int_{-\infty}^0 u^s \, ds$ *and* $r_{j,n} = o_p(h_j^{p+2})$.

PROOF. The conditional bias of $\hat{\beta}_{j,p}(h_j)$ can be expressed as, for $j = 0, 1$,

$$\text{Bias}\big(\hat{\beta}_{j,p}(h_j)|X\big) = \big(X_p(c)'W_{j,h_j}(c)X_p(c)\big)^{-1}X_p(c)'W_{j,h_j}(c)\big(m_j - X_p(c)\beta_{j,p}\big),$$

where $m_j = (m_j(X_1), \dots, m_j(X_n))'$ and $\beta_{j,p} = (m_j(c), \dots, m_j^{(p)}(c)/p!)'$. Define, for a nonnegative integer k, $(p+1)\times(p+1)$ matrices $S_{1,k,p}(h_1)$ and $(p+1)\times 1$ vectors $c_{j,k,p}(h_j)$ as

$$S_{j,k,p}(h_j) = \big(s_{j,k+\ell_1+\ell_2}(h_j)\big)_{0\le \ell_1,\ell_2\le p}, \qquad c_{j,k,p}(h_1) = \big(s_{j,k+\ell}(h_j)\big)_{0\le \ell\le p},$$
$$s_{j,k}(h_j) = \sum_{i=1}^{n} K_{j,h_j}(X_i - c)(X_i - c)^k. \tag{E.1}$$

Note that $S_{j,0,p}(h_j) = X_p(c)'W_{j,h_j}(c)X_p(c)$. The argument made by Fan, Gijbels, Hu, and Huang (1996) can be generalized to yield

$$s_{j,k}(h_j) = nh_j^k\big\{f(c)\mu_{j,k} + h_j f^{(1)}(c)\mu_{j,k+1} + o_p(h_j)\big\}. \tag{E.2}$$

Then it follows that

$$S_{j,0,p}(h_1) = nH_p(h_j)\big\{f(c)S_{j,0,p} + h_j f^{(1)}(c)S_{j,1,p} + o_p(h_j)\big\}H_p(h_j).$$

By using the fact that $(A+hB)^{-1} = A^{-1} - hA^{-1}BA^{-1} + o(h)$, we obtain

$$S_{j,0,p}^{-1}(h_j) = n^{-1}H_p^{-1}(h_j)\left\{\frac{1}{f(c)}S_{j,0,p}^{-1} - h_j\frac{f^{(1)}(c)}{f(c)^2}S_{j,0,p}^{-1}S_{j,1,p}S_{j,0,p}^{-1} + o_p(h_j)\right\} \times H_p^{-1}(h_j). \tag{E.3}$$

Next, we consider $X_p(c)'W_{j,h_j}(c)\{m_j - X_p(c)\beta_{j,p}\}$. A Taylor expansion of $m_j(\cdot)$ yields

$$X_p(c)'W_{j,h_j}(c)\big\{m_j - X_p(c)\beta_{j,p}\big\} = \frac{m_j^{(p+1)}(c)}{(p+1)!}c_{j,p+1,p}(h_j) + \frac{m_j^{(p+2)}(c)}{(p+2)!}c_{j,p+2,p}(h_j) + nH_p(h_j)r_{j,n}. \tag{E.4}$$

The definition of $c_{j,k,j}$ in (E.1), in conjunction with (E.2), yields

$$c_{j,k,p}(h_j) = nh_j^k H_p(h_j)\big\{f(c)c_{j,k,p} + h_j f^{(1)}(c)c_{j,k+1,p} + o_p(h_j)\big\}. \tag{E.5}$$

Combining this with (E.3) and (E.4) gives

$$\begin{aligned}
&X_p(c)'W_{j,h_j}(c)\big\{m_j - X_p(c)\beta_{j,p}\big\}\\
&= nh_j^{p+1}H_p(h_j)\Bigg\{\frac{m_j^{(p+1)}(c)}{(p+1)!}f(c)c_{j,p+1,p}\\
&\quad + h_j\bigg(\frac{m_j^{(p+1)}(c)}{(p+1)!}f^{(1)}(c) + \frac{m_j^{(p+2)}(c)}{(p+2)!}f(c)\bigg)c_{j,p+2,p} + o_p(h_j)\Bigg\}.
\end{aligned} \tag{E.6}$$

Combining (E.3) and (E.6) gives the required result. □

E.2 *Proofs of Theorem 2*

Recall that the objective function is

$$\text{MMSE}_n^p(h) = \{d_1[\hat{m}_1^{(3)}(c)h_1^2 - \hat{m}_0^{(3)}(c)h_0^2]\}^2 + [\hat{d}_{2,1}(c)h_1^3 - \hat{d}_{2,0}(c)h_0^3]^2 + \frac{w}{n\hat{f}(c)}\left\{\frac{\hat{\sigma}_1^2(c)}{h_1^3} + \frac{\hat{\sigma}_0^2(c)}{h_0^3}\right\}.$$

To begin with, we show that $\hat{h}_1$ and $\hat{h}_0$ satisfy Assumption 2. If we choose a sequence of h_1 and h_0 to satisfy Assumption 2, then $\text{MMSE}_n^p(h)$ converges to 0. Assume to the contrary that either one or both of $\hat{h}_1$ and $\hat{h}_0$ do not satisfy Assumption 2. Since $m_0^{(3)}(c)^3 d_{2,1}(c)^2 \neq m_1^{(3)}(c)^3 d_{2,0}(c)^2$ by assumption, $\hat{m}_0^{(3)}(c)^3 \hat{d}_{2,1}(c)^2 \neq \hat{m}_1^{(3)}(c)^3 \hat{d}_{2,0}(c)^2$ with probability approaching 1. Without loss of generality, we assume this as well. Then at least one of the first-order bias term, the second-order bias term and the variance term of $\text{MMSE}_n^p(\hat{h})$ does not converge to zero in probability. Then $\text{MMSE}_n^p(\hat{h}) > \text{MMSE}_n^p(h)$ holds for some n. This contradicts the definition of $\hat{h}$. Hence $\hat{h}$ satisfies Assumption 2.

We first consider the case in which $m_1^{(3)}(c)m_0^{(3)}(c) < 0$. In this case, with probability approaching 1, $\hat{m}_1^{(3)}(c)\hat{m}_0^{(3)}(c) < 0$, so that we assume this without loss of generality. When this holds, note that the leading terms are the first term and the last term of $\text{MMSE}_n^p(\hat{h})$ since $\hat{h}_1$ and $\hat{h}_0$ satisfy Assumption 2. Define the plug-in version of $\text{AMSE}_{1n}(h)$ provided in Definition 2 by

$$\text{AMSE}_{1n}^p(h) = \{d_1[\hat{m}_1^{(3)}(c)h_1^2 - \hat{m}_0^{(3)}(c)h_0^2]\}^2 + \frac{w}{n\hat{f}(c)}\left\{\frac{\hat{\sigma}_1^2(c)}{h_1^3} + \frac{\hat{\sigma}_0^2(c)}{h_0^3}\right\}.$$

A calculation yields $\tilde{h}_1 = \tilde{C}_1 n^{-1/7}$ and $\tilde{h}_0 = \tilde{C}_0 n^{-1/7}$ where

$$\tilde{C}_1 = \left\{\frac{3w\hat{\sigma}_1^2(c)}{4d_1^2\hat{f}(c)\hat{m}_1^{(3)}(c)[\hat{m}_1^{(3)}(c) - \hat{\lambda}_1^2\hat{m}_0^{(3)}(c)]}\right\}^{1/7}, \quad \hat{\lambda}_1 = \left\{-\frac{\hat{\sigma}_0^2(c)\hat{m}_1^{(3)}(c)}{\hat{\sigma}_1^2(c)\hat{m}_0^{(3)}(c)}\right\}^{1/5},$$

and $\tilde{C}_0 = \tilde{C}_1\hat{\lambda}_1$. With this choice, $\text{AMSE}_{1n}^p(\tilde{h})$, and hence $\text{MMSE}_n^p(\tilde{h})$ converges at the rate of $n^{-4/7}$. Note that if $\hat{h}_1$ or $\hat{h}_0$ converges at the rate slower than $n^{-1/7}$, then the bias term converges at the rate slower than $n^{-4/7}$. If $\hat{h}_1$ or $\hat{h}_0$ converges at the rate faster than $n^{-1/7}$, then the variance term converges at the rate slower than $n^{-4/7}$. Thus, the minimizer of $\text{MMSE}_n^p(h)$, $\hat{h}_1$, and $\hat{h}_0$ converges to 0 at rate $n^{-1/7}$.

Thus, we can write $\hat{h}_1 = \hat{C}_1 n^{-1/7} + o_p(n^{-1/7})$ and $\hat{h}_0 = \hat{C}_0 n^{-1/7} + o_p(n^{-1/7})$ for some $O_P(1)$ sequences $\hat{C}_1$ and $\hat{C}_0$ that are bounded away from 0 and ∞ as $n \to \infty$. Using this expression,

$$\text{MMSE}_n^p(\hat{h}) = n^{-4/7}\{d_1[\hat{m}_1^{(3)}(c)\hat{C}_1^2 - \hat{m}_0^{(3)}(c)\hat{C}_0^2]\}^2 + \frac{w}{n^{4/7}\hat{f}(c)}\left\{\frac{\hat{\sigma}_1^2(c)}{\hat{C}_1^3} + \frac{\hat{\sigma}_0^2(c)}{\hat{C}_0^3}\right\} + o_p(n^{-4/7}).$$

Note that

$$\mathrm{MMSE}_n^p(\tilde{h}) = n^{-4/7}\{d_1[\hat{m}_1^{(3)}(c)\tilde{C}_1^2 - \hat{m}_0^{(3)}(c)\tilde{C}_0^2]\}^2 + \frac{w}{n^{4/7}\hat{f}(c)}\left\{\frac{\hat{\sigma}_1^2(c)}{\tilde{C}_1^3} + \frac{\hat{\sigma}_0^2(c)}{\tilde{C}_0^3}\right\} + O_P(n^{-6/7}).$$

Since $\hat{h}$ is the optimizer, $\mathrm{MMSE}_n^p(\hat{h})/\mathrm{MMSE}_n^p(\tilde{h}) \leq 1$. Thus,

$$\frac{\{d_1[\hat{m}_1^{(3)}(c)\hat{C}_1^2 - \hat{m}_0^{(3)}(c)\hat{C}_0^2]\}^2 + \frac{w}{\hat{f}(c)}\left\{\frac{\hat{\sigma}_1^2(c)}{\hat{C}_1^3} + \frac{\hat{\sigma}_0^2(c)}{\hat{C}_0^3}\right\} + o_P(1)}{\{d_1[\hat{m}_1^{(3)}(c)\tilde{C}_1^2 - \hat{m}_0^{(3)}(c)\tilde{C}_0^2]\}^2 + \frac{w}{\hat{f}(c)}\left\{\frac{\hat{\sigma}_1^2(c)}{\tilde{C}_1^3} + \frac{\hat{\sigma}_0^2(c)}{\tilde{C}_0^3}\right\} + O_P(n^{-2/7})} \leq 1.$$

Note that the denominator converges to

$$\{d_1[m_1^{(3)}(c)C_1^{*2} - m_0^{(3)}(c)C_0^{*2}]\}^2 + \frac{w}{f(c)}\left\{\frac{\sigma_1^2(c)}{C_1^{*3}} + \frac{\sigma_0^2(c)}{C_0^{*3}}\right\},$$

where C_1^* and C_0^* are the unique optimizers of

$$\{d_1[m_1^{(3)}(c)C_1^2 - m_0^{(3)}(c)C_0^2]\}^2 + \frac{w}{f(c)}\left\{\frac{\sigma_1^2(c)}{C_1^3} + \frac{\sigma_0^2(c)}{C_0^3}\right\},$$

with respect to C_1 and C_0. This implies that $\hat{C}_1$ and $\hat{C}_0$ also converge to the same respective limit C_1^* and C_0^* because the inequality will be violated otherwise.

Next, we consider the case with $m_1^{(3)}(c)m_0^{(3)}(c) > 0$. In this case, with probability approaching 1, $\hat{m}_1^{(3)}(c)\hat{m}_0^{(3)}(c) > 0$, so that we assume this without loss of generality.

When these conditions hold, define $h_0 = \hat{\lambda}_2 h_1$ where $\hat{\lambda}_2 = \{\hat{m}_1^{(3)}(c)/\hat{m}_0^{(3)}(c)\}^{1/2}$. This sets the first-order bias term of $\mathrm{MMSE}_n^p(h)$ equal to 0. Define the plug-in version of $\mathrm{AMSE}_{2n}(h)$ by

$$\mathrm{AMSE}_{2n}^p(h) = \{\hat{d}_{2,1}(c)h_1^3 - \hat{d}_{2,0}(c)h_0^3\}^2 + \frac{w}{n\hat{f}(c)}\left\{\frac{\hat{\sigma}_1^2(c)}{h_1^3} + \frac{\hat{\sigma}_0^2(c)}{h_0^3}\right\}.$$

Choosing h_1 to minimize $\mathrm{AMSE}_{2n}^p(h)$, we define $\tilde{h}_1 = \tilde{C}_1 n^{-1/9}$ and $\tilde{h}_0 = \tilde{C}_0 n^{-1/9}$ where

$$\hat{\theta}_2 = \left\{\frac{w[\hat{\sigma}_1^2(c) + \hat{\sigma}_0^2(c)/\hat{\lambda}_2^3]}{2\hat{f}(c)[\hat{d}_{2,1}(c) - \hat{\lambda}_2^3\hat{d}_{2,0}(c)]^2}\right\}^{1/9} \quad \text{and} \quad \tilde{C}_0 = \tilde{C}_1\hat{\lambda}_2. \tag{E.7}$$

Then $\mathrm{MMSE}_n^p(\tilde{h})$ can be written as

$$\mathrm{MMSE}_n^p(\tilde{h}) = n^{-6/9}\{\hat{d}_{2,1}(c)\tilde{C}_1^3 - \hat{d}_{2,0}(c)\tilde{C}_0^3\}^2 + n^{-6/9}\frac{w}{\hat{f}(c)}\left\{\frac{\hat{\sigma}_1^2(c)}{\tilde{C}_1^3} + \frac{\hat{\sigma}_0^2(c)}{\tilde{C}_{0^3}}\right\}.$$

In order to match this rate of convergence, both $\hat{h}_1$ and $\hat{h}_0$ need to converge at the rate slower than or equal to $n^{-1/9}$ because the variance term needs to converge at the rate

$n^{-6/9}$ or faster. In order for the first-order bias term to match this rate,

$$\hat{m}_1^{(3)}(c)\hat{h}_1^2 - \hat{m}_0^{(3)}(c)\hat{h}_0^2 \equiv B_{1n} = n^{-3/9}b_{1n},$$

where $b_{1n} = O_P(1)$ so that under the assumption that $m_0^{(3)}(c) \neq 0$, with probability approaching 1, $\hat{m}_0^{(3)}(c)$ is bounded away from 0 so that assuming this without loss of generality, we have $\hat{h}_0^2 = \hat{\lambda}_2^2\hat{h}_1^2 - B_{1n}/\hat{m}_0^{(3)}(c)$. Substituting this expression to the second term and the third term of MMSE_n^p, we have

$$\mathrm{MMSE}_n^p(\hat{h}) = \{d_1B_{1n}\}^2 + \left\{\hat{d}_{2,1}(c)\hat{h}_1^3 - \hat{d}_{2,0}(c)\left[\hat{\lambda}_2^2\hat{h}_1^2 - B_{1n}/\hat{m}_0^{(2)}(c)\right]^{3/2}\right\}^2 + \frac{w}{n\hat{f}(c)}\left\{\frac{\hat{\sigma}_1^2(c)}{\hat{h}_1} + \frac{\hat{\sigma}_0^2(c)}{\left[\hat{\lambda}_2^2\hat{h}_1^2 - B_{1n}/\hat{m}_0^{(2)}(c)\right]^{3/2}}\right\}.$$

Suppose $\hat{h}_1$ is of order slower than $n^{-1/9}$. Then because $\hat{m}_0^{(3)}(c)^3\hat{d}_{2,1}(c)^2 \neq \hat{m}_1^{(3)}(c)^3\hat{d}_{2,0}(c)^2$ and this holds even in the limit, the second-order bias term is of order slower than $n^{-6/9}$. If $\hat{h}_1$ converges to 0 faster than $n^{-1/9}$, then the variance term converges at the rate slower than $n^{-6/9}$. Therefore, we can write $\hat{h}_1 = \hat{C}_1n^{-1/9} + o_p(n^{-1/9})$ for some $O_P(1)$ sequence $\hat{C}_1$ that is bounded away from 0 and ∞ as $n \to \infty$ and as before $\hat{h}_0^2 = \hat{\lambda}_2^2\hat{h}_1^2 - B_{1n}/\hat{m}_0^{(3)}(c)$. Using this expression, we can write

$$\begin{aligned}\mathrm{MMSE}_n^p(\hat{h}) &= n^{-6/9}\{d_1b_{1n}\}^2 \\ &\quad + n^{-6/9}\left\{\hat{d}_{2,1}(c)\hat{C}_1^3 + o_p(1) - \hat{d}_{2,0}(c)\left[\hat{\lambda}_2^2\hat{C}_1^2 + o_p(1) - n^{-1/9}b_{1n}/\hat{m}_0^{(3)}(c)\right]^{3/2}\right\}^2 \\ &\quad + n^{-6/9}\frac{w}{\hat{f}(c)}\left\{\frac{\hat{\sigma}_1^2(c)}{\hat{C}_1^3 + o_p(1)} + \frac{\hat{\sigma}_0^2(c)}{\left[\hat{\lambda}_2^2\hat{C}_1^2 + o_p(1) - n^{-1/9}b_{1n}/\hat{m}_0^{(3)}(c)\right]^{3/2}}\right\}.\end{aligned}$$

Thus, b_{1n} converges in probability to 0. Otherwise the first-order bias term remains and that contradicts the definition of $\hat{h}_1$.

Since $\hat{h}$ is the optimizer, $\mathrm{MMSE}_n^p(\hat{h})/\mathrm{MMSE}_n^p(\tilde{h}) \leq 1$. Thus,

$$\frac{o_p(1) + \left\{\hat{d}_{2,1}(c)\hat{C}_1^3 - \hat{d}_{2,0}(c)\left[\hat{\lambda}_2^2\hat{C}_1^2 + o_p(1)\right]^{3/2}\right\}^2 + \frac{w}{\hat{f}(c)}\left\{\frac{\hat{\sigma}_1^2(c)}{\hat{C}_1^3 + o_p(1)} + \frac{\hat{\sigma}_0^2(c)}{\left[\hat{\lambda}_2^2\hat{C}_1^2 + o_p(1)\right]^{3/2}}\right\}}{\left\{\hat{d}_{2,1}(c)\tilde{C}_1^3 - \hat{d}_{2,0}(c)\tilde{C}_0^3\right\}^2 + \frac{w}{\hat{f}(c)}\left\{\frac{\hat{\sigma}_1^2(c)}{\tilde{C}_1^3} + \frac{\hat{\sigma}_0^2(c)}{\tilde{C}_0^3}\right\}} \leq 1.$$

If $\hat{C}_1 - \tilde{C}_1$ does not converge to 0 in probability, then the ratio is not less than one at some point. Hence $\hat{C}_1 - \tilde{C}_1 = o_p(1)$. Therefore, $\hat{h}_0/\tilde{h}_0 \xrightarrow{p} 1$ as well.

The results shown above also imply that $\mathrm{MMSE}_n^p(\hat{h})/\mathrm{MSE}_n(h^*) \xrightarrow{p} 1$ in both cases. □

E.3 *Proofs of Theorem 3*

Since the results follow in the same manner as those of Theorems 1 and 2 of CCT, we outline the proof of Theorem 3. As in Calonico, Cattaneo, and Farrell (forthcoming), we restrict our attention to the case where $h_j/h_{p+1,j} \to \rho_{j,p+1}$, $h_j/h_{p+2,j} \to \rho_{j,p+2}$, and $h_{p+1,j}/h_{p+2,j} \to \rho_{j,p+1,p+2}$ with $\rho_{j,p+1}$ $\rho_{j,p+2}$ and $\rho_{j,p+1,p+2}$ bounded for $j = 0, 1$. (See Remark 2 of Calonico, Cattaneo, and Farrell (forthcoming).)

First, we note that, by Lemmas S.A.1 and S.A.2 of CCT,

$$\begin{aligned}
V^{(0)}_{j,\nu,p}(h_j) &= O_p\left(\frac{1}{nh_j^{2\nu+1}}\right),\\
V^{(1)}_{j,\nu,p,q}(h_j, h_{p+1,j}) &= O_p\left(\frac{h_j^{2(p-\nu+1)}}{nh_{p+1,j}^{2p+3}}\right),\\
V^{(2)}_{j,\nu,p,q}(h_j, h_{p+1,j}, h_{p+2,j}) &= O_p\left(\frac{h_j^{2(p-\nu+2)}}{nh_{p+2,j}^{2p+5}}\right),\\
C^{(0,1)}_{j,\nu,p,q}(h_j, h_{p+1,j}) &= O_p\left(\frac{h_j^{p-\nu+1}\min\{h_j, h_{p+1,j}\}}{nh_j^{\nu+1}h_{p+1,j}^{p+2}}\right),\\
C^{(0,2)}_{j,\nu,p,q}(h_j, h_{p+1,j}, h_{p+2,j}) &= O_p\left(\frac{h_j^{p-\nu+2}\min\{h_j, h_{p+2,j}\}}{nh_j^{\nu+1}h_{p+2,j}^{p+3}}\right),\\
C^{(1,2)}_{j,\nu,p,q}(h_j, h_{p+1,j}, h_{p+2,j}) &= O_p\left(\frac{h_j^{2p-2\nu+3}\min\{h_{p+1,j}, h_{p+2,j}\}}{nh_{p+1,j}^{p+2}h_{p+2,j}^{p+3}}\right).
\end{aligned} \tag{E.8}$$

These, together with Lemma E.1 and the notation introduced in Appendix B of the main text, imply

$$\begin{aligned}
E\big[\hat{\tau}^{\text{bc}}_{j,\nu,p,q}(h_j, h_{p+1,j}, h_{p+2,j})|X\big] - \nu! e'_\nu \beta_{j,p} &= O_p\big(h_j^{p+3-\nu} + h_j^{p+1-\nu}h_{p+1,j}^{q-p} + h_j^{p+2-\nu}h_{p+2,j}^{q-p-1}\big),\\
V^{\text{bc}}_{j,\nu,p,q}(h_j, h_{p+1,j}, h_{p+2,j}) &= O_p\left(\frac{1}{nh_j^{2\nu+1}} + \frac{h_j^{2(p-\nu+1)}}{nh_{p+1,j}^{2p+3}} + \frac{h_j^{2(p-\nu+2)}}{nh_{p+2,j}^{2p+5}}\right).
\end{aligned}$$

The rest of the proof follows in the same manner as Lemma S.A.4 of CCT. Observe that

$$\frac{\hat{\tau}^{\text{bc}}_{j,\nu,p,q}(h_j, h_{p+1,j}, h_{p+2,j}) - \nu! e'_\nu \beta_{j,p}}{\sqrt{V^{\text{bc}}_{j,\nu,p,q}(h_j, h_{p+1,j}, h_{p+2,j})}} = \xi_{j,1,n} + \xi_{j,2,n},$$

where

$$\xi_{j,1,n} = \frac{\hat{\tau}^{\text{bc}}_{j,\nu,p,q}(h_j, h_{p+1,j}, h_{p+2,j}) - E\big[\hat{\tau}^{\text{bc}}_{j,\nu,p,q}(h_j, h_{p+1,j}, h_{p+2,j})|X\big]}{\sqrt{V^{\text{bc}}_{j,\nu,p,q}(h_j, h_{p+1,j}, h_{p+2,j})}}$$

and

$$\xi_{j,2,n} = \frac{E\big[\hat{\tau}^{\mathrm{bc}}_{j,\nu,p,q}(h_j, h_{p+1,j}, h_{p+2,j})|X\big] - \nu! e_\nu' \beta_{j,p}}{\sqrt{V^{\mathrm{bc}}_{j,\nu,p,q}(h_j, h_{p+1,j}, h_{p+2,j})}}.$$

First, we note that

$$\begin{aligned}
\xi_{j,2,n}^2 &= O_p\left(\min\left\{nh_j^{2\nu+1}, \frac{nh_{p+1,j}^{2p+3}}{h_j^{2p+2-2\nu}}, \frac{nh_{p+2,j}^{2p+5}}{h_j^{2p+4-2\nu}}\right\}\right) \\
&\quad \times O_p\big(\max\{h_j^{2p+6-2\nu}, h_j^{2p+2-2\nu} h_{p+1,j}^{2q-2p}, h_j^{2p+4-2\nu} h_{p+2,j}^{2q-2p-2}\}\big) \\
&= O_p\left(n \min\left\{h_j^{2p+3}, h_{p+1,j}^{2p+3}, \frac{h_{p+2,j}^{2p+5}}{h_j^2}\right\}\right) \\
&\quad \times O_p\big(\max\{h_j^4, h_{p+1,j}^{2(q-p)}, h_j^2 h_{p+2,j}^{2(q-p-1)}\}\big).
\end{aligned}$$

It follows that $\xi_{j,2,n}^2 = o_p(1)$ since we use $(p,q) = (1,3)$, and h_2 and h_3 of order $n^{-1/9}$ for the sharp RD design and $(p,q) = (2,4)$, and h_3 and h_4 of order $n^{-1/11}$ for the sharp RK design.

Next, we show $\xi_{j,1,n} \xrightarrow{d} N(0,1)$. Note that

$$\xi_{j,1,n} = \sum_{i=1}^n \frac{\zeta_{j,1,n,i}\varepsilon_i}{\zeta_{j,2,n}} + o_p(1),$$

where

$$\begin{aligned}
\varepsilon_{j,i} &= Y_i - m_j(X_i), \\
\zeta_{j,1,n,i} &= h_j^{-\nu} e_\nu' H_p(h_j) S_{j,0,p}^{-1}(h_j) K_{j,h_j}(X_i - c)\phi_p(X_i - c) \\
&\quad - h_j^{p+1-\nu} h_{p+1,j}^{-p-1} \vartheta_{j,\nu,p,p+1} \\
&\quad \times \big\{e_{p+1}' H_q(h_{p+1,j}) S_{j,0,q}^{-1}(h_{p+1,j}) K_{j,h_{p+1,j}}(X_i - c)\phi_q(X_i - c)\big\} \\
&\quad - h_j^{p+2-\nu} h_{p+2,j}^{-p-2} \vartheta_{j,\nu,p,p+2} \\
&\quad \times \big\{e_{p+2}' H_q(h_{p+2,j}) S_{j,0,q}^{-1}(h_{p+2,j}) K_{j,h_{p+2,j}}(X_i - c)\phi_q(X_i - c)\big\}, \\
\phi_p(x) &= \big(1, x, \dots, x^p\big)',
\end{aligned}$$

and we have by (E.8)

$$\zeta_{j,2,n} = \sum_{i=1}^n E\big[\zeta_{j,1,n}^2 \varepsilon_i^2\big] = O\left(\frac{1}{nh_j^{2\nu+1}} + \frac{h_j^{2(p-\nu+1)}}{nh_{p+1,j}^{2p+3}} + \frac{h_j^{2(p-\nu+2)}}{nh_{p+2,j}^{2p+5}}\right).$$

Then the required result follows because the Lyapunov's condition is satisfied

$$\sum_{i=1}^{n} E\left|\frac{\zeta_{j,1,n,i}\varepsilon_i}{\zeta_{j,2,n}}\right|^4 = O\left(\frac{1}{nh_j}\min\{1, \rho_{j,p+1,n}^{-4p-6}, \rho_{j,p+1,n}^{-4p-8}\}\right)$$
$$+ O\left(\frac{1}{nh_{p+1j}}\min\{\rho_{j,p+1,n}^{4p+6}, 1, \rho_{j,p+2,n}^{-4}\rho_{j,p+1,p+2,n}^{-4p-6}\}\right)$$
$$+ O\left(\frac{1}{nh_{p+2,j}}\min\{\rho_{p+2,j}^{-4p+10}, \rho_{j,p+2,n}^{4}\rho_{j,p+1,p+2,n}^{4p+6}, 1\}\right)$$
$$\to 0,$$

where $\rho_{j,p+1,n} = h_j/h_{p+1,j}$, $\rho_{j,p+2,n} = h_j/h_{p+2,j}$, and $\rho_{j,p+1,p+2,n} = h_{p+1,j}/h_{p+2,j}$. □

References

Calonico, S., M. D. Cattaneo, and M. H. Farrell (forthcoming), "On the effect of bias estimation on coverage accuracy in nonparametric inference." *Journal of the American Statistical Association*. [17, 18]

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014a), "Robust nonparametric bias-corrected inference in the regression discontinuity design." *Econometrica*, 82, 2295–2326. [3]

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014b), "Supplement to 'Robust nonparametric confidence intervals for regression-discontinuity design'." *Econometrica Supplemental Material*, 82, http://dx.doi.org/10.3982/ECTA11757. [2]

Fan, J. and I. Gijbels (1996), *Local Polynomial Modeling and Its Applications*. Chapman & Hall, London. [2, 3, 4, 13]

Fan, J., I. Gijbels, T.-C. Hu, and L.-S. Huang (1996), "A study of variable bandwidth selection for local polynomial regression." *Statistica Sinica*, 6, 113–127. [14]

Imbens, G. W. and K. Kalyanaraman (2012), "Optimal bandwidth choice for the regression discontinuity estimator." *Review of Economic Studies*, 79, 933–959. [3]

Jones, M. C. (1994), "On kernel density derivative estimation." *Communications in Statistics. Theory and Methods*, 23, 2133–2139. [3]

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analaysis*. Chapman & Hall, London. [3]

Wand, M. P. and M. C. Jones (1994), *Kernel Smoothing*. Chapman & Hall, London. [3]

Co-editor Rosa L. Matzkin handled this manuscript.

Manuscript received 30 June, 2015; final version accepted 29 April, 2017; available online 21 June, 2017.