Supplement to "Simultaneous Selection of Optimal Bandwidths for the Sharp Regression Discontinuity Estimator"

Yoichi Arai and Hidehiko Ichimura

A Introduction

We present a detailed procedure to obtain RMSE* provided in Tables 1 and 2, a detailed algorithm to implement the proposed method, and a proof of Lemma 1, in this supplemental material.

B A Procedure to Obtain RMSE*

We describe how RMSE* is computed for the LLR estimators based on the MMSE bandwidths, the IND bandwidths, and the IK 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.

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

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

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

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 p.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 $c \leq X_i \leq c + h_{2,+}$ and $c - h_{2,-} \leq X_i < c$, 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} in page 12 in the main text.

C Implementation

To obtain the proposed bandwidths, 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. We obtain these pilot estimates in a number of steps.

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 kernel density estimation is chosen using the normal scale rule with

[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.

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 **?** and **?** for the normal scale rules). The first derivative of the density is estimated using the method proposed by **?**. 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 **?**, $L(u) = -15u(1 - u^2)1_{\{|u|<1\}}/4$. 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 **?**, **?**, and **?**, 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

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

where $j = 0, 1$ (see **?**, 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.

[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.

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 IK, we use the uniform kernel, which yields $C_{2,3} = 5.2088$ and $C_{3,3} = 4.8227$. 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 $c \leq X_i \leq c + h_{2,1}$. 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 $\hat{\beta}_{2,1} = ({X_a}'X_a)^{-1}X_aY_a$. 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 $\hat{\beta}_{3,1}$, $\hat{\beta}_{3,1} = ({X_b}'X_b)^{-1}X_bY_b$, $Y_b = (Y_1, \ldots, Y_{n_b})'$, 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 $c \leq X_i \leq c + h_{3,1}$, 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 $\hat{\sigma}_1^2(c) = \sum_{i=1}^{n_1}(Y_i - \hat{Y}_i)^2/(n_1 - 4)$, where $\hat{Y}_i$ denotes the fitted values from the regression used to estimate the second derivative.[3] $\hat{\beta}_{2,0}$, $\hat{\beta}_{3,0}$ 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 (8) in the main text and to use numerical minimization over the compact region to obtain $\hat{h}_1$ and $\hat{h}_0$. Unlike $AMSE_{1n}(h)$ and $AMSE_{2n}(h)$ 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.

[3] Clearly, the estimator is not a consistent estimator of the conditional variance, but we do not need to estimate it consistently. One can use a non-parametric method to consistently estimate it, but it produces almost identical simulation results.

D Proof of Lemma 1

The LLR estimator can be expressed as $\left(\hat{\alpha}_{h_1}(c), \hat{\beta}_{h_1}(c)\right)' = \left(X(c)'W_1(c)X(c)\right)^{-1} X(c)'W_1(c)Y$, where $X(c)$ is an $n \times 2$ matrix whose ith row is given by $(1, X_i - c)$, $Y = (Y_1, \ldots, Y_n)'$, $W_1(c) = \text{diag}(K_{h_1}(X_i - c))$ and $K_{h_1}(\cdot) = K(\cdot/h_1)\mathbb{I}\{\cdot \geq 0\}/h_1$. The LLR estimator of $m_1(c)$ can also be written as $\hat{\alpha}_{h_1}(c) = e_1' \left(X(c)'W_1(c)X(c)\right)^{-1} X(c)'W_1(c)Y$, where e_1 is a 2×1 vector having one in the first entry and zero in the other entry. Similarly, the LLR estimator for $m_0(c)$, denoted by $\hat{\alpha}_{h_0}(c)$, can be obtained by replacing $W_1(c)$ with $W_0(c)$, where $W_0(c) = \text{diag}(K_{h_0}(X_i - c))$ and $K_{h_0}(\cdot) = K(\cdot/h_0)\mathbb{I}\{\cdot < 0\}/h_0$.

A contribution to the MSE from a variance component is standard. See **?** for the details. Here we consider the contribution made by the bias component. We present the proof only for $\hat{\alpha}_{h_1}(c)$. The proof for $\hat{\alpha}_{h_0}$ is parallel and hence is omitted. Denote $\hat{\gamma}_1 = \left(\hat{\alpha}_{h_1}(c),\ \hat{\beta}_{h_1}(c)\right)'$. The conditional bias is given by

$$\text{Bias}(\hat{\gamma}_1 | X) = \left(X(c)'W_1(c)X(c)\right)^{-1} X(c)W_1(c)(m_1 - X(c)\gamma_1),$$

where $m_1 = (m_1(X_1), \ldots, m_1(X_n))'$ and $\gamma_1 = (m_1(c), m_1^{(1)}(c))'$. Define, for $j = 0, 1$ and an integer k,

$$S_{n,k,j} = \begin{bmatrix} s_{n,k,j} & s_{n,k+1,j} \\ s_{n,k+1,j} & s_{n,k+2,j} \end{bmatrix}, \quad c_{n,k,j} = \begin{bmatrix} s_{n,k,j} \\ s_{n,k+1,j} \end{bmatrix}, \quad s_{n,k,j} = \sum_{i=1}^{n} K_{h_j}(X_i - c)(X_i - c)^k,$$
$$S_{k,1} = \begin{bmatrix} \mu_{k,0} & \mu_{k+1,0} \\ \mu_{k+1,0} & \mu_{k+2,0} \end{bmatrix}, \quad \text{and} \quad c_{k,1} = \begin{bmatrix} \mu_{k,0} \\ \mu_{k+1,0} \end{bmatrix}. \tag{D.1}$$

Note that $S_{n,0,1} = X(c)'W_1(c)X(c)$. The argument made by **?** can be generalized to yield

$$s_{n,k,1} = nh^k \left\{ f(c)\mu_{k,0} + hf^{(1)}(c)\mu_{k+1,0} + o_p(h) \right\}. \tag{D.2}$$

Then, it follows that

$$S_{n,0,1} = nH \left\{ f(c)S_{0,1} + hf^{(1)}(c)S_{1,1} + o_p(h) \right\} H,$$

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

$$S_{n,0,1}^{-1} = n^{-1}H^{-1}\left\{\frac{1}{f(c)}A_{0,1} - \frac{hf^{(1)}(c)}{f(c)^2}A_{1,1} + o_p(h)\right\}H^{-1}, \tag{D.3}$$

where

$$A_{0,1} = \begin{bmatrix} \mu_{2,0} & -\mu_{1,0} \\ -\mu_{1,0} & \mu_{0,0}^{-1} \end{bmatrix},$$

$$A_{1,1} = \frac{1}{\mu_{0,0}\mu_{2,0} - \mu_{1,0}^2}\begin{bmatrix} -\mu_{1,0}(\mu_{2,0}^2 - \mu_{1,0}\mu_{3,0}) & \mu_{2,0}(\mu_{2,0}^2 - \mu_{1,0}\mu_{3,0}) \\ \mu_{2,0}(\mu_{2,0}^2 - \mu_{1,0}\mu_{3,0}) & \mu_{1,0}^3 - 2\mu_{0,0}\mu_{1,0}\mu_{2,0} + \mu_{0,0}^2\mu_{3,0} \end{bmatrix}.$$

Next, we consider $X(c)W_1(c)\{m_1 - X(c)\gamma_1\}$. A Taylor expansion of $m_1(\cdot)$ yields

$$X(c)W_1(c)\{m_1 - X(c)\gamma_1\} = \frac{m_1^{(2)}(c)}{2}c_{n,2,1} + \frac{m_1^{(3)}(c)}{3!}c_{n,3,1} + o_p\left(nh^3\right). \tag{D.4}$$

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

$$c_{n,k,1} = nh^k H\left\{f(c)c_{k,1} + hf^{(1)}(c)c_{k+1,1} + o_p(h)\right\}. \tag{D.5}$$

Combining this with (D.3) and (D.4) and extracting the first element gives

$$\text{Bias}(\hat{\alpha}_{h_1}(c)|X) = \frac{b_1 m_1^{(2)}(c)}{2}h_1^2 + b_{2,1}(c)h_1^3 + o_p\left(h_1^3\right).$$

This expression gives the required result. ■