Supplementary Material for:

"New Control Function Approaches in Threshold Regression with Endogeneity"[13]

Ping Yu, Qin Liao,
University of Hong Kong, University of Hong Kong
Peter C. B. Phillips
Yale University, University of Auckland
University of Southampton & Singapore Management University

This Supplement contains five sections. SD.1 provides detailed calculations relating to the inconsistency of KST's STR estimator. SD.2 develops asymptotics for CF-I $\widehat{\gamma}$ and GMM-I2 $\widehat{\beta}$. SD.3 contains detailed analysis and discussion concerning CF and GMM-1 estimation of β. SD.4 provides some simplified asymptotic results for estimation of θ. SD.5 reports further simulation results relating to the performance of the IDKE estimator of γ and some comparisons of the two CF approaches when q is exogenous.

SD.1 Inconsistency of KST's STR Estimator

In this section, we first discuss the inconsistency of $\widehat{\gamma}$ when κ_0 and δ_0 are known in the simple example of Section 2.2, and then collect details of calculation omitted in the main text of Section 2.2 and point out the key problems in their proof using the framework of our simple example, and finally compare M-estimators and Z-estimators in the general sense to sharpen the discussion given in Section 2.3 of the main paper.

Inconsistency When κ_0 and δ_0 are Known

When δ_0 and κ_0 are known, Q_n and Q depend only on γ. Since $Q_n(\gamma)$ and $Q(\gamma)$ are apparently asymmetric about γ_0, we discuss $\gamma \in [\underline{\gamma}, \gamma_0]$ and $\gamma \in [\gamma_0, \overline{\gamma}]$ separately. Specifically, when $\gamma \in [\underline{\gamma}, \gamma_0]$,

$$\begin{aligned} Q(\gamma) = & -\mathbb{E}\left[\Phi_i^\gamma \kappa_0^2 \left(\lambda_{1i}^0 - \lambda_{1i}^\gamma\right)^2\right] + \mathbb{E}\left[\kappa_0^2\left(\lambda_{2i}^0 - \lambda_{2i}^\gamma\right)^2\left(1-\Phi_i^0\right)\right] \\ & + \mathbb{E}\left[\left(\delta_0 + \kappa_0\lambda_{1i}^0 - \kappa_0\lambda_{2i}^\gamma\right)^2\left(\Phi_i^0 - \Phi_i^\gamma\right)\right] + 2\mathbb{E}\left[\left(\delta_0 + \kappa_0\lambda_{1i}^0 - \kappa_0\lambda_{2i}^\gamma\right)\Phi_i^\gamma\kappa_0\left(\lambda_{1i}^0 - \lambda_{1i}^\gamma\right)\right], \end{aligned}$$

where $\Phi_i^\gamma = \Phi(\gamma + z_i)$, $\phi_i^\gamma = \phi(\gamma + z_i)$, $\lambda_{1i}^0 = \lambda_{1i}^{\gamma_0}$, $\lambda_{2i}^0 = \lambda_{2i}^{\gamma_0}$, $\Phi_i^0 = \Phi_i^{\gamma_0}$ and $\phi_i^0 = \phi_i^{\gamma_0}$. When $\gamma \in [\gamma_0, \overline{\gamma}]$,

$$\begin{aligned} Q(\gamma) = & \mathbb{E}\left[\kappa_0^2\left(\lambda_{1i}^0 - \lambda_{1i}^\gamma\right)^2\Phi_i^0\right] - \mathbb{E}\left[\kappa_0^2\left(\lambda_{2i}^0 - \lambda_{2i}^\gamma\right)^2\left(1-\Phi_i^\gamma\right)\right] \\ & + \mathbb{E}\left[\left(-\delta_0 - \kappa_0\lambda_{1i}^\gamma + \kappa_0\lambda_{2i}^0\right)^2\left(\Phi_i^\gamma - \Phi_i^0\right)\right] + 2\mathbb{E}\left[\left(-\delta_0 - \kappa_0\lambda_{1i}^\gamma + \kappa_0\lambda_{2i}^0\right)\left(1-\Phi_i^\gamma\right)\kappa_0\left(\lambda_{2i}^0 - \lambda_{2i}^\gamma\right)\right]. \end{aligned}$$

which is a quadratic function of (δ_0, κ_0). KST assume $\delta_0 = c_\delta n^{-\alpha}$ and $\kappa_0 = c_\kappa n^{-\alpha}$, so $Q_n(\gamma)$ is scaled by $n^{2\alpha}$ to avoid asymptotic degeneration. Actually, in the fixed-threshold-effect framework of Chan (1993), $Q(\gamma)$ takes exactly the same form as in KST's framework as long as we understand δ_0 and κ_0 as fixed numbers; in other words, our arguments also show that the KST estimator is inconsistent in Chan (1993)'s framework. To avoid further notation, we use κ_0 to represent $n^\alpha\kappa_0$ and and δ_0 for $n^\alpha\delta_0$.

[13] We thank the Co-Editor, Simon Lee, and three referees for helpful comments on earlier versions of this paper. Thanks go also to Chirok Han, Bruce Hansen, Shengjie Hong, Zhongxiao Jia, Hiroyuki Kasahara, Seojeong Lee, Hongyi Li, Jaimie Lien, Sangsoo Park, Wenwu Wang, Jason Wu, Zhiguo Xiao, Chuancun Yin and seminar participants at CUHK, ESWC2020, QNU, UWAC2018, Korea University and Tsinghua University for helpful suggestions. Phillips and Yu acknowledge support from the GRF of Hong Kong Government under Grant No. 17520716, and Phillips acknowledges research support under NSF Grant No. SES 18-50860 and the Kelly Fund at the University of Auckland.

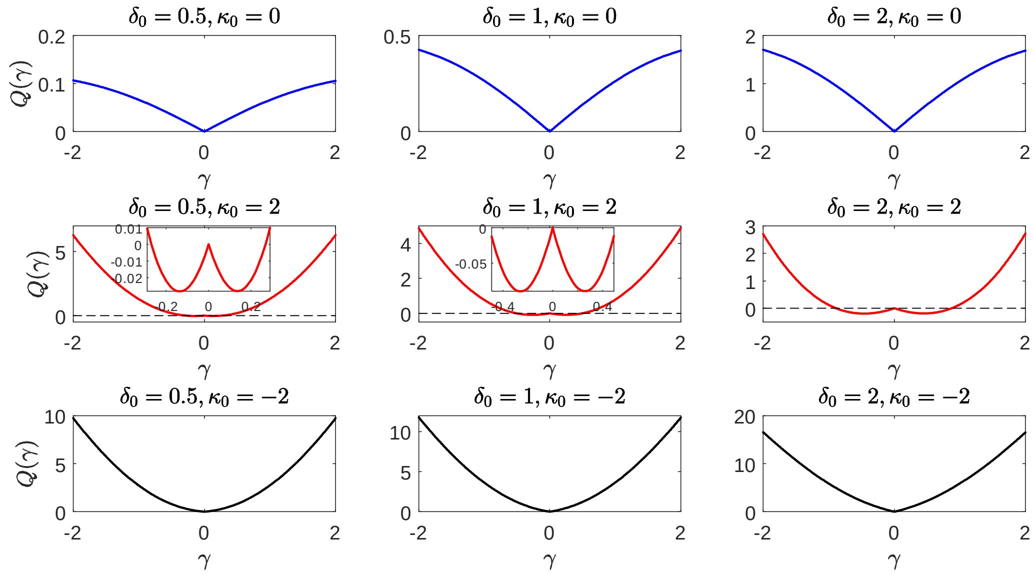

Figure 3: $Q(\gamma)$ When $\delta_0 = 0.5, 1, 2$ and $\kappa_0 = 0, \pm 2$

We plot $Q(\gamma)$ in Figure 3 for both the exogenous case and the endogenous case when $z \sim \mathcal{N}(0,1)$. In the exogenous case, $\kappa_0 = 0$, and in the endogenous case, $\kappa_0 = \pm 2$, with corresponding correlations $Corr(q,u) = 0$ and $Corr(q,u) \approx \pm 0.63$, respectively. In the figure, we set $\delta_0 = 0.5, 1, 2$, representing small, medium and large threshold effects, respectively. Figure 3 confirms that $\widehat{\gamma}$ is indeed consistent when q is exogenous, as shown in CH. On the contrary, $\widehat{\gamma}$ need not be consistent when q is endogenous. In fact, $Q(\gamma)$ is locally maximized rather than minimized at γ_0 when $\kappa_0 = 2$, and the argmin function $\arg\min_\gamma Q(\gamma)$ is not even unique in this case – as the three figures with $\kappa_0 = 2$, especially the inset magnified portions reveal in detail.[14] Interestingly, when $\kappa_0 = -2$, $\widehat{\gamma}$ is consistent. The intuition is as follows: when $\kappa_0 < 0$, the threshold effect based on (7) is $\delta_0 + \kappa_0 \left(\lambda_{1i}^0 - \lambda_{2i}^0\right)$, where $\lambda_{\ell i}^0 = \lambda_{\ell i}^{\gamma_0}$; and since $\lambda_{1i}^0 - \lambda_{2i}^0 = -\lambda(\gamma_0 + z_i) - \lambda(-z_i - \gamma_0) = -\lambda(z_i) - \lambda(-z_i) = -\frac{\phi(z_i)}{\Phi(z_i)(1-\Phi(z_i))} < 0$ when $\gamma_0 = 0$, the original threshold effect $\delta_0 (> 0)$ is strengthened when $\kappa_0 < 0$.

Whether the limiting outcome $\arg\min_\gamma Q(\gamma) = \gamma_0$ is realized depends on the relative magnitude of κ_0 and δ_0. If κ_0/δ_0 is close to zero, then the endogeneity is neglectable, and $Q(\gamma)$ should be minimized at γ_0. In Section 2.2, we discussed the relationship between $\arg\min_\gamma Q(\gamma)$ and κ_0/δ_0 when κ_0 and δ_0 are unknown and need to be estimated.

Details of Calculation in the Simple Example of Section 2.2

In the case where parameters κ_0 and δ_0 are unknown it is possible that the misspecification of $Q_n(\cdot)$ may be completely absorbed through the parameters (δ, κ), so that although (δ, κ) are not consistently estimated by using Q_n, the threshold parameter γ may be. In other words, although the minimized functional $\arg\min_\gamma Q(\gamma, \delta_0, \kappa_0) \neq \gamma_0$, it may be that in some special case $\arg\min_\gamma Q(\gamma, \delta_\gamma, \kappa_\gamma) = \gamma_0$. We proceed to explore this possibility in what follows and the main text.

[14] Nonuniqueness of $\arg\min_\gamma Q(\gamma)$ is the consequence of the special symmetric setup where $z \sim \mathcal{N}(0,1)$ and $\gamma_0 = 0$. If $\gamma_0 - \pi_0 z$ is not symmetric, then $Q(\gamma)$ would not be symmetric and $\arg\min_\gamma Q(\gamma)$ is unique.

When κ_0 and δ_0 are unknown, based on KST, we first find $(\delta_\gamma, \kappa_\gamma) = \arg\min_{(\delta,\kappa)} Q(\gamma, \delta, \kappa)$ by minimizing

$$\begin{aligned} &Q(\gamma,\delta,\kappa) - Q(\gamma,\delta_0,\kappa_0) \\ &= \mathbb{E}\left[\Phi_i^\gamma \left[(\delta_0-\delta) + (\kappa_0\lambda_{1i}^\gamma - \kappa\lambda_{1i}^\gamma)\right]^2\right] + \mathbb{E}\left[(1-\Phi_i^\gamma)\left(\kappa_0\lambda_{2i}^\gamma - \kappa\lambda_{2i}^\gamma\right)^2\right] + 2\mathbb{E}\left[\left(\Phi_i^0 - \Phi_i^\gamma\right)\delta_0\left(\kappa_0\lambda_{2i}^\gamma - \kappa\lambda_{2i}^\gamma\right)\right]. \end{aligned}$$

Solving the first order conditions (FOCs) with respect to (δ, κ), we have

$$\delta_\gamma - \delta_0 = -\frac{\mathbb{E}\left[\Phi_i^\gamma \lambda_{1i}^\gamma\right]}{\mathbb{E}\left[\Phi_i^\gamma\right]}(\kappa_\gamma - \kappa_0), \tag{32}$$

and

$$\kappa_\gamma - \kappa_0 = \frac{\delta_0 \mathbb{E}\left[\left(\Phi_i^0 - \Phi_i^\gamma\right)\lambda_{2i}^\gamma\right]}{\mathbb{E}\left[\phi_i^\gamma \lambda_{2i}^\gamma\right] - \mathbb{E}\left[\phi_i^\gamma \lambda_{1i}^\gamma\right] - \mathbb{E}\left[\phi_i^\gamma\right]^2 \Big/ \mathbb{E}\left[\Phi_i^\gamma\right]}. \tag{33}$$

We then plug $(\delta_\gamma, \kappa_\gamma)$ into $Q(\gamma, \delta, \kappa)$, giving

$$\begin{aligned} &Q(\gamma,\delta_\gamma,\kappa_\gamma) = S(\gamma,\delta_\gamma,\kappa_\gamma) - S(\gamma_0,\delta_0,\kappa_0) \\ &= \mathbb{E}\left[\Phi_i^\gamma\left(\delta_0 - \delta_\gamma + \kappa_0\lambda_{1i}^0 - \kappa_\gamma\lambda_{1i}^\gamma\right)^2\right] - 2\mathbb{E}\left[\left(\delta_0 - \delta_\gamma + \kappa_0\lambda_{1i}^0 - \kappa_\gamma\lambda_{1i}^\gamma\right)\kappa_0\left(\lambda_{1i}^0 - \lambda_{1i}^\gamma\right)\Phi_i^\gamma\right] \\ &\quad + \mathbb{E}\left[\left(\kappa_0\lambda_{2i}^0 - \kappa_\gamma\lambda_{2i}^\gamma\right)^2\left(1-\Phi_i^0\right)\right] + \mathbb{E}\left[\left(\delta_0 + \kappa_0\lambda_{1i}^0 - \kappa_\gamma\lambda_{2i}^\gamma\right)^2\left(\Phi_i^0 - \Phi_i^\gamma\right)\right] \\ &\quad + 2\mathbb{E}\left[\left(\delta_0 + \kappa_0\lambda_{1i}^0 - \kappa_\gamma\lambda_{2i}^\gamma\right)\Phi_i^\gamma\kappa_0\left(\lambda_{1i}^0 - \lambda_{1i}^\gamma\right)\right]. \end{aligned} \tag{34}$$

Notice from (32) and (33) that $(\delta_\gamma, \kappa_\gamma) = (\delta_0, \kappa_0)$ when $\gamma = \gamma_0$. So taking the derivative of (34) with respect to γ and evaluating at γ_0 gives

$$dQ(\gamma,\delta_\gamma,\kappa_\gamma)/d\gamma|_{\gamma=\gamma_0} = -\delta_0\mathbb{E}\left[\phi_i^0\Delta_i^0\right] + \kappa_0\mathbb{E}\left[\phi_i^0\Delta_i^0\left(\lambda_{1i}^0 + \lambda_{2i}^0 - 2(\gamma_0 + z_i)\right)\right], \tag{35}$$

where $\Delta_i^0 = \delta_0 + \kappa_0\left(\lambda_{1i}^0 - \lambda_{2i}^0\right)$ is the threshold effect at γ_0 implied by the KST objective function. By definition, $Q(\gamma_0, \delta_0, \kappa_0) = 0$. Hence, inspired by the $Q(\gamma)$ in Figure 3, if $dQ(\gamma,\delta_\gamma,\kappa_\gamma)/d\gamma|_{\gamma=\gamma_0} > 0$, then we can find a $\gamma^* \in \left[\underline{\gamma}, \gamma_0\right]$ such that $Q(\gamma^*, \delta_{\gamma^*}, \kappa_{\gamma^*}) < Q\left(\gamma_0, \delta_{\gamma_0}, \kappa_{\gamma_0}\right)$. From (35), given $z \sim \mathcal{N}(0,1)$ and $\gamma_0 = 0$, $dQ(\gamma,\delta_\gamma,\kappa_\gamma)/d\gamma|_{\gamma=\gamma_0}$ depends only on δ_0 and κ_0. Define $f(x) = \phi(x)\Delta(x)(\lambda_1(x) + \lambda_2(x) - 2x)$ with $\Delta(x) = \delta_0 + \kappa_0(\lambda_1(x) - \lambda_2(x))$; then, it follows that $\Delta(x) = \Delta(-x)$ and it is not hard to see[15] that $f(x) = -f(-x)$. Since $\gamma_0 + z_i \sim \mathcal{N}(0,1)$ is symmetric around 0, the term involving $\mathbb{E}\left[\phi_i^0\Delta_i^0(\gamma_0 + z_i)\right]$ in (35) disappears. In consequence, as long as $\kappa_0 \in \left(\delta_0\frac{\mathbb{E}[\phi^0]}{\mathbb{E}\left[\phi^0\left(\lambda_2^0 - \lambda_1^0\right)\right]}, \infty\right)$ if $\delta_0 > 0$, or $\kappa_0 \in \left(-\infty, \delta_0\frac{\mathbb{E}[\phi^0]}{\mathbb{E}\left[\phi^0\left(\lambda_2^0 - \lambda_1^0\right)\right]}\right)$ if $\delta_0 < 0$, $dQ(\gamma,\delta_\gamma,\kappa_\gamma)/d\gamma|_{\gamma=\gamma_0} > 0$. Notice that if $z \sim \mathcal{N}(0,1)$ and $\gamma_0 = 0$, $\frac{\mathbb{E}[\phi^0]}{\mathbb{E}\left[\phi^0\left(\lambda_2^0 - \lambda_1^0\right)\right]} \approx 0.587$. So we have $\gamma^* < 0$ if $\kappa_0/\delta_0 > 0.587$ regardless of whether δ_0 is positive or negative.[16]

Problems in the Consistency Proof of KST

KST mimic the proof idea of CH to prove the consistency of $\widehat{\gamma}$. However, there is a fundamental difference between the KST and CH frameworks. Unlike CH, the regressors in (8) depend also on γ through inverse

[15] Note that

$$\lambda_1(x) - \lambda_2(x) = -\lambda(x) - \lambda(-x) = \lambda_1(-x) - \lambda_2(-x)$$

so that $\Delta(x) = \Delta(-x)$, and

$$\lambda_1(x) + \lambda_2(x) = -\lambda(x) + \lambda(-x) = -\{\lambda_2(-x) + \lambda_1(-x).$$

[16] The break point, such as 0.587 in the above calculation, critically depends on the data generating process. For example, if β_{20} is unknown, the break point is smaller.

Mills ratios. In KST's proof of their Proposition 1, the objective function is

$$S_n(\gamma) = Y'(I - P^*(\gamma))Y,$$

where Y is the vector stacking y_i, I is the $n \times n$ identity matrix,

$$P^*(\gamma) = X_\gamma(\gamma)\left(X_\gamma(\gamma)' X_\gamma(\gamma)\right)^{-1} X_\gamma(\gamma)' + X_\perp(\gamma)\left(X_\perp(\gamma)' X_\perp(\gamma)\right)^{-1} X_\perp(\gamma)'$$

is the projection matrix on $X(\gamma) := (X_\gamma(\gamma), X_\perp(\gamma))$, $X_\gamma(\gamma)$ is the matrix stacking $(\mathbf{x}_i', \lambda_{1i}^\gamma, \lambda_{2i}^\gamma)\,1(q_i \le \gamma)$, and $X_\perp(\gamma)$ is the matrix stacking $(\mathbf{x}_i', \lambda_{1i}^\gamma, \lambda_{2i}^\gamma)\,1(q_i > \gamma)$, so we are assuming $\mathbf{x}$ is exogenous and π is known to simplify the discussion. KST write Y as

$$Y = G(\gamma_0)\overline{\beta} + G_0(\gamma_0)\overline{\delta} + \widetilde{\mathbf{e}},$$

where $G(\gamma_0)$ is the matrix stacking $\left(\mathbf{x}_i', \lambda_{1i}^0, \lambda_{2i}^0\right)$, denoted as $\left(X, \boldsymbol{\lambda}_1^0, \boldsymbol{\lambda}_2^0\right)$, $G_0(\gamma_0)$ is the matrix stacking $\left(\mathbf{x}_i', \lambda_{1i}^0, \lambda_{2i}^0\right)1(q_i \le \gamma_0)$, $\widetilde{\mathbf{e}}$ is the vector stacking $\widetilde{e}_i = u_i - \kappa_0\lambda_{1i}^0 1(q_i \le \gamma_0) - \kappa_0\lambda_{2i}^0 1(q_i > \gamma_0)$ which is e_i^* in (9), $\overline{\beta} = \left(\beta_{20}', 0, \kappa_0\right)'$ and $\overline{\delta} = \left(\delta_0', \kappa_0, -\kappa_0\right)'$. They claim

$$S_n(\gamma) = \left(G_0(\gamma_0)\overline{\delta} + \widetilde{\mathbf{e}}\right)'(I - P^*(\gamma))\left(G_0(\gamma_0)\overline{\delta} + \widetilde{\mathbf{e}}\right),$$

in other words, it is implicitly assumed that $G(\gamma_0)$ falls in the space spanned by $X(\gamma)$. But this assumption is not satisfied. Although X falls in the span of $X(\gamma)$, $\boldsymbol{\lambda}_1^0$ and $\boldsymbol{\lambda}_2^0$ do not. This is because $\lambda_{\ell i}^\gamma$ is a nonlinear function of γ and so the spaces spanned by $\boldsymbol{\lambda}_\ell^0$ and $\boldsymbol{\lambda}_\ell^\gamma$ are different.

Even if the expression for S_n were correct, its probability limit given in KST is not correct. Specifically, since $\left(G_0(\gamma_0)\overline{\delta} + \widetilde{\mathbf{e}}\right)' I \left(G_0(\gamma_0)\overline{\delta} + \widetilde{\mathbf{e}}\right)$ does not depend on γ, minimizing S_n is equivalent to maximizing $\left(G_0(\gamma_0)\overline{\delta} + \widetilde{\mathbf{e}}\right)' P^*(\gamma)\left(G_0(\gamma_0)\overline{\delta} + \widetilde{\mathbf{e}}\right)$. Normalizing this new objective function by $n^{-1+2\alpha}$ gives their function

$$\begin{aligned} &S_n^*(\gamma) \\ &= n^{-1+2\alpha}\widetilde{\mathbf{e}}' P^*(\gamma)\widetilde{\mathbf{e}} + 2n^{-1+\alpha}c'G_0'(\gamma_0)P^*(\gamma)\widetilde{\mathbf{e}} + n^{-1}c'G_0'(\gamma_0)P^*(\gamma)G_0(\gamma_0)c \\ &=: T_{1n}(\gamma) + T_{2n}(\gamma) + T_{3n}(\gamma), \end{aligned} \tag{36}$$

where $c = n^\alpha\overline{\delta} := (c_\delta', c_\kappa, -c_\kappa)'$. KST claim in Lemma I.A.3 that $T_{1n}(\gamma) \xrightarrow{p} 0$ and $T_{2n}(\gamma) \xrightarrow{p} 0$ uniformly in γ, so only T_{3n} counts. But this is incorrect: both T_{1n} and T_{2n} contribute to the probability limit of $S_n^*(\gamma)$ unless $c_\kappa = 0$. Essentially, they neglect the fact that $\widetilde{e}_i$ is a function of γ_0.

If we redefine $G(\gamma_0) = X$, $G_0(\gamma_0) = \left(\mathbf{x}_i'1(q_i \le \gamma_0), \lambda_{1i}^0 1(q_i \le \gamma_0) + \lambda_{2i}^0 1(q_i > \gamma_0)\right)$, $\overline{\beta} = \beta_{20}$, $\overline{\delta} = \left(\delta_0', \kappa_0\right)'$, $P^*(\gamma) = X(\gamma)\left(X(\gamma)'X(\gamma)\right)^{-1}X(\gamma)'$ with $X(\gamma)$ redefined as the matrix stacking $(\mathbf{x}_i'1(q_i \le \gamma), \mathbf{x}_i'1(q_i > \gamma)$, $\lambda_{1i}^\gamma 1(q_i \le \gamma) + \lambda_{2i}^\gamma 1(q_i > \gamma))$, then the objective function $S_n^*(\gamma)$ is now valid. So the key difference between KST's $S_n^*(\gamma)$ and our correctly defined $S_n^*(\gamma)$ is that the element of their $G_0(\gamma_0)c$ is

$$\mathbf{x}_i'c_\delta 1(q_i \le \gamma_0) + c_\kappa\left(\lambda_{1i}^0 - \lambda_{2i}^0\right)1(q_i \le \gamma_0),$$

whereas ours is

$$\mathbf{x}_i'c_\delta 1(q_i \le \gamma_0) + c_\kappa\left[\lambda_{1i}^0 1(q_i \le \gamma_0) + \lambda_{2i}^0 1(q_i > \gamma_0)\right].$$

In other words, our element of $G_0(\gamma_0)c$ minus theirs is $c_\kappa\lambda_{2i}^0$. Also, our $P^*(\gamma)$ projects onto a smaller space than theirs.

To calculate the probability limit of our $S_n^*(\gamma)$, denoted as $S^*(\gamma)$, we return to our simple example where $\mathbf{x} = 1$ and β_{20} is known, so $X(\gamma)$ stacks $(1(q_i \le \gamma), \lambda_{1i}^\gamma 1(q_i \le \gamma) + \lambda_{2i}^\gamma 1(q_i > \gamma))$ and the element of $G_0(\gamma_0)c$

is $c_\delta 1(q_i \le \gamma_0) + c_\kappa \left[\lambda_{1i}^0 1(q_i \le \gamma_0) + \lambda_{2i}^0 1(q_i > \gamma_0)\right]$. As in KST, we restrict $\gamma \in [\gamma_0, \overline{\gamma}]$.

Lemma 5

$$\begin{aligned} T_{1n}(\gamma) &\xrightarrow{p} T_1(\gamma) = c_\kappa^2 A(\gamma) M(\gamma)^{-1} A(\gamma), \\ T_{2n}(\gamma) &\xrightarrow{p} T_2(\gamma) = 2c_\kappa A(\gamma)' M(\gamma)^{-1} B(\gamma), \\ T_{3n}(\gamma) &\xrightarrow{p} T_3(\gamma) = B(\gamma)' M(\gamma)^{-1} B(\gamma), \end{aligned} \tag{37}$$

where

$$A(\gamma) = \begin{pmatrix} \mathbb{E}\left[\phi^0 - \phi^\gamma - \lambda_2^0\left(\Phi^\gamma - \Phi^0\right)\right] \\ \mathbb{E}\left[\phi^0 \lambda_1^\gamma - \phi^\gamma \lambda_1^\gamma + \lambda_1^\gamma \lambda_2^0\left(\Phi^\gamma - \Phi^0\right)\right] + \mathbb{E}\left[\lambda_2^\gamma\left(\lambda_2^\gamma - \lambda_2^0\right)(1-\Phi^\gamma)\right] \end{pmatrix},$$

$$B(\gamma) = \begin{pmatrix} \mathbb{E}\left[c_\delta \Phi^0 + c_\kappa \Phi^0 \lambda_1^0 + c_\kappa \lambda_2^0\left(\Phi^\gamma - \Phi^0\right)\right] \\ \mathbb{E}\left[c_\delta \Phi^0 \lambda_1^\gamma + c_\kappa \Phi^0 \lambda_1^0 \lambda_1^\gamma + c_\kappa \lambda_2^0 \lambda_1^\gamma\left(\Phi^\gamma - \Phi^0\right)\right] + \mathbb{E}\left[c_\kappa \lambda_2^0 \lambda_2^\gamma (1-\Phi^\gamma)\right] \end{pmatrix},$$

$$M(\gamma) = \begin{pmatrix} \mathbb{E}\left[1(q \le \gamma)\right] & \mathbb{E}\left[\lambda_1^\gamma 1(q \le \gamma)\right] \\ \mathbb{E}\left[\lambda_1^\gamma 1(q \le \gamma)\right] & \mathbb{E}\left[\lambda_1^{\gamma 2} 1(q \le \gamma) + \lambda_2^{\gamma 2} 1(q > \gamma)\right] \end{pmatrix}.$$

Proof. In this simple example, $X(\gamma)$ stacks $\left(1(q_i \le \gamma), \lambda_1^\gamma 1(q_i \le \gamma) + \lambda_2^\gamma 1(q_i > \gamma)\right)$, so

$$\frac{1}{n} X(\gamma)' X(\gamma) \xrightarrow{p} \begin{pmatrix} \mathbb{E}\left[1(q \le \gamma)\right] & \mathbb{E}\left[\lambda_1^\gamma 1(q \le \gamma)\right] \\ \mathbb{E}\left[\lambda_1^\gamma 1(q \le \gamma)\right] & \mathbb{E}\left[\lambda_1^{\gamma 2} 1(q \le \gamma) + \lambda_2^{\gamma 2} 1(q > \gamma)\right] \end{pmatrix} = M(\gamma).$$

Since $G_0(\gamma_0) = X(\gamma_0)$,

$$\begin{aligned} n^{-1} c' G_0'(\gamma_0) X(\gamma) &\xrightarrow{p} (c_\delta, c_\kappa) \mathbb{E}\left[\begin{pmatrix} 1(q_i \le \gamma_0) \\ \lambda_1^0 1(q_i \le \gamma_0) + \lambda_2^0 1(q_i > \gamma_0) \end{pmatrix} \left(1(q_i \le \gamma), \lambda_1^\gamma 1(q_i \le \gamma) + \lambda_2^\gamma 1(q_i > \gamma)\right)\right] \\ &= \left(\mathbb{E}\left[c_\delta \Phi^0 + c_\kappa \Phi^0 \lambda_1^0 + c_\kappa \lambda_2^0\left(\Phi^\gamma - \Phi^0\right)\right], \mathbb{E}\left[c_\delta \Phi^0 \lambda_1^\gamma + c_\kappa \Phi^0 \lambda_1^0 \lambda_1^\gamma + c_\kappa \lambda_2^0 \lambda_1^\gamma\left(\Phi^\gamma - \Phi^0\right)\right] + \mathbb{E}\left[c_\kappa \lambda_2^0 \lambda_2^\gamma (1-\Phi^\gamma)\right]\right) = B(\gamma)'. \end{aligned}$$

Next,

$$\begin{aligned} n^{-1+\alpha} X(\gamma)' \widetilde{\mathbf{e}} &= \begin{pmatrix} n^{-1+\alpha} \sum_{i=1}^n 1(q_i \le \gamma) \widetilde{e}_i \\ n^{-1+\alpha} \sum_{i=1}^n \left[\lambda_1^\gamma 1(q_i \le \gamma) + \lambda_2^\gamma 1(q_i > \gamma)\right] \widetilde{e}_i \end{pmatrix} \\ &\xrightarrow{p} c_\kappa \begin{pmatrix} \mathbb{E}\left[\phi^0 - \phi^\gamma - \lambda_2^0\left(\Phi^\gamma - \Phi^0\right)\right] \\ \mathbb{E}\left[\phi^0 \lambda_1^\gamma - \phi^\gamma \lambda_1^\gamma - \lambda_1^\gamma \lambda_2^0\left(\Phi^\gamma - \Phi^0\right)\right] + \mathbb{E}\left[\lambda_2^\gamma\left(\lambda_2^\gamma - \lambda_2^0\right)(1-\Phi^\gamma)\right] \end{pmatrix} = c_\kappa A(\gamma). \end{aligned}$$

In summary, by the CMT we have

$$\begin{aligned} T_{1n}(\gamma) &= \left(n^{-1+\alpha} \widetilde{\mathbf{e}}' X(\gamma)\right)\left(n^{-1} X(\gamma)' X(\gamma)\right)^{-1}\left(n^{-1+\alpha} X(\gamma)' \widetilde{\mathbf{e}}\right) \xrightarrow{p} c_\kappa^2 A(\gamma) M(\gamma)^{-1} A(\gamma), \\ T_{2n}(\gamma) &= 2\left(n^{-1} c' G_0'(\gamma_0) X(\gamma)\right)\left(n^{-1} X(\gamma)' X(\gamma)\right)^{-1}\left(n^{-1+\alpha} X(\gamma)' \widetilde{\mathbf{e}}\right) \xrightarrow{p} 2c_\kappa B(\gamma)' M(\gamma)^{-1} A(\gamma), \\ T_{3n}(\gamma) &= \left(n^{-1} c' G_0'(\gamma_0) X(\gamma)\right)\left(n^{-1} X(\gamma)' X(\gamma)\right)^{-1}\left(n^{-1} X(\gamma)' G_0(\gamma_0) c\right) \xrightarrow{p} B(\gamma)' M(\gamma)^{-1} B(\gamma). \end{aligned}$$

■

Figure 4 shows $S^*(\gamma)$ and its three components when $c_\delta = 1$ and $c_\kappa = 0, 0.5, 1, 2$. Since KST consider only T_3 and $\arg\max_\gamma T_3(\gamma) = \gamma_0 = 0$, they claim $\widehat{\gamma}$ is consistent.[17] This claim is correct *only* if $c_\kappa = 0$ (i.e., q is exogenous) where $T_1(\gamma) = T_2(\gamma) = 0$, so $S^*(\gamma) = T_3(\gamma)$. Since $\arg\max_\gamma T_3(\gamma) = \gamma_0$ as shown in CH, $\arg\max_\gamma S^*(\gamma) = \gamma_0$ in that case. When $c_\kappa = 0.5$, neither T_1 nor T_2 is zero, but since T_3 dominates, $\arg\max_\gamma S^*(\gamma)$ is still γ_0. But when $c_\kappa = 1$ or 2, i.e., there is strong endogeneity, T_1 and T_2 are not

[17] Although their T_3 is different from our T_3, it is still correct that $\arg\max_\gamma T_3(\gamma) = \gamma_0$ for their T_3.

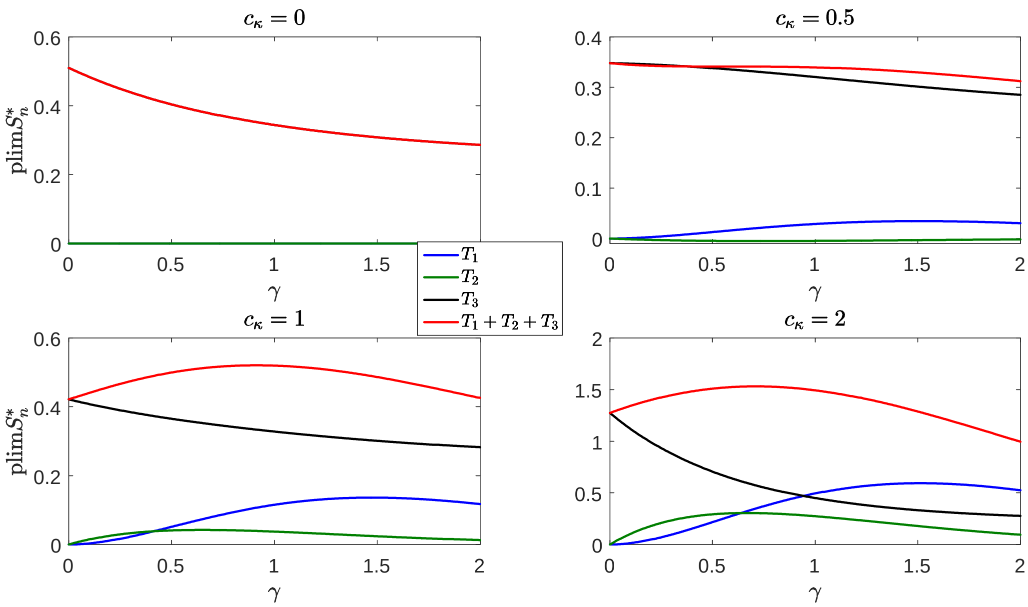

Figure 4: Components of plim$S_n^*(\gamma)$ When $c_\delta = 1$ and $c_\kappa = 0, 0.5, 1, 2$

neglectable. As a result, $\arg\max_\gamma S^*(\gamma) \neq \gamma_0$. These results are consistent with the information in the middle graph of the first row in Figure 1.

Given the negative result above, a natural question is whether we can modify KST's objective function (pursuing their approach) to obtain a consistent estimator of γ. The answer is – No. First, we strengthen the key assumption in KST.

Assumption K$'$: $\mathbf{x} = x$, and $v_x \perp 1(q \leq \gamma)|\mathbf{z}$ for any $\gamma \in \Gamma$.

Assumption K only requires $v_x \perp 1(q \leq \gamma_0)|\mathbf{z}$, so Assumption K$'$ strengthens Assumption K. Since $q = \pi'\mathbf{z} + v_q$, if γ can take any value in $\mathbb{R}$ (rather than only in Γ), then $v_x \perp 1(q \leq \gamma)|\mathbf{z}$ is the same as $v_x \perp 1(v_q \leq a)|\mathbf{z}$ for any $a \in \mathbb{R}$, which is equivalent to $v_x \perp v_q|\mathbf{z}$. For any γ,

$$
\begin{aligned}
\mathbb{E}[y|\mathbf{z}, q \leq \gamma] &= \mathbb{E}\left[\beta_1'(\Pi_x'\mathbf{z} + v_x)1(q \leq \gamma_0) + \beta_2'(\Pi_x'\mathbf{z} + v_x)1(q > \gamma_0) + u|\mathbf{z}, q \leq \gamma\right] \\
&= \beta_1'(\Pi_x'\mathbf{z})\,\mathbb{E}[1(q \leq \gamma_0)|\mathbf{z}, q \leq \gamma] + \mathbb{E}\left[\beta_1' v_x 1(q \leq \gamma_0)|\mathbf{z}, q \leq \gamma\right] \\
&\quad + \beta_2'(\Pi_x'\mathbf{z})\,\mathbb{E}[1(q > \gamma_0)|\mathbf{z}, q \leq \gamma] + \mathbb{E}\left[\beta_2' v_x 1(q > \gamma_0)|\mathbf{z}, q \leq \gamma\right] + \mathbb{E}[u|\mathbf{z}, q \leq \gamma]
\end{aligned}
$$

$$
\begin{aligned}
&= \kappa \cdot \lambda_1(\gamma - \pi'\mathbf{z}) + \begin{cases} \beta_1'(\Pi_x'\mathbf{z}) + \mathbb{E}\left[\beta_1' v_x|\mathbf{z}, q \leq \gamma\right] 1(q \leq \gamma), \text{ if } \gamma \leq \gamma_0, \\ \beta_2'(\Pi_x'\mathbf{z}) \frac{\Phi(\gamma - \pi'\mathbf{z}) - \Phi(\gamma_0 - \pi'\mathbf{z})}{\Phi(\gamma - \pi'\mathbf{z})} + \mathbb{E}\left[\beta_2' v_x|\mathbf{z}, \gamma_0 < q \leq \gamma\right] \frac{\Phi(\gamma - \pi'\mathbf{z}) - \Phi(\gamma_0 - \pi'\mathbf{z})}{\Phi(\gamma - \pi'\mathbf{z})} \\ + \beta_1'(\Pi_x'\mathbf{z}) \frac{\Phi(\gamma_0 - \pi'\mathbf{z})}{\Phi(\gamma - \pi'\mathbf{z})} + \mathbb{E}\left[\beta_1' v_x|\mathbf{z}, q \leq \gamma_0\right] \frac{\Phi(\gamma_0 - \pi'\mathbf{z})}{\Phi(\gamma - \pi'\mathbf{z})}, \text{ if } \gamma > \gamma_0, \end{cases} \\
&= \kappa \cdot \lambda_1(\gamma - \pi'\mathbf{z}) + \begin{cases} \beta_1'(\Pi_x'\mathbf{z}), & \text{if } \gamma \leq \gamma_0, \\ \beta_1'(\Pi_x'\mathbf{z}) \frac{\Phi(\gamma_0 - \pi'\mathbf{z})}{\Phi(\gamma - \pi'\mathbf{z})} + \beta_2'(\Pi_x'\mathbf{z}) \frac{\Phi(\gamma - \pi'\mathbf{z}) - \Phi(\gamma_0 - \pi'\mathbf{z})}{\Phi(\gamma - \pi'\mathbf{z})}, & \text{if } \gamma > \gamma_0, \end{cases} \\
&\neq \beta_1'(\Pi_x'\mathbf{z}) + \kappa \cdot \lambda_1(\gamma - \pi'\mathbf{z}),
\end{aligned}
$$

as claimed in KST unless $\gamma \leq \gamma_0$, where the conditional mean when $\gamma > \gamma_0$ is a weighted average of

$\beta_1'\left(\Pi_x'\mathbf{z}\right)+\kappa\cdot\lambda_1\left(\gamma-\pi'\mathbf{z}\right)$ and $\beta_2'\left(\Pi_x'\mathbf{z}\right)+\kappa\cdot\lambda_1\left(\gamma-\pi'\mathbf{z}\right)$. Similarly, for any γ,

$$
\begin{aligned}
\mathbb{E}\left[y|\mathbf{z},q>\gamma\right] &= \kappa\cdot\lambda_2\left(\gamma-\pi'\mathbf{z}\right)+\begin{cases}\beta_1'\left(\Pi_x'\mathbf{z}\right)\frac{\Phi\left(\gamma_0-\pi'\mathbf{z}\right)-\Phi\left(\gamma-\pi'\mathbf{z}\right)}{1-\Phi\left(\gamma-\pi'\mathbf{z}\right)}+\beta_2'\left(\Pi_x'\mathbf{z}\right)\frac{1-\Phi\left(\gamma_0-\pi'\mathbf{z}\right)}{1-\Phi\left(\gamma-\pi'\mathbf{z}\right)}, & \text{if } \gamma<\gamma_0,\\ \beta_2'\left(\Pi_x'\mathbf{z}\right), & \text{if } \gamma\geq\gamma_0,\end{cases}\\
&\neq \beta_2'\left(\Pi_x'\mathbf{z}\right)+\kappa\cdot\lambda_2\left(\gamma-\pi'\mathbf{z}\right)
\end{aligned}
$$

as claimed in KST unless $\gamma\geq\gamma_0$. Based on such an analysis, it seems that the correct objective function should be

$$
S_n\left(\theta,\kappa\right)=\begin{cases}\sum_{i=1}^n\left[y_i-\beta_1'\left(\widehat{\Pi}_x'\mathbf{z}_i\right)\frac{\Phi\left(\gamma_0-\widehat{\pi}'\mathbf{z}_i\right)-\Phi\left(\gamma-\widehat{\pi}'\mathbf{z}_i\right)}{1-\Phi\left(\gamma-\widehat{\pi}'\mathbf{z}_i\right)}-\beta_2'\left(\widehat{\Pi}_x'\mathbf{z}_i\right)\frac{1-\Phi\left(\gamma_0-\widehat{\pi}'\mathbf{z}_i\right)}{1-\Phi\left(\gamma-\widehat{\pi}'\mathbf{z}_i\right)}-\kappa\cdot\lambda_2\left(\gamma-\mathbf{z}_i'\widehat{\pi}\right)\right]^2 1(q_i>\gamma)\\ +\sum_{i=1}^n\left[y_i-\beta_1'\left(\widehat{\Pi}_x'\mathbf{z}_i\right)-\kappa\cdot\lambda_1\left(\gamma-\mathbf{z}_i'\widehat{\pi}\right)\right]^2 1(q_i\leq\gamma),\text{ if }\gamma\leq\gamma_0,\\ \sum_{i=1}^n\left[y_i-\beta_1'\left(\widehat{\Pi}_x'\mathbf{z}_i\right)\frac{\Phi\left(\gamma_0-\widehat{\pi}'\mathbf{z}_i\right)}{\Phi\left(\gamma-\widehat{\pi}'\mathbf{z}_i\right)}-\beta_2'\left(\widehat{\Pi}_x'\mathbf{z}_i\right)\frac{\Phi\left(\gamma-\widehat{\pi}'\mathbf{z}_i\right)-\Phi\left(\gamma_0-\widehat{\pi}'\mathbf{z}_i\right)}{\Phi\left(\gamma-\widehat{\pi}'\mathbf{z}_i\right)}-\kappa\cdot\lambda_1\left(\gamma-\mathbf{z}_i'\widehat{\pi}\right)\right]^2 1(q_i\leq\gamma)\\ +\sum_{i=1}^n\left[y_i-\beta_2'\left(\widehat{\Pi}_x'\mathbf{z}_i\right)-\kappa\cdot\lambda_2\left(\gamma-\mathbf{z}_i'\widehat{\pi}\right)\right]^2 1(q_i>\gamma),\text{ if }\gamma>\gamma_0,\end{cases}
$$

where $\widehat{\Pi}_x$ and $\widehat{\pi}$ are from the first-stage regression. However, the objective function depends on the true value γ_0, so this approach is infeasible.

Comparison of M-Estimators and Z-Estimators

Different from estimation based on a sum of squares, in estimation based on GMM (or a square of sums), we can introduce unknown parameters to the implied moments from the conditional moments. From YLP, we know the following identification results for γ using moment conditions. First, if q is exogenous (in the sense of $\mathbb{E}\left[u|\mathbf{z},q\right]=0$) and independent of $\left(\mathbf{z}',\mathbf{x}'\right)'$, then the identifying set based on $\mathbb{E}\left[\mathbf{z}u1(q\leq\gamma)\right]=\mathbf{0}$ and $\mathbb{E}\left[\mathbf{z}u1(q>\gamma)\right]=\mathbf{0}$ is Γ, i.e., the moment conditions do not have identification power at all. Second, if q is exogenous and not independent of $\left(\mathbf{z}',\mathbf{x}'\right)'$, then the identifying set using $\mathbb{E}\left[\mathbf{z}u1(q\leq\gamma)\right]=\mathbf{0}$ is $\left[\underline{\gamma},\gamma_0\right]$, using $\mathbb{E}\left[\mathbf{z}u1(q>\gamma)\right]=\mathbf{0}$ is $\left[\gamma_0,\overline{\gamma}\right]$, and using both is γ_0. What is the lesson here? Even if the unknown parameter γ is introduced to the unconditional moment conditions which are implied by the conditional moment conditions, the identifying set contains γ_0, i.e., the unconditional moment conditions will not contradict the truth. These identification results contrast dramatically with those when using sum of squares to identify γ, e.g., the formulae of $\mathbb{E}[y|\mathbf{z},q\leq\gamma]$ and $\mathbb{E}[y|\mathbf{z},q>\gamma]$ are implied by $\mathbb{E}[y|\mathbf{z},q]$, but the resulting estimator is inconsistent (contradicting the truth).

To further understand why the KST estimator is not consistent whereas our CF estimators are, we go back to the simple example above. It is not hard to show that the identifying set by the implied moment conditions from (14),

$$
\mathbb{E}\left[\begin{array}{l}\left(y-\frac{1-2e^{-1}}{1-e^{-1}}\theta\right)1\left(y\leq\theta\right)\\ \left(y-2\theta\right)1\left(y>\theta\right)\end{array}\right]=\mathbf{0},\tag{38}
$$

is $(0,\infty)$ which at least contains the true value θ_0, and the moment conditions originated from (15)

$$
\mathbb{E}\left[\begin{array}{l}\left(y-\theta+\frac{e^{-1/\theta}}{1-e^{-1/\theta}}\right)1\left(y\leq1\right)\\ \left(y-1-\theta\right)1\left(y>1\right)\end{array}\right]=\mathbf{0}.\tag{39}
$$

can point identify θ_0. Why is it that the identification result based on (14) is not consistent with that based on the resulting moments (38), whereas using (15) and the resulting moment (39) generates the same identification result? The key point is still that the conditioning set in (14) depends on the unknown parameter. Due to this fact, the first order condition (FOC) in minimizing $S_1(\theta)$ is not $\mathbb{E}\left[\left(y-\frac{1-2e^{-1}}{1-e^{-1}}\theta\right)1(y\leq\theta)\right]+\mathbb{E}\left[(y-2\theta)1(y>\theta)\right]=0$ because we need to take a derivative with respect to the θ in $1(y\leq\theta)$ and $1(y>\theta)$; if this were the case, $\min_\theta S_1(\theta)$ should be $(0,\infty)$ rather than 1.24. On the other hand, the FOCs in minimizing $S_{21}(\theta)$ and $S_{22}(\theta)$ are equivalent to (39); this is why using (15) and (39) generates the same identification result. Although γ is a nonregular parameter so that the limit objective function of KST is not differentiable at γ_0 (see the limit objective functions in Figure 1), the calculation in this simple example with a regular parameter θ provides some intuition regarding why identification results based on sum of squares and square of sums may not be the same when the conditioning set depends on the unknown parameter and would be the same otherwise. Since the identifying set using square of sums or GMM will not contradict the truth, this explains why KST fails but our two CF approaches work – their identification result may differ from that based on GMM whereas ours will be consistent with GMM.

It seems that estimation based on square of sums is more robust because unknown parameters can be introduced to the implied moment conditions. But robustness does not come without a cost. Given identification, sum of squares may have higher efficiency than square of sums. For example, the CH estimator when q is exogenous and our CF-I and CF-II estimators when q is endogenous are all n consistent. However, as is well known, the estimator based on GMM is at most $\sqrt{n}$ consistent. In summary, using square of sums (i.e., the Z-estimator) and sum of squares (i.e., the M-estimator) to identify γ illustrates the classical wisdom of trade-off between efficiency and robustness. Using square of sums is more robust but less efficient (the moment conditions may not provide enough information to identify the truth and under identification, can only provide a $\sqrt{n}$-consistent estimator), whereas using sum of squares is less robust but more efficient (under identification, it can, although not always, provide an n-consistent estimator).

SD.2 Asymptotics for CF-I $\widehat{\gamma}$ and GMM-I2 $\widehat{\beta}$

In this Section, we will discuss asymptotics for CF-I $\widehat{\gamma}$ and GMM-I2 $\widehat{\beta}$. Because these two estimators are parallel to our recommended estimators CF-II $\widehat{\gamma}$ and GMM-II2 $\widehat{\beta}$, we collect them together in this section.

Consistency of CF-I $\widehat{\gamma}$ When Endogeneity Takes Nonlinear Forms

First, the following assumptions are imposed.

Assumption C-I:

1. $\{w_i\}_{i=1}^n$ are strictly stationary and ergodic; $\theta\in\Theta$ with Θ being compact; $\left(\beta_{10}',\sigma_{10}\right)'\neq\left(\beta_{20}',\sigma_{20}\right)'$.
2. $\mathbb{E}\left[\mathbf{v}_i|\mathcal{F}_{i-1}\right]=\mathbf{0}$, $\mathbb{E}\left[\mathbf{v}_i|\mathcal{F}_{i-1},q_i\right]=\mathbb{E}\left[\mathbf{v}_i|v_{qi}\right]=\mathbf{g}_1\left(v_{qi};\varphi_0\right)$ and $\mathbb{E}\left[u_i|\mathcal{F}_{i-1},q_i\right]=\mathbb{E}\left[u_i|v_{qi}\right]=g_2\left(v_{qi};\kappa_0\right)$.
3. $\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]>0$, $\mathbb{E}\left[\|\mathbf{z}_i\|^2\right]<\infty$ and $\mathbb{E}\left[\|\mathbf{v}_i\|^2\right]<\infty$.
4. $\mathbb{E}\left[\left\|\mathbf{g}_1\left(v_{qi};\varphi_0\right)\right\|^2\right]<\infty$, $\mathbb{E}\left[g_2\left(v_{qi};\kappa_0\right)^2\right]<\infty$ and $\mathbb{E}\left[\left(e_i^0\right)^2\right]<\infty$.
5. For all $\gamma\in\Gamma$, $f(\gamma)\leq\overline{f}<\infty$, $f(\gamma)$ is continuous at γ_0 and $f:=f(\gamma_0)>0$, $P\left(q<\underline{\gamma}\right)>0$ and $P\left(q>\overline{\gamma}\right)>0$, where $f(\cdot)$ is the density function of q.
6. $\mathbf{g}_1\left(v_q;\varphi\right)$ and $g_2\left(v_q;\kappa\right)$ are Lipschitz functions in each of their arguments, i.e., there is a positive constant $C<\infty$ such that

$$\begin{aligned}\left\|\mathbf{g}_1\left(v_{q1};\varphi\right)-\mathbf{g}_1\left(v_{q2};\varphi\right)\right\| &\leq C\left|v_{q1}-v_{q2}\right|,\\ \left\|\mathbf{g}_1\left(v_q;\varphi_1\right)-\mathbf{g}_1\left(v_q;\varphi_2\right)\right\| &\leq C_1\left(v_q\right)\left\|\varphi_1-\varphi_2\right\|,\\ \left\|g_2\left(v_{q1};\kappa\right)-g_2\left(v_{q2};\kappa\right)\right\| &\leq C\left|v_{q1}-v_{q2}\right|,\end{aligned}$$

$$\|g_2(v_q;\kappa_1) - g_2(v_q;\kappa_2)\| \leq C_2(v_q)\|\kappa_1 - \kappa_2\|$$

with $\mathbb{E}\left[C_1(v_q)^2\right] < C$ and $\mathbb{E}\left[C_2(v_q)^2\right] < C$.

7. $P\left(\beta_1'\mathbf{g}_i + \beta_1'\mathbf{g}_1(v_{qi};\varphi) + \sigma_1 g_2(v_{qi};\kappa) \neq \beta_{10}'\mathbf{g}_i + \beta_{10}'\mathbf{g}_1(v_{qi};\varphi_0) + \sigma_{10}g_2(v_{qi};\kappa_0)|q\right) > 0$ for any $\theta_\ell \neq \theta_{\ell 0}$ and any q value in its support, where $\theta_\ell = \left(\beta_\ell', \sigma_\ell, \varphi', \kappa'\right)'$.

Obviously, C-I.1,3,5 are the same as C-II.1,3,5, and other assumptions are the counterparts of those in C-II. Especially, C-I.4 is implied by $\mathbb{E}\left[\|\mathbf{v}_i\|^2\right] < \infty$ and $\mathbb{E}\left[u_i^2\right] < \infty$, so it has some overlap with C-I.3; C-II.6 guarantees that replacing v_{qi} by $\widehat{v}_{qi}$ will not affect the consistency of $\widehat{\gamma}$. C-II.7 is the key assumption for identification of γ when $\widehat{v}_{qi}$ is replaced by v_{qi}, e.g., it excludes the case where $\mathbf{g}_1(v_{qi};\varphi)$ takes a linear form in ψ and $g_2(v_{qi};\kappa)$ takes a linear form in κ, and implicitly assumes $\delta_\beta'\mathbb{E}\left[\mathbf{g}_i\mathbf{g}_i'|q\right]\delta_\beta > 0$.

Lemma 6 *Under Assumption C-I,* $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$.

Proof. The proof follows the same lines of those of Lemma 1. First, we show that $S_n(\boldsymbol{\theta})$ and $\widetilde{S}_n(\boldsymbol{\theta})$ have the same probability limit, where $\widetilde{S}_n(\boldsymbol{\theta})$ is the same as $S_n(\boldsymbol{\theta})$ except replacing $\widehat{\mathbf{g}}_i$ by $\mathbf{g}_i$ and $\widehat{v}_{qi}$ by v_{qi}. Since

$$\begin{aligned}
& n^{-1}\left(S_n(\boldsymbol{\theta}) - \widetilde{S}_n(\boldsymbol{\theta})\right) \\
= & -2n^{-1}\sum_{i=1}^n \left[y_i - \left(\beta_1'\mathbf{g}_i + \beta_1'\mathbf{g}_{1,i,\varphi} + \sigma_1 g_{2,i,\kappa} + \beta_1'\Delta\widehat{\mathbf{g}}_i/2 + \beta_1'\Delta\widehat{\mathbf{g}}_{1,i,\varphi}/2 + \sigma_1\Delta\widehat{g}_{2,i,\kappa}/2\right)1(q_i \leq \gamma)\right. \\
& \left. - \left(\beta_2'\mathbf{g}_i + \beta_2'\mathbf{g}_{1,i,\varphi} + \sigma_2 g_{2,i,\kappa} + \beta_2'\Delta\widehat{\mathbf{g}}_i/2 + \beta_2'\Delta\widehat{\mathbf{g}}_{1,i,\varphi}/2 + \sigma_2\Delta\widehat{g}_{2,i,\kappa}/2\right)1(q_i > \gamma)\right] \\
& \cdot\left[\left(\beta_1'\Delta\widehat{\mathbf{g}}_i + \beta_1'\Delta\widehat{\mathbf{g}}_{1,i,\varphi} + \sigma_1\Delta\widehat{g}_{2,i,\kappa}\right)1(q_i \leq \gamma) + \left(\beta_2'\Delta\widehat{\mathbf{g}}_i + \beta_2'\Delta\widehat{\mathbf{g}}_{1,i,\varphi} + \sigma_2\Delta\widehat{g}_{2,i,\kappa}\right)1(q_i > \gamma)\right] \\
= & -2n^{-1}\sum_{i=1}^n \left[y_i - \left(\beta_1'\mathbf{g}_i + \beta_1'\mathbf{g}_{1,i,\varphi} + \sigma_1 g_{2,i,\kappa}\right)1(q_i \leq \gamma) - \left(\beta_2'\mathbf{g}_i + \beta_2'\mathbf{g}_{1,i,\varphi} + \sigma_2 g_{2,i,\kappa}\right)1(q_i > \gamma)\right] \\
& \cdot\left[\left(\beta_1'\Delta\widehat{\mathbf{g}}_i + \beta_1'\Delta\widehat{\mathbf{g}}_{1,i,\varphi} + \sigma_1\Delta\widehat{g}_{2,i,\kappa}\right)1(q_i \leq \gamma) + \left(\beta_2'\Delta\widehat{\mathbf{g}}_i + \beta_2'\Delta\widehat{\mathbf{g}}_{1,i,\varphi} + \sigma_2\Delta\widehat{g}_{2,i,\kappa}\right)1(q_i > \gamma)\right] \\
& + n^{-1}\sum_{i=1}^n \left[\left(\beta_1'\Delta\widehat{\mathbf{g}}_i + \beta_1'\Delta\widehat{\mathbf{g}}_{1,i,\varphi} + \sigma_1\Delta\widehat{g}_{2,i,\kappa}\right)1(q_i \leq \gamma) + \left(\beta_2'\Delta\widehat{\mathbf{g}}_i + \beta_2'\Delta\widehat{\mathbf{g}}_{1,i,\varphi} + \sigma_2\Delta\widehat{g}_{2,i,\kappa}\right)1(q_i > \gamma)\right]^2,
\end{aligned}$$

where $\mathbf{g}_{1,i,\varphi} = \mathbf{g}_1(v_{qi};\varphi)$, $g_{2,i,\kappa} = g_2(v_{qi};\kappa)$, and

$$\begin{aligned}
\Delta\widehat{\mathbf{g}}_i &= \widehat{\mathbf{g}}_i - \mathbf{g}_i = \left(\widehat{\Pi} - \Pi\right)'\mathbf{z}_i \leq \left\|\widehat{\Pi} - \Pi\right\|\|\mathbf{z}_i\| = o_p(1)\|\mathbf{z}_i\|, \\
\|\Delta\widehat{\mathbf{g}}_{1,i,\varphi}\| &= \|\mathbf{g}_1(\widehat{v}_{qi};\varphi) - \mathbf{g}_1(v_{qi};\varphi)\| \leq C|\widehat{v}_{qi} - v_{qi}| \leq C\|\widehat{\pi} - \pi\|\|\mathbf{z}_i\| = o_p(1)\|\mathbf{z}_i\|, \\
\Delta\widehat{g}_{2,i,\kappa} &= g_2(\widehat{v}_{qi};\kappa) - g_2(v_{qi};\kappa) \leq C|\widehat{v}_{qi} - v_{qi}| \leq C\|\widehat{\pi} - \pi\|\|\mathbf{z}_i\| = o_p(1)\|\mathbf{z}_i\|,
\end{aligned}$$

with the first inequalities for $\Delta\widehat{\mathbf{g}}_{1,i,\varphi}$ and $\Delta\widehat{g}_{2,i,\kappa}$ from C-I.6, the second inequalities for $\Delta\widehat{\mathbf{g}}_{1,i,\varphi}$ and $\Delta\widehat{g}_{2,i,\kappa}$ from $\widehat{v}_{qi} - v_{qi} = (\pi - \widehat{\pi})'\mathbf{z}_i$, and the last equalities for all three terms from C-I.2-3, we need only show that

$$n^{-1}\sum_{i=1}^n \left|y_i - \left(\beta_1'\mathbf{g}_i + \beta_1'\mathbf{g}_{1,i,\varphi} + \sigma_1 g_{2,i,\kappa}\right)1(q_i \leq \gamma) - \left(\beta_2'\mathbf{g}_i + \beta_2'\mathbf{g}_{1,i,\varphi} + \sigma_2 g_{2,i,\kappa}\right)1(q_i > \gamma)\right|\|\mathbf{z}_i\| = O_p(1) \tag{40}$$

uniformly in $\boldsymbol{\theta}$ and

$$n^{-1}\sum_{i=1}^n \|\mathbf{z}_i\|^2 = O_p(1). \tag{41}$$

(41) is straightforward by the ergodic theorem and C-I.3. As for (40), we apply Lemma 2.4 of Newey and McFadden (1994). Its almost sure continuity condition is implied by the continuity of $\mathbf{g}_{1,i,\varphi}$ in φ and $g_{2,i,\kappa}$

in κ (C-I.6) and the distribution of q (C-I.5). To check the summand is dominated by a function of w_i with finite first moment, we need only show $\mathbb{E}\left[y_i^2\right]<\infty$, $\mathbb{E}\left[\|\mathbf{g}_i\|^2\right]<\infty$, $\mathbb{E}\left[\left\|\mathbf{g}_{1,i,\varphi_0}\right\|^2\right]<\infty$ and $\mathbb{E}\left[g_{2,i,\kappa_0}^2\right]<\infty$ by the Cauchy-Schwarz inequality and

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{g}_{1,i,\varphi}\|^2\right] &\leq C\left(\mathbb{E}\left[\left\|\mathbf{g}_{1,i,\varphi_0}\right\|^2\right]+\|\varphi-\varphi_0\|^2\right),\\ \mathbb{E}\left[g_{2,i,\kappa}^2\right] &\leq C\left(\mathbb{E}\left[g_{2,i,\kappa_0}^2\right]+\|\kappa-\kappa_0\|^2\right)\end{aligned}$$

for any φ and κ (which is implied by C-I.6). $\mathbb{E}\left[\|\mathbf{g}_i\|^2\right]<\infty$ is implied by C-I.3, $\mathbb{E}\left[\left\|\mathbf{g}_{1,i,\varphi_0}\right\|^2\right]<\infty$ and $\mathbb{E}\left[g_{2,i,\kappa_0}^2\right]<\infty$ are assumed in C-I.4, and $\mathbb{E}\left[y_i^2\right]<\infty$ is implied by these two results and $\mathbb{E}\left[\left(e_i^0\right)^2\right]<\infty$ (C-I.4).

Second, we prove the consistency of $\widehat{\boldsymbol{\theta}}$ by applying Theorem 2.1 of Newey and McFadden (1994). For this purpose, we need only show that $\widetilde{S}_n(\boldsymbol{\theta})$ converges uniformly in probability to $\overline{S}(\boldsymbol{\theta})$ which is continuous and minimized uniquely at $\boldsymbol{\theta}_0$. By Lemma 2.4 of Newey and McFadden (1994) and the analysis above, $\widetilde{S}_n(\boldsymbol{\theta})$ converges uniformly in probability to

$$\overline{S}(\boldsymbol{\theta})=\mathbb{E}\left[\left\{y_i-\left(\beta_1'\mathbf{g}_i+\beta_1'\mathbf{g}_{1,i,\varphi}+\sigma_1 g_{2,i,\kappa}\right)1\left(q_i\leq\gamma\right)-\left(\beta_2'\mathbf{g}_i+\beta_2'\mathbf{g}_{1,i,\varphi}+\sigma_2 g_{2,i,\kappa}\right)1\left(q_i>\gamma\right)\right\}^2\right],$$

which is continuous in $\boldsymbol{\theta}$. From Section 2.2.2 of Newey and McFadden (1994), $\overline{S}(\boldsymbol{\theta})$ is minimized uniquely at $\boldsymbol{\theta}_0$ if $g(w_i;\boldsymbol{\theta})=g(w_i;\boldsymbol{\theta}_0)$ implies $\boldsymbol{\theta}=\boldsymbol{\theta}_0$, where $g(w_i;\boldsymbol{\theta}):=\left(\beta_1'\mathbf{g}_i+\beta_1'\mathbf{g}_1(v_{qi};\varphi)+\sigma_1 g_2(v_{qi};\kappa)\right)1(q_i\leq\gamma)-\left(\beta_2'\mathbf{g}_i+\beta_2'\mathbf{g}_1(v_{qi};\varphi)+\sigma_2 g_2(v_{qi};\kappa)\right)1(q_i>\gamma)$. The remaining analysis is exactly the same as in the proof of Lemma 1 so omitted. ■

Asymptotics for CF-I $\widehat{\gamma}$

The notations such as $D_0, V_0^{\pm}$ and $\bar{c}$ in Section 3.3 still apply with $\check{\mathbf{x}}_i$, κ_ℓ and $e_{\ell i}$ adapted to CF-I, where $e_i^0=e_{1i}1(q_i\leq\gamma_0)+e_{2i}1(q_i>\gamma_0)$ with $e_{\ell i}=\beta_\ell'\mathbf{v}_i+\sigma_\ell u_i-\kappa_\ell v_{qi}$. We impose the following assumptions which strengthen Assumption C-I.

Assumption I:

1. Conditions 1, 3-8, and 10 hold as in Assumption II.
2. $\mathbb{E}\left[\mathbf{v}_i|\mathcal{F}_{i-1}\right]=\mathbf{0}$, $\mathbb{E}\left[\mathbf{v}_i|\mathcal{F}_{i-1},q_i\right]=\mathbb{E}\left[\mathbf{v}_i|v_{qi}\right]=\varphi v_{qi}$ and $\mathbb{E}\left[u_i|\mathcal{F}_{i-1},q_i\right]=\mathbb{E}\left[u_i|v_{qi}\right]=\kappa v_{qi}$.
9. $\bar{c}'D_0\bar{c}>0, \bar{c}'V_0^{\pm}\bar{c}>0$ and $f>0$, where $\bar{c}=\begin{pmatrix} I_d & \mathbf{0}\\ \varphi' & \kappa\end{pmatrix}c$.

Because Assumption I is parallel to Assumption II, the comments on the latter can be applied to the former. We only provide more explanations on I.9 here. As in II.9, the assumption $\bar{c}'D_0\bar{c}>0$ in I.9 excludes the CTR in the augmented regression. When $c_\sigma=0$, the CTR in the original regression implies the CTR in the augmented regression. Suppose $\mathbf{x}=(1,q)'$, and then $c_\beta=(-\gamma_0,1)$ implies a CTR. Now, it is not hard to see that $\bar{c}=(-\gamma_0,1,1)$ if $c_\sigma=0$, so $\bar{c}'D_0\bar{c}=(-\gamma_0,1,1)\,\mathbb{E}\left[(1,\pi'\mathbf{z},v_q)'(1,\pi'\mathbf{z},v_q)\middle|q=\gamma_0\right](-\gamma_0,1,1)'=\mathbb{E}\left[(q-\gamma_0)^2|q=\gamma_0\right]=0$. When $c_\sigma\neq0$, however, even if the original regression is a CTR, the augmented regression need not be. If $\mathbf{x}=(1,q)'$ and $c_\beta=(-\gamma_0,1)$, then $\bar{c}=(-\gamma_0,1,1+\kappa c_\sigma)$. As a result, $\bar{c}'D_0\bar{c}=\mathbb{E}\left[(q-\gamma_0+\kappa c_\sigma v_q)^2|q=\gamma_0\right]=\kappa^2c_\sigma^2\mathbb{E}\left[v_q^2|q=\gamma_0\right]$ is positive in general. Also, $D_0>0$ implies that $\dim(\mathbf{z}_i)\geq\dim(\mathbf{x}_i)=d$ since $\mathbb{E}\left[\mathbf{g}_i\mathbf{g}_i'|q_i=\gamma_0\right]=\Pi'\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'|q_i=\gamma_0\right]\Pi>0$ implies $\dim(\mathbf{z}_i)\geq d$.

Theorem 3 *Under Assumption I, $a_n(\widehat{\gamma}-\gamma_0)$ has the same form asymptotic distribution as in Theorem 1 with new definitions of $D_0, V_0^{\pm}$ and $\bar{c}$.*

Proof. The proof is exactly the same as that of Theorem 1 except that $\widehat{e}_i = -\widehat{r}_i'\beta_2 + \widehat{r}_{qi}\kappa_2 + e_i^0$, and $\widehat{\mathbf{e}} = -\widehat{\mathbf{r}}\beta_2 + \widehat{\mathbf{r}}_q\kappa_2 + \mathbf{e}^0$ stacks $\widehat{e}_i$, where $\widehat{r}_i' = \mathbf{v}_i' - \widehat{\mathbf{v}}_i' = \mathbf{z}_i'\left(\widehat{\Pi} - \Pi\right) = (\widehat{r}_{xi}', \widehat{r}_{qi})$, and $\widehat{\mathbf{r}} = V - \widehat{V} = (\widehat{\mathbf{r}}_x, \widehat{\mathbf{r}}_q)$ stacks $\widehat{r}_i'$. ■

As in CF-II, extra randomness in the generated regressors would not affect the asymptotic distribution of $\widehat{\gamma}$. If $\delta_\sigma = 0$, then $\overline{c} = \begin{pmatrix} I_d \\ \varphi' \end{pmatrix} c_\beta$ is simplified. It seems hard to determine the relative magnitude of its particular coefficients ϕ and ω compared to those of LS even under the simplifications in Section 3.3. First, consider $\overline{c}'D_0\overline{c}$ vs. $c_\beta'\mathbb{E}\left[\mathbf{x}_i\mathbf{x}_i'|q_i = \gamma_0\right]c_\beta$. Since

$$\begin{aligned}
\overline{c}'D_0\overline{c} &= \left(c_\beta', c_\beta'\varphi + c_\sigma\kappa\right)\begin{pmatrix} \mathbb{E}\left[\mathbf{g}_i\mathbf{g}_i'|q_i = \gamma_0\right] & \mathbb{E}\left[\mathbf{g}_i v_{qi}|q_i = \gamma_0\right] \\ \mathbb{E}\left[v_{qi}\mathbf{g}_i'|q_i = \gamma_0\right] & \mathbb{E}\left[v_{qi}^2|q_i = \gamma_0\right] \end{pmatrix}\begin{pmatrix} c_\beta \\ c_\beta'\varphi + c_\sigma\kappa \end{pmatrix} \\
&= c_\beta'\mathbb{E}\left[\mathbf{g}_i\mathbf{g}_i'|q_i = \gamma_0\right]c_\beta + 2c_\beta'\mathbb{E}\left[\mathbf{g}_i v_{qi}|q_i = \gamma_0\right]\left(c_\beta'\varphi + c_\sigma\kappa\right) + \left(c_\beta'\varphi + c_\sigma\kappa\right)^2\mathbb{E}\left[v_{qi}^2|q_i = \gamma_0\right]
\end{aligned}$$

where $\mathbb{E}\left[v_{qi}^2|q_i = \gamma_0\right] = \mathbb{E}\left[(\gamma_0 - \pi'\mathbf{z}_i)^2|v_{qi} = \gamma_0 - \pi'\mathbf{z}_i\right] > 0$, the threshold effect information in the conditional mean δ_β (or $c_\beta = n^\alpha\delta_\beta$) is re-explored in δ_κ in the sense that $n^\alpha\delta_\kappa = c_\beta'\varphi + c_\sigma\kappa$ involves c_β. On the other hand, the variation of $\mathbf{g}_i$ is smaller than that of $\mathbf{x}_i$, i.e., $c_\beta'\mathbb{E}\left[\mathbf{g}_i\mathbf{g}_i'|q_i = \gamma_0\right]c_\beta$ tends to be smaller than $c_\beta'\mathbb{E}\left[\mathbf{x}_i\mathbf{x}_i'|q_i = \gamma_0\right]c_\beta$. So we lose some information in the first term $c_\beta'\mathbb{E}\left[\mathbf{g}_i\mathbf{g}_i'|q_i = \gamma_0\right]c_\beta$ and gain some information from the third term $\left(c_\beta'\varphi + c_\sigma\kappa\right)^2\mathbb{E}\left[v_{qi}^2|q_i = \gamma_0\right]$ compared with least squares, and the overall effect is not clear. Here, note that although $\mathbb{E}\left[\mathbf{g}_i v_{qi}\right] = \mathbb{E}\left[\mathbb{E}\left[\mathbf{g}_i v_{qi}|\mathbf{z}_i\right]\right] = \mathbf{0}$, the second term $2c_\beta'\mathbb{E}\left[\mathbf{g}_i v_{qi}|q_i = \gamma_0\right]\left(c_\beta'\varphi + c_\sigma\kappa\right) = 2c_\beta'\Pi'\mathbb{E}\left[\mathbb{E}\left[\mathbf{z}_i v_{qi}|v_{qi} = \gamma_0 - \pi'\mathbf{z}_i, \mathbf{z}_i\right]|v_{qi} = \gamma_0 - \pi'\mathbf{z}_i\right]\left(c_\beta'\varphi + c_\sigma\kappa\right) = 2c_\beta'\Pi'\mathbb{E}\left[\mathbf{z}_i\left(\gamma_0 - \pi'\mathbf{z}_i\right)|v_{qi} = \gamma_0 - \pi'\mathbf{z}_i\right]\left(c_\beta'\varphi + c_\sigma\kappa\right)$ is generally nonzero unless q is exogenous; actually, when q is exogenous, $\varphi = \mathbf{0}$ and $\kappa = 0$, so both the second and third terms disappear. Second, consider $e_{\ell i} = \beta_\ell'\mathbf{v}_i + \sigma_\ell u_i - \kappa_\ell v_{qi}$ vs. $\sigma_\ell u$. Although the variance of $\sigma_\ell u - \sigma_\ell\kappa v_q$ is smaller than that of $\sigma_\ell u$, it is hard to compare the variances of e_{1i} and $\sigma_1 u_i$ and the ratios $\frac{\mathbb{E}\left[e_{2i}^2\right]}{\mathbb{E}\left[e_{1i}^2\right]}$ and $\frac{\sigma_2^2}{\sigma_1^2}$.

Likelihood Ratio Tests

The LR statistic still takes the form

$$LR_n(\gamma) = \frac{S_n(\gamma) - S_n(\widehat{\gamma})}{2\widehat{\eta}^2},$$

where $\widehat{\eta}^2$ is a consistent estimator of $\eta^2 = \overline{c}'V_0^-\overline{c}/\overline{c}'D_0\overline{c}$.

Corollary 2 *Under the assumptions in Theorem 3, $LR_n(\gamma_0)$ has the same form asymptotic distribution as in Theorem 1 with the new definition of $\phi = \overline{c}'V_0^+\overline{c}/\overline{c}'V_0^-\overline{c}$.*

Proof. The proof is exactly the same as that of Corollary 1. ■

If the model is homoskedastic, then $\phi = 1$ and $\widehat{\eta}^2$ in $LR_n(\gamma)$ can be replaced by an estimate of $\mathbb{E}\left[e_i^{02}\right]$, such as $S_n(\widehat{\gamma})/n$ as in CH. However, the model is generally heteroskedastic. For example,

$$\begin{aligned}
V_0^- &= \mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i'\left(\beta_1'\left(\mathbf{v}_i - \varphi v_{qi}\right) + \sigma_1\left(u_i - \kappa v_{qi}\right)\right)^2|q_i = \gamma_0-\right], \\
V_0^+ &= \mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i'\left(\beta_2'\left(\mathbf{v}_i - \varphi v_{qi}\right) + \sigma_2\left(u_i - \kappa v_{qi}\right)\right)^2|q_i = \gamma_0+\right].
\end{aligned}$$

Even if $\mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i'\left\|\mathbf{v}_i - \varphi v_{qi}\right\|^2|q_i = \gamma\right]$ and $\mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i'\left(u_i - \kappa v_{qi}\right)^2|q_i = \gamma\right]$ are continuous at γ_0, we have $V_0^- \neq V_0^+$ if $\beta_1 \neq \beta_2$ and/or $\sigma_1 \neq \sigma_2$.

For inference based on $LR_n(\gamma)$, we need to estimate η^2 and ϕ. The estimation method is exactly the same as in CF-II except with new definitions of r_{1i}, r_{2i} and r_{3i}. The nuisance parameter ϕ is not easy to simplify unless $\mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i' e_{1i}^2 | q_i = \gamma_0 -\right] = \mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i' | q_i = \gamma_0\right]\mathbb{E}\left[e_{1i}^2\right]$ and $\mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i' e_{2i}^2 | q_i = \gamma_0 +\right] = \mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i' | q_i = \gamma_0\right]\mathbb{E}\left[e_{2i}^2\right]$, which would hold under the homoskedasticity assumption $Var\left(\left(\mathbf{v}', u\right)' \middle| \mathbf{z}, q\right) = Var\left(\left(\mathbf{v}', u\right)' | v_q\right) = \mathbb{E}\left[Var\left(\left(\mathbf{v}', u\right)' | v_q\right)\right]$ because

$$
\begin{aligned}
\mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i' e_{\ell i}^2 | q_i\right] &= \mathbb{E}\left[\mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i' e_{\ell i}^2 | \mathbf{z}_i, q_i\right] | q_i\right] = \mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i' \left(\beta_\ell', \sigma_\ell\right) Var\left(\left(\mathbf{v}_i', u_i\right)' | v_{qi}\right)\left(\beta_\ell', \sigma_\ell\right)' | q_i\right] \\
&= \mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i' | q_i\right]\left(\beta_\ell', \sigma_\ell\right)\mathbb{E}\left[Var\left(\left(\mathbf{v}_i', u_i\right)' | v_{qi}\right)\right]\left(\beta_\ell', \sigma_\ell\right)' = \mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i' | q_i\right]\mathbb{E}\left[e_{\ell i}^2\right].
\end{aligned}
$$

In this case, $\phi = \mathbb{E}\left[e_{2i}^2\right]/\mathbb{E}\left[e_{1i}^2\right]$ can be estimated by $\widehat{\phi} = \widehat{\sigma}_2^2/\widehat{\sigma}_1^2$ if the homoskedasticity assumptions $\mathbb{E}\left[e_{1i}^2 | q_i\right] = \mathbb{E}\left[e_{1i}^2\right]$ when $q_i \le \gamma_0$ and $\mathbb{E}\left[e_{2i}^2 | q_i\right] = \mathbb{E}\left[e_{2i}^2\right]$ when $q_i > \gamma_0$ hold, where $\widehat{\sigma}_\ell^2 = \left|\mathcal{I}_\ell\right|^{-1}\sum_{i\in\mathcal{I}_\ell}\widehat{e}_{\ell i}^2$ is a consistent estimator of $\mathbb{E}\left[\left(\beta_\ell'\left(\mathbf{v}_i - \varphi v_{qi}\right) + \sigma_\ell\left(u_i - \kappa v_{qi}\right)\right)^2\right]$, $\widehat{e}_{1i} = y_i - \widehat{\check{\mathbf{x}}}_i'\widehat{\overline{\beta}}_1$ is defined only for $i \in \mathcal{I}_1 := \{i | q_i \le \widehat{\gamma}\}$, and $\widehat{e}_{2i} = y_i - \widehat{\check{\mathbf{x}}}_i'\widehat{\overline{\beta}}_2$ is defined only for $i \in \mathcal{I}_2 := \{i | q_i > \widehat{\gamma}\}$ with some consistent estimators of $\overline{\beta}_\ell$ in Section 4. If $\mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i' e_{1i}^2 | q_i = \gamma_0 -\right] = \mathbb{E}\left[\check{\mathbf{x}}_i\check{\mathbf{x}}_i' | q_i = \gamma_0\right]\mathbb{E}\left[e_{1i}^2\right]$, then $\eta^2 = \mathbb{E}\left[e_{1i}^2\right]$ can be estimated by $\widehat{\sigma}_1^2$ under the homoskedasticity assumption $\mathbb{E}\left[e_{1i}^2 | q_i\right] = \mathbb{E}\left[e_{1i}^2\right]$ when $q_i \le \gamma_0$. Given the estimates of η^2 and ϕ, the LR-CI of γ takes the same form as in Section 3.4.

Asymptotics for GMM-I2 $\widehat{\beta}$

The following theorem states the asymptotic distribution of GMM-I2 $\widehat{\beta}$.

Theorem 4 *Under Assumption I,*

$$
n^{1/2}\left(\widehat{\overline{\beta}}_\ell - \overline{\beta}_\ell\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \Sigma_\ell\right),
$$

where

$$
\Sigma_\ell = \left(G_\ell' W_\ell G_\ell\right)^{-1}\left(G_\ell' W_\ell \Omega_\ell W_\ell G_\ell\right)\left(G_\ell' W_\ell G_\ell\right)^{-1}.
$$

In Σ_ℓ, $G_1 = \mathbb{E}\left[\check{\mathbf{z}}_i\check{\mathbf{x}}_{i,\le\gamma_0}'\right]$, $G_2 = \mathbb{E}\left[\check{\mathbf{z}}_i\check{\mathbf{x}}_{i,>\gamma_0}'\right]$ *and* $G_\ell' W_\ell G_\ell > 0$,

$$
\Omega_\ell = \Omega_1^\ell + \Omega_{12}^\ell + \Omega_{21}^\ell + \Omega_2^\ell,
$$

with

$$
\begin{aligned}
\Omega_1^1 &= \mathbb{E}\left[\check{\mathbf{z}}_i\check{\mathbf{z}}_{i,\le\gamma_0}' e_{1i}^2\right] > 0, \Omega_1^2 = \mathbb{E}\left[\check{\mathbf{z}}_i\check{\mathbf{z}}_{i,>\gamma_0}' e_{2i}^2\right] > 0, \\
\Omega_2^1 &= \mathbb{E}\left[\check{\mathbf{z}}_i\mathbf{z}_{i,\le\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\left(\overline{\kappa}_1'\mathbf{v}_i\mathbf{v}_i'\overline{\kappa}_1\right)\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\check{\mathbf{z}}_{i,\le\gamma_0}'\right], \\
\Omega_2^2 &= \mathbb{E}\left[\check{\mathbf{z}}_i\mathbf{z}_{i,>\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\left(\overline{\kappa}_2'\mathbf{v}_i\mathbf{v}_i'\overline{\kappa}_2\right)\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\check{\mathbf{z}}_{i,>\gamma_0}'\right], \\
\Omega_{21}^1 &= \mathbb{E}\left[\check{\mathbf{z}}_i\mathbf{z}_{i,\le\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\check{\mathbf{z}}_{i,\le\gamma_0}'\overline{\kappa}_1'\mathbf{v}_i e_{1i}\right], \Omega_{12}^1 = \Omega_{21}^{1\prime}, \\
\Omega_{21}^2 &= \mathbb{E}\left[\check{\mathbf{z}}_i\mathbf{z}_{i,>\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\check{\mathbf{z}}_{i,>\gamma_0}'\overline{\kappa}_2'\mathbf{v}_i e_{2i}\right], \Omega_{12}^2 = \Omega_{21}^{2\prime},
\end{aligned}
$$

and $\overline{\kappa}_\ell := \left(-\beta_{\ell x}', \kappa_\ell - \beta_{\ell q}\right)'$. *When* $W_\ell = \Omega_\ell^{-1}$, Σ_ℓ *reduces to* $\left(G_\ell'\Omega_\ell^{-1}G_\ell\right)^{-1}$. *The asymptotic covariance matrix between* $n^{1/2}\left(\widehat{\overline{\beta}}_1 - \overline{\beta}_1\right)$ *and* $n^{1/2}\left(\widehat{\overline{\beta}}_2 - \overline{\beta}_2\right)$ *is*

$$
C_{GMM} = \left(G_1' W_1 G_1\right)^{-1} G_1' W_1 \Omega_{12} W_2 G_2 \left(G_2' W_2 G_2\right)^{-1},
$$

with $\Omega_{12}=\mathbb{E}\left[\breve{\mathbf{z}}_i\mathbf{z}'_{i,\le\gamma_0}\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}'_i\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\mathbf{z}'_i\left(\overline{\kappa}'_1\mathbf{v}_i\mathbf{v}'_i\overline{\kappa}_2\right)\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}'_i\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\breve{\mathbf{z}}'_{i,>\gamma_0}\right]$. *The* $\mathcal{N}(\mathbf{0},\Sigma_\ell)$ *limit distribution is independent of* $\zeta(\phi)$ *in Theorem 3.*

Proof. In the proof of Theorem 2, we need only replace $\widehat{\mathbf{r}}\kappa_1$ by $-\widehat{\mathbf{r}}\beta_1+\widehat{\mathbf{r}}_q\kappa_1=\widehat{\mathbf{r}}\begin{pmatrix}-\beta_{1x}\\ \kappa_1-\beta_{1q}\end{pmatrix}=:\widehat{\mathbf{r}}\overline{\kappa}_1$, so we just replace κ_1 in GMM-II2 by $\overline{\kappa}_1$ in GMM-I2. The two influence functions $\mathbf{z}_i\mathbf{v}'_i\overline{\kappa}_1$ and $\breve{\mathbf{z}}_{i,\le\gamma_0}e_{1i}$ are correlated, so we also have two cross terms in the asymptotic variance matrix. Specifically, the cross terms are

$$\mathbb{E}\left[\breve{\mathbf{z}}_i\mathbf{z}'_{i,\le\gamma_0}\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}'_i\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\mathbf{v}'_i\overline{\kappa}_1e_{1i}\breve{\mathbf{z}}'_{i,\le\gamma_0}\right]$$

and its transpose.

The covariance matrix in GMM-I2 just replaces κ_ℓ in GMM-II2 by $\overline{\kappa}_\ell$.

The asymptotic independence between $\mathcal{N}(\mathbf{0},\Sigma_\ell)$ and $\zeta(\phi)$ follows from Proposition 3. ■

Compared with GMM-II2, GMM-I2 has some extra terms in Σ_ℓ, i.e., Ω^ℓ_{12} and Ω^ℓ_{21}. Take $\widehat{\overline{\beta}}_1$ as an example, where the extra terms arise because the two random components in the influence functions $\mathbf{z}_i\mathbf{v}'_i\overline{\kappa}_1$ and $\breve{\mathbf{z}}_{i,\le\gamma_0}e_{1i}$ are correlated in CF-I. To see why, note that in general

$$\begin{aligned}&\mathbb{E}\left[\mathbf{v}_ie_{1i}|\mathbf{z}_i,q_i\right]=\mathbb{E}\left[\mathbf{v}_i\left(\beta'_1\left(\mathbf{v}_i-\varphi v_{qi}\right)+\sigma_1\left(u_i-\kappa v_{qi}\right)\right)|\mathbf{z}_i,q_i\right]\\&=\mathbb{E}\left[\left(\mathbf{v}_i-\varphi v_{qi}\right)\left(\mathbf{v}_i-\varphi v_{qi}\right)'|v_q\right]\beta_1+\sigma_1\mathbb{E}\left[\left(\mathbf{v}_i-\varphi v_{qi}\right)\left(u_i-\kappa v_{qi}\right)|v_q\right]\\&=Var\left(\mathbf{v}|v_q\right)\beta_1+\sigma_1Cov\left(\mathbf{v},u|v_q\right)\ne\mathbf{0},\end{aligned}$$

so that

$$\mathbb{E}\left[\mathbf{z}_i\breve{\mathbf{z}}'_{i,\le\gamma_0}\mathbf{v}'_i\overline{\kappa}_1e_{1i}\right]=\mathbb{E}\left[\mathbb{E}\left[\mathbf{z}_i\breve{\mathbf{z}}'_{i,\le\gamma_0}\mathbf{v}'_i\overline{\kappa}_1e_{1i}\middle|\mathbf{z}_i,q_i\right]\right]=\mathbb{E}\left[\mathbf{z}_i\breve{\mathbf{z}}'_{i,\le\gamma_0}\overline{\kappa}_1\left[Var\left(\mathbf{v}|v_q\right)\beta_1+\sigma_1Cov\left(\mathbf{v},u|v_q\right)\right]\right]\ne\mathbf{0}.$$

Even in CH's framework, $\Omega^1_{12}\ne\mathbf{0}$ and $\Omega^1_{21}\ne\mathbf{0}$. Specifically, $\breve{\mathbf{z}}_i=\mathbf{z}_i$, $\kappa_1=0$, $v_{qi}=0$, and $e_{1i}=\beta'_1\mathbf{v}+\sigma_1u=v'_{xi}\beta_{1x}+\sigma_1u$, so that

$$\begin{aligned}\Omega^1_{21}&=\mathbb{E}\left[\mathbf{z}_{i,\le\gamma_0}\mathbf{z}'_i\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}'_i\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\mathbf{z}'_{i,\le\gamma_0}\left(-v'_{xi}\beta_{1x}\right)e_{1i}\right]\\&=-\mathbb{E}\left[\mathbf{z}_{i,\le\gamma_0}\mathbf{z}'_i\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}'_i\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\mathbf{z}'_{i,\le\gamma_0}\left(\sigma_1uv'_{xi}\beta_{1x}+\beta'_{1x}v_{xi}v'_{xi}\beta_{1x}\right)\right]\ne\mathbf{0},\end{aligned}$$

and similarly $\Omega^1_{12}\ne\mathbf{0}$.

The comments in the main text on the estimation of Σ_ℓ and CI construction of β in GMM-II2 can also be applied to GMM-I2.

SD.3 CF and GMM-1 Estimators for β

In this section, we provide detailed discussions on the CF and GMM-1 estimators of β. Before discussing the GMM-1 estimators we first show that the moment conditions in CH and KST will not generate consistent estimates of β in general when q is endogenous.

CF Estimators

The following theorem gives the asymptotic properties of $\left(\widehat{\overline{\beta}}'_1,\widehat{\overline{\beta}}'_2\right)'$. Different from $\widehat{\gamma}$, the extra randomness in the generated regressors now affects the asymptotic efficiency of $\left(\overline{\beta}'_1,\overline{\beta}'_2\right)'$.

Theorem 5 *Under Assumption I for CF-I and Assumption II for CF-II,*

$$n^{1/2}\left(\widehat{\overline{\beta}}_\ell-\overline{\beta}_\ell\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\Sigma_\ell\right),$$

where $\Sigma_\ell=M_\ell^{-1}\Omega_\ell M_\ell^{-1}$ *with* $M_1=M_0$ *and* $M_2=M-M_0$. *In CF-I,*

$$\Omega_\ell=V_\ell+\Omega_{12}^\ell+\Omega_{21}^\ell+\Omega_2^\ell,$$

with

$$\begin{aligned}
V_1 &= \mathbb{E}\left[\breve{\mathbf{x}}_i\breve{\mathbf{x}}_i'e_{1i}^2 1(q_i\le\gamma_0)\right]>0, V_2=\mathbb{E}\left[\breve{\mathbf{x}}_i\breve{\mathbf{x}}_i'e_{2i}^2 1(q_i>\gamma_0)\right]>0,\\
\Omega_2^1 &= \mathbb{E}\left[\breve{\mathbf{x}}_i\mathbf{z}_{i,\le\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\left(\overline{\kappa}_1'\mathbf{v}_i\mathbf{v}_i'\overline{\kappa}_1\right)\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\breve{\mathbf{x}}_{i,\le\gamma_0}'\right],\\
\Omega_2^2 &= \mathbb{E}\left[\breve{\mathbf{x}}_i\mathbf{z}_{i,>\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\left(\overline{\kappa}_2'\mathbf{v}_i\mathbf{v}_i'\overline{\kappa}_2\right)\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\breve{\mathbf{x}}_{i,>\gamma_0}'\right],\\
\Omega_{21}^1 &= \mathbb{E}\left[\breve{\mathbf{x}}_i\mathbf{z}_{i,\le\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\breve{\mathbf{x}}_{i,\le\gamma_0}'\overline{\kappa}_1\mathbf{v}_ie_{1i}\right],\Omega_{12}^1=\Omega_{21}^{1\prime},\\
\Omega_{21}^2 &= \mathbb{E}\left[\breve{\mathbf{x}}_i\mathbf{z}_{i,>\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\breve{\mathbf{x}}_{i,>\gamma_0}'\overline{\kappa}_2\mathbf{v}_ie_{2i}\right],\Omega_{12}^2=\Omega_{21}^{2\prime};
\end{aligned}$$

and $\overline{\kappa}_\ell:=\left(-\beta_{\ell x}',\kappa_\ell-\beta_{\ell q}\right)'$, *and in CF-II,*

$$\Omega_\ell=V_\ell+\Omega_2^\ell,$$

with Ω_2^ℓ *taking the same form as in CF-I but replacing* $\overline{\kappa}_\ell$ *by* κ_ℓ.

In CF-I, the asymptotic covariance matrix between $n^{1/2}\left(\widehat{\overline{\beta}}_1-\overline{\beta}_1\right)$ *and* $n^{1/2}\left(\widehat{\overline{\beta}}_2-\overline{\beta}_2\right)$ *is*

$$C_{CF}=M_1^{-1}\Omega_{12}M_2^{-1}$$

with $\Omega_{12}=\mathbb{E}\left[\breve{\mathbf{x}}_i\mathbf{z}_{i,\le\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\left(\overline{\kappa}_1'\mathbf{v}_i\mathbf{v}_i'\overline{\kappa}_2\right)\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\breve{\mathbf{x}}_{i,>\gamma_0}'\right]$, *and in CF-II, again simply replace* $\overline{\kappa}_\ell$ *by* κ_ℓ *in the above formula.*

Proof of Theorem 5. Take $\widehat{\overline{\beta}}_1$ as an example since $\widehat{\overline{\beta}}_2$ can be similarly analyzed. Note that in CF-II,

$$\begin{aligned}
n^{1/2}\left(\widehat{\overline{\beta}}_1-\overline{\beta}_1\right) &= n^{1/2}\left(\widehat{X}_1'\widehat{X}_1\right)^{-1}\widehat{X}_1'\left(\breve{X}\overline{\beta}_2+\breve{X}_0\overline{c}n^{-\alpha}-\widehat{X}_1\overline{\beta}_1+\mathbf{e}^0\right)\\
&= \left(\frac{1}{n}\widehat{X}_1'\widehat{X}_1\right)^{-1}\frac{1}{\sqrt{n}}\widehat{X}_1'\left(\left(\breve{X}_1-\widehat{X}_1\right)\overline{\beta}_1-\breve{X}_b\overline{c}n^{-\alpha}+\mathbf{e}^0\right)\\
&= \left(\frac{1}{n}\widehat{X}_1'\widehat{X}_1\right)^{-1}\frac{1}{\sqrt{n}}\widehat{X}_1'\left(\widehat{\mathbf{r}}\kappa_1-\breve{X}_b\overline{c}n^{-\alpha}+\mathbf{e}^0\right)
\end{aligned}$$

where $\breve{X}_1$ and $\breve{X}_b$ are matrices stacking $\breve{\mathbf{x}}_i'1(q_i\le\widehat{\gamma})$ and $\breve{\mathbf{x}}_i'\left(1\left(q_i\le\widehat{\gamma}\right)-1\left(q_i\le\gamma_0\right)\right)$, respectively. By Lemma 2, the continuity of M_γ and consistency of $\widehat{\gamma}$, $\frac{1}{n}\widehat{X}_1'\widehat{X}_1\xrightarrow{p}M_0$. We now study the asymptotic distribution of $\frac{1}{\sqrt{n}}\widehat{X}_1'\widehat{\mathbf{r}}\kappa_1$, $\frac{n^{-\alpha}}{\sqrt{n}}\sum_{i=1}^n\widehat{\breve{\mathbf{x}}}_i\breve{\mathbf{x}}_i'\overline{c}1\left(\gamma_0<q_i\le\widehat{\gamma}\right)$ and $\frac{1}{\sqrt{n}}\widehat{X}_1'\mathbf{e}^0$.

First,

$$\frac{1}{\sqrt{n}}\widehat{X}_1'\widehat{\mathbf{r}}\kappa_1=\frac{1}{\sqrt{n}}\widehat{X}_1'Z\left(\widehat{\Pi}-\Pi\right)\kappa_1=\frac{1}{\sqrt{n}}\widehat{X}_1'Z\left(Z'Z\right)^{-1}Z'V\kappa_1=\left(\frac{1}{n}\widehat{X}_1'Z\right)\left(\frac{1}{n}Z'Z\right)^{-1}\frac{1}{\sqrt{n}}Z'V\kappa_1,$$

where $\frac{1}{n}\widehat{X}_1'Z \xrightarrow{p} \mathbb{E}\left[\breve{\mathbf{x}}_{i,\leq\gamma_0}\mathbf{z}_i'\right], \frac{1}{n}Z'Z \xrightarrow{p} \mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]$, and $\frac{1}{\sqrt{n}}Z'V\kappa_1 \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\left(\mathbf{v}_i'\kappa_1\right)^2\right]\right)$ with the influence function $\mathbf{z}_i\mathbf{v}_i'\kappa_1$. Second, by Lemma A.10 of Hansen (2000), uniformly on $v \in [-\overline{v}, \overline{v}]$,

$$n^{-2\alpha}\sum_{i=1}^n \breve{\mathbf{x}}_i\breve{\mathbf{x}}_i' 1\left(q_i \leq \gamma_0 + v/a_n\right) - 1\left(q_i \leq \gamma_0\right) = O_p(1),$$

and

$$\begin{aligned} n^{-2\alpha}\sum_{i=1}^n \widehat{r}_i\breve{\mathbf{x}}_i' 1\left(q_i \leq \gamma_0 + v/a_n\right) - 1\left(q_i \leq \gamma_0\right) &= \left(\widehat{\Pi} - \Pi\right)' n^{-2\alpha}\sum_{i=1}^n \mathbf{z}_i\breve{\mathbf{x}}_i' 1\left(q_i \leq \gamma_0 + v/a_n\right) - 1\left(q_i \leq \gamma_0\right) \\ &= O_p\left(n^{-1/2}\right) O_p(1) = O_p\left(n^{-1/2}\right). \end{aligned}$$

Since $a_n\left(\widehat{\gamma} - \gamma_0\right) = O_p(1)$, we have

$$\begin{aligned} \frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{\breve{\mathbf{x}}}_i\breve{\mathbf{x}}_i'\overline{c}n^{-\alpha}\left(1\left(q_i \leq \widehat{\gamma}\right) - 1\left(q_i \leq \gamma_0\right)\right) &= n^{-1/2}O_p\left(n^{2\alpha} + n^{2\alpha}n^{-1/2}\right) n^{-\alpha} \\ &= O_p(a_n^{-1/2}) = o_p(1). \end{aligned}$$

Third, by Lemma 2,

$$\frac{1}{\sqrt{n}}\widehat{X}_1'\mathbf{e}^0 - \frac{1}{\sqrt{n}}\breve{X}_1\mathbf{e}^0 = o_p(1).$$

By Lemma A.4 of Hansen (2000),

$$\frac{1}{\sqrt{n}}\breve{X}_1\mathbf{e}^0 \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbb{E}\left[\breve{\mathbf{x}}_i\breve{\mathbf{x}}_{i,\leq\gamma_0}' e_{1i}^2\right]\right),$$

with the influence function $\breve{\mathbf{x}}_{i,\leq\gamma_0}e_{1i}$. Because the two influence functions $\mathbf{z}_i\mathbf{v}_i'\kappa_1$ and $\breve{\mathbf{x}}_{i,\leq\gamma_0}e_{1i}$ are not correlated, the asymptotic variance of $n^{1/2}\left(\widehat{\overline{\beta}}_1 - \overline{\beta}_1\right)$ is Σ_1.

In CF-I, we need only replace $\widehat{\mathbf{r}}\kappa_1$ by $-\widehat{\mathbf{r}}\beta_1 + \widehat{\mathbf{r}}_q\kappa_1 = \widehat{\mathbf{r}}\begin{pmatrix} -\beta_{1x} \\ \kappa_1 - \beta_{1q} \end{pmatrix} =: \widehat{\mathbf{r}}\overline{\kappa}_1$, so we just replace κ_1 in CF-II by $\overline{\kappa}_1$ in CF-I. The two influence functions $\mathbf{z}_i\mathbf{v}_i'\overline{\kappa}_1$ and $\breve{\mathbf{x}}_{i,\leq\gamma_0}e_{1i}^0$ are correlated, so we also have two cross terms in the asymptotic variance matrix. Specifically, the cross terms are

$$\mathbb{E}\left[\breve{\mathbf{x}}_{i,\leq\gamma_0}\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\mathbf{v}_i'\overline{\kappa}_1 e_{1i}\breve{\mathbf{x}}_{i,\leq\gamma_0}'\right],$$

and its transpose.

To study the asymptotic covariance matrix between $\widehat{\overline{\beta}}_1$ and $\widehat{\overline{\beta}}_2$ in CF-II, note that the influence function for $\widehat{\overline{\beta}}_1$ is

$$M_1^{-1}\left(\mathbb{E}\left[\breve{\mathbf{x}}_{i,\leq\gamma_0}\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbf{z}_i\mathbf{v}_i'\kappa_1 + \breve{\mathbf{x}}_{i,\leq\gamma_0}e_{1i}\right),$$

and for $\widehat{\overline{\beta}}_2$ is

$$M_2^{-1}\left(\mathbb{E}\left[\breve{\mathbf{x}}_{i,>\gamma_0}\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbf{z}_i\mathbf{v}_i'\kappa_2 + \breve{\mathbf{x}}_{i,>\gamma_0}e_{2i}\right).$$

Note that the second parts of the two influence functions are not correlated, so we need only consider the correlation between the first parts. Specifically, the covariance matrix is

$$M_1^{-1}\mathbb{E}\left[\breve{\mathbf{x}}_{i,\leq\gamma_0}\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\left(\mathbf{v}_i'\kappa_1\right)\left(\mathbf{v}_i'\kappa_2\right)\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\breve{\mathbf{x}}_{i,>\gamma_0}'\right]M_2^{-1}.$$

The covariance matrix in CF-I just replaces κ_ℓ by $\overline{\kappa}_\ell$. ■

Because GMM-2 is an extension of CF, most of the comments on the GMM-2 estimators remain relevant here. For example, the randomness of the generated regressors will not disappear in the asymptotic distributions of $\widehat{\widehat{\beta}}_\ell$, the matrices Ω_2^ℓ and C_{CF} can be simplified when the first-stage regression is homoskedastic, the GMM-I1 estimator has two more cross terms than the GMM-II1 estimator in the asymptotic variance matrices, the estimates $\widehat{\widehat{\beta}}_1$ and $\widehat{\widehat{\beta}}_2$ are not asymptotically independent, the estimator $\underline{\widehat{\boldsymbol{\theta}}}$ is asymptotically independent of our CF estimators $\widehat{\gamma}$, and so $\widehat{\gamma}$ does not affect the asymptotic distribution of $\underline{\widehat{\boldsymbol{\theta}}}$. Estimation of Σ_ℓ is needed for inference and a plug in procedure may be used for this purpose but for brevity is not pursued here.

Overviewing the asymptotic properties of $\widehat{\boldsymbol{\theta}}$ in the two CF approaches, it seems that CF-II is the more convenient and appealing from many perspectives. In particular, it is more robust to discreteness in x, it has a firmer theoretical foundation, the γ estimator should be more efficient than the LS estimator, the nuisance parameters ϕ and η^2 in the LR statistic can be simplified under weaker homoskedasticity assumptions, and the asymptotic variance matrix of $\underline{\widehat{\boldsymbol{\theta}}}$ is easier to estimate.

Finally, we detail the asymptotic variances of CF-I and CF-II under DGP2 in our simulations of Section 5.3. Because δ plays the role of β_1, we only report Ω_1 here. In both CF-I and CF-II, the matrix V_1 can be simplified as

$$V_1 = \mathbb{E}\left[\check{\mathbf{x}}_i \check{\mathbf{x}}_i' 1(q_i \le \gamma_0)\right] \mathbb{E}\left[e_{1i}^2\right]$$

where $\check{\mathbf{x}}_i = (\Pi_x' \mathbf{z}_i, v_{qi})'$ and $e_{1i} = (\delta + \psi_x) e_{xi} + e_{ui}$ in CF-I and $\check{\mathbf{x}}_i = (x_i, v_{xi}, v_{qi})'$ and $e_{1i} = e_{ui}$ in CF-II. In CF-I, Ω_2^1 can be simplified as

$$\Omega_2^1 = (-\delta, \kappa_1) \mathbb{E}\left[\mathbf{v}_i \mathbf{v}_i'\right] (-\delta, \kappa_1)' \, \mathbb{E}\left[\check{\mathbf{x}}_{i,\le\gamma_0} \mathbf{z}_i'\right] \mathbb{E}\left[\mathbf{z}_i \mathbf{z}_i'\right]^{-1} \mathbb{E}\left[\mathbf{z}_i \check{\mathbf{x}}_{i,\le\gamma_0}'\right],$$

and in CF-II, Ω_2^1 can be simplified as

$$\Omega_2^1 = \mathbb{E}\left[\left(\psi' \mathbf{v}_i\right)^2\right] \mathbb{E}\left[\check{\mathbf{x}}_{i,\le\gamma_0} \mathbf{z}_i'\right] \mathbb{E}\left[\mathbf{z}_i \mathbf{z}_i'\right]^{-1} \mathbb{E}\left[\mathbf{z}_i \check{\mathbf{x}}_{i,\le\gamma_0}'\right].$$

Also, in CF-I, the cross term

$$\begin{aligned}
\Omega_{21}^1 &= \mathbb{E}\left[\check{\mathbf{x}}_{i,\le\gamma_0} \mathbf{z}_i'\right] \mathbb{E}\left[\mathbf{z}_i \mathbf{z}_i'\right]^{-1} \mathbb{E}\left[\mathbf{z}_i \check{\mathbf{x}}_{i,\le\gamma_0}' (-\delta, \kappa_1) \mathbf{v}_i e_{1i}\right] \\
&= \mathbb{E}\left[\check{\mathbf{x}}_{i,\le\gamma_0} \mathbf{z}_i'\right] \mathbb{E}\left[\mathbf{z}_i \mathbf{z}_i'\right]^{-1} \mathbb{E}\left[\mathbf{z}_{i,\le\gamma_0} (\Pi_x' \mathbf{z}_i, v_{qi}) \left((\kappa_1 - \varphi\delta) v_{qi} - \delta e_{xi}\right) \left((\delta + \psi_x) e_{xi} + e_{ui}\right)\right] \\
&= -\delta (\delta + \psi_x) \mathbb{E}\left[e_{xi}^2\right] \mathbb{E}\left[\check{\mathbf{x}}_{i,\le\gamma_0} \mathbf{z}_i'\right] \mathbb{E}\left[\mathbf{z}_i \mathbf{z}_i'\right]^{-1} \mathbb{E}\left[\mathbf{z}_i \check{\mathbf{x}}_{i,\le\gamma_0}'\right] \ne \mathbf{0}.
\end{aligned}$$

In GMM-2, $G_1 = \mathbb{E}\left[\check{\mathbf{z}}_i \check{\mathbf{x}}_{i,\le\gamma_0}'\right]$, where $\check{\mathbf{z}}_i = (\mathbf{z}_i', q_i)'$ in GMM-I2 and $\check{\mathbf{z}}_i = (\mathbf{z}_i', x_i, q_i)'$ in GMM-II2. For Ω_1^1, Ω_2^1 and Ω_{21}^1 in GMM-I2 and Ω_1^1 and Ω_2^1 in GMM-II2, just replace the $\check{\mathbf{x}}_i$ in V_1, Ω_2^1 and Ω_{21}^1 of CF-I and V_1 and Ω_2^1 of CF-II by the corresponding $\check{\mathbf{z}}_i$. Due to extra effects from the generated regressors, the 2SLS estimator is not efficient – it is actually the same as the CF estimator. The optimal W is Ω_1^{-1} and the resulting Σ_1 is $\left(G_1' \Omega_1^{-1} G_1\right)^{-1}$. In simulations, we use the optimal W.

Inconsistency of CH and KST Estimators of β when q is Endogenous

To discuss these two methods, we first assume that there is no threshold effect in the conditional variance.

Assumption H: $\delta_\sigma = 0$.

As mentioned in the Introduction, this simplification produces the setup in CH and KST, and we absorb σ_2

in u under Assumption H. Section 3.3 of CH estimates β based on the moment conditions

$$\mathbb{E}\left[\mathbf{z}_i u_i 1(q_i \leq \gamma_0)\right] = \mathbf{0} \text{ and } \mathbb{E}\left[\mathbf{z}_i u_i 1(q_i > \gamma_0)\right] = \mathbf{0};$$

this method is labeled CH's GMM. When q is exogenous, $\mathbb{E}\left[u_i|\mathbf{z}_i\right] = 0$ implies these moment conditions. However, when q is endogenous, say in the setup of CF-I,

$$\mathbb{E}\left[\mathbf{z}_i u_i 1(q_i \leq \gamma_0)\right] = \mathbb{E}\left[\mathbf{z}_i\left(\kappa v_{qi}\right) 1(\pi'\mathbf{z}_i + v_{qi} \leq \gamma_0)\right] \neq \mathbf{0};$$

similarly in the setup of CF-II and for the moments $\mathbb{E}\left[\mathbf{z}_i u_i 1(q_i > \gamma_0)\right]$. So the estimator of β based on these moment conditions is not consistent even if the estimator of γ were consistent.

In the KST setup where $\mathbf{x} = x$ and $v_x \perp 1(q \leq \gamma_0)|\mathcal{F}_{i-1}$, if their $\widehat{\gamma}$ estimator were consistent, then their estimator of β would be consistent. This is because their estimator uses the moment conditions

$$\mathbb{E}\left[\breve{\mathbf{z}}_i e^*_{1i} 1(q_i \leq \gamma_0)\right] = \mathbf{0} \text{ and } \mathbb{E}\left[\breve{\mathbf{z}}_i e^*_{2i} 1(q_i > \gamma_0)\right] = \mathbf{0},$$

where

$$\breve{\mathbf{z}}_i = \left(\mathbf{z}'_i, \lambda_1\left(\gamma_0 - \pi'\mathbf{z}\right), \lambda_2\left(\gamma_0 - \pi'\mathbf{z}\right)\right)',$$

and

$$\begin{aligned} e^*_{1i} &= y_i - \beta'_1 g_{xi} - \kappa \cdot \lambda_1\left(\gamma_0 - \pi'\mathbf{z}_i\right), \\ e^*_{2i} &= y_i - \beta'_2 g_{xi} - \kappa \cdot \lambda_2\left(\gamma_0 - \pi'\mathbf{z}_i\right). \end{aligned}$$

This works because $\mathbb{E}\left[e^*_{1i}|\mathbf{z}_i, q_i \leq \gamma_0\right] = 0$ and $\mathbb{E}\left[e^*_{2i}|\mathbf{z}_i, q_i > \gamma_0\right] = 0$ as shown in KST, so any function of $\mathbf{z}_i$, say, $h\left(\mathbf{z}_i\right)$, in each regime can serve as instruments. Using $\lambda_1\left(\gamma_0 - \pi'\mathbf{z}_i\right)$ and $\lambda_2\left(\gamma_0 - \pi'\mathbf{z}_i\right)$ as instruments is natural because they appear as the control functions in the STR estimator. Of course, if $\widehat{\gamma}$ is not consistent as shown in Section 2, then these moment conditions need not hold when γ_0 is replaced by $\text{plim}_{n\to\infty}\widehat{\gamma}$ and the resulting estimator of β need not be consistent. If we use our CF estimators $\widehat{\gamma}$, and estimate β based on $\mathbb{E}\left[h\left(\mathbf{z}_i\right) e^*_{1i} 1(q_i \leq \gamma_0)\right] = \mathbf{0}$ and $\mathbb{E}\left[h\left(\mathbf{z}_i\right) e^*_{2i} 1(q_i > \gamma_0)\right] = \mathbf{0}$, then we require some critical assumptions. First, $\dim(h\left(\mathbf{z}_i\right)) \geq d + 1$. Second, $q \notin \mathbf{x}$. This assumption can be relaxed; see the discussion around equation (7). Third, $v_q \sim \mathcal{N}\left(0, 1\right)$. This assumption is quite strong; see footnote 1. Fourth, $v_x \perp 1(q \leq \gamma_0)|\mathcal{F}_{i-1}$. This assumption is too strong to hold in practice unless x is exogenous and such that $v_x = \mathbf{0}$. In summary, estimation of β based on these moment conditions seems unattractive.

GMM-1 Estimators

We first state and prove the asymptotic properties of the GMM-1 estimators of β.

Theorem 6 *Under Assumption I (for CF-I $\widehat{\gamma}$) or Assumption II (for CF-II $\widehat{\gamma}$), and Assumption H,*

$$n^{1/2}\left(\widehat{\beta} - \beta\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \Sigma\right),$$

where $\Sigma = \left(G'WG\right)^{-1}\left(G'W\Omega WG\right)\left(G'WG\right)^{-1}$ *with* $G = \left(\mathbb{E}\left[\mathbf{z}_i\mathbf{x}'_{i,\leq\gamma_0}\right], \mathbb{E}\left[\mathbf{z}_i\mathbf{x}'_{i,>\gamma_0}\right]\right)$, $\Omega = \mathbb{E}\left[\mathbf{z}_i\mathbf{z}'_i u_i^2\right]$, *and* $G'WG > 0$. *When* $W = \Omega^{-1}$, Σ *reduces to* $\left(G'\Omega^{-1}G\right)^{-1}$.

Proof of Theorem 6. Note that

$$n^{1/2}\left(\widehat{\beta}-\beta\right)=\left[\left(\frac{1}{n}X^{*\prime}Z\right)W_n\left(\frac{1}{n}Z'X^*\right)\right]^{-1}\left[\left(\frac{1}{n}X^{*\prime}Z\right)W_n\frac{1}{\sqrt{n}}Z'\left(\mathbf{u}+\left(X_0-X_{\widehat{\gamma}}\right)\delta_n\right)\right],$$

where X_γ stacks $\mathbf{x}'_{i,\leq\gamma}$, and $X_0 = X_{\gamma_0}$. By Lemma 1 of Hansen (1996), the consistency of $\widehat{\gamma}$ and the continuity of $G_\gamma = \left(\mathbb{E}\left[\mathbf{z}_i\mathbf{x}'_{i,\leq\gamma}\right],\mathbb{E}\left[\mathbf{z}_i\mathbf{x}'_{i,>\gamma}\right]\right)$ in γ, we can show $\frac{1}{n}Z'X^* \xrightarrow{p} G$. Next, by Lemma A.10 of Hansen (2000), uniformly on $v\in[-\overline{v},\overline{v}]$,

$$n^{-2\alpha}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{x}_i'1\left(q_i\leq\gamma_0+v/a_n\right)-1\left(q_i\leq\gamma_0\right)=O_p(1).$$

Since $a_n\left(\widehat{\gamma}-\gamma_0\right)=O_p(1)$, we have

$$\begin{aligned}\frac{1}{\sqrt{n}}Z'\left(X_0-X_{\widehat{\gamma}}\right)\delta_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{x}_i'\left(1\left(q_i\leq\gamma_0\right)-1\left(q_i\leq\widehat{\gamma}\right)\right)cn^{-\alpha}=n^{-1/2}O_p\left(n^{2\alpha}\right)n^{-\alpha}\\ &= O_p(a_n^{-1/2})=o_p(1).\end{aligned}$$

Since $\frac{1}{\sqrt{n}}Z'\mathbf{u}\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\Omega\right)$, by Slutsky's theorem the result in the theorem follows. ■

This result follows from standard GMM asymptotics. It is easy to obtain the asymptotic variance of $\widehat{\beta}_1$, $\widehat{\beta}_2$ and $\widehat{\delta}_n$ as $(I_d,\mathbf{0})\Sigma(I_d,\mathbf{0})'$, $(\mathbf{0},I_d)\Sigma(\mathbf{0},I_d)'$ and $(I_d,-I_d)\Sigma(I_d,-I_d)'$ for these three estimates, respectively. Different from CH's GMM where $\widehat{\beta}_1(\widehat{\beta}_2)$ uses only information in the data with $q_i\leq\widehat{\gamma}$ $(q_i>\widehat{\gamma})$, the GMM-1 estimates $\widehat{\beta}_1$ and $\widehat{\beta}_2$ use information in all data points. As a result, $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are not asymptotically independent; this is similar to the CF estimators of β. Estimation of the asymptotic variance matrix by its sample analog is a standard econometric exercise and is omitted here.

We next discuss why the GMM-1 estimates are hard to extend to the $\delta_\sigma\neq 0$ case. In this case, $\mathbb{E}\left[\mathbf{z}_iu_i\right]=\mathbf{0}$ can be written as

$$\mathbb{E}\left[\mathbf{z}_i\left(\frac{y_i-\mathbf{x}_i'\beta_1}{\sigma_1}1(q_i\leq\gamma_0)+\frac{y_i-\mathbf{x}_i'\beta_2}{\sigma_2}1(q_i>\gamma_0)\right)\right]=\mathbf{0}.$$

Obviously, σ_1 and σ_2 cannot be identified separately. Instead, define $\varrho_\sigma=\sigma_1/\sigma_2$ and then

$$\mathbb{E}\left[\mathbf{z}_i\left(\frac{y_i-\mathbf{x}_i'\beta_1}{\varrho_\sigma}1(q_i\leq\gamma_0)+\left(y_i-\mathbf{x}_i'\beta_2\right)1(q_i>\gamma_0)\right)\right]=\mathbf{0}.$$

The moment conditions are nonlinear in ϱ_σ. We can first estimate ϱ_σ and then estimate β, or estimate ϱ_σ and β jointly. In CF-I, it seems hard to estimate ϱ_σ. The square root of $1/\widehat{\phi}=\widehat{\sigma}_1^2/\widehat{\sigma}_2^2$ in Section 3.4 estimates the ratio $\mathbb{E}\left[e_1^2\right]/\mathbb{E}\left[e_2^2\right]=\mathbb{E}\left[\left(\beta_1'\mathbf{v}+\sigma_1u-\kappa_1v_q\right)^2\right]/\mathbb{E}\left[\left(\beta_2'\mathbf{v}+\sigma_2u-\kappa_2v_q\right)^2\right]$ which is generally different from σ_1^2/σ_2^2.[18] In CF-II, ϱ_σ can be estimated by the ratio of any component of $\widehat{\kappa}_1$ and the corresponding component of $\widehat{\kappa}_2$ given that $\kappa_1=\sigma_1\psi$ and $\kappa_2=\sigma_2\psi$.[19] Given $\widehat{\varrho}_\sigma$, the moment conditions are linear in β so $\widehat{\beta}$ can be easily solved out. However, the randomness in $\widehat{\varrho}_\sigma$ will affect the asymptotic variance of $\widehat{\beta}$ in a complicated way. In joint estimation of ϱ_σ and β, we can concentrate on ϱ_σ and then grid search over ϱ_σ, but estimation requires $2d+1$ instruments. In sum, GMM estimation of β when $\delta_\sigma\neq 0$ involves several complications and seems messy for practical work.

[18] If $\varphi=\mathbf{0}$ and $\kappa\neq 0$, then $\kappa_1/\kappa_2=\sigma_1/\sigma_2$, but this case is very special.

[19] ϱ_σ can also be estimated by the square root of $\widehat{\sigma}_1^2/\widehat{\sigma}_2^2$ in Section 3.4 because $\widehat{\sigma}_1^2/\widehat{\sigma}_2^2$ estimates the ratio $E\left[e_1^2\right]/E\left[e_2^2\right]=E\left[\sigma_1^2\left(u-\psi'\mathbf{v}\right)^2\right]/E\left[\sigma_2^2\left(u-\psi'\mathbf{v}\right)^2\right]=\sigma_1^2/\sigma_2^2$ under the homoskedasticity assumption $E\left[e_1^2|q\right]=E\left[e_1^2\right]$ when $q\leq\gamma_0$ and $E\left[e_2^2|q\right]=E\left[e_2^2\right]$ when $q>\gamma_0$.

Finally, we detail the asymptotic variances of GMM-1 under DGP2 in our simulations of Section 5.3. Now, the 2SLS estimator is efficient, and the asymptotic variance can be simplified as

$$\mathbb{E}\left[u^2\right]\left(G'WG\right)^{-1}$$

with $\mathbb{E}\left[u^2\right]=\left(\psi_x\varphi+\psi_q\right)^2+\psi_x^2+1$, $G=\mathbb{E}\left[\mathbf{z}_i x_{i,\leq\gamma_0}\right]$ and $W=\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}$.

SD.4 Simplified Asymptotic Theory

We first decompose $\breve{\mathbf{x}}_i$ into $(\breve{\mathbf{x}}_{1i}',\breve{\mathbf{x}}_{2i}')'$, where the coefficients of $\breve{\mathbf{x}}_{2i}$ remain the same across the two regimes while the coefficients of $\breve{\mathbf{x}}_{1i}$ change. Correspondingly, write $\widehat{\breve{\mathbf{x}}}_i$, $\overline{\beta}_\ell$ and $\overline{c}$ as $\left(\widehat{\breve{\mathbf{x}}}_{1i}',\widehat{\breve{\mathbf{x}}}_{2i}'\right)'$, $\left(\overline{\beta}_{\ell 1}',\overline{\beta}_c'\right)$ and $(\overline{c}_1',\mathbf{0}')'$, and denote $D_{10}=\mathbb{E}\left[\breve{\mathbf{x}}_{1i}\breve{\mathbf{x}}_{1i}'|q_i=\gamma_0\right]$, $V_{10}^-=\mathbb{E}\left[\breve{\mathbf{x}}_{1i}\breve{\mathbf{x}}_{1i}'e_{1i}^2|q_i=\gamma-\right]$, $V_{10}^+=\mathbb{E}\left[\breve{\mathbf{x}}_{1i}\breve{\mathbf{x}}_{1i}'e_{2i}^2|q_i=\gamma_0+\right]$. Then in Theorem 1 and Corollary 1, we need only replace ω and ϕ by $\omega_1=\frac{\overline{c}_1'V_{10}^-\overline{c}_1}{\left(\overline{c}_1'D_{10}\overline{c}_1\right)^2 f}$ and $\phi_1=\overline{c}_1'V_{10}^+\overline{c}_1/\overline{c}_1'V_{10}^-\overline{c}_1$, respectively.

For $\overline{\beta}_\ell$ estimators, because part of $\overline{\beta}_1$ and $\overline{\beta}_2$ are the same, we can employ such information to improve the efficiency of $\overline{\beta}_\ell$ estimators. Consider GMM-2 estimators first because the GMM-II2 estimator is the most efficient and will be used in practice. Now, the moment conditions are

$$\mathbb{E}\left[\begin{array}{c}\breve{\mathbf{z}}_i\left(y_i-\breve{\mathbf{x}}_{1i}'\overline{\beta}_{11}-\breve{\mathbf{x}}_{2i}'\overline{\beta}_c\right)1(q_i\leq\gamma_0)\\ \breve{\mathbf{z}}_i\left(y_i-\breve{\mathbf{x}}_{1i}'\overline{\beta}_{21}-\breve{\mathbf{x}}_{2i}'\overline{\beta}_c\right)1(q_i>\gamma_0)\end{array}\right]=\mathbf{0} \tag{42}$$

and we estimate $\overline{\beta}:=\left(\overline{\beta}_{11}',\overline{\beta}_{21}',\overline{\beta}_c\right)$ jointly by minimizing

$$n\overline{m}_n\left(\overline{\beta}\right)'W_n\overline{m}_n\left(\overline{\beta}\right),$$

to have

$$\widehat{\overline{\beta}}=\left(\widehat{G}'W_n\widehat{G}\right)^{-1}\widehat{G}'W_n\frac{1}{n}\sum_{i=1}^n\left(\begin{array}{c}\breve{\mathbf{z}}_i y_i 1(q_i\leq\widehat{\gamma})\\ \breve{\mathbf{z}}_i y_i 1(q_i>\widehat{\gamma})\end{array}\right),$$

where $W_n\xrightarrow{p}W>0$, $\widehat{G}=\frac{1}{n}\sum_{i=1}^n\left(\begin{array}{ccc}\breve{\mathbf{z}}_i\widehat{\breve{\mathbf{x}}}_{1i}'1(q_i\leq\widehat{\gamma}) & \mathbf{0} & \breve{\mathbf{z}}_i\widehat{\breve{\mathbf{x}}}_{2i}'1(q_i\leq\widehat{\gamma})\\ \mathbf{0} & \breve{\mathbf{z}}_i\widehat{\breve{\mathbf{x}}}_{1i}'1(q_i>\widehat{\gamma}) & \breve{\mathbf{z}}_i\widehat{\breve{\mathbf{x}}}_{2i}'1(q_i>\widehat{\gamma})\end{array}\right)=\left(\begin{array}{c}\widehat{Z}_1'\left(\widehat{X}_1,\mathbf{0},\widehat{X}_2\right)\\ \widehat{Z}_2'\left(\mathbf{0},\widehat{X}_1,\widehat{X}_2\right)\end{array}\right)$ with $\widehat{X}_\ell$ stacking $\widehat{\breve{\mathbf{x}}}_{\ell i}'$, and

$$\overline{m}_n\left(\overline{\beta}\right)=\frac{1}{n}\sum_{i=1}^n\left(\begin{array}{c}\breve{\mathbf{z}}_i\left(y_i-\widehat{\breve{\mathbf{x}}}_{1i}'\overline{\beta}_{11}-\widehat{\breve{\mathbf{x}}}_{2i}'\overline{\beta}_c\right)1(q_i\leq\widehat{\gamma})\\ \breve{\mathbf{z}}_i\left(y_i-\widehat{\breve{\mathbf{x}}}_{1i}'\overline{\beta}_{21}-\widehat{\breve{\mathbf{x}}}_{2i}'\overline{\beta}_c\right)1(q_i>\widehat{\gamma})\end{array}\right).$$

For identification, we need only $2\dim(\breve{\mathbf{z}}_i)\geq 2\dim(\breve{\mathbf{x}}_{1i})+\dim(\breve{\mathbf{x}}_{2i})$ rather than $\dim(\breve{\mathbf{z}}_i)\geq\dim(\breve{\mathbf{x}}_i)$ (or equivalently, $\dim(\mathbf{z}_i)\geq\dim(\mathbf{x}_i)$).

First consider GMM-II2. Note that δ_σ is either zero or not, so $\breve{\mathbf{x}}_{1i}$ will either include the whole $\mathbf{v}_i$ or not at all. First suppose $\check{\mathbf{x}}_{1i}$ include $\mathbf{v}_i$. Then by the analysis in the proof of Theorem 2,

$$
\begin{aligned}
& n^{1/2}\left(\widehat{\overline{\beta}}-\overline{\beta}\right) \\
&= \left(\widehat{G}'W_n\widehat{G}\right)^{-1}\widehat{G}'W_n\tfrac{1}{\sqrt{n}}\begin{pmatrix} \widehat{Z}_1'\left(\check{X}_2\overline{\beta}_c+\check{X}_1\overline{\beta}_{21}+\check{X}_{10}\overline{c}_1n^{-\alpha}-\widehat{X}_1\overline{\beta}_{11}-\widehat{X}_2\overline{\beta}_c+\mathbf{e}^0\right) \\ \widehat{Z}_2'\left(\check{X}_2\overline{\beta}_c+\check{X}_1\overline{\beta}_{11}-\check{X}_{10}\overline{c}_1n^{-\alpha}-\widehat{X}_1\overline{\beta}_{21}-\widehat{X}_2\overline{\beta}_c+\mathbf{e}^0\right) \end{pmatrix} \\
&= \left(\widehat{G}'W_n\widehat{G}\right)^{-1}\left[\widehat{G}'W_n\tfrac{1}{\sqrt{n}}\begin{pmatrix} \widehat{Z}_1'\left[\left(\check{X}_2-\widehat{X}_2\right)\overline{\beta}_c+\left(\check{X}_1-\widehat{X}_1\right)\overline{\beta}_{11}+\mathbf{e}^0\right] \\ \widehat{Z}_2'\left[\left(\check{X}_2-\widehat{X}_2\right)\overline{\beta}_c+\left(\check{X}_1-\widehat{X}_1\right)\overline{\beta}_{21}+\mathbf{e}^0\right] \end{pmatrix}\right]+o_p(1) \\
&= \left(G'WG\right)^{-1}G'W\tfrac{1}{\sqrt{n}}\begin{pmatrix} \widehat{Z}_1'\left(\widehat{\mathbf{r}}\kappa_1+\mathbf{e}^0\right) \\ \widehat{Z}_2'\left(\widehat{\mathbf{r}}\kappa_2+\mathbf{e}^0\right) \end{pmatrix}+o_p(1)+o_p(1) \\
&\xrightarrow{d} \left[G'WG\right]^{-1}G'W\times\mathcal{N}\left(\mathbf{0},\begin{matrix}\Omega_1 & \Omega_{12} \\ \Omega_{21} & \Omega_2\end{matrix}\right)=:\left(G'WG\right)^{-1}G'W\times\mathcal{N}\left(\mathbf{0},\Omega\right),
\end{aligned}
$$

where $\check{X}_\ell$ stacks $\check{\mathbf{x}}_{\ell i}'$, $\check{X}_{10}$ stacks $\check{\mathbf{x}}_{1i}'1(q_i \le \gamma_0)$, $G = \begin{pmatrix} \mathbb{E}\left[\check{\mathbf{z}}_i\check{\mathbf{x}}_{1i,\le\gamma_0}'\right] & \mathbf{0} & \mathbb{E}\left[\check{\mathbf{z}}_i\check{\mathbf{x}}_{2i,\le\gamma_0}'\right] \\ \mathbf{0} & \mathbb{E}\left[\check{\mathbf{z}}_i\check{\mathbf{x}}_{1i,>\gamma_0}'\right] & \mathbb{E}\left[\check{\mathbf{z}}_i\check{\mathbf{x}}_{2i,>\gamma_0}'\right] \end{pmatrix}$, $\check{X}_2-\widehat{X}_2=\mathbf{0}$, Ω_1, Ω_2 and Ω_{12} are defined in GMM-II2 of Theorem 2, and $\Omega_{21}=\Omega_{12}'$. Second, suppose $\check{\mathbf{x}}_{1i}$ does not include $\mathbf{v}_i$ (i.e., $\sigma_1=\sigma_2=\sigma$). Now,

$$
\begin{aligned}
& n^{1/2}\left(\widehat{\overline{\beta}}-\overline{\beta}\right) \\
&= \left[\widehat{G}'W_n\widehat{G}\right]^{-1}\left[\widehat{G}'W_n\tfrac{1}{\sqrt{n}}\begin{pmatrix} \widehat{Z}_1'\left[\left(\check{X}_2-\widehat{X}_2\right)\overline{\beta}_c+\left(\check{X}_1-\widehat{X}_1\right)\overline{\beta}_{11}+\mathbf{e}^0\right] \\ \widehat{Z}_2'\left[\left(\check{X}_2-\widehat{X}_2\right)\overline{\beta}_c+\left(\check{X}_1-\widehat{X}_1\right)\overline{\beta}_{21}+\mathbf{e}^0\right] \end{pmatrix}\right]+o_p(1) \\
&= \left[G'WG\right]^{-1}G'W\tfrac{1}{\sqrt{n}}\begin{pmatrix} \widehat{Z}_1'\left(\widehat{\mathbf{r}}\kappa+\mathbf{e}^0\right) \\ \widehat{Z}_2'\left(\widehat{\mathbf{r}}\kappa+\mathbf{e}^0\right) \end{pmatrix}+o_p(1) \\
&\xrightarrow{d} \left[G'WG\right]^{-1}G'W\times\mathcal{N}\left(\mathbf{0},\begin{matrix}\Omega_1 & \Omega_{12} \\ \Omega_{21} & \Omega_2\end{matrix}\right)=:\left(G'WG\right)^{-1}G'W\times\mathcal{N}\left(\mathbf{0},\Omega\right),
\end{aligned}
$$

where $\check{X}_1-\widehat{X}_1=\mathbf{0}$, and $\kappa=\sigma\psi$ is the common κ_1 and κ_2, so in Ω_1,Ω_2 and Ω_{12}, replace both κ_1 and κ_2 by κ and note that $e_{1i}=e_{2i}$. In practice, we can set $W_n=\text{diag}\left(\widehat{Z}_1'\widehat{Z}_1,\widehat{Z}_2'\widehat{Z}_2\right)$ to get an initial estimator of $\overline{\beta}$, and then set W_n as a consistent estimator of Ω^{-1} to obtain the optimal estimator.

In GMM-I2, the analysis is similar. $\check{\mathbf{x}}_{1i}$ will either include v_q or not. In the former case, suppose $\check{\mathbf{x}}_{1i}'=(\mathbf{z}_i'\Pi_1,v_q)$ and $\check{\mathbf{x}}_{2i}'=\mathbf{z}_i'\Pi_2$, the error terms corresponding to $\mathbf{z}_i'\Pi_1$ and $\mathbf{z}_i'\Pi_2$ in the first stage are $\mathbf{v}_{1i}$ and $\mathbf{v}_{2i}$, and correspondingly, $\widehat{\mathbf{r}}$ is decomposed as $(\widehat{\mathbf{r}}_1',\widehat{\mathbf{r}}_2')'$. We need only replace $\widehat{\mathbf{r}}\kappa_\ell$ by $-\widehat{\mathbf{r}}_2\beta_c-\widehat{\mathbf{r}}_1\overline{\beta}_{\ell 1}+\widehat{\mathbf{r}}_q\kappa_\ell=:\widehat{\mathbf{r}}\overline{\kappa}_\ell$. With this new definition of $\overline{\kappa}_\ell$, the asymptotic distribution takes the same form as in GMM-II2 with Ω_1, Ω_2 and Ω_{12} defined in GMM-I2 of Theorem 2. In the latter case, suppose $\check{\mathbf{x}}_{1i}'=\mathbf{z}_i'\Pi_1$ and $\check{\mathbf{x}}_{2i}'=(\mathbf{z}_i'\Pi_2,v_q)$. Then we need only replace $\widehat{\mathbf{r}}\kappa_\ell$ by $(-\widehat{\mathbf{r}}_2,\widehat{\mathbf{r}}_q)\beta_c-\widehat{\mathbf{r}}_1\overline{\beta}_{\ell 1}=:\widehat{\mathbf{r}}\overline{\kappa}_\ell$, where the last component of β_c is κ - the common κ_1 and κ_2. With this new definition of $\overline{\kappa}_\ell$, the remaining specification of the asymptotic distribution is the same as in the former case.

The analyses for CF-I and CF-II estimators are parallel to those for GMM-I2 and GMM-II2 estimators. Since

$$\widehat{\overline{\beta}}=\left(\widehat{X}'\widehat{X}\right)^{-1}\widehat{X}'Y,$$

where $\widehat{X}=\left(\widehat{X}_{1,\le\widehat{\gamma}},\widehat{X}_{1,>\widehat{\gamma}},\widehat{X}_2\right)$ with $\widehat{X}_{1,\le\widehat{\gamma}}$ stacking $\widehat{\check{\mathbf{x}}}_{1i}'1(q_i\le\widehat{\gamma})$ and $\widehat{X}_{1,>\widehat{\gamma}}$ stacking $\widehat{\check{\mathbf{x}}}_{1i}'1(q_i>\widehat{\gamma})$, the moment conditions corresponding to (42) are

$$\mathbb{E}\left[\begin{pmatrix} \check{\mathbf{x}}_{1i,\le\gamma_0} \\ \check{\mathbf{x}}_{1i,>\gamma_0} \\ \check{\mathbf{x}}_{2i} \end{pmatrix}\left(y_i - \check{\mathbf{x}}'_{1i,\le\gamma_0}\overline{\beta}_{11} - \check{\mathbf{x}}'_{1i,>\gamma_0}\overline{\beta}_{21} - \check{\mathbf{x}}'_{2i}\overline{\beta}_c\right)\right] = \mathbf{0},$$

or equivalently,

$$\mathbb{E}\left[\begin{pmatrix} \check{\mathbf{x}}_{1i}\left(y_i - \check{\mathbf{x}}'_{1i}\overline{\beta}_{11} - \check{\mathbf{x}}'_{2i}\overline{\beta}_c\right)1(q_i \le \gamma_0) \\ \check{\mathbf{x}}_{1i}\left(y_i - \check{\mathbf{x}}'_{1i}\overline{\beta}_{21} - \check{\mathbf{x}}'_{2i}\overline{\beta}_c\right)1(q_i > \gamma_0) \\ \check{\mathbf{x}}_{2i}\left(y_i - \check{\mathbf{x}}'_{1i,\le\gamma_0}\overline{\beta}_{11} - \check{\mathbf{x}}'_{1i,>\gamma_0}\overline{\beta}_{21} - \check{\mathbf{x}}'_{2i}\overline{\beta}_c\right) \end{pmatrix}\right] = \mathbf{0},$$

which is weaker than

$$\mathbb{E}\left[\begin{pmatrix} \check{\mathbf{x}}_{i}\left(y_i - \check{\mathbf{x}}'_{1i}\overline{\beta}_{11} - \check{\mathbf{x}}'_{2i}\overline{\beta}_c\right)1(q_i \le \gamma_0) \\ \check{\mathbf{x}}_{i}\left(y_i - \check{\mathbf{x}}'_{1i}\overline{\beta}_{21} - \check{\mathbf{x}}'_{2i}\overline{\beta}_c\right)1(q_i > \gamma_0) \end{pmatrix}\right] = \mathbf{0}.$$

Now,

$$\begin{aligned}
& n^{1/2}\left(\widehat{\overline{\beta}} - \overline{\beta}\right) \\
&= \left(\tfrac{1}{n}\widehat{X}'\widehat{X}\right)^{-1}\tfrac{1}{\sqrt{n}}\begin{pmatrix} \widehat{X}'_{1,\le\widehat{\gamma}}\left(\check{X}_2\overline{\beta}_c + \check{X}_1\overline{\beta}_{21} + \check{X}_{10}\overline{c}_1 n^{-\alpha} - \widehat{X}_{1,\le\widehat{\gamma}}\overline{\beta}_{11} - \widehat{X}_{1,>\widehat{\gamma}}\overline{\beta}_{21} - \widehat{X}_2\overline{\beta}_c + \mathbf{e}^0\right) \\ \widehat{X}'_{1,>\widehat{\gamma}}\left(\check{X}_2\overline{\beta}_c + \check{X}_1\overline{\beta}_{11} - \check{X}_{10}\overline{c}_1 n^{-\alpha} - \widehat{X}_{1,\le\widehat{\gamma}}\overline{\beta}_{11} - \widehat{X}_{1,>\widehat{\gamma}}\overline{\beta}_{21} - \widehat{X}_2\overline{\beta}_c + \mathbf{e}^0\right) \\ \widehat{X}'_2\left(\check{X}_2\overline{\beta}_c + \check{X}_1\overline{\beta}_{21} + \check{X}_{10}\overline{c}_1 n^{-\alpha} - \widehat{X}_1\overline{\beta}_{21} - \widehat{X}_{1,\le\widehat{\gamma}}\overline{c}_1 n^{-\alpha} - \widehat{X}_2\overline{\beta}_c + \mathbf{e}^0\right) \end{pmatrix} \\
&= \overline{M}^{-1}\tfrac{1}{\sqrt{n}}\begin{pmatrix} \widehat{X}'_{1,\le\widehat{\gamma}}\left(\left(\check{X}_2 - \widehat{X}_2\right)\overline{\beta}_c + \left(\check{X}_1 - \widehat{X}_1\right)\overline{\beta}_{11} + \mathbf{e}^0\right) \\ \widehat{X}'_{1,>\widehat{\gamma}}\left(\left(\check{X}_2 - \widehat{X}_2\right)\overline{\beta}_c + \left(\check{X}_1 - \widehat{X}_1\right)\overline{\beta}_{21} + \mathbf{e}^0\right) \\ \widehat{X}'_2\left(\left(\check{X}_2 - \widehat{X}_2\right)\overline{\beta}_c + \left(\check{X}_1 - \widehat{X}_1\right)\overline{\beta}_{21} + \mathbf{e}^0\right) \end{pmatrix} + o_p(1),
\end{aligned}$$

where $\overline{M} = \mathbb{E}\left[\left(\check{\mathbf{x}}'_{1i,\le\gamma_0}, \check{\mathbf{x}}'_{1i,>\gamma_0}, \check{\mathbf{x}}'_{2i}\right)'\left(\check{\mathbf{x}}'_{1i,\le\gamma_0}, \check{\mathbf{x}}'_{1i,>\gamma_0}, \check{\mathbf{x}}'_{2i}\right)\right]$, and in the third term,

$$\begin{aligned}
& \frac{1}{\sqrt{n}}\widehat{X}'_2\left(\check{X}_{10} - \widehat{X}_{1,\le\widehat{\gamma}}\right)\overline{c}_1 n^{-\alpha} \\
= \;& \frac{1}{\sqrt{n}}\widehat{X}'_2\left(\check{X}_{10} - \check{X}_{1,\le\widehat{\gamma}}\right)\overline{c}_1 n^{-\alpha} + \frac{1}{\sqrt{n}}\widehat{X}'_2\left(\check{X}_{1,\le\widehat{\gamma}} - \widehat{X}_{1,\le\widehat{\gamma}}\right)\overline{c}_1 n^{-\alpha} \\
= \;& o_p(1) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{\check{\mathbf{x}}}_{2i}\left(\check{\mathbf{x}}'_{1i} - \widehat{\check{\mathbf{x}}}'_{1i}\right)1\left(q_i \le \widehat{\gamma}\right)\overline{c}_1 n^{-\alpha} \\
= \;& o_p(1) + O_p\left(n^{-\alpha}\right) = o_p(1),
\end{aligned}$$

which will not disappear in the fixed-threshold-effect framework of Chan (1993). Note also that by re-expressing $\check{X}_2\overline{\beta}_c + \check{X}_1\overline{\beta}_{21} + \check{X}_{10}\overline{c}_1 n^{-\alpha} - \widehat{X}_1\overline{\beta}_{21} - \widehat{X}_{1,\le\widehat{\gamma}}\overline{c}_1 n^{-\alpha} - \widehat{X}_2\overline{\beta}_c$ as $\check{X}_2\overline{\beta}_c + \check{X}_1\overline{\beta}_{11} - \check{X}_{10}\overline{c}_1 n^{-\alpha} - \widehat{X}_1\overline{\beta}_{11} + \widehat{X}_{1,\le\widehat{\gamma}}\overline{c}_1 n^{-\alpha} - \widehat{X}_2\overline{\beta}_c$, the components $\left(\check{X}_1 - \widehat{X}_1\right)\overline{\beta}_{21}$ in the third term can be replaced by $\left(\check{X}_1 - \widehat{X}_1\right)\overline{\beta}_{11}$; this will not affect the asymptotic distribution because $\overline{\beta}_{11} - \overline{\beta}_{21} = \overline{c}_1 n^{-\alpha} = o(1)$. In other words, in the statements of all theorems in Section 4, κ_1 and κ_2 (or $\overline{\kappa}_1$ and $\overline{\kappa}_2$) can be exchanged with each other.

First consider CF-II. If $\check{\mathbf{x}}_{1i}$ include $\mathbf{v}_i$, then

$$n^{1/2}\left(\widehat{\overline{\beta}} - \overline{\beta}\right) = \overline{M}^{-1}\frac{1}{\sqrt{n}}\begin{pmatrix} \widehat{X}'_{1,\le\widehat{\gamma}}\left(\widehat{\mathbf{r}}\kappa_1 + \mathbf{e}^0\right) \\ \widehat{X}'_{1,>\widehat{\gamma}}\left(\widehat{\mathbf{r}}\kappa_2 + \mathbf{e}^0\right) \\ \widehat{X}'_2\left(\widehat{\mathbf{r}}\kappa_2 + \mathbf{e}^0\right) \end{pmatrix} + o_p(1)$$

$$\xrightarrow{d} \overline{M}^{-1} \cdot \mathcal{N}\left(\mathbf{0}, \begin{array}{ccc} \Omega_{1,\le\gamma_0} & \mathbf{0} & \Omega_{12,\le\gamma_0} \\ \mathbf{0} & \Omega_{1,>\gamma_0} & \Omega_{12,>\gamma_0} \\ \Omega_{21,\le\gamma_0} & \Omega_{21,>\gamma_0} & \Omega_2 \end{array}\right) =: \overline{M}^{-1} \cdot \mathcal{N}(\mathbf{0}, \Omega),$$

where $\Omega_{1,\le\gamma_0}$ and $\Omega_{1,>\gamma_0}$ are the same as Ω_1 and Ω_2 in CF-II of Theorem 5 but replacing $\breve{\mathbf{x}}_i$ by $\breve{\mathbf{x}}_{1i}$,

$$\begin{aligned}
\Omega_2 &= \mathbb{E}\left[\breve{\mathbf{x}}_{2i}\breve{\mathbf{x}}_{2i}'\left(e_i^0\right)^2\right] + \mathbb{E}\left[\breve{\mathbf{x}}_{2i}\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\left(\kappa_2'\mathbf{v}_i\mathbf{v}_i'\kappa_2\right)\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\breve{\mathbf{x}}_{2i}'\right], \\
\Omega_{12,\le\gamma_0} &= \mathbb{E}\left[\breve{\mathbf{x}}_{1i,\le\gamma_0}\breve{\mathbf{x}}_{2i}'e_{1i}^2\right] + \mathbb{E}\left[\breve{\mathbf{x}}_{1i,\le\gamma_0}\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\left(\kappa_1'\mathbf{v}_i\mathbf{v}_i'\kappa_2\right)\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\breve{\mathbf{x}}_{2i}'\right], \\
\Omega_{12,>\gamma_0} &= \mathbb{E}\left[\breve{\mathbf{x}}_{1i,>\gamma_0}\breve{\mathbf{x}}_{2i}'e_{2i}^2\right] + \mathbb{E}\left[\breve{\mathbf{x}}_{1i,>\gamma_0}\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\left(\kappa_2'\mathbf{v}_i\mathbf{v}_i'\kappa_2\right)\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\breve{\mathbf{x}}_{2i}'\right]
\end{aligned}$$

and $\Omega_{21,\le\gamma_0} = \Omega_{12,\le\gamma_0}'$, $\Omega_{21,>\gamma_0} = \Omega_{12,>\gamma_0}'$. If $\breve{\mathbf{x}}_{1i}$ does not include $\mathbf{v}_i$, then

$$n^{1/2}\left(\widehat{\overline{\beta}} - \overline{\beta}\right) = \overline{M}^{-1}\frac{1}{\sqrt{n}}\left(\begin{array}{c} \widehat{X}_{1,\le\widehat{\gamma}}'\left(\widehat{\mathbf{r}}\kappa + \mathbf{e}^0\right) \\ \widehat{X}_{1,>\widehat{\gamma}}'\left(\widehat{\mathbf{r}}\kappa + \mathbf{e}^0\right) \\ \widehat{X}_2'\left(\widehat{\mathbf{r}}\kappa + \mathbf{e}^0\right) \end{array}\right) + o_p(1),$$

so we need only replace κ_1 and κ_2 everywhere in Ω by κ and note that $e_i^0 = e_{1i} = e_{2i}$.

In CF-I, the analysis can combine those of GMM-I2 and CF-II. Specifically, the asymptotic distribution takes the same form as in CF-II with only Ω redefined: $\Omega_{1,\le\gamma_0}$ and $\Omega_{1,>\gamma_0}$ take the same form as Ω_1 and Ω_2 in CF-I of Theorem 5 but replacing $\breve{\mathbf{x}}_i$ by $\breve{\mathbf{x}}_{1i}$,

$$\begin{aligned}
\Omega_2 &= \mathbb{E}\left[\left(\mathbb{E}\left[\breve{\mathbf{x}}_{2i}\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbf{z}_i'\mathbf{v}_i\overline{\kappa}_2 + \breve{\mathbf{x}}_{2i}e_i^0\right)\left(\mathbb{E}\left[\breve{\mathbf{x}}_{2i}\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbf{z}_i'\mathbf{v}_i\overline{\kappa}_2 + \breve{\mathbf{x}}_{2i}e_i^0\right)'\right], \\
\Omega_{12,\le\gamma_0} &= \mathbb{E}\left[\left(\mathbb{E}\left[\breve{\mathbf{x}}_{1i}\mathbf{z}_{i,\le\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbf{z}_i'\mathbf{v}_i\overline{\kappa}_1 + \breve{\mathbf{x}}_{1i,\le\gamma_0}e_{1i}\right)\left(\mathbb{E}\left[\breve{\mathbf{x}}_{2i}\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbf{z}_i'\mathbf{v}_i\overline{\kappa}_2 + \breve{\mathbf{x}}_{2i}e_i^0\right)'\right], \\
\Omega_{12,>\gamma_0} &= \mathbb{E}\left[\left(\mathbb{E}\left[\breve{\mathbf{x}}_{1i}\mathbf{z}_{i,>\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbf{z}_i'\mathbf{v}_i\overline{\kappa}_2 + \breve{\mathbf{x}}_{1i,>\gamma_0}e_{2i}\right)\left(\mathbb{E}\left[\breve{\mathbf{x}}_{2i}\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbf{z}_i'\mathbf{v}_i\overline{\kappa}_2 + \breve{\mathbf{x}}_{2i}e_i^0\right)'\right],
\end{aligned}$$

$\Omega_{21,\le\gamma_0} = \Omega_{12,\le\gamma_0}'$, $\Omega_{21,>\gamma_0} = \Omega_{12,>\gamma_0}'$, where $\overline{\kappa}_\ell$ is understood as the $\overline{\kappa}_\ell$ in GMM-I2. Note that the cross terms in $\Omega_{1,\le\gamma_0}$, $\Omega_{1,>\gamma_0}$, Ω_2, $\Omega_{12,\le\gamma_0}$ and $\Omega_{12,>\gamma_0}$ will not disappear.

We next check GMM-1. Now, we decompose $\mathbf{x}_i$ (rather than $\breve{\mathbf{x}}_i$) as $(\mathbf{x}_{1i}', \mathbf{x}_{2i}')'$, and the moment conditions are

$$\mathbb{E}\left[\mathbf{z}_i\left(y_i - \mathbf{x}_{1i,\le\gamma_0}'\overline{\beta}_{11} - \mathbf{x}_{1i,>\gamma_0}'\overline{\beta}_{21} - \mathbf{x}_{2i}'\overline{\beta}_c\right)\right] = \mathbf{0},$$

so

$$\widehat{\beta} = \left[\left(X^{*\prime}Z\right)W_n\left(Z'X^*\right)\right]^{-1}\left[\left(X^{*\prime}Z\right)W_n\left(Z'Y\right)\right],$$

where X^* stacks $\left(\mathbf{x}_{1i,\le\widehat{\gamma}}', \mathbf{x}_{1i,>\widehat{\gamma}}', \mathbf{x}_{2i}'\right)$, and Z stacks $\mathbf{z}_i'$. In the statement of the asymptotic distribution of $\widehat{\beta}$ in Theorem 6, we need only redefine $G = \left(\mathbb{E}\left[\mathbf{z}_i\mathbf{x}_{1i,\le\gamma_0}'\right], \mathbb{E}\left[\mathbf{z}_i\mathbf{x}_{1i,>\gamma_0}'\right], \mathbb{E}\left[\mathbf{z}_i\mathbf{x}_{2i}'\right]\right)$.

Finally, we specify the formulae above to the simulations in Section 5.3; only CF-II and GMM-II2 can be simplified where $\kappa_1 = \kappa_2 = \psi$. In CF-II,

$$n^{1/2}\left(\begin{array}{c} \widehat{\delta} - \delta \\ \widehat{\psi} - \psi \end{array}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma),$$

where

$$\Sigma = \overline{M}^{-1}\Omega\overline{M}^{-1} = \mathbb{E}\left[e_{ui}^2\right]\overline{M}^{-1} + \overline{M}^{-1}\Omega_2\overline{M}^{-1},$$

with

$$\begin{aligned}
\overline{M} &= \mathbb{E}\left[\left(x_{i,\le\gamma_0},\mathbf{v}_i'\right)'\left(x_{i,\le\gamma_0},\mathbf{v}_i'\right)\right],\\
\Omega &= \mathbb{E}\left[\left(\begin{array}{c} x_{i,\le\gamma_0}e_{ui}+\mathbb{E}\left[x_{i,\le\gamma_0}\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbf{z}_i\mathbf{v}_i'\psi \\ \mathbf{v}_ie_{ui}+\mathbb{E}\left[\mathbf{v}_i\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbf{z}_i\mathbf{v}_i'\psi \end{array}\right)\left(\begin{array}{c} x_{i,\le\gamma_0}e_{ui}+\mathbb{E}\left[x_{i,\le\gamma_0}\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbf{z}_i\mathbf{v}_i'\psi \\ \mathbf{v}_ie_{ui}+\mathbb{E}\left[\mathbf{v}_i\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbf{z}_i\mathbf{v}_i'\psi \end{array}\right)'\right]\\
&= \mathbb{E}\left[e_{ui}^2\right]\mathbb{E}\left[\left(\begin{array}{c} x_{i,\le\gamma_0} \\ \mathbf{v}_i \end{array}\right)\left(x_{i,\le\gamma_0},\mathbf{v}_i'\right)\right]+\mathbb{E}\left[\left(\psi'\mathbf{v}_i\right)^2\right]\mathbb{E}\left[\left(\begin{array}{c} x_{i,\le\gamma_0}\mathbf{z}_i' \\ \mathbf{v}_i\mathbf{z}_i' \end{array}\right)\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\left(\mathbf{z}_ix_{i,\le\gamma_0},\mathbf{z}_i\mathbf{v}_i'\right)\right]\\
&=: \Omega_1+\Omega_2.
\end{aligned}$$

In GMM-II2,

$$n^{1/2}\left(\begin{array}{c} \widehat{\delta}-\delta \\ \widehat{\psi}-\psi \end{array}\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\Sigma\right),$$

where $\Sigma=\left(G'WG\right)^{-1}\left(G'W\Omega WG\right)\left(G'WG\right)^{-1}$ with $G=\mathbb{E}\left[\begin{array}{cc} \breve{\mathbf{z}}_ix_{i,\le\gamma_0} & \breve{\mathbf{z}}_i\mathbf{v}_{i,\le\gamma_0}' \\ \mathbf{0} & \breve{\mathbf{z}}_i\mathbf{v}_{i,>\gamma_0}' \end{array}\right]$, and

$$\begin{aligned}
\Omega &= \mathbb{E}\left[\left(\begin{array}{c} \breve{\mathbf{z}}_{i,\le\gamma_0}e_{ui}+\mathbb{E}\left[\breve{\mathbf{z}}_{i,\le\gamma_0}\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbf{z}_i\mathbf{v}_i'\psi \\ \breve{\mathbf{z}}_{i,>\gamma_0}e_{ui}+\mathbb{E}\left[\breve{\mathbf{z}}_{i,>\gamma_0}\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbf{z}_i\mathbf{v}_i'\psi \end{array}\right)\left(\begin{array}{c} \breve{\mathbf{z}}_{i,\le\gamma_0}e_{ui}+\mathbb{E}\left[\breve{\mathbf{z}}_{i,\le\gamma_0}\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbf{z}_i\mathbf{v}_i'\psi \\ \breve{\mathbf{z}}_{i,>\gamma_0}e_{ui}+\mathbb{E}\left[\breve{\mathbf{z}}_{i,>\gamma_0}\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbf{z}_i\mathbf{v}_i'\psi \end{array}\right)'\right]\\
&= \mathbb{E}\left[e_{ui}^2\right]\mathbb{E}\left[\begin{array}{cc} \breve{\mathbf{z}}_i\breve{\mathbf{z}}_{i,\le\gamma_0}' & \mathbf{0} \\ \mathbf{0} & \breve{\mathbf{z}}_i\breve{\mathbf{z}}_{i,>\gamma_0}' \end{array}\right]+\mathbb{E}\left[\left(\psi'\mathbf{v}_i\right)^2\right]\mathbb{E}\left[\left(\begin{array}{c} \breve{\mathbf{z}}_{i,\le\gamma_0}\mathbf{z}_i' \\ \breve{\mathbf{z}}_{i,>\gamma_0}\mathbf{z}_i' \end{array}\right)\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\left(\mathbf{z}_i\breve{\mathbf{z}}_{i,\le\gamma_0}',\mathbf{z}_i\breve{\mathbf{z}}_{i,>\gamma_0}'\right)\right]\\
&=: \Omega_1+\Omega_2.
\end{aligned}$$

Note that, different from CF-II, $\Omega_1\neq G$, where G is not even square. In GMM-II2, the 2SLS estimator is still the same as the CF-II estimator, and we instead use the optimal $W=\Omega^{-1}$.

SD.5 Further Simulation Results

Performance of the IDKE

In this subsection, we report the risk of the IDKE of γ in YP under DGP1 and DGP2 where no instruments are used. We do not report these simulation results in the main text because it is not fair to compare the IDKE and the two CF estimators given that more data (i.e., $\mathbf{z}_i$'s) are used in the latter. The main theme of this paper is how to estimate γ and β when $\mathbf{z}$ is available and q is endogenous; deviating from this theme too far seems undesirable. Also, the performance of the IDKE has already been studied in YP and YLP.

First, recall that the IDKE of YP is defined as

$$\widehat{\gamma}=\underset{\gamma}{\operatorname{argmax}}\ \frac{1}{n}\sum_{i=1}^n\left[\frac{1}{n-1}\sum_{j=1,j\neq i}^n y_jK_{h,ij}^{\gamma-}-\frac{1}{n-1}\sum_{j=1,j\neq i}^n y_jK_{h,ij}^{\gamma+}\right]^2,$$

where

$$K_{h,ij}^{\gamma\pm}=k_h(x_j-x_i,x_i)\cdot k_h^{\pm}\left(q_j-\gamma\right),$$

with $k_h\left(\cdot,\cdot\right)$ and $k_h^{\pm}=\frac{1}{h}k_{\pm}\left(\frac{\cdot}{h}\right)$ being rescaled boundary kernels. Note that for a given γ, either $\frac{1}{n-1}\sum_{j=1,j\neq i}^n y_jK_{h,ij}^{\gamma-}$ or $\frac{1}{n-1}\sum_{j=1,j\neq i}^n y_jK_{h,ij}^{\gamma+}$ has $n-1$ summands since x_i falls in either the $q_i\le\gamma$ regime or $q_i>\gamma$ regime rather than both. For simplicity, we replace $\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}$ by $\frac{1}{n}\sum_{j=1}^{n}$. YP use $\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}$ to convert a V-statistic to a U-statistic, but this conversion is not important in practice. To avoid using the boundary kernel $k_h(\cdot,\cdot)$ (which is needed to judge whether x_i is near the boundary of x's support), we replace $k_h(\cdot,\cdot)$ by the rescaled normal density $\frac{1}{h}\phi\left(\frac{\cdot}{h}\right)$; also, we replace the compact-supported boundary kernel $k_+(\cdot)$ by the half normal density $2\phi(\cdot)1(\cdot>0)$ and set $k_-(\cdot)=k_+(-\cdot)$. Under DGP1, the IDKE reduces to the DKE because q is the only covariate, where the DKE of γ is defined as

$$\widehat{\gamma}=\underset{\gamma}{\operatorname{argmax}}\left[\frac{1}{n}\sum_{j=1}^{n}y_j k_h^+(q_j-\gamma)-\frac{1}{n}\sum_{j=1}^{n}y_j k_h^-(q_j-\gamma)\right]^2.$$

As to the bandwidth selection, we use the Matlab function kde.m of Botev et al. (2010) to choose the bandwidth h in DKE, and use their Matlab function kde2d.m to choose the bandwidths $\mathbf{h}=(h_1,h_2)'$ for $(x,q)'$ in IDKE. To check the robustness of our bandwidth selection, we also tried $h/2$ and $2h$ for the DKE and $\mathbf{h}/2$ and $2\mathbf{h}$ for IDKE.

Table 12 reports the MAD of the DKE of γ under DGP1. Note that because $\begin{pmatrix} v_{qi} \\ q_i \end{pmatrix}\sim\mathcal{N}\left(0,\begin{pmatrix}1&1\\1&2\end{pmatrix}\right)$,

$$\mathbb{E}\left[u_i|q_i\right]=\kappa\mathbb{E}\left[v_{qi}|q_i\right]=\frac{\kappa}{2}q_i$$

is a continuous function of q_i, so the DKE can be applied. The new error term is $e_i:=u_i-\frac{\kappa}{2}q_i=e_{ui}+\frac{\kappa}{2}v_{qi}+\frac{\kappa}{2}z_i$ whose variance is $\sigma^2=1+\frac{\kappa^2}{2}$. Based on the approximation of shrinking threshold effects in YLP, $\phi=1$ and $\omega=\frac{\sigma^2}{fc_\beta^2}$ in DKE, while for our CF estimators, $\phi=1$ and $\omega=\frac{1}{fc_\beta^2}$ which is smaller than that in DKE, so our CF estimators are expected to be more efficient. Table 13 reports the MAD of the IDKE of γ under DGP2. Note that because $\begin{pmatrix} v_{qi} \\ x_i \\ q_i \end{pmatrix}\sim\mathcal{N}\left(0,\begin{pmatrix}1&1&1\\1&3&2\\1&2&2\end{pmatrix}\right)$ and $\begin{pmatrix} e_{xi} \\ x_i \\ q_i \end{pmatrix}\sim\mathcal{N}\left(0,\begin{pmatrix}1&1&0\\1&3&2\\0&2&2\end{pmatrix}\right)$,

$$\begin{aligned}\mathbb{E}\left[u_i|x_i,q_i\right] &= \mathbb{E}\left[\psi_x v_{xi}+\psi_q v_{qi}|\varphi v_{qi}+e_{xi}+\Pi_x'\mathbf{z}_i, v_{qi}+\pi'\mathbf{z}_i\right]\\ &= 2\kappa\mathbb{E}\left[v_{qi}|v_{qi}+e_{xi}-z_i, v_{qi}-z_i\right]+\kappa\mathbb{E}\left[e_{xi}|x_i,q_i\right]\\ &= \kappa q_i+\kappa(x_i-q_i)=\kappa x_i\end{aligned}$$

is a continuous function of $(x_i,q_i)'$, so the IDKE can be applied. The new error term is $e_i:=u_i-\kappa x_i=e_{ui}+(2-\kappa)v_{qi}+(1-\kappa)e_{xi}+\kappa z_i$ whose variance is $\sigma^2=3(\kappa-1)^2+3$. Now, $\phi=1$ and $\omega=\frac{\sigma^2\mathbb{E}[x_i^2 f(x_i,0)^2 f^2(x_i)|q_i=0]}{fc_\beta^2\left(\mathbb{E}[x_i^2 f(x_i,0)f(x_i)|q_i=0]\right)^2}\approx\frac{1.8\sigma^2}{fc_\beta^2}$ in IDKE, while for our CF-II estimator, $\phi=1$ and $\omega=\frac{1}{fc_\beta^2\mathbb{E}\left[x_i^2|q_i=0\right]}=\frac{1}{fc_\beta^2}$ which is much smaller than that in IDKE, so our CF-II estimator is expected to be more efficient. In CF-I, $\phi\neq1$, so it is hard to compare with CF-II and IDKE.

Porter and Yu (2015) suggest a smaller (than optimal in density estimation) bandwidth in γ estimation, but it is not always the case in our simulation. Under DGP1, the bandwidth should be smaller when κ is larger (to induce a smaller bias) because we use the local constant estimator while the endogeneity increases the slope of q from 0 to $\frac{\kappa}{2}$. Under DGP2, the bandwidth for x should be smaller when δ and/or κ are larger (to induce a smaller bias) because the endogeneity increases the slope of x from δ to $\delta+\kappa$, while the bandwidth for q should be large (to increase efficiency) because its slope is zero even under endogeneity. As a result, we also tried the bandwidth $\mathbf{h}=(h_1/2,2h_2)$ in our simulation.

We next compare the risks of DKE and IDKE with our two CF approaches. Our comparison is based on the best performances of the DKE and IDKE among the few bandwidths (which are blacked in Tables 12 and

13). Comparing Table 12 and Table 1, we can see that the risk of DKE is between that of the two CFs and 2SLS in all cases. Also, endogeneity is harmful and larger δ is beneficial. When n increases from 200 to 800, the MAD decreases although need not decrease exactly in an order of 4 (as implied by the n consistency) because we did not use the optimal bandwidth (which is unknown) in our estimation. Comparing Table 13 and Table 4, we can see that the risk of IDKE is between that of CF-II and CF-I (also 2SLS) in all cases. When $n = 200$, weak and strong endogeneities induce similar risks, while when $n = 800$, strong endogeneity is indeed harmful. In all cases, larger δ and n are beneficial although the implied convergence rate (by comparing the risks of $n = 200$ and $n = 800$) is smaller than n.

$n \rightarrow$	200						800					
$\delta \rightarrow$	0.5		1		2		0.5		1		2	
$\kappa \rightarrow$	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ
h	**0.359**	0.714	**0.161**	0.650	**0.037**	0.553	**0.091**	0.390	0.020	0.150	0.009	0.070
$h/2$	0.424	**0.560**	0.194	**0.417**	0.081	**0.287**	0.149	**0.305**	0.028	**0.086**	0.011	**0.025**
$2h$	0.420	1.068	0.188	1.182	0.039	1.230	0.093	1.027	**0.020**	1.076	**0.008**	1.124

Table 12: MAD for the DKE of γ Under DGP1

$n \rightarrow$	200						800					
$\delta \rightarrow$	0.5		1		2		0.5		1		2	
$\kappa \rightarrow$	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ
$\mathbf{h}$	0.511	0.611	0.556	0.621	0.571	0.623	0.345	0.486	0.283	0.543	0.245	0.563
$\mathbf{h}/2$	0.562	0.695	0.752	0.794	0.913	0.794	0.451	0.569	0.582	0.734	0.688	0.806
$2\mathbf{h}$	**0.437**	**0.393**	**0.367**	**0.355**	**0.330**	**0.316**	0.317	**0.328**	0.234	**0.314**	0.188	**0.299**
$\left(\frac{h_1}{2}, 2h_2\right)$	0.437	0.542	0.455	0.532	0.424	0.501	**0.262**	0.374	**0.216**	0.400	**0.153**	0.387

Table 13: MAD for the IDKE of γ Under DGP2

Comparison When q is Exogenous

In this subsection, we compare the performance of the two CF approaches when q is exogenous. We use DGP2 but set $\pi = \mathbf{0}$, $\varphi = 0$, $\psi_q = 0$ and $z = q$. The formulae for ϕ and η^2 are the same as before. In this DGP, $\kappa_1 = \kappa_2 = 0$ in CF-I and $\kappa_1 = \kappa_2 = \psi_x$ in CF-II; since $v_q = 0$, $\breve{\mathbf{x}}_i$ would exclude v_q in both CF-I and CF-II, i.e.,

$$\breve{\mathbf{x}}_i = \mathbf{g}_i = \Pi_x' \mathbf{z}_i \text{ in CF-I and } \breve{\mathbf{x}}_i = (x_i, v_{xi})' \text{ in CF-II.}$$

In CF-I,

$$\begin{aligned} V_1 &= \mathbb{E}\left[g_{i,\leq\gamma_0}^2\right]\mathbb{E}\left[e_{1i}^2\right], \\ \Omega_2^1 &= \mathbb{E}\left[v_{xi}^2\right]\delta^2\mathbb{E}\left[g_i\mathbf{z}_{i,\leq\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i g_{i,\leq\gamma_0}\right], \\ \Omega_{21}^1 &= -\delta\mathbb{E}\left[v_{xi}e_{1i}\right]\mathbb{E}\left[g_i\mathbf{z}_{i,\leq\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i g_{i,\leq\gamma_0}\right], \end{aligned}$$

and in CF-II,

$$\begin{aligned} V_1 &= \mathbb{E}\left[\breve{\mathbf{x}}_i\breve{\mathbf{x}}_{i,\leq\gamma_0}'\right]\mathbb{E}\left[e_{1i}^2\right], \\ \Omega_2^1 &= \psi_x^2\mathbb{E}\left[v_{xi}^2\right]\mathbb{E}\left[\breve{\mathbf{x}}_{i,\leq\gamma_0}\mathbf{z}_i'\right]\mathbb{E}\left[\mathbf{z}_i\mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i\breve{\mathbf{x}}_{i,\leq\gamma_0}'\right]. \end{aligned}$$

We can still estimate δ and ψ_x jointly in CF-II and the resulting asymptotic variance matrix Σ takes the same formula as before except that $\mathbf{v}_i$ is replaced by v_{xi} and $\psi' \mathbf{v}_i$ is replaced by $\psi_x v_{xi}$. The asymptotic variance formula in GMM-1 is the same as before. In GMM-2,

$$\breve{\mathbf{z}}_i = \mathbf{z}_i \text{ in GMM-I2 and } \breve{\mathbf{z}}_i = \left(\mathbf{z}_i', x_i\right)' \text{ in GMM-II2.}$$

We estimate δ and ψ_x jointly in GMM-II2 and adjust the asymptotic variance matrix correspondingly.

Note that when q is exogenous, we can also estimate δ by CH's GMM which is based on the moment conditions

$$\mathbb{E}\left[\mathbf{z}_i \left(y_i - x_i\delta\right) 1\left(q_i \le \gamma_0\right)\right] = \mathbf{0};$$

we compare its performance with the GMM-1 estimator which is based on the moment conditions

$$\mathbb{E}\left[\mathbf{z}_i \left(y_i - x_i\delta 1\left(q_i \le \gamma_0\right)\right)\right] = \mathbf{0},$$

and the GMM-I2 estimator which is based on the moment conditions

$$\mathbb{E}\left[\mathbf{z}_i e_{1i} 1\left(q_i \le \gamma_0\right)\right] = \mathbf{0},$$

where $e_{1i} = y_i - g_i\delta = \delta e_{xi} + u_i = \left(\delta + \psi_x\right) e_{xi} + e_{ui}$. The former two asymptotic variances take the form

$$\mathbb{E}\left[u^2\right]\left(G'WG\right)^{-1},$$

where $\mathbb{E}\left[u^2\right] = \psi_x^2 + 1$ and $G = \mathbb{E}\left[\mathbf{z}_i x_{i,\le\gamma_0}\right]$. In CH's GMM, $W = \mathbb{E}\left[\mathbf{z}_i \mathbf{z}_{i,\le\gamma_0}'\right]^{-1}$ (and $\mathbb{E}\left[u^2\right]$ can be estimated based on $\mathbb{E}\left[u^2 | q \le \gamma_0\right]$) and in GMM-1, $W = \mathbb{E}\left[\mathbf{z}_i \mathbf{z}_i'\right]^{-1}$. So CH's GMM is more efficient because only the data points with $q_i \le \gamma_0$ are informative for δ whereas GMM-1 uses some redundant data. In GMM-I2, the asymptotic variance takes the form

$$\left(\mathbb{E}\left[g_i \mathbf{z}_{i,\le\gamma_0}'\right]\Omega^{-1}\mathbb{E}\left[\mathbf{z}_{i,\le\gamma_0} g_i\right]\right)^{-1},$$

where

$$\begin{aligned}\Omega &= \mathbb{E}\left[e_{1i}^2\right]\mathbb{E}\left[\mathbf{z}_i \mathbf{z}_{i,\le\gamma_0}'\right] + \mathbb{E}\left[v_{xi}^2\right]\delta^2\mathbb{E}\left[\mathbf{z}_i \mathbf{z}_{i,\le\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i \mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i \mathbf{z}_{i,\le\gamma_0}'\right] - 2\delta\mathbb{E}\left[v_{xi}e_{1i}\right]\mathbb{E}\left[\mathbf{z}_i \mathbf{z}_{i,\le\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i \mathbf{z}_i'\right]^{-1}\mathbb{E}\left[\mathbf{z}_i \mathbf{z}_{i,\le\gamma_0}'\right] \\ &= \mathbb{E}\left[\mathbf{z}_i \mathbf{z}_{i,\le\gamma_0}'\right]\left(\mathbb{E}\left[e_{1i}^2\right]\mathbb{E}\left[\mathbf{z}_i \mathbf{z}_{i,\le\gamma_0}'\right]^{-1} + \mathbb{E}\left[v_{xi}^2\right]\delta^2\mathbb{E}\left[\mathbf{z}_i \mathbf{z}_i'\right]^{-1} - 2\delta\mathbb{E}\left[v_{xi}e_{1i}\right]\mathbb{E}\left[\mathbf{z}_i \mathbf{z}_i'\right]^{-1}\right)\mathbb{E}\left[\mathbf{z}_i \mathbf{z}_{i,\le\gamma_0}'\right].\end{aligned}$$

It seems that even in this simple case, it is hard to compare the asymptotic variance of GMM-I2 with those of CH's GMM and GMM-1. If the randomness of the generated regressors can be neglected, then the asymptotic variance reduces to

$$\begin{aligned}\mathbb{E}\left[e_{1i}^2\right]\left(\mathbb{E}\left[g_i \mathbf{z}_{i,\le\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i \mathbf{z}_{i,\le\gamma_0}'\right]^{-1}\mathbb{E}\left[\mathbf{z}_{i,\le\gamma_0} g_i\right]\right)^{-1} &= \mathbb{E}\left[e_{1i}^2\right]\left(\mathbb{E}\left[x_i \mathbf{z}_{i,\le\gamma_0}'\right]\mathbb{E}\left[\mathbf{z}_i \mathbf{z}_{i,\le\gamma_0}'\right]^{-1}\mathbb{E}\left[\mathbf{z}_{i,\le\gamma_0} x_i\right]\right)^{-1} \\ &= \mathbb{E}\left[e_{1i}^2\right]\left(G'WG\right)^{-1} \text{ with } W = \mathbb{E}\left[\mathbf{z}_i \mathbf{z}_{i,\le\gamma_0}'\right]^{-1}.\end{aligned}$$

Since it is hard to compare the magnitude of $\mathbb{E}\left[e_{1i}^2\right] = \mathbb{E}\left[\left(\delta e_{xi} + u_i\right)^2\right] = \left(\delta + \psi_x\right)^2 + 1$ and $\mathbb{E}\left[u_i^2\right] = \psi_x^2 + 1$, it is still hard to compare the asymptotic variance of GMM-I2 with that of CH's GMM.

Additional References

Botev, Z.I. et al., 2010, Kernel Density Estimation via Diffusion, *The Annals of Statistics*, 38, 2916–2957.

Porter, J. and P. Yu, 2015, Regression Discontinuity with Unknown Discontinuity Points: Testing and Estimation, *Journal of Econometrics*, 189, 132-147.

$n \rightarrow$	200						800					
$\delta \rightarrow$	0.5		1		2		0.5		1		2	
$\kappa \rightarrow$	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ
CF-I	0.695	0.752	0.487	0.620	0.381	0.558	0.431	0.479	0.301	0.396	0.238	0.354
CF-II	0.199	0.201	0.054	0.053	0.021	0.021	0.046	0.049	0.013	0.013	0.005	0.005

Table 14: MAD for Two CF Estimators of γ

$n \rightarrow$	200						800					
$\delta \rightarrow$	0.5		1		2		0.5		1		2	
$\kappa \rightarrow$	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ
CF-I	0.973	0.985	0.972	0.971	0.982	0.973	0.977	0.975	0.972	0.973	0.977	0.971
CF-II	0.971	0.974	0.972	0.986	0.988	0.983	0.977	0.976	0.980	0.977	0.987	0.981

Table 15: Coverage of Nominal 95% Confidence Intervals for γ

$n \rightarrow$	200						800					
$\delta \rightarrow$	0.5		1		2		0.5		1		2	
$\kappa \rightarrow$	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ
CF-I	2.486	2.709	1.752	2.330	1.478	2.196	1.451	1.679	1.023	1.338	0.885	1.250
CF-II	1.085	1.080	0.271	0.268	0.091	0.093	0.243	0.250	0.064	0.065	0.023	0.023

Table 16: Length of Nominal 95% Confidence Intervals for γ

$n \rightarrow$	200						800					
$\delta \rightarrow$	0.5		1		2		0.5		1		2	
$\kappa \rightarrow$	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ
CF-I	0.131	0.161	0.151	0.206	0.188	0.363	0.057	0.067	0.060	0.086	0.082	0.147
CH's GMM	0.117	0.137	0.115	0.156	0.110	0.243	0.053	0.060	0.052	0.069	0.055	0.112
GMM-I1	0.153	0.177	0.154	0.208	0.167	0.347	0.067	0.076	0.061	0.087	0.069	0.144
GMM-I2	0.132	0.162	0.153	0.209	0.190	0.366	0.058	0.067	0.060	0.086	0.082	0.149
CF-II	0.086	0.092	0.086	0.098	0.091	0.148	0.041	0.045	0.040	0.049	0.043	0.069
CH's GMM	0.106	0.121	0.107	0.142	0.111	0.220	0.051	0.057	0.051	0.068	0.055	0.105
GMM-II1	0.119	0.136	0.118	0.156	0.123	0.257	0.058	0.063	0.056	0.075	0.061	0.120
GMM-II2	0.086	0.091	0.086	0.093	0.090	0.114	0.041	0.045	0.040	0.047	0.043	0.052

Table 17: RMSE for Eight Estimators of δ

$n \rightarrow$	200						800					
$\delta \rightarrow$	0.5		1		2		0.5		1		2	
$\kappa \rightarrow$	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ
CF-I	0.904	0.879	0.905	0.886	0.936	0.896	0.930	0.913	0.942	0.938	0.944	0.929
CH's GMM	0.919	0.872	0.913	0.875	0.951	0.878	0.935	0.923	0.937	0.953	0.933	0.933
GMM-I1	0.928	0.910	0.938	0.931	0.963	0.934	0.923	0.915	0.954	0.948	0.953	0.937
GMM-I2	0.905	0.875	0.897	0.872	0.919	0.880	0.923	0.914	0.939	0.932	0.923	0.918
CF-II	0.944	0.926	0.941	0.950	0.941	0.946	0.945	0.941	0.952	0.957	0.945	0.955
CH's GMM	0.949	0.931	0.937	0.937	0.959	0.947	0.951	0.933	0.943	0.964	0.936	0.956
GMM-II1	0.956	0.938	0.942	0.954	0.954	0.954	0.935	0.939	0.951	0.959	0.937	0.954
GMM-II2	0.946	0.929	0.943	0.953	0.942	0.943	0.945	0.946	0.950	0.953	0.948	0.957

Table 18: Coverage of Nominal 95% Confidence Intervals for δ

$n \rightarrow$	200						800					
$\delta \rightarrow$	0.5		1		2		0.5		1		2	
$\kappa \rightarrow$	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ	0.2δ	δ
CF-I	0.427	0.488	0.481	0.668	0.653	1.158	0.207	0.235	0.229	0.318	0.302	0.541
CH's GMM	0.400	0.423	0.404	0.522	0.429	0.822	0.199	0.216	0.202	0.270	0.214	0.426
GMM-I1	0.508	0.577	0.515	0.732	0.623	1.226	0.232	0.264	0.239	0.335	0.282	0.553
GMM-I2	0.427	0.487	0.477	0.658	0.634	1.133	0.207	0.234	0.226	0.313	0.289	0.527
CF-II	0.323	0.342	0.323	0.396	0.335	0.573	0.161	0.171	0.162	0.199	0.167	0.286
CH's GMM	0.398	0.445	0.401	0.561	0.427	0.890	0.198	0.221	0.200	0.278	0.211	0.440
GMM-II1	0.448	0.498	0.447	0.622	0.475	1.000	0.219	0.244	0.221	0.307	0.234	0.488
GMM-II2	0.323	0.340	0.323	0.373	0.333	0.428	0.161	0.170	0.162	0.187	0.166	0.214

Table 19: Length of Nominal 95% Confidence Intervals for δ