

SUPPLEMENT TO “ON HECKITS, LATE, AND NUMERICAL EQUIVALENCE”  
(*Econometrica*, Vol. 87, No. 2, March 2019, 677–696)

PATRICK KLINE

Department of Economics, UC Berkeley and NBER

CHRISTOPHER R. WALTERS

Department of Economics, UC Berkeley and NBER

*Proof of Proposition 2*

WE BEGIN BY REWRITING the IV and control function estimates of  $LATE_z$  in matrix form. The IV estimator is given by

$$\begin{aligned} \widehat{LATE}_z^{IV} &= \frac{[\hat{P}(z)\bar{Y}_1^z + (1 - \hat{P}(z))\bar{Y}_0^z] - [\hat{P}(z-1)\bar{Y}_1^{z-1} + (1 - \hat{P}(z-1))\bar{Y}_0^{z-1}]}{\hat{P}(z) - \hat{P}(z-1)} \\ &= \Psi_1^z(\hat{P})'\bar{Y}_1 - \Psi_0^z(\hat{P})'\bar{Y}_0, \end{aligned}$$

where  $\bar{Y}_d \equiv (\bar{Y}_d^0, \bar{Y}_d^1, \dots, \bar{Y}_d^K)'$  is the  $(K + 1) \times 1$  vector of sample average outcomes for each value of  $z$  conditional on  $D_i = d$  and  $\hat{P}$  is the vector of propensity score estimates. The  $(K + 1) \times 1$  vector  $\Psi_1^z(\hat{P})$  has  $-\hat{P}(z-1)/[\hat{P}(z) - \hat{P}(z-1)]$  at entry  $z-1$ ,  $\hat{P}(z)/[\hat{P}(z) - \hat{P}(z-1)]$  at entry  $z$ , and zeros elsewhere, and the  $(K + 1) \times 1$  vector  $\Psi_0^z(\hat{P})$  has  $(1 - \hat{P}(z-1))/[\hat{P}(z) - \hat{P}(z-1)]$  at entry  $z-1$ ,  $-(1 - \hat{P}(z))/[\hat{P}(z) - \hat{P}(z-1)]$  at entry  $z$ , and zeros elsewhere:

$$\begin{aligned} \Psi_1^z(\hat{P}) &= \left( 0, \dots, 0, \frac{-\hat{P}(z-1)}{\hat{P}(z) - \hat{P}(z-1)}, \frac{\hat{P}(z)}{\hat{P}(z) - \hat{P}(z-1)}, 0, \dots, 0 \right)', \\ \Psi_0^z(\hat{P}) &= \left( 0, \dots, 0, \frac{(1 - \hat{P}(z-1))}{\hat{P}(z) - \hat{P}(z-1)}, \frac{-(1 - \hat{P}(z))}{\hat{P}(z) - \hat{P}(z-1)}, 0, \dots, 0 \right)'. \end{aligned}$$

The second-step control function estimates with  $L = K$  can be rewritten

$$(\hat{\alpha}_d, \hat{\gamma}_{d1}, \dots, \hat{\gamma}_{dK}) = \arg \min_{\alpha_d, \gamma_{d1}, \dots, \gamma_{dK}} \sum_i 1\{D_i = d\} \left[ Y_i - \alpha_d - \sum_{\ell=1}^K 1\{Z_i = z\} \gamma_{d\ell} \lambda_{d\ell}(\hat{P}(z)) \right]^2.$$

This is a saturated OLS regression of  $Y_i$  on  $Z_i$  for each treatment category. The coefficient estimates satisfy

$$\hat{\alpha}_d + \sum_{\ell=1}^K \hat{\gamma}_{d\ell} \lambda_{d\ell}(\hat{P}(z)) = \bar{Y}_d^z, \quad d \in \{0, 1\}, z \in \{0, 1, \dots, K\}.$$

Patrick Kline: [pkline@berkeley.edu](mailto:pkline@berkeley.edu)

Christopher R. Walters: [crwalters@econ.berkeley.edu](mailto:crwalters@econ.berkeley.edu)

Letting  $\hat{\Delta}_d = (\hat{\alpha}_d, \hat{\gamma}_{d1}, \dots, \hat{\gamma}_{dK})'$  denote the control function estimates for treatment value  $d$ , we can write this system in matrix form as

$$\Lambda_d(\hat{P})\hat{\Delta}_d = \bar{Y}_d,$$

where the matrix  $\Lambda_d(\hat{P})$  has ones in its first column and  $\lambda_{dj-1}(\hat{P}(k))$  in row  $k$  and column  $j > 1$ :

$$\Lambda_d(\hat{P}) = \begin{bmatrix} 1 & \lambda_{d1}(\hat{P}(1)) & \cdots & \lambda_{dK}(\hat{P}(1)) \\ 1 & \lambda_{d1}(\hat{P}(2)) & \cdots & \lambda_{dK}(\hat{P}(2)) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_{d1}(\hat{P}(K)) & \cdots & \lambda_{dK}(\hat{P}(K)) \end{bmatrix}.$$

The control function estimates are therefore given by

$$\hat{\Delta}_d = \Lambda_d(\hat{P})^{-1}\bar{Y}_d.$$

The values of  $\lambda_{dk}(\hat{P}(z))$  are well-defined for all  $(k, z)$  whenever  $0 < \hat{P}(z) < 1 \forall z$ , and  $\Lambda_d(\hat{P})$  is full rank if  $\hat{P}(z) \neq \hat{P}(z')$  whenever  $z \neq z'$ . These requirements hold if Conditions 1 and 2 are true for every pair of instrument values, so the matrix  $\Lambda_d(\hat{P})$  is invertible under the conditions of Proposition 2 and the control function estimate  $\hat{\Delta}_d$  exists.

In matrix form, the control function estimate of  $LATE_z$  is given by

$$\widehat{LATE}_z^{CF} = Y^z(\hat{P})'(\hat{\Delta}_1 - \hat{\Delta}_0),$$

where the  $(K+1) \times 1$  vector  $Y^z(\hat{P})$  has first entry equal to unity and  $k$ th entry  $\Gamma_{k-1}(\hat{P}(z-1), \hat{P}(z))$  for  $k > 1$ :

$$Y^z(\hat{P}) = \left( 1, \frac{\hat{P}(z)\lambda_{11}(\hat{P}(z)) - \hat{P}(z-1)\lambda_{11}(\hat{P}(z-1))}{\hat{P}(z) - \hat{P}(z-1)}, \dots, \frac{\hat{P}(z)\lambda_{1K}(\hat{P}(z)) - \hat{P}(z-1)\lambda_{1K}(\hat{P}(z-1))}{\hat{P}(z) - \hat{P}(z-1)} \right)'$$

Plugging in the formulas for  $\hat{\Delta}_1$  and  $\hat{\Delta}_0$  yields

$$\widehat{LATE}_z^{CF} = Y^z(\hat{P})'\Lambda_1(\hat{P})^{-1}\bar{Y}_1 - Y^z(\hat{P})'\Lambda_0(\hat{P})^{-1}\bar{Y}_0.$$

The IV and control functions are therefore identical if  $\Psi_d^z(\hat{P})' = Y^z(\hat{P})'\Lambda_d(\hat{P})^{-1}$  for  $d \in \{0, 1\}$ , or equivalently, if  $\Lambda_d(\hat{P})'\Psi_d^z(\hat{P}) = Y^z(\hat{P})$  for  $d \in \{0, 1\}$ .

For  $d = 1$ , we have

$$\begin{aligned} \Lambda_1(\hat{P})'\Psi_1^z(\hat{P}) &= \left( 1, \frac{\hat{P}(z)\lambda_{11}(\hat{P}(z)) - \hat{P}(z-1)\lambda_{11}(\hat{P}(z-1))}{\hat{P}(z) - \hat{P}(z-1)}, \dots, \right. \\ &\quad \left. \frac{\hat{P}(z)\lambda_{1K}(\hat{P}(z)) - \hat{P}(z-1)\lambda_{1K}(\hat{P}(z-1))}{\hat{P}(z) - \hat{P}(z-1)} \right)' \\ &= Y^z(\hat{P}). \end{aligned}$$

For  $d = 0$ , we have

$$\begin{aligned}
A_0(\hat{P})' \Psi_0^z(\hat{P}) &= \left( 1, \frac{\lambda_{01}(\hat{P}(z-1))(1 - \hat{P}(z-1)) - \lambda_{01}(\hat{P}(z))(1 - \hat{P}(z))}{\hat{P}(z) - \hat{P}(z-1)}, \dots, \right. \\
&\quad \left. \frac{\lambda_{01}(\hat{P}(z-1))(1 - \hat{P}(z-1)) - \lambda_{01}(\hat{P}(z))(1 - \hat{P}(z))}{\hat{P}(z) - \hat{P}(z-1)} \right)' \\
&= \left( 1, \frac{\lambda_{11}(\hat{P}(z))\hat{P}(z) - \lambda_{11}(\hat{P}(z-1))\hat{P}(z-1)}{\hat{P}(z) - \hat{P}(z-1)}, \dots, \right. \\
&\quad \left. \frac{\lambda_{1K}(\hat{P}(z))\hat{P}(z) - \lambda_{1K}(\hat{P}(z-1))\hat{P}(z-1)}{\hat{P}(z) - \hat{P}(z-1)} \right)' \\
&= Y^z(\hat{P}),
\end{aligned}$$

where the second equality follows from the fact that  $p'\lambda_{1\ell}(p') - p\lambda_{1\ell}(p) = (1 - p)\lambda_{0\ell}(p) - (1 - p')\lambda_{0\ell}(p')$  for any  $p, p'$ , and  $\ell$ . This implies that  $\widehat{LATE}_z^{\text{IV}}$  and  $\widehat{LATE}_z^{\text{CF}}$  are equal to the same linear combination of  $\bar{Y}_1$  and  $\bar{Y}_0$ , so these estimates are identical for any  $z$ .

### *Proof of Proposition 3*

The unrestricted control function estimates come from the regression

$$\begin{aligned}
Y_i &= \alpha_0(0)(1 - D_i)(1 - X_i) + \gamma_0(0)(1 - D_i)(1 - X_i)\lambda_0(\hat{P}(0, Z_i)) \\
&\quad + \alpha_0(1)(1 - D_i)X_i + \gamma_0(1)(1 - D_i)X_i\lambda_0(\hat{P}(1, Z_i)) \\
&\quad + \alpha_1(0)D_i(1 - X_i) + \gamma_1(0)D_i(1 - X_i)\lambda_1(\hat{P}(0, Z_i)) \\
&\quad + \alpha_1(1)D_iX_i + \gamma_1(1)D_iX_i\lambda_1(\hat{P}(1, Z_i)) + \epsilon_i.
\end{aligned}$$

We can write this equation in matrix form as

$$Y = W\Delta + \epsilon,$$

where  $W$  is the matrix of regressors and  $\Delta = (\alpha_0(0), \gamma_0(0), \alpha_0(1), \gamma_0(1), \alpha_1(0), \gamma_1(0), \alpha_1(1), \gamma_1(1))'$  collects the control function coefficients. Under the conditions of Proposition 3,  $W'W$  has full rank and the unrestricted control function estimates are

$$\hat{\Delta}_u = (W'W)^{-1}W'Y.$$

The estimator in equation (13) imposes three restrictions:  $\alpha_1(1) - \alpha_1(0) = \alpha_0(1) - \alpha_0(0)$ ,  $\gamma_1(1) = \gamma_1(0)$ , and  $\gamma_0(1) = \gamma_0(0)$ . The resulting estimates can be written

$$(\hat{\Delta}_r, \hat{\varrho}) = \arg \min_{\Delta, \varrho} (Y - W\Delta)'(Y - W\Delta) - \varrho C\Delta,$$

where

$$C = \begin{bmatrix} -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and  $\varrho$  is a Lagrange multiplier. Then

$$\begin{aligned} \hat{\Delta}_r &= \hat{\Delta}_u + (W'W)^{-1}C'\hat{\varrho}, \\ \hat{\varrho} &= -(C(W'W)^{-1}C')^{-1}C\hat{\Delta}_u. \end{aligned}$$

For any estimate  $\hat{\Delta}$ , the corresponding estimate of LATE for compliers with  $X_i = 1$  is  $Y(\hat{P})'\hat{\Delta}$ , with

$$Y(\hat{P}) = (0, 0, -1, -\Gamma(\hat{P}(1, 0), \hat{P}(1, 1)), 0, 0, 1, \Gamma(\hat{P}(1, 0), \hat{P}(1, 1)))'.$$

The restricted estimate of  $LATE(1)$  is therefore

$$\begin{aligned} \widehat{LATE}_r^{CF}(1) &= Y(\hat{P})'(\hat{\Delta}_u + (W'W)^{-1}C'\hat{\varrho}) \\ &= \widehat{LATE}_1^{CF}(1) + Y(\hat{P})'(W'W)^{-1}C'(C(W'W)^{-1}C')^{-1}\zeta, \end{aligned}$$

where  $\zeta = -C\hat{\Delta}_u$  is the constraint vector evaluated at the unrestricted estimates:

$$\zeta = ([\hat{\alpha}_0(0) - \hat{\alpha}_0(1)] - [\hat{\alpha}_1(0) - \hat{\alpha}_1(1)], \hat{\gamma}_1(0) - \hat{\gamma}_1(1), \hat{\gamma}_0(0) - \hat{\gamma}_0(1))'.$$

Write  $\Omega = (W'W)^{-1}C'(C(W'W)^{-1}C')^{-1}$ , and let  $\nu_k$  denote the  $3 \times 1$  vector equal to the transpose of the  $k$ th row of  $\Omega$ . Using the fact that a scalar is equal to its trace, we can then write the difference in restricted and unrestricted LATE estimates as

$$\begin{aligned} \widehat{LATE}_r^{CF}(1) - \widehat{LATE}_1^{CF}(1) &= \text{tr}(Y(\hat{P})'\Omega\zeta) \\ &= \text{tr}(\Omega\zeta Y(\hat{P})') \\ &= \varphi'\zeta, \end{aligned}$$

where  $\varphi = \nu_7 - \nu_3 + \Gamma(\hat{P}(1, 0), \hat{P}(1, 1))(\nu_8 - \nu_4) \equiv (\varphi_1, \varphi_2, \varphi_3)'$ . Then

$$\begin{aligned} \widehat{LATE}_r^{CF}(1) - \widehat{LATE}_1^{CF}(1) &= \varphi_1([\hat{\alpha}_0(0) - \hat{\alpha}_0(1)] - [\hat{\alpha}_1(0) - \hat{\alpha}_1(1)]) \\ &\quad + \varphi_2(\hat{\gamma}_1(0) - \hat{\gamma}_1(1)) + \varphi_3(\hat{\gamma}_0(0) - \hat{\gamma}_0(1)) \\ &= \varphi_1(\widehat{LATE}_1^{CF}(1) - \widehat{LATE}_0^{CF}(1)) \\ &\quad + (\hat{\gamma}_1(0) - \hat{\gamma}_1(1))(\varphi_2 + \varphi_1\Gamma(\hat{P}(1, 0), \hat{P}(1, 1))) \\ &\quad + (\hat{\gamma}_0(0) - \hat{\gamma}_0(1))(\varphi_3 - \varphi_1\Gamma(\hat{P}(1, 0), \hat{P}(1, 1))). \end{aligned}$$

This implies

$$\widehat{LATE}_r^{CF}(1) = w\widehat{LATE}_1^{CF}(1) + (1-w)\widehat{LATE}_0^{CF}(1) + b_1(\hat{\gamma}_1(1) - \hat{\gamma}_1(0)) + b_0(\hat{\gamma}_0(1) - \hat{\gamma}_0(0)),$$

where  $w = 1 + \varphi_1$ ,  $b_1 = -(\varphi_2 + \varphi_1\Gamma(\hat{P}(1, 0), \hat{P}(1, 1)))$ , and  $b_0 = \varphi_1\Gamma(\hat{P}(1, 0), \hat{P}(1, 1)) - \varphi_3$ . Furthermore, note that the elements of  $\varphi$  only depend on sample moments of  $D_i$ ,  $X_i$ , and  $\hat{P}(X_i, Z_i)$ , so the proposition follows.

*Proof of Proposition 4*

The log-likelihood function for model (16) is

$$\begin{aligned} & \log \mathcal{L}(P(0), P(1), \alpha_0, \alpha_1, \rho_0, \rho_1) \\ &= \sum_i D_i \log \left( \int_0^{P(Z_i)} [Y_i F_{\epsilon|U}(\alpha_1|u; \rho_1) + (1 - Y_i)(1 - F_{\epsilon|U}(\alpha_1|u; \rho_1))] du \right) \\ &+ \sum_i (1 - D_i) \log \left( \int_{P(Z_i)}^1 [Y_i F_{\epsilon|U}(\alpha_0|u; \rho_0) + (1 - Y_i)(1 - F_{\epsilon|U}(\alpha_0|u; \rho_0))] du \right). \end{aligned}$$

We first rewrite this likelihood in terms of the six identified parameters of the LATE model, which are given by

$$\begin{aligned} \pi_{at} &= P(0), \\ \pi_c &= P(1) - P(0), \\ \mu_{1at} &= \frac{\int_0^{P(0)} F_{\epsilon|U}(\alpha_1|u; \rho_1) du}{P(0)}, \\ \mu_{0nt} &= \frac{\int_{P(1)}^1 F_{\epsilon|U}(\alpha_0|u; \rho_0) du}{1 - P(1)}, \\ \mu_{dc} &= \frac{\int_{P(0)}^{P(1)} F_{\epsilon|U}(\alpha_d|u; \rho_d) du}{P(1) - P(0)}, \quad d \in \{0, 1\}. \end{aligned}$$

Note that since  $F_{\epsilon|U}(\cdot|u; \rho)$  is a CDF, we have  $\mu_{dg} \in [0, 1] \forall (d, g)$ . Substituting these parameters into the likelihood function yields

$$\begin{aligned} & \log \mathcal{L}(\pi_{at}, \pi_c, \mu_{1at}, \mu_{0nt}, \mu_{1c}, \mu_{0c}) \\ &= \sum_i D_i Z_i \log(\pi_{at} [Y_i \mu_{1at} + (1 - Y_i)(1 - \mu_{1at})] + \pi_c [Y_i \mu_{1c} + (1 - Y_i)(1 - \mu_{1c})]) \\ &+ \sum_i D_i (1 - Z_i) \log(\pi_{at} [Y_i \mu_{1at} + (1 - Y_i)(1 - \mu_{1at})]) \\ &+ \sum_i (1 - D_i) Z_i \log((1 - \pi_{at} - \pi_c) [Y_i \mu_{0nt} + (1 - Y_i)(1 - \mu_{0nt})]) \\ &+ \sum_i (1 - D_i) (1 - Z_i) \log((1 - \pi_{at} - \pi_c) [Y_i \mu_{0nt} + (1 - Y_i)(1 - \mu_{0nt})]) \\ &+ \pi_c [Y_i \mu_{0c} + (1 - Y_i)(1 - \mu_{0c})]. \end{aligned}$$

We first consider interior solutions. The first-order conditions are

$$\begin{aligned}
[\mu_{1at}] &: \sum_i \left( \frac{D_i(2Y_i - 1)Z_i\pi_{at}}{\pi_{at}[Y_i\mu_{1at} + (1 - Y_i)(1 - \mu_{1at})] + \pi_c[Y_i\mu_{1c} + (1 - Y_i)(1 - \mu_{1c})]} \right. \\
&\quad \left. + \frac{D_i(2Y_i - 1)(1 - Z_i)\pi_{at}}{\pi_{at}[Y_i\mu_{1at} + (1 - Y_i)(1 - \mu_{1at})]} \right) = 0, \\
[\mu_{0nt}] &: \sum_i \left( \frac{(1 - D_i)(2Y_i - 1)Z_i(1 - \pi_{at} - \pi_c)}{(1 - \pi_{at} - \pi_c)[Y_i\mu_{0nt} + (1 - Y_i)(1 - \mu_{0nt})]} \right. \\
&\quad \left. + \frac{(1 - D_i)(2Y_i - 1)(1 - Z_i)(1 - \pi_{at} - \pi_c)}{(1 - \pi_{at} - \pi_c)[Y_i\mu_{0nt} + (1 - Y_i)(1 - \mu_{0nt})] + \pi_c[Y_i\mu_{0c} + (1 - Y_i)(1 - \mu_{0c})]} \right) = 0, \\
[\mu_{1c}] &: \sum_i \frac{D_iZ_i(2Y_i - 1)\pi_c}{\pi_{at}[Y_i\mu_{1at} + (1 - Y_i)(1 - \mu_{1at})] + \pi_c[Y_i\mu_{1c} + (1 - Y_i)(1 - \mu_{1c})]} = 0, \\
[\mu_{0c}] &: \sum_i \frac{(1 - D_i)(1 - Z_i)(2Y_i - 1)\pi_c}{(1 - \pi_{at} - \pi_c)[Y_i\mu_{0nt} + (1 - Y_i)(1 - \mu_{0nt})] + \pi_c[Y_i\mu_{0c} + (1 - Y_i)(1 - \mu_{0c})]} \\
&= 0, \\
[\pi_{at}] &: \sum_i \frac{D_iZ_i[Y_i\mu_{1at} + (1 - Y_i)(1 - \mu_{1at})]}{\pi_{at}[Y_i\mu_{1at} + (1 - Y_i)(1 - \mu_{1at})] + \pi_c[Y_i\mu_{1c} + (1 - Y_i)(1 - \mu_{1c})]} \\
&\quad + \sum_i \frac{D_i(1 - Z_i)}{\pi_{at}} - \sum_i \frac{(1 - D_i)Z_i[Y_i\mu_{0nt} + (1 - Y_i)(1 - \mu_{0nt})]}{(1 - \pi_{at} - \pi_c)[Y_i\mu_{0nt} + (1 - Y_i)(1 - \mu_{0nt})]} \\
&\quad - \sum_i \frac{(1 - D_i)(1 - Z_i)[Y_i\mu_{0nt} + (1 - Y_i)(1 - \mu_{0nt})]}{(1 - \pi_{at} - \pi_c)[Y_i\mu_{0nt} + (1 - Y_i)(1 - \mu_{0nt})] + \pi_c[Y_i\mu_{0c} + (1 - Y_i)(1 - \mu_{0c})]} \\
&= 0, \\
[\pi_c] &: \sum_i \frac{D_iZ_i[Y_i\mu_{1c} + (1 - Y_i)(1 - \mu_{1c})]}{[\pi_{at}[Y_i\mu_{1at} + (1 - Y_i)(1 - \mu_{1at})] + \pi_c[Y_i\mu_{1c} + (1 - Y_i)(1 - \mu_{1c})]} \\
&\quad - \sum_i \frac{(1 - D_i)Z_i[Y_i\mu_{0nt} + (1 - Y_i)(1 - \mu_{0nt})]}{(1 - \pi_{at} - \pi_c)[Y_i\mu_{0nt} + (1 - Y_i)(1 - \mu_{0nt})]} \\
&\quad - \sum_i \frac{(1 - D_i)(1 - Z_i)(2Y_i - 1)(\mu_{0nt} - \mu_{0c})}{(1 - \pi_{at} - \pi_c)[Y_i\mu_{0nt} + (1 - Y_i)(1 - \mu_{0nt})] + \pi_c[Y_i\mu_{0c} + (1 - Y_i)(1 - \mu_{0c})]} \\
&= 0.
\end{aligned}$$

Under Conditions 1 and 2, we can compute  $\hat{\mu}_{1at}^{IV}$ ,  $\hat{\mu}_{0nt}^{IV}$ ,  $\hat{\mu}_{1c}^{IV}$ , and  $\hat{\mu}_{0c}^{IV}$ . Setting  $\hat{\pi}_c^{IV} = \hat{P}(1) - \hat{P}(0)$  and  $\hat{\pi}_{at}^{IV} = \hat{P}(0)$  and plugging the IV parameter estimates into the FIML first-order conditions, we see that these conditions are satisfied. Thus, at interior solutions, maximum likelihood and IV estimators of all parameters are equal, and it follows that  $\hat{\mu}_{dc}^{ML} = \hat{\mu}_{dc}^{IV}$  for  $d \in \{0, 1\}$ .

Next, we consider corner solutions, which occur when at least one parameter lies outside  $[0, 1]$  at the unconstrained solution to the first-order conditions. Note that  $\hat{\mu}_{1at}^{IV}$ ,  $\hat{\mu}_{0at}^{IV}$ , and  $\hat{\pi}_{at}^{IV}$  are sample means of binary variables, so these estimates are always in the unit interval.  $\hat{\pi}_c^{IV}$  is the difference in empirical treatment rates between the two values of  $Z_i$ ; without loss of generality, we assume that  $Z_i = 1$  refers to the group with the higher treatment rate, so  $\hat{\pi}_c^{IV} \in (0, 1)$ . Thus, a constraint binds if and only if  $\hat{\mu}_{dc}^{IV}$  is outside  $[0, 1]$  for  $d = 0$ ,  $d = 1$ , or both. In these cases, at least one of the maximum likelihood complier means fails to match the corresponding IV estimate because the IV estimate is outside the FIML parameter space. This establishes that the FIML and IV estimates match if and only if both  $\hat{\mu}_{1c}^{IV}$  and  $\hat{\mu}_{0c}^{IV}$  are in  $[0, 1]$ , which completes the proof.

---

*Co-editor Ulrich K. Müller handled this manuscript.*

*Manuscript received 19 June, 2017; final version accepted 12 November, 2018; available online 16 November, 2018.*