Supplement to "Testing firm conduct"

(*Quantitative Economics*, Vol. 15, No. 3, July 2024, 571–606)

Marco Duarte
Department of Economics, University of North Carolina at Chapel Hill

Lorenzo Magnolfi
Department of Economics, University of Wisconsin-Madison

Mikkel Sølvsten
Department of Economics and Business Economics, Aarhus University

Christopher Sullivan
Department of Economics, University of Wisconsin-Madison

Appendix A: Proofs

As a service to the reader, we collect here the key notational conventions and give a formulaic description of the asymptotic RV variance σ^2_{RV}. Additionally, we present a lemma that serves as the foundation for multiple subsequent proofs.

For any variable $\boldsymbol{y}$, we let $y = \boldsymbol{y} - \mathbf{w}E[\mathbf{w}'\mathbf{w}]^{-1}E[\mathbf{w}'\boldsymbol{y}]$, $\hat{y} = \boldsymbol{y} - \mathbf{w}(\mathbf{w}'\mathbf{w})^{-1}\mathbf{w}'\boldsymbol{y}$. Predicted markups are $\Delta^z_m = z\Gamma_m$ where $\Gamma_m = E[z'z]^{-1}E[z'\Delta_m]$. The GMM objective functions are $Q_m = g'_m W g_m$ where $g_m = E[z_i(p_i - \Delta_{mi})]$ and $W = E[z_i z'_i]^{-1}$ with sample analogs of $\hat{Q}_m = \hat{g}'_m \hat{W}\hat{g}_m$ where $\hat{g}_m = n^{-1}\hat{z}'(\hat{p} - \hat{\Delta}_m)$ and $\hat{W} = n(\hat{z}'\hat{z})^{-1}$. The RV test statistic is $T^{\text{RV}} = \sqrt{n}(\hat{Q}_1 - \hat{Q}_2)/\hat{\sigma}_{\text{RV}}$ where

$$\hat{\sigma}^2_{\text{RV}} = 4\big[\hat{g}'_1\hat{W}^{1/2}\hat{V}^{\text{RV}}_{11}\hat{W}^{1/2}\hat{g}_1 + \hat{g}'_2\hat{W}^{1/2}\hat{V}^{\text{RV}}_{22}\hat{W}^{1/2}\hat{g}_2 - 2\hat{g}'_1\hat{W}^{1/2}\hat{V}^{\text{RV}}_{12}\hat{W}^{1/2}\hat{g}_2\big] \tag{31}$$

and the variance estimators are $\hat{V}^{\text{RV}}_{\ell k} = \frac{1}{n}\sum_{i=1}^n \hat{\psi}_{\ell i}\hat{\psi}'_{ki}$ for the influence function

$$\hat{\psi}_{mi} = \hat{W}^{1/2}\big(\hat{z}_i(\hat{p}_i - \hat{\Delta}_{mi}) - \hat{g}_m\big) - \frac{1}{2}\hat{W}^{3/4}\big(\hat{z}_i\hat{z}'_i - \hat{W}^{-1}\big)\hat{W}^{3/4}\hat{g}_m. \tag{32}$$

The AR statistic is $T^{\text{AR}}_m = n\hat{\pi}'_m(\hat{V}^{\text{AR}}_{mm})^{-1}\hat{\pi}_m$ for $\hat{\pi}_m = \hat{W}\hat{g}_m$, $\hat{V}^{\text{AR}}_{\ell k} = \frac{1}{n}\sum_{i=1}^n \hat{\phi}_{\ell i}\hat{\phi}'_{ki}$,

$$\hat{\phi}_{mi} = \hat{W}\big(\hat{z}_i(\hat{p}_i - \hat{\Delta}_{mi}) - \hat{g}_m\big) - \hat{W}\big(\hat{z}_i\hat{z}'_i - \hat{W}^{-1}\big)\hat{W}\hat{g}_m. \tag{33}$$

Also, $\pi_m = Wg_m$. $\hat{V}^{\text{AR}}_{mm}$ is the White heteroskedasticity-robust variance estimator, since we also have $\hat{\phi}_{mi} = \hat{W}\hat{z}_i(\hat{p}_i - \hat{\Delta}_{mi} - \hat{z}'_i\hat{\pi}_m)$.

Marco Duarte: duartema@unc.edu
Lorenzo Magnolfi: magnolfi@wisc.edu
Mikkel Sølvsten: miso@econ.au.dk
Christopher Sullivan: cjsullivan@wisc.edu

Available at http://qeconomics.org. https://doi.org/10.3982/QE2319

To state the following lemma and give a formulation of σ^2_{RV}, we introduce population versions of $\hat{\psi}_{mi}$ and $\hat{\phi}_{mi}$ along with notation for their variances. Let $\psi_{mi} = W^{1/2} z_i(p_i - \Delta_{mi}) - \frac{1}{2} W^{3/4} z_i z_i' W^{3/4} g_m - \frac{1}{2} W^{1/2} g_m$ and $\phi_{mi} = W z_i e_{mi}$. Also, let $V^{\mathrm{RV}}_{\ell k} = E[\psi_{\ell i}\psi'_{ki}]$, $V^{\mathrm{AR}}_{\ell k} = E[\phi_{\ell i}\phi'_{ki}]$, and $V^{\mathrm{RV}} = E[(\psi'_{1i}, \psi'_{2i})'(\psi'_{1i}, \psi'_{2i})]$, which is a matrix with V^{RV}_{11}, V^{RV}_{12}, and V^{RV}_{22} as its entries. Finally,

$$\sigma^2_{\mathrm{RV}} = 4\big[g_1' W^{1/2} V^{\mathrm{RV}}_{11} W^{1/2} g_1 + g_2' W^{1/2} V^{\mathrm{RV}}_{22} W^{1/2} g_2 - 2 g_1' W^{1/2} V^{\mathrm{RV}}_{12} W^{1/2} g_2\big]. \tag{34}$$

Lemma A.1. *Suppose Assumptions* 1 *and* 2 *hold. For* $\ell, k, m \in \{1, 2\}$, *we have*

$$\text{(i)} \quad \sqrt{n}\begin{pmatrix} \hat{W}^{1/2}\hat{g}_1 - W^{1/2} g_1 \\ \hat{W}^{1/2}\hat{g}_2 - W^{1/2} g_2 \end{pmatrix} \xrightarrow{d} N\big(0, V^{\mathrm{RV}}\big), \qquad \text{(ii)} \quad \hat{V}^{\mathrm{RV}}_{\ell k} \xrightarrow{p} V^{\mathrm{RV}}_{\ell k}, \tag{35}$$

$$\text{(iii)} \quad \sqrt{n}(\hat{\pi}_m - \pi_m) \xrightarrow{d} N\big(0, V^{\mathrm{AR}}_m\big), \qquad \text{(iv)} \quad \hat{V}^{\mathrm{AR}}_m \xrightarrow{p} V^{\mathrm{AR}}_m. \tag{36}$$

Proof. See Appendix J. □

Remark 1. From parts (iii) and (iv), it immediately follows that $T^{\mathrm{AR}}_m \xrightarrow{d} \chi^2_{d_z}$ under $H^{\mathrm{AR}}_{0,m}$ so that the AR tests are asymptotically valid when Assumptions 1 and 2 hold. When Assumption ND also holds, it follows from parts (i), (ii), and a first-order Taylor approximation that $T^{\mathrm{RV}} \xrightarrow{d} N(0, 1)$ under H^{RV}_0 so that the RV test is asymptotically valid. Details of this step can be found in Rivers and Vuong (2002), Hall and Pelletier (2011), and are omitted. When Assumption ND fails to hold, a first-order Taylor approximation does not capture the behavior of T^{RV} as discussed in Section 5.

For the next two proofs, we rely on the following sequence of equalities:

$$E\big[\big(\Delta^z_{0i} - \Delta^z_{mi}\big)^2\big] = E\big[\big(z_i'(\Gamma_0 - \Gamma_m)\big)^2\big] \tag{37}$$

$$= (\Gamma_0 - \Gamma_m)' E\big[z_i z_i'\big](\Gamma_0 - \Gamma_m) \tag{38}$$

$$= E\big[z_i(\Delta_{0i} - \Delta_{mi})\big]' E\big[z_i z_i'\big]^{-1} E\big[z_i(\Delta_{0i} - \Delta_{mi})\big] \tag{39}$$

$$= E\big[z_i(p_i - \Delta_{mi})\big]' E\big[z_i z_i'\big]^{-1} E\big[z_i(p_i - \Delta_{mi})\big] \tag{40}$$

$$= \pi_m E\big[z_i z_i'\big]\pi_m = Q_m. \tag{41}$$

The first equality follows from $\Delta^z_{mi} = z_i\Gamma_m$, the third is implied by $\Gamma_m = E[z'z]^{-1}E[z'\Delta_m] = E[z_i z_i']^{-1}E[z_i\Delta_{mi}]$, the fourth is a consequence of $E[z_i\omega_{0i}] = 0$, $W = E[z_i z_i']^{-1}$, and $\Delta_{0i} = p_i - \omega_{0i}$, and the fifth and final equalities follow from $\pi_m = W g_m$ and $g_m = E[z_i(p_i - \Delta_{mi})]$.

Proof of Lemma 1. In our parametric framework, the falsifiable condition in Equation (28) of Berry and Haile (2014) is[39]

$$E[\boldsymbol{p}_i - \boldsymbol{\Delta}_{mi} | z_i, \mathbf{w}_i] = \mathbf{w}_i\tau + E[\omega_{0i}|z_i, \mathbf{w}_i] \quad \text{a.s.} \tag{42}$$

[39]See Section 6, Case 2 in Berry and Haile (2014) for a discussion of their nonparametric environment.

This equation, together with residualization against $\mathbf{w}_i$ and integration over the distribution of $\mathbf{w}_i$, then implies that

$$E[p_i - \Delta_{mi} | z_i] = E[\omega_{0i} | z_i] \quad \text{a.s.} \tag{43}$$

Integrating the above against z_i, using the assumption $E[z_i \omega_{0i}] = 0$, and relying on the law of iterated expectations, then yields

$$E\big[z_i(p_i - \Delta_{mi})\big] = 0 \tag{44}$$

Finally, we note the equivalence

$$E\big[(\Delta_{0i}^z - \Delta_{mi}^z)^2\big] = 0 \quad \Leftrightarrow \quad E\big[z_i(p_i - \Delta_{mi})\big] = 0 \tag{45}$$

which follows from (37), (40), and $E[z_i z_i']$ being positive definite. Thus, we have shown that Equation (28) of Berry and Haile (2014) implies (3). □

Proof of Proposition 1. For (i), we need to show the equivalence

$$\pi_m = 0 \quad \Leftrightarrow \quad E\big[(\Delta_{0i}^z - \Delta_{mi}^z)^2\big] = 0 \tag{46}$$

This equivalence is a consequence of Equations (37) and (41) in addition to $E[z_i z_i']$ being positive definite. For (ii), we we need to show the equivalence

$$Q_1 - Q_2 = 0 \quad \Leftrightarrow \quad E\big[(\Delta_{0i}^z - \Delta_{1i}^z)^2\big] - E\big[(\Delta_{0i}^z - \Delta_{2i}^z)^2\big] = 0. \tag{47}$$

This equivalence is a consequence of Equations (37) and (41). □

Proof of Proposition 2. For (i), we use Equations (38) and (41) in addition to the definition of the local alternative in Equation (13) to write

$$\sqrt{n}(Q_1 - Q_2) = \sqrt{n}\big((\Gamma_0 - \Gamma_1) - (\Gamma_0 - \Gamma_2)\big)' E\big[z_i z_i'\big]\big((\Gamma_0 - \Gamma_1) + (\Gamma_0 - \Gamma_2)\big) \tag{48}$$

$$= q' E\big[z_i z_i'\big]\big((\Gamma_0 - \Gamma_1) + (\Gamma_0 - \Gamma_2)\big). \tag{49}$$

Assumption 2, part (iii), implies that $(\Gamma_0 - \Gamma_1) + (\Gamma_0 - \Gamma_2)$ is bounded. We therefore assume essentially without loss of generality that $\sqrt{n}(Q_1 - Q_2)$ is a constant, say c.[40] From Equations (37) and (41), we can also write

$$c = \sqrt{n}\Big(E\big[(\Delta_{0i}^z - \Delta_{1i}^z)^2\big] - E\big[(\Delta_{0i}^z - \Delta_{2i}^z)^2\big]\Big) \tag{50}$$

$$= E\big[(\Delta_{0i}^{\mathrm{RV},z} - \Delta_{1i}^{\mathrm{RV},z})^2\big] - E\big[(\Delta_{0i}^{\mathrm{RV},z} - \Delta_{2i}^{\mathrm{RV},z})^2\big]. \tag{51}$$

As in Remark 1, a first-order Taylor expansion, Lemma A.1, parts (i) and (ii), together with consistency of $\hat{\sigma}_{\mathrm{RV}}^2$ now leads to

$$T^{\mathrm{RV}} = \frac{c}{\sigma_{\mathrm{RV}}} + \frac{g_1' W^{1/2} \hat{W}^{1/2} \hat{g}_1 - g_2' W^{1/2} \hat{W}^{1/2} \hat{g}_2}{\sigma_{\mathrm{RV}}} + o_p(1) \xrightarrow{d} N(c, 1). \tag{52}$$

[40] This assumption is essentially without loss of generality since the boundedness of $(\Gamma_0 - \Gamma_1) + (\Gamma_0 - \Gamma_2)$ and, therefore, of $\sqrt{n}(Q_1 - Q_2)$, allows us to otherwise argue along subsequences where $\sqrt{n}(Q_1 - Q_2)$ converges to a constant.

For (ii), we first note that Assumption 3 implies that

$$V_{mm}^{\mathrm{AR}} = E\big[z_i z_i'\big]^{-1} E\big[e_{mi}^2 z_i z_i'\big] E\big[z_i z_i'\big]^{-1} = \sigma_m^2 E\big[z_i z_i'\big]. \tag{53}$$

Thus, we have from Equations (38) and (41) that

$$\sigma_m^2 n \pi_m' \big(V_{mm}^{\mathrm{AR}}\big)^{-1} \pi_m = n(\Gamma_0 - \Gamma_m)' E\big[z_i z_i'\big](\Gamma_0 - \Gamma_m) = q_m' E\big[z_i z_i'\big] q_m, \tag{54}$$

which in turn yields that $n\pi_m'(V_{mm}^{\mathrm{AR}})^{-1}\pi_m = E[(\Delta_{0i}^{\mathrm{AR},z} - \Delta_{mi}^{\mathrm{AR},z})^2]/\sigma_m^2$ is a constant, say c_m. From Lemma A.1, parts (iii) and (iv), and continuous mapping, we then have

$$T_m^{\mathrm{AR}} = n\hat{\pi}_m' \big(\hat{V}_{mm}^{\mathrm{AR}}\big)^{-1} \hat{\pi}_m = n\hat{\pi}_m' \big(V_{mm}^{\mathrm{AR}}\big)^{-1} \hat{\pi}_m + o_p(1) \xrightarrow{d} \chi_{d_z}^2(c_m). \qquad \square$$

Proof of Proposition 3 and Corollary 1. From Equations (37) and (38), we have equivalence between $E[(\Delta_{0i}^z - \Delta_{mi}^z)^2] = 0$ and $\Gamma_0 - \Gamma_m = 0$. Furthermore, we recall that $\pi_m = \Gamma_0 - \Gamma_m$. Thus, we only need to show that $\sigma_{\mathrm{RV}}^2 = 0$ if and only if $\pi_m = 0$ for all $m \in \{1, 2\}$. Rewriting Equation (34) in matrix notation, we have

$$\sigma_{\mathrm{RV}}^2 = 4 \begin{pmatrix} W^{-1/2}\pi_1 \\ W^{-1/2}\pi_2 \end{pmatrix}' \begin{bmatrix} V_{11}^{\mathrm{RV}} & -V_{12}^{\mathrm{RV}} \\ -V_{12}^{\mathrm{RV}} & V_{22}^{\mathrm{RV}} \end{bmatrix} \begin{pmatrix} W^{-1/2}\pi_1 \\ W^{-1/2}\pi_2 \end{pmatrix}. \tag{55}$$

Therefore, the claims to be proven follow from positive definiteness of the variance matrix $V^{\mathrm{RV}} = E[(\psi_{1i}', \psi_{2i}')'(\psi_{1i}', \psi_{2i}')]$, which is the matrix with V_{11}^{RV}, V_{12}^{RV}, and V_{22}^{RV} as its entries.

To show that V^{RV} is positive definite, we take an arbitrary nonzero, nonrandom vector $v = (v_1', v_2')' \in \mathbb{R}^{2d_z}$. V^{RV} is positive definite if $E(v_1'\psi_{1i} + v_2'\psi_{2i})^2 > 0$ for any such v. For certain implied nonrandom $u = (u_1', u_2')'$ and $t = (t_1', t_2')'$ where t is nonzero, we have $E(v_1'\psi_{1i} + v_2'\psi_{2i})^2 = E(u_1' z_i \cdot u_2' z_i + t_1' z_i \cdot e_{1i} + t_2' z_i \cdot e_{2i})^2$. Because $\sigma_{12}^2 < \sigma_1^2\sigma_2^2$ (Assumption 3), we have that $E(t_1' z_i \cdot e_{1i} + t_2' z_i \cdot e_{2i})^2 > 0$. Therefore, we can only have $E(v_1'\psi_{1i} + v_2'\psi_{2i})^2 = 0$ if e_{1i} or e_{2i} is a linear function of z_i almost surely. However, because $E[z_i e_{mi}] = 0$, such dependence is ruled out by $\sigma_m^2 > 0$ (Assumption 3). $\square$

Proof of Proposition 4. The proof proceeds in three steps. Step (1) provides a function of the data $(\tilde{\Psi}_-', \tilde{\Psi}_+')'$ and two constants μ_-, μ_+ such that $(\tilde{\Psi}_-', \tilde{\Psi}_+')' \xrightarrow{d} (\Psi_-', \Psi_+')'$. All of $(\tilde{\Psi}_-', \tilde{\Psi}_+')'$, μ_-, and μ_+ are also functions of the DGP. Step (2) establishes parts (ii)–(iv). Step (3) shows that $|T^{\mathrm{RV}}| = |\tilde{\Psi}_-'\tilde{\Psi}_+|/(\|\tilde{\Psi}_-\|^2 + \|\tilde{\Psi}_+\|^2 + 2\rho\tilde{\Psi}_-'\tilde{\Psi}_+ + o_p(1))^{1/2}$ and $F = (\|\tilde{\Psi}_-\|^2 + \|\tilde{\Psi}_+\|^2 - 2\rho\tilde{\Psi}_-'\tilde{\Psi}_+)/(2d_z) + o_p(1)$. Part (i) follows from continuous mapping, step (1), and step (3).

Step (1) We first provide definitions of μ_- and μ_+. Let $\tau_+ = 2(\sigma_1^2 + \sigma_2^2 + 2\sigma_{12})^{1/2}$ and $\tau_- = 2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})^{1/2}$ and use these two positive constants to define

$$\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \left(\frac{1}{\tau_-} + \frac{1}{\tau_+}\right) \begin{pmatrix} \sqrt{n}W^{1/2}g_1 \\ -\sqrt{n}W^{1/2}g_2 \end{pmatrix} + \left(\frac{1}{\tau_-} - \frac{1}{\tau_+}\right) \begin{pmatrix} -\sqrt{n}W^{1/2}g_2 \\ \sqrt{n}W^{1/2}g_1 \end{pmatrix}. \tag{56}$$

Defining $\kappa = 1 + \mathbf{1}\{\|\mu_1\| \le \|\mu_2\|\}$, we then let $\mu_- = \|\mu_\kappa\| - \|\mu_{3-\kappa}\|$ and $\mu_+ = \|\mu_1\| + \|\mu_2\|$. Since the instruments are weak for testing, we have that $\sqrt{n}W^{1/2}g_m = E[z_i z_i']^{1/2} q_m$ so μ_- and μ_+ do not depend on n.

To introduce $\tilde{\Psi}_-$ and $\tilde{\Psi}_+$, we let $\mathcal{Q}_m \in \mathbb{R}^{d_z \times d_z}$ be a nonrandom orthogonal matrix ($\mathcal{Q}_m \mathcal{Q}_m' = \mathcal{Q}_m' \mathcal{Q}_m = I_{d_z}$) such that $\mathcal{Q}_m \mu_m = \|\mu_m\| \boldsymbol{e}_1$. With these matrices in hand, we define $\tilde{\Psi}_- = \tilde{\mu}_\kappa - \tilde{\mu}_{3-\kappa}$ and $\tilde{\Psi}_+ = \tilde{\mu}_1 + \tilde{\mu}_2$ where

$$\begin{pmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{pmatrix} = \left(\frac{1}{\tau_-} + \frac{1}{\tau_+}\right) \begin{pmatrix} \sqrt{n}\mathcal{Q}_1 \hat{W}^{1/2} \hat{g}_1 \\ -\sqrt{n}\mathcal{Q}_2 \hat{W}^{1/2} \hat{g}_2 \end{pmatrix} + \left(\frac{1}{\tau_-} - \frac{1}{\tau_+}\right) \begin{pmatrix} -\sqrt{n}\mathcal{Q}_1 \hat{W}^{1/2} \hat{g}_2 \\ \sqrt{n}\mathcal{Q}_2 \hat{W}^{1/2} \hat{g}_1 \end{pmatrix}. \tag{57}$$

The preceding definitions imply that $(\tilde{\Psi}_-', \tilde{\Psi}_+')' = \sqrt{n} A (\hat{g}_1' \hat{W}^{1/2}, \hat{g}_2' \hat{W}^{1/2})'$ for a particular nonrandom matrix $A \in \mathbb{R}^{2d_z \times 2d_z}$ with $\sqrt{n} A (g_1' W^{1/2}, g_2' W^{1/2})' = (\mu_- \boldsymbol{e}_1', \mu_+ \boldsymbol{e}_1')'$. Since μ_- and μ_+ do not depend on n, it therefore follows from Lemma A.1, part (i), that $(\tilde{\Psi}_-', \tilde{\Psi}_+')' \xrightarrow{d} (\Psi_-', \Psi_+')'$ provided that $A V^{\text{RV}} A' = \left[\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right] \otimes I_{d_z}$ where ρ is the correlation between $e_{\kappa i} - e_{3-\kappa,i}$ and $e_{1i} + e_{2i}$. To see why this convergence occurs, note first that weak instruments for testing (Assumption 4) implies that $g_m = O(n^{-1/2})$ for $m = 1, 2$, which in turn yields that the second part of ψ_{mi} is $O_p(n^{-1/2})$. From Assumption 3, we then have

$$V_{\ell k}^{\text{RV}} = W^{1/2} E\big[e_{\ell i} e_{ki} z_i z_i'\big] W^{1/2} + O\big(n^{-1/2}\big) = \sigma_{\ell k} I_{d_z} + O\big(n^{-1/2}\big), \tag{58}$$

where we write σ_{mm} for σ_m^2. Thus, we have that $V^{\text{RV}} = \left[\begin{smallmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{smallmatrix}\right] \otimes I_{d_z}$. Furthermore, we note that A takes the form

$$A = \begin{bmatrix} (-1)^{3-\kappa} I & (-1)^\kappa I \\ I & I \end{bmatrix} \begin{bmatrix} \mathcal{Q}_1 & 0 \\ 0 & \mathcal{Q}_2 \end{bmatrix} \begin{bmatrix} I & I \\ I & -I \end{bmatrix} \begin{bmatrix} I\tau_-^{-1} & 0 \\ 0 & I\tau_+^{-1} \end{bmatrix} \begin{bmatrix} I & -I \\ I & I \end{bmatrix} \tag{59}$$

and leave the verification of $A(\left[\begin{smallmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{smallmatrix}\right] \otimes I_{d_z}) A' = \left[\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right] \otimes I_{d_z}$ to the reader.

Step (2) Part (iv) is an immediate implication of the triangle inequality and the definitions $\mu_- = \|\mu_\kappa\| - \|\mu_{3-\kappa}\|$ and $\mu_+ = \|\mu_1\| + \|\mu_2\|$. For part (ii), we have that $\mu_- = 0$ if and only if $\|\mu_1\|^2 - \|\mu_2\|^2 = 0$. In turn, we have that

$$\begin{aligned} \|\mu_1\|^2 - \|\mu_2\|^2 &= n\big((\tau_-^{-1} + \tau_+^{-1})^2 - (\tau_-^{-1} - \tau_+^{-1})^2\big)(Q_1 - Q_2) && (60) \\ &= 4n(\tau_+ \tau_-)^{-1}(Q_1 - Q_2), && (61) \end{aligned}$$

from which part (ii) is immediate. For part (iii), we have that $\mu_+ = 0$ if and only if $\|\mu_1\|^2 + \|\mu_2\|^2 = 0$. We now have

$$\begin{aligned} \|\mu_1\|^2 + \|\mu_2\|^2 &= n \begin{pmatrix} W^{1/2} g_1 \\ W^{1/2} g_2 \end{pmatrix}' A' A \begin{pmatrix} W^{1/2} g_1 \\ W^{1/2} g_2 \end{pmatrix} \\ &= n \begin{pmatrix} W^{-1/2} \pi_1 \\ W^{-1/2} \pi_2 \end{pmatrix}' A' A \begin{pmatrix} W^{-1/2} \pi_1 \\ W^{-1/2} \pi_2 \end{pmatrix}, && (62) \end{aligned}$$

so part (iii) follows from the positive definiteness of both W and the matrix $A'A = 4\left[\begin{smallmatrix} \tau_+^{-2} + \tau_-^{-2} & \tau_+^{-2} - \tau_-^{-2} \\ \tau_+^{-2} - \tau_-^{-2} & \tau_+^{-2} + \tau_-^{-2} \end{smallmatrix}\right] \otimes I_{d_z}$.

Step (3) For T^{RV}, we first consider the numerator. Here, we observe that

$$\tilde{\Psi}_-'\tilde{\Psi}_+ = \tilde{\mu}_\kappa'\tilde{\mu}_\kappa - \tilde{\mu}_{3-\kappa}'\tilde{\mu}_{3-\kappa} = 4n(\tau_+\tau_-)^{-1}(\hat{Q}_\kappa - \hat{Q}_{3-\kappa}) \tag{63}$$

so that $\sqrt{n}|\hat{Q}_1 - \hat{Q}_2| = (\tau_+\tau_-/4)|\tilde{\Psi}_-'\tilde{\Psi}_+|/\sqrt{n}$. For the denominator, we initially note that Lemma A.1, part (ii), and Equation (58) yields that

$$(\tau_+\tau_-/4)^{-2}n\hat{\sigma}_{\mathrm{RV}}^2 = \frac{4^3 n}{\tau_+^2\tau_-^2}\left[\sigma_1^2\hat{Q}_1 + \sigma_2^2\hat{Q}_2 - 2\sigma_{12}\hat{g}_1'\hat{W}\hat{g}_2\right] + o_p(1). \tag{64}$$

Similarly, we can calculate that

$$\begin{aligned}
&\|\tilde{\Psi}_-\|^2 + \|\tilde{\Psi}_+\|^2 + 2\rho\tilde{\Psi}_-'\tilde{\Psi}_+ \\
&= n\begin{pmatrix}\hat{W}^{1/2}\hat{g}_1\\ \hat{W}^{1/2}\hat{g}_2\end{pmatrix}' A'(\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\otimes I_{d_z})A\begin{pmatrix}\hat{W}^{1/2}\hat{g}_1\\ \hat{W}^{1/2}\hat{g}_2\end{pmatrix} && (65)\\
&= 4^2 n\begin{pmatrix}\hat{W}^{1/2}\hat{g}_1\\ \hat{W}^{1/2}\hat{g}_2\end{pmatrix}'\left(\frac{4}{\tau_+^2\tau_-^2}\begin{bmatrix}\sigma_1^2 & -\sigma_{12}\\ -\sigma_{12} & \sigma_2^2\end{bmatrix}\otimes I_{d_z}\right)\begin{pmatrix}\hat{W}^{1/2}\hat{g}_1\\ \hat{W}^{1/2}\hat{g}_2\end{pmatrix} && (66)\\
&= \frac{4^3 n}{\tau_+^2\tau_-^2}\left[\sigma_1^2\hat{Q}_1 + \sigma_2^2\hat{Q}_2 - 2\sigma_{12}\hat{g}_1'\hat{W}\hat{g}_2\right], && (67)
\end{aligned}$$

where the second equality follows from the last sentences of steps (1) and (2) together with

$$\begin{aligned}
&\begin{bmatrix}\tau_+^{-2}+\tau_-^{-2} & \tau_+^{-2}-\tau_-^{-2}\\ \tau_+^{-2}-\tau_-^{-2} & \tau_+^{-2}+\tau_-^{-2}\end{bmatrix}\begin{bmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{bmatrix}\begin{bmatrix}\tau_+^{-2}+\tau_-^{-2} & \tau_+^{-2}-\tau_-^{-2}\\ \tau_+^{-2}-\tau_-^{-2} & \tau_+^{-2}+\tau_-^{-2}\end{bmatrix}\\
&= \begin{bmatrix}\sigma_1^2 & -\sigma_{12}\\ -\sigma_{12} & \sigma_2^2\end{bmatrix}\frac{4}{\tau_+^2\tau_-^2}. && (68)
\end{aligned}$$

Thus, we have the desired conclusion

$$\begin{aligned}
\left|T^{\mathrm{RV}}\right| &= \sqrt{n}\frac{|\hat{Q}_1 - \hat{Q}_2|}{\hat{\sigma}_{\mathrm{RV}}} = \frac{4n(\tau_+\tau_-)^{-1/2}|\hat{Q}_1 - \hat{Q}_2|}{\left((\tau_+\tau_-/4^2)^{-1}n\hat{\sigma}_{\mathrm{RV}}^2\right)^{1/2}} && (69)\\
&= \left|\tilde{\Psi}_-'\tilde{\Psi}_+\right|/\left(\|\tilde{\Psi}_-\|^2 + \|\tilde{\Psi}_+\|^2 + 2\kappa\tilde{\Psi}_-'\tilde{\Psi}_+ + o_p(1)\right)^{1/2} && (70)
\end{aligned}$$

For F, we first note that standard arguments lead to

$$\begin{aligned}
2d_zF &= n\big(1-\hat{\rho}^2\big)\frac{\hat{\sigma}_2^2\hat{g}_1'\hat{W}\hat{g}_1 + \hat{\sigma}_1^2\hat{g}_2'\hat{W}\hat{g}_2 - 2\hat{\sigma}_{12}\hat{g}_1'\hat{W}\hat{g}_2}{\hat{\sigma}_1^2\hat{\sigma}_2^2 - \hat{\sigma}_{12}^2} && (71)\\
&= n\big(1-\rho^2\big)\frac{\sigma_2^2\hat{Q}_1 + \sigma_1^2\hat{Q}_2 - 2\sigma_{12}\hat{g}_1'\hat{W}\hat{g}_2}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2} + o_p(1). && (72)
\end{aligned}$$

Similarly, we can calculate that

$$
\begin{aligned}
&\|\tilde{\Psi}_-\|^2 + \|\tilde{\Psi}_+\|^2 - 2\rho\tilde{\Psi}_-'\tilde{\Psi}_+ \\
&= n\begin{pmatrix}\hat{W}^{1/2}\hat{g}_1\\ \hat{W}^{1/2}\hat{g}_2\end{pmatrix}' A'(\begin{bmatrix}1 & -\rho\\ -\rho & 1\end{bmatrix}\otimes I_{d_z})A\begin{pmatrix}\hat{W}^{1/2}\hat{g}_1\\ \hat{W}^{1/2}\hat{g}_2\end{pmatrix} && (73)\\
&= n(1-\rho^2)\begin{pmatrix}\hat{W}^{1/2}\hat{g}_1\\ \hat{W}^{1/2}\hat{g}_2\end{pmatrix}'(\begin{bmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{bmatrix}^{-1}\otimes I_{d_z})\begin{pmatrix}\hat{W}^{1/2}\hat{g}_1\\ \hat{W}^{1/2}\hat{g}_2\end{pmatrix} && (74)\\
&= n(1-\rho^2)\frac{\sigma_2^2\hat{Q}_1 + \sigma_1^2\hat{Q}_2 - 2\sigma_{12}\hat{g}_1'\hat{W}\hat{g}_2}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}, && (75)
\end{aligned}
$$

which follows from inverting the equality in the last sentence of step (1). □

Appendix B: Testing with nonconstant marginal cost

In this Appendix, we discuss how the results of the paper are preserved in a more general setting where marginal cost depends on quantity sold. While in the paper, we derive results assuming constant marginal cost, this is an important extension. It is well known going back to Bresnahan (1982) and Lau (1982) that the requirements for testing conduct are greater when marginal cost is nonconstant. Thus, the reader may wonder if the results of our paper would be qualitatively different in a more general setting.

Let $\boldsymbol{q}_i$ denote quantity, and consider a separable cost function:

$$
\boldsymbol{c}_i = \bar{\boldsymbol{c}}(\boldsymbol{q}_i, \mathbf{w}_i; \tau) + \omega_i, \tag{76}
$$

where $\bar{c}$ is specified up to some cost parameters τ. The specification of marginal cost that we adopt in the paper is a special case where $\bar{\boldsymbol{c}}(\boldsymbol{q}_i, \mathbf{w}_i; \tau) = \mathbf{w}_i\tau$.

There are two additional cases to consider. The first concerns the researcher knowing the true value of τ. We can define $\bar{\boldsymbol{\Delta}}_{mi} = \boldsymbol{\Delta}_{mi} + \bar{\boldsymbol{c}}(\boldsymbol{q}_i, \mathbf{w}_i; \tau)$: this term is pinned down by a model of conduct m, and a cost function $\bar{\boldsymbol{c}}$. In this case, we can test alternative pairs of models of conduct and cost functions with the methods described in the paper. In particular, that can be done by replacing Δ_m in the paper with $\bar{\boldsymbol{\Delta}}_m$. The set of instruments z must include $\mathbf{w}$ in this case, but since $\bar{\boldsymbol{c}}$ is fully specified, there is no additional requirement on instruments.

Alternatively, when τ is unknown to the researcher, she can estimate it under a model of conduct m either as a preliminary step, or simultaneously with testing, as she would in the case of constant marginal cost. However, excluded instruments are now needed to estimate τ, and thus must be sufficient to identify τ under the true model of conduct. If a researcher pursues a sequential approach, the researcher can construct $\boldsymbol{\omega}_{mi} = \boldsymbol{p}_i - \boldsymbol{\Delta}_{mi} - \bar{\boldsymbol{c}}(\boldsymbol{q}_i, \mathbf{w}_i; \tau_m)$ and perform testing after having estimated τ_m under each model.

The results in the paper are still applicable as long as the researcher adjusts the standard errors of the RV test statistic and the effective F-statistic. As there may exist models, falsified by a set of instruments under constant marginal cost, which are not falsified by those same instruments with a more flexible marginal cost specification, degeneracy is more likely to occur when cost is nonconstant. This is akin to the classic example of demand shifters in Bresnahan (1982). More specifically, Corollary 3 in Magnolfi, Quint, Sullivan, and Waldfogel (2022) shows that falsification does not necessarily require rotators, but rather two sets of economically distinct variables—intuitively, after using some exogenous variation to pin down the slope of marginal cost implied by a given model, we need to have some residual exogenous variation to falsify models of conduct. For instance, two cost shifters of the same rivals will not be enough, but cost shifters of two different firms may be sufficient to falsify models of conduct. While this is an additional requirement as compared to falsification in the constant marginal cost case, conditional on having this additional variation, our discussion of falsification in the constant marginal cost case goes through. As this setup places additional requirements on the instruments, evaluating the quality of the inference on conduct using our diagnostic is even more important in this setting. However, beyond the extra burden, the testing problem is fundamentally unchanged, therefore the results in this paper apply.

This discussion makes it explicit that testing firm conduct also jointly tests models of marginal cost. In fact, $\bar{\boldsymbol{\Delta}}_m$ generalizes the term $\breve{\boldsymbol{\Delta}}_m$ defined in Appendix E below. In that Appendix, we show that cost misspecification can be incorporated in $\breve{\boldsymbol{\Delta}}_m$, and thus be understood as markup misspecification. Here, misspecification of $\bar{c}$ is manifested as misspecification of $\bar{\boldsymbol{\Delta}}_m$. If one is flexible in specifying $\bar{c}$, misspecification of $\bar{\boldsymbol{\Delta}}_m$ largely concerns misspecification of $\boldsymbol{\Delta}_m$ and, therefore, conduct. Finally, this formulation shows that the methods described in the paper can be used to test models of cost, even when conduct is known.

Appendix C: Standard error adjustments

This Appendix extends all our previously introduced statistics to take into account uncertainty stemming from preliminary demand estimation as well as dependence across observations. We suppose that Δ_m is a function of demand parameters θ_0^D that are estimated using a GMM estimator $\hat{\theta}^D$. We therefore let W^D denote the GMM weight matrix and $h(\theta^D) = \frac{1}{n}\sum_{i=1}^n h_i(\theta^D)$ the GMM sample moment function used. Furthermore, we let $H = \nabla_\theta h(\hat{\theta}^D)$ be the gradient of the sample moment function h and let $G_m = -\frac{1}{n}\hat{z}'\nabla_\theta\hat{\Delta}_m(\hat{\theta}^D)$ be the gradient of $\hat{g}_m$. Both gradients are with respect to θ^D.

RV test

The RV statistic with a two-step adjustment replaces $\hat{V}_{\ell k}^{\mathrm{RV}}$ in the definition of $\hat{\sigma}_{\mathrm{RV}}^2$ with $\tilde{V}_{\ell k}^{\mathrm{RV}} = \frac{1}{n}\sum_{i=1}^n \tilde{\psi}_{\ell i}\tilde{\psi}_{ki}'$. Here, the influence function $\tilde{\psi}_{mi}$ adjusts $\hat{\psi}_{mi}$ to account for preliminary demand estimation:

$$\tilde{\psi}_{mi} = \hat{\psi}_{mi} - \hat{W}^{1/2}G_m\Phi\big(h_i(\hat{\theta}^D) - h(\hat{\theta}^D)\big), \tag{77}$$

where $\Phi = (H'W^DH)^{-1}H'W^D$. This is a standard adjustment for first-step estimation based on the asymptotic approximation $\hat{\theta}^D - \theta_0^D = -\Phi h(\theta_0^D) + o_p(n^{-1/2})$.

AR test

Analogously to the above, the AR statistics with a two-step adjustment replaces $\hat{V}^{\mathrm{AR}}_{mm}$ with $\tilde{V}^{\mathrm{AR}}_{mm}$ where $\tilde{V}^{\mathrm{AR}}_{\ell k} = \frac{1}{n}\sum_{i=1}^{n} \tilde{\phi}_{\ell i}\tilde{\phi}'_{ki}$. Here, the influence function $\tilde{\phi}_{mi}$ adjusts $\hat{\phi}_{mi} = \hat{W}\hat{z}_i(\hat{p}_i - \hat{\Delta}_{mi} - \hat{z}_i'\hat{\pi}_i)$ to account for preliminary demand estimation using the same approximation to $\hat{\theta}^D$ as above:

$$\tilde{\phi}_{mi} = \hat{\phi}_{mi} - \hat{W} G_m \Phi_m\big(h_i(\hat{\theta}^D) - h(\hat{\theta}^D)\big). \tag{78}$$

F-statistic

The F-statistic with a two-step adjustment replaces $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, and $\hat{\sigma}_{12}$ in the definition of F and $\hat{\rho}^2$ with $\tilde{\sigma}_m^2 = d_z^{-1}\,\mathrm{trace}(\tilde{V}^{\mathrm{AR}}_{mm}\hat{W}^{-1})$ for $m \in \{1, 2\}$ and $\tilde{\sigma}_{12} = d_z^{-1}\,\mathrm{trace}(\tilde{V}^{\mathrm{AR}}_{\ell k}\hat{W}^{-1})$. Here, $\tilde{V}^{\mathrm{AR}}_{\ell k}$ and $\tilde{\phi}_{mi}$ were introduced in the preceding paragraph.

An extension of Proposition 4 that accounts for two-step estimation can be established under homoskedasticity, that is, when $\hat{W}^{-1/2}\tilde{V}^{\mathrm{AR}}_{\ell k}\hat{W}^{-1/2}$ for $\ell, k \in \{1, 2\}$ converge in probability to diagonal matrices. In the absence of homoskedasticity, F is still informative about the strength of the instruments, but the exact thresholds for size control reported in Table 1 may only be approximations to the true thresholds.

Dependence

Dependent data, for example, cluster sampling, is easily accommodated by adjustments to $\hat{V}^{\mathrm{RV}}_{\ell k}$ and $\hat{V}^{\mathrm{AR}}_{\ell k}$. If we let c_{ij} take the value one if observations i and j are deemed dependent and zero otherwise, then the variance estimators used in the paper can be replaced by

$$\check{V}^{\mathrm{RV}}_{\ell k} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}\hat{\psi}_{\ell i}\hat{\psi}'_{kj} \quad \text{and} \quad \check{V}^{\mathrm{AR}}_{\ell k} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}\hat{\phi}_{\ell i}\hat{\phi}'_{kj}. \tag{79}$$

Combinations of two-step estimation and dependence are also handled by simply replacing $\hat{\psi}$ and $\hat{\phi}$ by $\tilde{\psi}$ and $\tilde{\phi}$ in the definitions of $\check{V}^{\mathrm{RV}}_{\ell k}$ and $\check{V}^{\mathrm{AR}}_{\ell k}$. Provided that suitable central limit theorem and laws of large numbers can be alluded to under the type of dependence considered, any claims on asymptotic validity under strong instruments made in the paper continue to hold. For clustered data, we refer to Hansen and Lee (2019) for such results.

Appendix D: Other model assessment tests

In this section, we discuss the two other model assessment procedures used in the empirical IO literature. Although the details of test performance differ across the three model assessment procedures, EB and Cox share with AR the undesirable property that inference on conduct is not valid under misspecification.

Estimation Based test (EB)

A test of Equation (3) can be constructed by viewing the problem as one of inference about a regression parameter. We refer to this approach as estimation based, or EB. One way to implement an estimation based approach, proposed in Pakes (2017), is to consider the equation[41]

$$p = \Delta_m \theta_m + \omega_m, \tag{80}$$

For each model m, the null and alternative hypotheses for model assessment are

$$H_{0,m}^{\text{EB}} : \theta_m = 1 \quad \text{and} \quad H_{A,m}^{\text{EB}} : \theta_m \neq 1. \tag{81}$$

Note also that under the null we have that $\omega_m = \omega_0$ so that $E[z_i \omega_{mi}] = 0$.

With this formulation, a natural testing procedure to consider is then based on the Wald statistic:

$$T_m^{\text{EB}} = (\hat{\theta}_m - 1)' \hat{V}_{\hat{\theta}_m}^{-1} (\hat{\theta}_m - 1), \tag{82}$$

where $\hat{\theta}_m$ is the 2SLS estimator applied to the sample counterpart of Equation (80) and $\hat{V}_{\hat{\theta}_m}$ is a consistent estimator of the variance of $\hat{\theta}_m$. The asymptotic null distribution of T_m^{EB} is a χ_1^2 distribution and the EB test at level α therefore rejects if T_m^{EB} exceeds the α-th quantile of that null distribution.

The EB test is similar to AR. We can, in general, show that if markups are misspecified, EB rejects the true model of conduct. To see this, note that

$$\operatorname{plim} n^{-1} T_m^{\text{EB}} = (\theta_m - 1)' V_{\theta_m}^{-1} (\theta_m - 1), \tag{83}$$

where $\theta_m = \operatorname{plim} \hat{\theta}_m$ is given as

$$\theta_m = 1 + E\big[\Delta_{mi}^z \Delta_{mi}^z\big]^{-1} E\big[\Delta_{mi}^z \big(\Delta_{0i}^z - \Delta_{mi}^z\big)\big] \tag{84}$$

Since $V_{\theta_m}^{-1}$ is strictly positive, $\operatorname{plim} n^{-1} T^{\text{EB}} = 0$ if and only if $\theta_m = 1$. Thus, EB asymptotically rejects any model m for which $E[\Delta_{mi}^z(\Delta_{0i}^z - \Delta_{mi}^z)] \neq 0$ as $\operatorname{plim} n^{-1} T_m^{\text{EB}} = \infty$ in that case. Generically with misspecification, $E[\Delta_{mi}^z(\Delta_{0i}^z - \Delta_{mi}^z)] \neq 0$ for $m = 1, 2$, and EB rejects both models. In the presence of misspecification, a researcher is not guaranteed to learn the true nature of conduct with this model assessment procedure.

Cox test (Cox)

The next testing procedure we consider is inspired by the Cox (1961) approach to testing nonnested hypotheses. To perform a Cox test, we formulate two different pairs of null and alternative hypotheses for each model m, based on the same moment conditions

[41] Alternatively, the procedure could be based on the regression $p = \Delta\theta + \omega$, where $\Delta = [\Delta_1, \Delta_2]$ is a n-by-2 vector of the implied markups for each of the two models. The analysis of this procedure is substantively identical, except for the fact that this procedure requires at least two valid instruments.

defined for RV. Specifically, for model m we formulate the null and alternative hypotheses:

$$H_{0,m}^{\text{Cox}} : g_m = 0 \quad \text{and} \quad H_{A,m}^{\text{Cox}} : g_{-m} = 0, \tag{85}$$

where $-m$ denotes the opposite of model m. To implement the Cox test in our environment, one can follow Smith (1992). With $\hat{g}_m$ as the finite sample analogue of the moment conditions, the test statistic for model m is

$$T_m^{\text{Cox}} = \frac{\sqrt{n}}{\hat{\sigma}_{\text{Cox}}}\big(\hat{g}'_{-m}\hat{W}\hat{g}_{-m} - \hat{g}'_m\hat{W}\hat{g}_m - (\hat{g}_{-m} - \hat{g}_m)'\hat{W}(\hat{g}_{-m} - \hat{g}_m)\big) \tag{86}$$

$$= \frac{2\sqrt{n}\hat{g}'_m\hat{W}(\hat{g}_{-m} - \hat{g}_m)}{\hat{\sigma}_{\text{Cox}}}, \tag{87}$$

where $\hat{\sigma}^2_{\text{Cox}} = 4\hat{g}'_{-m}\hat{V}^{\text{AR}}_{mm}\hat{g}_{-m}$ is a consistent estimator of the asymptotic variance of the numerator of T_m^{Cox} under the null. As shown in Smith (1992), this statistic is asymptotically distributed according to a standard normal distribution under the null hypothesis. Under the alternative, the mean of T_m^{Cox} is negative, so the test rejects for values of T_m^{Cox} below the αth quantile of a standard normal distribution. As for the case of RV, the asymptotic normal limit distribution requires that Assumption ND is satisfied.

The Cox test maintains—under the null and the alternative—that either of the two candidate models is correctly specified. Thus, in the presence of misspecification, one is neither under the null nor the alternative making the properties of the test hard to characterize.

In practice, as $n \to \infty$, the Cox test statistic diverges. To see this, note that the plim of T_m^{Cox} is given as

$$\operatorname{plim} T_m^{\text{Cox}} = \lim_{n\to\infty} \frac{2\sqrt{n}g'_m W(g_{-m} - g_m)}{\sigma_{\text{Cox}}} \tag{88}$$

$$= \lim_{n\to\infty} \frac{2\sqrt{n}\big(\|W^{1/2}g_1\|\|W^{1/2}g_2\|\cos(\vartheta) - \|W^{1/2}g_1\|^2\big)}{\sigma_{\text{Cox}}}, \tag{89}$$

where $\sigma^2_{\text{Cox}} = 4g_{-m}V^{\text{AR}}_{mm}g_{-m}$ and ϑ is the angle between $W^{1/2}g_1$ and $W^{1/2}g_2$. Suppose now that model 1 is the true model, both models are misspecified, and model 1 has the better fit, that is, $0 < \|W^{1/2}g_2\|^{-1}\|W^{1/2}g_1\| < 1$. While the RV test will select in favor of model 1 in this case, the behavior of the Cox test depends on the angle ϑ:

$$\operatorname{plim} T_1^{\text{Cox}} = \begin{cases} +\infty, & \text{if } \cos(\vartheta) > \dfrac{\|W^{1/2}g_1\|}{\|W^{1/2}g_2\|}, \\ -\infty, & \text{if } \cos(\vartheta) < \dfrac{\|W^{1/2}g_1\|}{\|W^{1/2}g_2\|}. \end{cases} \tag{90}$$

If treated as a two-sided test, Cox therefore rejects the true model with probability approaching one for all g_1 and g_2 except in the knife edge case of $\cos(\vartheta) = \|W^{1/2}g_2\|^{-1}\|W^{1/2}g_1\|$. If treated as a one-sided test, Cox may still reject the true model if $\cos(\vartheta)$ is sufficiently small. By similar derivations and the ordering $\|W^{1/2}g_1\|^{-1} \times \|W^{1/2}g_2\| > 1 > \cos(\vartheta)$, it follows that $\operatorname{plim} T_2^{\text{Cox}} = -\infty$, that is, the worse fitting model, model 2, is rejected with probability approaching one in large samples. In summary, if considered as a two-sided test the Cox test will reject both models in large samples, while as a one-sided test, it can also lead the researcher to reject the true model of conduct even when the true model has better asymptotic fit than the wrong model.

Appendix E: Marginal cost misspecification

In Section 4.2, we consider the consequences of misspecification of markups, for example, because of misspecification of the demand system. It is also possible that a researcher misspecifies marginal cost. Here, we show that testing with misspecified marginal costs can be reexpressed as testing with misspecifed markups so that the results in the previous section apply. As a leading example, we consider the case where the researcher specifies $\mathbf{w_a}$, which are a subset of $\mathbf{w}$. This could happen in practice because the researcher does not observe all the variables that determine marginal cost or does not specify those variables flexibly enough in constructing $\mathbf{w_a}$.[42]

To perform testing with misspecified costs, the researcher would residualize $\boldsymbol{p}$, $\boldsymbol{\Delta}_1$, $\boldsymbol{\Delta}_2$, and z with respect to $\mathbf{w_a}$ instead of $\mathbf{w}$. Let $y^{\mathbf{a}}$ denote a generic variable $\boldsymbol{y}$ residualized with respect to $\mathbf{w_a}$. Thus, with cost misspecification, both RV and AR depend on the moment

$$E\big[z_i^{\mathbf{a}}\big(p_i^{\mathbf{a}} - \Delta_{mi}^{\mathbf{a}}\big)\big] = E\big[z_i^{\mathbf{a}}\big(\Delta_{0i}^{\mathbf{a}} - \breve{\Delta}_{mi}^{\mathbf{a}}\big)\big], \tag{91}$$

where $\breve{\Delta}_{mi}^{\mathbf{a}} = \Delta_{mi}^{\mathbf{a}} - \mathrm{w}^{\mathbf{a}}\tau$ and $\mathrm{w}^{\mathbf{a}}$ is $\mathbf{w}$ residualized with respect to $\mathbf{w_a}$. When price is residualized with respect to cost shifters $\mathbf{w_a}$, the true cost shifters are not fully controlled for and the fit of model m depends on the distance between $\Delta_0^{\mathbf{a}z}$ and $\breve{\Delta}_m^{\mathbf{a}z}$. For example, suppose model 1 is the true model and demand is correctly specified so that $\boldsymbol{\Delta}_0 = \boldsymbol{\Delta}_1$. Model 1 is still falsified as $\breve{\Delta}_1^{\mathbf{a}} = \Delta_0^{\mathbf{a}} - \mathrm{w}^{\mathbf{a}}\tau$. Thus, when performing testing with misspecified cost, it is as if the researcher performs testing with markups that have been misspecified by $-\mathrm{w}^{\mathbf{a}}\tau$. We formalize the implications of cost misspecification on testing in the following lemma.

Lemma 2. *Suppose* $\mathbf{w_a}$ *is a subset of* $\mathbf{w}$, *all variables* $y^{\boldsymbol{a}}$ *are residualized with respect to* $\mathbf{w_a}$, *and Assumptions* 1, 2, *and ND are satisfied. Then, with probability approaching one as* $n \to \infty$:

(i) RV *rejects the null of equal fit in favor of model* 1 *if* $E[(\Delta_{0i}^{\mathbf{a}z} - \breve{\Delta}_{1i}^{\mathbf{a}z})^2] < E[(\Delta_{0i}^{\mathbf{a}z} - \breve{\Delta}_{2i}^{\mathbf{a}z})^2]$,

(ii) AR *rejects the null of perfect fit for any model* m *with* $E[(\Delta_{0i}^{\mathbf{a}z} - \breve{\Delta}_{mi}^{\mathbf{a}z})^2] \neq 0$.

Thus, the effects of cost misspecification can be fully understood as markup misspecification.

[42] Under a mild exogeneity condition, the results here extend to the broader case where $\mathbf{w} \neq \mathbf{w_a}$.

Appendix F: Simulations

We construct three Monte Carlo experiments that illustrate the finite sample performance of the RV test and the ability of the F-statistic to detect low power and size distortions generated by weak instruments.

Monte Carlo environment

We consider an environment that mirrors the market for yogurt in our empirical application, importing the same market structure, product characteristics, and cost shifters. We also adopt the estimated demand model of Table 3, column 3, and the estimated cost parameters implied by the zero retail margin model as the true demand system and cost function in the data generating process. The cost shocks, ω, are drawn from a normal distribution with the same mean and standard deviation as the empirical distribution of the estimated shocks. For each of the three experiments, we perform 1000 simulations; in each simulation, we sample 500 markets uniformly at random from the set of 5034 yogurt markets to limit computational burden, and we take draws from the distribution of the cost unobservables. We solve for prices and quantities according to a true model of conduct, which differs across the three experiments as discussed below. Note that in all three experiments, we assume that the researcher knows the true demand model, which allows us to avoid reestimating demand for thousands of simulations.

Experiment 1

In this experiment, we illustrate how weak instruments for power affect the RV test, and how the F-statistic diagnoses weak instruments for power in finite samples. We consider as the true model of conduct Nash price setting, and we test the Nash price setting hypothesis against a model of joint profit maximization. To vary the power of the test, we consider instruments that are found to be strong in our application, the “NoProd” (number of own and rivals’ products). We then inject noise by adding a vector of random values $\boldsymbol{u}$ drawn from a uniform distribution to construct a sequence of instruments with varying strength: $z'(\lambda_1) = (1-\lambda_1)z + \lambda_1\boldsymbol{u}$ with values of λ_1 ranging from 0 to 0.9. As λ_1 increases, the instruments become more noisy.

Figure F.1 shows the simulated rejection probability of the RV test and the fraction of simulations for which the F-statistic diagnoses that the instruments are strong for best-case power of 0.95. As we increase the noise in the instruments, the power of the test reduces monotonically, from always rejecting to power essentially equal to size. Appropriately, our proposed F-statistic detects this reduction in power by increasingly diagnosing instruments as weak for power.

Experiment 2

To perform the second experiment, we need an environment where we can approach the region of degeneracy while staying in the null space. Example 3 from the paper provides an economically meaningful environment in which we can accomplish this goal. In particular, we consider “rule-of-thumb” pricing models, and we assume that prices are set

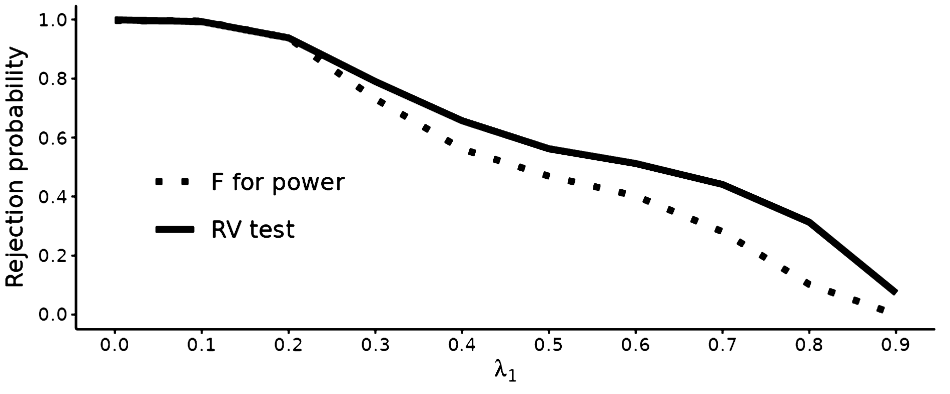

FIGURE F.1. Power reduction from weak instruments. The figure plots the rejection probabilities of the RV test (solid line) and the fraction of simulations when the F-statistic exceeds the critical value for best-case power of 0.95 (dashed line) as a function of λ_1. As λ_1 increases, we inject more noise into the NoProd instruments.

by firms as two times cost, or $\boldsymbol{p} = 2\boldsymbol{c}_0 = 2(\mathbf{w}\gamma + \boldsymbol{\omega})$. We then test between two alternative rule-of-thumb models: $\boldsymbol{\Delta}_1 = 0.5\boldsymbol{c}_0 + \lambda_2 \boldsymbol{z} + 0.01\tilde{\boldsymbol{e}}_1$ and $\boldsymbol{\Delta}_2 = 1.5\boldsymbol{c}_0 + \lambda_2 \boldsymbol{z} + 0.01\tilde{\boldsymbol{e}}_2$, where the instrument $\boldsymbol{z}$ is the sum of rivals' products, λ_2 determines the correlation between the instrument and markups for models 1 and 2, and $(\tilde{\boldsymbol{e}}_1, \tilde{\boldsymbol{e}}_2)$ are random draws from a standard normal distribution. By considering alternatives where markups are specified as multipliers of costs plus the instruments (rescaled by a parameter), we can stay under the null (as $\Gamma_1 = \Gamma_2$) while varying the strength of the instruments for size: as λ_2 gets closer to zero, we move toward the space of degeneracy (as $\Gamma_1 - \Gamma_0$ and $\Gamma_2 - \Gamma_0$ approach zero). Random components $(\tilde{\boldsymbol{e}}_1, \tilde{\boldsymbol{e}}_2)$ are included to satisfy Assumption 3.

Figure F.2 shows the simulated rejection probabilities of the RV test and the fraction of simulations for which the F-statistic diagnoses that the instruments are strong for

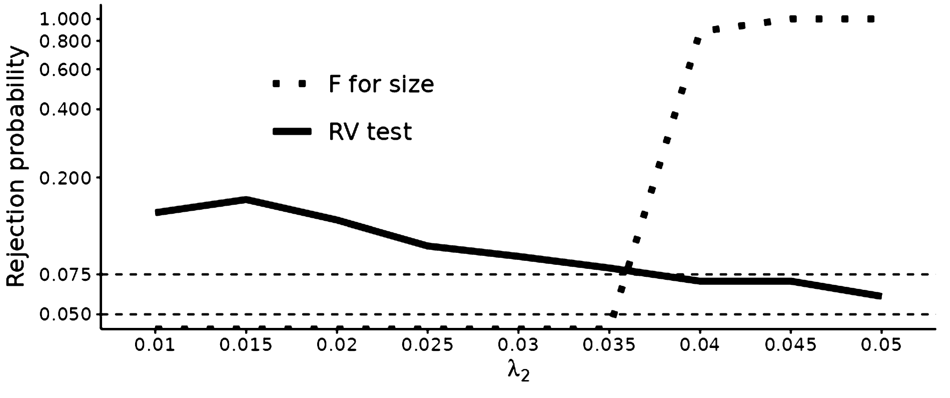

FIGURE F.2. Size distortions from weak instruments. The figure plots the rejection probabilities of the RV test (solid line) and the fraction of simulations when the F-statistic exceeds the critical value for worst-case size of 0.075 (dashed line) as a function of λ_2. As λ_2 increases, we move away from the space of degeneracy, while remaining under the null.

TABLE F.1. Fraction of simulations with correctly signed RV test statistics.

	$\Pr(T^{\mathrm{RV}} < 0)$	
$\tilde{\beta}_I^p$	JPM vs. Nash price setting (1)	JPM vs. perfect competition (2)
5.09	0.88	0.99
4.71	0.96	1
4.52	0.98	1
4.33	1	1
4.14	1	1
3.95	1	1
3.57	1	0.97

Note: The table reports the fraction of simulations for which T^{RV} is correctly signed, that is, suggests better fit of the joint profit maximization (JPM) model relative to Nash price setting (column 1) and perfect competition (column 2). Different rows correspond to different values of the demand parameter $\tilde{\beta}_I^p$ used in testing. Data are generated for the true value $\tilde{\beta}_I^p = 4.33$.

size. Near degeneracy, that is, for small values of λ_2, we find large size distortions for the RV test (around 0.15). As λ_2 increases, and we move away from the space of degeneracy, size approaches the level of the test. Our diagnostic appropriately detects when instruments are weak for size.

Experiment 3

In the third experiment, we explore the degree to which the RV test is affected by misspecification of demand. To do so, we test between different models of conduct while using a misspecified demand model to construct markups. Prices and shares are simulated from a joint profit maximization model, using the estimated demand model of Table 3, column 3. To approximate the effects of misspecification of demand for testing conduct, we construct markups for joint profit maximization, Nash price setting, and perfect competition, while altering the value of the coefficient $\tilde{\beta}_I^p$ of the interaction between prices and income, in the spirit of the misspecification introduced in Example 1 of the paper. We vary $\tilde{\beta}_I^p$ by adding values that correspond to ± 0.5, ± 1, ± 2 times the standard error for the coefficient (0.378).

Table F.1 reports the fraction of simulations for which the RV test has the correct sign, that is, the true model of joint profit maximization (JPM) with potentially misspecified demand fits the data better than the wrong model with misspecified demand. We find that the RV statistic has the correct sign in virtually all simulations, showing that the RV test may still be able to conclude for the true JPM model with misspecified demand.

APPENDIX G: ADDITIONAL EMPIRICAL DETAILS

This Appendix provides additional details for our empirical application.

Code details

Figure G.1 below provides the definition of the demand estimation problem in `PyBLP` and the testing problem in `pyRVtest`.

Panel A. Demand estimation code

```
>>> problem = pyblp.Problem(
...             product_formulations = (
...             pyblp.Formulation('prices+size+light+flavor_Plain+log_nflavors',
...                               absorb='brand+quarter+C(store)'
...                               ),
...             pyblp.Formulation('0+prices+light+size')
...             ),
...             agent_formulation = pyblp.Formulation('0 + log_income + age'),
...             product_data = product_data,
...             agent_data = agent_data
...             )
Initializing the problem ...
Absorbing demand-side fixed effects ...
Initialized the problem after 00:00:23.

Dimensions:
=======================================================
 T      N      F      I      K1    K2    D    MD    ED
----  ------  ---  -------  ----  ----  ---  ----  ----
5034  205123  105  5034000   5     3     2    10    3
=======================================================

Formulations:
===================================================================================
       Column Indices:             0         1      2          3            4
-----------------------------  ----------  -----  -----  ------------  ------------
 X1: Linear Characteristics      prices     size  light  flavor_Plain  log_nflavors
X2: Nonlinear Characteristics    prices    light  size
       d: Demographics         log_income   age
===================================================================================
```

Panel B. Testing code

```
>>> testing_problem = pyRVtest.Problem(
...     cost_formulation = (
...     pyRVtest.Formulation('1+freight_cost', absorb = 'C(brand)+C(quarter)+C(city)' )
...     ),
...     instrument_formulation = (
...     pyRVtest.Formulation('0+n_prod_own+n_prod_oppo'),
...     pyRVtest.Formulation('0+incxlight+agexlight+agexsize'),
...     pyRVtest.Formulation('0+cost_instr0'),
...     pyRVtest.Formulation('0+diff_instr0_pca+diff_instr1_pca+diff_instr2_pca+diff_instr3_pca+ diff_instr4_pca')
...     ),
...     model_formulations = (
...         pyRVtest.ModelFormulation(model_downstream='monopoly', ownership_downstream='firm_ids'),
...         pyRVtest.ModelFormulation(model_downstream='bertrand', ownership_downstream='firm_ids'),
...         pyRVtest.ModelFormulation(model_downstream='monopoly', ownership_downstream='firm_ids',
...                         model_upstream='bertrand',  ownership_upstream='firm_ids'),
...         pyRVtest.ModelFormulation(model_downstream='monopoly', ownership_downstream='firm_ids',
...                           model_upstream='bertrand',  ownership_upstream='firm_ids',
...                                       vertical_integration='vertical_ids'
...                           )
...                         ),
...     product_data = product_data,
...     demand_results = pyblp_results
...     )
Initializing the problem ...
Initialized the problem after 00:00:04.

Dimensions:
=============================================
 T      N      M    L   d_z0  d_z1  d_z2  d_z3
----  ------  ---  ---  ----  ----  ----  ----
5034  205123   4    4    2     3     1     5
=============================================

Formulations:
==============================================================================================
Column Indices:          0               1               2               3               4
---------------  --------------  --------------  --------------  --------------  --------------
w: Marginal Cost   freight_cost
z0: Instruments     n_prod_own      n_prod_oppo
z1: Instruments      incxlight       agexlight        agexsize
z2: Instruments     cost_instr0
z3: Instruments   diff_instr0_pca diff_instr1_pca diff_instr2_pca diff_instr3_pca diff_instr4_pca
==============================================================================================

Models:
==============================================================
                          0         1         2          3
--------------------  --------  --------  --------  ------------
 Model - Downstream   monopoly  bertrand  monopoly    monopoly
  Model - Upstream      None      None    bertrand    bertrand
Firm IDs - Downstream monopoly  firm_ids  monopoly    monopoly
 Firm IDs - Upstream    None      None    firm_ids    firm_ids
      VI Index          None      None      None    vertical_ids
 Cost Scaling Column    None      None      None        None
      Unit Tax          None      None      None        None
    Advalorem Tax       None      None      None        None
   Advalorem Payer      None      None      None        None
User Supplied Markups   None      None      None        None
==============================================================
```

Panels A and B report the definitions of the demand estimation problem in `PyBLP` and the testing problem in `pyRVtest`, respectively.

FIGURE G.1. Code for demand estimation and testing. Panels A and B report the definitions of the demand estimation problem in `PyBLP` and the testing problem in `pyRVtest`, respectively.

Description of the models of conduct

Following Villas-Boas (2007), the markups in market t for a model m among those we consider can be written in the following form:

$$\boldsymbol{\Delta}_{mt} = \underbrace{-\left(\boldsymbol{\Omega}^{r}_{mt} \odot \boldsymbol{D}^{r}_{t}\right)^{-1} \boldsymbol{s}_t}_{=\boldsymbol{\Delta}^{\text{downstream}}_{mt}} \underbrace{-\left(\boldsymbol{\Omega}^{w}_{mt} \odot \boldsymbol{D}^{w}_{t}\right)^{-1} \boldsymbol{s}_t}_{=\boldsymbol{\Delta}^{\text{upstream}}_{mt}},$$

where $\boldsymbol{\Omega}^{r}_{mt}$ and $\boldsymbol{\Omega}^{w}_{mt}$ are ownership matrices, $\boldsymbol{D}^{w}_{t}$ is the Jacobian of retail share $\boldsymbol{s}_t$ with respect to wholesale price, and $\boldsymbol{D}^{r}_{t}$ is the Jacobian of retail share with respect to retail price. The markup $\boldsymbol{\Delta}_{mt}$ implied by each model is the sum of downstream markups $\boldsymbol{\Delta}^{\text{downstream}}_{mt}$ and upstream markups $\boldsymbol{\Delta}^{\text{upstream}}_{mt}$. We can derive each model by using different assumptions on the ownership matrices:

1. *Zero wholesale margin*: Set $\boldsymbol{\Omega}^{w}_{mt}$ to a matrix of zeros, set $\boldsymbol{\Omega}^{r}_{mt}$ to a matrix of ones.
2. *Zero retail margin*: Set $\boldsymbol{\Omega}^{w}_{mt}$ to a matrix of zeros, and set $\boldsymbol{\Omega}^{r}_{mt}$ to a matrix with element (i, j) that is equal to one if products i and j are produced by the same manufacturer, and to zero otherwise.

TABLE G.1. Cox and EB test results.

	2	3	4
Cox Test			
1. Zero wholesale margin	−4.72, 3.99	4.34, −6.53	4.35, -6.55
2. Zero retail margin		9.47, −12.42	9.47, −12.46
3. Linear pricing			0.01, −0.01
4. Hybrid model			
EB Test			
1. Zero wholesale margin	253.16, 319.1	253.16, 338.14	253.16, 337.97
2. Zero retail margin		319.1, 338.14	319.1, 337.97
3. Linear pricing			338.14, 337.97
4. Hybrid model			

Note: Each cell reports the respective pair of test statistics T_i, T_j for row model i and column model j, with NoProd instruments. For 95% confidence, the critical value for Cox is ±1.95, and EB is 5.98. Standard errors account for two-step estimation error and clustering at the market level; see Appendix C.

3. *Linear pricing*: Set $\mathbf{\Omega}^r_{mt}$ to a matrix of ones, and set $\mathbf{\Omega}^w_{mt}$ to a matrix with element (i, j) that is equal to one if products i and j are produced by the same manufacturer, and to zero otherwise.
4. *Hybrid model*: Set $\mathbf{\Omega}^r_{mt}$ to a matrix of ones, and set $\mathbf{\Omega}^w_{mt}$ to a matrix with element (i, j) that is equal to one if products i and j are produced by the same manufacturer and i is not a private label, and to zero otherwise.
5. *Wholesale collusion*: Set $\mathbf{\Omega}^r_{mt}$ and $\mathbf{\Omega}^w_{mt}$ to matrices of ones.

Assessment test results

We present in Table G.1 the test results from the two other model assessment procedures defined in Appendix D. With NoProd instruments, the EB and Cox tests reject all models. Similar to the AR results in Table 4, the Cox and EB results illustrate the pitfalls of using model assessment tests for conduct, formally discussed in Section 4.2.

Pooling all instrument sets

We present in Table G.2 the test results obtained by pooling all instrument sets used in Section 7. Two observations emerge from the table. First, the F-statistics with the pooled instruments, although above the corresponding critical values for 95% best-case power, are below those for the NoProd instrument set. This confirms our observation that pooling instruments may result in lower F-statistics, and potentially weak instruments. Second, the MCS for both sets of pooled instruments now includes not only model 2, but also model 1 at a confidence level $\alpha = 0.05$. Hence, pooled instrument sets lead to less sharp conclusions in this example.

TABLE G.2. RV test results with alternative instruments.

Models	T^{RV} 2	T^{RV} 3	T^{RV} 4	F-statistics 2	F-statistics 3	F-statistics 4	MCS p-values
IVs: No. Products + Demo + Cost + Diff ($d_z = 11$)							
1. Zero wholesale margin	1.664	−8.827	−8.838	105.3	70.7	70.8	0.096
2. Zero retail margin		−5.655	−5.653		85.6	86.2	1.0
3. Linear pricing			−1.781			101.1	0.00
4. Hybrid model							0.00

Note: Table reports the RV test statistics T^{RV} and the effective F-statistic for all pairs of models, and the MCS p-values (details on their computation are in Appendix F). A negative RV test statistic suggests better fit of the row model. With MCS p-values below 0.05 a row model is rejected from the model confidence set. All F-statistics exceed the appropriate critical values for a worst-case size of 0.075 and best-case power of 0.95. Both T^{RV} and the F-statistics account for two-step estimation error and clustering at the market level; see Appendix C for details.

APPENDIX H: MCS DETAILS

In this Appendix, we provide further details on the construction of the model confidence set. The construction is iterative where in each step a p-value for a model of worst fit is computed. For a set of instruments z_ℓ, the MCS M_ℓ^* is the collection of models with a p-value above the significance level.

For a researcher testing a set of candidate models M, we now describe the construction of the MCS p-values at each step of the algorithm. At iteration k, there are $M_k \subset M$ models under current consideration. Denoting $|M_k|$ as the cardinality of M_k, this results in $K = \binom{|M_k|}{2}$ distinct pairs of models and, therefore, RV test statistics. To make notation compact, we define Π as a one-to-one mapping from unique model pairs to $\{1, \dots, K\}$, and let $T^{\text{RV}}_{\Pi(m_1,m_2)} = \sqrt{n}(\hat{Q}_{m_1} - \hat{Q}_{m_2})/\hat{\sigma}_{\text{RV},\Pi(m_1,m_2)}$. In this iteration, we find the pair of models (m_1, m_2) associated with the largest test statistic in magnitude, denoted $\bar{T}^{\text{RV}}$. If $\bar{T}^{\text{RV}}$ is positive (negative), then m_1 (m_2) is the model of worst fit for which we compute the MCS p-value. This model will be dropped from M_{k+1} in the next iteration.

For computation of the MCS p-values, we utilize that $(T_1^{\text{RV}}, \dots, T_K^{\text{RV}})'$ is asymptotically normal with zero mean and a variance Σ under no degeneracy and a null of equal fit. This observation follows by extending Lemma A.1 to $|M_k|$ models and alluding to a first-order Taylor approximation as in Remark 1. We estimate Σ using $\hat{\Sigma}$. The diagonal entries of $\hat{\Sigma}$ are all one. For any off-diagonal element $\hat{\Sigma}_{ij}$ with $(i, j) = (\Pi(m_1, m_2), \Pi(m_3, m_4))$, we have

$$\hat{\Sigma}_{ij} = 4\hat{\sigma}^{-1}_{\text{RV},i}\hat{\sigma}^{-1}_{\text{RV},j}\big(\hat{g}'_{m_1}\hat{W}^{1/2}\hat{V}^{\text{RV}}_{m_1m_3}\hat{W}^{1/2}\hat{g}_{m_3} - \hat{g}'_{m_2}\hat{W}^{1/2}\hat{V}^{\text{RV}}_{m_2m_3}\hat{W}^{1/2}\hat{g}_{m_3} \tag{92}$$

$$- \hat{g}'_{m_1}\hat{W}^{1/2}\hat{V}^{\text{RV}}_{m_1m_4}\hat{W}^{1/2}\hat{g}_{m_4} + \hat{g}'_{m_2}\hat{W}^{1/2}\hat{V}^{\text{RV}}_{m_2m_4}\hat{W}^{1/2}\hat{g}_{m_4}\big). \tag{93}$$

We take 99,999 draws from this distribution and for each draw compute the max of the absolute value of the K simulated test statistics denoted $\bar{T}^{\text{RV,sim}}$. The p-value is then computed as the fraction of draws for which $\bar{T}^{\text{RV,sim}} > |\bar{T}^{\text{RV}}|$.

This procedure is computationally expedient. In particular, one does not need to bootstrap both demand estimation and the testing procedure, which can be prohibitively costly. Instead, using the adjustments derived in Appendix C, $\hat{\Sigma}$ can incorporate adjustments for demand estimation and clustering.

Appendix I: Simultaneous versus sequential demand estimation: Implications for testing

Consider the sequential setting in the paper, where demand is estimated in a preliminary step, and then testing is performed with the RV test. The lack of fit for a model of conduct m is measured by the GMM objective function formed with the supply moments: $\hat{g}_m(\theta^D) = \frac{1}{n}\hat{z}'(\hat{p} - \hat{\Delta}_m(\theta^D))$, and the hat notation denotes that these are finite sample objects, residualized against the observed cost shifters $\mathbf{w}$. We augment the framework in the paper by allowing the supply moments to depend on demand parameters θ^D through markups $\Delta_m = \Delta_m(\theta^D)$. Demand moments are instead $\hat{h}(\theta^D) = \frac{1}{n}\mathbf{z}^{D\prime}\boldsymbol{\xi}(\theta^D)$, where $\boldsymbol{\xi}(\theta^D)$ is the value of the demand unobservable implied by θ^D, and $\mathbf{z}^D$ is a vector of demand instruments.[43] We assume that the dimension of $\mathbf{z}^D$ is sufficient to identify θ^D from demand moments alone.

We define the probability limit of the demand parameter estimates obtained from sequential estimation as

$$\theta^{\mathrm{Dseq}} = \arg\min_{\theta^D} Q^D(\theta^D),$$

where $Q^D(\theta^D) = h(\theta^D)' W^D h(\theta^D)$, $h(\theta^D) = \operatorname{plim} \hat{h}(\theta^D)$, and W^D is the probability limit of the weight matrix used in demand estimation. We construct the RV test statistic in a second step from a measure of fit based on supply moments. The asymptotic value of the measure of fit for each model m is

$$Q_m(\theta^{\mathrm{Dseq}}) = g_m(\theta^{\mathrm{Dseq}})' W g_m(\theta^{\mathrm{Dseq}}).$$

Alternatively, we could pursue a simultaneous approach, where we use the supply moments to estimate demand under each model of conduct. Stacking the demand and supply moments, the simultaneous GMM objective function is

$$Q_m^{\mathrm{sim}}(\theta^D) = \big(g_m(\theta^D)', h(\theta^D)'\big) W^{\mathrm{sim}} \big(g_m(\theta^D)', h(\theta^D)'\big)'$$

We maintain the following assumption on W^{sim} for tractability.

Assumption 5. W^{sim} is block diagonal, so that $W^{\mathrm{sim}} = \mathrm{diag}(W, W^D)$.

The probability limit of the demand parameters obtained from simultaneous estimation is

$$\theta_m^{\mathrm{Dsim}} = \arg\min_{\theta^D} Q_m(\theta^D) + Q^D(\theta^D).$$

While the value of θ_m^{Dsim} is model specific, θ^{Dseq} is not. We now establish the following.

Proposition 5. *Let Assumption 5 hold, $\boldsymbol{w}$ be correctly specified, θ_m^{Dsim} be unique, and $\theta^{\mathrm{Dseq}} \neq \theta_m^{\mathrm{Dsim}}$ for model m. Then $E[(\Delta_0^z - \Delta_m^z(\theta_m^{\mathrm{Dsim}}))^2] < E[(\Delta_0^z - \Delta_m^z(\theta^{\mathrm{Dseq}}))^2]$.*

[43] Demand moments could include micromoments; our discussion in this section would still apply.

The consequences of the proposition depend on the model of conduct m and the testing environment. Suppose $m = 1$ is the true model and there is no misspecification. Then $\theta^{\text{Dseq}} = \theta_1^{\text{Dsim}}$, and $E[(\Delta_0^z - \Delta_1^z(\theta_1^{\text{Dsim}}))^2] = E[(\Delta_0^z - \Delta_1^z(\theta^{\text{Dseq}}))^2] = 0$. However, predicted markups for $m = 2$ also get closer to the true ones. It follows that $Q_1 - Q_2$ is smaller in magnitude under simultaneous estimation. Instead, when demand is misspecified, the proposition implies that the MSE in predicted markups is smaller with a simultaneous approach for each model. We summarize these findings as follows.

Corollary 2. *Let Assumption 5 hold, θ_m^{Dsim} be unique, and $\theta^{\text{Dseq}} \neq \theta_m^{\text{Dsim}}$ for model m.*

(i) *Without misspecification, the probability limit of the RV numerator is smaller in magnitude under simultaneous estimation, but has the same sign as under sequential estimation.*

(ii) *With misspecification, the sign and magnitude of probability limit of the RV numerator could differ under simultaneous and sequential estimation.*

While we focus on the numerator, the mode of demand estimation will also affect the denominator of T^{RV}. How these forces affect the power of RV is ambiguous even under our simplifying assumptions and we leave it to future research.

The demand estimation literature (see Conlon and Gortmaker (2020)) has discussed how simultaneous and sequential estimation affect estimates and standard errors for the demand parameters. Our results show that simultaneous estimation does not guarantee demand parameter estimates closer to the true demand parameters. Instead, the supply moments pull the demand parameters toward values, which improve the fit of the predicted markups under the assumed conduct model. Moreover, efficiency considerations that apply to estimation do not immediately apply to our setup as one of the models we consider is not correctly specified.

Appendix J: Proof of Appendix results

Remark 2. As a prologue to the proof of Lemma A.1, we remind the reader that the first- order properties of $\hat{W}^{-1} = n^{-1}\hat{z}'\hat{z}$ and the infeasible $\check{W}^{-1} = n^{-1}z'z$ are the same. This follows from the equality $n^{-1}\hat{z}'\hat{z} = n^{-1}z'\hat{z}$, which in turn leads to

$$\begin{aligned}\hat{W}^{-1} &= \check{W}^{-1} + \underbrace{n^{-1}z'\mathbf{w}}_{=O_p(n^{-1/2})}\underbrace{\left(E\left[\mathbf{w}'\mathbf{w}\right]^{-1}E\left[\mathbf{w}'z\right] - \left(\mathbf{w}'\mathbf{w}\right)^{-1}\mathbf{w}'z\right)}_{=O_p(n^{-1/2})} \\ &= \check{W}^{-1} + O_p\left(n^{-1}\right). \qquad (94)\end{aligned}$$

The same argument applied to $\hat{g}_m = n^{-1}\hat{z}'(\hat{p} - \hat{\Delta}_m)$ yields $\hat{g}_m = \check{g}_m + O_p(n^{-1})$ where $\check{g}_m = n^{-1}z'(p - \Delta_m)$.

Remark 3. For a matrix A with all singular values strictly below 1, the proof of Lemma A.1 relies on the binomial series expansion $(I + A)^{-1/2} = \sum_{j=0}^{\infty}\binom{-1/2}{j}A^j = I - \frac{1}{2}A + \frac{3}{8}A^2 - \frac{5}{16}A^3 + \cdots$, where the generalized binomial coefficient is $\binom{\alpha}{j} = (j!)^{-1}\prod_{k=1}^{j}(\alpha - k + 1)$.

PROOF OF LEMMA A.1. We prove parts (i) and (ii) in three steps and then comment on the modifications to these steps that derive (iii) and (iv). Step (1) shows that $\sum_{i=1}^n(\psi_{1i}', \psi_{2i}')'/\sqrt{n} \xrightarrow{d} N(0, V^{\mathrm{RV}})$ and $\check{V}_{\ell k}^{\mathrm{RV}} := \frac{1}{n}\sum_{i=1}^n \psi_{\ell i}\psi_{ki}' \xrightarrow{p} V_{\ell k}^{\mathrm{RV}}$ for $\ell, k \in \{1, 2\}$, step (2) establishes that $\sqrt{n}(\hat{W}^{1/2}\hat{g}_m - W^{1/2}g_m) - \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_{mi} = o_p(1)$ for $m \in \{1, 2\}$, and step (3) proofs that $\mathrm{trace}((\hat{V}_{\ell k}^{\mathrm{RV}} - \check{V}_{\ell k}^{\mathrm{RV}})'(\hat{V}_{\ell k}^{\mathrm{RV}} - \check{V}_{\ell k}^{\mathrm{RV}})) = o_p(1)$ for $\ell, k \in \{1, 2\}$. The combination of steps (1) and (2) establishes (i) while steps (1) and (3) yield (ii).

Step (1) From Assumption 2, parts (i) and (iii), it follows from the standard central limit theorem for i.i.d. data that $\frac{1}{\sqrt{n}}\sum_{i=1}^n(\psi_{1i}', \psi_{2i}')u \xrightarrow{d} N(0, u'V^{\mathrm{RV}}u)$ for any nonrandom $u \in \mathbb{R}^{2d_z}$ with $\|u\| = 1$. The Cramér–Wold device therefore yields $\frac{1}{\sqrt{n}}\sum_{i=1}^n(\psi_{1i}', \psi_{2i}')' \xrightarrow{d} N(0, V^{\mathrm{RV}})$. Additionally, a standard law of large numbers applied elementwise implies $\check{V}_{\ell k} \xrightarrow{p} V_{\ell k}$ for $\ell, k \in \{1, 2\}$.

Step (2) From standard variance calculations, it follows that

$$\check{W} - W = O_p\big(n^{-1/2}\big) \quad \text{and} \quad \check{g}_m - g_m = O_p\big(n^{-1/2}\big). \tag{95}$$

In turn, Equation (95) together with Remarks 2 and 3 implies that

$$\hat{W}^{1/2} - W^{1/2} = W^{1/4}\big(\big(I + W^{1/2}\big(\hat{W}^{-1} - W^{-1}\big)W^{1/2}\big)^{-1/2} - I\big)W^{1/4} \tag{96}$$

$$= -\frac{1}{2}W^{1/4}\big(W^{1/2}\big(\hat{W}^{-1} - W^{-1}\big)W^{1/2}\big)W^{1/4} + O_p\big(n^{-1}\big) \tag{97}$$

Combining Equations (95) and (96), we then arrive at

$$\sqrt{n}\big(\hat{W}^{1/2}\hat{g}_m - W^{1/2}g_m\big)$$

$$= W^{1/2}(\hat{g}_m - g_m) + \big(\hat{W}^{1/2} - W^{1/2}\big)g_m + O_p\big(n^{-1}\big) \tag{98}$$

$$= W^{1/2}(\hat{g}_m - g_m) - \frac{1}{2}W^{3/4}\big(\hat{W}^{-1} - W^{-1}\big)W^{3/4}g_m + O_p\big(n^{-1}\big) \tag{99}$$

$$= W^{1/2}(\check{g}_m - g_m) - \frac{1}{2}W^{3/4}\big(\check{W}^{-1} - W^{-1}\big)W^{3/4}g_m + O_p\big(n^{-1}\big) \tag{100}$$

$$= \frac{1}{n}\sum_{i=1}^n \psi_{mi} + O_p\big(n^{-1}\big). \tag{101}$$

Step (3) Letting $R_m = \frac{1}{n}\sum_{i=1}^n(\hat{\psi}_{mi} - \psi_{mi})'(\hat{\psi}_{mi} - \psi_{mi})$, it follows from matrix analogs of the Cauchy–Schwarz inequality that

$$\mathrm{trace}\big(\big(\hat{V}_{\ell k}^{\mathrm{RV}} - \check{V}_{\ell k}^{\mathrm{RV}}\big)'\big(\hat{V}_{\ell k}^{\mathrm{RV}} - \check{V}_{\ell k}^{\mathrm{RV}}\big)\big) \leq 4\big(\mathrm{trace}\big(\check{V}_{\ell\ell}^{\mathrm{RV}}\big)R_k + \mathrm{trace}\big(\check{V}_{kk}^{\mathrm{RV}}\big)R_\ell + R_\ell R_k\big). \tag{102}$$

Therefore, it suffices to show that $R_m = o_p(1)$ for $m \in \{1, 2\}$, since we have from step (1) that $\check{V}_{mm}^{\mathrm{RV}} = O_p(1)$ for $m \in \{1, 2\}$. To further compartmentalize the problem, we note that

$$R_m \leq \frac{3}{n}\sum_{i=1}^n \big\|\hat{W}^{1/2}\hat{z}_i(\hat{p}_i - \hat{\Delta}_{mi}) - W^{1/2}z_i(p_i - \Delta_{mi})\big\|^2 \tag{103}$$

$$+\frac{3}{4n}\sum_i \|\hat{W}^{3/4}\hat{z}_i\hat{z}_i'\hat{W}^{3/4}\hat{g}_m - W^{3/4}z_iz_i'W^{3/4}g_m\|^2$$
$$+\frac{3}{4}\|\hat{W}^{1/2}\hat{g}_m - W^{1/2}g_m\|^2. \tag{104}$$

Argumentation analogous to Remark 2 combined with Equation (95) yields $R_m = o_p(1)$.

Finally, note that to derive (iii) and (iv), one can follow the same line of argument. The only real difference is that Equation (96) gets replaced by

$$\hat{W} - W = -W^{-1}(\hat{W}^{-1} - W^{-1})W^{-1} + O_p(n^{-1}). \qquad \square$$

As a prelude to the proof of Lemma 2, we establish the following analogous result.

LEMMA 3. *Suppose that Assumptions* 1, 2, *and ND are satisfied. Then, with probability approaching one as* $n \to \infty$,

(i) RV *rejects the null of equal fit in favor of model* 1 *if* $E[(\Delta_{0i}^z - \Delta_{1i}^z)^2] < E[(\Delta_{0i}^z - \Delta_{2i}^z)^2]$;

(ii) AR *rejects the null of perfect fit for model* m *with* $E[(\Delta_{0i}^z - \Delta_{mi}^z)^2] \neq 0$.

PROOF OF LEMMA 3. For (i), suppose for concreteness that $E[(\Delta_{0i}^z - \Delta_{1i}^z)^2] < E[(\Delta_{0i}^z - \Delta_{2i}^z)^2]$. We need to show that $\Pr(T^{\mathrm{RV}} < -q_{1-\alpha/2}(N)) \to 1$ where $q_{1-\alpha/2}(N)$ is the $(1-\alpha/2)$-th quantile of a standard normal distribution. From Proposition 1, part (ii), we have $Q_1 < Q_2$. Lemma A.1, parts (i) and (ii), together with Remark 1 yield $T^{\mathrm{RV}}/\sqrt{n} = \frac{\hat{Q}_1 - \hat{Q}_2}{\hat{\sigma}_{\mathrm{RV}}} \xrightarrow{p} \frac{Q_1 - Q_2}{\sigma_{\mathrm{RV}}} < 0$. Therefore, $\Pr(T^{\mathrm{RV}}/\sqrt{n} < -q_{1-\alpha/2}(N_{0,1})/\sqrt{n}) \to 1$.

For (ii), suppose that $E[(\Delta_{0i}^z - \Delta_{mi}^z)^2] \neq 0$ for some $m \in \{1, 2\}$. We need to show that $\Pr(\text{reject } H_{0,m}^{\mathrm{AR}}) \to 1$. From Proposition 1, part (i), it follows that $\pi_m \neq 0$ and since V_{mm}^{AR} is positive definite we have $\pi_m'(V_{mm}^{\mathrm{AR}})^{-1}\pi_m > 0$. Using Lemma A.1, parts (iii) and (iv), we have $(\hat{\pi}_m, \hat{V}_{mm}^{\mathrm{AR}}) \xrightarrow{p} (\pi_m, V_{mm}^{\mathrm{AR}})$ so that $T_m^{\mathrm{AR}}/n = \hat{\pi}_m'(\hat{V}_{mm}^{\mathrm{AR}})^{-1}\hat{\pi}_m \xrightarrow{p} \pi_m'(V_{mm}^{\mathrm{AR}})^{-1}\pi_m > 0$. Therefore, $\Pr(\text{reject } H_{0,m}^{\mathrm{AR}}) = \Pr(T_m^{\mathrm{AR}}/n > q_{1-\alpha}(\chi_{d_z}^2)/n) \to 1$ where $q_{1-\alpha}(\chi_{d_z}^2)/n) \to 1$ where $q_{d_z}(1-\alpha)$ denotes the $(1-\alpha)$-th quantile of a $\chi_{d_z}^2$ distribution. $\square$

PROOF OF LEMMA 2. Since $\mathbf{w_a}$ is a subset of $\mathbf{w}$, the proof follows immediately by replacing any variable y residualized with respect to $\mathbf{w}$ in the proof of Lemma 3, with $y^{\mathbf{a}}$, which has been residualized with respect to $\mathbf{w_a}$. $\square$

PROOF OF PROPOSITION 5. By definition of θ_m^{Dsim} and because θ^{Dseq} is the minimizer of Q^{dem}, it must be that $Q_m^{\mathrm{sim}}(\theta_m^{\mathrm{Dsim}}) < Q_m^{\mathrm{sim}}(\theta^{\mathrm{Dseq}})$. In turn, this implies that $Q_m(\theta_m^{\mathrm{Dsim}}) < Q_m(\theta^{\mathrm{Dseq}})$—if this inequality did not hold, we would need to have $Q^D(\theta^{\mathrm{Dseq}}) > Q^D(\theta_m^{\mathrm{Dsim}})$. Since each objective function is assumed to have a unique minimizer, this last inequality is a contradiction. Because we assumed that cost shifters are correctly specified, we have $Q_m(\Delta(\theta^D)) = E[(\Delta_0^z - \Delta_m^z(\theta^D))^2]$. Therefore,

$$E\big[\big(\Delta_0^z - \Delta_m^z(\theta_m^{\mathrm{Dsim}})\big)^2\big] < E\big[\big(\Delta_0^z - \Delta_m^z(\theta^{\mathrm{Dseq}})\big)^2\big]. \qquad \square$$

References

Berry, Steven and Philip Haile (2014), "Identification in differentiated products markets using market level data." *Econometrica*, 82 (5), 1749–1797. [2, 3]

Bresnahan, Timothy (1982), "The oligopoly solution concept is identified." *Economics Letters*, 10, 87–92. [7, 8]

Conlon, Christopher and Jeff Gortmaker (2020), "Best practices for differentiated products demand estimation with pyblp." *RAND Journal of Economics*, 51 (4), 1108–1161. [20]

Cox, David (1961), "Tests of separate families of hypotheses." In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 105–123. [10]

Hall, Alastair and Denis Pelletier (2011), "Nonnested testing in models estimated via generalized method of moments." *Econometric Theory*, 27 (2), 443–456. [2]

Hansen, Bruce and Seojeong Lee (2019), "Asymptotic theory for clustered samples." *Journal of Econometrics*, 210 (2), 268–290. [9]

Lau, Lawrence (1982), "On identifying the degree of competitiveness from industry price and output data." *Economics Letters*, 10 (1-2), 93–99. [7]

Magnolfi, Lorenzo, Daniel Quint, Christopher Sullivan, and Sarah Waldfogel (2022), "Falsifying models of firm conduct." Technical report, Working paper. [8]

Pakes, Ariel (2017), "Empirical tools and competition analysis: Past progress and current problems." *International Journal of Industrial Organization*, 53, 241–266. [10]

Rivers, Douglas and Quang Vuong (2002), "Model selection tests for nonlinear dynamic models." *Econometrics Journal*, 5, 1–39. [2]

Smith, Richard (1992), "Non-nested tests for competing models estimated by generalized method of moments." *Econometrica*, 60 (4), 973–980. [11]

Villas-Boas, Sofia (2007), "Vertical relationships between manufacturers and retailers: Inference with limited data." *Review of Economic Studies*, 74 (2), 625–652. [16]

Co-editor Stéphane Bonhomme handled this manuscript.

Manuscript received 16 January, 2023; final version accepted 8 May, 2024; available online 9 May, 2024.