Econometrica Supplementary Material

SUPPLEMENT TO "SPARSE MODELS AND METHODS FOR OPTIMAL INSTRUMENTS WITH AN APPLICATION TO EMINENT DOMAIN"

(*Econometrica*, Vol. 80, No. 6, November 2012, 2369–2429)

BY A. BELLONI, D. CHEN, V. CHERNOZHUKOV, AND C. HANSEN

THIS SUPPLEMENT PROVIDES technical results and proofs.

S1. TOOLS

S1.1. *Lyapunov CLT, Rosenthal Inequality, and Von Bahr–Esseen Inequality*

LEMMA S1—Lyapunov CLT: *Let* $\{X_{i,n}, i = 1, \ldots, n\}$ *be independent zero-mean random variables with variance* $s_{i,n}^2$, $n = 1, 2, \ldots$. *Define* $s_n^2 = \sum_{i=1}^n s_{i,n}^2$. *If, for some* $\mu > 0$, *Lyapunov's condition holds*:

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\mu}} \sum_{i=1}^n \mathrm{E}\big[|X_{i,n}|^{2+\mu}\big] = 0$$

then as n goes to infinity,

$$\frac{1}{s_n} \sum_{i=1}^n X_{i,n} \to_d \mathcal{N}(0,\ 1).$$

LEMMA S2—Rosenthal Inequality: *Let* $X_1, \ldots, X_n$ *be independent zero-mean random variables*; *then, for* $r \geq 2$,

$$\mathrm{E}\left(\left|\sum_{i=1}^n X_i\right|^r\right) \leq C(r) \max\left[\sum_{i=1}^n \mathrm{E}\big(|X_i|^r\big), \left\{\sum_{i=1}^n \mathrm{E}\big(X_i^2\big)\right\}^{r/2}\right].$$

This is due to Rosenthal (1970).

COROLLARY S1: *Let* $r \geq 2$, *and consider the case of independent zero-mean variables* X_i *with* $\mathrm{E}\mathbb{E}_n(X_i^2) = 1$ *and* $\mathrm{E}\mathbb{E}_n(|X_i|^r)$ *bounded by* C. *Then, for any* $\ell_n \to \infty$,

$$\Pr\left(\frac{\left|\sum_{i=1}^n X_i\right|}{n} > \ell_n n^{-1/2}\right) \leq \frac{2C(r)C}{\ell_n^r} \to 0.$$

 DOI: 10.3982/ECTA9626

To verify the corollary, we use Rosenthal's inequality $\mathrm{E}(|\sum_{i=1}^n X_i|^r) \leq Cn^{r/2}$, and the result follows by Markov inequality,

$$\mathrm{P}\left(\frac{\left|\sum_{i=1}^n X_i\right|}{n} > c\right) \leq \frac{C(r)Cn^{r/2}}{c^r n^r} \leq \frac{C(r)C}{c^r n^{r/2}}.$$

LEMMA S3—von Bahr–Essen Inequality: *Let $X_1, \dots, X_n$ be independent zero-mean random variables. Then, for $1 \leq r \leq 2$,*

$$\mathrm{E}\left(\left|\sum_{i=1}^n X_i\right|^r\right) \leq \left(2 - n^{-1}\right) \cdot \sum_{k=1}^n \mathrm{E}\left(|X_k|^r\right).$$

This result is due to von Bahr and Esseen (1965).

COROLLARY S2: *Let $r \in [1, 2]$, and consider the case of independent zero-mean variables X_i with $\mathrm{E}\mathbb{E}_n(|X_i|^r)$ bounded by C. Then, for any $\ell_n \to \infty$,*

$$\mathrm{P}\left\{\frac{\left|\sum_{i=1}^n X_i\right|}{n} > \ell_n n^{-(1-1/r)}\right\} \leq \frac{2C}{\ell_n^r} \to 0.$$

The corollary follow by Markov and von Bahr–Esseen's inequalities,

$$\mathrm{P}\left(\frac{\left|\sum_{i=1}^n X_i\right|}{n} > c\right) \leq \frac{C\mathrm{E}\left(\left|\sum_{i=1}^n X_i\right|^r\right)}{c^r n^r} \leq \frac{\sum_{i=1}^n \mathrm{E}(|X_i|^r)}{c^r n^r} \leq C\frac{\bar{\mathrm{E}}(|X_i|^r)}{c^r n^{r-1}}.$$

S1.2. *A Symmetrization-Based Probability Inequality*

Next we proceed to use symmetrization arguments to bound the empirical process. Let $\|f\|_{\mathbb{P}_n,2} = \sqrt{\mathbb{E}_n[f(Z_i)^2]}$, $\mathbb{G}_n(f) = \sqrt{n}\mathbb{E}_n[f(Z_i) - \mathrm{E}[Z_i]]$, and, for a random variable Z, let $q(Z, 1-\tau)$ denote its $(1-\tau)$-quantile. The proof follows standard symmetrization arguments.

LEMMA S4—Maximal Inequality via Symmetrization: *Let $Z_1, \dots, Z_n$ be arbitrary independent stochastic processes and $\mathcal{F}$ a finite set of measurable functions. For any $\tau \in (0, 1/2)$, and $\delta \in (0, 1)$ with probability at least $1 - 4\tau - 4\delta$, we* *have*

$$\sup_{f\in\mathcal{F}}\big|\mathbb{G}_n(f)\big| \le 4\sqrt{2\log\big(2|\mathcal{F}|/\delta\big)}q\Big(\sup_{f\in\mathcal{F}}\|f\|_{\mathbb{P}_n,2}, 1-\tau\Big)$$
$$\vee 2\sup_{f\in\mathcal{F}} q\big(\big|\mathbb{G}_n(f)\big|, 1/2\big).$$

PROOF: Let

$$e_{1n} = \sqrt{2\log\big(2|\mathcal{F}|/\delta\big)}q\Big(\max_{f\in\mathcal{F}}\sqrt{\mathbb{E}_n\big[f(Z_i)^2\big]}, 1-\tau\Big),$$
$$e_{2n} = \max_{f\in\mathcal{F}} q\bigg(\big|\mathbb{G}_n\big(f(Z_i)\big)\big|, \frac{1}{2}\bigg),$$

and the event $\mathcal{E} = \{\max_{f\in\mathcal{F}}\sqrt{\mathbb{E}_n[f^2(Z_i)]} \le q(\max_{f\in\mathcal{F}}\sqrt{\mathbb{E}_n[f^2(Z_i)]}, 1-\tau)\}$, which satisfies $\mathrm{P}(\mathcal{E}) \ge 1-\tau$. By the symmetrization Lemma 2.3.7 of van der Vaart and Wellner (1996) (by definition of e_{2n}, we have $\beta_n(x) \ge 1/2$ in Lemma 2.3.7), we obtain

$$\mathbb{P}\Big\{\max_{f\in\mathcal{F}}\big|\mathbb{G}_n\big(f(Z_i)\big)\big| > 4e_{1n}\vee 2e_{2n}\Big\}$$
$$\le 4\mathbb{P}\Big\{\max_{f\in\mathcal{F}}\big|\mathbb{G}_n\big(\varepsilon_i f(Z_i)\big)\big| > e_{1n}\Big\}$$
$$\le 4\mathbb{P}\Big\{\max_{f\in\mathcal{F}}\big|\mathbb{G}_n\big(\varepsilon_i f(Z_i)\big)\big| > e_{1n}|\mathcal{E}\Big\} + 4\tau,$$

where ε_i are independent Rademacher random variables, $\mathrm{P}(\varepsilon_i = 1) = \mathrm{P}(\varepsilon_i = -1) = 1/2$.

Thus, a union bound yields

$$\mathbb{P}\Big\{\max_{f\in\mathcal{F}}\big|\mathbb{G}_n\big(f(Z_i)\big)\big| > 4e_{1n}\vee 2e_{2n}\Big\}$$
$$\le 4\tau + 4|\mathcal{F}|\max_{f\in\mathcal{F}}\mathbb{P}\big\{\big|\mathbb{G}_n\big(\varepsilon_i f(Z_i)\big)\big| > e_{1n}|\mathcal{E}\big\}. \tag{S.1}$$

We then condition on the values of $Z_1, \dots, Z_n$ and $\mathcal{E}$, denoting the conditional probability measure as $\mathbb{P}_\varepsilon$. Conditional on $Z_1, \dots, Z_n$, by the Hoeffding inequality, the symmetrized process $\mathbb{G}_n(\varepsilon_i f(Z_i))$ is sub-Gaussian for the $L_2(\mathbb{P}_n)$ norm, namely, for $f \in \mathcal{F}$, $\mathbb{P}_\varepsilon\{|\mathbb{G}_n(\varepsilon_i f(Z_i))| > x\} \le 2\exp(-x^2/\{2\mathbb{E}_n[f^2(Z_i)]\})$. Hence, under the event $\mathcal{E}$, we can bound

$$\mathbb{P}_\varepsilon\big\{\big|\mathbb{G}_n\big(\varepsilon_i f(Z_i)\big)\big| > e_{1n}|Z_1, \dots, Z_n, \mathcal{E}\big\} \le 2\exp\big(-e_{1n}^2/[2\mathbb{E}_n\big[f^2(Z_i)\big]\big)$$
$$\le 2\exp\big(-\log(2|\mathcal{F}|/\delta)\big).$$

Taking the expectation over $Z_1, \dots, Z_n$ does not affect the right hand side bound. Plugging in this bound yields the result. *Q.E.D.*

S2. PROOF OF THEOREM 6

To show part (a), note that, by a standard argument,

$$\sqrt{n}(\tilde{\alpha} - \alpha) = M^{-1}\mathbb{G}_n[A_i \epsilon_i] + o_{\mathrm{P}}(1).$$

From the proof of Theorem 4, we have that

$$\sqrt{n}(\widehat{\alpha} - \alpha_a) = Q^{-1}\mathbb{G}_n\big[D(x_i)\epsilon_i\big] + o_{\mathrm{P}}(1).$$

The conclusion follows. The consistency of $\widehat{\Sigma}$ for Σ can be demonstrated similarly to the proof of consistency of $\widehat{\Omega}$ and $\widehat{Q}$ in the proof of Theorem 4.

To show part (b), let α denote the true value as before, which by assumption coincides with the estimand of the baseline IV estimator by the standard argument, $\tilde{\alpha} - \alpha = o_{\mathrm{P}}(1)$. The baseline IV estimator is consistent for this quantity. Under the alternative hypothesis, the estimand of $\widehat{\alpha}$ is

$$\begin{aligned}\alpha_a &= \bar{\mathrm{E}}\big[D(x_i)d_i'\big]^{-1}\bar{\mathrm{E}}\big[D(x_i)y_i\big] \\ &= \alpha + \bar{\mathrm{E}}\big[D(x_i)D(x_i)'\big]^{-1}\bar{\mathrm{E}}\big[D(x_i)\epsilon_i\big].\end{aligned}$$

Under the alternative hypothesis, $\|\bar{\mathrm{E}}[D(x_i)\epsilon_i]\|_2$ is bounded away from zero uniformly in n. Hence, since the eigenvalues of Q are bounded away from zero uniformly in n, $\|\alpha - \alpha_a\|_2$ is also bounded away from zero uniformly in n. Thus, it remains to show that $\widehat{\alpha}$ is consistent for α_a. We have that

$$\widehat{\alpha} - \alpha_a = \mathbb{E}_n\big[\widehat{D}(x_i)d_i'\big]^{-1}\mathbb{E}_n\big[\widehat{D}(x_i)\epsilon_i\big] - \bar{\mathrm{E}}\big[D(x_i)D(x_i)'\big]^{-1}\bar{\mathrm{E}}\big[D(x_i)\epsilon_i\big],$$

so that

$$\begin{aligned}\|\widehat{\alpha} - \alpha_a\|_2 &\le \big\|\mathbb{E}_n\big[\widehat{D}(x_i)d_i'\big]^{-1} - \bar{\mathrm{E}}\big[D(x_i)D(x_i)'\big]^{-1}\big\| \big\|\mathbb{E}_n\big[\widehat{D}(x_i)\epsilon_i\big]\big\|_2 \\ &\quad + \big\|\bar{\mathrm{E}}\big[D(x_i)D(x_i)\big]^{-1}\big\| \big\|\mathbb{E}_n\big[\widehat{D}(x_i)\epsilon_i\big] - \bar{\mathrm{E}}\big[D(x_i)\epsilon_i\big]\big\|_2 \\ &= o_{\mathrm{P}}(1),\end{aligned} \tag{S.2}$$

provided that (i) $\|\mathbb{E}_n[\widehat{D}(x_i)d_i']^{-1} - \bar{\mathrm{E}}[D(x_i)D(x_i)']^{-1}\| = o_{\mathrm{P}}(1)$, which is shown in the proof of Theorem 4; (ii) $\|\bar{\mathrm{E}}[D(x_i)D(x_i)]^{-1}\| = \|Q^{-1}\|$ is bounded from above uniformly in n, which follows from the assumption on Q in Theorem 4; and (iii)

$$\big\|\mathbb{E}_n\big[\widehat{D}(x_i)\epsilon_i\big] - \bar{\mathrm{E}}\big[D(x_i)\epsilon_i\big]\big\|_2 = o_{\mathrm{P}}(1), \quad \big\|\bar{\mathrm{E}}\big[D(x_i)\epsilon_i\big]\big\|_2 = O(1),$$

where $\|\bar{\mathrm{E}}[D(x_i)\epsilon_i]\|_2 = O(1)$ is assumed. To show the last claim, note that

$$\begin{aligned}&\big\|\mathbb{E}_n\big[\widehat{D}(x_i)\epsilon_i\big] - \bar{\mathrm{E}}\big[D(x_i)\epsilon_i\big]\big\|_2 \\ &\quad\le \big\|\mathbb{E}_n\big[\widehat{D}(x_i)\epsilon_i\big] - \mathbb{E}_n\big[D(x_i)\epsilon_i\big]\big\|_2 + \big\|\mathbb{E}_n\big[D(x_i)\epsilon_i\big] - \bar{\mathrm{E}}\big[D(x_i)\epsilon_i\big]\big\|_2 \\ &\quad\le \sqrt{k_e}\max_{1\le l\le k_e}\big\|D_l(x_i) - \widehat{D}_l(x_i)\big\|_{2,n}\|\epsilon_i\|_{2,n} + o_{\mathrm{P}}(1) = o_{\mathrm{P}}(1),\end{aligned}$$

since k_e is fixed, $\|\mathbb{E}_n[D(x_i)\epsilon_i] - \bar{\mathrm{E}}[D(x_i)\epsilon_i]\|_2 = o_{\mathrm{P}}(1)$ by von Bahr–Esseen inequality (von Bahr and Esseen (1965)) and SM, $\|\epsilon_i\|_{2,n} = O_{\mathrm{P}}(1)$ follows by the Markov inequality and assumptions on the moments of ϵ_i, and $\max_{1\le l\le k_e}\|D_l(x_i) - \widehat{D}_l(x_i)\|_{2,n} = o_{\mathrm{P}}(1)$ follows from Theorems 1 and 2. *Q.E.D.*

S3. PROOF OF THEOREM 7

We introduce additional superscripts a and b on all variables; and n gets replaced by either n_a or n_b. The proof is divided in steps. Assuming

$$\max_{1\le l\le k_e}\big\|\widehat{D}^k_{il} - D^k_{il}\big\|_{2,n_k} \lesssim_{\mathrm{P}} \sqrt{\frac{s\log(p\vee n)}{n}} = o_{\mathrm{P}}(1), \quad k = a, b, \tag{S.3}$$

Step 1 establishes bounds for the intermediary estimates on each subsample. Step 2 establishes the result for the final estimator $\widehat{\alpha}_{ab}$ and consistency of the matrices estimates. Finally, Step 3 establishes (S.3).

Step 1. We have that

$$\begin{aligned}\sqrt{n_k}(\widehat{\alpha}_k - \alpha_0) &= \mathbb{E}_{n_k}\big[\widehat{D}^k_i d^{k\prime}_i\big]^{-1}\sqrt{n}\mathbb{E}_{n_k}\big[\widehat{D}^k_i\epsilon^k_i\big] \\ &= \big\{\mathbb{E}_{n_k}\big[\widehat{D}^k_i d^{k\prime}_i\big]\big\}^{-1}\big(\mathbb{G}_{n_k}\big[D^k_i\epsilon^k_i\big] + o_{\mathrm{P}}(1)\big) \\ &= \big\{\mathbb{E}_{n_k}\big[D_iD_i'\big] + o_{\mathrm{P}}(1)\big\}^{-1}\big(\mathbb{G}_{n_k}\big[D^k_i\epsilon^k_i\big] + o_{\mathrm{P}}(1)\big),\end{aligned}$$

since

$$\mathbb{E}_{n_k}\big[\widehat{D}^k_i d^{k\prime}_i\big] = \mathbb{E}_{n_k}\big[D_iD_i'\big] + o_{\mathrm{P}}(1), \tag{S.4}$$

$$\sqrt{n}\mathbb{E}_{n_k}\big[\widehat{D}^k_i\epsilon^k_i\big] = \mathbb{G}_{n_k}\big[D^k_i\epsilon^k_i\big] + o_{\mathrm{P}}(1). \tag{S.5}$$

Indeed, (S.4) follows similarly to Step 2 in the proof of Theorems 4 and 5 and condition (S.3). The relation (S.5) follows from $\mathrm{E}[\epsilon^k_i|x^k_i] = 0$ for both $k = a$ and $k = b$, Chebyshev inequality, and

$$\begin{aligned}&\mathrm{E}\big[\big\|\sqrt{n}\mathbb{E}_{n_k}\big[\big(\widehat{D}^k_i - D^k_i\big)\epsilon^k_i\big]\big\|^2_2|x^k_i, i = 1,\dots,n, k^c\big] \\ &\quad\lesssim \sqrt{k_e}\max_{1\le l\le k_e}\big\|\big(\widehat{D}^k_{il} - D^k_{il}\big)\big\|^2_{2,n_k} \to_{\mathrm{P}} 0,\end{aligned}$$

where $\mathrm{E}[\cdot|x^k_i, i = 1,\dots,n, k^c]$ denotes the estimate computed conditional x^k_i, $i = 1,\dots,n$ and on the sample k^c, where $k^c = \{a, b\}\setminus k$. The bound follows from the fact that (a)

$$\widehat{D}_{il}^k - D_{il}^k = f(x_i^k)'(\widehat{\beta}_l^{k^c} - \beta_{0l}^{k^c}) - a_l(x_i^i), \quad 1 \le i \le n_k,$$

by Condition AS, where $(\widehat{\beta}_l^{k^c} - \beta_{0l}^{k^c})$ are independent of $\{\epsilon_i^k, 1 \le i \le n_k\}$, by the independence of the two subsamples k and k^c, (b) $\{\epsilon_i^k, x_i, 1 \le i \le n_k\}$ are independent across i and independent from the sample k^c, (c) $\{\epsilon_i^k, 1 \le i \le n_k\}$ have conditional mean equal to zero, conditional on x_i^k, $i = 1, \dots, n$, and have conditional variance bounded from above, uniformly in n, conditional on x_i^k, $i = 1, \dots, n$, by Condition SM, and that (d) $\max_{1 \le l \le k_e} \|\widehat{D}_{il}^k - D_{il}^k\|_{2,n_k} \to_{\mathrm{P}} 0$.

Using the same arguments as in Step 1 in the proof of Theorems 4 and 5,

$$\sqrt{n_k}(\widehat{\alpha}_k - \alpha_0) = (\mathbb{E}_{n_k}[D_i^k D_i^{k\prime}])^{-1} \mathbb{G}_{n_k}[D_i^k \epsilon_i^k] + o_{\mathrm{P}}(1) = O_{\mathrm{P}}(1).$$

Step 2. Now, putting together terms, we get

$$\begin{aligned}
\sqrt{n}(\widehat{\alpha}_{ab} - \alpha_0) &= \left((n_a/n)\mathbb{E}_{n_a}[\widehat{D}_i^a \widehat{D}_i^{a\prime}] + (n_b/n)\mathbb{E}_{n_b}[\widehat{D}_i^b \widehat{D}_i^{b\prime}]\right)^{-1} \\
&\quad \times \left((n_a/n)\mathbb{E}_{n_a}[\widehat{D}_i^a \widehat{D}_i^{a\prime}]\sqrt{n}(\widehat{\alpha}_a - \alpha_0)\right. \\
&\quad \left. + (n_b/n)\mathbb{E}_{n_b}[\widehat{D}_i^b \widehat{D}_i^{b\prime}]\sqrt{n}(\widehat{\alpha}_b - \alpha_0)\right) \\
&= \left((n_a/n)\mathbb{E}_{n_a}[D_i^a D_i^{a\prime}] + (n_b/n)\mathbb{E}_{n_b}[D_i^b D_i^{b\prime}]\right)^{-1} \\
&\quad \times \left((n_a/n)\mathbb{E}_{n_a}[D_i^a D_i^{a\prime}]\sqrt{n}(\widehat{\alpha}_a - \alpha_0)\right. \\
&\quad \left. + (n_b/n)\mathbb{E}_{n_b}[D_i^b D_i^{b\prime}]\sqrt{n}(\widehat{\alpha}_b - \alpha_0)\right) + o_{\mathrm{P}}(1) \\
&= \{\mathbb{E}_n[D_i D_i']\}^{-1} \times \{(1/\sqrt{2}) \times \mathbb{G}_{na}[D_i^a \epsilon_i^a] \\
&\quad + (1/\sqrt{2})\mathbb{G}_{nb}[D_i^b \epsilon_i^b]\} + o_{\mathrm{P}}(1) \\
&= \{\mathbb{E}_n[D_i D_i']\}^{-1} \times \mathbb{G}_n[D_i \epsilon_i] + o_{\mathrm{P}}(1),
\end{aligned}$$

where we are also using the fact that

$$\mathbb{E}_{n_k}[\widehat{D}_i^k \widehat{D}_i^{k\prime}] - \mathbb{E}_{n_k}[D_i^k D_i^{k\prime}] = o_{\mathrm{P}}(1),$$

which is shown similarly to the proofs given in Theorem 4 for showing that $\mathbb{E}_n[D_i D_i'] - \mathbb{E}_n[D_i D_i'] = o_{\mathrm{P}}(1)$. The conclusion of the theorem follows by an application of Lyapunov CLT, similarly to the proof of Theorem 4 in the main text.

Step 3. In this step, we establish (S.3). For every observation i and $l = 1, \dots, k_e$, by Condition AS, we have

$$\text{(S.6)} \qquad D_{il} = f_i' \beta_{l0} + a_l(x_i),$$

$$\|\beta_{l0}\|_0 \le s, \quad \max_{1 \le l \le k_e} \|a_l(x_i)\|_{2,n} \le c_s \lesssim_{\mathrm{P}} \sqrt{s/n}.$$

Under our conditions, the sparsity bound for Lasso by Lemma 9 implies that, for all $\delta_l^k = \widehat{\beta}_l^k - \beta_{l0}$, $k = a, b$, and $l = 1, \dots, k_e$,

$$\|\delta_l^k\|_0 \lesssim_{\mathrm{P}} s.$$

Therefore, by Condition SE, we have, for $M = \mathbb{E}_{n_k}[f_i^k f_i^{k\prime}]$, $k = a, b$, that with probability going to 1, for n large enough,

$$0 < \kappa' \leq \phi_{\min}(\|\delta_l^k\|_0)(M) \leq \phi_{\max}(\|\delta_l^k\|_0)(M) \leq \kappa'' < \infty,$$

where κ' and κ'' are some constants that do not depend on n. Thus,

$$\begin{aligned} \|D_{il}^k - \widehat{D}_{il}^k\|_{2,n_k} &= \|f_i^{k\prime}\beta_{l0} + a_l(x_i^k) - f_i^{k\prime}\widehat{\beta}_l^{k^c}\|_{2,n_k} \\ &= \|f_i^{k\prime}(\beta_{l0} - \widehat{\beta}_l^{k^c})\|_{2,n_k} + \|a_l(x_i^k)\|_{2,n_k} \\ &\leq \sqrt{\kappa''/\kappa'}\|f_i^{k^c\prime}(\widehat{\beta}_l^{k^c} - \beta_{l0})\|_{2,n_{k^c}} + c_s\sqrt{n/n_k}, \end{aligned} \tag{S.7}$$

where the last inequality holds with probability going to 1 by Condition SE imposed on matrices $M = \mathbb{E}_{n_k}[f_i^k f_i^{k\prime}]$, $k = a, b$, and by

$$\|a_l(x_i^k)\|_{2,n_k} \leq \sqrt{n/n_k}\|a_l(x_i)\|_{2,n} \leq \sqrt{n/n_k}c_s.$$

Then, in view of (S.6), $\sqrt{n/n_k} = \sqrt{2} + o(1)$, and condition $s \log p = o(n)$, the result (S.3) holds by (S.7) combined with Theorem 1 for Lasso and Theorem 2 for post-Lasso, which imply

$$\max_{1 \leq l \leq k_e} \|f_i^{k\prime}(\widehat{\beta}_l^k - \beta_{l0})\|_{2,n_k} \lesssim_{\mathrm{P}} \sqrt{\frac{s \log(p \vee n)}{n}} = o_{\mathrm{P}}(1), \quad k = a, b.$$

Q.E.D.

S4. PROOF OF LEMMA 3

Part 1. The first condition in Condition RF(iv) is assumed, and we omit the proof for the third condition since it is analogous to the proof for the second condition.

Note that $\max_{1 \leq j \leq p} \mathbb{E}_n[f_{ij}^4 v_{il}^4] \leq (\mathbb{E}_n[v_{il}^8])^{1/2} \max_{1 \leq j \leq p} (\mathbb{E}_n[f_{ij}^8])^{1/2} \lesssim_{\mathrm{P}} 1$, since $\max_{1 \leq j \leq p} \sqrt{\mathbb{E}_n[f_{ij}^8]} \lesssim_{\mathrm{P}} 1$ by assumption and $\max_{1 \leq l \leq k_e} \sqrt{\mathbb{E}_n[v_{il}^8]} \lesssim_{\mathrm{P}} 1$ by the bounded k_e, Markov inequality, and the assumption that $\bar{\mathrm{E}}[v_{il}^8]$ are uniformly bounded in n and l.

Thus, applying Lemma S4,

$$\max_{1 \leq l \leq k_e, 1 \leq j \leq p} |\mathbb{E}_n[f_{ij}^2 v_{il}^2] - \bar{\mathrm{E}}[f_{ij}^2 v_{il}^2]| \lesssim_{\mathrm{P}} \sqrt{\frac{\log(pk_e)}{n}} \lesssim_{\mathrm{P}} \sqrt{\frac{\log p}{n}} \to_{\mathrm{P}} 0.$$

Part 2. To show (a), we note that, by simple union bounds and tail properties of Gaussian variable, we have that $\max_{ij} f_{ij}^2 \lesssim_{\mathrm{P}} \log(p \vee n)$, so we need $\log(p \vee n)\frac{s\log(p\vee n)}{n} \to 0$.

Therefore, $\max_{1\le j\le p} \mathbb{E}_n[f_{ij}^4 v_{il}^4] \le \mathbb{E}_n[v_{il}^4] \max_{1\le i\le n, 1\le j\le p} f_{ij}^4 \lesssim_{\mathrm{P}} \log^2(p \vee n)$. Thus, applying Lemma S4,

$$\max_{1\le l\le k_e, 1\le j\le p} \left|\mathbb{E}_n\left[f_{ij}^2 v_{il}^2\right] - \bar{\mathrm{E}}\left[f_{ij}^2 v_{il}^2\right]\right| \lesssim_{\mathrm{P}} \log(p \vee n)\sqrt{\frac{\log(pk_e)}{n}}$$
$$\lesssim_{\mathrm{P}} \sqrt{\frac{\log^3(p \vee n)}{n}} \to_{\mathrm{P}} 0,$$

since $\log p = o(n^{1/3})$. The remaining moment conditions of Condition RF(ii) follow immediately from the definition of the conditionally bounded moments since, for any $m > 0$, $\bar{\mathrm{E}}[|f_{ij}|^m]$ is bounded, uniformly in $1 \le j \le p$, uniformly in n, for the i.i.d. Gaussian regressors of Lemma 1 of Belloni, Chen, Chernozhukov, and Hansen (2012). The proof of (b) for arbitrary bounded i.i.d. regressors of Lemma 2 of Belloni et al. (2012) is similar. *Q.E.D.*

S5. PROOF OF LEMMA 4

The first two conditions of Condition SM(iii) follow from the assumed rate $s^2 \log^2(p \vee n) = o(n)$ since we have $q_\epsilon = 4$. To show part (a), we note that, by simple union bounds and tail properties of Gaussian variable, we have that $\max_{1\le i\le n, 1\le j\le p} f_{ij}^2 \lesssim_{\mathrm{P}} \log(p \vee n)$, so we need $\log(p \vee n)\frac{s\log(p\vee n)}{n} \to 0$. Applying union bound, Gaussian concentration inequalities (Ledoux and Talagrand (1991)), and that $\log^2 p = o(n)$, we have $\max_{1\le j\le p} \mathbb{E}_n[f_{ij}^4] \lesssim_{\mathrm{P}} 1$. Thus Condition SM(iii)(c) holds by $\max_j \mathbb{E}_n[f_{ij}^2 \epsilon_i^2] \le (\mathbb{E}_n[\epsilon_i^4])^{1/2} \max_{1\le j\le p}(\mathbb{E}_n[f_{ij}^4])^{1/2} \lesssim_{\mathrm{P}} 1$. Part (b) follows because regressors are bounded and the moment assumption on ϵ. *Q.E.D.*

S6. PROOF OF LEMMA 11

Step 1. Here we consider the initial option, in which $\widehat{\gamma}_{jl}^2 = \mathbb{E}_n[f_{ij}^2(d_{il} - \mathbb{E}_n d_{il})^2]$. Let us define $\tilde{d}_{il} = d_{il} - \bar{\mathrm{E}}[d_{il}]$, $\tilde{\gamma}_{jl}^2 = \mathbb{E}_n[f_{ij}^2 \tilde{d}_{il}^2]$, and $\gamma_{jl}^2 = \bar{\mathrm{E}}[f_{ij}^2 \tilde{d}_{il}^2]$. We want to show that

$$\Delta_1 = \max_{1\le l\le k_e, 1\le j\le p} \left|\widehat{\gamma}_{jl}^2 - \tilde{\gamma}_{jl}^2\right| \to_{\mathrm{P}} 0,$$
$$\Delta_2 = \max_{1\le l\le k_e, 1\le j\le p} \left|\tilde{\gamma}_{jl}^2 - \gamma_{jl}^2\right| \to_{\mathrm{P}} 0,$$

which would imply that $\max_{1\le j\le p, 1\le l\le k_e} |\widehat{\gamma}_{jl}^2 - \gamma_{jl}^2| \to_{\mathrm{P}} 0$, and then, since γ_{jl}^2's are uniformly bounded from above by Condition RF and bounded below by $\gamma_{jl}^{02} =$

$\bar{\mathrm{E}}[f_{ij}^2 v_{il}^2]$, which are bounded away from zero. The asymptotic validity of the initial option then follows.

We have that $\Delta_2 \to_{\mathrm{P}} 0$ by Condition RF, and, since $\mathbb{E}_n[\tilde{d}_{il}] = \mathbb{E}_n[d_{il}] - \bar{\mathrm{E}}[d_{il}]$, we have

$$\begin{aligned} \Delta_1 &= \max_{1\le l\le k_e, 1\le j\le p} \big|\mathbb{E}_n\big[f_{ij}^2\big\{(\tilde{d}_{il} - \mathbb{E}_n\tilde{d}_{il})^2 - \tilde{d}_{il}^2\big\}\big]\big| \\ &\le \max_{1\le l\le k_e, 1\le j\le p} 2\big|\mathbb{E}_n\big[f_{ij}^2\tilde{d}_{il}\big]\mathbb{E}_n[\tilde{d}_{il}]\big| \\ &\quad + \max_{1\le l\le k_e, 1\le j\le p} \big|\mathbb{E}_n\big[f_{ij}^2\big](\mathbb{E}_n\tilde{d}_{il})^2\big| \to_{\mathrm{P}} 0. \end{aligned}$$

Indeed, we have for the first term that

$$\max_{1\le l\le k_e, 1\le j\le p} \big|\mathbb{E}_n\big[f_{ij}^2\tilde{d}_{il}\big]|\mathbb{E}_n[\tilde{d}_{il}]\big| \le \max_{i\le n, j\le p} |f_{ij}|\sqrt{\mathbb{E}_n\big[f_{ij}^2\tilde{d}_{il}^2\big]}O_{\mathrm{P}}(1/\sqrt{n}) \to_{\mathrm{P}} 0,$$

where we first used Holder inequality and $\max_{1\le l\le k_e} |\mathbb{E}_n[\tilde{d}_{il}]| \lesssim_{\mathrm{P}} \sqrt{k_e}/\sqrt{n}$ by the Chebyshev inequality and by $\bar{\mathrm{E}}[\tilde{d}_{il}^2]$ being uniformly bounded by Condition RF; then we invoked Condition RF to claim convergence to zero. Likewise, by Condition RF,

$$\max_{1\le l\le k_e, 1\le j\le p} \big|\mathbb{E}_n\big[f_{ij}^2\big](\mathbb{E}_n\tilde{d}_{il})^2\big| \le \max_{1\le j\le p} \big|f_{ij}^2\big| O_{\mathrm{P}}(1/n) \to_{\mathrm{P}} 0.$$

Step 2. Here we consider the refined option, in which $\widehat{\gamma}_{jl}^2 = \mathbb{E}_n[f_{ij}^2\widehat{v}_{il}^2]$. The residual here, $\widehat{v}_{il} = d_{il} - \widehat{D}_{il}$, can be based on any estimator that obeys

$$\max_{1\le l\le k_e} \|\widehat{D}_{il} - D_{il}\|_{2,n} \lesssim_{\mathrm{P}} \sqrt{\frac{s\log(p\vee n)}{n}}. \tag{S.8}$$

Such estimators include the Lasso and Post-Lasso estimators based on the initial option. Below we establish that the penalty levels, based on the refined option using any estimator obeying (S.8), are asymptotically valid. Thus by Theorems 1 and 2, the Lasso and Post-Lasso estimators based on the refined option also obey (S.8). This establishes that we can iterate on the refined option a bounded number of times, without affecting the validity of the approach.

Recall that $\widehat{\gamma}_{jl}^{02} = \mathbb{E}_n[f_{ij}^2 v_{il}^2]$ and define $\gamma_{jl}^{02} := \bar{\mathrm{E}}[f_{ij}^2 v_{il}^2]$, which is bounded away from zero and from above by assumption. Hence, by Condition RF, it suffices to show that $\max_{1\le j\le p, 1\le l\le k_e} |\widehat{\gamma}_{jl}^2 - \gamma_{jl}^{02}| \to_{\mathrm{P}} 0$, which implies the loadings are asymptotic valid with $u' = 1$. This, in turn, follows from

$$\Delta_1 = \max_{1\le l\le k_e, 1\le j\le p} \big|\widehat{\gamma}_{jl}^2 - \widehat{\gamma}_{jl}^{02}\big| \to_{\mathrm{P}} 0,$$

$$\Delta_2 = \max_{1\le l\le k_e, 1\le j\le p} \big|\widehat{\gamma}_{jl}^{02} - \gamma_{jl}^{02}\big|^2 \to_{\mathrm{P}} 0,$$

which we establish below.

Now note that we have proven $\Delta_2 \to_{\mathrm{P}} 0$ in Step 3 of the proof of Theorem 1. As for Δ_1, we note that

$$\Delta_1 \leq 2 \max_{1\leq l\leq k_e, 1\leq j\leq p} \left|\mathbb{E}_n\left[f_{ij}^2 v_{jl}(\widehat{D}_{il} - D_{il})\right]\right| + \max_{1\leq l\leq k_e, 1\leq j\leq p} \mathbb{E}_n\left[f_{ij}^2(\widehat{D}_{il} - D_{il})^2\right].$$

The first term is bounded by Holder and Lyapunov inequalities,

$$\max_{i\leq n, j\leq p} |f_{ij}|\left(\mathbb{E}_n\left[f_{ij}^2 v_{il}^2\right]\right)^{1/2} \max_{l\leq k_e} \|\widehat{D}_{il} - D_{il}\|_{2,n} \lesssim_{\mathrm{P}} \max_{i\leq n, j\leq p} |f_{ij}|\left(\mathbb{E}_n\left[f_{ij}^2 v_{il}^2\right]\right)^{1/2} \sqrt{\frac{s\log(p\vee n)}{n}} \to_{\mathrm{P}} 0,$$

where the conclusion is by Condition RF. The second term is of stochastic order

$$\max_{i\leq n, j\leq p} \left|f_{ij}^2\right| \frac{s\log(p\vee n)}{n} \to_{\mathrm{P}} 0,$$

which converges to zero by Condition RF. *Q.E.D.*

S7. ADDITIONAL SIMULATION RESULTS

In this section, we present simulation results to complement the results given in the paper. The simulations use the same model as the simulations in the paper:

$$y_i = \beta x_i + e_i,$$
$$x_i = z_i'\Pi + v_i,$$
$$(e_i, v_i) \sim N\left(0, \begin{pmatrix} \sigma_e^2 & \sigma_{ev} \\ \sigma_{ev} & \sigma_v^2 \end{pmatrix}\right) \text{ i.i.d.},$$

where $\beta = 1$ is the parameter of interest, and $z_i = (z_{i1}, z_{i2}, \dots, z_{i100})' \sim N(0, \Sigma_Z)$ is a 100×1 vector with $\mathrm{E}[z_{ih}^2] = \sigma_z^2$ and $\mathrm{Corr}(z_{ih}, z_{ij}) = 0.5^{|j-h|}$. In all simulations, we set $\sigma_e^2 = 2$ and $\sigma_z^2 = 0.3$.

For the other parameters, we consider various settings. We provide results for sample sizes, n, of 100, 250, and 500; and we consider three different values for $\mathrm{Corr}(e, v)$: 0, 0.3, and 0.6. We also consider four values of σ_v^2 that are chosen to benchmark four different strengths of instruments. The four values of σ_v^2 are found as $\sigma_v^2 = \frac{n\Pi'\Sigma_Z\Pi}{F^*\Pi'\Pi}$ for F^*: 2.5, 10, 40, and 160. We use two different designs for the first-stage coefficients, Π. The first sets the first S elements of

Π equal to 1 and the remaining elements equal to zero. We refer to this design as the "cut-off" design. The second model sets the coefficient on $z_{ih} = 0.7^{h-1}$ for $h = 1, \ldots, 100$. We refer to this design as the "exponential" design. In the cut-off case, we consider S of 5, 25, 50, and 100 to cover different degrees of sparsity.

For each setting of the simulation parameter values, we report results from seven different estimation procedures. A simple possibility when presented with many instrumental variables is to just estimate the model using 2SLS and all of the available instruments. It is well known that this will result in poor finite-sample properties unless there are many more observations than instruments; see, for example, Bekker (1994). The limited information maximum likelihood estimator (LIML) and its modification by Fuller (1977) (FULL)[1] are both robust to many instruments as long as the presence of many instruments is accounted for when constructing standard errors for the estimators; see Bekker (1994) and Hansen, Hausman, and Newey (2008), for example. We report results for these estimators in rows labeled 2SLS(100), LIML(100), and FULL(100), respectively.[2] For LASSO, we consider variable selection based on two different sets of instruments. In the first scenario, we use LASSO to select among the base 100 instruments and report results for the IV estimator based on the LASSO (LASSO) and Post-LASSO (Post-LASSO) forecasts. In the second, we use LASSO to select among 120 instruments formed by augmenting the base 100 instruments by the first 20 principal components constructed from the sampled instruments in each replication. We then report results for the IV estimator based on the LASSO (LASSO-F) and Post-LASSO (Post-LASSO-F) forecasts. In all cases, we use the refined data-dependent penalty loadings given in the paper.[3] For each estimator, we report root-truncated-mean-squared-error (RMSE),[4] median bias (Med. Bias), median absolute deviation (MAD), and rejection frequencies for 5% level tests (rp(0.05)).[5] For computing rejection frequencies, we estimate conventional 2SLS standard errors for 2SLS(100), LASSO, and Post-LASSO, and the many-instrument robust standard errors of Hansen, Hausman, and Newey (2008) for LIML(100) and FULL(100).

[1]Fuller (1977) required a user-specified parameter. We set this parameter equal to 1, which produces a higher-order unbiased estimator.

[2]With $n = 100$, we randomly select 99 instruments for use in FULL(100) and LIML(100).

[3]Specifically, we start by finding a value of λ for which LASSO selected only one instrument. We use this instrument to construct an initial set of residuals for use in defining the refined penalty loadings. Using these loadings, we run another LASSO step and select a new set of instruments. We then compute another set of residuals and use these residuals to recompute the loadings. We then run a final LASSO step. We report the results for the IV estimator of β based on the instruments chosen in this final LASSO step.

[4]We truncate the squared error at 1e12.

[5]In cases where LASSO selects no instruments, Med. Bias, and MAD use only the replications where LASSO selects a non-empty set of instruments, and we set the confidence interval equal to $(-\infty, \infty)$ and thus fail to reject.

TABLE S.I

2SLS SIMULATION RESULTS. EXPONENTIAL DESIGN. $N = 100$[a]

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
						$F^* = 2.5$									
2SLS(100)		0.037	0.000	0.025	0.056		0.108	0.103	0.103	0.852		0.205	0.202	0.202	1.000
LIML(100)		6.836	−0.051	0.345	0.024		8.442	0.135	0.378	0.121		32.289	0.181	0.318	0.230
FULL(100)		2.753	−0.051	0.345	0.024		2.792	0.135	0.377	0.121		1.973	0.181	0.316	0.230
LASSO	499	0.059	0.059	0.059	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	499	0.056	0.056	0.056	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	238	0.145	−0.004	0.084	0.022	231	0.151	0.085	0.105	0.058	261	0.184	0.154	0.155	0.174
Post-LASSO-F	238	0.136	0.003	0.081	0.018	231	0.144	0.081	0.101	0.060	261	0.177	0.146	0.147	0.188
						$F^* = 10$									
2SLS(100)		0.065	−0.003	0.043	0.058		0.176	0.168	0.168	0.758		0.335	0.335	0.335	0.998
LIML(100)		6.638	−0.039	0.479	0.038		228.717	0.113	0.527	0.122		14.189	0.298	0.588	0.210
FULL(100)		3.657	−0.039	0.479	0.038		3.889	0.113	0.526	0.122		4.552	0.298	0.588	0.210
LASSO	79	0.137	−0.008	0.087	0.034	77	0.136	0.047	0.081	0.056	93	0.142	0.051	0.094	0.104
Post-LASSO	79	0.132	−0.008	0.082	0.030	77	0.133	0.042	0.085	0.056	93	0.140	0.060	0.094	0.114
LASSO-F	9	0.152	−0.015	0.094	0.050	13	0.160	0.069	0.097	0.078	12	0.212	0.136	0.152	0.242
Post-LASSO-F	9	0.142	−0.010	0.080	0.048	13	0.150	0.068	0.094	0.092	12	0.198	0.129	0.144	0.244

(*Continues*)

TABLE S.I—*Continued*

Estimator	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 40$								
2SLS(100)		0.096	−0.004	0.068	0.044		0.217	0.193	0.193	0.530		0.391	0.381	0.381	0.988
LIML(100)		19.277	−0.037	0.599	0.022		21.404	0.070	0.718	0.082		6.796	0.158	0.607	0.140
FULL(100)		6.273	−0.037	0.598	0.020		6.567	0.070	0.718	0.082		5.444	0.158	0.607	0.140
LASSO	0	0.129	−0.001	0.086	0.042	0	0.135	0.016	0.091	0.042	0	0.137	0.042	0.094	0.088
Post-LASSO	0	0.126	−0.002	0.084	0.038	0	0.135	0.025	0.088	0.066	0	0.136	0.052	0.094	0.096
LASSO-F	0	0.127	0.001	0.087	0.050	0	0.134	0.029	0.093	0.056	0	0.148	0.073	0.100	0.114
Post-LASSO-F	0	0.122	0.000	0.081	0.042	0	0.132	0.038	0.090	0.056	0	0.147	0.082	0.101	0.136
							$F^* = 160$								
2SLS(100)		0.128	0.003	0.086	0.064		0.178	0.144	0.147	0.206		0.307	0.285	0.285	0.710
LIML(100)		79.113	0.014	0.540	0.038		45.371	−0.016	0.516	0.044		8.024	0.012	0.408	0.116
FULL(100)		10.092	0.014	0.540	0.038		17.107	−0.016	0.516	0.044		7.637	0.012	0.408	0.116
LASSO	0	0.139	0.001	0.093	0.054	0	0.129	0.009	0.087	0.048	0	0.136	0.023	0.087	0.086
Post-LASSO	0	0.139	−0.001	0.094	0.058	0	0.128	0.019	0.086	0.046	0	0.136	0.031	0.089	0.090
LASSO-F	0	0.138	0.000	0.092	0.036	0	0.128	0.018	0.090	0.044	0	0.139	0.038	0.089	0.094
Post-LASSO-F	0	0.137	−0.008	0.092	0.054	0	0.126	0.026	0.086	0.044	0	0.139	0.047	0.093	0.100

[a]Results are based on 500 simulation replications and 100 instruments. The first-stage coefficients were set equal to 0.7^{j-1} for $j = 1, \ldots, 100$ in this design. Corr(e, v) is the correlation between first-stage and structural errors. F^* measures the strength of the instruments as outlined in the text. 2SLS(100), LIML(100), and FULL(100) are respectively the 2SLS, LIML, and Fuller(1) estimator using all 100 potential instruments. Many-instrument robust standard errors are computed for LIML(100) and FULL(100) to obtain testing rejection frequencies. LASSO and Post-LASSO respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 100 instruments. LASSO-F and Post-LASSO-F respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 120 instruments formed by augmenting the original 100 instruments with the first 20 principal components. We report root-mean-squared-error (RMSE), median bias (Med. Bias), mean absolute deviation (MAD), and rejection frequency for 5% level tests (rp(0.05)). "Select 0" is the number of cases in which LASSO chose no instruments. In these cases, RMSE, Med. Bias, and MAD use only the replications where LASSO selects a non-empty set of instruments, and we set the confidence interval equal to $(-\infty, \infty)$ and thus fail to reject.

TABLE S.II

2SLS SIMULATION RESULTS. EXPONENTIAL DESIGN. $N = 250$[a]

Estimator	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 2.5$								
2SLS(100)		0.023	0.001	0.016	0.054		0.068	0.064	0.064	0.858		0.131	0.129	0.129	1.000
LIML(100)		28.982	0.009	0.158	0.032		5.126	0.034	0.152	0.050		2.201	0.072	0.145	0.130
FULL(100)		0.310	0.009	0.151	0.030		0.272	0.034	0.144	0.048		0.266	0.076	0.134	0.130
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	409	0.070	0.015	0.044	0.002	400	0.083	0.041	0.051	0.012	399	0.115	0.106	0.106	0.088
Post-LASSO-F	409	0.068	0.007	0.039	0.002	400	0.082	0.045	0.057	0.012	399	0.113	0.104	0.104	0.094
							$F^* = 10$								
2SLS(100)		0.042	0.000	0.027	0.064		0.112	0.104	0.104	0.774		0.211	0.208	0.208	1.000
LIML(100)		0.850	−0.014	0.137	0.036		0.759	0.020	0.126	0.034		1.111	0.015	0.119	0.076
FULL(100)		0.338	−0.014	0.134	0.036		0.255	0.021	0.123	0.034		0.271	0.019	0.114	0.076
LASSO	88	0.086	−0.001	0.060	0.036	90	0.079	0.026	0.054	0.040	78	0.092	0.046	0.067	0.128
Post-LASSO	88	0.083	−0.002	0.058	0.034	90	0.077	0.024	0.054	0.036	78	0.092	0.052	0.069	0.144
LASSO-F	33	0.096	−0.005	0.063	0.038	34	0.096	0.042	0.060	0.072	29	0.122	0.092	0.096	0.252
Post-LASSO-F	33	0.091	−0.004	0.057	0.040	34	0.092	0.043	0.055	0.058	29	0.115	0.086	0.090	0.282

(*Continues*)

TABLE S.II—*Continued*

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 40$								
2SLS(100)		0.066	0.001	0.045	0.072		0.130	0.115	0.115	0.476		0.249	0.244	0.244	0.994
LIML(100)		0.138	0.003	0.084	0.066		0.126	−0.008	0.082	0.046		0.115	−0.003	0.072	0.044
FULL(100)		0.137	0.003	0.084	0.066		0.125	−0.007	0.081	0.044		0.114	−0.001	0.072	0.044
LASSO	0	0.091	0.000	0.062	0.076	0	0.085	0.007	0.056	0.056	0	0.086	0.023	0.057	0.084
Post-LASSO	0	0.089	−0.002	0.061	0.080	0	0.084	0.012	0.054	0.060	0	0.085	0.031	0.061	0.098
LASSO-F	0	0.088	0.001	0.060	0.062	0	0.084	0.015	0.054	0.058	0	0.091	0.042	0.064	0.114
Post-LASSO-F	0	0.086	−0.001	0.059	0.068	0	0.081	0.021	0.053	0.064	0	0.092	0.052	0.065	0.134
							$F^* = 160$								
2SLS(100)		0.077	0.000	0.052	0.062		0.113	0.089	0.092	0.220		0.187	0.171	0.171	0.680
LIML(100)		0.096	0.000	0.064	0.060		0.090	0.006	0.062	0.044		0.092	0.002	0.058	0.060
FULL(100)		0.095	0.000	0.064	0.060		0.090	0.006	0.061	0.044		0.091	0.003	0.057	0.060
LASSO	0	0.083	0.005	0.056	0.072	0	0.081	0.003	0.056	0.050	0	0.084	0.014	0.057	0.068
Post-LASSO	0	0.082	0.006	0.054	0.070	0	0.080	0.004	0.053	0.052	0	0.085	0.018	0.058	0.076
LASSO-F	0	0.082	0.002	0.055	0.064	0	0.081	0.009	0.051	0.050	0	0.086	0.024	0.058	0.074
Post-LASSO-F	0	0.081	0.004	0.054	0.062	0	0.081	0.013	0.051	0.056	0	0.087	0.030	0.060	0.084

[a] Results are based on 500 simulation replications and 100 instruments. The first-stage coefficients were set equal to 0.7^{j-1} for $j = 1, \dots, 100$ in this design. Corr(e, v) is the correlation between first-stage and structural errors. F^* measures the strength of the instruments as outlined in the text. 2SLS(100), LIML(100), and FULL(100)are respectively the 2SLS, LIML, and Fuller(1) estimator using all 100 potential instruments. Many-instrument robust standard errors are computed for LIML(100) and FULL(100) to obtain testing rejection frequencies. LASSO and Post-LASSO respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 100 instruments. LASSO-F and Post-LASSO-F respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 120 instruments formed by augmenting the original 100 instruments with the first 20 principal components. We report root-mean-squared-error (RMSE), median bias (Med. Bias), mean absolute deviation (MAD), and rejection frequency for 5% level tests (rp(0.05)). "Select 0" is the number of cases in which LASSO chose no instruments. In these cases, RMSE, Med. Bias, and MAD use only the replications where LASSO selects a non-empty set of instruments, and we set the confidence interval equal to $(-\infty, \infty)$ and thus fail to reject.

TABLE S.III

2SLS SIMULATION RESULTS. EXPONENTIAL DESIGN. $N = 500$[a]

	$\mathrm{Corr}(e, v) = 0$					$\mathrm{Corr}(e, v) = 0.3$					$\mathrm{Corr}(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 2.5$								
2SLS(100)		0.016	0.002	0.010	0.050		0.047	0.045	0.045	0.828		0.093	0.093	0.093	1.000
LIML(100)		1.085	0.020	0.116	0.034		1.899	0.011	0.102	0.052		2.729	0.038	0.103	0.122
FULL(100)		0.188	0.019	0.103	0.030		0.165	0.013	0.093	0.046		0.166	0.042	0.092	0.122
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	461	0.046	0.001	0.026	0.000	463	0.054	0.011	0.035	0.002	467	0.080	0.065	0.065	0.028
Post-LASSO-F	461	0.045	0.003	0.030	0.000	463	0.048	0.011	0.033	0.004	467	0.074	0.067	0.067	0.030
							$F^* = 10$								
2SLS(100)		0.028	−0.002	0.019	0.052		0.079	0.074	0.074	0.758		0.152	0.150	0.150	1.000
LIML(100)		0.681	−0.012	0.085	0.026		2.662	0.003	0.082	0.040		7.405	0.004	0.077	0.068
FULL(100)		0.191	−0.011	0.083	0.020		0.182	0.005	0.079	0.036		0.154	0.008	0.073	0.070
LASSO	119	0.061	−0.005	0.038	0.036	105	0.059	0.016	0.041	0.040	106	0.066	0.037	0.048	0.124
Post-LASSO	119	0.059	−0.005	0.037	0.030	105	0.057	0.015	0.040	0.038	106	0.065	0.038	0.048	0.132
LASSO-F	46	0.063	−0.001	0.039	0.024	52	0.063	0.030	0.040	0.048	46	0.084	0.063	0.066	0.222
Post-LASSO-F	46	0.059	−0.004	0.037	0.026	52	0.059	0.027	0.039	0.060	46	0.078	0.061	0.064	0.236

(*Continues*)

TABLE S.III—*Continued*

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 40$								
2SLS(100)		0.045	0.002	0.030	0.074		0.095	0.085	0.085	0.532		0.176	0.172	0.172	0.988
LIML(100)		0.085	0.004	0.054	0.060		0.084	0.003	0.049	0.044		0.072	0.008	0.044	0.064
FULL(100)		0.084	0.004	0.053	0.058		0.083	0.004	0.049	0.044		0.072	0.010	0.044	0.064
LASSO	0	0.060	0.001	0.040	0.060	0	0.058	0.006	0.036	0.048	0	0.060	0.019	0.041	0.078
Post-LASSO	0	0.058	0.002	0.039	0.054	0	0.057	0.007	0.037	0.050	0	0.061	0.020	0.042	0.078
LASSO-F	0	0.058	0.000	0.040	0.052	0	0.059	0.010	0.040	0.048	0	0.064	0.031	0.045	0.112
Post-LASSO-F	0	0.057	0.001	0.038	0.050	0	0.057	0.013	0.038	0.070	0	0.066	0.033	0.044	0.156
							$F^* = 160$								
2SLS(100)		0.054	0.004	0.036	0.066		0.082	0.063	0.064	0.246		0.134	0.124	0.124	0.684
LIML(100)		0.065	0.006	0.045	0.058		0.069	0.000	0.046	0.058		0.063	0.002	0.045	0.048
FULL(100)		0.065	0.006	0.045	0.054		0.068	0.001	0.045	0.058		0.063	0.003	0.045	0.048
LASSO	0	0.060	0.003	0.042	0.054	0	0.059	0.006	0.039	0.066	0	0.059	0.010	0.040	0.070
Post-LASSO	0	0.060	0.003	0.040	0.054	0	0.059	0.007	0.040	0.054	0	0.059	0.010	0.041	0.068
LASSO-F	0	0.060	0.004	0.042	0.058	0	0.060	0.010	0.039	0.066	0	0.061	0.017	0.041	0.082
Post-LASSO-F	0	0.059	0.003	0.040	0.058	0	0.059	0.012	0.039	0.062	0	0.061	0.021	0.041	0.090

[a]Results are based on 500 simulation replications and 100 instruments. The first-stage coefficients were set equal to 0.7^{j-1} for $j = 1, \dots, 100$ in this design. Corr(e, v) is the correlation between first-stage and structural errors. F^* measures the strength of the instruments as outlined in the text. 2SLS(100), LIML(100), and FULL(100) are respectively the 2SLS, LIML, and Fuller(1) estimator using all 100 potential instruments. Many-instrument robust standard errors are computed for LIML(100) and FULL(100) to obtain testing rejection frequencies. LASSO and Post-LASSO respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 100 instruments. LASSO-F and Post-LASSO-F respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 120 instruments formed by augmenting the original 100 instruments with the first 20 principal components. We report root-mean-squared-error (RMSE), median bias (Med. Bias), mean absolute deviation (MAD), and rejection frequency for 5% level tests (rp(0.05)). "Select 0" is the number of cases in which LASSO chose no instruments. In these cases, RMSE, Med. Bias, and MAD use only the replications where LASSO selects a non-empty set of instruments, and we set the confidence interval equal to $(-\infty, \infty)$ and thus fail to reject.

TABLE S.IV

2SLS SIMULATION RESULTS. CUT-OFF DESIGN, $S = 5$. $N = 100$[a]

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 2.5$								
2SLS(100)		0.026	0.000	0.018	0.054		0.077	0.073	0.073	0.812		0.146	0.144	0.144	1.000
LIML(100)		46.872	0.013	0.257	0.018		48.070	0.067	0.226	0.122		1.851	0.141	0.235	0.235
FULL(100)		2.298	0.013	0.257	0.018		1.447	0.067	0.226	0.122		1.307	0.141	0.235	0.235
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	395	0.088	−0.002	0.063	0.012	391	0.090	0.047	0.065	0.020	396	0.113	0.090	0.090	0.058
Post-LASSO-F	395	0.080	−0.002	0.048	0.010	391	0.085	0.045	0.053	0.026	396	0.110	0.090	0.092	0.072
							$F^* = 10$								
2SLS(100)		0.044	−0.001	0.032	0.040		0.121	0.110	0.110	0.728		0.224	0.221	0.221	1.000
LIML(100)		57.489	0.009	0.385	0.022		13.839	0.095	0.373	0.124		13.556	0.191	0.397	0.194
FULL(100)		2.170	0.009	0.385	0.022		5.258	0.095	0.373	0.124		2.386	0.191	0.396	0.194
LASSO	206	0.082	−0.008	0.055	0.028	232	0.086	0.027	0.058	0.034	215	0.084	0.046	0.059	0.068
Post-LASSO	206	0.081	−0.007	0.052	0.036	232	0.087	0.026	0.059	0.048	215	0.082	0.040	0.058	0.070
LASSO-F	27	0.089	−0.001	0.051	0.040	22	0.105	0.042	0.065	0.084	34	0.115	0.082	0.088	0.206
Post-LASSO-F	27	0.081	−0.001	0.047	0.050	22	0.096	0.040	0.060	0.090	34	0.109	0.079	0.081	0.224

(*Continues*)

TABLE S.IV—*Continued*

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 40$								
2SLS(100)		0.066	−0.007	0.046	0.064		0.126	0.110	0.110	0.432		0.229	0.221	0.221	0.948
LIML(100)		16.626	0.007	0.358	0.030		5.794	0.098	0.367	0.060		2.223	0.054	0.295	0.144
FULL(100)		4.540	0.007	0.358	0.030		5.405	0.098	0.367	0.060		2.129	0.054	0.295	0.144
LASSO	0	0.079	−0.009	0.053	0.040	0	0.079	0.012	0.051	0.046	0	0.084	0.019	0.056	0.080
Post-LASSO	0	0.078	−0.009	0.054	0.044	0	0.078	0.012	0.051	0.046	0	0.081	0.014	0.055	0.074
LASSO-F	0	0.080	−0.006	0.054	0.048	0	0.082	0.021	0.054	0.060	0	0.091	0.041	0.062	0.112
Post-LASSO-F	0	0.077	−0.009	0.054	0.046	0	0.080	0.023	0.052	0.056	0	0.087	0.041	0.060	0.116
							$F^* = 160$								
2SLS(100)		0.073	0.003	0.049	0.054		0.106	0.073	0.078	0.190		0.160	0.146	0.146	0.534
LIML(100)		2.958	0.037	0.246	0.034		9.206	0.010	0.268	0.048		23.035	0.007	0.237	0.064
FULL(100)		2.885	0.037	0.245	0.034		3.339	0.010	0.268	0.048		7.825	0.007	0.237	0.064
LASSO	0	0.077	0.001	0.050	0.046	0	0.079	0.005	0.053	0.060	0	0.074	0.005	0.049	0.036
Post-LASSO	0	0.076	0.000	0.051	0.044	0	0.079	0.005	0.054	0.058	0	0.074	0.004	0.049	0.036
LASSO-F	0	0.077	0.001	0.050	0.046	0	0.080	0.010	0.054	0.062	0	0.075	0.018	0.051	0.048
Post-LASSO-F	0	0.076	−0.001	0.049	0.046	0	0.079	0.010	0.054	0.060	0	0.075	0.014	0.051	0.044

[a]Results are based on 500 simulation replications and 100 instruments. The first 5 first-stage coefficients were set equal to 1 and the remaining 95 to zero in this design. Corr(e, v) is the correlation between first-stage and structural errors. F^* measures the strength of the instruments as outlined in the text. 2SLS(100), LIML(100), and FULL(100) are respectively the 2SLS, LIML, and Fuller(1) estimator using all 100 potential instruments. Many-instrument robust standard errors are computed for LIML(100) and FULL(100) to obtain testing rejection frequencies. LASSO and Post-LASSO respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 100 instruments. LASSO-F and Post-LASSO-F respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 120 instruments formed by augmenting the original 100 instruments with the first 20 principal components. We report root-mean-squared-error (RMSE), median bias (Med. Bias), mean absolute deviation (MAD), and rejection frequency for 5% level tests (rp(0.05)). "Select 0" is the number of cases in which LASSO chose no instruments. In these cases, RMSE, Med. Bias, and MAD use only the replications where LASSO selects a nonempty set of instruments, and we set the confidence interval equal to $(-\infty, \infty)$ and thus fail to reject.

TABLE S.V

2SLS SIMULATION RESULTS. CUT-OFF DESIGN, $S = 5$. $N = 250$[a]

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 2.5$								
2SLS(100)		0.016	0.001	0.010	0.040		0.049	0.047	0.047	0.844		0.094	0.093	0.093	1.000
LIML(100)		1.295	−0.001	0.088	0.028		1.612	0.016	0.100	0.044		10.351	0.026	0.082	0.128
FULL(100)		0.202	−0.001	0.085	0.026		0.215	0.017	0.095	0.040		0.165	0.029	0.079	0.126
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	491	0.049	−0.029	0.033	0.002	493	0.057	0.047	0.047	0.000	491	0.073	0.063	0.063	0.008
Post-LASSO-F	491	0.042	−0.010	0.027	0.002	493	0.056	0.045	0.045	0.004	491	0.073	0.056	0.056	0.008
							$F^* = 10$								
2SLS(100)		0.029	−0.002	0.019	0.056		0.075	0.071	0.071	0.718		0.141	0.139	0.139	0.998
LIML(100)		0.555	0.000	0.074	0.034		3.082	0.006	0.071	0.038		0.136	0.001	0.057	0.070
FULL(100)		0.157	0.000	0.073	0.034		0.163	0.007	0.070	0.038		0.121	0.004	0.056	0.072
LASSO	239	0.052	−0.002	0.037	0.024	253	0.051	0.020	0.036	0.038	249	0.050	0.029	0.038	0.050
Post-LASSO	239	0.049	−0.003	0.037	0.024	253	0.050	0.019	0.036	0.046	249	0.048	0.025	0.037	0.050
LASSO-F	83	0.059	−0.003	0.036	0.036	88	0.058	0.021	0.036	0.074	78	0.067	0.050	0.053	0.160
Post-LASSO-F	83	0.053	−0.005	0.036	0.032	88	0.055	0.022	0.036	0.076	78	0.061	0.048	0.049	0.170

(*Continues*)

TABLE S.V—*Continued*

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 40$								
2SLS(100)		0.040	0.001	0.027	0.054		0.079	0.070	0.070	0.416		0.142	0.138	0.138	0.964
LIML(100)		0.068	−0.002	0.043	0.048		0.069	0.000	0.046	0.044		0.058	−0.001	0.036	0.024
FULL(100)		0.067	−0.002	0.043	0.048		0.069	0.001	0.046	0.044		0.057	0.001	0.036	0.022
LASSO	0	0.048	0.002	0.032	0.042	0	0.049	0.005	0.034	0.042	0	0.048	0.014	0.035	0.058
Post-LASSO	0	0.048	0.001	0.032	0.036	0	0.049	0.004	0.033	0.052	0	0.048	0.012	0.033	0.056
LASSO-F	0	0.048	0.001	0.033	0.040	0	0.050	0.013	0.033	0.050	0	0.052	0.025	0.038	0.076
Post-LASSO-F	0	0.047	0.001	0.032	0.042	0	0.049	0.011	0.033	0.056	0	0.051	0.026	0.038	0.082
							$F^* = 160$								
2SLS(100)		0.045	−0.001	0.030	0.042		0.063	0.045	0.047	0.158		0.104	0.096	0.096	0.560
LIML(100)		0.054	0.000	0.036	0.048		0.054	0.000	0.036	0.048		0.051	0.001	0.035	0.036
FULL(100)		0.053	0.000	0.036	0.048		0.054	0.001	0.036	0.046		0.051	0.001	0.034	0.036
LASSO	0	0.047	0.001	0.030	0.048	0	0.048	0.004	0.032	0.048	0	0.049	0.007	0.033	0.050
Post-LASSO	0	0.047	0.000	0.031	0.046	0	0.048	0.004	0.031	0.042	0	0.049	0.007	0.032	0.050
LASSO-F	0	0.047	0.000	0.030	0.056	0	0.049	0.006	0.032	0.048	0	0.049	0.011	0.034	0.056
Post-LASSO-F	0	0.047	−0.001	0.030	0.050	0	0.048	0.005	0.031	0.052	0	0.049	0.013	0.033	0.058

[a]Results are based on 500 simulation replications and 100 instruments. The first 5 first-stage coefficients were set equal to 1 and the remaining 95 to zero in this design. Corr(e, v) is the correlation between first-stage and structural errors. F^* measures the strength of the instruments as outlined in the text. 2SLS(100), LIML(100), and FULL(100) are respectively the 2SLS, LIML, and Fuller(1) estimator using all 100 potential instruments. Many-instrument robust standard errors are computed for LIML(100) and FULL(100) to obtain testing rejection frequencies. LASSO and Post-LASSO respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 100 instruments. LASSO-F and Post-LASSO-F respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 120 instruments formed by augmenting the original 100 instruments with the first 20 principal components. We report root-mean-squared-error (RMSE), median bias (Med. Bias), mean absolute deviation (MAD), and rejection frequency for 5% level tests (rp(0.05)). "Select 0" is the number of cases in which LASSO chose no instruments. In these cases, RMSE, Med. Bias, and MAD use only the replications where LASSO selects a nonempty set of instruments, and we set the confidence interval equal to $(-\infty, \infty)$ and thus fail to reject.

TABLE S.VI

2SLS SIMULATION RESULTS. CUT-OFF DESIGN, $S = 5$. $N = 500$[a]

Estimator	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 2.5$								
2SLS(100)		0.012	0.001	0.008	0.052		0.034	0.032	0.032	0.838		0.066	0.065	0.065	1.000
LIML(100)		2.472	0.006	0.062	0.024		3.056	0.004	0.067	0.060		1.026	0.017	0.059	0.120
FULL(100)		0.122	0.006	0.058	0.022		0.127	0.005	0.059	0.056		0.106	0.021	0.055	0.122
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	498	0.040	0.039	0.039	0.000
Post-LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	498	0.037	0.037	0.037	0.000
							$F^* = 10$								
2SLS(100)		0.020	0.000	0.014	0.050		0.053	0.050	0.050	0.750		0.097	0.096	0.096	0.996
LIML(100)		0.205	0.001	0.047	0.044		0.127	0.004	0.048	0.056		0.072	−0.001	0.037	0.060
FULL(100)		0.100	0.001	0.047	0.040		0.088	0.005	0.047	0.050		0.067	0.002	0.038	0.060
LASSO	300	0.032	0.000	0.021	0.008	306	0.034	0.012	0.024	0.022	324	0.036	0.021	0.025	0.042
Post-LASSO	300	0.030	−0.001	0.021	0.014	306	0.034	0.014	0.024	0.030	324	0.034	0.019	0.023	0.044
LASSO-F	121	0.036	−0.002	0.023	0.020	136	0.038	0.017	0.026	0.042	118	0.043	0.030	0.033	0.140
Post-LASSO-F	121	0.034	0.002	0.019	0.016	136	0.037	0.016	0.026	0.070	118	0.041	0.029	0.031	0.154

(*Continues*)

TABLE S.VI—*Continued*

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 40$								
2SLS(100)		0.027	−0.003	0.019	0.044		0.056	0.048	0.048	0.426		0.099	0.097	0.097	0.950
LIML(100)		0.044	−0.001	0.030	0.036		0.043	0.001	0.029	0.034		0.041	−0.001	0.026	0.040
FULL(100)		0.044	−0.001	0.030	0.036		0.043	0.002	0.029	0.036		0.041	0.000	0.026	0.040
LASSO	0	0.034	0.000	0.022	0.048	0	0.033	0.009	0.023	0.040	0	0.035	0.009	0.023	0.066
Post-LASSO	0	0.034	−0.002	0.023	0.044	0	0.032	0.006	0.023	0.036	0	0.034	0.008	0.023	0.064
LASSO-F	0	0.034	0.000	0.023	0.050	0	0.034	0.011	0.025	0.040	0	0.036	0.016	0.024	0.086
Post-LASSO-F	0	0.034	−0.002	0.023	0.048	0	0.033	0.010	0.023	0.046	0	0.036	0.016	0.025	0.084
							$F^* = 160$								
2SLS(100)		0.033	−0.005	0.021	0.058		0.047	0.035	0.035	0.188		0.071	0.064	0.064	0.536
LIML(100)		0.037	−0.003	0.024	0.046		0.036	0.003	0.025	0.044		0.036	−0.001	0.024	0.040
FULL(100)		0.037	−0.003	0.024	0.046		0.036	0.003	0.025	0.044		0.036	0.000	0.024	0.040
LASSO	0	0.034	−0.004	0.025	0.048	0	0.034	0.005	0.025	0.030	0	0.034	0.003	0.024	0.048
Post-LASSO	0	0.034	−0.004	0.025	0.046	0	0.034	0.005	0.025	0.030	0	0.034	0.003	0.024	0.046
LASSO-F	0	0.034	−0.005	0.024	0.050	0	0.034	0.007	0.025	0.032	0	0.035	0.008	0.024	0.050
Post-LASSO-F	0	0.034	−0.005	0.023	0.052	0	0.034	0.007	0.025	0.032	0	0.035	0.009	0.023	0.052

[a]Results are based on 500 simulation replications and 100 instruments. The first 5 first-stage coefficients were set equal to 1 and the remaining 95 to zero in this design. Corr(e, v) is the correlation between first-stage and structural errors. F^* measures the strength of the instruments as outlined in the text. 2SLS(100), LIML(100), and FULL(100) are respectively the 2SLS, LIML, and Fuller(1) estimator using all 100 potential instruments. Many-instrument robust standard errors are computed for LIML(100) and FULL(100) to obtain testing rejection frequencies. LASSO and Post-LASSO respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 100 instruments. LASSO-Fand Post-LASSO-F respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 120 instruments formed by augmenting the original 100 instruments with the first 20 principal components. We report root-mean-squared-error (RMSE), median bias (Med. Bias), mean absolute deviation (MAD), and rejection frequency for 5% level tests (rp(0.05)). "Select 0" is the number of cases in which LASSO chose no instruments. In these cases, RMSE, Med. Bias, and MAD use only the replications where LASSO selects a nonempty set of instruments, and we set the confidence interval equal to $(-\infty, \infty)$ and thus fail to reject.

TABLE S.VII

2SLS SIMULATION RESULTS. CUT-OFF DESIGN, $S = 25$. $N = 100$[a]

Estimator	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 2.5$								
2SLS(100)		0.020	0.000	0.013	0.044		0.048	0.045	0.045	0.678		0.090	0.088	0.088	1.000
LIML(100)		4.259	0.013	0.150	0.020		5.353	0.027	0.141	0.102		1.394	0.063	0.137	0.205
FULL(100)		1.615	0.013	0.150	0.020		1.394	0.027	0.141	0.102		1.268	0.063	0.137	0.205
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	486	0.043	0.018	0.034	0.002	493	0.024	0.009	0.025	0.000	486	0.032	0.013	0.028	0.000
Post-LASSO-F	486	0.044	−0.011	0.032	0.000	493	0.022	−0.018	0.018	0.000	486	0.027	0.016	0.022	0.000
							$F^* = 10$								
2SLS(100)		0.026	−0.002	0.018	0.054		0.049	0.042	0.042	0.386		0.087	0.083	0.083	0.906
LIML(100)		1.269	−0.001	0.138	0.026		2.468	0.018	0.142	0.086		4.565	0.038	0.139	0.136
FULL(100)		1.181	−0.001	0.138	0.026		1.616	0.018	0.142	0.086		2.000	0.039	0.139	0.136
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	55	0.037	−0.005	0.024	0.038	46	0.038	0.006	0.023	0.058	46	0.041	0.011	0.027	0.074
Post-LASSO-F	55	0.035	−0.004	0.022	0.052	46	0.035	0.008	0.024	0.058	46	0.040	0.015	0.028	0.078

(*Continues*)

TABLE S.VII—*Continued*

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 40$								
2SLS(100)		0.031	−0.002	0.020	0.076		0.040	0.027	0.029	0.170		0.061	0.053	0.053	0.448
LIML(100)		3.549	0.007	0.088	0.034		2.640	0.021	0.092	0.040		1.916	0.008	0.078	0.088
FULL(100)		2.810	0.007	0.088	0.034		2.333	0.021	0.092	0.040		1.837	0.008	0.078	0.088
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	3	0.038	−0.001	0.025	0.058	1	0.036	0.007	0.024	0.054	0	0.040	0.011	0.028	0.086
Post-LASSO-F	3	0.036	−0.005	0.024	0.066	1	0.034	0.008	0.023	0.056	0	0.037	0.011	0.025	0.082
							$F^* = 160$								
2SLS(100)		0.032	−0.001	0.022	0.050		0.032	0.014	0.023	0.072		0.041	0.029	0.031	0.176
LIML(100)		2.393	0.005	0.055	0.020		1.946	0.001	0.058	0.022		1.295	−0.001	0.053	0.030
FULL(100)		2.375	0.005	0.055	0.020		1.932	0.001	0.058	0.022		1.289	−0.001	0.053	0.030
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	499	0.055	0.055	0.055	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	499	0.049	0.049	0.049	0.000
LASSO-F	0	0.038	−0.001	0.025	0.050	0	0.035	0.003	0.023	0.060	0	0.035	0.006	0.024	0.036
Post-LASSO-F	0	0.036	−0.003	0.024	0.060	0	0.032	0.002	0.021	0.046	0	0.034	0.005	0.022	0.048

[a]Results are based on 500 simulation replications and 100 instruments. The first 25 first-stage coefficients were set equal to 1 and the remaining 75 to zero in this design. Corr(e, v) is the correlation between first-stage and structural errors. F^* measures the strength of the instruments as outlined in the text. 2SLS(100), LIML(100), and FULL(100) are respectively the 2SLS, LIML, and Fuller(1) estimator using all 100 potential instruments. Many-instrument robust standard errors are computed for LIML(100) and FULL(100) to obtain testing rejection frequencies. LASSO and Post-LASSO respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 100 instruments. LASSO-F and Post-LASSO-F respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 120 instruments formed by augmenting the original 100 instruments with the first 20 principal components. We report root-mean-squared-error (RMSE), median bias (Med. Bias), mean absolute deviation (MAD), and rejection frequency for 5% level tests (rp(0.05)). "Select 0" is the number of cases in which LASSO chose no instruments. In these cases, RMSE, Med. Bias, and MAD use only the replications where LASSO selects a non-empty set of instruments, and we set the confidence interval equal to $(-\infty, \infty)$ and thus fail to reject.

TABLE S.VIII

2SLS SIMULATION RESULTS. CUT-OFF DESIGN, $S = 25$. $N = 250$[a]

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 2.5$								
2SLS(100)		0.012	0.000	0.008	0.052		0.031	0.029	0.029	0.712		0.058	0.057	0.057	1.000
LIML(100)		0.355	0.000	0.024	0.054		1.183	0.002	0.025	0.034		0.045	0.002	0.020	0.058
FULL(100)		0.052	0.000	0.023	0.050		0.047	0.002	0.025	0.034		0.040	0.003	0.020	0.058
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	499	0.000	0.000	0.000	0.000	499	0.014	−0.014	0.014	0.000	500	*	*	*	0.000
Post-LASSO-F	499	0.003	0.003	0.003	0.000	499	0.007	0.007	0.007	0.000	500	*	*	*	0.000
							$F^* = 10$								
2SLS(100)		0.017	0.001	0.012	0.044		0.030	0.027	0.027	0.406		0.054	0.052	0.052	0.908
LIML(100)		0.026	0.001	0.018	0.040		0.024	0.001	0.018	0.036		0.025	0.000	0.015	0.054
FULL(100)		0.026	0.001	0.018	0.040		0.024	0.001	0.018	0.036		0.024	0.000	0.015	0.052
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	1	0.022	−0.001	0.015	0.050	0	0.022	0.006	0.014	0.044	1	0.024	0.009	0.015	0.100
Post-LASSO-F	1	0.021	0.000	0.014	0.046	0	0.021	0.008	0.015	0.054	1	0.023	0.012	0.016	0.118

(*Continues*)

TABLE S.VIII—*Continued*

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 40$								
2SLS(100)		0.018	0.001	0.012	0.048		0.026	0.018	0.019	0.176		0.038	0.034	0.034	0.468
LIML(100)		0.021	0.001	0.013	0.052		0.022	0.001	0.015	0.056		0.022	0.000	0.014	0.054
FULL(100)		0.021	0.001	0.013	0.052		0.022	0.001	0.015	0.056		0.022	0.000	0.014	0.054
LASSO	359	0.020	0.006	0.015	0.002	348	0.023	0.007	0.017	0.022	356	0.023	0.007	0.016	0.016
Post-LASSO	359	0.019	0.005	0.013	0.008	348	0.022	0.004	0.016	0.020	356	0.022	0.006	0.016	0.016
LASSO-F	0	0.021	0.001	0.015	0.048	0	0.022	0.005	0.016	0.056	0	0.022	0.007	0.015	0.060
Post-LASSO-F	0	0.020	0.002	0.013	0.052	0	0.022	0.006	0.017	0.068	0	0.022	0.009	0.016	0.090
							$F^* = 160$								
2SLS(100)		0.019	−0.001	0.013	0.040		0.022	0.010	0.016	0.088		0.027	0.018	0.019	0.188
LIML(100)		0.019	0.000	0.013	0.036		0.020	0.001	0.014	0.058		0.020	0.000	0.013	0.074
FULL(100)		0.019	0.000	0.013	0.036		0.020	0.001	0.014	0.058		0.020	0.000	0.013	0.074
LASSO	51	0.020	−0.001	0.015	0.040	47	0.021	0.004	0.014	0.044	37	0.022	0.007	0.014	0.066
Post-LASSO	51	0.020	−0.001	0.013	0.026	47	0.020	0.004	0.014	0.048	37	0.021	0.006	0.014	0.060
LASSO-F	0	0.020	−0.001	0.014	0.042	0	0.022	0.002	0.015	0.058	0	0.021	0.003	0.014	0.060
Post-LASSO-F	0	0.020	−0.001	0.013	0.040	0	0.021	0.003	0.014	0.050	0	0.021	0.006	0.014	0.070

[a]Results are based on 500 simulation replications and 100 instruments. The first 25 first-stage coefficients were set equal to 1 and the remaining 75 to zero in this design. Corr(e, v) is the correlation between first-stage and structural errors. F^* measures the strength of the instruments as outlined in the text. 2SLS(100), LIML(100), and FULL(100) are respectively the 2SLS, LIML, and Fuller(1) estimator using all 100 potential instruments. Many-instrument robust standard errors are computed for LIML(100) and FULL(100) to obtain testing rejection frequencies. LASSO and Post-LASSO respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 100 instruments. LASSO-F and Post-LASSO-F respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 120 instruments formed by augmenting the original 100 instruments with the first 20 principal components. We report root-mean-squared-error (RMSE), median bias (Med. Bias), mean absolute deviation (MAD), and rejection frequency for 5% level tests (rp(0.05)). "Select 0" is the number of cases in which LASSO chose no instruments. In these cases, RMSE, Med. Bias, and MAD use only the replications where LASSO selects a non-empty set of instruments, and we set the confidence interval equal to $(-\infty, \infty)$ and thus fail to reject.

TABLE S.IX

2SLS SIMULATION RESULTS. CUT-OFF DESIGN, $S = 25$. $N = 500$[a]

Estimator	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 2.5$								
2SLS(100)		0.009	0.000	0.006	0.052		0.021	0.020	0.020	0.652		0.041	0.040	0.040	0.998
LIML(100)		0.036	0.001	0.016	0.044		0.071	−0.001	0.016	0.042		0.025	0.000	0.013	0.052
FULL(100)		0.031	0.001	0.016	0.042		0.030	−0.001	0.016	0.040		0.023	0.001	0.013	0.054
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
							$F^* = 10$								
2SLS(100)		0.012	0.000	0.008	0.046		0.022	0.019	0.019	0.406		0.037	0.036	0.036	0.888
LIML(100)		0.017	−0.001	0.011	0.044		0.017	0.000	0.012	0.044		0.017	−0.001	0.012	0.040
FULL(100)		0.017	−0.001	0.011	0.044		0.017	0.000	0.012	0.044		0.017	−0.001	0.012	0.040
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	0	0.014	0.000	0.009	0.048	0	0.015	0.004	0.011	0.066	0	0.016	0.006	0.010	0.096
Post-LASSO-F	0	0.014	−0.001	0.008	0.060	0	0.015	0.004	0.010	0.078	0	0.017	0.009	0.011	0.144

(*Continues*)

TABLE S.IX—*Continued*

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 40$								
2SLS(100)		0.013	−0.001	0.009	0.044		0.017	0.012	0.013	0.160		0.026	0.022	0.022	0.426
LIML(100)		0.014	−0.002	0.010	0.034		0.015	0.000	0.009	0.046		0.014	−0.001	0.009	0.056
FULL(100)		0.014	−0.002	0.010	0.034		0.014	0.000	0.009	0.048		0.014	0.000	0.009	0.056
LASSO	0	0.014	0.000	0.010	0.038	0	0.014	0.004	0.009	0.056	0	0.016	0.009	0.012	0.108
Post-LASSO	0	0.013	0.000	0.009	0.040	0	0.014	0.003	0.009	0.056	0	0.015	0.007	0.010	0.090
LASSO-F	0	0.014	0.000	0.010	0.040	0	0.014	0.002	0.009	0.048	0	0.015	0.004	0.010	0.064
Post-LASSO-F	0	0.014	−0.001	0.010	0.044	0	0.014	0.003	0.009	0.042	0	0.015	0.007	0.010	0.092
							$F^* = 160$								
2SLS(100)		0.014	0.000	0.010	0.062		0.015	0.007	0.011	0.080		0.018	0.013	0.014	0.138
LIML(100)		0.014	0.000	0.010	0.062		0.014	0.001	0.010	0.048		0.013	0.000	0.009	0.036
FULL(100)		0.014	0.000	0.010	0.062		0.014	0.001	0.010	0.048		0.013	0.000	0.009	0.036
LASSO	0	0.014	−0.001	0.010	0.060	0	0.014	0.002	0.009	0.056	0	0.014	0.005	0.010	0.064
Post-LASSO	0	0.014	−0.001	0.010	0.068	0	0.014	0.002	0.009	0.054	0	0.014	0.004	0.010	0.056
LASSO-F	0	0.014	0.000	0.010	0.056	0	0.014	0.001	0.010	0.048	0	0.014	0.003	0.010	0.052
Post-LASSO-F	0	0.014	−0.001	0.010	0.056	0	0.014	0.002	0.010	0.054	0	0.014	0.005	0.010	0.064

[a]Results are based on 500 simulation replications and 100 instruments. The first 25 first-stage coefficients were set equal to 1 and the remaining 75 to zero in this design. Corr(e, v) is the correlation between first-stage and structural errors. F^* measures the strength of the instruments as outlined in the text. 2SLS(100), LIML(100), and FULL(100) are respectively the 2SLS, LIML, and Fuller(1) estimator using all 100 potential instruments. Many-instrument robust standard errors are computed for LIML(100) and FULL(100) to obtain testing rejection frequencies. LASSO and Post-LASSO respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 100 instruments. LASSO-F and Post-LASSO-F respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 120 instruments formed by augmenting the original 100 instruments with the first 20 principal components. We report root-mean-squared-error (RMSE), median bias (Med. Bias), mean absolute deviation (MAD), and rejection frequency for 5% level tests (rp(0.05)). "Select 0" is the number of cases in which LASSO chose no instruments. In these cases, RMSE, Med. Bias, and MAD use only the replications where LASSO selects a non-empty set of instruments, and we set the confidence interval equal to $(-\infty, \infty)$ and thus fail to reject.

TABLE S.X

2SLS SIMULATION RESULTS. CUT-OFF DESIGN, $S = 50$. $N = 100$[a]

Estimator	Corr$(e, v) = 0$ Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Corr$(e, v) = 0.3$ Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Corr$(e, v) = 0.6$ Select 0	RMSE	Med. Bias	MAD	rp(0.05)
$F^* = 2.5$															
2SLS(100)		0.017	0.000	0.012	0.058		0.036	0.032	0.032	0.546		0.065	0.063	0.063	0.988
LIML(100)		1.649	0.005	0.108	0.016		25.295	0.014	0.101	0.090		74.420	0.038	0.096	0.165
FULL(100)		1.204	0.005	0.108	0.016		1.580	0.014	0.101	0.090		0.884	0.038	0.096	0.165
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
$F^* = 10$															
2SLS(100)		0.019	−0.002	0.013	0.050		0.031	0.024	0.024	0.256		0.052	0.049	0.049	0.702
LIML(100)		4.640	−0.003	0.082	0.016		1.284	0.004	0.079	0.062		5.798	0.016	0.080	0.098
FULL(100)		2.054	−0.003	0.082	0.016		1.156	0.004	0.079	0.062		1.542	0.016	0.080	0.098
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	472	0.030	−0.001	0.025	0.002	468	0.032	−0.001	0.019	0.006	477	0.022	−0.007	0.017	0.000
Post-LASSO-F	472	0.030	−0.007	0.023	0.004	468	0.031	0.002	0.016	0.008	477	0.024	−0.005	0.014	0.002

(*Continues*)

TABLE S.X—*Continued*

Estimator	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 40$								
2SLS(100)		0.023	−0.001	0.017	0.064		0.026	0.015	0.018	0.114		0.034	0.027	0.027	0.264
LIML(100)		0.374	0.003	0.046	0.034		0.662	0.012	0.049	0.034		0.495	0.002	0.041	0.058
FULL(100)		0.374	0.003	0.046	0.034		0.657	0.012	0.049	0.034		0.495	0.002	0.041	0.058
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	400	0.028	−0.005	0.019	0.014	400	0.027	0.004	0.018	0.006	391	0.031	0.001	0.024	0.010
Post-LASSO-F	400	0.027	−0.006	0.019	0.014	400	0.025	0.004	0.016	0.010	391	0.028	0.002	0.017	0.014
							$F^* = 160$								
2SLS(100)		0.022	0.000	0.014	0.056		0.022	0.008	0.015	0.068		0.025	0.014	0.017	0.110
LIML(100)		0.419	0.002	0.031	0.026		1.272	0.000	0.033	0.018		4.169	−0.001	0.032	0.038
FULL(100)		0.419	0.002	0.031	0.026		1.224	0.000	0.033	0.018		4.017	−0.001	0.032	0.038
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	356	0.025	−0.003	0.018	0.010	357	0.026	0.002	0.017	0.014	350	0.023	−0.001	0.015	0.006
Post-LASSO-F	356	0.025	−0.001	0.017	0.016	357	0.023	0.002	0.017	0.012	350	0.023	−0.001	0.016	0.006

[a]Results are based on 500 simulation replications and 100 instruments. The first 50 first-stage coefficients were set equal to 1 and the remaining 50 to zero in this design. Corr(e, v) is the correlation between first-stage and structural errors. F^* measures the strength of the instruments as outlined in the text. 2SLS(100), LIML(100), and FULL(100) are respectively the 2SLS, LIML, and Fuller(1) estimator using all 100 potential instruments. Many-instrument robust standard errors are computed for LIML(100) and FULL(100) to obtain testing rejection frequencies. LASSO and Post-LASSO respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 100 instruments. LASSO-F and Post-LASSO-F respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 120 instruments formed by augmenting the original 100 instruments with the first 20 principal components. We report root-mean-squared-error (RMSE), median bias (Med. Bias), mean absolute deviation (MAD), and rejection frequency for 5% level tests (rp(0.05)). "Select 0" is the number of cases in which LASSO chose no instruments. In these cases, RMSE, Med. Bias, and MAD use only the replications where LASSO selects a non-empty set of instruments, and we set the confidence interval equal to $(-\infty, \infty)$ and thus fail to reject.

TABLE S.XI

2SLS SIMULATION RESULTS. CUT-OFF DESIGN, $S = 50$. $N = 250$[a]

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 2.5$								
2SLS(100)		0.010	0.001	0.007	0.062		0.023	0.021	0.021	0.594		0.042	0.041	0.041	0.990
LIML(100)		0.023	0.001	0.014	0.060		0.021	0.002	0.013	0.020		0.020	0.002	0.013	0.036
FULL(100)		0.023	0.001	0.014	0.060		0.021	0.002	0.013	0.020		0.020	0.002	0.012	0.036
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
							$F^* = 10$								
2SLS(100)		0.012	0.000	0.009	0.038		0.019	0.015	0.015	0.248		0.032	0.030	0.030	0.712
LIML(100)		0.016	0.000	0.012	0.046		0.015	0.000	0.011	0.032		0.016	0.000	0.010	0.058
FULL(100)		0.016	0.000	0.012	0.046		0.015	0.000	0.011	0.032		0.016	0.000	0.010	0.060
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	85	0.015	0.001	0.011	0.030	92	0.014	0.001	0.009	0.028	94	0.015	0.003	0.010	0.042
Post-LASSO-F	85	0.014	0.001	0.010	0.032	92	0.013	0.002	0.009	0.030	94	0.015	0.004	0.011	0.046

(*Continues*)

TABLE S.XI—*Continued*

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 40$								
2SLS(100)		0.013	0.000	0.009	0.054		0.016	0.010	0.012	0.106		0.021	0.017	0.017	0.278
LIML(100)		0.014	0.001	0.009	0.050		0.014	0.001	0.010	0.048		0.014	−0.001	0.010	0.048
FULL(100)		0.014	0.001	0.009	0.050		0.014	0.001	0.010	0.048		0.014	−0.001	0.010	0.048
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	1	0.015	0.000	0.009	0.058	0	0.015	0.001	0.010	0.050	0	0.015	0.002	0.009	0.064
Post-LASSO-F	1	0.014	0.001	0.009	0.068	0	0.015	0.002	0.010	0.056	0	0.014	0.003	0.009	0.048
							$F^* = 160$								
2SLS(100)		0.013	0.000	0.009	0.044		0.015	0.005	0.010	0.082		0.017	0.010	0.012	0.112
LIML(100)		0.014	0.000	0.009	0.046		0.014	0.000	0.010	0.050		0.014	0.001	0.009	0.054
FULL(100)		0.014	0.000	0.009	0.046		0.014	0.000	0.010	0.050		0.014	0.001	0.009	0.052
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	0	0.015	0.000	0.010	0.058	0	0.015	0.002	0.010	0.062	0	0.015	0.001	0.010	0.048
Post-LASSO-F	0	0.014	0.000	0.009	0.056	0	0.015	0.002	0.010	0.048	0	0.014	0.002	0.009	0.052

[a]Results are based on 500 simulation replications and 100 instruments. The first 50 first-stage coefficients were set equal to 1 and the remaining 50 to zero in this design. Corr(e, v) is the correlation between first-stage and structural errors. F^* measures the strength of the instruments as outlined in the text. 2SLS(100), LIML(100), and FULL(100) are respectively the 2SLS, LIML, and Fuller(1) estimator using all 100 potential instruments. Many-instrument robust standard errors are computed for LIML(100) and FULL(100) to obtain testing rejection frequencies. LASSO and Post-LASSO respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 100 instruments. LASSO-F and Post-LASSO-F respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 120 instruments formed by augmenting the original 100 instruments with the first 20 principal components. We report root-mean-squared-error (RMSE), median bias (Med. Bias), mean absolute deviation (MAD), and rejection frequency for 5% level tests (rp(0.05)). "Select 0" is the number of cases in which LASSO chose no instruments. In these cases, RMSE, Med. Bias, and MAD use only the replications where LASSO selects a non-empty set of instruments, and we set the confidence interval equal to $(-\infty, \infty)$ and thus fail to reject.

TABLE S.XII

2SLS SIMULATION RESULTS. CUT-OFF DESIGN, $S = 50$. $N = 500$[a]

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 2.5$								
2SLS(100)		0.007	0.001	0.005	0.058		0.016	0.014	0.014	0.552		0.029	0.028	0.028	0.994
LIML(100)		0.015	0.001	0.009	0.052		0.013	0.000	0.009	0.034		0.012	0.000	0.008	0.044
FULL(100)		0.015	0.001	0.009	0.052		0.013	0.000	0.009	0.034		0.012	0.000	0.008	0.044
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
							$F^* = 10$								
2SLS(100)		0.009	0.000	0.006	0.052		0.015	0.011	0.012	0.262		0.022	0.020	0.020	0.656
LIML(100)		0.011	0.000	0.007	0.052		0.011	0.000	0.007	0.066		0.011	−0.001	0.007	0.058
FULL(100)		0.011	0.000	0.007	0.052		0.011	0.000	0.007	0.066		0.011	−0.001	0.007	0.054
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	1	0.010	0.000	0.007	0.042	0	0.011	0.002	0.007	0.054	1	0.011	0.003	0.007	0.070
Post-LASSO-F	1	0.010	0.000	0.006	0.052	0	0.010	0.003	0.007	0.066	1	0.011	0.003	0.007	0.088

(*Continues*)

TABLE S.XII—*Continued*

Estimator	Corr$(e, v) = 0$ Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Corr$(e, v) = 0.3$ Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Corr$(e, v) = 0.6$ Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 40$								
2SLS(100)		0.009	0.000	0.006	0.032		0.011	0.006	0.008	0.108		0.015	0.012	0.012	0.248
LIML(100)		0.009	0.000	0.006	0.032		0.010	0.000	0.007	0.066		0.009	0.000	0.006	0.046
FULL(100)		0.009	0.000	0.006	0.032		0.010	0.000	0.007	0.064		0.009	0.000	0.006	0.042
LASSO	497	0.005	−0.001	0.004	0.000	500	*	*	*	0.000	497	0.014	−0.006	0.009	0.000
Post-LASSO	497	0.010	−0.001	0.007	0.000	500	*	*	*	0.000	497	0.012	−0.005	0.008	0.000
LASSO-F	0	0.010	0.000	0.006	0.044	0	0.010	0.001	0.007	0.052	0	0.010	0.002	0.007	0.046
Post-LASSO-F	0	0.009	0.000	0.006	0.038	0	0.010	0.002	0.007	0.058	0	0.010	0.002	0.007	0.056
							$F^* = 160$								
2SLS(100)		0.010	0.000	0.007	0.060		0.010	0.003	0.007	0.076		0.012	0.007	0.009	0.128
LIML(100)		0.010	0.000	0.007	0.064		0.010	0.000	0.007	0.044		0.010	0.000	0.007	0.046
FULL(100)		0.010	0.000	0.007	0.064		0.010	0.000	0.007	0.044		0.010	0.000	0.007	0.048
LASSO	349	0.011	−0.001	0.008	0.014	358	0.010	0.003	0.007	0.004	351	0.012	0.004	0.007	0.024
Post-LASSO	349	0.010	−0.001	0.007	0.016	358	0.010	0.003	0.008	0.006	351	0.011	0.003	0.007	0.024
LASSO-F	0	0.011	0.000	0.007	0.066	0	0.010	0.001	0.007	0.048	0	0.010	0.002	0.007	0.056
Post-LASSO-F	0	0.010	0.000	0.007	0.066	0	0.010	0.001	0.007	0.042	0	0.010	0.002	0.007	0.064

[a]Results are based on 500 simulation replications and 100 instruments. The first 50 first-stage coefficients were set equal to 1 and the remaining 50 to zero in this design. Corr(e, v) is the correlation between first-stage and structural errors. F^* measures the strength of the instruments as outlined in the text. 2SLS(100), LIML(100), and FULL(100) are respectively the 2SLS, LIML, and Fuller(1) estimator using all 100 potential instruments. Many-instrument robust standard errors are computed for LIML(100) and FULL(100) to obtain testing rejection frequencies. LASSO and Post-LASSO respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 100 instruments. LASSO-F and Post-LASSO-F respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 120 instruments formed by augmenting the original 100 instruments with the first 20 principal components. We report root-mean-squared-error (RMSE), median bias (Med. Bias), mean absolute deviation (MAD), and rejection frequency for 5% level tests (rp(0.05)). "Select 0" is the number of cases in which LASSO chose no instruments. In these cases, RMSE, Med. Bias, and MAD use only the replications where LASSO selects a non-empty set of instruments, and we set the confidence interval equal to $(-\infty, \infty)$ and thus fail to reject.

TABLE S.XIII

2SLS SIMULATION RESULTS. CUT-OFF DESIGN, $S = 100$. $N = 100$[a]

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 2.5$								
2SLS(100)		0.013	0.000	0.009	0.068		0.024	0.021	0.021	0.378		0.043	0.041	0.041	0.916
LIML(100)		6.463	0.006	0.065	0.012		2.172	0.008	0.061	0.076		0.697	0.017	0.060	0.145
FULL(100)		0.950	0.006	0.065	0.012		0.701	0.008	0.061	0.076		0.555	0.017	0.060	0.145
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
							$F^* = 10$								
2SLS(100)		0.014	0.000	0.008	0.054		0.019	0.012	0.013	0.152		0.030	0.027	0.027	0.486
LIML(100)		1.387	0.002	0.044	0.026		2.774	0.001	0.044	0.040		1.250	0.007	0.046	0.074
FULL(100)		1.171	0.002	0.044	0.026		0.539	0.001	0.044	0.040		1.131	0.007	0.046	0.074
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000

(*Continues*)

TABLE S.XIII—*Continued*

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 40$								
2SLS(100)		0.016	0.000	0.010	0.074		0.017	0.007	0.010	0.088		0.021	0.014	0.015	0.176
LIML(100)		0.203	0.002	0.026	0.034		1.345	0.005	0.027	0.022		0.803	0.002	0.023	0.050
FULL(100)		0.202	0.002	0.026	0.034		1.139	0.005	0.027	0.022		0.782	0.002	0.023	0.050
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
							$F^* = 160$								
2SLS(100)		0.015	0.000	0.010	0.054		0.015	0.004	0.011	0.054		0.016	0.007	0.011	0.086
LIML(100)		0.376	0.001	0.019	0.028		0.221	0.000	0.019	0.010		0.277	−0.001	0.020	0.030
FULL(100)		0.376	0.001	0.019	0.028		0.221	0.000	0.019	0.010		0.277	−0.001	0.020	0.030
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000

[a]Results are based on 500 simulation replications and 100 instruments. All first-stage coefficients were set equal to 1 in this design. Corr(e, v) is the correlation between first-stage and structural errors. F^* measures the strength of the instruments as outlined in the text. 2SLS(100), LIML(100), and FULL(100) are respectively the 2SLS, LIML, and Fuller(1) estimator using all 100 potential instruments. Many-instrument robust standard errors are computed for LIML(100) and FULL(100) to obtain testing rejection frequencies. LASSO and Post-LASSO respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 100 instruments. LASSO-F and Post-LASSO-F respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 120 instruments formed by augmenting the original 100 instruments with the first 20 principal components. We report root-mean-squared-error (RMSE), median bias (Med. Bias), mean absolute deviation (MAD), and rejection frequency for 5% level tests (rp(0.05)). "Select 0" is the number of cases in which LASSO chose no instruments. In these cases, RMSE, Med. Bias, and MAD use only the replications where LASSO selects a non-empty set of instruments, and we set the confidence interval equal to $(-\infty, \infty)$ and thus fail to reject.

TABLE S.XIV

2SLS SIMULATION RESULTS. CUT-OFF DESIGN, $S = 100$. $N = 250$[a]

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 2.5$								
2SLS(100)		0.008	0.000	0.005	0.050		0.015	0.013	0.013	0.390		0.027	0.026	0.026	0.926
LIML(100)		0.013	0.000	0.008	0.048		0.011	0.000	0.007	0.034		0.012	0.001	0.008	0.042
FULL(100)		0.013	0.000	0.008	0.046		0.011	0.000	0.007	0.034		0.012	0.001	0.008	0.044
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
							$F^* = 10$								
2SLS(100)		0.009	0.001	0.006	0.068		0.012	0.009	0.009	0.158		0.019	0.016	0.016	0.456
LIML(100)		0.011	0.001	0.007	0.064		0.010	0.001	0.007	0.048		0.010	0.000	0.007	0.056
FULL(100)		0.011	0.001	0.007	0.064		0.010	0.001	0.007	0.046		0.010	0.001	0.007	0.054
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000

(*Continues*)

TABLE S.XIV—*Continued*

Estimator	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 40$								
2SLS(100)		0.009	0.000	0.007	0.046		0.011	0.005	0.007	0.090		0.013	0.009	0.010	0.150
LIML(100)		0.010	0.000	0.007	0.044		0.010	0.000	0.007	0.050		0.010	0.000	0.006	0.050
FULL(100)		0.010	0.000	0.007	0.044		0.010	0.000	0.007	0.050		0.010	0.000	0.006	0.050
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	487	0.011	0.003	0.009	0.000	496	0.010	−0.003	0.009	0.000	494	0.008	−0.006	0.007	0.000
Post-LASSO-F	487	0.010	0.001	0.006	0.002	496	0.007	−0.004	0.008	0.000	494	0.009	−0.008	0.008	0.000
							$F^* = 160$								
2SLS(100)		0.010	0.000	0.007	0.054		0.010	0.003	0.006	0.080		0.011	0.005	0.007	0.088
LIML(100)		0.010	0.000	0.007	0.052		0.010	0.000	0.006	0.064		0.010	0.000	0.006	0.058
FULL(100)		0.010	0.000	0.007	0.052		0.010	0.000	0.006	0.062		0.010	0.000	0.006	0.058
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	475	0.010	−0.001	0.006	0.000	478	0.016	−0.001	0.009	0.006	477	0.011	0.003	0.009	0.002
Post-LASSO-F	475	0.009	−0.005	0.007	0.000	478	0.016	−0.002	0.007	0.008	477	0.011	0.003	0.007	0.002

[a]Results are based on 500 simulation replications and 100 instruments. All first-stage coefficients were set equal to 1 in this design. Corr(e, v) is the correlation between first-stage and structural errors. F^* measures the strength of the instruments as outlined in the text. 2SLS(100), LIML(100), and FULL(100) are respectively the 2SLS, LIML, and Fuller(1) estimator using all 100 potential instruments. Many-instrument robust standard errors are computed for LIML(100) and FULL(100) to obtain testing rejection frequencies. LASSO and Post-LASSO respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 100 instruments. LASSO-F and Post-LASSO-F respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 120 instruments formed by augmenting the original 100 instruments with the first 20 principal components. We report root-mean-squared-error (RMSE), median bias (Med. Bias), mean absolute deviation (MAD), and rejection frequency for 5% level tests (rp(0.05)). "Select 0" is the number of cases in which LASSO chose no instruments. In these cases, RMSE, Med. Bias, and MAD use only the replications where LASSO selects a non-empty set of instruments, and we set the confidence interval equal to $(-\infty, \infty)$ and thus fail to reject.

TABLE S.XV

2SLS SIMULATION RESULTS. CUT-OFF DESIGN, $S = 100$. $N = 500$[a]

	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
Estimator	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 2.5$								
2SLS(100)		0.006	0.000	0.004	0.046		0.011	0.009	0.009	0.396		0.019	0.018	0.018	0.920
LIML(100)		0.008	0.001	0.005	0.048		0.008	0.000	0.006	0.052		0.007	0.000	0.005	0.036
FULL(100)		0.008	0.001	0.005	0.044		0.008	0.000	0.006	0.052		0.007	0.000	0.005	0.040
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO-F	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
							$F^* = 10$								
2SLS(100)		0.006	0.000	0.004	0.048		0.009	0.006	0.006	0.180		0.012	0.011	0.011	0.412
LIML(100)		0.007	0.000	0.004	0.048		0.007	0.000	0.005	0.062		0.007	−0.001	0.005	0.050
FULL(100)		0.007	0.000	0.004	0.046		0.007	0.000	0.005	0.062		0.007	−0.001	0.005	0.052
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	340	0.006	0.000	0.004	0.004	325	0.007	0.000	0.004	0.026	316	0.007	0.000	0.004	0.014
Post-LASSO-F	340	0.006	0.000	0.004	0.008	325	0.007	0.000	0.004	0.022	316	0.007	0.001	0.004	0.014

(*Continues*)

TABLE S.XV—*Continued*

Estimator	Corr$(e, v) = 0$					Corr$(e, v) = 0.3$					Corr$(e, v) = 0.6$				
	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)	Select 0	RMSE	Med. Bias	MAD	rp(0.05)
							$F^* = 40$								
2SLS(100)		0.006	0.000	0.004	0.038		0.008	0.003	0.006	0.092		0.009	0.006	0.007	0.164
LIML(100)		0.007	0.000	0.004	0.040		0.007	0.000	0.005	0.078		0.007	0.000	0.005	0.048
FULL(100)		0.007	0.000	0.004	0.040		0.007	0.000	0.005	0.078		0.007	0.000	0.005	0.048
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	0	0.007	0.000	0.005	0.036	2	0.007	0.000	0.005	0.070	0	0.007	0.001	0.005	0.060
Post-LASSO-F	0	0.007	0.000	0.005	0.034	2	0.007	0.001	0.005	0.064	0	0.007	0.001	0.005	0.066
							$F^* = 160$								
2SLS(100)		0.007	0.000	0.005	0.052		0.007	0.002	0.005	0.060		0.008	0.003	0.005	0.090
LIML(100)		0.007	0.000	0.005	0.052		0.007	0.000	0.005	0.058		0.007	0.000	0.005	0.044
FULL(100)		0.007	0.000	0.005	0.052		0.007	0.000	0.005	0.060		0.007	0.000	0.005	0.044
LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
Post-LASSO	500	*	*	*	0.000	500	*	*	*	0.000	500	*	*	*	0.000
LASSO-F	0	0.007	0.000	0.005	0.048	0	0.007	0.000	0.005	0.050	0	0.007	0.001	0.005	0.056
Post-LASSO-F	0	0.007	0.000	0.005	0.056	0	0.007	0.000	0.005	0.056	0	0.007	0.001	0.005	0.048

[a] Results are based on 500 simulation replications and 100 instruments. All first-stage coefficients were set equal to 1 in this design. Corr(e, v) is the correlation between first-stage and structural errors. F^* measures the strength of the instruments as outlined in the text. 2SLS(100), LIML(100), and FULL(100) are respectively the 2SLS, LIML, and Fuller(1) estimator using all 100 potential instruments. Many-instrument robust standard errors are computed for LIML(100) and FULL(100) to obtain testing rejection frequencies. LASSO and Post-LASSO respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 100 instruments. LASSO-F and Post-LASSO-F respectively correspond to IV using LASSO or Post-LASSO with the refined data-driven penalty to select among the 120 instruments formed by augmenting the original 100 instruments with the first 20 principal components. We report root-mean-squared-error (RMSE), median bias (Med. Bias), mean absolute deviation (MAD), and rejection frequency for 5% level tests (rp(0.05)). "Select 0" is the number of cases in which LASSO chose no instruments. In these cases, RMSE, Med. Bias, and MAD use only the replications where LASSO selects a non-empty set of instruments, and we set the confidence interval equal to $(-\infty, \infty)$ and thus fail to reject.

REFERENCES

BEKKER, P. A. (1994): "Alternative Approximations to the Distributions of Instrumental Variables Estimators," *Econometrica*, 63, 657–681. [11]

BELLONI, A., D. CHEN, V. CHERNOZHUKOV, AND C. HANSEN (2012): "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain," *Econometrica* (forthcoming). [8]

FULLER, W. A. (1977): "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45, 939–954. [11]

HANSEN, C., J. HAUSMAN, AND W. K. NEWEY (2008): "Estimation With Many Instrumental Variables," *Journal of Business & Economic Statistics*, 26, 398–422. [11]

LEDOUX, M., AND M. TALAGRAND (1991): *Probability in Banach Spaces: Isoperimetry and Processes*. Ergebnisse der Mathematik undihrer Grenzgebiete. Berlin: Springer. [8]

ROSENTHAL, H. P. (1970): "On the Subspaces of L^p $(p > 2)$ Spanned by Sequences of Independent Random Variables," *Israel Journal of Mathematics*, 9, 273–303. [1]

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Berlin: Springer. [3]

VON BAHR, B., AND C.-G. ESSEEN (1965): "Inequalities for the rth Absolute Moment of a Sum of Random Variables, $1 \leq r \leq 2$," *The Annals of Mathematical Statistics*, 36, 299–303. [2,5]

Duke University Fuqua School of Business, Durham, NC 27708, U.S.A.; abn5@duke.edu,

D-GESS, ETH Zurich, IFW E 48.3, Haldeneggsteig 4, CH-8092, Zurich, Switzerland; chendan@ethz.ch,

Dept. of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.; vchern@mit.edu,

and

University of Chicago Booth School of Business, Chicago, IL 60637, U.S.A.; chansen1@chicagobooth.edu.

Manuscript received October, 2010; final revision received June, 2012.