SUPPLEMENTARY APPENDIX FOR "INFERENCE ON TREATMENT EFFECTS AFTER SELECTION AMONGST HIGH-DIMENSIONAL CONTROLS"

A. BELLONI, V. CHERNOZHUKOV, AND C. HANSEN

Abstract. In this supplementary appendix we provide additional results, omitted proofs and extensive simulations that complement the analysis of the main text (arXiv:1201.0224).

1. Split-Sample Estimation and Inference

In this section we discuss a variant of the double selection estimator based on sample splitting. The motivation for the split-sample estimator is that its use allows us to relax the requirement $s^2 \log^2(p \vee n) = o(n)$ that is assumed in the full-sample counterpart to the milder condition

$$s \log(p \vee n) = o(n).$$

To define the estimator, divide the sample randomly into (approximately) equal parts a and b with sizes $n_a = \lceil n/2 \rceil$ and $n_b = n - n_a$. We use superscripts a and b for variables in the first and second subsample respectively. We let the index $k = a, b$ refer to one of the subsamples and let $k^c = \{a, b\} \setminus \{k\}$ refer to the other.

For each subsample $k = a, b$, the model $\widehat{I}^k$ is selected based on the subsample k independently from the subsample k^c. In what follows the model $\widehat{I}^k$ is used to fit the subsample k^c. A constructive way to obtain $\widehat{I}^a$ and $\widehat{I}^b$ is to apply the double selection method for each subsample to select the sets of controls $\widehat{I}^a := \widehat{I}^a_1 \cup \widehat{I}^a_2 \cup \widehat{I}^a_3$ and $\widehat{I}^b := \widehat{I}^b_2 \cup \widehat{I}^b_2 \cup \widehat{I}^b_3$.

Then we form estimates in the two subsamples

$$(\check\alpha^a, \check\beta^a) = \operatorname*{argmin}_{\alpha \in \mathbb{R}, \beta \in \mathbb{R}^p} \{\mathbb{E}_{n_a}[(y_i - d_i\alpha - x_i'\beta)^2] \ : \ \beta_j = 0, \forall j \notin \widehat{I}^b\}, \text{and}$$

$$(\check\alpha^b, \check\beta^b) = \operatorname*{argmin}_{\alpha \in \mathbb{R}, \beta \in \mathbb{R}^p} \{\mathbb{E}_{n_b}[(y_i - d_i\alpha - x_i'\beta)^2] \ : \ \beta_j = 0, \forall j \notin \widehat{I}^a\}.$$

Date: First version: May 2010, This version of October 17, 2013.

For an index i in the subsample k, we define the residuals

$$\widehat{\zeta}_i^o := [y_i - d_i\check{\alpha}_k - x_i'\check{\beta}_k]\{n_k/(n_k - \widehat{s}^{k^c} - 1)\}^{1/2} \tag{1.1}$$

$$\widehat{v}_i := d_i - x_i'\widehat{\beta}_k \text{ and} \tag{1.2}$$

$$\widehat{\zeta}_i := \widehat{\zeta}_i^o \; 1\{|\widehat{\zeta}_i^o| \vee |\widehat{v}_i| \leqslant Cn^{1/2}/[(\widehat{s}^{k^c} \vee n^{1/2})\log n]^{1/2}\} \tag{1.3}$$

where $\widehat{\beta}_k \in \arg\min_\beta\{\mathbb{E}_{n_k}[(d_i - x_i'\beta)^2] : \beta_j = 0, \forall j \notin \widehat{I}^{k^c}\}$ and $\widehat{s}^{k^c} = |\widehat{I}^{k^c}|$.

Finally, we combine the estimates into the split-sample estimator based on $\widehat{I}^a$ and $\widehat{I}^b$ is defined as

$$\check{\alpha}_{ab} = \{(n_a/n)\Upsilon^a + (n_b/n)\Upsilon^b\}^{-1}\{(n_a/n)\Upsilon^a\check{\alpha}_a + (n_b/n)\Upsilon^b\check{\alpha}_b\}, \tag{1.4}$$

where $\Upsilon^k = D^{k\prime}\mathcal{M}_{\widehat{I}^{k^c}}D^k/n_k$.

We state below sufficient conditions for the analysis of the split-sample method.

Condition ASTESS (P). *(i) $\{(y_i, d_i, z_i), i = 1, \ldots, n\}$ are i.n.i.d. vectors on $(\Omega, \mathcal{F}, \mathrm{P})$ that obey the model (2.2)-(2.3), and the vector $x_i = P(z_i)$ is a dictionary of transformations of z_i, which may depend on n but not on P. (ii) The true parameter value α_0, which may depend on P, is bounded, $\|\alpha_0\| \leqslant C$. (iii) Functions m and g admit an approximately sparse form. Namely there exists $s \geqslant 1$ and β_{m0} and β_{g0}, which depend on n and P, such that*

$$m(z_i) = x_i'\beta_{m0} + r_{mi}, \quad \|\beta_{m0}\|_0 \leqslant s, \quad \{\bar{\mathrm{E}}[r_{mi}^2]\}^{1/2} \leqslant C\sqrt{s/n}, \tag{1.5}$$

$$g(z_i) = x_i'\beta_{g0} + r_{gi}, \quad \|\beta_{g0}\|_0 \leqslant s, \quad \{\bar{\mathrm{E}}[r_{gi}^2]\}^{1/2} \leqslant C\sqrt{s/n}. \tag{1.6}$$

(iv) The sparsity index obeys $s\log(p \vee n)/n \leqslant C\delta_n$. (v) For each subsample $k = a, b$, the model $\widehat{I}^{k^c}$ satisfies condition HLMS. (vi) We have $\bar{\mathrm{E}}[|v_i^q| + |\zeta_i^q|] \leqslant C$ for some $q > 4$ and $n^{2/q}s\log(p \vee n)/n \leqslant C\delta_n$.

The Conditions ASTESS(i)-(iii) agree with the corresponding conditions in ASTE. The remaining conditions ASTESS(iv)-(v) are implied by Condition ASTE. We note that Condition ASTESS(vi) is needed only for obtaining consistent estimates of the asymptotic variance. Such conditions are mild since they do not require uniform estimation of the functions g and m.

The next result establishes that the split-sample estimator $\check{\alpha}_{ab}$ has similar large sample properties to the full-sample double-selection estimator under weaker growth condition.

Theorem SA.1 (Inference on Treatment Effects, Split Sample)**.** *Let $\{\mathrm{P}_n\}$ be a sequence of data-generating processes. Assume conditions ASTESS(P)(i-v), SM(P), and SE(P) hold for $\mathbb{P}_n = \mathrm{P}_n$ for each n and each subsample. The split sample estimator $\check{\alpha}_{ab}$ based on $\widehat{I}^a$ and $\widehat{I}^b$ obeys,*

$$([\bar{\mathrm{E}}v_i^2]^{-1}\bar{\mathrm{E}}[v_i^2\zeta_i^2][\bar{\mathrm{E}}v_i^2]^{-1})^{-1/2}\sqrt{n}(\check{\alpha}_{ab} - \alpha_0) \rightsquigarrow N(0, 1).$$

Moreover, if Condition ASTESS(P)(vi) also holds, the result continues to apply if $\bar{\mathrm{E}}[v_i^2]$ and $\bar{\mathrm{E}}[v_i^2\zeta_i^2]$ are replaced by $\mathbb{E}_n[\widehat{v}_i^2]$ and $\mathbb{E}_n[\widehat{v}_i^2\widehat{\zeta}_i^2]$ for $\widehat{\zeta}_i$ and $\widehat{v}_i$ defined in (1.3) and (1.2).

Proof. We use the same notation as in the proof of Theorem 1 with the addition of sub/superscripts indicating the appropriate subsample $k = a, b$, where $k^c = \{a, b\} \setminus \{k\}$.

Step 0.(Combining) In this step we combine both subsample estimators. Letting $\Upsilon^k = D^{k\prime}\mathcal{M}_{\widehat{I}^{k^c}} D^k/n_k$, for $k = a, b$, so that we have

$$\begin{aligned}
\sqrt{n}(\check{\alpha}_{ab} - \alpha_0) &= ((n_a/n)\Upsilon^a + (n_b/n)\Upsilon^b)^{-1} \times \\
&\times \ ((n_a/n)\Upsilon^a\sqrt{n}(\check{\alpha}_a - \alpha_0) + (n_b/n)\Upsilon^b\sqrt{n}(\check{\alpha}_b - \alpha_0)) \\
&= (V'V/n + o_P(1))^{-1} \times \\
&\times \ ((n_a/n)\Upsilon^a\sqrt{n}(\check{\alpha}_a - \alpha_0) + (n_b/n)\Upsilon^b\sqrt{n}(\check{\alpha}_b - \alpha_0)) + o_P(1) \\
&= \{V'V/n\}^{-1} \times \{(1/\sqrt{2}) \times \mathbb{G}_{na}[v_i\zeta_i] + (1/\sqrt{2})\mathbb{G}_{nb}[v_i\zeta_i]\} + o_P(1) \\
&= \{V'V/n\}^{-1} \times \mathbb{G}_n[v_i\zeta_i] + o_P(1)
\end{aligned}$$

where we are also using the fact that

$$\mathbb{E}_{n_k}[\widehat{v}_i^2] - \mathbb{E}_{n_k}[v_i^2] = o_P(1), \quad k = a, b$$

which follows similarly to the proofs given in Step 5.

For $\sigma_n^2 := [\bar{\mathrm{E}}v_i^2]^{-1}\bar{\mathrm{E}}[v_i^2\zeta_i^2][\bar{\mathrm{E}}v_i^2]^{-1}$, define

$$Z_n = \sigma_n^{-1}\sqrt{n}(\check{\alpha}_{ab} - \alpha_0) = \mathbb{G}_n[z_{i,n}] + o_{\mathrm{P}}(1),$$

where $z_{i,n} = \sigma_n^{-1}v_i\zeta_i/\sqrt{n}$ are i.n.i.d. with mean zero. We have that for some small enough $\delta > 0$

$$\bar{\mathrm{E}}|z_{i,n}|^{2+\delta} \lesssim \bar{\mathrm{E}}\left[|v_i|^{2+\delta}|\zeta_i|^{2+\delta}\right] \lesssim \sqrt{\bar{\mathrm{E}}|v_i|^{4+2\delta}}\sqrt{\bar{\mathrm{E}}|\zeta_i|^{4+2\delta}} \lesssim 1,$$

by Condition SM(ii).

This condition verifies the Lyapunov condition and thus implies that $Z_n \to_d N(0, 1)$.

Step 1.(Main) For the subsample $k = a, b$ write $\check{\alpha}_k = \left[D^{k\prime}\mathcal{M}_{\widehat{I}^{k^c}} D^k/n_k\right]^{-1}[D^{k\prime}\mathcal{M}_{\widehat{I}^{k^c}} Y^k/n_k]$ so that

$$\sqrt{n_k}(\check{\alpha}_k - \alpha_0) = \left[D^{k\prime}\mathcal{M}_{\widehat{I}^{k^c}} D^k/n_k\right]^{-1}[D^{k\prime}\mathcal{M}_{\widehat{I}^{k^c}}(g^k + \zeta^k)/\sqrt{n_k}] =: ii_k^{-1} \cdot i_k.$$

By Steps 2 and 3, $ii_k = V^{k\prime}V^k/n_k + o_P(1)$ and $i_k = V^{k\prime}\zeta^k/\sqrt{n_k} + o_P(1)$. Next note that $V^{k\prime}V^k/n_k = \mathrm{E}[V^{k\prime}V^k/n_k] + o_P(1)$ by Chebyshev, and we have that $\bar{\mathrm{E}}_k[v_i^2\zeta_i^2]$ and $\mathrm{E}[V^{k\prime}V^k/n_k]$ are bounded from above and away from zero by assumption.

Step 2. (Behavior of i_k.) Decompose

$$i_k = V^{k\prime}\zeta^k/\sqrt{n_k} + \underbrace{m^{k\prime}\mathcal{M}_{\widehat{I}^{k^c}} g^k/\sqrt{n_k}}_{=:i_{k,a}} + \underbrace{m^{k\prime}\mathcal{M}_{\widehat{I}^{k^c}} \zeta^k/\sqrt{n_k}}_{=:i_{k,b}} + \underbrace{V^{k\prime}\mathcal{M}_{\widehat{I}^{k^c}} g^k/\sqrt{n_k}}_{=:i_{k,c}} - \underbrace{V^{k\prime}\mathcal{P}_{\widehat{I}^{k^c}} \zeta^k/\sqrt{n_k}}_{=:i_{k,d}}.$$

First, note that by Condition ASTESS we have

$$|i_{k,a}| = |m^{k\prime}\mathcal{M}_{\widehat{I}^{k^c}} g^k/\sqrt{n_k}| \leqslant \|\mathcal{M}_{\widehat{I}^{k^c}} m^k\| \ \|\mathcal{M}_{\widehat{I}^{k^c}} g^k\|/\sqrt{n_k} = o_P(1).$$

Second, by the split sample construction, we have that $\widehat{I}^{k^c}$ is independent from ζ^k, and by assumption of the model m^k is also independent of ζ^k. Thus by Chebyshev inequality

$$|i_{k,b}| \lesssim_P \|\mathcal{M}_{\widehat{I}^{k^c}} m^k/\sqrt{n_k}\| = o_P(1),$$

where the last relation follows by ASTESS.

Third, using similar independence arguments, by Chebyshev and Condition ASTESS, conclude

$$|i_{k,c}| \lesssim_P \|\mathcal{M}_{\widehat{I}^{k^c}} g^k/\sqrt{n_k}\| = o_P(1).$$

Fourth, using that $\widehat{s}^{k^c} \lesssim_P s$ by ASTESS so that $\phi_{\min}^{-1}(\widehat{s}^{k^c}) \lesssim_P 1$ by condition SE, we have that

$$|i_{k,d}| \leqslant |\tilde{\beta}_{V^k}(\widehat{I}^{k^c})' X^{k\prime}\zeta^k/\sqrt{n_k}| \lesssim_P \sqrt{s/n} = o_P(1),$$

by Chebyshev since $\|X^k\tilde{\beta}_{V^k}(\widehat{I}^{k^c})/\sqrt{n_k}\| \lesssim_P \sqrt{s/n_k}$ because of the independence of the two subsamples k and k^c.

Step 3.(Behavior of ii_k.) Since $ii_k = (m^k + V^k)'\mathcal{M}_{\widehat{I}^{k^c}}(m^k+V^k)/n_k$, decompose

$$ii_k = V^{k\prime}V^k/n_k + \underbrace{m^{k\prime}\mathcal{M}_{\widehat{I}^{k^c}} m^k/n_k}_{=:ii_{k,a}} + \underbrace{2m^{k\prime}\mathcal{M}_{\widehat{I}^{k^c}} V^k/n_k}_{=:ii_{k,b}} - \underbrace{V^{k\prime}\mathcal{P}_{\widehat{I}^{k^c}} V^k/n_k}_{=:ii_{k,c}}.$$

Then $|ii_{k,a}| = o_P(1)$ by Condition ASTESS, $|ii_{k,b}| = o_P(1)$ by reasoning similar to deriving the bound for $|i_{k,b}|$, and $|ii_{k,c}| = o_P(1)$ by reasoning similar to deriving the bound for $|i_{k,d}|$.

Step 4.(Auxiliary Bounds.) Note that

$$\begin{aligned}\|g^k - X^k\check{\beta}_k\| &= \|g^k - \mathcal{P}_{\widehat{I}^{k^c}}(Y^k - D^k\check{\alpha}_k)\| \\ &\leqslant \|\mathcal{M}_{\widehat{I}^{k^c}} g^k\| + |\check{\alpha}_k - \alpha_0|\|\mathcal{P}_{\widehat{I}^{k^c}} D^k\| + \|\mathcal{P}_{\widehat{I}^{k^c}}\zeta^k\|.\end{aligned}$$

By condition ASTESS $\|\mathcal{M}_{\widehat{I}^{k^c}} g^k\| = o_P(n^{1/4})$ and by condition SM(ii) we have $\|\mathcal{P}_{\widehat{I}^{k^c}} D^k/\sqrt{n_k}\| \leqslant \|D^k/\sqrt{n_k}\| \lesssim_P 1$, and by Step 1 we have $|\check{\alpha}_k - \alpha_0| \lesssim_P n^{-1/2}$. Moreover,

$$\begin{aligned}\|\mathcal{P}_{\widehat{I}^{k^c}}\zeta^k\| &= \|X^k[\widehat{I}^{k^c}](X^k[\widehat{I}^{k^c}]'X^k[\widehat{I}^{k^c}])^{-1}X^k[\widehat{I}^{k^c}]'\zeta^k\| \\ &\leqslant [\sqrt{\phi_{\max,k}(\widehat{s}^{k^c})}/\phi_{\min,k}(\widehat{s}^{k^c})]\|X^k[\widehat{I}^{k^c}]'\zeta^k/\sqrt{n_k}\|.\end{aligned}$$

We have $\sqrt{\phi_{\max,k}(\widehat{s}^{k^c})}/\phi_{\min,k}(\widehat{s}^{k^c}) \lesssim_P 1$ by condition SE, and $\|X^k[I^{k^c}]'\zeta^k/\sqrt{n_k}\| \lesssim_P \sqrt{\widehat{s}^{k^c}}$ by condition SM(ii), the independence between the selected components $\widehat{I}^{k^c}$ and ζ^k since they are based on different subsamples, and applying Chebyshev inequality.

Finally, collecting terms we have

$$\|g^k - X^k\check{\beta}_k\|/\sqrt{n^k} \lesssim_P o(n^{-1/4}) + \sqrt{\widehat{s}^{k^c}/n_k}$$

Similarly, we have $\|m^k - X^k\widehat{\beta}_k\|/\sqrt{n^k} \lesssim_P o(n^{-1/4}) + \sqrt{\widehat{s}^{k^c}/n_k}$.

Step 5.(Variance Estimation.) Since $\widehat{s}^k \lesssim_P s = o(n)$, $(n_k - \widehat{s}^k - 1)/n_k = o_P(1)$, so we can use n as the denominator. Recall the definitions $\widehat{\zeta}_i^o = y_i - d_i\check{\alpha}_k - x_i'\check{\beta}_k$, $\widehat{v}_i = d_i - x_i'\widehat{\beta}_k$ and $\widehat{\zeta}_i = \widehat{\zeta}_i^o 1\{|\widehat{\zeta}_i^o| \vee |\widehat{v}_i| \leqslant H_k\}$ if i belongs to subsample k where $H_k = C\sqrt{n/[(\widehat{s}^{k^c} \vee n^{1/2})\log n]}$. For notational convenience let $A_i = \{|\widehat{\zeta}_i^o| \vee |\widehat{v}_i| \leqslant H_k\}$. Since $q > 4$, $\widehat{s}^{k^c} \lesssim_P s$, and $n^{2/q} s \log(n \vee p) = o(n)$, we have $n^{1/q} = o_P(H_k)$. Hence consider

$$\begin{aligned} \mathbb{E}_n[\widehat{v}_i^2] &= (n_a/n) D^{a\prime} \mathcal{M}_{\widehat{I}^b} D^a / n_a + (n_b/n) D^{b\prime} \mathcal{M}_{\widehat{I}^a} D^b / n_b \\ &= (n_a/n) ii_a + (n_b/n) ii_b = V'V/n + o_P(1) = \bar{\mathrm{E}}[v_i^2] + o_P(1) \end{aligned}$$

by Step 3 and $\bar{\mathrm{E}}[|v_i|^q] \lesssim 1$ for some $q > 4$ by condition SM(ii).

By Condition ASTESS(vi), for each subsample $k = a, b$, we have

$$\mathbb{E}_{n_k}[v_i^2 \zeta_i^2] - \bar{\mathrm{E}}_k[v_i^2 \zeta_i^2] \to_P 0$$

by Vonbahr-Esseen's inequality in von Bahr and Esseen (1965) since

$$\bar{\mathrm{E}}_k[|v_i \zeta_i|^{2+\delta}] \leqslant (\bar{\mathrm{E}}_k[|v_i|^{4+2\delta}] \bar{\mathrm{E}}_k[| \zeta_i|^{4+2\delta}])^{1/2}$$

is uniformly bounded for $4 + 2\delta \leqslant q$. Thus it suffices to show that

$$\mathbb{E}_{n_k}[\widehat{v}_i^2 \widehat{\zeta}_i^2] - \mathbb{E}_{n_k}[v_i^2 \zeta_i^2] \to_P 0.$$

By the triangle inequality

$$\begin{aligned} |\mathbb{E}_{n_k}[\widehat{v}_i^2 \widehat{\zeta}_i^2 - v_i^2 \zeta_i^2]| &\leqslant |\mathbb{E}_{n_k}[(\widehat{v}_i^2 \widehat{\zeta}_i^2 - v_i^2 \zeta_i^2) 1\{A_i\}]| + |\mathbb{E}_{n_k}[(\widehat{v}_i^2 \widehat{\zeta}_i^2 - v_i^2 \zeta_i^2) 1\{A_i^c\}]| \\ &\leqslant |\mathbb{E}_{n_k}[(\widehat{v}_i^2 - v_i^2) \zeta_i^2 1\{A_i\}]| + |\mathbb{E}_{n_k}[v_i^2 (\widehat{\zeta}_i^2 - \zeta_i^2) 1\{A_i\}]| + \\ &\quad + |\mathbb{E}_{n_k}[(\widehat{v}_i^2 - v_i^2)(\widehat{\zeta}_i^2 - \zeta_i^2) 1\{A_i\}]| + o_P(1) \end{aligned}$$

since $|\mathbb{E}_{n_k}[(\widehat{v}_i^2 \widehat{\zeta}_i^2 - v_i^2 \zeta_i^2) 1\{A_i^c\}]| = o_P(1)$ by Step 6. Then,

$$\begin{aligned} |\mathbb{E}_{n_k}[v_i^2 (\widehat{\zeta}_i^2 - \zeta_i^2) 1\{A_i\}]| &\leqslant \underbrace{2\mathbb{E}_{n_k}[\{d_i(\alpha_0 - \check{\alpha}_k)\}^2 v_i^2]}_{=:iii_1} + \underbrace{2\mathbb{E}_{n_k}[\{x_i' \check{\beta}_k - g_i\}^2 v_i^2]}_{=:iii_2} \\ &\quad \underbrace{+ 2\max_{i \leqslant n} |v_i| \{\mathbb{E}_{n_k}[\zeta_i^2 v_i^2]\}^{1/2} \{\mathbb{E}_{n_k}[d_i^2 (\alpha_0 - \check{\alpha}_k)^2]\}^{1/2}}_{=:iii_3} \\ &\quad \underbrace{+ 2\max_{i \leqslant n} |v_i| \{\mathbb{E}_{n_k}[\zeta_i^2 v_i^2]\}^{1/2} \{\mathbb{E}_{n_k}[(g_i - x_i' \check{\beta}_k)^2]\}^{1/2}}_{=:iii_4} \end{aligned}$$

As a consequence of Condition SM(ii) we have $\mathrm{E}[\max_{i \leqslant n} d_i^2] \lesssim n^{2/q}$, $\mathrm{E}[\max_{i \leqslant n} \zeta_i^2] \lesssim n^{2/q}$, $\mathrm{E}[\max_{i \leqslant n} v_i^2] \lesssim n^{2/q}$, thus by Markov inequality we have $\max_{i \leqslant n} |d_i| + |\zeta_i| + |v_i| \lesssim_P n^{1/q}$.

We have the following relations:

$$\begin{aligned} iii_1 &\leqslant |\alpha_0 - \check{\alpha}_k|^2 \mathbb{E}_{n_k}[d_i^2] \max_{i \leqslant n} v_i^2 \lesssim_P n^{-1} n^{2/q} = o_P(1), \\ iii_2 &\leqslant \max_{i \leqslant n} v_i^2 \mathbb{E}_{n_k}[\{x_i' \check{\beta}_k - g_i\}^2] \lesssim_P n^{2/q} \{o(n^{-1/4}) + \sqrt{\widehat{s}^{k^c}/n}\}^2 = o_P(1), \\ iii_3 &\lesssim_P n^{1/q} \sqrt{1/n} = o_P(1), \\ iii_4 &\lesssim_P n^{1/q} \{o(n^{-1/4}) + \sqrt{\widehat{s}^{k^c}/n}\} = o_P(1), \end{aligned}$$

since $\mathbb{E}_{n_k}[\zeta_i^2 v_i^2] \lesssim_P 1$, $\mathbb{E}_{n_k}[\{x_i' \check{\beta}_k - g_i\}^2] \lesssim_P \{o(n^{-1/4}) + \sqrt{\widehat{s}^{k^c}/n}\}^2$ by Step 4, $\widehat{s}^{k^c} \lesssim_P s$, and $|\check{\alpha}_k - \alpha_0|^2 \lesssim_P 1/n$ by Step 1.

Similarly, $\mathbb{E}_{n_k}[(\widehat{v}_i^2 - v_i^2) \zeta_i^2] = o_P(1)$.

Finally, since $\max_{i\leqslant n}\|1\{A_i\}(\widehat{v}_i,\widehat{\zeta}_i,\zeta_i,v_i)'\|_\infty^2 \lesssim_P (H_k^2 \vee n^{2/q}) \lesssim_P H_k^2$, we have

$$\begin{aligned}
|\mathbb{E}_{n_k}[(\widehat{v}_i^2-v_i^2)(\widehat{\zeta}_i^2-\zeta_i^2)1\{A_i\}]| &\leqslant \{\mathbb{E}_{n_k}[(\widehat{v}_i^2-v_i^2)^2 1\{A_i\}]\mathbb{E}_{n_k}[(\widehat{\zeta}_i^2-\zeta_i^2)^2 1\{A_i\}]\}^{1/2}\\
&\leqslant \{\mathbb{E}_{n_k}[2(\widehat{v}_i^2+v_i^2)(\widehat{v}_i-v_i)^2 1\{A_i\}]\mathbb{E}_{n_k}[2(\widehat{\zeta}_i^2+\zeta_i^2)(\widehat{\zeta}_i-\zeta_i)^2 1\{A_i\}]\}^{1/2}\\
&\lesssim_P (H_k^2\vee n^{2/q})\{\mathbb{E}_{n_k}[(\widehat{v}_i-v_i)^2]\mathbb{E}_{n_k}[(\widehat{\zeta}_i-\zeta_i)^2]\}^{1/2}\\
&\lesssim_P H_k^2\{o(n^{-1/4})+\sqrt{\widehat{s}^{k^c}/n}\}^2\\
&\lesssim \tfrac{n}{(\widehat{s}^{k^c}\vee n^{1/2})\log n}\{o(n^{-1/2})+\widehat{s}^{k^c}/n\}=o(1).
\end{aligned}$$

Step 6.(Controlling large terms) By definition of the event A_i we have

$$\begin{aligned}
H_k^2\mathbb{E}_{n_k}[1\{A_i^c\}] &\leqslant \mathbb{E}_{n_k}[\widehat{\zeta}_i^{o2}1\{A_i^c\}]\\
&\leqslant 4\mathbb{E}_{n_k}[\zeta_i^2 1\{A_i^c\}]+4\mathbb{E}_{n_k}[d_i^2(\check{\alpha}_k-\alpha_0)^2 1\{A_i^c\}]+4\mathbb{E}_{n_k}[\{x_i'\check{\beta}_k-g_i\}^2 1\{A_i^c\}]\\
&\lesssim_P n^{2/q}\mathbb{E}_{n_k}[1\{A_i^c\}]+n^{2/q-1}\mathbb{E}_{n_k}[1\{A_i^c\}]+\mathbb{E}_{n_k}[\{x_i'\check{\beta}_k-g_i\}^2].
\end{aligned}$$

Since $n^{1/q}=o_P(H_k)$, and $\mathbb{E}_{n_k}[\{x_i'\check{\beta}_k-g_i\}^2]\lesssim_P o(n^{-1/2})+\widehat{s}^{k^c}/n$, we have

$$\mathbb{E}_{n_k}[1\{A_i^c\}]\lesssim_P \{o(n^{-1/2})+\widehat{s}^{k^c}/n\}/H_k^2.$$

Therefore,

$$\mathbb{E}_{n_k}[\tilde{\zeta}_i^2\tilde{v}_i^2 1\{A_i^c\}]\lesssim_P n^{4/q}\mathbb{E}_{n_k}[1\{A_i^c\}]\lesssim_P n^{4/q}\{o(n^{-1/2})+\widehat{s}^{k^c}/n\}/H_k^2.$$

Finally note that

$$\frac{n^{4/q}\{o(n^{-1/2})+\widehat{s}^{k^c}/n\}}{H_k^2}\quad\lesssim \frac{n^{2/q}}{n^{1/2}}\frac{n^{2/q}(\widehat{s}^{k^c}\vee n^{1/2})\log n}{n}+\frac{n^{2/q}\widehat{s}^{k^c}\log n}{n}\frac{n^{2/q}(\widehat{s}^{k^c}\vee n^{1/2})}{n}=o_P(1)$$

since $q>4$, $\widehat{s}^{k^c}\lesssim_P s$, and $n^{2/q}s\log(n\vee p)=o(n)$ by ASTESS. Also, by construction, we have $\mathbb{E}_{n_k}[\widehat{\zeta}_i^2\widehat{v}_i^2 1\{A_i^c\}]=0$.

□

2. Proof of Lemma 1

We establish the result for Lasso (the proof for other feasible Lasso estimators is similar).

By Lemma 7 in Belloni, Chen, Chernozhukov, and Hansen (2012), under our choice of penalty level and loadings, we have that the condition $\lambda/n \geqslant 2c\|\widehat{\Psi}^{-1}\mathbb{E}_n[\tilde{x}_i\epsilon_i]\|_\infty$ holds with probability $1-o(1)$. Thus, the conclusion of Lemma 11 of Belloni, Chen, Chernozhukov, and Hansen (2012) holds with probability $1-o(1)$, namely for $c_s=(\mathbb{E}_n[r_i^2])^{1/2}$

$$\widehat{s}\leqslant s+\left(\min_{m\in\mathcal{H}}\phi_{\max}(m)\right)\|\widehat{\Psi}^{-1}\|_\infty\left(\frac{2\bar{c}}{\kappa_{\bar{c}}}+\frac{4\bar{c}nc_s}{\lambda\sqrt{s}}\right)^2 \tag{2.7}$$

where $\bar{c}=(c+1)/(c-1)$,

$$\mathcal{H}=\left\{m\in\mathbf{N}: m\geqslant 2s\phi_{\max}(m)\|\widehat{\Psi}^{-1}\|_\infty\left(\frac{2\bar{c}}{\kappa_{\bar{c}}}+\frac{4\bar{c}nc_s}{\lambda\sqrt{s}}\right)^2\right\},$$

$$\kappa_{\bar c} \geqslant \max_{m\in\mathbf{N}} \frac{\sqrt{\phi_{\min}(m+s)}}{\|\widehat\Psi\|_\infty}\left(1-\sqrt{\frac{\phi_{\max}(m+s)}{\phi_{\min}(m+s)}}\bar c\sqrt{s/m}\right).$$

By Condition SE, with probability $1-o(1)$ for n sufficiently large we have $\kappa_{\bar c} > \kappa'/2\|\widehat\Psi\|_\infty$ so that with the same probability

$$\frac{2\bar c}{\kappa_{\bar c}} \lesssim 1. \tag{2.8}$$

Moreover, by condition RF we have with probability $1-o(1)$ that

$$\max\{\|\widehat\Psi\|_\infty,\ \|\widehat\Psi^{-1}\|_\infty\} \lesssim 1. \tag{2.9}$$

Finally, since $\lambda \gtrsim \sqrt{n\log(p\vee n)}$ we have

$$\frac{4\bar c n c_s}{\lambda\sqrt{s}} \lesssim \frac{\sqrt{n}c_s}{\sqrt{s\log(p\vee n)}} \lesssim 1 \quad \text{with probability } 1-o(1) \tag{2.10}$$

since $c_s \lesssim_P \sqrt{s/n}$ by condition ASM and Chebyshev inequality.

Therefore, for some constant $\tilde C$, we have $\tilde C s \in \mathcal{H}$, so that $\min_{m\in\mathcal{H}}\phi_{\max}(m) \leqslant \kappa''$ for n sufficiently large with probability $1-o(1)$ by Condition SE. In turn combining this bound with (2.8), (2.9) and (2.10) into (2.7) we have that $\widehat s \lesssim s$ holds with probability $1-o(1)$ which is the first statement of (i).

To show the second statement in (i), note that

$$\min_{\beta\in\mathbb{R}^p:\ \beta_j=0\ \forall j\notin\widehat T}\sqrt{\mathbb{E}_n[f(\tilde z_i)-\tilde x_i'\beta]^2} \leqslant \sqrt{\mathbb{E}_n[f(\tilde z_i)-\tilde x_i'\widehat\beta]^2}$$

where $\widehat\beta$ is the Lasso estimator. Again by Lemma 7 in Belloni, Chen, Chernozhukov, and Hansen (2012) we have that the assumptions of Lemma 6 in Belloni, Chen, Chernozhukov, and Hansen (2012) hold with probability $1-o(1)$. Using Condition SE to bound $\kappa_{\bar c}$ from below and Condition RF to bound $\|\widehat\Psi\|_\infty$ from above with probability $1-o(1)$ as before, and $\lambda \lesssim \sigma\sqrt{n\log(p\vee n)}$, it follows from Lemma 6 in Belloni, Chen, Chernozhukov, and Hansen (2012) that with probability $1-o(1)$ that

$$\sqrt{\mathbb{E}_n[f(\tilde z_i)-\tilde x_i'\widehat\beta]^2} \lesssim \sigma\sqrt{\frac{s\log(p\vee n)}{n}}.$$

The results regarding Post-Lasso in (ii) follow similarly by invoking Lemma 8 in Belloni, Chen, Chernozhukov, and Hansen (2012).

3. Verification of Conditions for the Examples from Section 4.1

3.1. **Verification for Example 1.** Let $\mathbf{P}$ be the collection of all regression models P that obey the conditions set forth above for all n for the given constants (p, b, B, q_x, q). Below we provide explicit bounds for κ', κ'', c, C, δ_n and Δ_n that appear in Conditions ASTE, SE and SM that depend only on (p, b, B, q_x, q) and n which in turn establish these conditions for any $\mathrm{P} \in \mathbf{P}$.

Condition ASTE(i) is assumed. Condition ASTE(ii) holds with $|\alpha_0| \leqslant C_1^{ASTE} = B$. Condition ASTE(iii) holds with $s = p$ and $r_{gi} = r_{mi} = 0$.

Condition ASTE(iv) holds with $\delta_{1n}^{ASTE} := p^2 \log^2(p \vee n)/n \to 0$ since $s = p$ is fixed. Finally, we verify ASTE(v). Because $\tilde{v}_i = v_i$, $\tilde{\zeta}_i = \zeta_i$ and the moment condition $\mathrm{E}[|v_i^q|] + \mathrm{E}[|\zeta_i^q|] \leqslant C_2^{ASTE} = 2B$ with $q > 4$, the first two requirements follow. To show the last requirement, note that because $\mathrm{E}[\|x_i\|^{q_x}] \leqslant B$ we have

(3.11)

$$\mathrm{P}\left(\max_{1\leqslant i\leqslant n} \|x_i\|_\infty > t_{1n}\right) \leqslant \mathrm{P}\left(\left[\sum_{i=1}^n \|x_i\|^{q_x}\right]^{1/q_x} > t_{1n}\right) \leqslant n\mathrm{E}[\|x_i\|^{q_x}]/t_{1n}^{q_x} \leqslant nB/t_{1n}^{q_x} =: \Delta_{1n}^{ASTE}.$$

Let $t_{1n} = (n \log n)^{1/q_x} B^{1/q_x}$ so that $\Delta_{1n}^{ASTE} = 1/\log n$. Thus we have with probability $1 - \Delta_{1n}^{ASTE}$

$$\max_{1\leqslant i\leqslant n} \|x_i\|_\infty^2 s n^{-1/2+2/q} \leqslant (n \log n)^{2/q_x} B^{2/q_x} p n^{-1/2+2/q} =: \delta_{2n}^{ASTE}.$$

It follows that $\delta_{2n}^{ASTE} \to 0$ by the assumption that $4/q_x + 4/q < 1$.

To verify Condition SE note that

$$\begin{aligned}
\mathrm{P}(\|\mathbb{E}_n[x_i x_i'] - \mathrm{E}[x_i x_i']\| > t_{2n}) &\leqslant \sum_{k=1}^p \sum_{j=1}^p \frac{\mathrm{E}[x_{ij}^2 x_{ik}^2]}{n t_{2n}^2} \leqslant \sum_{k=1}^p \sum_{j=1}^p \frac{\mathrm{E}[x_{ij}^4] + \mathrm{E}[x_{ik}^4]}{2n t_{2n}^2} \\
&\leqslant \frac{p\mathrm{E}[\|x_i\|^4]}{n t_{2n}^2} \leqslant \frac{pB^{4/q_x}}{n t_{2n}^2} =: \Delta_{1n}^{SE}.
\end{aligned}$$

Setting $t_{2n} := b/2$ we have $\Delta_{1n}^{SE} = (2/b)^2 B^{4/q_x} p/n \to 0$ since p is fixed. Then, with probability $1 - \Delta_{1n}^{SE}$ we have

$$\begin{aligned}
\lambda_{\min}(\mathbb{E}_n[x_i x_i']) &\geqslant \lambda_{\min}(\mathrm{E}[x_i x_i']) - \|\mathbb{E}_n[x_i x_i'] - \mathrm{E}[x_i x_i']\| \geqslant b/2 =: \kappa', \\
\lambda_{\max}(\mathbb{E}_n[x_i x_i']) &\leqslant \lambda_{\max}(\mathrm{E}[x_i x_i']) + \|\mathbb{E}_n[x_i x_i'] - \mathrm{E}[x_i x_i']\| \leqslant \mathrm{E}[\|x_i\|^2] + b/2 \leqslant 2B^{2/q_x} =: \kappa''.
\end{aligned}$$

In the verification of Condition SM note that the second and third requirements in Condition SM(i) hold with $c_1^{SM} = b$ and $C_1^{SM} = B^{2/q}$. Condition SM(iii) holds with $\delta_{1n}^{SM} := \log^3 p/n \to 0$ since p is fixed.

The first requirement in Condition SM(i) and Condition SM(ii) hold by the stated moment assumptions, for $\epsilon_i = v_i$ and $\epsilon_i = \zeta_i$, $\tilde{y}_i = d_i$ and $\tilde{y}_i = y_i$,

$$\begin{aligned}
\mathrm{E}[|\epsilon_i^q|] &\leqslant B =: A_1\\
\mathrm{E}[|d_i^q|] &\leqslant 2^{q-1}\mathrm{E}[|x_i'\beta_{m0}|^q] + 2^{q-1}\mathrm{E}[|v_i^q|] \leqslant 2^{q-1}\mathrm{E}[\|x_i\|^q]\|\beta_{m0}\|^q + 2^{q-1}\mathrm{E}[|v_i^q|]\\
&\leqslant 2^{q-1}(B^{q/q_x}B^q + B) =: A_2\\
\mathrm{E}[d_i^4] &\leqslant 2^3(B^{4/q_x}B^4 + B) =: A_2'\\
\mathrm{E}[y_i^4] &\leqslant 3^3\|\alpha_0\|^4\mathrm{E}[d_i^4] + 3^3\|\beta_{g0}\|^4\mathrm{E}[\|x_i\|^4] + 3^3\mathrm{E}[\zeta_i^4]\\
&\leqslant 3^3B^4 2^3 A_2' + 3^3B^4B^{4/q_x} + 3^3B^{4/q} =: A_3\\
\max_{1\leqslant j\leqslant p}\mathrm{E}[x_{ij}^2\tilde{y}_i^2] &\leqslant \max_{1\leqslant j\leqslant p}(\mathrm{E}[x_{ij}^4])^{1/2}(\mathrm{E}[\tilde{y}_i^4])^{1/2} \leqslant B^{2/q_x}(\mathrm{E}[\tilde{y}_i^4])^{1/2} \leqslant B^{2/q_x}(A_2'\vee A_3)^{1/2} =: A_4\\
\max_{1\leqslant j\leqslant p}\mathrm{E}[|x_{ij}\epsilon_i|^3] &= \max_{1\leqslant j\leqslant p}\mathrm{E}[|x_{ij}^3|\mathrm{E}[|\epsilon_i^3| \mid x_i]] \leqslant B^{3/q}\max_{1\leqslant j\leqslant p}\mathrm{E}[|x_{ij}^3|] \leqslant B^{3/q+3/q_x} =: A_5\\
\max_{1\leqslant j\leqslant p} 1/\mathrm{E}[x_{ij}^2] &\leqslant 1/\lambda_{\min}(\mathrm{E}[x_ix_i']) \leqslant 1/b =: A_6
\end{aligned}$$

since $4 < q \leqslant q_x$. Thus these conditions hold with $C_2^{SM} = A_2 \vee (A_1 + (A_2'\vee A_3)^{1/2} + A_4 + A_5 + A_6)$.

Next we show Condition SM(iv). By (3.11) we have $\max_{1\leqslant i\leqslant n}\|x_i\|_\infty^2 \leqslant (n\log n)^{2/q_x}B^{2/q_x}$ with probability $1-\Delta_{1n}^{ASTE}$, thus with the same probability

$$\max_{i\leqslant n}\|x_i\|_\infty^2\frac{s\log(n\vee p)}{n} \leqslant (B\log n)^{2/q_x}\frac{n^{2/q_x}p\log(p\vee n)}{n} =: \delta_{1n}^{SM} \to 0$$

since $q_x > 4$ and $s = p$ is fixed.

Next for $\epsilon_i = v_i$ and $\epsilon_i = \zeta_i$ we have

$$\mathrm{P}\left(\max_{1\leqslant j\leqslant p}|(\mathbb{E}_n-\mathrm{E})[x_{ij}^2\epsilon_i^2]| > \delta_{2n}^{SM}\right) \leqslant \sum_{j=1}^p\frac{\mathrm{E}[x_{ij}^4\epsilon_i^4]}{n(\delta_{2n}^{SM})^2} \leqslant \frac{pB^{4/q+4/q_x}}{n(\delta_{2n}^{SM})^2} =: \Delta_{1n}^{SM}$$

by the union bound, Chebyshev inequality and by $\mathrm{E}[x_{ij}^4\epsilon_i^4] = \mathrm{E}[x_{ij}^4\mathrm{E}[\epsilon_i^4 \mid x_i]] \leqslant B^{4/q+4/q_x}$. Letting $\delta_{2n}^{SM} = B^{2/q+2/q_x}n^{-1/4} \to 0$ we have $\Delta_{1n}^{SM} = p/n^{1/2} \to 0$ since p, B, q and q_x are fixed.

Next for $\tilde{y}_i = d_i$ and $\tilde{y}_i = y_i$ we have

$$\mathrm{P}\left(\max_{1\leqslant j\leqslant p}|(\mathbb{E}_n-\mathrm{E})[x_{ij}^2\tilde{y}_i^2]| > \delta_{3n}^{SM}\right) \leqslant \sum_{j=1}^p\frac{\mathrm{E}[x_{ij}^4\tilde{y}_i^4]}{n(\delta_{3n}^{SM})^2} \leqslant \frac{pB^{4/q_x}A_8^{4/q}}{n(\delta_{3n}^{SM})^2} =: \Delta_{2n}^{SM}$$

by the union bound, Chebyshev inequality and by

$$\mathrm{E}[x_{ij}^4\tilde{y}_i^4] \leqslant \mathrm{E}[x_{ij}^{\tilde{q}}]^{4/\tilde{q}}\mathrm{E}[\tilde{y}_i^q]^{4/q} \leqslant \mathrm{E}[x_{ij}^{q_x}]^{4/q_x}\mathrm{E}[\tilde{y}_i^q]^{4/q} \leqslant B^{4/q_x}A_8^{4/q}$$

holding by Hölder inequality where $4 < \tilde{q} \leqslant q_x$ such that $4/q + 4/\tilde{q} = 1$, and

$$\begin{aligned}
\mathrm{E}[\tilde{y}_i^q] &\leqslant (1+3^{q-1}|\alpha_0|^q)\mathrm{E}[d_i^q] + 3^{q-1}\|\beta_{g0}\|^q\mathrm{E}[\|x_i\|^q] + 3^{q-1}\mathrm{E}[\zeta_i^q]\\
&\leqslant 3^q(A_2 + B^qA_2 + B^qB^{q/q_x} + B) =: A_8.
\end{aligned}$$

Letting $\delta_{3n}^{SM} = B^{4/q_x}A_8^{4/q}n^{-1/4} \to 0$ we have $\Delta_{2n}^{SM} = p/n^{1/2} \to 0$ since p, B, q and q_x are fixed.

Finally, we set $c = c_1^{SM}$, $C = \max\{C_1^{ASTE}, C_2^{ASTE}, C_1^{SM}, C_2^{SM}\}$, $\delta_n = \max\{\delta_{1n}^{ASTE}, \delta_{2n}^{ASTE}, \delta_{1n}^{SM} + \delta_{2n}^{SM} + \delta_{3n}^{SM}\} \to 0$, and $\Delta_n = \max\{\Delta_{1n}^{ASTE} + \Delta_{1n}^{SM} + \Delta_{2n}^{SM}, \Delta_{1n}^{SE}\} \to 0$. □

3.2. **Verification for Examples 2 and 3.** We will make use of the following technical lemmas in the verification of examples 2 and 3.

Lemma SA.1. *Let $f_{ij} \sim N(0, \sigma_j^2)$, $\sigma_j \leqslant \sigma$, independent across $i = 1, \ldots, n$, where $j = 1, \ldots, p$. Then, for some universal constant $\bar{C} \geqslant 1$, we have that for any $k \geqslant 2$ and $\gamma \in (0,1)$*

$$P\left(\max_{1 \leqslant j \leqslant p} \{\mathbb{E}_n[|f_{ij}^k|]\}^{1/k} \geqslant \sigma \bar{C}\sqrt{k} + \sigma n^{-1/k}\sqrt{2\log(2p/\gamma)} \right) \leqslant \gamma.$$

Proof. Note that $P(\mathbb{E}_n[|f_{ij}^k|] > M) = P(\|f_{\cdot j}\|_k^k > Mn) = P(\|f_{\cdot j}\|_k > (Mn)^{1/k})$.

Since $|\|f\|_k - \|g\|_k| \leqslant \|f - g\|_k \leqslant \|f - g\|$, we have that $\|\cdot\|_k$ is 1-Lipschitz for $k \geqslant 2$. Moreover,

$$\begin{aligned} \mathrm{E}[\|f_{\cdot j}\|_k] &\leqslant (\mathrm{E}[\|f_{\cdot j}\|_k^k])^{1/k} = (\sum_{i=1}^n \mathrm{E}[|f_{ij}^k|])^{1/k} = n^{1/k}(\mathrm{E}[|f_{1j}^k|])^{1/k} \\ &= n^{1/k}\{\sigma_j^k 2^{k/2}\Gamma((k+1)/2)/\Gamma(1/2)\}^{1/k} \leqslant n^{1/k}\sigma\sqrt{k}\bar{C}. \end{aligned}$$

By Ledoux and Talagrand (1991), page 21 equation (1.6), we have

$$P(\|f_{\cdot j}\|_k > (Mn)^{1/k}) \leqslant 2\exp(-\{(Mn)^{1/k} - \mathrm{E}[\|f_{\cdot j}\|_k]\}^2/2\sigma_j^2).$$

Setting $M := \{\sigma\sqrt{k}\bar{C} + \sigma n^{-1/k}\sqrt{2\log(2p/\gamma)}\}^k$, so that $(Mn)^{1/k} = n^{1/k}\sigma\sqrt{k}\bar{C} + \sigma\sqrt{2\log(2p/\gamma)}$ we have by the union bound and $\sigma \geqslant \sigma_j$

$$P(\max_{1 \leqslant j \leqslant p} \mathbb{E}_n[|f_{ij}^k|] \geqslant M) \leqslant p \max_{1 \leqslant j \leqslant p} P(\mathbb{E}_n[|f_{ij}^k|] \geqslant M) \leqslant \gamma.$$

□

Lemma SA.2 (Uniform Approximation)**.** *Let $h_i = x_i'\theta_h + \rho_i$ be a function whose coefficients $\theta_h \in S_A^a(p)$, and $\underline{\kappa} \leqslant \lambda_{\min}(\mathrm{E}[x_i x_i']) \leqslant \lambda_{\max}(\mathrm{E}[x_i x_i']) \leqslant \bar{\kappa}$. For $s = A^{1/a}n^{1/2a}$, $a > 1$, define β_{h0} as in (4.31), $r_{hi} = h_i - x_i'\beta_{h0}$, for $i = 1, \ldots, n$. Then we have*

$$|r_{hi}| \leqslant \|x_i\|_\infty (\bar{\kappa}/\underline{\kappa})^{3/2} \left\{ \frac{2a-1}{a-1}\sqrt{s^2/n} + 5\sqrt{s\mathrm{E}[\rho_i^2]/\underline{\kappa}} \right\} + |\rho_i|.$$

Proof. Let T_h denote the support of β_{h0} and S denote the support of the s largest components of θ_h. Note that $|T_h| = |S| = s$. First we establish some auxiliary bounds on the $\|\theta_h[T_h^c]\|$ and $\|\theta_h[T_h^c]\|_1$. By the optimality of T_h and β_{h0} we have that

$$\sqrt{\mathrm{E}[(h_i - x_i'\beta_{h0})^2]} \leqslant \sqrt{\mathrm{E}[(x_i[S^c]'\theta_h[S^c] + \rho_i)^2]} \leqslant \sqrt{\bar{\kappa}}\|\theta_h[S^c]\| + \sqrt{\mathrm{E}[\rho_i^2]} \text{ and}$$

$$\sqrt{\mathrm{E}[(h_i - x_i'\beta_{h0})^2]} = \sqrt{\mathrm{E}[\{x_i'(\theta_h - \beta_{h0}) + \rho_i\}^2]} \geqslant \sqrt{\underline{\kappa}}\|\theta_h[T_h^c]\| - \sqrt{\mathrm{E}[\rho_i^2]}.$$

Thus we have $\|\theta_h[T_h^c]\| \leqslant \sqrt{\bar\kappa/\underline{\kappa}}\|\theta_h[S^c]\| + 2\sqrt{\mathrm{E}[\rho_i^2]/\underline{\kappa}}$. Moreover, since $\theta_h \in S_A^a(p)$, we have

$$\|\theta_h[S^c]\|^2 = \sum_{j=s+1}^{\infty} \theta_{h(j)}^2 \leqslant A^2 \sum_{j=s+1}^{\infty} j^{-2a} \leqslant A^2 s^{-2a+1}/[2a-1] \leqslant A^2 s^{-2a+1}$$

since $a > 1$. Combining these relations we have

$$\begin{aligned}\|\theta_h[T_h^c]\| &\leqslant \sqrt{\bar\kappa/\underline{\kappa}} A s^{-a+1/2} + 2\sqrt{\mathrm{E}[\rho_i^2]/\underline{\kappa}} \\ &= \sqrt{\bar\kappa/\underline{\kappa}}\sqrt{s/n} + 2\sqrt{\mathrm{E}[\rho_i^2]/\underline{\kappa}}.\end{aligned}$$

The second bound follows by observing that

$$\begin{aligned}\|\theta_h[T_h^c]\|_1 &\leqslant \sqrt{s}\|\theta_h[T_h^c \cap S]\| + \|\theta_h[S^c]\|_1 \leqslant \sqrt{s}\|\theta_h[T_h^c]\| + A s^{-a+1}/[a-1] \\ &\leqslant \sqrt{s^2/n}\sqrt{\bar\kappa/\underline{\kappa}} + 2\sqrt{s\mathrm{E}[\rho_i^2]/\underline{\kappa}} + (s/\sqrt{n})/[a-1] \\ &\leqslant \sqrt{s^2/n}\sqrt{\bar\kappa/\underline{\kappa}}\ a/[a-1] + 2\sqrt{s\mathrm{E}[\rho_i^2]/\underline{\kappa}}.\end{aligned}$$

By the first-order optimality condition of the problem (4.31) that defines β_{h0}, we have

$$\mathrm{E}[x_i[T_h]x_i[T_h]'](\beta_{h0}[T_h] - \theta_h[T_h]) = \mathrm{E}[x_i[T_h]x_i[T_h^c]']\theta_h[T_h^c] + \mathrm{E}[x_i[T_h]\rho_i].$$

Thus, since $\|\mathrm{E}[x_i[T_h]\rho_i]\| = \sup_{\|\eta\|=1} \mathrm{E}[\eta' x_i[T_h]\rho_i] \leqslant \sup_{\|\eta\|=1} \sqrt{\mathrm{E}[(\eta' x_i[T_h])^2]}\sqrt{\mathrm{E}[\rho_i^2]}$ we have

$$\begin{aligned}\underline{\kappa}\|\beta_{h0} - \theta_h[T_h]\| &\leqslant \bar\kappa\|\theta_h[T_h^c]\| + \sqrt{\bar\kappa \mathrm{E}[\rho_i^2]} \\ &\leqslant \sqrt{s/n}\ (\bar\kappa^{3/2}/\sqrt{\underline{\kappa}}) + \sqrt{\mathrm{E}[\rho_i^2]}\ \sqrt{\bar\kappa}(1 + 2\sqrt{\bar\kappa/\underline{\kappa}})\end{aligned}$$

where the last inequality follows from the definition of $s = A^{1/a}n^{1/2a}$. Therefore

$$\begin{aligned}|r_{hi}| &= |h_i - x_i'\beta_{h0}| = |x_i'(\theta_h - \beta_{h0})| + |\rho_i| \\ &\leqslant \|x_i\|_\infty \|\theta_h - \beta_{h0}\|_1 + |\rho_i| \\ &\leqslant \sqrt{s}\|x_i\|_\infty\|\theta_{hT_h} - \beta_{h0}\| + \|x_i\|_\infty\|\theta_{hT_h^c}\|_1 + |\rho_i| \\ &\leqslant \|x_i\|_\infty\{\sqrt{s^2/n}\ (\bar\kappa/\underline{\kappa})^{3/2} + \sqrt{s\mathrm{E}[\rho_i^2]/\underline{\kappa}}\ \sqrt{\bar\kappa/\underline{\kappa}}(1 + 2\sqrt{\bar\kappa/\underline{\kappa}})\} + \\ &\quad + \|x_i\|_\infty(\sqrt{s^2/n}\sqrt{\bar\kappa/\underline{\kappa}}\ a/[a-1] + 2\sqrt{s\mathrm{E}[\rho_i^2]/\underline{\kappa}}) + |\rho_i| \\ &\leqslant \|x_i\|_\infty(\bar\kappa/\underline{\kappa})^{3/2}\{\tfrac{2a-1}{a-1}\sqrt{s^2/n} + 5\sqrt{s\mathrm{E}[\rho_i^2]/\underline{\kappa}}\} + |\rho_i|.\end{aligned}$$

$\square$

Example 2. Let $\mathbf{P}$ be the collection of all regression models P that obey the conditions set forth above for all n for the given constants $(\underline{\kappa}, \bar\kappa, a, A, B, \chi)$ and sequences p_n and $\bar\delta_n$. Below we provide explicit bounds for κ', κ'', c, C, δ_n and Δ_n that appear in Conditions ASTE, SE and SM that depend only on $(\underline{\kappa}, \bar\kappa, a, A, B, \chi)$, p, $\bar\delta_n$ and n which in turn establish these conditions for any $\mathrm{P} \in \mathbf{P}$. In what follows we exploit Gaussianity of w_i and use that $(\mathrm{E}[|\eta' w_i|^k])^{1/k} \leqslant G_k(\mathrm{E}[|\eta' w_i|^2])^{1/2}$ for any vector η, $\|\eta\| < \infty$, where the constant G_k depends on k only.

Conditions ASTE(i) is assumed. Condition ASTE(ii) holds with $|\alpha_0| \leqslant B =: C_1^{ASTE}$. Because $\theta_m, \theta_g \in S_A^a(p)$, Condition ASTE(iii) holds with

$$s = A^{1/a} n^{1/2a}, \quad r_{mi} = m(z_i) - \sum_{j=1}^{p} z_{ij}\beta_{m0j}, \quad \text{and} \quad r_{gi} = g(z_i) - \sum_{j=1}^{p} z_{ij}\beta_{g0j}$$

where $\|\beta_{m0}\|_0 \leqslant s$ and $\|\beta_{g0}\|_0 \leqslant s$. Indeed, we have

$$\mathrm{E}[r_{mi}^2] \leqslant \mathrm{E}\left[\left(\sum_{j\geqslant s+1} \theta_{m(j)} z_{i(j)}\right)^2\right] \leqslant \bar{\kappa} \sum_{j\geqslant s+1} \theta_{m(j)}^2 \leqslant \bar{\kappa} A^2 s^{-2a+1}/[2a-1] \leqslant \bar{\kappa} s/n$$

where the first inequality follows by the definition of β_{m0} in (4.31), the second inequality follows from $\theta_m \in S_A^a(p)$, and the last inequality because $s = A^{1/a} n^{1/2a}$. Similarly we have $\mathrm{E}[r_{gi}^2] \leqslant \mathrm{E}[(\sum_{j\geqslant s+1} \theta_{g(j)} z_{i(j)})^2] \leqslant \bar{\kappa} A^2 s^{-2a+1}/[2a-1] \leqslant \bar{\kappa} s/n$. Thus let $C_2^{ASTE} := \sqrt{\bar{f}}$.

Condition ASTE(iv) holds with $\delta_{1n}^{ASTE} := A^{2/a} n^{1/a-1} \log^2(p \vee n) \to 0$ since $s = A^{1/a} n^{1/2a}$, A is fixed, and the assumed condition $n^{(1-a)/a} \log^2(p \vee n) \log^2 n \leqslant \bar{\delta}_n \to 0$.

The moment restrictions in Condition ASTE(v) are satisfied by the Gaussianity. Indeed, we have for $q = 4/\chi$ (where $\chi < 1$ by assumption)

$$\begin{aligned} \mathrm{E}[|\tilde{\zeta}_i|^q] &\leqslant 2^{q-1}\mathrm{E}[|\zeta_i^q|] + 2^{q-1}\mathrm{E}[|r_{gi}^q|] \leqslant 2^{q-1} G_q^q (\mathrm{E}[\zeta_i^2]^{q/2} + \mathrm{E}[r_{gi}^2]^{q/2}) \\ &\leqslant 2^{q-1} G_q^q \{\bar{\kappa}^{q/2} + \bar{\kappa}^{q/2}(s/n)^{q/2}\} \\ &\leqslant 2^q G_q^q \bar{\kappa}^{q/2} =: C_3^{ASTE} \end{aligned}$$

for $s \leqslant n$, i.e., $n \geqslant n_{01}^{ASTE} := A^{2/[2a-1]}$. Similarly, $\mathrm{E}[|\tilde{v}_i|^q] \leqslant C_3^{ASTE}$. Moreover,

$$\begin{aligned} |\mathrm{E}[\tilde{\zeta}_i^2 \tilde{v}_i^2] - \mathrm{E}[\zeta_i^2 v_i^2]| &\leqslant \mathrm{E}[\zeta_i^2 r_{mi}^2] + \mathrm{E}[r_{gi}^2 v_i^2] + \mathrm{E}[r_{mi}^2 r_{gi}^2] \\ &\leqslant \sqrt{\mathrm{E}[\zeta_i^4]\mathrm{E}[r_{mi}^4]} + \sqrt{\mathrm{E}[r_{gi}^4]\mathrm{E}[v_i^4]} + \sqrt{\mathrm{E}[r_{mi}^4]\mathrm{E}[r_{gi}^4]} \\ &\leqslant G_4^2 \bar{\kappa} \mathrm{E}[r_{mi}^2] + G_4^2 \bar{\kappa} \mathrm{E}[r_{gi}^2] + G_4^2 \mathrm{E}[r_{mi}^2]\mathrm{E}[r_{gi}^2] \\ &\leqslant G_4^2 \bar{\kappa}^2 \{2 + \bar{\kappa} s/n\} s/n =: \delta_{2n}^{ASTE} \to 0. \end{aligned}$$

Next note that by Gaussian tail bounds and $\lambda_{\max}(\mathrm{E}[w_i w_i']) \leqslant \bar{\kappa}$ we have

$$\begin{aligned} \max_{i\leqslant n} \|x_i\|_\infty &\leqslant \|\mathrm{E}[x_i]\|_\infty + \max_{i\leqslant n} \|x_i - \mathrm{E}[x_i]\|_\infty \\ &\leqslant \sqrt{\bar{\kappa}} + \sqrt{2\bar{\kappa}\log(pn)} \ \text{ with probability at least } 1 - \Delta_{1n}^{ASTE} \end{aligned} \tag{3.12}$$

where $\Delta_{1n}^{ASTE} = 1/\sqrt{2\bar{\kappa}\log(pn)}$. The last requirement in Condition ASTE(v) holds with $q = 4/\chi$

$$\max_{i\leqslant n} \|x_i\|_\infty^2 s n^{-1/2+2/q} \leqslant 6\bar{\kappa}\log(pn) A^{1/a} n^{\frac{1}{2a}-\frac{1}{2}+\chi/2} =: \delta_{3n}^{ASTE}$$

with probability $1 - \Delta_{1n}^{ASTE}$. By the assumption on a, p, χ, and n, $\delta_{3n}^{ASTE} \to 0$.

To verify Condition SE with $\ell_n = \log n$ note that the minimal and maximal eigenvalues of $\mathrm{E}[x_i x_i']$ are bounded away from zero by $\underline{\kappa} > 0$ and from above by $\bar{\kappa} < \infty$ uniformly in n. Also, let $\mu = \mathrm{E}[x_i]$ so that $x_i = \tilde{x}_i + \mu$ where $\tilde{x}_i$ is zero mean. By constriction $\mathrm{E}[x_i x_i'] = \mathrm{E}[\tilde{x}_i \tilde{x}_i'] + \mu\mu'$ and $\|\mu\| \leqslant \sqrt{\bar{\kappa}}$.

For any $\eta \in \mathbb{R}^p$, $\|\eta\|_0 \leqslant k := s \log n$ and $\|\eta\| = 1$, we have that

$$\mathbb{E}_n[(\eta' x_i)^2] - \mathrm{E}[(\eta' x_i)^2] = \mathbb{E}_n[(\eta' \tilde{x}_i)^2] - \mathrm{E}[(\eta' \tilde{x}_i)^2] + 2\eta' \mathbb{E}_n[\tilde{x}_i] \cdot \eta' \mu.$$

Moreover, by Gaussianity of x_i, with probability $1 - \Delta_{1n}^{SE}$, where $\Delta_{1n}^{SE} = 1/\sqrt{2\bar{\kappa} \log(pn)}$,

$$\begin{array}{rl} |\eta' \mathbb{E}_n[\tilde{x}_i]| & \leqslant \|\eta\|_1 \|\mathbb{E}_n[\tilde{x}_i]\|_\infty \leqslant \sqrt{k}\sqrt{2\bar{\kappa}\log(pn)}/\sqrt{n} \\ |\eta'\mu| & \leqslant \|\eta\| \; \|\mu\| \leqslant \sqrt{\bar{\kappa}}. \end{array}$$

By the sub-Gaussianity of $\tilde{x}_i = (\mathrm{E}[x_i x_i'] - \mu\mu')^{-1/2}\Psi_i$, where $\Psi_i \sim N(0, I_p)$, by Theorem 3.2 in Rudelson and Zhou (2011) (restated in Lemma SA.8) with $\tau = 1/6$, $k = s \log n$, $\alpha = \sqrt{8/3}$, provided that

$$n \geqslant N_n := 80(\alpha^4/\tau^2)(s \log n) \;\; \log(12ep/[\tau s \log n]),$$

we have

$$(1-\tau)^2 \mathrm{E}[(\eta' \tilde{x}_i)^2] \leqslant \mathbb{E}_n[(\eta' \tilde{x}_i)^2] \leqslant (1+\tau)^2 \mathrm{E}[(\eta' \tilde{x}_i)^2]$$

with probability $1 - \Delta_{1n}^{SE}$, where $\Delta_{1n}^{SE} = 2\exp(-\tau^2 n / 80\alpha^4)$. Note that under ASTE(iv) we have $\Delta_{1n}^{SE} \to 0$ and

$$n_{01}^{SE} := \max\{n : n \leqslant N_n\} \leqslant \max\{(12e/\tau)^{2a} A^{-2}, \; 80^2(\alpha^8/\tau^4) A^{2/a}, n^*\}$$

where n^* is the smallest n such that $\bar{\delta}_n < 1$.

Therefore, with probability $1 - \Delta_{1n}^{SE}$ and $n \geqslant n_{01}^{SE}$, we have for any $\eta \in \mathbb{R}^p$, $\|\eta\|_0 \leqslant k$ and $\|\eta\| = 1$,

$$\begin{array}{rl} \mathbb{E}_n[(\eta' x_i)^2] & \geqslant \mathrm{E}[(\eta' x_i)^2] - |\mathbb{E}_n[(\eta' x_i)^2] - \mathrm{E}[(\eta' x_i)^2]| \\ & \geqslant \mathrm{E}[(\eta' x_i)^2] - |\mathbb{E}_n[(\eta' \tilde{x}_i)^2] - \mathrm{E}[(\eta' \tilde{x}_i)^2]| - 2|\eta' \mathbb{E}_n[\tilde{x}_i]| \cdot |\eta'\mu| \\ & \geqslant \mathrm{E}[(\eta' x_i)^2]\{1 - 2\tau - \tau^2\} - 2\bar{\kappa}\sqrt{2k\log(pn)}/\sqrt{n} \\ & \geqslant \mathrm{E}[(\eta' x_i)^2]/2 - 2\bar{\kappa}\sqrt{2k\log(pn)}/\sqrt{n} \end{array}$$

since $\tau = 1/6$ and $\mathrm{E}[(\eta' \tilde{x}_i)^2] \leqslant \mathrm{E}[(\eta' x_i)^2]$. So for $n \geqslant n_{02}^{SE} := 288k(\bar{\kappa}/\underline{\kappa})^2 \log(pn)$ we have

$$\phi_{\min}(s \log n)[\mathbb{E}_n[x_i x_i']] \geqslant \underline{\kappa}/3 =: \kappa'.$$

Similarly, we have

$$\begin{array}{rl} \mathbb{E}_n[(\eta' x_i)^2] & \leqslant \mathrm{E}[(\eta' x_i)^2] + |\mathbb{E}_n[(\eta' x_i)^2] - \mathrm{E}[(\eta' x_i)^2]| \\ & \leqslant \mathrm{E}[(\eta' x_i)^2] + |\mathbb{E}_n[(\eta' \tilde{x}_i)^2] - \mathrm{E}[(\eta' \tilde{x}_i)^2]| + 2|\eta' \mathbb{E}_n[\tilde{x}_i]| \cdot |\eta'\mu| \\ & \leqslant \mathrm{E}[(\eta' x_i)^2]\{1 + 2\tau + \tau^2\} + 2\bar{\kappa}\sqrt{2k\log(pn)}/\sqrt{n} \\ & \leqslant 2\mathrm{E}[(\eta' x_i)^2] + 2\bar{\kappa}\sqrt{2k\log(pn)}/\sqrt{n} \end{array}$$

since $\tau = 1/6$ and $\mathrm{E}[(\eta' \tilde{x}_i)^2] \leqslant \mathrm{E}[(\eta' x_i)^2]$. So for $n \geqslant n_{03}^{SE} := 2k\log(pn)$ we have

$$\phi_{\max}(s \log n)[\mathbb{E}_n[x_i x_i']] \leqslant 4\bar{\kappa} =: \kappa''.$$

The second and third requirements in Conditions SM(i) holds by the Gaussianity of w_i, $\mathrm{E}[\zeta_i \mid x_i, v_i] = 0$, $\mathrm{E}[v_i \mid x_i] = 0$, and the assumption that the minimal and maximum eigenvalues of the covariance matrix (operator) $\mathrm{E}[w_i w_i']$ are bounded below and above by positive absolute constants.

The first requirement in Condition SM(i) and Condition SM(ii) also hold by Gaussianity. Indeed, we have for $\epsilon_i = v_i$ and $\epsilon_i = \zeta_i$, $\tilde{y}_i = d_i$ and $\tilde{y}_i = y_i$

$$\begin{aligned}
\mathrm{E}[|v_i^q|] + \mathrm{E}[|\zeta_i^q|] &\leqslant 2^{q-1} G_q^q \{(\mathrm{E}[v_i^2])^{q/2} + (\mathrm{E}[\zeta_i^2])^{q/2}\} \leqslant 2^q G_q^q \bar{\kappa}^{q/2} =: A_1 \\
\mathrm{E}[|d_i^q|] &\leqslant 2^{q-1}\mathrm{E}[|\theta_m' z|^q] + 2^{q-1}\mathrm{E}[|v_i^q|] \leqslant 2^{q-1} G_q^q (\mathrm{E}[|\theta_m' z|^2])^{q/2} + 2^{q-1} G_q^q (\mathrm{E}[v_i^2])^{q/2} \\
&\leqslant 2^{q-1} G_q^q \|\theta_m\|^q \bar{\kappa}^{q/2} + 2^{q-1} G_q^q \bar{\kappa}^{q/2} \leqslant 2^q G_q^q \bar{\kappa}^{q/2}(1 + (2A)^q) =: A_2 \\
\mathrm{E}[d_i^2] &\leqslant 2\mathrm{E}[|\theta_m' z_i|^2] + 2\mathrm{E}[v_i^2] \leqslant 2\bar{\kappa}\|\theta_m\|^2 + 2\bar{\kappa} \leqslant 2\bar{\kappa}(4A^2 + 1) =: A_2' \\
\mathrm{E}[y_i^2] &\leqslant 3|\alpha_0|^2 \mathrm{E}[d_i^2] + 3\mathrm{E}[|\theta_m' z|^2] + 3\mathrm{E}[\zeta_i^2] \leqslant 3B^2 A_2' + 3A_2' + 3\bar{\kappa} =: A_3 \\
\max_{1\leqslant j\leqslant p} \mathrm{E}[x_{ij}^2 \tilde{y}_i^2] &\leqslant \max_{1\leqslant j\leqslant p} (\mathrm{E}[x_{ij}^4])^{1/2} (\mathrm{E}[\tilde{y}_i^4])^{1/2} \leqslant G_4^4 \max_{1\leqslant j\leqslant p} \mathrm{E}[x_{ij}^2]\mathrm{E}[\tilde{y}_i^2] \\
&\leqslant G_4^4 \bar{\kappa}(A_2' \vee A_3) =: A_4 \\
\max_{1\leqslant j\leqslant p} \mathrm{E}[|x_{ij}\epsilon_i|^3] &\leqslant \max_{1\leqslant j\leqslant p} (\mathrm{E}[x_{ij}^6])^{1/2} (\mathrm{E}[\epsilon_i^6])^{1/2} \leqslant G_6^6 \max_{1\leqslant j\leqslant p} (\mathrm{E}[x_{ij}^2])^{3/2} (\mathrm{E}[\epsilon_i^2])^{3/2} \\
&\leqslant G_6^6 \bar{\kappa}^3 =: A_5 \\
\max_{1\leqslant j\leqslant p} 1/\mathrm{E}[x_{ij}^2] &\leqslant 1/\lambda_{\min}(\mathrm{E}[w_i w_i']) \leqslant 1/\underline{\kappa} =: A_6
\end{aligned}$$

because $\|\theta_m\| \leqslant 2A$ and $\|\theta_g\| \leqslant 2A$ since $\theta_m, \theta_g \in S_A^a(p)$. Thus the first requirement in Condition SM(i) holds with $C_2^{SM} = A_2$. Condition SM(ii) holds with $C_3^{SM} = A_1 + (A_2' \vee A_3) + A_4 + A_5 + A_6$.

Condition SM(iii) is assumed.

To verify Condition SM(iv) note that for $\epsilon_i = v_i$ and $\epsilon_i = \zeta_i$, by (3.12), with probability $1 - \Delta_{1n}^{ASTE}$,

$$\begin{aligned}
\max_{j\leqslant p} \sqrt{\mathbb{E}_n[x_{ij}^4 \epsilon_i^4]} &\leqslant \max_{j\leqslant p} \sqrt[4]{\mathbb{E}_n[x_{ij}^8]}\sqrt[4]{\mathbb{E}_n[\epsilon_i^8]} \\
&\leqslant \{\sqrt{\bar{\kappa}} + \sqrt{2\bar{\kappa}\log(pn)}\}\ \max_{j\leqslant p} \sqrt[4]{\mathbb{E}_n[x_{ij}^4]}\sqrt[4]{\mathbb{E}_n[\epsilon_i^8]}.
\end{aligned} \tag{3.13}$$

By Lemma SA.1 with $k = 4$ we have with probability $1 - \Delta_{1n}^{SM}$, where $\Delta_{1n}^{SM} = 1/n$

$$\begin{aligned}
\max_{j\leqslant p} \sqrt[4]{\mathbb{E}_n[x_{ij}^4]} &\leqslant \|\mathrm{E}[x_i]\|_\infty + \max_{j\leqslant p} \sqrt[4]{\mathbb{E}_n[(x_{ij} - \mathrm{E}[x_{ij}])^4]} \\
&\leqslant \sqrt{\bar{\kappa}} + \sqrt{\bar{\kappa}} 2\bar{C} + \sqrt{\bar{\kappa}} n^{-1/4}\sqrt{2\log(2pn)} \leqslant 4\bar{C}\sqrt{\bar{\kappa}}
\end{aligned} \tag{3.14}$$

for $n \geqslant n_{01}^{SM} = 4\log^2(2pn)$. Also, Lemma SA.1 with $k = 8$ and $p = 1$ we have with probability $1 - \Delta_{1n}^{SM}$ that

$$\sqrt[4]{\mathbb{E}_n[\epsilon_i^8]} \leqslant 2\bar{\kappa} 8\bar{C}^2 + 2\bar{\kappa} n^{-1/4} 2\log(2n) \leqslant 20\bar{C}^2\bar{\kappa} \tag{3.15}$$

for $n \geqslant n_{02}^{SM} = 16\log^4(2n)$. Moreover, we have

$$\max_{1\leqslant j\leqslant p} \sqrt{\mathrm{E}[x_{ij}^4 \epsilon_i^4]} \leqslant \max_{1\leqslant j\leqslant p} \sqrt[4]{\mathrm{E}[x_{ij}^8]}\sqrt[4]{\mathrm{E}[\epsilon_i^8]} \leqslant G_8^4 \bar{\kappa}^2.$$

Applying Lemma SA.4, for $\tau = 2\Delta_{1n}^{ASTE} + \Delta_{1n}^{SM}$, with probability $1 - 8\tau$ we have

$$\max_{j \leqslant p} |(\mathbb{E}_n - \bar{\mathrm{E}})[x_{ij}^2 \epsilon_i^2]| \leqslant 4\sqrt{\frac{2\log(2p/\tau)}{n}} \sqrt{Q(\max_{1\leqslant j\leqslant p} \mathbb{E}_n[x_{ij}^4 \epsilon_i^4], 1-\tau)} \vee \frac{2\sqrt{2}G_8^4\bar{\kappa}^2}{\sqrt{n}}$$

where by (3.13), (3.14) and (3.15) we have

$$Q(\max_{1\leqslant j\leqslant p}\sqrt{\mathbb{E}_n[x_{ij}^4\epsilon_i^4]}, 1-\tau) \quad \leqslant \bar{\kappa}^2\sqrt{2\log(pn)}80\bar{C}^3.$$

So we let $\delta_{1n}^{SM} = 640\bar{C}^3\bar{\kappa}^2\sqrt{\frac{\log(2p/\tau)}{n}}\sqrt{\log(pn)} \vee 2\sqrt{2}\frac{G_8^4\bar{\kappa}^2}{\sqrt{n}} \to 0$ under the condition that $\log^2(p \vee n)/n \leqslant \bar{\delta}_n$.

Similarly for $\tilde{y}_i = d_i$ and $\tilde{y}_i = y_i$, by Lemma SA.1, we have with probability $1 - \Delta_{1n}^{SM}$, for $n \geqslant n_{02}^{SM}$ we have

$$\begin{aligned} \sqrt[8]{\mathbb{E}_n[\tilde{y}_i^8]} \quad & \leqslant |\mathrm{E}[\tilde{y}_i]| + \sqrt[8]{\mathbb{E}_n[(\tilde{y}_i - \mathrm{E}[\tilde{y}_i])^8]} \\ & \leqslant [A_2' \vee A_3]^{1/2} + (20\bar{C}^2\mathrm{E}[\tilde{y}_i^2])^{1/2} \leqslant 6\bar{C}[A_2' \vee A_3]^{1/2}. \end{aligned} \tag{3.16}$$

Moreover, $\sqrt[4]{\mathrm{E}[\tilde{y}_i^8]} \leqslant G_8^2\mathrm{E}[\tilde{y}_i^2] \leqslant G_8^2[A_2' \vee A_3]$. Therefore by Lemma SA.4, for $\tau = 2\Delta_{1n}^{ASTE} + \Delta_{2n}^{SM}$, with probability $1 - 8\tau$ we have by the arguments in (3.13), (3.14), and (3.16)

$$\max_{j\leqslant p}|(\mathbb{E}_n - \bar{\mathrm{E}})[x_{ij}^2\tilde{y}_i^2]| \leqslant 4\sqrt{\frac{2\log(2p/\tau)}{n}}\sqrt{6\bar{\kappa}\log(pn)}4\bar{C}\sqrt{\bar{\kappa}}(36\bar{C}^2[A_2' \vee A_3]) \vee \frac{2\sqrt{2}G_8^4\bar{\kappa}[A_2' \vee A_3]}{\sqrt{n}} =: \delta_{2n}^{SM}$$

where $\delta_{2n}^{SM} \to 0$ under the condition $\log^2(p \vee n)/n \leqslant \bar{\delta}_n \to 0$.

We have that the last term in Condition SM(iv) satisfies with probability $1 - \Delta_{1n}^{ASTE}$

$$\max \|x_i\|_\infty^2 \frac{s\log(p \vee n)}{n} \leqslant 6\bar{\kappa}\log(pn)A^{1/a}n^{-1+1/2a}\log(p \vee n) =: \delta_{3n}^{SM}.$$

Under ASTE(iv) and $s = A^{1/a}n^{1/2a}$ we have $\delta_{3n}^{SM} \to 0$.

Finally, we set $n_0 = \max\{n_{01}^{ASTE}, n_{01}^{SE}, n_{02}^{SE}, n_{03}^{SE}, n_{01}^{SM}, n_{02}^{SM}\}$, $C = \max\{C_1^{ASTE}, C_2^{ASTE}, 2C_3^{ASTE}, C_1^{SM}, C_2^{SM}\}$, $\delta_n = \max\{\bar{\delta}_n, \delta_{1n}^{ASTE}, \delta_{2n}^{ASTE}, \delta_{1n}^{SM} + \delta_{2n}^{SM} + \delta_{3n}^{SM}\} \to 0$, and $\Delta_n = \max\{33\Delta_{1n}^{ASTE} + 16\Delta_{1n}^{SM}, \Delta_{1n}^{SE}\} \to 0$.

□

Example 3. Let $\mathbf{P}$ be the collection of all regression models P that obey the conditions set forth above for all n for the given constants $(\underline{f}, \bar{f}, a, A, b, B, q)$ and the sequence $\bar{\delta}_n$. Below we provide explicit bounds for κ', κ'', c, C, δ_n and Δ_n that appear in Conditions ASTE, SE and SM that depend only on $(\underline{f}, \bar{f}, a, A, b, B, q)$ and $\bar{\delta}_n$ which in turn establish these conditions for all $\mathrm{P} \in \mathbf{P}$.

Conditions ASTE(i) is assumed. Condition ASTE(ii) holds with $|\alpha_0| \leqslant B =: C_1^{ASTE}$. Because $\theta_m, \theta_g \in S_A^a(p)$, Condition ASTE(iii) holds with

$$s = A^{1/a}n^{\frac{1}{2a}}, \quad r_{mi} = m(z_i) - \sum_{j=1}^p \beta_{m0j}P_j(z_i) \quad \text{and} \quad r_{gi} = g(z_i) - \sum_{j=1}^p \beta_{g0j}P_j(z_i)$$

where $\|\beta_{m0}\|_0 \leqslant s$ and $\|\beta_{g0}\|_0 \leqslant s$. Indeed, we have

$$\mathrm{E}[r_{mi}^2] \leqslant \mathrm{E}\left[\left(\sum_{j\geqslant s+1}\theta_{m(j)}P_{(j)}(z_i)\right)^2\right] \leqslant \bar{f}\sum_{j\geqslant s+1}\theta_{m(j)}^2 \leqslant \bar{f}A^2s^{-2a+1}/[2a-1] = \bar{f}s/n$$

where the first inequality follows by the definition of β_{m0} in (4.31), the second inequality follows from the upper bound on the density and orthogonality of the basis, the third inequality follows from $\theta_m \in S_A^a(p)$, and the last inequality because $s = A^{1/a}n^{1/2a}$. Similarly we have $\mathrm{E}[r_{gi}^2] \leqslant \mathrm{E}[(\sum_{j\geqslant s+1}\theta_{g(j)}z_{i(j)})^2] \leqslant \bar{f}A^2s^{-2a+1}/[2a-1] = \bar{f}s/n$. Let $C_2^{ASTE} = \sqrt{\bar{f}}$.

Condition ASTE(iv) holds with $\delta_{1n}^{ASTE} := A^{2/a}n^{1/a-1}\log^2(p\vee n) \to 0$ since $s = A^{1/a}n^{1/2a}$, A is fixed, and the assumed condition $n^{(1-a)/a}\log^2(p\vee n) \leqslant \bar{\delta}_n \to 0$.

Next we establish the moment restrictions in Condition ASTE(v). Because $\underline{f} \leqslant \lambda_{\min}(\mathrm{E}[x_ix_i']) \leqslant \lambda_{\max}(\mathrm{E}[x_ix_i']) \leqslant \bar{f}$, by the assumption on the density and orthonormal basis, and $\max_{i\leqslant n}\|x_i\|_\infty \leqslant B$, by Lemma SA.2 with $\rho_i = 0$ we have

$$\max_{1\leqslant i\leqslant n}|r_{mi}|\vee|r_{gi}| \leqslant \max_{1\leqslant i\leqslant n}\|x_i\|_\infty(\bar{f}/\underline{f})^{3/2}\frac{2a-1}{a-1}\sqrt{s^2/n} \leqslant B(\bar{f}/\underline{f})^{3/2}\frac{2a-1}{a-1}\sqrt{s^2/n} =: \delta_{2n}^{ASTE}$$

where $\delta_{2n}^{ASTE} \to 0$ under $s = A^{1/a}n^{1/2a}$ and $a > 1$.

Thus we have

$$\begin{aligned}\mathrm{E}[|\tilde{\zeta}_i|^q] &\leqslant 2^{q-1}\mathrm{E}[|\zeta_i^q|] + 2^{q-1}\mathrm{E}[|r_{gi}^q|] \leqslant 2^{q-1}B + 2^{q-1}(\delta_{2n}^{ASTE})^q\\ &\leqslant 2^{q-1}B + 2^{q-1}(\delta_{2n_0}^{ASTE})^q =: C_3^{ASTE}.\end{aligned}$$

Similarly, $\mathrm{E}[|\tilde{v}_i|^q] \leqslant C_3^{ASTE}$. Moreover, since $\delta_{2n}^{ASTE} \to 0$ we have

$$\begin{aligned}|\mathrm{E}[\tilde{\zeta}_i^2\tilde{v}_i^2] - \mathrm{E}[\zeta_i^2v_i^2]| &\leqslant \mathrm{E}[\zeta_i^2r_{mi}^2] + \mathrm{E}[r_{gi}^2v_i^2] + \mathrm{E}[r_{mi}^2r_{gi}^2]\\ &\leqslant \sqrt{\mathrm{E}[\zeta_i^4]\mathrm{E}[r_{mi}^4]} + \sqrt{\mathrm{E}[r_{gi}^4]\mathrm{E}[v_i^4]} + \sqrt{\mathrm{E}[r_{mi}^4]\mathrm{E}[r_{gi}^4]}\\ &\leqslant 2B^{2/q}(\delta_{2n}^{ASTE})^2 + (\delta_{2n}^{ASTE})^4 =: \delta_{3n}^{ASTE} \to 0.\end{aligned}$$

Finally, the last requirement holds because $(1-a)/a + 4/q < 0$ implies

$$\max_{i\leqslant n}\|x_i\|_\infty^2 sn^{-1/2+2/q} \leqslant B^2A^{1/a}n^{1/2a-1/2+2/q} =: \delta_{4n}^{ASTE} \to 0,$$

since $s = A^{1/a}n^{1/2a}$ and $\max_{i\leqslant n}\|x_i\|_\infty \leqslant B$.

To show Condition SE with $\ell_n = \log n$ note that regressors are uniformly bounded, and minimal and maximal eigenvalues of $\mathrm{E}[x_ix_i']$ are bounded below by $\underline{f}$ and above by $\bar{f}$ uniformly in n. Thus Condition SE follows by Corollary 4 in the supplementary material in Belloni and Chernozhukov (2013) (restated in Lemma SA.7) which is based on Rudelson and Vershynin (2008). Let

$$\delta_{1n}^{SE} := 2\bar{C}B\sqrt{s\log n}\log(1+s\log n)\sqrt{\log(p\vee n)}\sqrt{\log n}/\sqrt{n}$$

and $\Delta_{1n}^{SE} := (2/\underline{f})(\delta_{1n}^{SE})^2 + \delta_{1n}^{SE}(2\bar{f}/\underline{f})$, where $\bar{C}$ is an universal constant. By this result and the Markov inequality, we have with probability $1 - \Delta_{1n}^{SE}$

$$\kappa' := \underline{f}/2 \leqslant \phi_{\min}(s \log n)[\mathbb{E}_n[x_i x_i']] \leqslant \phi_{\max}(s \log n)[\mathbb{E}_n[x_i x_i']] \leqslant 2\bar{f} =: \kappa''.$$

We need to show that $\Delta_{1n}^{SE} \to 0$ which follows from $\delta_{1n}^{SE} \to 0$. We have that

$$\delta_{1n}^{SE} \leqslant \frac{2\bar{C}B(1+A)^2\sqrt{n^{1/2a}}\log^2(n)\sqrt{\log(p\vee n)}}{\sqrt{n}} = 2\bar{C}B(1+A)^2\sqrt{\frac{n^{1/2a}\log^4 n}{n^{2/3}}}\sqrt{\frac{\log(p\vee n)}{n^{1/3}}}.$$

By assumption we have $\log^3 p/n \leqslant \bar{\delta}_n \to 0$ and $a > 1$ we have $\delta_{1n}^{SE} \to 0$.

The second and third requirements in Condition SM(i) hold with $C_1^{SM} = B^{2/q}$ and $c_1^{SM} = b$ by assumption. Condition SM(iii) is assumed.

The first requirement in Condition SM(i) and Condition SM(ii) follow by, for $\epsilon_i = v_i$ and $\epsilon_i = \zeta_i$, $\tilde{y}_i = d_i$ and $\tilde{y}_i = y_i$

$$\begin{aligned}
\mathrm{E}[|v_i^q|] + \mathrm{E}[|\zeta_i^q|] &\leqslant 2B =: A_1\\
\mathrm{E}[|d_i^q|] &\leqslant 2^{q-1}\mathrm{E}[|\theta_m' x_i|^q] + 2^{q-1}\mathrm{E}[|v_i^q|] \leqslant 2^{q-1}\|\theta_m\|_1^q \mathrm{E}[\|x_i\|_\infty^q] + 2^{q-1}B\\
&\leqslant 2^{q-1}(2A)^q B^q + 2^{q-1}B =: A_2\\
\mathrm{E}[d_i^2] &\leqslant 2\bar{f}\|\theta_m\|^2 + 2\mathrm{E}[v_i^2] \leqslant 8\bar{f}A^2 + 2B^{2/q} =: A_2'\\
\mathrm{E}[y_i^2] &\leqslant 3|\alpha_0|^2\mathrm{E}[d_i^2] + 3\|\theta_g\|_1^2\mathrm{E}[\|x_i\|_\infty^2] + 3\mathrm{E}[\zeta_i^2]\\
&\leqslant 3B^2A_2' + 12A^2B^2 + 3B^{2/q} =: A_3\\
\max_{1\leqslant j\leqslant p}\mathrm{E}[x_{ij}^2\tilde{y}_i^2] &\leqslant B^2\mathrm{E}[\tilde{y}_i^2] \leqslant B^2(A_2' \vee A_3) =: A_4\\
\max_{1\leqslant j\leqslant p}\mathrm{E}[|x_{ij}\epsilon_i|^3] &\leqslant B^3\mathrm{E}[|\epsilon_i^3|] \leqslant B^3B^{3/q} =: A_5\\
\max_{1\leqslant j\leqslant p} 1/\mathrm{E}[x_{ij}^2] &\leqslant 1/\lambda_{\min}(\mathrm{E}[x_i x_i']) \leqslant 1/\underline{f} =: A_6
\end{aligned}$$

where we used that $\max_{i\leqslant n}\|x_i\|_\infty \leqslant B$, the moment assumptions of the disturbances, $\|\theta_m\| \leqslant \|\theta_m\|_1 \leqslant 2A$, $\|\theta_g\|_1 \leqslant 2A$ since $\theta_m, \theta_g \in S_A^a(p)$ for $a > 1$. Thus the first requirement in Condition SM(i) holds with $C_2^{SM} = A_2$. Condition SM(ii) holds with $C_3^{SM} := A_1 + (A_2' \vee A_3) + A_4 + A_5 + A_6$.

To verify Condition SM(iv) note that for $\epsilon_i = v_i$ and $\epsilon_i = \zeta_i$ we have by Lemma SA.4 with probability $1 - 8\tau$, where $\tau = 1/\log n$,

$$\begin{aligned}
\max_{1\leqslant j\leqslant p}|(\mathbb{E}_n - \bar{\mathrm{E}})[x_{ij}^2\epsilon_i^2]| &\leqslant 4\sqrt{\tfrac{2\log(2p/\tau)}{n}}Q(\max_{1\leqslant j\leqslant p}\sqrt{\mathbb{E}_n[x_{ij}^4\epsilon_i^4]}, 1-\tau) \vee \frac{2\max_{1\leqslant j\leqslant p}\sqrt{2\mathrm{E}[x_{ij}^4\epsilon_i^4]}}{\sqrt{n}}\\
&\leqslant 4\sqrt{\tfrac{2\log(2p/\tau)}{n}}B^2Q(\sqrt{\mathbb{E}_n[\epsilon_i^4]}, 1-\tau) \vee \frac{2B^2\sqrt{2\mathrm{E}[\epsilon_i^4]}}{\sqrt{n}}\\
&\leqslant 4\sqrt{\tfrac{2\log(2p\log n)}{n}}B^2B^{2/q}\log n =: \delta_{1n}^{SM}
\end{aligned}$$

where we used $\mathrm{E}[\epsilon_i^4] \leqslant B^{4/q}$ and the Markov inequality. By the definition of τ and the assumed rate $\log^3(p\vee n)/n \leqslant \bar{\delta}_n \to 0$, we have $\delta_{1n}^{SM} \to 0$.

Similarly, we have for $\tilde y_i = d_i$ and $\tilde y_i = y_i$, with probability $1-8\tau$

$$\begin{aligned}
\max_{1\leqslant j\leqslant p}|(\mathbb{E}_n-\bar{\mathrm{E}})[x_{ij}^2\tilde y_i^2]| &\leqslant 4\sqrt{\tfrac{2\log(2p/\tau)}{n}}Q(\max_{1\leqslant j\leqslant p}\sqrt{\mathbb{E}_n[x_{ij}^4\tilde y_i^4]},1-\tau)\vee \frac{2\max_{1\leqslant j\leqslant p}\sqrt{2\mathrm{E}[x_{ij}^4\tilde y_i^4]}}{\sqrt{n}}\\
&\leqslant 4\sqrt{\tfrac{2\log(2p/\tau)}{n}}B^2Q(\sqrt{\mathbb{E}_n[\tilde y_i^4]},1-\tau)\vee\tfrac{2B^2\sqrt{2\mathrm{E}[\tilde y_i^4]}}{\sqrt{n}}\\
&\leqslant 4\sqrt{\tfrac{2\log(2p\log n)}{n}}B^2A_7\log n =: \delta_{2n}^{SM}
\end{aligned}$$

where we used the Markov inequality and

$$\begin{aligned}
\mathrm{E}[\tilde y_i^4] &\leqslant \mathrm{E}[d_i^4]+3^3|\alpha_0|^4\mathrm{E}[d_i^4]+3^3\|\theta_g\|_1^4\mathrm{E}[\|x_i\|_\infty^4]+3^3\mathrm{E}[\zeta_i^4]\\
&\leqslant A_2^{4/q}+3^3B^4A_2^{4/q}+3^3(2A)^4B^4+3^3B^{4/q}=:A_7.
\end{aligned}$$

By the definition of τ and the assumed rate $\log^3(p\vee n)/n\leqslant\bar\delta_n\to 0$, we have $\delta_{2n}^{SM}\to 0$.

The last term in the requirement of Condition SM(iv), because $\max_{i\leqslant n}\|x_i\|_\infty\leqslant B$ and Condition ASTE(iv) holds, is bounded by $\delta_{3n}^{SM}:=B^2A^{1/a}n^{1/2a}\log(p\vee n)/n\to 0$.

Finally, we set $c=c_1^{SM}$, $C=\max\{C_1^{ASTE},C_2^{ASTE},2C_3^{ASTE},C_1^{SM},C_2^{SM},C_3^{SM}\}$,

$$\delta_n=\max\{\bar\delta_n,\delta_{1n}^{ASTE},\delta_{2n}^{ASTE},\delta_{3n}^{ASTE},\delta_{4n}^{ASTE},\delta_{1n}^{SM}+\delta_{2n}^{SM}+\delta_{3n}^{SM}\}\to 0,$$

$\Delta_n=\max\{16/\log n,\Delta_{1n}^{SE}\}\to 0.$ $\square$

4. Additional Tools

4.1. **Inequalities based on Symmetrization.** Next we proceed to use symmetrization arguments to bound the empirical process. In what follows for a random variable Z let $Q(Z,1-\tau)$ denote its $(1-\tau)$-quantile.

Lemma SA.3 (Maximal inequality via symmetrization)**.** *Let $Z_1,\ldots,Z_n$ be arbitrary independent stochastic processes and $\mathcal{F}$ a finite set of measurable functions. For any $\tau\in(0,1/2)$, and $\delta\in(0,1)$ we have that with probability at least $1-4\tau-4\delta$*

$$\max_{f\in\mathcal{F}}|\mathbb{G}_n(f(Z_i))|\leqslant\left\{4\sqrt{2\log(2|\mathcal{F}|/\delta)}\;Q\left(\max_{f\in\mathcal{F}}\sqrt{\mathbb{E}_n[f(Z_i)^2]},1-\tau\right)\right\}\vee 2\max_{f\in\mathcal{F}}Q\left(|\mathbb{G}_n(f(Z_i))|,\frac12\right).$$

Proof. Let

$$e_{1n}=\sqrt{2\log(2|\mathcal{F}|/\delta)}\;Q\left(\max_{f\in\mathcal{F}}\sqrt{\mathbb{E}_n[f(Z_i)^2]},1-\tau\right),\quad e_{2n}=\max_{f\in\mathcal{F}}Q\left(|\mathbb{G}_n(f(Z_i))|,\frac12\right)$$

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

$$\begin{aligned}
\mathbb{P}\{\max_{f\in\mathcal{F}}|\mathbb{G}_n(f(Z_i))|>4e_{1n}\vee 2e_{2n}\} &\leqslant 4\mathbb{P}\{\max_{f\in\mathcal{F}}|\mathbb{G}_n(\varepsilon_if(Z_i))|>e_{1n}\}\\
&\leqslant 4\mathbb{P}\{\max_{f\in\mathcal{F}}|\mathbb{G}_n(\varepsilon_if(Z_i))|>e_{1n}|\mathcal{E}\}+4\tau
\end{aligned}$$

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

Thus a union bound yields

$$\mathbb{P}\left\{\max_{f\in\mathcal{F}}|\mathbb{G}_n(f(Z_i))| > 4e_{1n}\vee 2e_{2n}\right\} \leqslant 4\tau + 4|\mathcal{F}|\max_{f\in\mathcal{F}}\mathbb{P}\left\{|\mathbb{G}_n(\varepsilon_i f(Z_i))| > e_{1n}|\mathcal{E}\right\}. \tag{4.17}$$

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 $f\in\mathcal{F}$, $\mathbb{P}_\varepsilon\{|\mathbb{G}_n(\varepsilon_i f(Z_i))| > x\} \leqslant 2\exp(-x^2/\{2\mathbb{E}_n[f^2(Z_i)]\})$. Hence, under the event $\mathcal{E}$, we can bound

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

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

The following specialization will be convenient.

Lemma SA.4. *Let $\tau\in(0,1)$ and $\{(x_i',\epsilon_i)'\in\mathbb{R}^p\times\mathbb{R}, i=1,\dots,n\}$ be random vectors that are independent across i. Then with probability at least $1-8\tau$*

$$\max_{1\leqslant j\leqslant p}|\mathbb{E}_n[x_{ij}^2\epsilon_i^2]-\bar{\mathrm{E}}[x_{ij}^2\epsilon_i^2]| \leqslant 4\sqrt{\frac{2\log(2p/\tau)}{n}\ Q\left(\max_{1\leqslant j\leqslant p}\mathbb{E}_n[x_{ij}^4\epsilon_i^4],1-\tau\right)} \vee 2\max_{1\leqslant j\leqslant p}\sqrt{\frac{2\bar{\mathrm{E}}[x_{ij}^4\epsilon_i^4]}{n}}$$

Proof. Let $Z_i = x_i\epsilon_i$, $f_j(Z_i) = x_{ij}^2\epsilon_i^2$, $\mathcal{F}=\{f_1,\dots,f_p\}$, so that $n^{-1/2}\mathbb{G}_n(f_j(Z_i)) = \mathbb{E}_n[x_{ij}^2\epsilon_i^2]-\bar{\mathrm{E}}[x_{ij}^2\epsilon_i^2]$. Also, for $\tau_1\in(0,1/2)$ and $\tau_2\in(0,1)$, let

$$e_{1n} = \sqrt{2\log(2p/\tau_1)}\sqrt{Q\left(\max_{1\leqslant j\leqslant p}\mathbb{E}_n[x_{ij}^4\epsilon_i^4],1-\tau_2\right)}\ \text{ and }\ e_{2n} = \max_{1\leqslant j\leqslant p}Q(|\mathbb{G}_n(x_{ij}^2\epsilon_i^2)|,1/2)$$

where we have $e_{2n}\leqslant \max_{1\leqslant j\leqslant p}\sqrt{2\bar{\mathrm{E}}[x_{ij}^4\epsilon_i^4]}$ by Chebyshev.

By Lemma SA.3 we have

$$P\left(\max_{1\leqslant j\leqslant p}|\mathbb{E}_n[x_{ij}^2\epsilon_i^2]-\bar{\mathrm{E}}[x_{ij}^2\epsilon_i^2]| > \frac{4e_{1n}\vee 2e_{2n}}{\sqrt{n}}\right) \leqslant 4\tau_1+4\tau_2.$$

The result follows by setting $\tau_1=\tau_2=\tau<1/2$. Note that for $\tau\geqslant 1/2$ the result is trivial. □

4.2. **Moment Inequality.** We shall be using the following result, which is based on Markov inequality and (von Bahr and Esseen 1965).

Lemma SA.5 (Vonbahr-Esseen's LLN)**.** *Let $r\in[1,2]$, and independent zero-mean random variables X_i with $\bar{\mathrm{E}}[|X_i|^r]\leqslant C$. Then for any $\ell_n>0$*

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

4.3. **Matrices Deviation Bounds.** In this section we collect matrices deviation bounds. We begin with a bound due to Rudelson (1999) for the case that $p < n$.

Lemma SA.6 (Essentially in Rudelson (1999))**.** *Let x_i, $i = 1, \ldots, n$, be independent random vectors in $\mathbb{R}^p$ and set*

$$\delta_n := \bar{C}\frac{\sqrt{\log(n \wedge p)}}{\sqrt{n}}\sqrt{\mathrm{E}[\max_{1\leqslant i\leqslant n}\|x_i\|^2]}.$$

for some universal constant $\bar{C}$. Then, we have

$$\mathrm{E}\left[\sup_{\|\alpha\|=1}\left|\mathbb{E}_n\left[(\alpha' x_i)^2 - \mathrm{E}[(\alpha' x_i)^2]\right]\right|\right] \leqslant \delta_n^2 + \delta_n \sup_{\|\alpha\|=1}\sqrt{\bar{\mathrm{E}}[(\alpha' x_i)^2]}.$$

Based on results in Rudelson and Vershynin (2008), the following lemma for bounded regressors was derived in the supplementary material of Belloni and Chernozhukov (2013)

Lemma SA.7 (Essentially in Theorem 3.6 of Rudelson and Vershynin (2008))**.** *Let x_i, $i = 1, \ldots, n$, be independent random vectors in $\mathbb{R}^p$ be such that $\sqrt{\mathrm{E}[\max_{1\leqslant i\leqslant n}\|x_i\|_\infty^2]} \leqslant K$. Let*

$$\delta_n := 2\left(\bar{C}K\sqrt{k}\log(1+k)\sqrt{\log(p\vee n)}\sqrt{\log n}\right)/\sqrt{n},$$

where $\bar{C}$ is the universal constant. Then,

$$\mathrm{E}\left[\sup_{\|\alpha\|_0\leqslant k, \|\alpha\|=1}\left|\mathbb{E}_n\left[(\alpha' x_i)^2 - \mathrm{E}[(\alpha' x_i)^2]\right]\right|\right] \leqslant \delta_n^2 + \delta_n \sup_{\|\alpha\|_0\leqslant k, \|\alpha\|=1}\sqrt{\bar{\mathrm{E}}[(\alpha' x_i)^2]}.$$

Proof. Let

$$V_k = \sup_{\|\alpha\|_0\leqslant k, \|\alpha\|=1}\left|\mathbb{E}_n\left[(\alpha' x_i)^2 - \mathrm{E}[(\alpha' x_i)^2]\right]\right|.$$

Then, by a standard symmetrization argument (Guédon and Rudelson (2007), page 804)

$$n\mathrm{E}[V_k] \quad \leqslant 2\mathrm{E}_x\mathrm{E}_\varepsilon\left[\sup_{\|\alpha\|_0\leqslant k, \|\alpha\|=1}\left|\sum_{i=1}^n \varepsilon_i(\alpha' x_i)^2\right|\right].$$

Letting

$$\phi(k) = \sup_{\|\alpha\|_0\leqslant k, \|\alpha\|\leqslant 1}\mathbb{E}_n[(\alpha' x_i)^2] \text{ and } \varphi(k) = \sup_{\|\alpha\|_0\leqslant k, \|\alpha\|=1}\bar{\mathrm{E}}[(\alpha' x_i)^2],$$

we have $\phi(k) \leqslant \varphi(k) + V_k$ and by Lemma 3.8 in Rudelson and Vershynin (2008) to bound the expectation in ε,

$$\begin{aligned} n\mathrm{E}[V_k] &\leqslant 2\left(\bar{C}\sqrt{k}\log(1+k)\sqrt{\log(p\vee n)}\sqrt{\log n}\right)\sqrt{n}\mathrm{E}_x\left[\max_{i\leqslant n}\|x_i\|_\infty\sqrt{\phi(k)}\right] \\ &\leqslant 2\left(\bar{C}\sqrt{k}\log(1+k)\sqrt{\log(p\vee n)}\sqrt{\log n}\right)\sqrt{n}\sqrt{\mathrm{E}_x\left[\max_{i\leqslant n}\|x_i\|_\infty^2\right]\mathrm{E}_x\left[\phi(k)\right]} \\ &\leqslant 2\left(\bar{C}K\sqrt{k}\log(1+k)\sqrt{\log(p\vee n)}\sqrt{\log n}\right)\sqrt{n}\sqrt{\varphi(k) + \mathrm{E}[V_k]}. \end{aligned}$$

The result follows by noting that for positive numbers v, A, B, $v \leqslant A(v+B)^{1/2}$ implies $v \leqslant A^2 + A\sqrt{B}$. $\square$

The following result establishes an approximation bound for sub-Gaussian regressors and was developed in Rudelson and Zhou (2011). Recall that a random vector $Z \in \mathbb{R}^p$ is isotropic if $\mathrm{E}[ZZ'] = I$, and it is called ψ_2 with a constant α if for every $w \in \mathbb{R}^p$ we have

$$\|Z'w\|_{\psi_2} := \inf\{t : \mathrm{E}[\exp(\ (Z'w)^2/t^2)] \leqslant 2\} \leqslant \alpha\|w\|_2.$$

Lemma SA.8 (Essentially in Theorem 3.2 of Rudelson and Zhou (2011))**.** *Let Ψ_i, $i = 1, \ldots, n$, be i.i.d. isotropic random vectors in $\mathbb{R}^p$ that are ψ_2 with a constant α. Let $x_i = \Sigma^{1/2}\Psi_i$ so that $\Sigma = \mathrm{E}[x_i x_i']$. For $m \leqslant p$ and $\tau \in (0,1)$ assume that*

$$n \geqslant \frac{80m\alpha^4}{\tau^2} \log\left(\frac{12ep}{m\tau}\right).$$

Then with probability at least $1 - 2\exp(-\tau^2 n/80\alpha^4)$, for all $u \in \mathbb{R}^p$, $\|u\|_0 \leqslant m$, we have

$$(1-\tau)\|\Sigma^{1/2}u\|_2 \leqslant \sqrt{\mathbb{E}_n[(x_i'u)^2]} \leqslant (1+\tau)\|\Sigma^{1/2}u\|_2.$$

For example, Lemma SA.8 covers the case of $x_i \sim N(0,\Sigma)$ by setting $\Psi_i \sim N(0,I)$ which is isotropic and ψ_2 with a constant $\alpha = \sqrt{8/3}$.

5. Empirical Example: Estimating the Effect of Abortion on Crime: Results in Levels

In this section, we expand on the discussion of the empirical section in the main paper by considering estimation of the effect of abortion on crime in levels. We consider both the original model of Donohue III and Levitt (2001) as well as the model from Donohue III and Levitt (2008) which responds to a criticism raised in Foote and Goetz (2008) which is similar to the conclusion we draw in the original data. The results using variable selection show that the results in Donohue III and Levitt (2008) also become imprecise once one considers a broad set of controls and selects among them using our variable selection technique.

For our analysis in this appendix, we follow Donohue III and Levitt (2001) and rely on the argument that abortion rates may be taken as exogenous relative to crime rates conditional upon a set of factors. Unlike Donohue III and Levitt (2001), we do not assume that the identity of these factors is known and allow for smooth, flexible trends to account for unobservable factors that may influence both abortion and crime but smoothly trend over time. Given the seemingly obvious importance of controlling for state and time effects, we account for these in all models we estimate by including a full set of state and time dummies. Thus, we estimate models of the form

$$y_{it} = \alpha a_{it} + w_{it}'\beta_y + \delta_{y,i} + \gamma_{y,t} + g(z_{it}, t) + \zeta_{it} \tag{5.18}$$

$$a_{it} = w_{it}'\beta_a + \delta_{a,i} + \gamma_{a,t} + m(z_{it}, t) + v_{it} \tag{5.19}$$

where $g(z,t)$ and $m(z,t)$ are smooth functions of observed variables z_{it} which includes w_{it}, time-invariant characteristics of $\{y_{it}, a_{it}, w_{it}\}_{t=1}^{T}$ such as initial conditions or state-level averages, and time. We use the same state-level data as Donohue III and Levitt (2001) but delete Alaska, Hawaii, and Washington, D.C. which gives a sample with 48 cross-sectional observations and 13 time series observations for a total of 624 observations. With these deletions, our baseline estimates using the same controls as in baseline results from Donohue III and Levitt (2001) are quite similar to those reported in Donohue III and Levitt (2001). Baseline estimates from Table IV of Donohue III and Levitt (2001) and our baseline estimates are given in the first and second row of Panel A of Table L.[1]

Note that interpreting estimates of the effect of abortion using the same controls as in Donohue III and Levitt (2001) as causal relies on the belief that there are no higher-order terms of the control variables, no interaction terms, and no additional excluded variables that are associated both to crime rates and the associated abortion rate. Allowing for such variables is important in that one might believe that there may be some feature of a state that is associated both with its growth rate in abortion and its growth rate in crime. For example, having an initially high-level of abortion could be associated with having high-growth rates in abortion and low growth rates in crime. Failure to control for this factor could then lead to misattributing the effect of this initial factor, perhaps driven by policy or state-level demographics, to the effect of abortion. In practice, it is common to account for this possibility by allowing state-specific trends (e.g. by specifying $g(z_{it}, t) = \kappa_{g,i} t$) in addition to state-specific intercepts. Results from estimating the baseline model augmented with state-specific trends are given in the third row in Table L Panel A. In this example, the inclusion of state-specific linear trends renders the results very imprecise. Of course, one might argue that including state-specific linear trends is too aggressive in a sample with only 13 time series observations. The linear trend specification is also very restrictive in imposing that any unobserved factors that relate to both abortion and crime exhibit constant growth over the 13 year time period. The assumption of constant growth becomes even more problematic when one expands the time period as in Foote and Goetz (2008) and Donohue III and Levitt (2008) discussed below.

We follow the Chamberlain (1985) approach and approximate $g(z_{it}, t)$ and $m(z_{it}, t)$ by a large number of controls. We approximate these functions by forming 27 factors to include in z_{it},

$$z_{it} = (a_{i0},\ \frac{1}{T}\sum_t a_{it},\ y_{i0},\ w_{i0}',\ \frac{1}{T}\sum_t w_{it}',\ w_{it}')',$$

forming nine smooth function of time,

$$f_t = (t,\ t^2,\ t^3,\ \sin(\pi\frac{t}{T}),\ \sin(2\pi\frac{t}{T}),\ \sin(3\pi\frac{t}{T}),\ \cos(\pi\frac{t}{T}),\ \cos(2\pi\frac{t}{T}),\ \cos(3\pi\frac{t}{T}))',$$

[1]Our estimates differ for three reasons. First, we delete Alaska, Hawaii, and Washington, D.C. Second, Donohue III and Levitt (2001) use population weighted estimates. Third, Donohue III and Levitt (2001) use an FGLS estimator based on an AR(1) model in the errors where the errors across states share the same AR coefficient.

and then supposing that

$$g(z_{it},t) \approx \sum_{r=1}^{27}\sum_{s=1}^{9} \beta_{g,r,s} z_{it,r} f_{t,s} = h_{it}'\beta_g \qquad \text{and}$$

$$m(z_{it},t) \approx \sum_{r=1}^{27}\sum_{s=1}^{9} \beta_{m,r,s} z_{it,r} f_{t,s} = h_{it}'\beta_m$$

where h_{it} is the vector containing all the interactions, and β_g and β_m are the vectors of coefficients for each equation. That is, we add an additional 243 control variables to the model and use the methods developed in this paper to search among these 243 additional control variables.[2] With this set of controls, the models we estimate are all more general than the baseline model using the same controls as in Donohue III and Levitt (2001) and are neither more nor less general than a model with state-specific trends in that we allow for nonlinearity in trends but do not allow for arbitrarily different state-specific coefficients. Rather, we restrict these coefficients to differ depending on values of observable covariates.

Controlling for a large set of variables as described above is desirable from the standpoint of making the belief underlying the causal interpretation of the abortion coefficient, that the abortion rate defined above may be taken as being as good as randomly assigned once the set of variables considered is controlled for, more plausible. As with the inclusion of state-specific trends, the downside is that controlling for many variables lessens our ability to identify the effect of interest and thus tends to make estimates far less precise. For example, the estimated abortion effects conditioning on the full set of 68 variables used in Donohue III and Levitt (2001) plus the 243 approximating functions (for a total of 311 control variables) are given in the fourth row of Table L Panel A. As expected, all coefficients are estimated very imprecisely. Of course, very few researchers would consider using 311 controls with only 624 observations due to exactly this issue.

We are faced with a trade-off between the precision of the estimate and the plausibility of the conditional exogeneity assumption. By including additional controls in the specification, we make the conditional exogeneity assumption more plausible. At the same time, we potentially reduce the precision of our estimate. The double selection method proposed in this paper offers one rigorous approach to achieving a balance. Thus, the approach complements the usual careful specification analysis by providing a researcher a simple-to-implement, data-driven way to search for a set of influential confounds from among a sensibly chosen broader set of potential confounding variables.

[2]To allow time effects, state effects, and w_{it} to enter each equation without shrinkage, we use our methods based on $\tilde{y}_{it}$, $\tilde{a}_{it}$ and $\tilde{h}_{it}$ where $\tilde{y}_{it}$ is the residual from the regression of y_{it} on w_{it} and a full set of state and time dummies and $\tilde{a}_{it}$ and $\tilde{h}_{it}$ are defined similarly.

Table L. Estimated Effects of Abortion on Crime Rates (Levels)

	Violent Crime		Property Crime		Murder	
	Effect	Std. Err.	Effect	Std. Err.	Effect	Std. Err.
	A. Donohue and Levitt (2001) Table IV					
DL (2001) Table IV	-0.129	0.024	-0.091	0.018	-0.121	0.047
Fixed Effects	-0.131	0.045	-0.091	0.016	-0.131	0.058
Fixed Effects + State Trends	-0.149	0.185	0.060	0.093	-0.383	0.207
All Controls	0.183	0.447	0.013	0.067	0.855	0.974
Post-Double-Selection	0.133	0.303	-0.053	0.044	-0.692	0.438
Polynomial Trend	0.321	0.349	-0.032	0.060	0.851	0.616
Post-Double-Selection, Polynomial Trend	0.013	0.251	-0.041	0.047	-0.178	0.276
	B. Donohue and Levitt (2008) Table III					
DL (2008) Table III	-0.160	0.088	-0.062	0.030	-0.248	0.100
DL (2008) Specification	-0.158	0.087	-0.057	0.026	-0.249	0.099
Fixed Effects	-0.186	0.063	-0.110	0.046	-0.061	0.078
All Controls	0.516	0.400	0.146	0.127	0.611	0.523
Post-Double-Selection	0.060	0.214	-0.025	0.086	0.460	0.322
Poynomial Trend	0.203	0.296	0.141	0.089	0.199	0.309
Post-Double-Selection, Polynomial Trend	-0.264	0.179	0.090	0.046	-0.088	0.192

Note: The table displays the estimated coefficient on the abortion rate, "Effect," and its estimated standard error. Numbers in the first row of Panel A are taken from Donohue III and Levitt (2001) Table IV, columns (2), (4), and (6). Numbers from the first row of Panel B are taken from Donohue III and Levitt (2008) Table III, column (8). The remaining rows are estimated by OLS of the crime rate on the abortion rate and different sets of controls described in the text and use standard errors clustered at the state-level. In Panel A, the row labeled "All Controls" uses 311 control variables as discussed in the text that include the 68 controls from the original specification of Donohue III and Levitt (2001) Table IV along with 243 variables meant to allow for flexible, smooth trends. The row labeled "Polynomial Trend" in Panel A restricts the set of controls added to allow for flexible trends to include only polynomial terms and uses only 149 total regressors, the 68 from the original specification and 81 added variables. In Panel B, the row labeled "All Controls" uses 713 control variables as discussed in the text that include the 473 controls from the original specification of Donohue III and Levitt (2008) Table III along with 240 variables meant to allow for flexible, smooth trends. The row labeled "Polynomial Trend" in Panel B restricts the set of controls added to allow for flexible trends to include only polynomial terms and uses only 553 total regressors, the 473 from the original specification and 80 added variables. The rows "Post-Double-Selection" report results from regressing the crime rates on the variables from the original Donohue III and Levitt (2001) and Donohue III and Levitt (2008) along with additional variables selected using the technique developed in this paper from among the set of variables considered in the corresponding "All Controls" row. The rows "Post-Double-Selection, Polynomial Trend" report results from regressing the crime rates on the variables from the original Donohue III and Levitt (2001) and Donohue III and Levitt (2008) along with additional variables selected using the technique developed in this paper from among the set of variables considered in the corresponding "Polynomial Trend" row. Further details are provided in the text.

In the abortion example, we use the post-double-Lasso estimator for each of our dependent variables. For violent crime, a total of 15 variables are selected: eight in the abortion equation[3] and seven in the crime equation.[4] For property crime, 16 variables are selected: ten in the abortion equation[5] and seven in the crime equation[6] with one occurring in both. For murder, ten variables

[3]The selected variables are average abortion times t, average abortion times $\cos(\pi\frac{t}{T})$, initial crime times t^2, initial crime times $\cos(2\pi\frac{t}{T})$, average income times t^3, average income times $\sin(\pi\frac{t}{T})$, average income times $\cos(2\pi\frac{t}{T})$, and initial poverty times $\cos(2\pi\frac{t}{T})$.

[4]The selected variables are average abortion times t^3, initial abortion times t^3, initial abortion times $\sin(\pi\frac{t}{T})$, initial poverty times $\sin(2\pi\frac{t}{T})$, initial poverty times $\cos(\pi\frac{t}{T})$, police$_{it}$ times t^3, and beer$_{it}$ times $\sin(3\pi\frac{t}{T})$.

[5]The selected variables are average abortion times $\cos(\pi\frac{t}{T})$, initial abortion times $\sin(3\pi\frac{t}{T})$, initial crime times $\cos(\pi\frac{t}{T})$, average income times t, average income times $\cos(\pi\frac{t}{T})$, initial poverty times $\cos(2\pi\frac{t}{T})$, initial beer times $\cos(2\pi\frac{t}{T})$, prison$_{it}$ times $\cos(\pi\frac{t}{T})$, income$_{it}$ times $\cos(\pi\frac{t}{T})$, and AFDC$_{it}$ times $\cos(2\pi\frac{t}{T})$.

[6]The selected variables are average abortion times t^3, initial crime times $\sin(2\pi\frac{t}{T})$, initial crime times $\cos(\pi\frac{t}{T})$, average police times $\cos(2\pi\frac{t}{T})$, average AFDC times t, initial AFDC times t, and initial AFDC times t^2.

are selected: eight in the abortion equation[7] and two in the crime equation.[8] It is interesting in looking at the selected variables that in all cases initial or average levels of abortion interacted with nonlinear trend terms and initial levels of crime interacted with nonlinear trend terms are selected. This selection illustrates the potential importance of allowing for nonlinear trends and also the potential that there may be omitted factors that are related to both abortion and crime.

Estimates of the causal effect of abortion on crime obtained by searching for confounding factors among our set of 243 potential controls are given in the fifth row of Panel A of Table L. Each of these estimates is obtained from the least squares regression of the crime rate on the abortion rate, a full set of state dummies, a full set of time dummies, the initial eight controls that vary across states and time from Donohue III and Levitt (2001) and the 15, 16, and ten controls selected by the post-double-Lasso procedure for violent crime, property crime, and murder respectively. The estimates for the effect of abortion on violent crime and the effect of abortion on murder are quite imprecise, producing 95% confidence intervals that encompass large positive and negative values. The estimated effect for property crime is roughly in line with the previous estimates though it is no longer significant and has a 95% confidence interval that includes negative as well as modest positive effects. For a quick benchmark relative to the simulation examples, we note that the R^2 obtained by regressing the crime rate on the selected variables are .2522, .3533, and .0554 for violent crime, property crime, and the murder rate respectively and that the R^2's from regressing the abortion rate on the selected variables are .9906, .9039, and .9863 for violent crime, property crime, and the murder rate respectively. These values correspond to regions of the R^2 space considered in the simulation where the post-double-selection procedure performed quite well, while the standard post-single-selection procedures performed quite poorly.

While the inclusion of trigonometric terms in our approximations allows for capturing some types of cyclicality, some researchers may feel more comfortable restricting attention to simpler trend specifications. To allow for this, we also present results in which the trigonometric functions are dropped from f_t, so that

$$f_t = (t,\ t^2,\ t^3).$$

That is, we approximate the functions as $g(z_{it}, t) \approx \sum_{r=1}^{27} \sum_{s=1}^{3} \beta_{g,r,s} z_{it,r} f_{t,s} = h_{it}' \beta_g$ and $m(z_{it}, t) \approx \sum_{r=1}^{27} \sum_{s=1}^{3} \beta_{m,r,s} z_{it,r} f_{t,s} = h_{it}' \beta_m$ which allows only cubic polynomial trends interacted with state-level characteristics. In this case, only 81 terms are considered in addition to the 68 controls from the original specification. Results using all 149 controls are given in the row "Polynomial Trends" in Table L Panel A, and results based on Lasso selection among the 81 added controls are given in the row "Post-Double-Selection, Polynomial Trends." Looking at these results we see that we

[7]The selected variables are average abortion times t^2, average abortion times $\cos(\pi \frac{t}{T})$, initial crime times t^3, initial crime times $\cos(2\pi \frac{t}{T})$, average income times t^3, average income times $\sin(\pi \frac{t}{T})$, average income times $\cos(2\pi \frac{t}{T})$, and average income times $\cos(3\pi \frac{t}{T})$.

[8]The variables selected are average abortion times $\sin(\pi \frac{t}{T})$ and initial abortion times $\sin(\pi \frac{t}{T})$.

would draw the same qualitative conclusion using this restricted specification as we would when allowing for trigonometric terms as well. Specifically, the estimated abortion effects become quite imprecise after allowing only for the polynomial terms in time.[9]

A similar conclusion was reached by Foote and Goetz (2008) who, without doing formal variable selection, found that inclusion of a linear trend interacted with the average crime rate from a period before the abortion rate should have been able to have an effect on the crime rate substantially attenuated the estimated effects from Donohue III and Levitt (2001) and also rendered them imprecise. It is interesting that we reach a similar conclusion through the use of formal variable selection procedures motivated by the desire to allow flexible, yet parsimonious trends in an effort to make the exogeneity assumption conditional on controls more plausible.

In a response to Foote and Goetz (2008), Donohue III and Levitt (2008) note that one problem with allowing flexible trends is that the short time series renders estimates of the treatment effect imprecise once flexible trends are allowed. Specifically, estimated treatment effects are imprecise in their preferred specification

$$y_{it} = \alpha a_{it} + \delta_i + \gamma_{d,t} + \kappa_i t + \varepsilon_{it} \tag{5.20}$$

where δ_i is a state-specific effect, κ_i is a state-specific coefficient on a linear trend, and $\gamma_{d,t}$ is Census division $\times$ time effect. To address this issue, Donohue III and Levitt (2008) extend the sample period to 1960-2003 to allow more precise estimates of the trends and thus more reliable estimates of the treatment effect. They find that the results in this longer sample with the full set of division times time interactions and state-specific trends are similar to the initial results in the shorter panel. Results from this analysis in Donohue III and Levitt (2008) are provided in the first row of Panel B of Table L. In the second row of Table L, Panel B, we report results from our estimates of the abortion effect using data from 1960-2003 using exactly the same methodology as Donohue III and Levitt (2008), and we report results from simple OLS regression of (5.20) in the third row.[10]

While (5.20) is certainly more general than the baseline model in Donohue III and Levitt (2001), state-specific linear trends are still quite restrictive, especially over a time period of 40 years. Specifically, it is a strong assumption that unobserved factors that are correlated to both state level abortion and crime rates exhibited constant growth over such a long time period. To allow

[9]In addition to the 68 original variables, the double-selection procedure selects ten total additional variables for the violent crime regression, eight additional variables for the property crime regression, and five additional variables for the murder regression. In each case, the mean of the abortion rate times t is selected and this variable accounts for most of the explanatory power among the selected additional regressors.

[10]Our results differ due to the exclusion of Alaska, Hawaii, and Washington, D.C. We also completed the data on abortion before 1985 by filling in 0 for all abortion rates before 1985.

for smooth, but flexible trends, we once again consider variable selection in a more general model

$$y_{it} = \alpha a_{it} + \delta_{y,i} + \gamma_{y,d,t} + \kappa_{y,i}t + g(z_{it}, t) + \zeta_{it} \tag{5.21}$$

$$a_{it} = \delta_{a,i} + \gamma_{a,d,t} + \kappa_{d,i}t + m(z_{it}, t) + v_{it} \tag{5.22}$$

where $g(z,t)$ and $m(z,t)$ are smooth functions of observed variables z_{it} which includes time-invariant characteristics of $\{y_{it}, a_{it}, w_{it}\}_{t=1}^{T}$ such as initial conditions or state-level averages and time. For this longer time period, we approximate g and m by setting

$$z_{it} = (a_{i1985},\ \frac{1}{44}\sum_{t=1960}^{2003} a_{it},\ y_{i1960},\ y_{i1961},\ w'_{i1985},\ \frac{1}{13}\sum_{t=1985}^{1997} w'_{it})',$$

$$f_t = (t^2,\ t^3,\ t^4,\ t^5,\ \sin(\pi\frac{t}{T}),\ \sin(2\pi\frac{t}{T}),\ \sin(3\pi\frac{t}{T}),\ \sin(4\pi\frac{t}{T}),$$
$$\cos(\pi\frac{t}{T}),\ \cos(2\pi\frac{t}{T}),\ \cos(3\pi\frac{t}{T}),\ \cos(4\pi\frac{t}{T}))',$$

and then supposing

$$g(z_{it}, t) \approx \sum_{r=1}^{20}\sum_{s=1}^{12} \beta_{g,r,s} z_{it,r} f_{t,s} = h'_{it}\beta_g \quad \text{and}$$
$$m(z_{it}, t) \approx \sum_{r=1}^{20}\sum_{s=1}^{12} \beta_{m,r,s} z_{it,r} f_{t,s} = h'_{it}\beta_m,$$

where h_{it} is the vector containing all the interactions, and β_g and β_m are the vectors of coefficients for each equation. Thus, we add an additional 240 control variables to (5.20).[11]

Estimates of the abortion effect using the full set of 713 controls consisting of the 473 controls in (5.20) augmented with the 240 additional controls for smooth nonlinear trends are given in the fourth row of Table L Panel B. As expected, the estimated abortion effects are extremely imprecise given this large set of controls.

To pare down the number of controls, we employ the Double-Selection procedure developed in this paper to search for a smaller set of relevant controls among the 240 potential additions. Based on this exercise, we select a total of 31 additional variables for the violence equation, 30 for the abortion equation, and 27 for the murder equation. R^2's from the regression of crime rates on the controls are .2806, .3451, and .0422 for violent crime, property crime, and the murder rate respectively; and the R^2's from regressing the abortion rate on the selected variables are .9618, .9461, and .9775 for violent crime, property crime, and the murder rate respectively. Estimates of the treatment effect controlling for the variables in Donohue III and Levitt (2008) and those selected by Double-Selection are given in the "Post-Double-Selection" row of Table L, Panel B.

[11] To allow for all the effects in (5.20) to enter each equation without shrinkage, we use our methods based on $\tilde{y}_{it}$, $\tilde{a}_{it}$ and $\tilde{h}_{it}$ where $\tilde{y}_{it}$ is the residual from the regression of y_{it} on a full set of state dummies, a full set of Census division cross time dummies, and a full set of state-specific trends and $\tilde{a}_{it}$ and $\tilde{h}_{it}$ are defined similarly.

As in the original data, we find that estimates of the abortion effect are relatively imprecise once parsimonious nonlinear trends are allowed for.

As in the previous specification, we report results using only interactions with the polynomial trend terms, i.e.

$$f_t = (\ t^2,\ t^3,\ t^4,\ t^5)',$$

in the final two rows of Panel B of Table L.[12] Using only the interactions with the polynomial terms adds 80 potential regressors to the 473 included in the original Donohue III and Levitt (2008) specification. Results using the full set of 553 regressors are reported in the row "Polynomial Trends" in Table L Panel B and show that once again using this broad set of regressors results in imprecise estimates of the regression coefficients. The lack of precision in the estimated abortion effect is qualitatively unchanged after using the double-selection procedure to select controls from among this restricted set, again illustrating that the baseline result is not driven by the inclusion of trigonometric terms in the set of approximating functions.[13]

We believe that the example in this section illustrates how one may use modern variable selection techniques to complement causal analysis in economics. In the abortion example, we are able to search among a large set of controls and transformations of variables when trying to estimate the effect of abortion on crime. Considering a large set of controls makes the underlying assumption of exogeneity of the abortion rate conditional on observables more plausible, while the methods we develop allow us to produce an end-model which is of manageable dimension. In this example, we see that inference about the treatment effects using the variable selection method differs substantively from inference drawn using the original set of controls. This statement is true whether one considers the data and model from Donohue III and Levitt (2001) or Donohue III and Levitt (2008). This difference is driven by the variable selection method's selecting different variables than are usually considered. Thus, it appears that the usual interpretation of there being a substantive causal effect of abortion on crime hinges on strong prior beliefs about the types of trends that may appear in the structural equation. In particular, inclusion of a modest number of smooth nonlinear trends interacted with time-invariant state-level characteristics substantively increases the variance of the estimated treatment effects.

[12]The results are qualitatively similar if one only allows up to a cubic term in the trends, i.e. if one considers $f(t) = (t^2,\ t^3)$.

[13]In addition to the 473 original variables, the double-selection procedure selects 12 total additional variables for the violent crime regression, 11 additional variables for the property crime regression, and 11 additional variables for the murder regression.

6. Additional Simulation Results

In this section, we present additional simulation results. All of the simulation results are based on the structural model

$$y_i = d_i'\alpha_0 + x_i'(c_y\beta_0) + \sigma_y(d_i, x_i)\zeta_i, \quad \zeta_i \sim N(0,1) \tag{6.23}$$

where $p = \dim(x_i) = 200$, the covariates $x \sim N(0, \Sigma)$ with $\Sigma_{kj} = (0.5)^{|j-k|}$, $\alpha_0 = .5$, and the sample size n is set to 100. In each design, we generate

$$d_i^* = x_i'(c_d\beta_1) + \sigma_d(x_i)v_i, \quad v_i \sim N(0,1) \tag{6.24}$$

with $\mathrm{E}[\zeta_i v_i] = 0$. Inference results for all designs are based on conventional t-tests with standard errors calculated using the heteroscedasticity consistent jackknife variance estimator discussed in MacKinnon and White (1985). We set λ according to the algorithm outlined in Appendix A with $1-\gamma = .95$. We draw new x's, ζ's and v's at every replication and draw new β_0's and β_1's at every replication in the random coefficient designs.

In the first thirteen designs, $\beta_1 = \beta_0$. We set the constants c_y and c_d to generate desired population values for the reduced form R^2's, i.e. the R^2's for equations (6.23) and (6.24). Let R_y^2 be the desired R^2 for the regression of y on x and R_d^2 be the desired R^2 from the regression of d on x. For each equation, we choose c_y and c_d to generate $R^2 = 0, .2, .4, .6$, and $.8$. In the heteroscedastic and binary designs discussed below, we choose c_y and c_d based on R^2 as if (6.23) held with $d_i = d_i^*$ and v_i and ζ_i were homoscedastic with variance equal to the average variance and label the results by R^2 as in the other cases. In the homoscedastic cases, we set $\sigma_y = \sigma_d = 1$; and in the heteroscedastic cases, the average of $\sigma_d(x_i)$ and the average of $\sigma_y(d_i, x_i)$ are both one. We set

$$c_d = \sqrt{\frac{R_d^2}{(1-R_d^2)\beta_0'\Sigma\beta_0}}$$

$$c_y = \frac{-(1-R_y^2)\alpha_0 c_d \beta_0'\Sigma\beta_0 + \sqrt{(1-R_y^2)R_y^2\beta_0'\Sigma\beta_0(\alpha_0^2+1)}}{(1-R_y^2)\beta_0'\Sigma\beta_0}$$

- Design 1. $d_i = d_i^*$, $\beta_0 = (1, 1/2, 1/3, 1/4, 1/5, 0, 0, 0, 0, 0, 1, 1/2, 1/3, 1/4, 1/5, 0, ..., 0)'$, $\sigma_y = \sigma_d = 1$.
- Design 2. $d_i = d_i^*$, $\beta_0 = (1, 1/4, 1/9, 1/16, 1/25, 0, 0, 0, 0, 0, 1, 1/4, 1/9, 1/16, 1/25, 0, ..., 0)'$, $\sigma_y = \sigma_d = 1$.
- Design 22. $d_i = d_i^*$, $\beta_{0,j} = (1/j)^2$, $\sigma_y = \sigma_d = 1$.
- Design 3. $d_i = d_i^*$, $\beta_0 = (1, 1/2, 1/3, 1/4, 1/5, 0, 0, 0, 0, 0, 1, 1/2, 1/3, 1/4, 1/5, 0, ..., 0)'$, $\sigma_d = \sqrt{\frac{(1+x_i'\beta_0)^2}{\frac{1}{n}\sum_{i=1}^n(1+x_i'\beta_0)^2}}$, $\sigma_y = \sqrt{\frac{(1+\alpha_0 d_i+x_i'\beta_0)^2}{\frac{1}{n}\sum_{i=1}^n(1+\alpha_0 d_i+x_i'\beta_0)^2}}$.

- Design 4. $d_i = d_i^*$, $\beta_0 = (1, 1/4, 1/9, 1/16, 1/25, 0, 0, 0, 0, 0, 1, 1/4, 1/9, 1/16, 1/25, 0, ..., 0)'$, $\sigma_d = \sqrt{\frac{(1+x_i'\beta_0)^2}{\frac{1}{n}\sum_{i=1}^n(1+x_i'\beta_0)^2}}$, $\sigma_y = \sqrt{\frac{(1+\alpha_0 d_i+x_i'\beta_0)^2}{\frac{1}{n}\sum_{i=1}^n(1+\alpha_0 d_i+x_i'\beta_0)^2}}$.
- Design 44. $d_i = d_i^*$, $\beta_{0,j} = (1/j)^2$, $\sigma_d = \sqrt{\frac{(1+x_i'\beta_0)^2}{\frac{1}{n}\sum_{i=1}^n(1+x_i'\beta_0)^2}}$, $\sigma_y = \sqrt{\frac{(1+\alpha_0 d_i+x_i'\beta_0)^2}{\frac{1}{n}\sum_{i=1}^n(1+\alpha_0 d_i+x_i'\beta_0)^2}}$.
- Design 5. $d_i = \mathbf{1}\{d_i^* > 0\}$, $\beta_0 = (1, 1/2, 1/3, 1/4, 1/5, 0, 0, 0, 0, 0, 1, 1/2, 1/3, 1/4, 1/5, 0, ..., 0)'$, $\sigma_y = \sigma_d = 1$.
- Design 6. $d_i = d_i^*$, $\beta_{0,j} \sim N(0, 1)$, $\sigma_y = \sigma_d = 1$.
- Design 7. $d_i = d_i^*$, $\widetilde{\beta}_0 = (1, 1/2, 1/3, 1/4, 1/5, 0, 0, 0, 0, 0, 1, 1/2, 1/3, 1/4, 1/5, 0, ..., 0)'$, $\beta_{0,j} \sim N(0, \widetilde{\beta}_{0,j}^2)$, $\sigma_y = \sigma_d = 1$.
- Design 72. $d_i = d_i^*$, $\widetilde{\beta}_0 = (1, 1/4, 1/9, 1/16, 1/25, 0, 0, 0, 0, 0, 1, 1/4, 1/9, 1/16, 1/25, 0, ..., 0)'$, $\beta_{0,j} \sim N(0, \widetilde{\beta}_{0,j}^2)$, $\sigma_y = \sigma_d = 1$.
- Design 722. $d_i = d_i^*$, $\widetilde{\beta}_{0,j} = (1/j)^2$, $\beta_{0,j} \sim N(0, \widetilde{\beta}_{0,j}^2)$, $\sigma_y = \sigma_d = 1$.
- Design 8. $d_i = d_i^*$, $\widetilde{\beta}_{0,j} = u_j z_{1,j} + (1-u_j) z_{2,j}$, $u_j \sim$ Bernoulli(.05), $z_{1,j} \sim N(0, 25)$, $z_{2,j} \sim N(0, .0025)$, $\sigma_y = \sigma_d = 1$
- Design 1001. $d_i = d_i^*$, $\beta_{0,j} = \mathbf{1}\{j \in \{2, 4, 6, ..., 38, 40\}\}$, $\sigma_y = \sigma_d = 1$.

In the last thirteen designs, we set the constants c_y and c_d according to

$$c_d = \sqrt{\frac{R_d^2}{(1-R_d^2)\beta_1'\Sigma\beta_1}}$$
$$c_y = \sqrt{\frac{R_d^2}{(1-R_d^2)\beta_0'\Sigma\beta_0}}$$

for $R_d^2 = 0, .2, .4, .6,$ and .8 and $R_y^2 = 0, .2, .4, .6,$ and .8.

- Design 1a. $d_i = d_i^*$, $\beta_0 = (1, 1/2, 1/3, 1/4, 1/5, 0, 0, 0, 0, 0, 1, 1/2, 1/3, 1/4, 1/5, 0, ..., 0)'$, $\beta_1 = (1, 1/2, 1/3, 1/4, 1/5, 1/6, 1/7, 1/8, 1/9, 1/10, 0, ..., 0)'$, $\sigma_y = \sigma_d = 1$.
- Design 2a. $d_i = d_i^*$, $\beta_0 = (1, 1/4, 1/9, 1/16, 1/25, 0, 0, 0, 0, 0, 1, 1/4, 1/9, 1/16, 1/25, 0, ..., 0)'$, $\beta_1 = (1, 1/4, 1/9, 1/16, 1/25, 1/36, 1/49, 1/64, 1/81, 1/100, 0, ..., 0)'$, $\sigma_y = \sigma_d = 1$.
- Design 22a. $d_i = d_i^*$, $\beta_{0,j} = (1/j)^2$, $\beta_{1,j} = (1/j)^2$, $\sigma_y = \sigma_d = 1$.
- Design 3a. $d_i = d_i^*$, $\beta_0 = (1, 1/2, 1/3, 1/4, 1/5, 0, 0, 0, 0, 0, 1, 1/2, 1/3, 1/4, 1/5, 0, ..., 0)'$, $\beta_1 = (1, 1/2, 1/3, 1/4, 1/5, 1/6, 1/7, 1/8, 1/9, 1/10, 0, ..., 0)'$, $\sigma_d = \sqrt{\frac{(1+x_i'\beta_1)^2}{\frac{1}{n}\sum_{i=1}^n(1+x_i'\beta_1)^2}}$, $\sigma_y = \sqrt{\frac{(1+\alpha_0 d_i+x_i'\beta_0)^2}{\frac{1}{n}\sum_{i=1}^n(1+\alpha_0 d_i+x_i'\beta_0)^2}}$.
- Design 4a. $d_i = d_i^*$, $\beta_0 = (1, 1/4, 1/9, 1/16, 1/25, 0, 0, 0, 0, 0, 1, 1/4, 1/9, 1/16, 1/25, 0, ..., 0)'$, $\beta_1 = (1, 1/4, 1/9, 1/16, 1/25, 1/36, 1/49, 1/64, 1/81, 1/100, 0, ..., 0)'$, $\sigma_d = \sqrt{\frac{(1+x_i'\beta_1)^2}{\frac{1}{n}\sum_{i=1}^n(1+x_i'\beta_1)^2}}$, $\sigma_y = \sqrt{\frac{(1+\alpha_0 d_i+x_i'\beta_0)^2}{\frac{1}{n}\sum_{i=1}^n(1+\alpha_0 d_i+x_i'\beta_0)^2}}$.
- Design 44a. $d_i = d_i^*$, $\beta_{0,j} = (1/j)^2$, $\beta_{1,j} = (1/j)^2$, $\sigma_d = \sqrt{\frac{(1+x_i'\beta_1)^2}{\frac{1}{n}\sum_{i=1}^n(1+x_i'\beta_1)^2}}$, $\sigma_y = \sqrt{\frac{(1+\alpha_0 d_i+x_i'\beta_0)^2}{\frac{1}{n}\sum_{i=1}^n(1+\alpha_0 d_i+x_i'\beta_0)^2}}$.

- Design 5a. $d_i = \mathbf{1}\{d_i^* > 0\}$, $\beta_0 = (1, 1/2, 1/3, 1/4, 1/5, 0, 0, 0, 0, 0, 1, 1/2, 1/3, 1/4, 1/5, 0, ..., 0)'$, $\beta_1 = (1, 1/2, 1/3, 1/4, 1/5, 1/6, 1/7, 1/8, 1/9, 1/10, 0, ..., 0)'$, $\sigma_y = \sigma_d = 1$.
- Design 6a. $d_i = d_i^*$, $\beta_{0,j} \sim N(0,1)$, $\beta_{1,j} \sim N(0,1)$, $\mathrm{E}[\beta_{0,j}\beta_{1,j}] = .8$, $\sigma_y = \sigma_d = 1$.
- Design 7a. $d_i = d_i^*$, $\widetilde{\beta}_0 = (1, 1/2, 1/3, 1/4, 1/5, 0, 0, 0, 0, 0, 1, 1/2, 1/3, 1/4, 1/5, 0, ..., 0)'$, $\widetilde{\beta}_1 = (1, 1/2, 1/3, 1/4, 1/5, 1/6, 1/7, 1/8, 1/9, 1/10, 0, ..., 0)'$, $\beta_{0,j} = \widetilde{\beta}_{0,j} z_{0,j}$, $\beta_{1,j} = \widetilde{\beta}_{1,j} z_{1,j}$, $z_{0,j} \sim N(0,1)$, $z_{1,j} \sim N(0,1)$, $\mathrm{E}[z_{0,j} z_{1,j}] = .8$, $\sigma_y = \sigma_d = 1$.
- Design 72a. $d_i = d_i^*$, $\widetilde{\beta}_0 = (1, 1/4, 1/9, 1/16, 1/25, 0, 0, 0, 0, 0, 1, 1/4, 1/9, 1/16, 1/25, 0, ..., 0)'$, $\widetilde{\beta}_1 = (1, 1/4, 1/9, 1/16, 1/25, 1/36, 1/49, 1/64, 1/81, 1/100, 0, ..., 0)'$, $\beta_{0,j} = \widetilde{\beta}_{0,j} z_{0,j}$, $\beta_{1,j} = \widetilde{\beta}_{1,j} z_{1,j}$, $z_{0,j} \sim N(0,1)$, $z_{1,j} \sim N(0,1)$, $\mathrm{E}[z_{0,j} z_{1,j}] = .8$, $\sigma_y = \sigma_d = 1$.
- Design 722a. $d_i = d_i^*$, $\widetilde{\beta}_{0,j} = (1/j)^2$, $\widetilde{\beta}_{1,j} = (1/j)^2$, $\beta_{0,j} = \widetilde{\beta}_{0,j} z_{0,j}$, $\beta_{1,j} = \widetilde{\beta}_{1,j} z_{1,j}$, $z_{0,j} \sim N(0,1)$, $z_{1,j} \sim N(0,1)$, $\mathrm{E}[z_{0,j} z_{1,j}] = .8$, $\sigma_y = \sigma_d = 1$.
- Design 8a. $d_i = d_i^*$, $\widetilde{\beta}_{0,j} = 5u_j z_{11,j} + .05(1-u_j) z_{12,j}$, $\widetilde{\beta}_{1,j} = 5u_j z_{21,j} + .05(1-u_j) z_{22,j}$, $u_j \sim \text{Bernoulli}(.05)$, $z_{11,j} \sim N(0,1)$, $z_{12,j} \sim N(0,1)$, $z_{21,j} \sim N(0,1)$, $z_{22,j} \sim N(0,1)$, $\sigma_y = \sigma_d = 1$
- Design 1001a. $d_i = d_i^*$, $\beta_{0,j} = \mathbf{1}\{j \in \{2, 4, 6, ..., 38, 40\}\}$, $\beta_{1,j} = \mathbf{1}\{j \in \{1, 3, 5, ..., 37, 39\}\}$, $\sigma_y = \sigma_d = 1$.

Results are summarized in figures and tables below. In the tables, we report results for the four estimators considered in the main text (Oracle, Double-Selection Oracle, Post-Lasso, and Double-Selection). We also report results for regular Lasso (Lasso), the union of the Double-Selection interval with the Post-Lasso interval (Double-Selection Union ADS), using the union of the set of variables selected by Double-Selection and the set of variables selected by running Lasso of y on d and x without penalizing d (Double-Selection + I3), and the split-sample procedure discussed in the text (Split-Sample). For Double-Selection Union ADS, the point estimate is taken as the midpoint of the union of the intervals.

Appendix Table 1. Simulation Results for Selected R^2 Values

	First Stage R^2 = .2		First Stage R^2 = .2		First Stage R^2 = .8		First Stage R^2 = .8	
	Structure R^2 = 0		Structure R^2 = .8		Structure R^2 = 0		Structure R^2 = .4	
	RMSE	Coverage	RMSE	Coverage	RMSE	Coverage	RMSE	Coverage
	Design 1 - Linear Decay with Cut-Off							
Oracle	0.104	0.068	0.106	0.049	0.106	0.052	0.106	0.049
Double-Selection Oracle	0.103	0.051	0.106	0.049	0.106	0.049	0.106	0.049
Lasso	0.137	0.198	0.430	0.886	0.405	1.000	0.496	1.000
Post-Lasso	0.135	0.191	0.164	0.166	0.403	1.000	0.493	1.000
Double-Selection	0.125	0.122	0.114	0.061	0.127	0.118	0.114	0.078
Double-Selection Union ADS	0.123	0.119	0.117	0.057	0.126	0.115	0.117	0.077
Double-Selection + I3	0.123	0.121	0.117	0.069	0.126	0.116	0.117	0.087
Split-Sample	0.137	0.194	0.309	0.386	0.229	0.632	0.235	0.554
	Design 2 - Quadratic Decay with Cut-Off							
Oracle	0.101	0.059	0.105	0.045	0.105	0.064	0.107	0.053
Double-Selection Oracle	0.101	0.059	0.105	0.045	0.104	0.051	0.104	0.042
Lasso	0.137	0.193	0.348	0.780	0.405	1.000	0.496	1.000
Post-Lasso	0.136	0.197	0.121	0.096	0.404	1.000	0.493	1.000
Double-Selection	0.120	0.112	0.108	0.051	0.113	0.078	0.107	0.062
Double-Selection Union ADS	0.119	0.106	0.110	0.045	0.113	0.078	0.110	0.062
Double-Selection + I3	0.119	0.109	0.110	0.058	0.113	0.080	0.110	0.069
Split-Sample	0.135	0.191	0.206	0.195	0.154	0.270	0.153	0.230
	Design 22 - Quadratic Decay							
Oracle	0.100	0.051	0.103	0.051	0.106	0.073	0.103	0.050
Double-Selection Oracle	0.101	0.051	0.103	0.051	0.102	0.050	0.102	0.052
Lasso	0.138	0.211	0.263	0.564	0.405	1.000	0.496	1.000
Post-Lasso	0.137	0.205	0.110	0.064	0.402	0.987	0.489	0.974
Double-Selection	0.107	0.063	0.107	0.058	0.109	0.074	0.104	0.062
Double-Selection Union ADS	0.108	0.063	0.108	0.055	0.108	0.072	0.106	0.061
Double-Selection + I3	0.108	0.068	0.107	0.060	0.109	0.074	0.106	0.068
Split-Sample	0.121	0.138	0.124	0.087	0.119	0.116	0.123	0.118
	Design 3 - Linear Decay with Cut-Off and Heteroscedasticity							
Oracle	0.143	0.070	0.150	0.075	0.145	0.068	0.150	0.075
Double-Selection Oracle	0.144	0.074	0.150	0.075	0.150	0.075	0.150	0.075
Lasso	0.168	0.142	0.536	0.746	0.411	0.990	0.500	0.999
Post-Lasso	0.167	0.140	0.257	0.236	0.410	0.990	0.500	0.999
Double-Selection	0.156	0.108	0.164	0.089	0.159	0.129	0.158	0.102
Double-Selection Union ADS	0.156	0.107	0.170	0.080	0.158	0.125	0.158	0.101
Double-Selection + I3	0.156	0.108	0.166	0.107	0.158	0.125	0.158	0.109
Split-Sample	0.175	0.198	0.398	0.411	0.260	0.609	0.274	0.567
	Design 4 - Quadratic Decay with Cut-Off and Heteroscedasticity							
Oracle	0.142	0.062	0.147	0.070	0.141	0.066	0.146	0.080
Double-Selection Oracle	0.142	0.062	0.147	0.070	0.145	0.067	0.146	0.070
Lasso	0.165	0.139	0.447	0.642	0.410	0.995	0.501	1.000
Post-Lasso	0.166	0.138	0.173	0.113	0.410	0.994	0.500	1.000
Double-Selection	0.152	0.092	0.150	0.075	0.147	0.086	0.147	0.082
Double-Selection Union ADS	0.152	0.090	0.155	0.070	0.147	0.085	0.149	0.080
Double-Selection + I3	0.152	0.093	0.154	0.078	0.147	0.085	0.149	0.088
Split-Sample	0.172	0.185	0.287	0.260	0.195	0.349	0.197	0.295

Appendix Table 1. Simulation Results for Selected R^2 Values

	First Stage R^2 = .2		First Stage R^2 = .2		First Stage R^2 = .8		First Stage R^2 = .8	
	Structure R^2 = 0		Structure R^2 = .8		Structure R^2 = 0		Structure R^2 = .4	
	RMSE	Coverage	RMSE	Coverage	RMSE	Coverage	RMSE	Coverage
	Design 44 - Quadratic Decay and Heteroscedasticity							
Oracle	0.163	0.080	0.166	0.084	0.158	0.080	0.164	0.088
Double-Selection Oracle	0.164	0.078	0.166	0.084	0.162	0.082	0.164	0.091
Lasso	0.173	0.131	0.382	0.460	0.410	0.996	0.503	0.999
Post-Lasso	0.175	0.139	0.178	0.097	0.409	0.994	0.501	0.993
Double-Selection	0.165	0.098	0.167	0.081	0.162	0.082	0.165	0.083
Double-Selection Union ADS	0.167	0.098	0.170	0.078	0.162	0.082	0.166	0.083
Double-Selection + I3	0.166	0.103	0.169	0.086	0.162	0.083	0.166	0.087
Split-Sample	0.177	0.170	0.205	0.160	0.177	0.168	0.183	0.172
	Design 5 - Binary Treatment							
Oracle	0.228	0.063	0.230	0.053	0.309	0.062	0.306	0.054
Double-Selection Oracle	0.225	0.055	0.230	0.053	0.306	0.054	0.306	0.054
Lasso	0.278	0.144	0.717	0.724	1.435	0.999	1.712	1.000
Post-Lasso	0.278	0.142	0.296	0.117	1.300	0.966	1.447	0.924
Double-Selection	0.259	0.101	0.239	0.055	0.364	0.109	0.339	0.080
Double-Selection Union ADS	0.260	0.100	0.246	0.053	0.364	0.109	0.349	0.079
Double-Selection + I3	0.260	0.105	0.246	0.061	0.374	0.123	0.349	0.094
Split-Sample	0.271	0.124	0.562	0.247	0.795	0.672	0.823	0.596
	Design 6 - Gaussian Random Coefficients							
Oracle	0.134	0.180	0.452	0.091	0.306	0.916	0.313	0.723
Double-Selection Oracle	0.134	0.180	0.452	0.091	0.481	0.125	0.476	0.136
Lasso	0.133	0.185	0.805	0.987	0.399	1.000	0.497	1.000
Post-Lasso	0.134	0.182	0.772	0.980	0.398	1.000	0.496	1.000
Double-Selection	0.139	0.191	0.646	0.899	0.389	1.000	0.464	1.000
Double-Selection Union ADS	0.137	0.188	0.659	0.899	0.387	1.000	0.465	1.000
Double-Selection + I3	0.137	0.189	0.653	0.913	0.387	1.000	0.464	1.000
Split-Sample	0.137	0.206	0.795	0.983	0.397	1.000	0.496	1.000
	Design 7 - Gaussian Random Coefficients, Linear Decay in Std. Dev. with Cut-Off							
Oracle	0.101	0.056	0.104	0.047	0.105	0.069	0.105	0.055
Double-Selection Oracle	0.101	0.056	0.104	0.047	0.103	0.050	0.103	0.050
Lasso	0.134	0.192	0.337	0.749	0.403	1.000	0.500	1.000
Post-Lasso	0.135	0.188	0.119	0.078	0.401	1.000	0.496	1.000
Double-Selection	0.119	0.106	0.109	0.053	0.112	0.083	0.108	0.063
Double-Selection Union ADS	0.117	0.102	0.112	0.050	0.112	0.082	0.112	0.062
Double-Selection + I3	0.118	0.103	0.112	0.061	0.112	0.082	0.111	0.074
Split-Sample	0.131	0.170	0.197	0.184	0.149	0.262	0.152	0.235
	Design 72 - Gaussian Random Coefficients, Quadratic Decay in Std. Dev. with Cut-Off							
Oracle	0.101	0.054	0.102	0.053	0.102	0.051	0.102	0.056
Double-Selection Oracle	0.101	0.054	0.102	0.053	0.102	0.051	0.102	0.056
Lasso	0.134	0.189	0.305	0.686	0.403	1.000	0.499	1.000
Post-Lasso	0.134	0.183	0.108	0.065	0.402	1.000	0.496	1.000
Double-Selection	0.117	0.095	0.105	0.060	0.105	0.066	0.103	0.057
Double-Selection Union ADS	0.116	0.092	0.107	0.057	0.105	0.065	0.106	0.056
Double-Selection + I3	0.116	0.094	0.107	0.065	0.104	0.066	0.106	0.062
Split-Sample	0.129	0.168	0.186	0.171	0.140	0.201	0.133	0.152

Appendix Table 1. Simulation Results for Selected R^2 Values

	First Stage R^2 = .2		First Stage R^2 = .2		First Stage R^2 = .8		First Stage R^2 = .8	
	Structure R^2 = 0		Structure R^2 = .8		Structure R^2 = 0		Structure R^2 = .4	
	RMSE	Coverage	RMSE	Coverage	RMSE	Coverage	RMSE	Coverage
	Design 722 - Gaussian Random Coefficients, Quadratic Decay in Std. Dev.							
Oracle	0.101	0.051	0.101	0.050	0.101	0.053	0.101	0.050
Double-Selection Oracle	0.101	0.051	0.101	0.050	0.101	0.052	0.101	0.050
Lasso	0.134	0.176	0.224	0.446	0.404	1.000	0.498	1.000
Post-Lasso	0.135	0.182	0.102	0.055	0.403	0.998	0.492	0.976
Double-Selection	0.106	0.070	0.103	0.053	0.103	0.062	0.102	0.057
Double-Selection Union ADS	0.106	0.070	0.105	0.053	0.103	0.062	0.104	0.053
Double-Selection + I3	0.106	0.073	0.104	0.059	0.103	0.063	0.104	0.060
Split-Sample	0.120	0.135	0.111	0.071	0.106	0.072	0.106	0.067
	Design 8 - Mixture of Normals							
Oracle	0.106	0.074	0.105	0.056	0.109	0.061	0.105	0.055
Double-Selection Oracle	0.103	0.053	0.105	0.056	0.106	0.055	0.105	0.051
Lasso	0.134	0.186	0.585	0.945	0.402	1.000	0.495	1.000
Post-Lasso	0.134	0.185	0.355	0.510	0.400	1.000	0.493	1.000
Double-Selection	0.129	0.138	0.216	0.214	0.231	0.576	0.228	0.500
Double-Selection Union ADS	0.128	0.135	0.226	0.213	0.230	0.574	0.230	0.499
Double-Selection + I3	0.128	0.139	0.223	0.252	0.230	0.577	0.230	0.507
Split-Sample	0.136	0.185	0.498	0.703	0.329	0.933	0.377	0.913
	Design 1001 - 20 Non-Overlapping Constant Coefficients							
Oracle	0.135	0.193	0.115	0.056	0.115	0.056	0.115	0.056
Double-Selection Oracle	0.135	0.193	0.115	0.056	0.115	0.056	0.115	0.056
Lasso	0.135	0.196	0.755	0.992	0.403	1.000	0.496	1.000
Post-Lasso	0.134	0.194	0.608	0.905	0.401	1.000	0.494	1.000
Double-Selection	0.139	0.185	0.329	0.399	0.327	0.910	0.330	0.861
Double-Selection Union ADS	0.136	0.177	0.354	0.399	0.325	0.910	0.333	0.861
Double-Selection + I3	0.137	0.179	0.349	0.492	0.325	0.911	0.332	0.864
Split-Sample	0.137	0.200	0.731	0.952	0.392	1.000	0.475	1.000

Appendix Table 2. Simulation Results for Selected R^2 Values

	First Stage R^2 = .4		First Stage R^2 = .4		First Stage R^2 = .8		First Stage R^2 = .8	
	Structure R^2 = .4		Structure R^2 = .8		Structure R^2 = .4		Structure R^2 = .4	
	RMSE	Coverage	RMSE	Coverage	RMSE	Coverage	RMSE	Coverage
			Design 1a - Linear Decay with Cut-Off					
Oracle	0.105	0.047	0.105	0.049	0.097	0.079	0.099	0.054
Double-Selection Oracle	0.104	0.047	0.105	0.049	0.110	0.051	0.110	0.049
Lasso	0.294	0.852	0.493	0.943	0.239	0.992	0.564	1.000
Post-Lasso	0.261	0.684	0.280	0.377	0.231	0.989	0.549	0.999
Double-Selection	0.110	0.058	0.112	0.060	0.112	0.061	0.111	0.062
Double-Selection Union ADS	0.112	0.049	0.115	0.052	0.112	0.050	0.111	0.055
Double-Selection + I3	0.112	0.065	0.114	0.069	0.109	0.066	0.111	0.069
Split-Sample	0.168	0.221	0.228	0.241	0.131	0.141	0.194	0.206
			Design 2a - Quadratic Decay with Cut-Off					
Oracle	0.103	0.045	0.105	0.045	0.101	0.053	0.105	0.043
Double-Selection Oracle	0.103	0.045	0.105	0.045	0.103	0.050	0.105	0.050
Lasso	0.289	0.812	0.408	0.878	0.238	0.997	0.570	1.000
Post-Lasso	0.258	0.648	0.190	0.145	0.235	0.999	0.567	0.999
Double-Selection	0.107	0.050	0.108	0.054	0.108	0.054	0.106	0.056
Double-Selection Union ADS	0.108	0.047	0.110	0.049	0.108	0.048	0.107	0.053
Double-Selection + I3	0.107	0.060	0.110	0.056	0.105	0.062	0.107	0.059
Split-Sample	0.130	0.102	0.160	0.112	0.125	0.076	0.155	0.122
			Design 22a - Quadratic Decay					
Oracle	0.103	0.053	0.103	0.051	0.104	0.063	0.104	0.047
Double-Selection Oracle	0.103	0.053	0.103	0.051	0.102	0.049	0.104	0.047
Lasso	0.344	0.845	0.404	0.901	0.330	1.000	0.770	1.000
Post-Lasso	0.201	0.255	0.124	0.092	0.330	0.999	0.668	0.835
Double-Selection	0.104	0.055	0.107	0.057	0.103	0.056	0.107	0.059
Double-Selection Union ADS	0.106	0.053	0.109	0.055	0.104	0.054	0.112	0.056
Double-Selection + I3	0.106	0.060	0.109	0.060	0.104	0.061	0.112	0.070
Split-Sample	0.117	0.096	0.132	0.103	0.112	0.089	0.137	0.144
			Design 3a - Linear Decay with Cut-Off and Heteroscedasticity					
Oracle	0.144	0.081	0.146	0.073	0.120	0.103	0.130	0.074
Double-Selection Oracle	0.144	0.074	0.146	0.073	0.153	0.080	0.153	0.079
Lasso	0.314	0.676	0.580	0.828	0.245	0.944	0.574	0.999
Post-Lasso	0.297	0.637	0.419	0.462	0.240	0.942	0.558	0.999
Double-Selection	0.153	0.081	0.157	0.084	0.149	0.085	0.150	0.087
Double-Selection Union ADS	0.157	0.076	0.162	0.071	0.151	0.073	0.153	0.073
Double-Selection + I3	0.155	0.089	0.159	0.093	0.146	0.086	0.150	0.096
Split-Sample	0.226	0.313	0.291	0.301	0.168	0.210	0.233	0.277
			Design 4a - Quadratic Decay with Cut-Off and Heteroscedasticity					
Oracle	0.144	0.065	0.147	0.063	0.138	0.067	0.145	0.062
Double-Selection Oracle	0.144	0.065	0.147	0.063	0.143	0.066	0.146	0.066
Lasso	0.309	0.671	0.527	0.736	0.242	0.946	0.576	1.000
Post-Lasso	0.298	0.634	0.380	0.295	0.241	0.945	0.572	1.000
Double-Selection	0.148	0.071	0.149	0.069	0.148	0.067	0.148	0.074
Double-Selection Union ADS	0.150	0.062	0.151	0.066	0.149	0.054	0.150	0.070
Double-Selection + I3	0.148	0.075	0.151	0.082	0.145	0.070	0.149	0.077
Split-Sample	0.193	0.216	0.215	0.186	0.167	0.144	0.201	0.163

Appendix Table 2. Simulation Results for Selected R^2 Values

	First Stage R^2 = .4		First Stage R^2 = .4		First Stage R^2 = .8		First Stage R^2 = .8	
	Structure R^2 = .4		Structure R^2 = .8		Structure R^2 = .4		Structure R^2 = .4	
	RMSE	Coverage	RMSE	Coverage	RMSE	Coverage	RMSE	Coverage
Design 44a - Quadratic Decay and Heteroscedasticity								
Oracle	0.166	0.097	0.166	0.091	0.161	0.099	0.166	0.093
Double-Selection Oracle	0.166	0.097	0.166	0.091	0.163	0.088	0.166	0.093
Lasso	0.408	0.707	0.563	0.691	0.336	0.982	0.803	0.997
Post-Lasso	0.358	0.541	0.220	0.140	0.336	0.982	0.785	0.955
Double-Selection	0.166	0.087	0.167	0.084	0.164	0.084	0.167	0.077
Double-Selection Union ADS	0.167	0.086	0.172	0.080	0.164	0.084	0.169	0.077
Double-Selection + I3	0.166	0.089	0.172	0.091	0.164	0.085	0.168	0.081
Split-Sample	0.215	0.234	0.206	0.163	0.176	0.157	0.193	0.185
Design 5a - Binary Treatment								
Oracle	0.243	0.058	0.242	0.047	0.288	0.055	0.288	0.045
Double-Selection Oracle	0.241	0.058	0.242	0.047	0.306	0.050	0.305	0.043
Lasso	0.597	0.637	0.848	0.789	0.857	0.933	1.709	0.995
Post-Lasso	0.478	0.415	0.390	0.170	0.797	0.840	1.308	0.695
Double-Selection	0.244	0.047	0.251	0.049	0.297	0.056	0.305	0.054
Double-Selection Union ADS	0.251	0.047	0.259	0.046	0.304	0.053	0.308	0.045
Double-Selection + I3	0.251	0.062	0.259	0.059	0.303	0.057	0.308	0.053
Split-Sample	0.385	0.187	0.490	0.185	0.407	0.162	0.539	0.184
Design 6a - Gaussian Random Coefficients								
Oracle	0.231	0.487	0.445	0.090	0.215	0.878	0.370	0.110
Double-Selection Oracle	0.205	0.286	0.539	0.107	0.534	0.138	0.518	0.000
Lasso	0.347	0.928	0.825	0.999	0.279	0.999	0.672	1.000
Post-Lasso	0.343	0.920	0.806	0.999	0.278	0.999	0.667	1.000
Double-Selection	0.301	0.761	0.672	0.953	0.254	0.968	0.603	1.000
Double-Selection Union ADS	0.304	0.761	0.687	0.952	0.256	0.968	0.611	1.000
Double-Selection + I3	0.302	0.775	0.682	0.958	0.255	0.969	0.607	1.000
Split-Sample	0.346	0.909	0.814	0.995	0.278	0.997	0.669	1.000
Design 7a - Gaussian Random Coefficients, Linear Decay in Std. Dev. with Cut-Off								
Oracle	0.102	0.054	0.104	0.053	0.099	0.056	0.103	0.052
Double-Selection Oracle	0.102	0.054	0.104	0.053	0.103	0.059	0.105	0.049
Lasso	0.318	0.768	0.356	0.790	0.303	0.999	0.708	1.000
Post-Lasso	0.214	0.282	0.113	0.065	0.303	0.998	0.672	0.830
Double-Selection	0.107	0.065	0.110	0.054	0.105	0.061	0.108	0.060
Double-Selection Union ADS	0.109	0.058	0.113	0.043	0.106	0.053	0.111	0.042
Double-Selection + I3	0.109	0.069	0.112	0.063	0.106	0.069	0.110	0.074
Split-Sample	0.117	0.076	0.136	0.079	0.110	0.073	0.132	0.085
Design 72a - Gaussian Random Coefficients, Quadratic Decay in Std. Dev. with Cut-Off								
Oracle	0.102	0.053	0.102	0.053	0.101	0.052	0.102	0.055
Double-Selection Oracle	0.102	0.053	0.102	0.053	0.102	0.052	0.102	0.053
Lasso	0.325	0.775	0.349	0.781	0.314	0.999	0.732	1.000
Post-Lasso	0.205	0.255	0.108	0.061	0.313	0.999	0.692	0.833
Double-Selection	0.106	0.064	0.109	0.069	0.106	0.065	0.109	0.064
Double-Selection Union ADS	0.108	0.059	0.110	0.059	0.107	0.060	0.109	0.046
Double-Selection + I3	0.108	0.071	0.109	0.071	0.107	0.068	0.108	0.063
Split-Sample	0.114	0.082	0.129	0.084	0.112	0.076	0.127	0.081

Appendix Table 2. Simulation Results for Selected R^2 Values

	First Stage R^2 = .4		First Stage R^2 = .4		First Stage R^2 = .8		First Stage R^2 = .8	
	Structure R^2 = .4		Structure R^2 = .8		Structure R^2 = .4		Structure R^2 = .4	
	RMSE	Coverage	RMSE	Coverage	RMSE	Coverage	RMSE	Coverage
	Design 722a - Gaussian Random Coefficients, Quadratic Decay in Std. Dev.							
Oracle	0.101	0.051	0.101	0.050	0.100	0.051	0.101	0.049
Double-Selection Oracle	0.101	0.051	0.101	0.050	0.101	0.050	0.101	0.050
Lasso	0.323	0.781	0.335	0.787	0.332	1.000	0.735	1.000
Post-Lasso	0.176	0.167	0.102	0.053	0.332	1.000	0.605	0.567
Double-Selection	0.103	0.054	0.103	0.053	0.102	0.055	0.102	0.056
Double-Selection Union ADS	0.105	0.052	0.104	0.052	0.103	0.055	0.105	0.045
Double-Selection + I3	0.105	0.062	0.104	0.056	0.103	0.060	0.105	0.059
Split-Sample	0.110	0.081	0.109	0.064	0.104	0.062	0.105	0.059
	Design 8a - Mixture of Normals							
Oracle	0.102	0.053	0.108	0.062	0.093	0.051	0.106	0.065
Double-Selection Oracle	0.105	0.049	0.108	0.059	0.104	0.056	0.109	0.058
Lasso	0.321	0.836	0.630	0.945	0.267	0.966	0.617	0.991
Post-Lasso	0.283	0.692	0.446	0.710	0.263	0.944	0.570	0.974
Double-Selection	0.150	0.165	0.207	0.217	0.144	0.244	0.231	0.337
Double-Selection Union ADS	0.155	0.158	0.222	0.210	0.147	0.232	0.248	0.308
Double-Selection + I3	0.153	0.181	0.218	0.272	0.145	0.255	0.241	0.387
Split-Sample	0.255	0.597	0.489	0.700	0.215	0.694	0.449	0.806
	Design 1001a - 20 Non-Overlapping Constant Coefficients							
Oracle	0.100	0.053	0.100	0.053	0.072	0.050	0.072	0.050
Double-Selection Oracle	0.131	0.050	0.131	0.050	0.131	0.050	0.131	0.050
Lasso	0.324	0.891	0.741	0.997	0.261	0.997	0.626	1.000
Post-Lasso	0.305	0.837	0.617	0.962	0.257	0.994	0.592	1.000
Double-Selection	0.188	0.319	0.281	0.328	0.150	0.343	0.262	0.466
Double-Selection Union ADS	0.191	0.319	0.298	0.324	0.152	0.341	0.269	0.463
Double-Selection + I3	0.190	0.337	0.294	0.379	0.151	0.348	0.265	0.485
Split-Sample	0.306	0.822	0.683	0.968	0.243	0.971	0.567	0.997

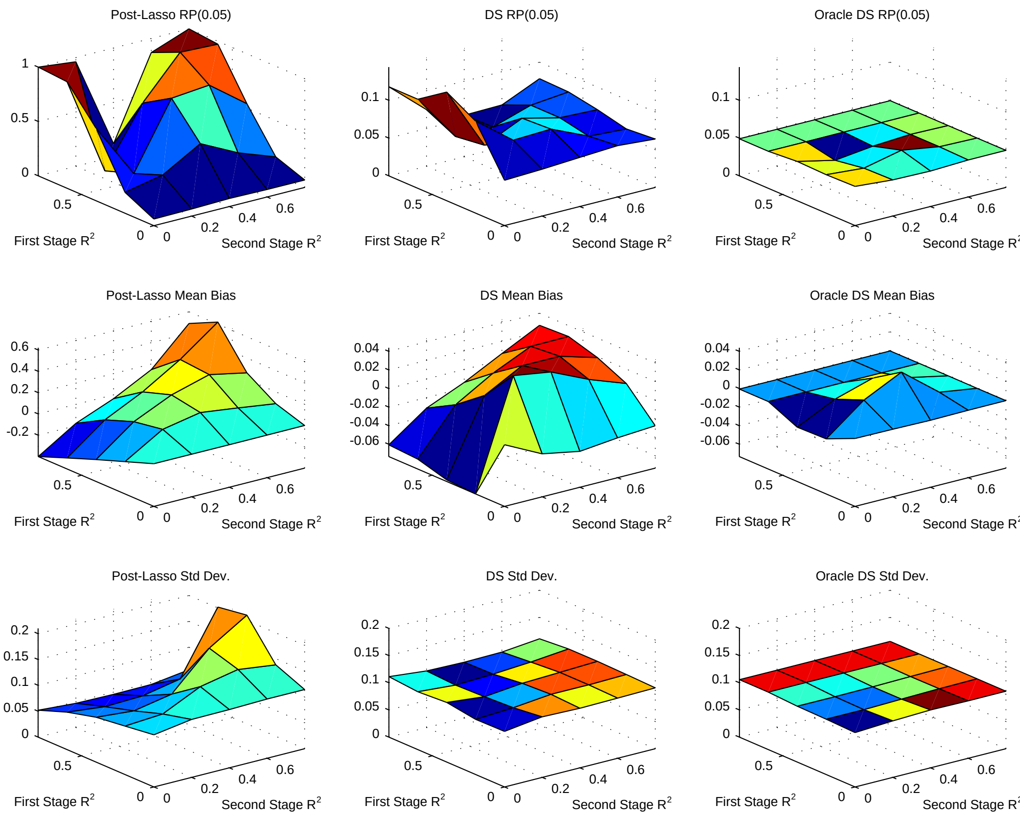

FIGURE 1. Design 1

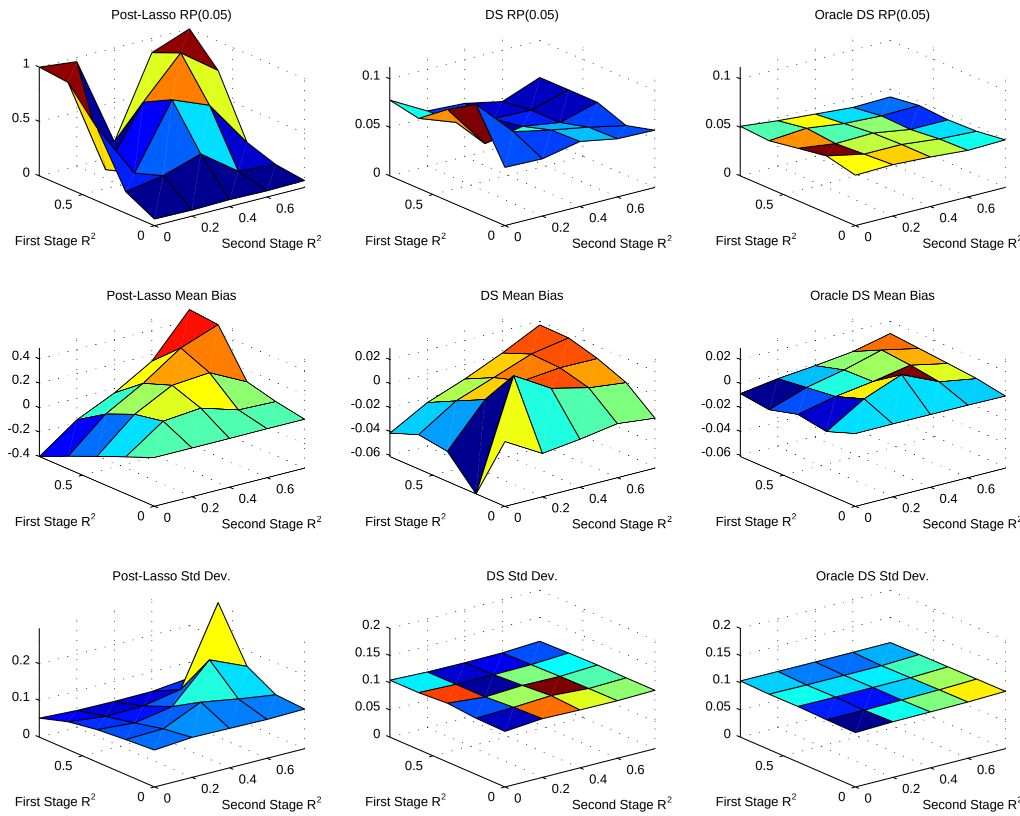

Figure 2. Design 2

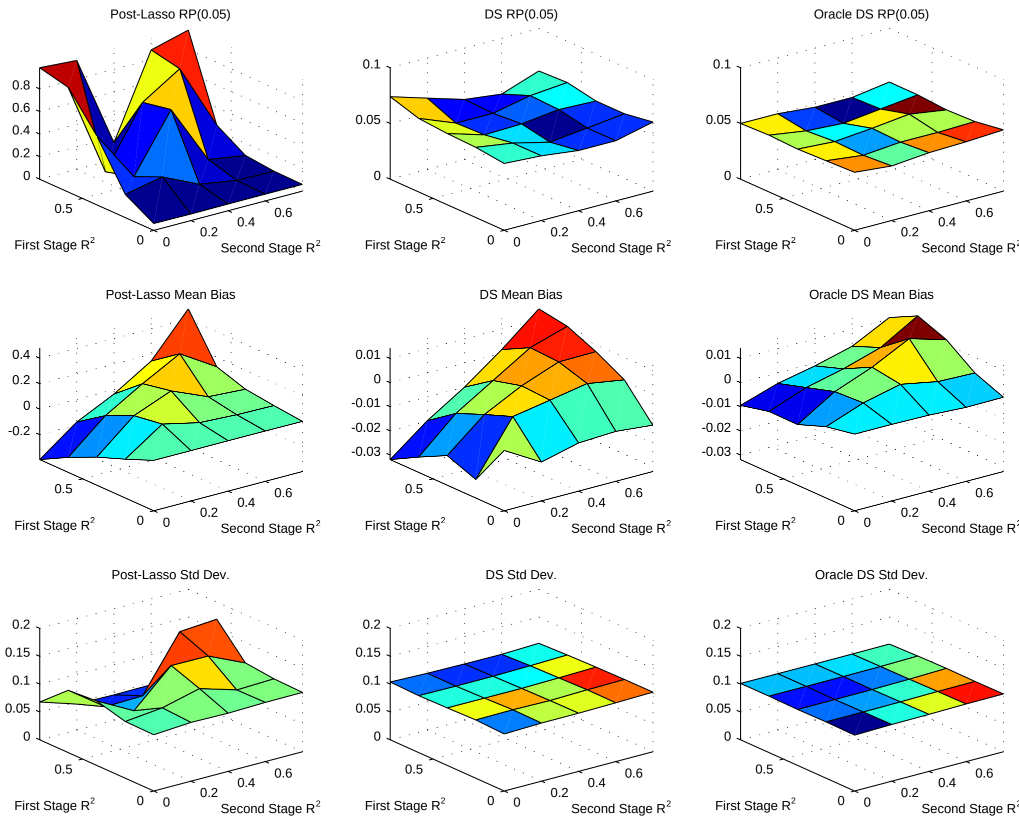

FIGURE 3. Design 22

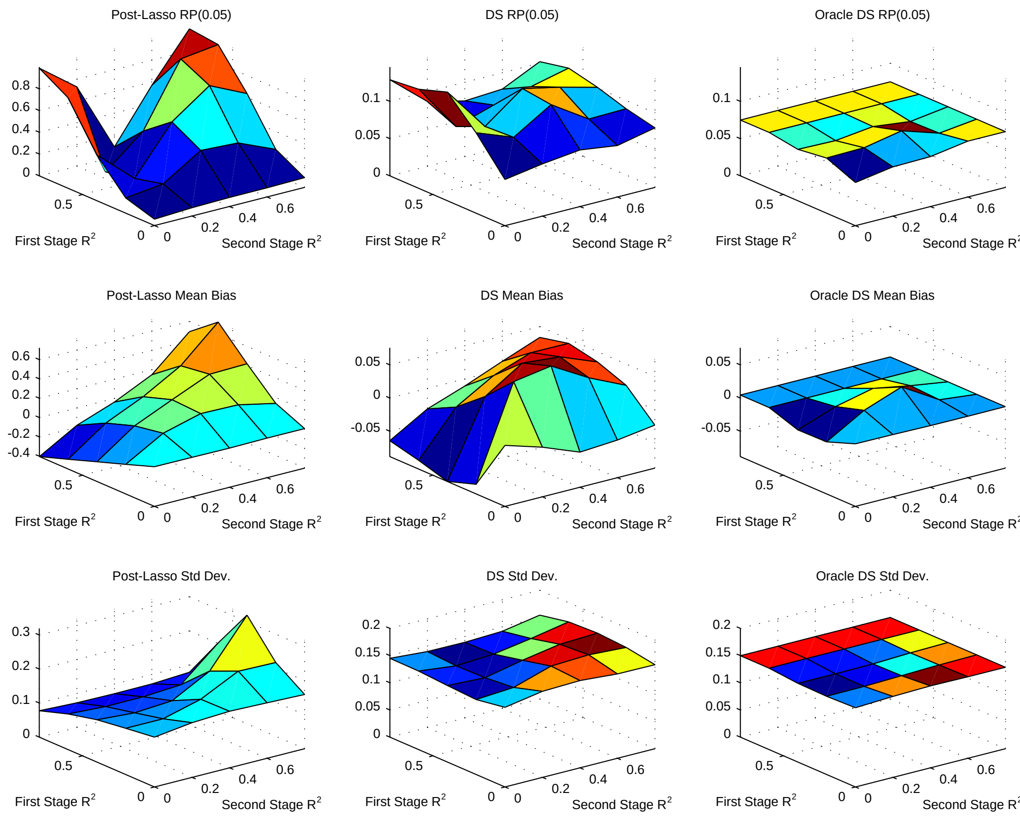

FIGURE 4. Design 3

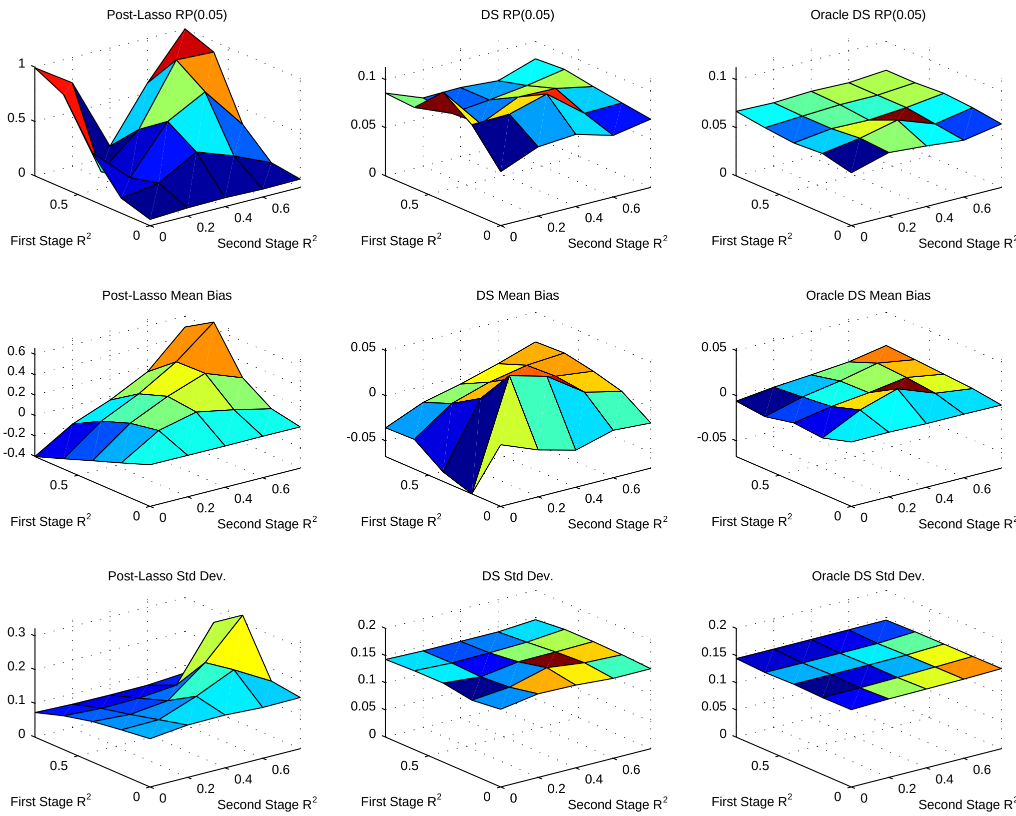

FIGURE 5. Design 4

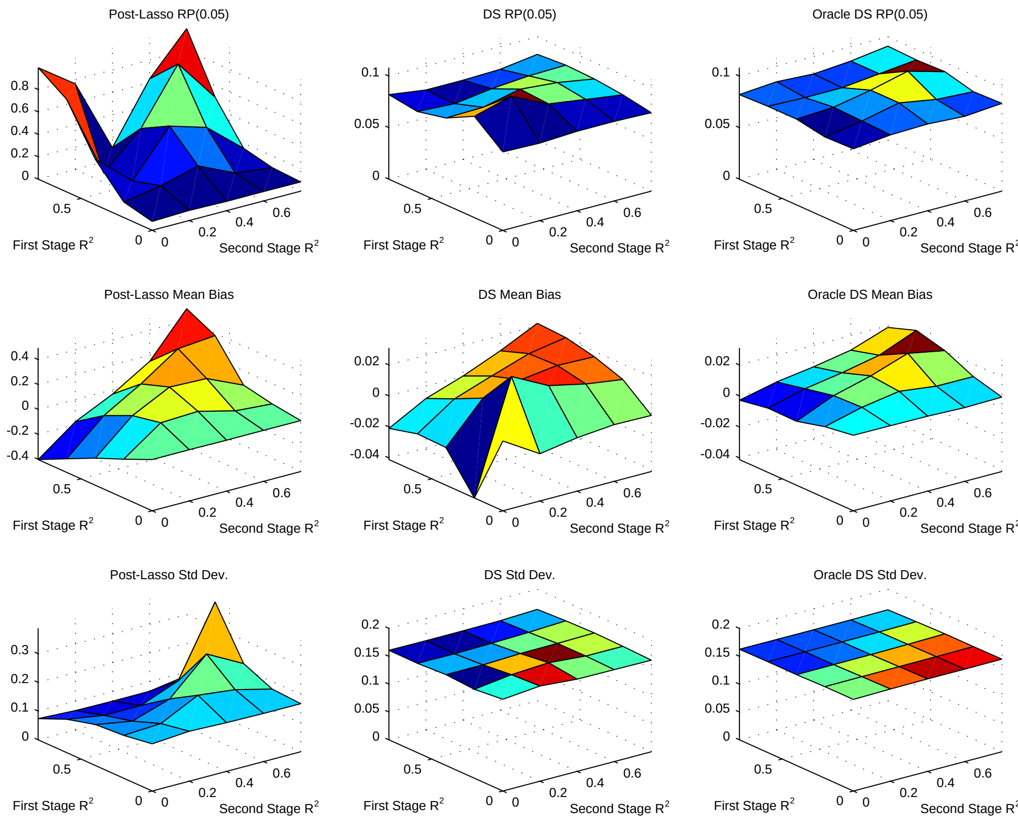

Figure 6. Design 44

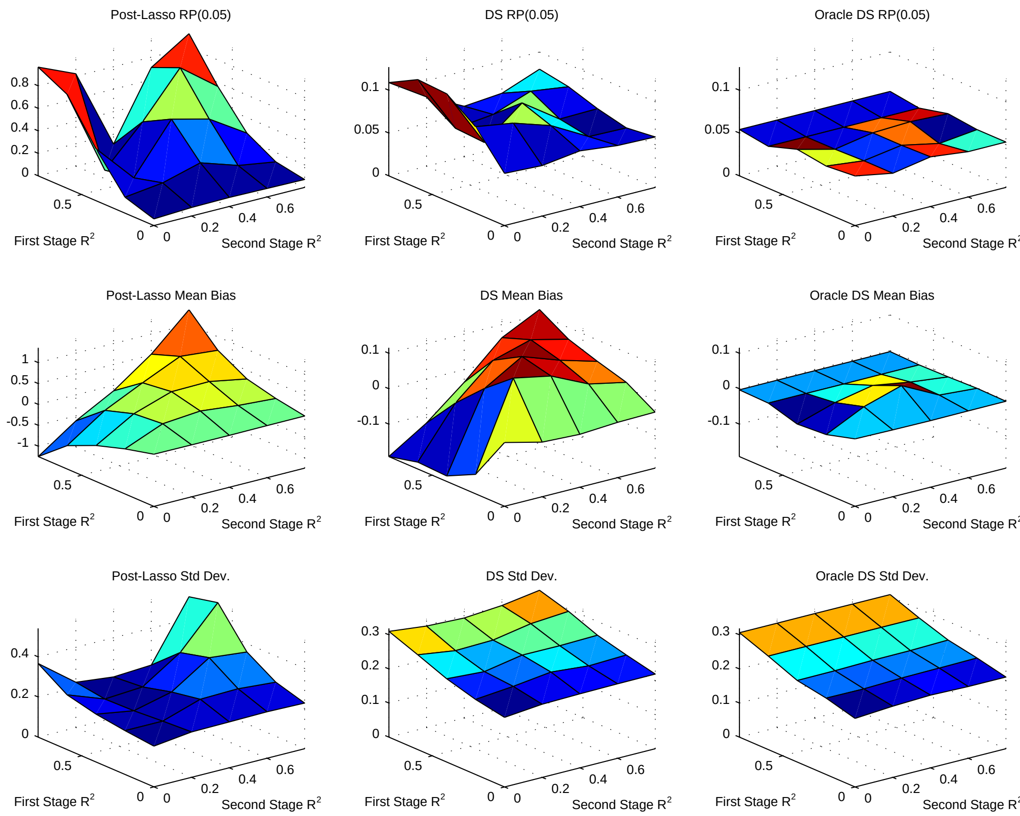

FIGURE 7. Design 5

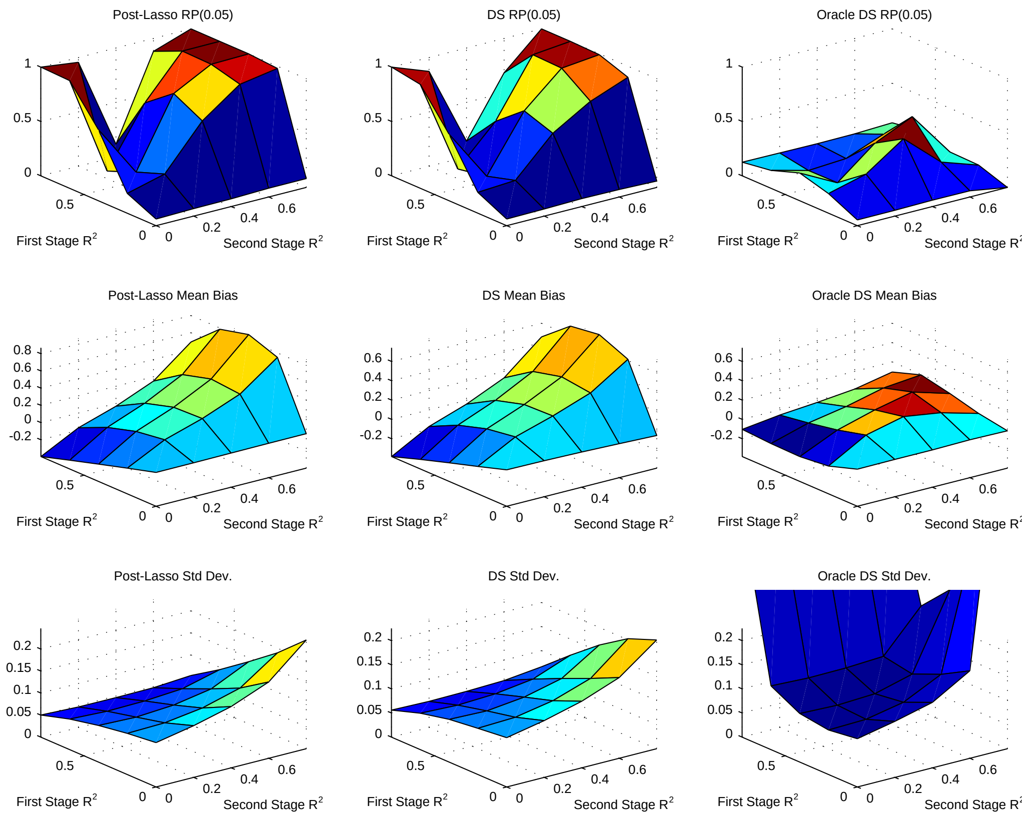

FIGURE 8. Design 6

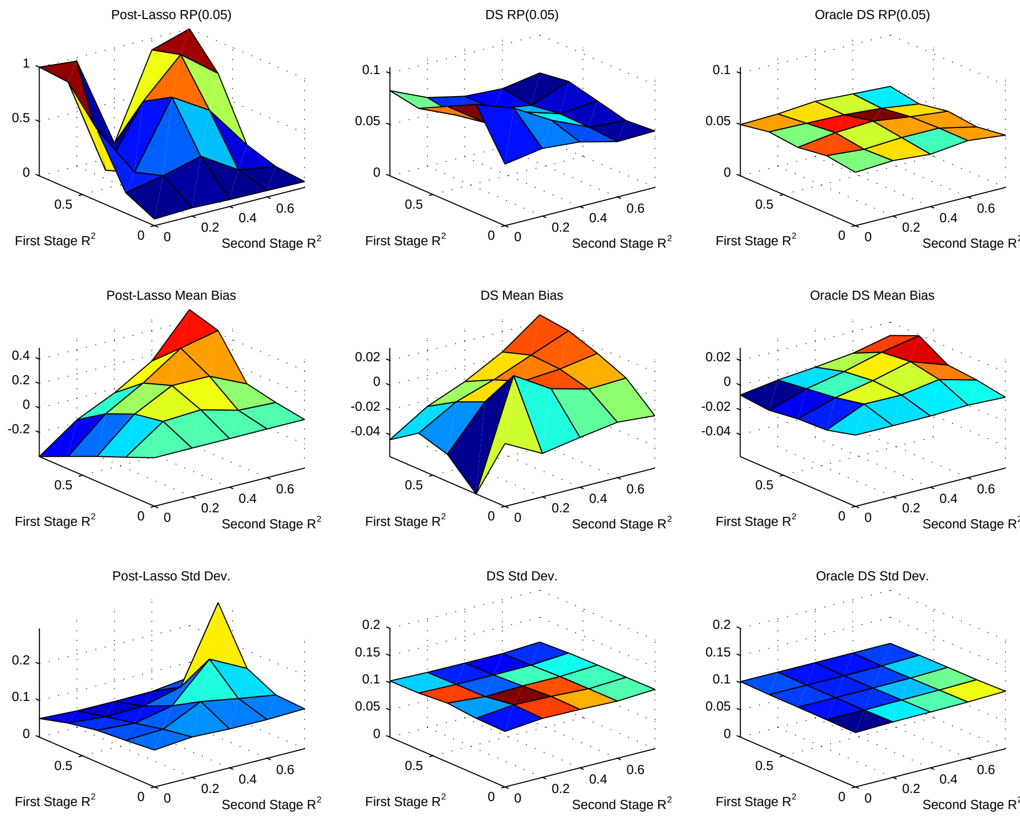

Figure 9. Design 7

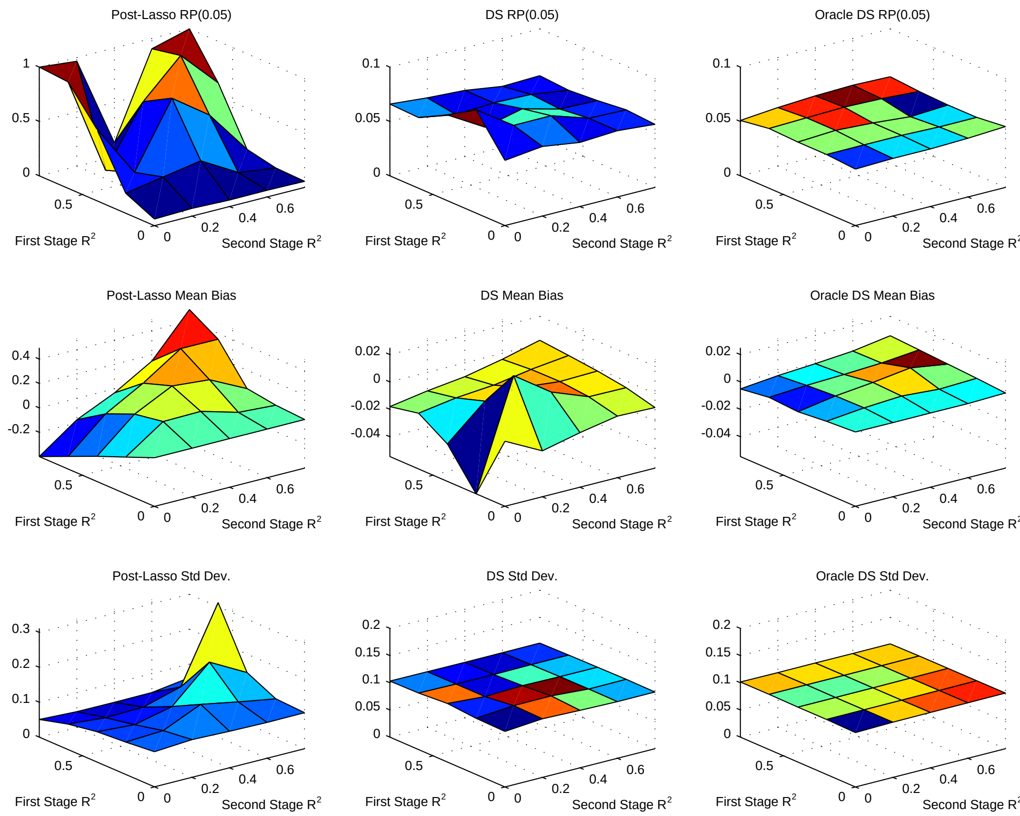

Figure 10. Design 72

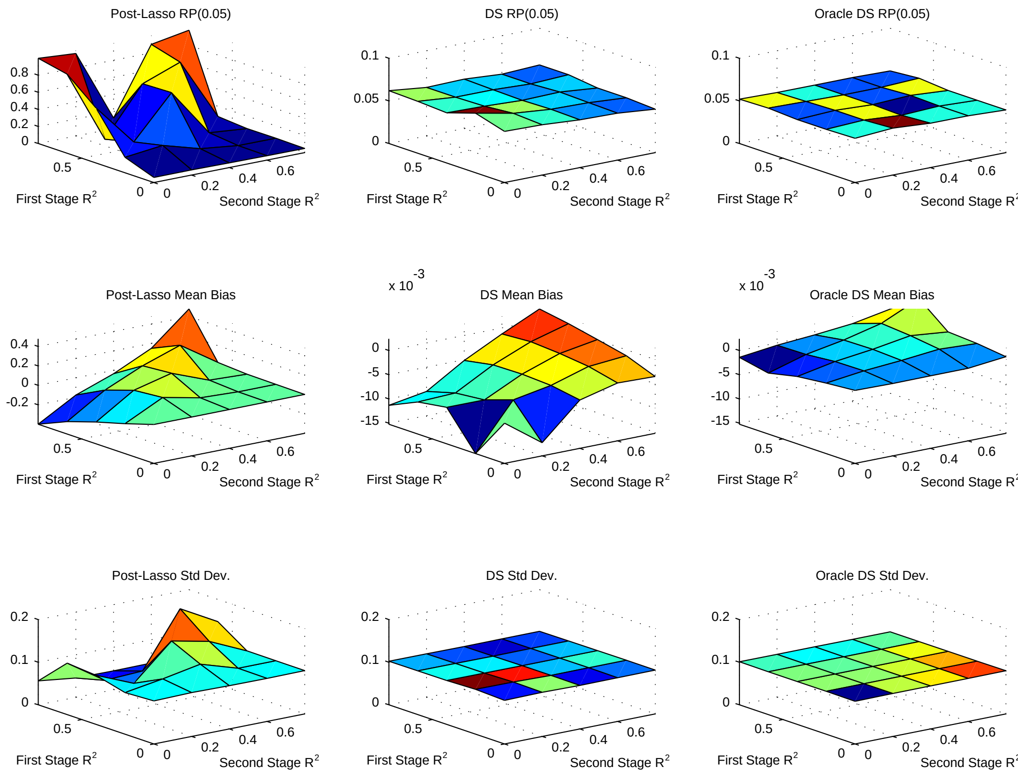

Figure 11. Design 722

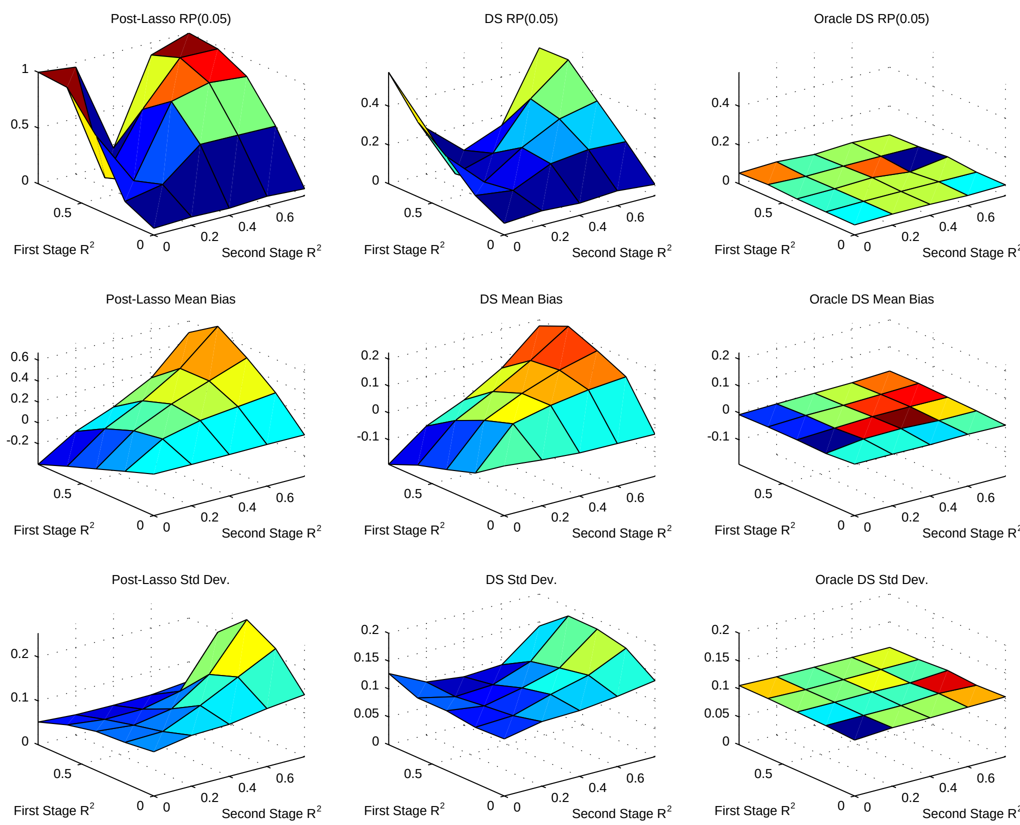

FIGURE 12. Design 8

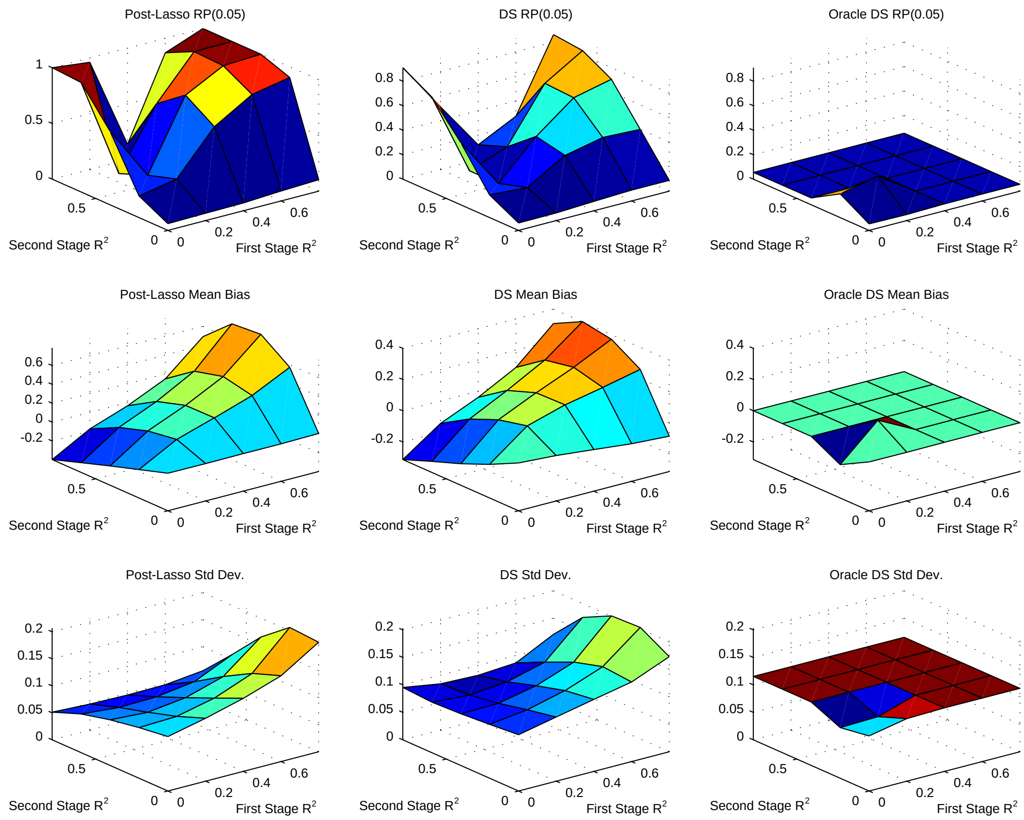

FIGURE 13. Design 1001

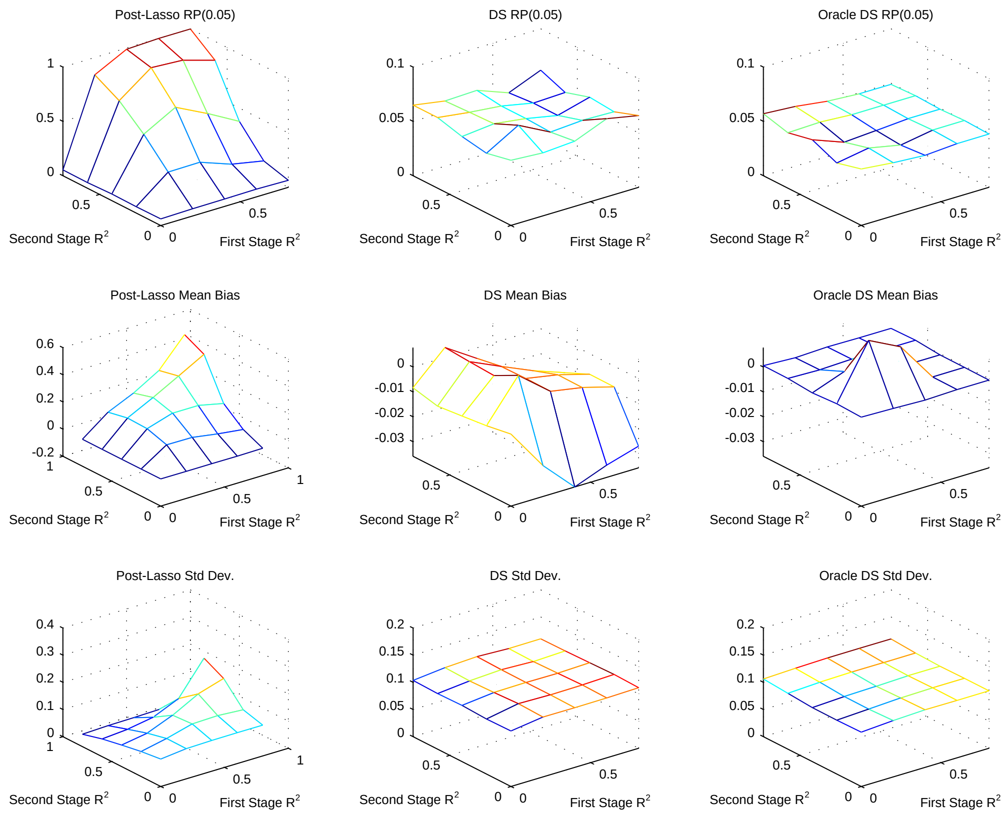

FIGURE 14. Design 1a

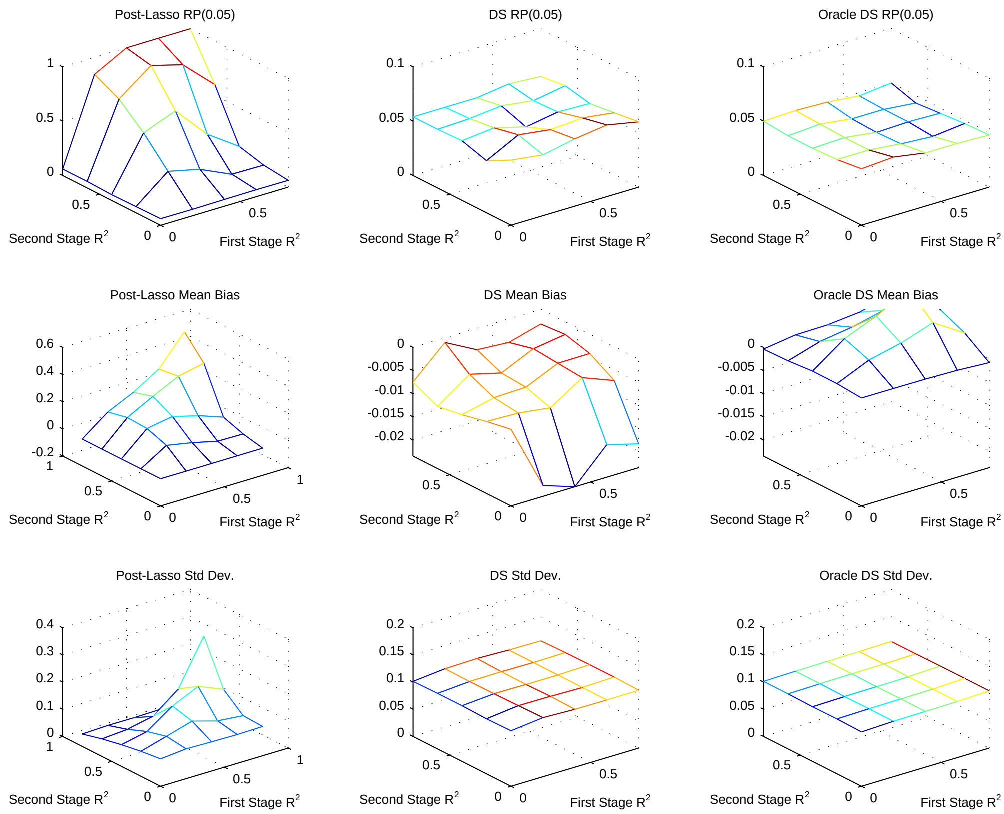

FIGURE 15. Design 2a

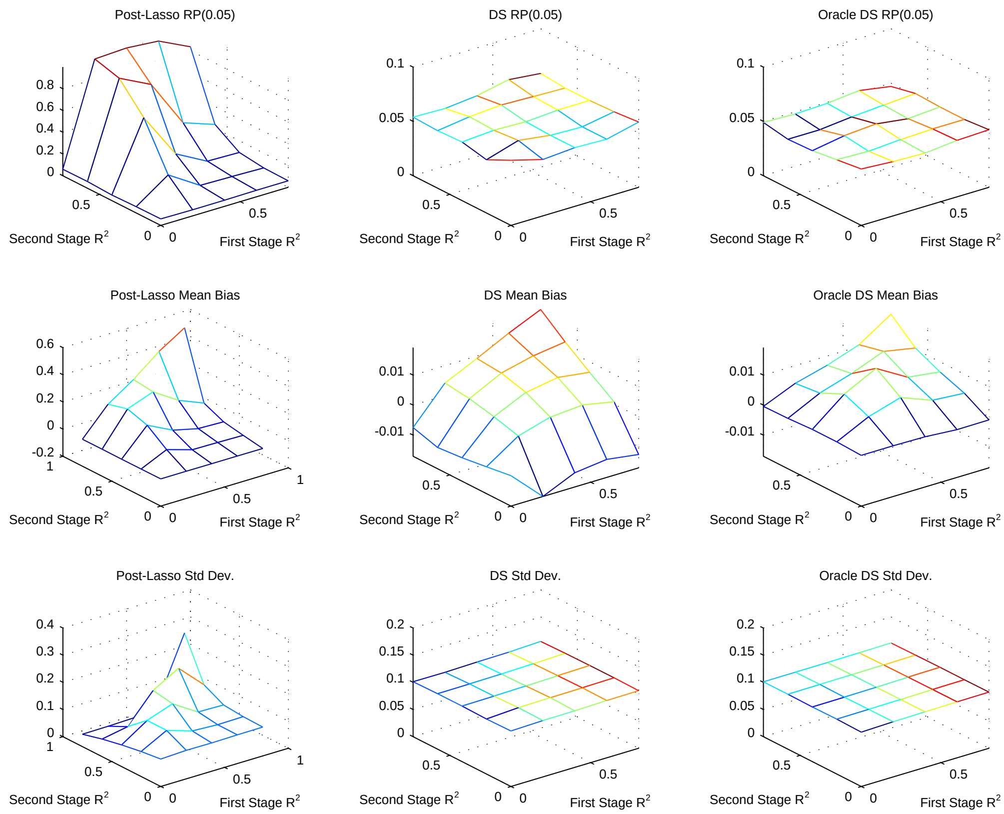

FIGURE 16. Design 22a

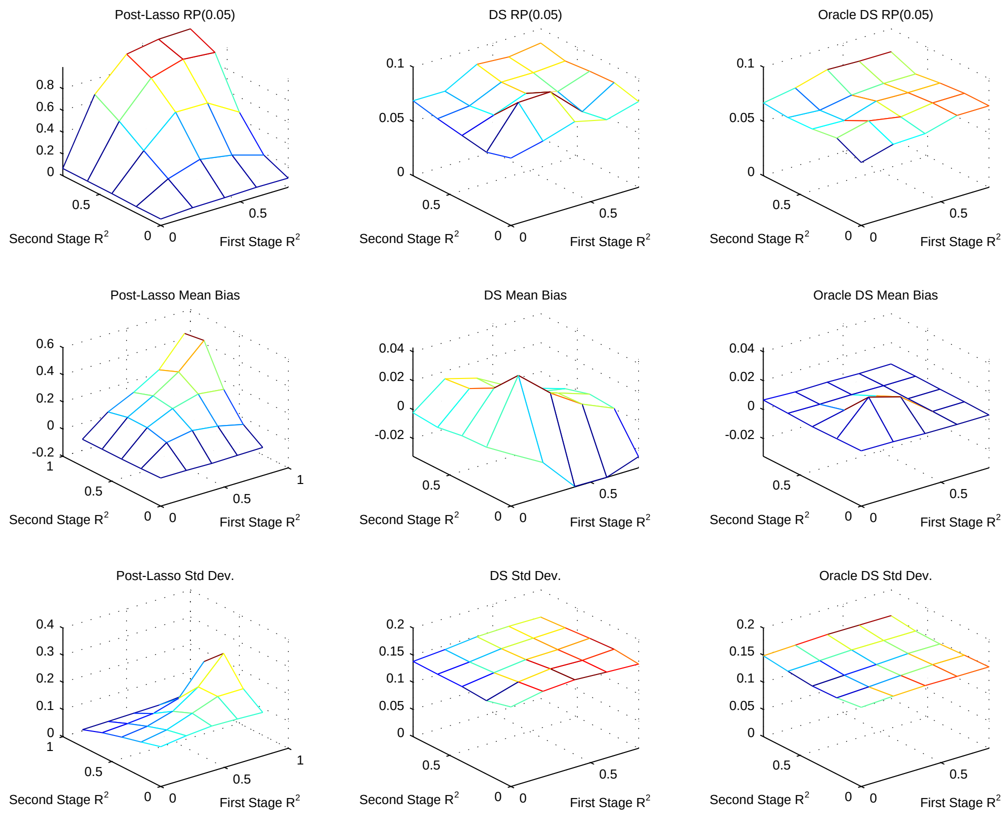

Figure 17. Design 3a

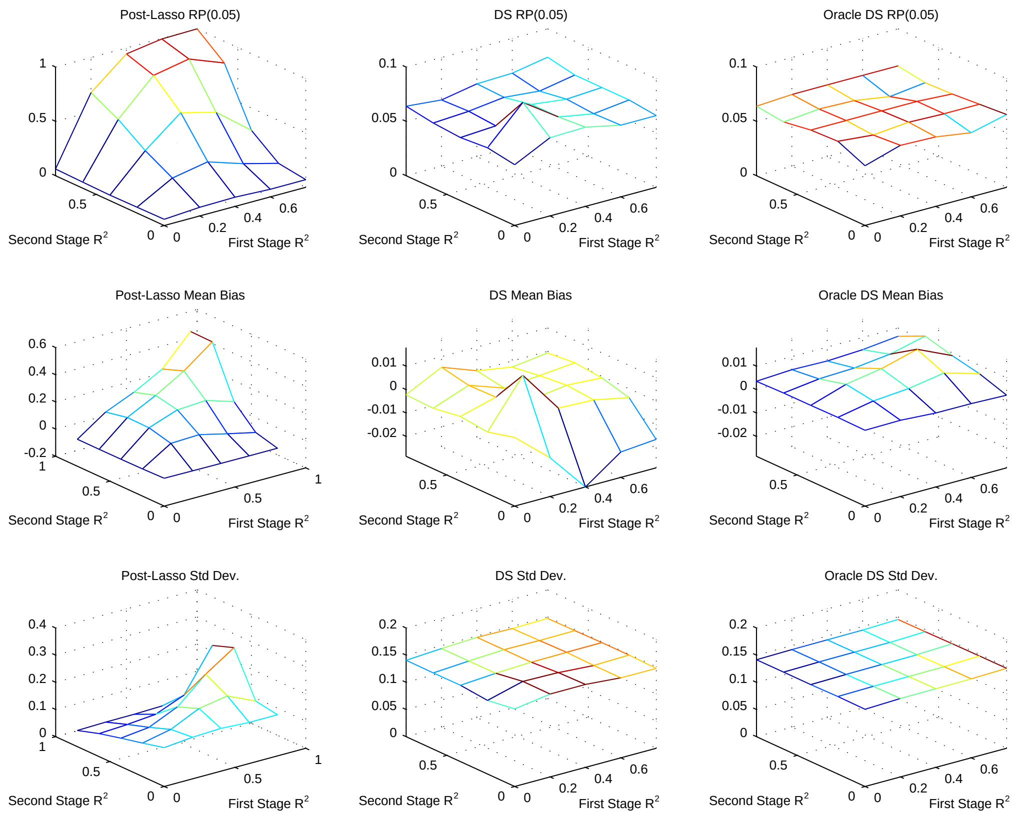

FIGURE 18. Design 4a

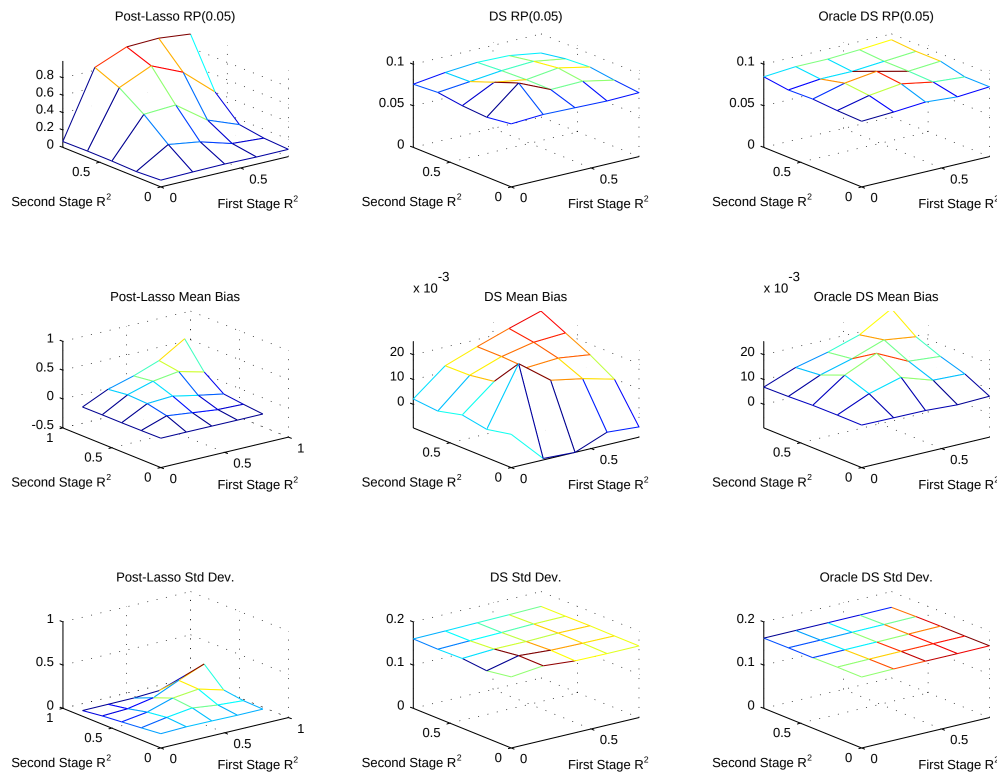

FIGURE 19. Design 44a

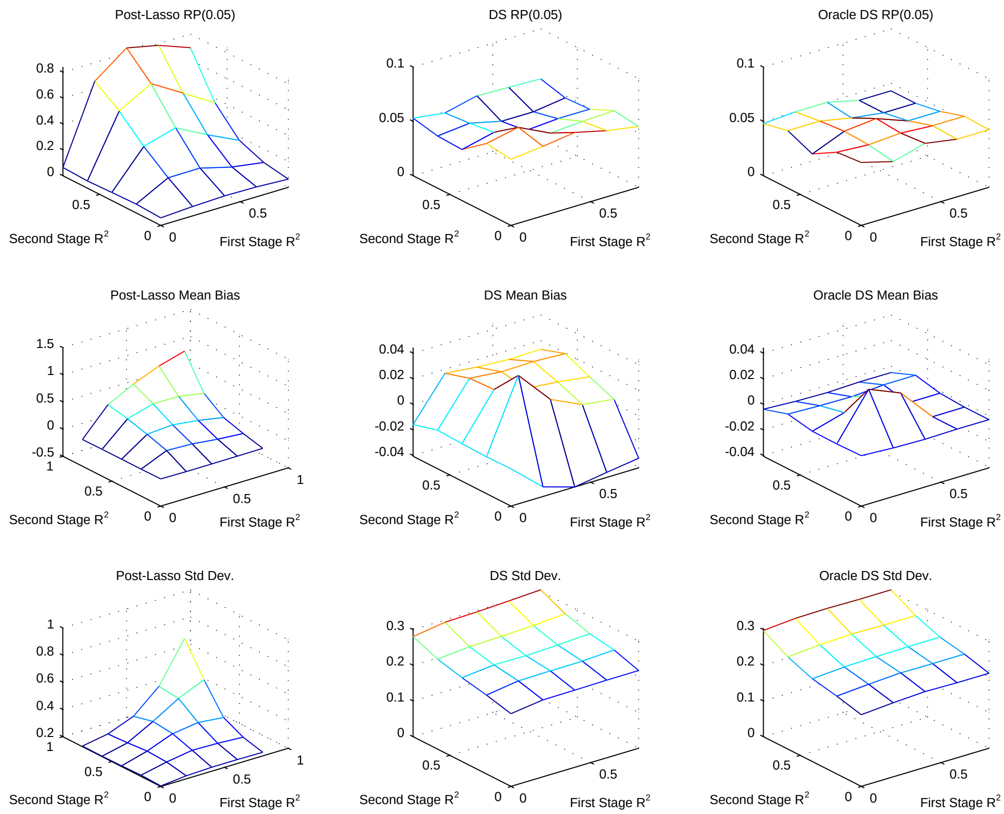

FIGURE 20. Design 5a

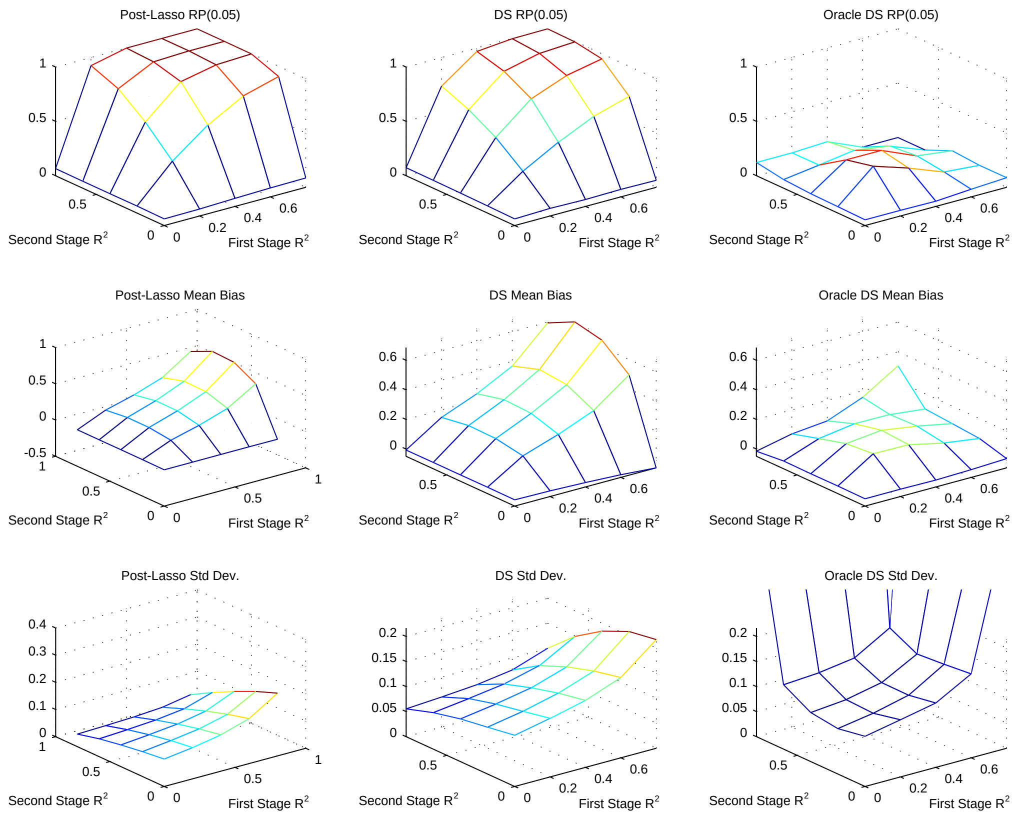

FIGURE 21. Design 6a

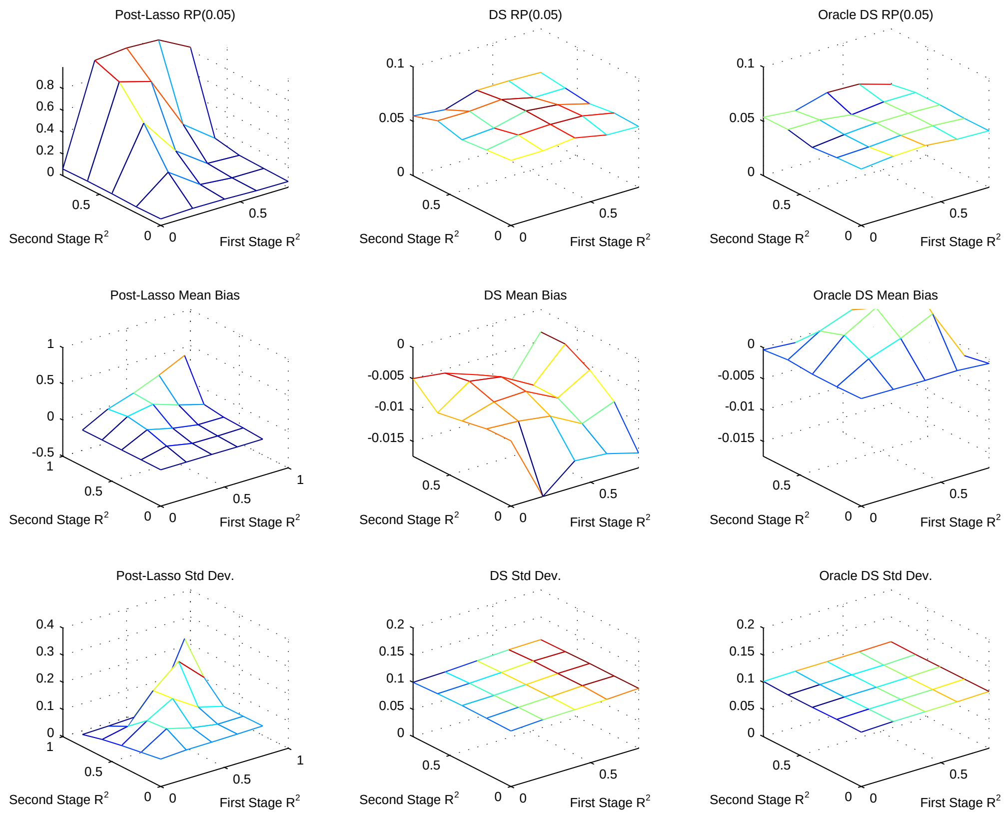

FIGURE 22. Design 7a

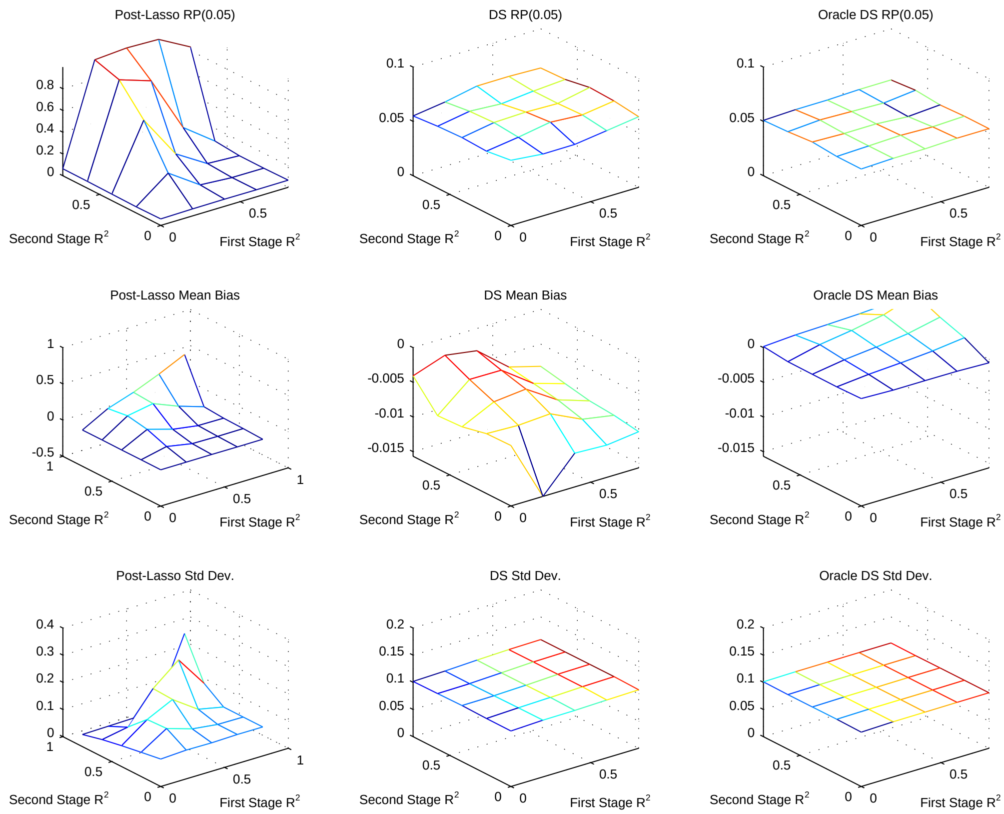

FIGURE 23. Design 72a

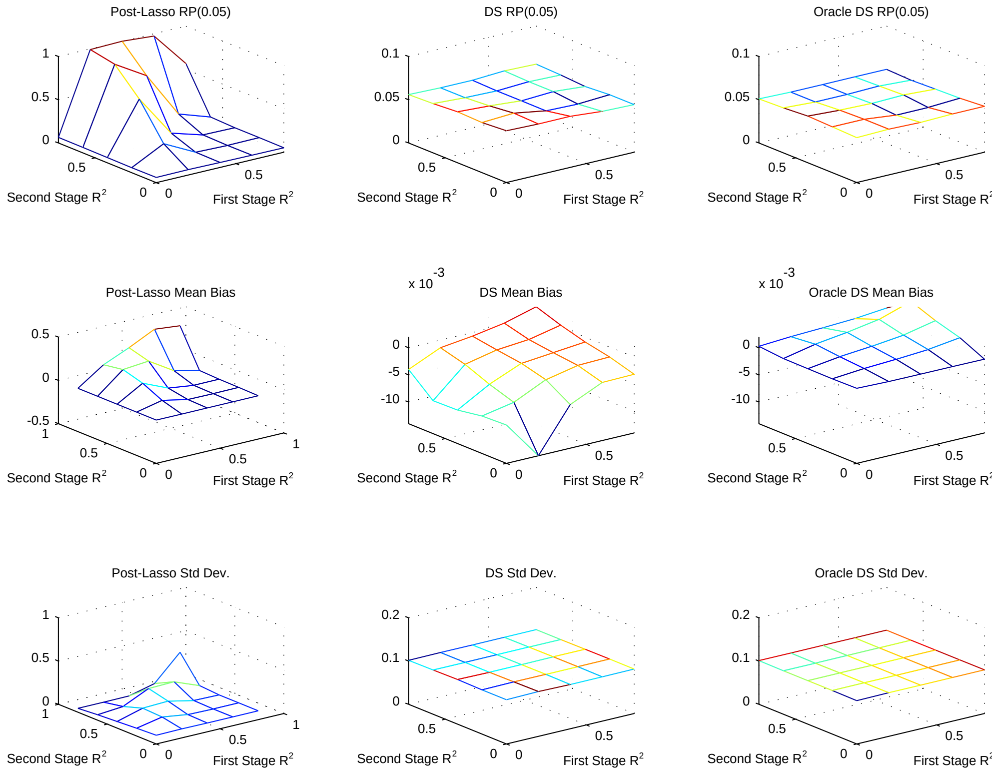

FIGURE 24. Design 722a

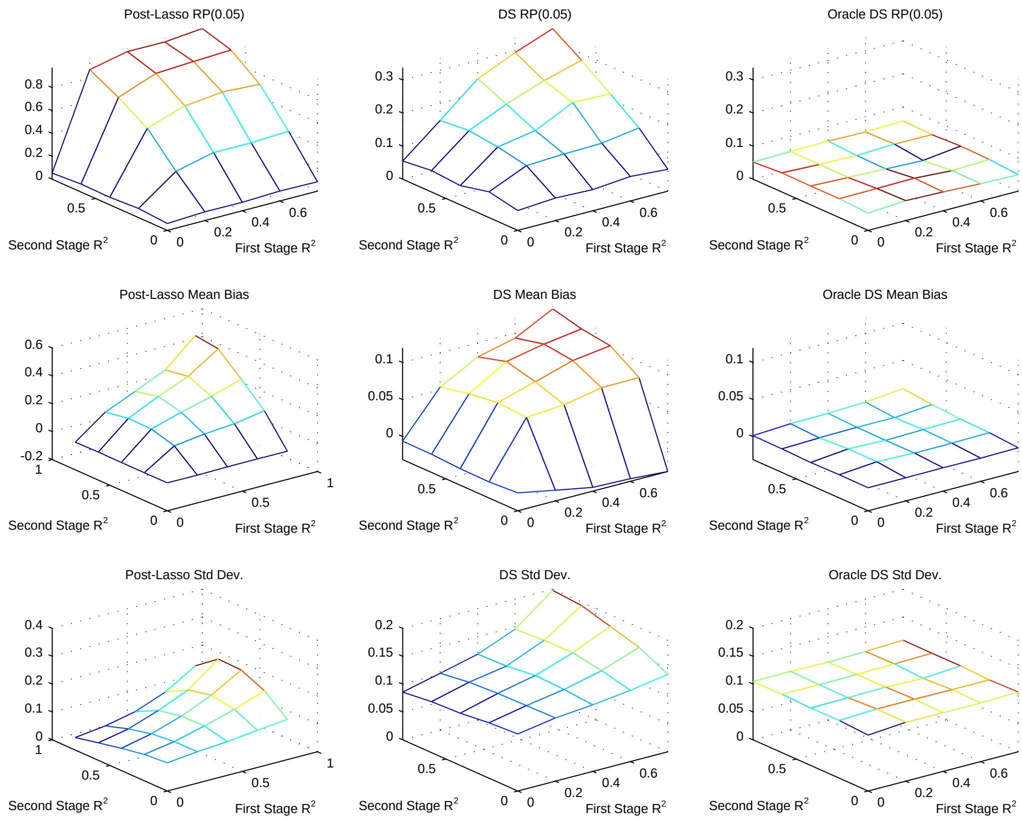

FIGURE 25. Design 8a

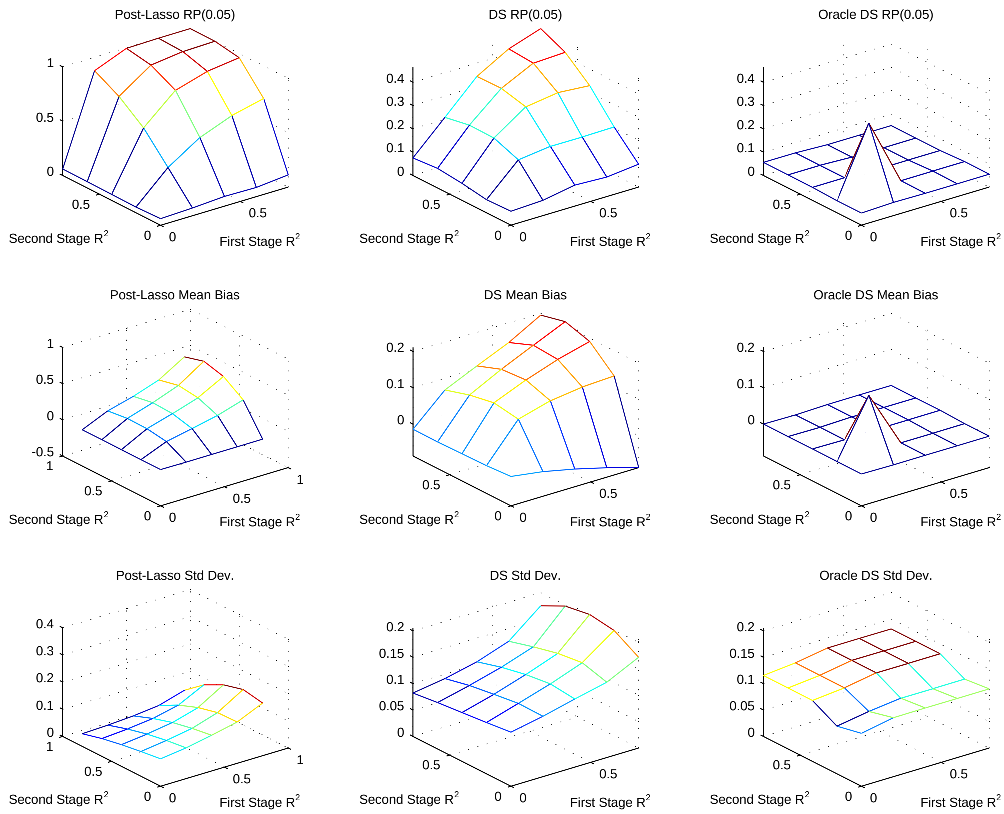

FIGURE 26. Design 1001a

References

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012): "Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain," *Econometrica*, 80, 2369–2429.

Belloni, A., and V. Chernozhukov (2013): "Least Squares After Model Selection in High-dimensional Sparse Models," *Bernoulli*, 19(2), 521–547.

Chamberlain, G. (1985): "Longitudinal Analysis of Labor Market Data," in *Heterogeneity, Omitted Variable Bias, and Duration Dependence*, ed. by J. J. Heckman, and B. Singer. Cambridge University Press.

Donohue III, J. J., and S. D. Levitt (2001): "The Impact of Legalized Abortion on Crime," *Quarterly Journal of Economics*, 116(2), 379–420.

——— (2008): "Measurement Error, Legalized Abortion, and the Decline in Crime: A Response to Foote and Goetz," *Quarterly Journal of Economics*, 123(1), 425–440.

Foote, C. L., and C. F. Goetz (2008): "The Impact of Legalized Abortion on Crime: Comment," *Quarterly Journal of Economics*, 123(1), 407–423.

Guédon, O., and M. Rudelson (2007): "L_p-moments of random vectors via majorizing measures," *Advances in Mathematics*, 208, 798–823.

Ledoux, M., and M. Talagrand (1991): *Probability in Banach Spaces (Isoperimetry and processes)*. Ergebnisse der Mathematik undihrer Grenzgebiete, Springer-Verlag.

MacKinnon, J. G., and H. White (1985): "Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties," *Journal of Econometrics*, 29, 305–325.

Rudelson, M. (1999): "Random vectors in the isotropic position," *Journal of Functional Analysis*, 164, 6072.

Rudelson, M., and R. Vershynin (2008): "On sparse reconstruction from Fourier and Gaussian measurements," *Communications on Pure and Applied Mathematics*, 61, 10251045.

Rudelson, M., and S. Zhou (2011): "Reconstruction from anisotropic random measurements," *ArXiv:1106.1151*.

van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes*. Springer Series in Statistics.

von Bahr, B., and C.-G. Esseen (1965): "Inequalities for the rth absolute moment of a sum of random variables, $1 \leqslant r \leqslant 2$," *Ann. Math. Statist*, 36, 299–303.