Supplementary Material for "LASSO for Stochastic Frontier Models with Many Efficient Firms"

William C. Horrace* Hyunseok Jung† Yoonseok Lee‡

This online Appendix contains proofs of the results in the main text of the article (Part A) and additional Monte Carlo simulation results (Part B).

A. Proofs

Let $\varkappa_{NT} = (\log N)/\sqrt{T}$. We first derive some technical lemmas.

Lemma A.1 *Suppose Assumption 2-(1) and 2-(2)-(ii) hold. Then, for some* $0 < C_x, C_v < \infty$*, as* $(N,T) \to \infty$*, we have*

$$\text{(a)} \quad \max_{1\le i\le N} \Pr\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\{x_{it} - E[x_{it}]\}\right\| \ge C_x \varkappa_{NT}\right) = o\left(N^{-1}\right), \text{ and}$$

$$\max_{1\le i\le N} \Pr\left(\left|\frac{1}{T}\sum_{t=1}^{T} v_{it}\right| \ge C_v \varkappa_{NT}\right) = o\left(N^{-1}\right);$$

$$\text{(b)} \quad \Pr\left(\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T}\{x_{it} - E[x_{it}]\}\right\| \ge C_x \varkappa_{NT}\right) = o(1), \text{ and}$$

$$\Pr\left(\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T} v_{it}\right| \ge C_v \varkappa_{NT}\right) = o(1).$$

*Department of Economics, Syracuse University, Syracuse, NY, 13244. whorrace@syr.edu

†Corresponding author: Department of Economics, University of Arkansas, Fayetteville, AR 72701. hj020@uark.edu

‡Department of Economics, Syracuse University, Syracuse, NY, 13244. ylee41@syr.edu

Proof of Lemma A.1 We only prove the first part of (a) since the proof for the second part of (a) is similar, and (a) implies (b), because

$$\begin{aligned}
\Pr\left(\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T}\{x_{it}-E[x_{it}]\}\right\|\ge C_x\varkappa_{NT}\right) &\le \sum_{i=1}^{N}\Pr\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\{x_{it}-E[x_{it}]\}\right\|\ge C_x\varkappa_{NT}\right)\\
&\le N\max_{1\le i\le N}\Pr\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\{x_{it}-E[x_{it}]\}\right\|\ge C_x\varkappa_{NT}\right)\\
&= N\cdot o(N^{-1})=o(1)
\end{aligned}$$

and similarly for the second part of (b), if (a) is true.

To prove the first result of (a), we let $M_T=\sqrt{T}/(\log T)^2$ and $\mathbf{1}_{it}=\mathbf{1}\{||x_{it}||<M_T\}$. We define

$$\begin{aligned}
\xi_{1,it} &= x_{it}\mathbf{1}_{it}-E\left[x_{it}\mathbf{1}_{it}\right],\\
\xi_{2,it} &= x_{it}\left(1-\mathbf{1}_{it}\right),\\
\xi_{3,it} &= -E\left[x_{it}\left(1-\mathbf{1}_{it}\right)\right].
\end{aligned}$$

Then, $x_{it}-E[x_{it}]=\xi_{1,it}+\xi_{2,it}+\xi_{3,it}$ and thus we have

$$\begin{aligned}
\max_{1\le i\le N}\Pr\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\{x_{it}-E[x_{it}]\}\right\|\ge C_x\varkappa_{NT}\right) &\le \max_{1\le i\le N}\Pr\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\xi_{1,it}\right\|+\left\|\frac{1}{T}\sum_{t=1}^{T}\xi_{2,it}\right\|\right.\\
&\qquad\left.+\left\|\frac{1}{T}\sum_{t=1}^{T}\xi_{3,it}\right\|\ge C_x\varkappa_{NT}\right).
\end{aligned}$$

We prove the first part of (a) by showing

$$\text{(a1)}\quad N\cdot\max_{1\le i\le N}\Pr\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\xi_{1,it}\right\|\ge\frac{C_x}{2}\varkappa_{NT}\right)=o\left(1\right),$$

$$\text{(a2)}\quad N\cdot\max_{1\le i\le N}\Pr\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\xi_{2,it}\right\|\ge\frac{C_x}{2}\varkappa_{NT}\right)=o\left(1\right),\text{ and}$$

$$\text{(a3)}\quad \max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T}\xi_{3,it}\right\|=o(\varkappa_{NT}).$$

To prove (a1), we let $\xi_{1,it}^{\varphi}=\varphi'\xi_{1,it}$ for some constant $p\times 1$ vector φ with $||\varphi||=1$. Then, by Assumption 2-(1)-(ii), $\xi_{1,it}^{\varphi}$ is a zero-mean strong mixing process, not necessarily stationary, with the mixing coefficients satisfying $\alpha[t]\le c_\alpha\rho^t$ for some $c_\alpha>0$ and $\rho\in(0,1)$. In addition, $\max_{1\le t\le T}|\xi_{1,it}^{\varphi}|\le 2M_T$ almost surely by construction. We define $v_N^2=\max_{1\le i\le N}\sup_{t\ge1}\{var(\xi_{1,it}^{\varphi})+2\sum_{s=t+1}^{\infty}|cov(\xi_{1,it}^{\varphi},\xi_{1,is}^{\varphi})|$, which is bounded by Assumption 2-(1)-(ii) and (iii), and the Davydov inequality. Then, by Lemma S1.1 of Su, Shi and Phillips (2016), there exists a constant $C_0>0$ such that for any $T\ge2$ and $C_x>0$,

$$\begin{aligned}N\cdot\max_{1\le i\le N}\Pr\left(\left|\frac{1}{T}\sum_{t=1}^{T}\xi_{1,it}^{\varphi}\right|\ge\frac{C_x}{2}\varkappa_{NT}\right)&\le N\exp\left(-\frac{C_0C_x^2T^2\varkappa_{NT}^2/4}{v_N^2T+4M_T^2+2C_xT\varkappa_{NT}M_T\left(\log T\right)^2/2}\right)\\&=\exp\left(-\left\{\frac{C_0C_x^2(\log N)^2/4}{v_N^2+4/(\log T)^4+C_x(\log N)}-\log N\right\}\right).\end{aligned}$$

Thus, by choosing C_x sufficiently large, it follows that

$$N\max_{1\le i\le N}\Pr\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\xi_{1,it}\right\|\ge\frac{C_x}{2}\varkappa_{NT}\right)\to0\quad\text{as }(N,T)\to\infty.$$

Next, by Assumption 2-(1)-(iii) and 2-(2)-(ii), and the Boole and Markov inequalities, we have

$$\begin{aligned}N\cdot\max_{1\le i\le N}\Pr\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\xi_{2,it}\right\|\ge\frac{C_x}{2}\varkappa_{NT}\right)&\le N\cdot\max_{1\le i\le N}\Pr\left(\max_{1\le t\le T}\|x_{it}\|\ge M_T\right)\\&\le NT\max_{1\le i\le N}\max_{1\le t\le T}\Pr\left(\|x_{it}\|\ge M_T\right)\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{NT}{M_T^q} \max_{1\leq i\leq N}\max_{1\leq t\leq T} E\,||x_{it}||^q \\
&= o(1).
\end{aligned}
$$

Lastly, by Assumption 2-(1)-(iii), and the Hölder and Markov inequalities,

$$
\begin{aligned}
\max_{1\leq i\leq N}\left\|\frac{1}{T}\sum_{t=1}^{T}\xi_{3,it}\right\| &\leq \max_{1\leq i\leq N}\max_{1\leq t\leq T} E\left\|x_{it}\mathbf{1}\left\{\|x_{it}\|\geq M_T\right\}\right\| \\
&\leq \max_{1\leq i\leq N}\max_{1\leq t\leq T}\left(E\left\|x_{it}\right\|^{q/2}\right)^{2/q}\max_{1\leq i\leq N}\max_{1\leq t\leq T}\left\{\Pr\left(||x_{it}||\geq M_T\right)\right\}^{(q-2)/q} \\
&\leq \max_{1\leq i\leq N}\max_{1\leq t\leq T}\left(E\left\|x_{it}\right\|^{q/2}\right)^{2/q}\max_{1\leq i\leq N}\max_{1\leq t\leq T}\left(\frac{E\left\|x_{it}\right\|^{q}}{M_T^q}\right)^{(q-2)/q} \\
&= O\left(M_T^{-(q-2)}\right) = o(\varkappa_{NT})
\end{aligned}
$$

where we use the fact that $M_T^{(q-2)}\varkappa_{NT} = T^{(q-3)/2}\log N/(\log T)^2 \to \infty$ for $q \geq 4$ in the last step. Then, the desired result follows by combining (a1), (a2) and (a3). ■

Proof of Lemma 1 First, note that

$$
\begin{aligned}
&\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\left\{x_{it}'(\beta_0-\hat{\beta})+v_{it}\right\}\right| \\
\leq\;& \left(\max_{1\leq i\leq N}\left\|\frac{1}{T}\sum_{t=1}^{T}\left\{x_{it}-E[x_{it}]\right\}\right\| + \max_{1\leq i\leq N}E||x_{it}||\right)\left\|\hat{\beta}-\beta_0\right\| + \max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}v_{it}\right|,
\end{aligned}
$$

where $\max_{1\leq i\leq N}E||x_{it}|| = O(1)$ and $\left\|\hat{\beta}-\beta_0\right\| = O_p((NT)^{-1/2})$ due to Assumption 2-(1)-(iii) and 2-(2)-(i), which implies for sufficiently large $0 < C < \infty$,

$$
\Pr\left(\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\left\{x_{it}'(\beta_0-\hat{\beta})+v_{it}\right\}\right| \geq C\varkappa_{NT}\right) = o\left(1\right) \tag{A.1}
$$

by Lemma A.1.

Recall $\eta = \min_{i \in \mathcal{S}^c} u_{0,i}$ and $\hat{\alpha}_i = T^{-1} \sum_{t=1}^{T} (y_{it} - x_{it}' \hat{\beta}) = T^{-1} \sum_{t=1}^{T} (\alpha_0 - u_{0,i} + x_{it}'(\beta_0 - \hat{\beta}) + v_{it})$ where $u_{0,i} = 0$ for all $i \in \mathcal{S}$. Thus, it follows that

$$
\begin{aligned}
& \min_{i \in \mathcal{S}} \hat{\alpha}_i - \max_{i \in \mathcal{S}^c} \hat{\alpha}_i \\
= \ & \min_{i \in \mathcal{S}} \left\{ \frac{1}{T} \sum_{t=1}^{T} (\alpha_0 + x_{it}'(\beta_0 - \hat{\beta}) + v_{it}) \right\} - \max_{i \in \mathcal{S}^c} \left\{ \frac{1}{T} \sum_{t=1}^{T} (\alpha_0 - u_{0,i} + x_{it}'(\beta_0 - \hat{\beta}) + v_{it}) \right\} \\
\geq \ & \min_{i \in \mathcal{S}^c} u_{0,i} + \left[\min_{i \in \mathcal{S}} \left\{ \frac{1}{T} \sum_{t=1}^{T} (x_{it}'(\beta_0 - \hat{\beta}) + v_{it}) \right\} - \max_{i \in \mathcal{S}^c} \left\{ \frac{1}{T} \sum_{t=1}^{T} (x_{it}'(\beta_0 - \hat{\beta}) + v_{it}) \right\} \right] \\
\geq \ & \eta - 2 \max_{1 \leq i \leq N} \left| \frac{1}{T} \sum_{t=1}^{T} (x_{it}'(\beta_0 - \hat{\beta}) + v_{it}) \right| \\
> \ & \frac{\eta}{2} - O_p(\varkappa_{NT}),
\end{aligned}
$$

which implies

$$
\Pr \left(\min_{i \in \mathcal{S}} \hat{\alpha}_i - \max_{i \in \mathcal{S}^c} \hat{\alpha}_i > 0 \right) \to 1 \tag{A.2}
$$

as $(N, T) \to \infty$ since $\eta > 0$ and $\eta / \varkappa_{NT} \to \infty$ by Assumption 2-(2)-(iii). (A.2), in turn, implies $\Pr(\hat{\alpha} = \max_{i \in \mathcal{S}} \hat{\alpha}_i) \to 1$ as $(N, T) \to \infty$ because $\hat{\alpha}$ is defined as $\max_{1 \leq i \leq N} \hat{\alpha}_i$.

By (A.2), we can let $\hat{\alpha} = \max_{i \in \mathcal{S}} \hat{\alpha}_i$ for sufficiently large (N, T), instead of $\hat{\alpha} = \max_{1 \leq i \leq N} \hat{\alpha}_i$. Hence, for sufficiently large (N, T), we have

$$
\begin{aligned}
|\hat{\alpha} - \alpha_0| &= \left| \max_{i \in \mathcal{S}} \left\{ \frac{1}{T} \sum_{t=1}^{T} \left(\alpha_0 + x_{it}'(\beta_0 - \hat{\beta}) + v_{it} \right) \right\} - \alpha_0 \right| \\
&\leq \max_{1 \leq i \leq N} \left| \frac{1}{T} \sum_{t=1}^{T} \left(x_{it}'(\beta_0 - \hat{\beta}) + v_{it} \right) \right| = O_p(\varkappa_{NT})
\end{aligned}
$$

from A.1, which proves Lemma 1.

Since $\hat{u}_i = \hat{\alpha} - \hat{\alpha}_i = (\hat{\alpha} - \alpha_0) + (\alpha_0 - \hat{\alpha}_i) = (\hat{\alpha} - \alpha_0) + (u_{0,i} + \alpha_{0,i} - \hat{\alpha}_i)$ so that

$|\hat{u}_i - u_{0,i}| \leq |\hat{\alpha} - \alpha_0| + |\hat{\alpha}_i - \alpha_{0,i}| \leq 2\max_{1\leq i\leq N} \left| \frac{1}{T}\sum_{t=1}^{T} x_{it}'(\beta_0 - \hat{\beta}) + v_{it} \right|$ by the results above, we also have

$$\Pr\left(|\hat{u}_i - u_{0,i}| \geq C\varkappa_{NT}\right) = o\left(1\right) \tag{A.3}$$

for sufficiently large $0 < C < \infty$. ■

Proof of Theorem 1 For Equation (5) in the main text, we form a Lagrangian as

$$\mathcal{L}\left(\alpha, \{u_i\}_{i=1}^N, \{\rho_i\}_{i=1}^N\right) = \sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - x_{it}'\widehat{\beta} - \alpha + u_i\right)^2 + \lambda\sum_{i=1}^{N}\pi_i u_i - \sum_{i=1}^{N}\rho_i u_i,$$

where $\rho_i \geq 0$, $u_i \geq 0$, and $\rho_i u_i = 0$ (complementary slackness) for all i. From the Karush-Kuhn-Tucker (KKT) conditions, we have

$$\widehat{\alpha}\left(\lambda\right) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - x_{it}'\widehat{\beta} + \widehat{u}_i\left(\lambda\right)\right) \tag{A.4}$$

$$\widehat{u}_i\left(\lambda\right) = \max\left\{0, \widehat{\alpha}\left(\lambda\right) - \frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - x_{it}'\widehat{\beta}\right) - \frac{\lambda}{2T}\hat{\pi}_i\right\}. \tag{A.5}$$

Recall $\delta = |\mathcal{S}|/N$ and let $\hat{\delta} = |\widehat{\mathcal{S}}|/N$. By plugging (A.5) into (A.4), we have

$$\begin{aligned}
\widehat{\alpha}\left(\lambda\right) &= \frac{1}{NT}\sum_{i\in\widehat{\mathcal{S}}}\sum_{t=1}^{T}\left(y_{it} - x_{it}'\widehat{\beta}\right) + \frac{1}{NT}\sum_{i\in\widehat{\mathcal{S}}^c}\sum_{t=1}^{T}\left(y_{it} - x_{it}'\widehat{\beta} + \widehat{u}_i\left(\lambda\right)\right) \\
&= \frac{1}{NT}\sum_{i\in\widehat{\mathcal{S}}}\sum_{t=1}^{T}\left(y_{it} - x_{it}'\widehat{\beta}\right) + \frac{1}{N}\sum_{i\in\widehat{\mathcal{S}}^c}\left(\widehat{\alpha}\left(\lambda\right) - \frac{\lambda}{2T}\hat{\pi}_i\right) \\
&= \frac{1}{NT}\sum_{i\in\widehat{\mathcal{S}}}\sum_{t=1}^{T}\left(x_{it}'\left(\beta_0 - \widehat{\beta}\right) + \alpha_0 - u_{0,i} + v_{it}\right) + \left(1 - \widehat{\delta}\right)\widehat{\alpha}\left(\lambda\right) - \frac{\lambda}{2NT}\sum_{i\in\widehat{\mathcal{S}}^c}\hat{\pi}_i
\end{aligned}$$

and hence

$$\widehat{\alpha}\left(\lambda\right)-\alpha_{0}=\frac{1}{\hat{\delta}NT}\sum_{i\in\widehat{\mathcal{S}}}\sum_{t=1}^{T}\left(x_{it}^{\prime}\left(\beta_{0}-\widehat{\beta}\right)-u_{0,i}+v_{it}\right)-\frac{\lambda}{2\hat{\delta}NT}\sum_{i\in\widehat{\mathcal{S}}^{c}}\hat{\pi}_{i}. \tag{A.6}$$

This shows that $\hat{\alpha}(\lambda)$ is estimated as a common intercept for the firms classified as fully efficient by the LASSO and also contains bias due to the use of shrinkage on $\hat{u}_i(\lambda)$. From (A.5), it follows that, for $i\in\widehat{\mathcal{S}}^{c}$ (i.e. $\widehat{u}_{i}\left(\lambda\right)>0$),

$$\begin{aligned}
\widehat{u}_{i}\left(\lambda\right) &= \widehat{\alpha}\left(\lambda\right)-\frac{1}{T}\sum_{t=1}^{T}\left(x_{it}^{\prime}\left(\beta_{0}-\widehat{\beta}\right)+\alpha_{0}-u_{0,i}+v_{it}\right)-\frac{\lambda}{2T}\hat{\pi}_{i}\\
&= \frac{1}{\hat{\delta}NT}\sum_{j\in\widehat{\mathcal{S}}}\sum_{t=1}^{T}\left(x_{jt}^{\prime}\left(\beta_{0}-\widehat{\beta}\right)-u_{0,j}+v_{jt}\right)-\frac{1}{T}\sum_{t=1}^{T}\left(x_{it}^{\prime}\left(\beta_{0}-\widehat{\beta}\right)-u_{0,i}+v_{it}\right)\\
&\quad -\frac{\lambda}{2\hat{\delta}NT}\sum_{j\in\widehat{\mathcal{S}}^{c}}\hat{\pi}_{j}-\frac{\lambda}{2T}\hat{\pi}_{i}.
\end{aligned}$$

We prove the theorem by showing $\mathcal{S}\subset\hat{\mathcal{S}}$ and $\mathcal{S}^c\subset\hat{\mathcal{S}}^c$ w.p.a.1.

(i) We first prove $\mathcal{S}\subset\hat{\mathcal{S}}$ w.p.a.1 by showing $\Pr\left(\max_{i\in\mathcal{S}}\hat{u}_i(\lambda)>0\right)\to 0$. Let $\hat{\tau}=\max_{i\in\mathcal{S}}\hat{u}_i$. Then, from (A.5), for any $C>0$, we have

$$\begin{aligned}
\Pr\left(\max_{i\in\mathcal{S}}\hat{u}_i(\lambda)>0\right) &= \Pr\left(\max_{i\in\mathcal{S}}\left\{\hat{\alpha}(\lambda)-\hat{\alpha}_i-\frac{\lambda}{T}\hat{\pi}_i\right\}>0\right)\\
&\leq \Pr\left(\max_{i\in\mathcal{S}}\left\{\hat{\alpha}(\lambda)-\hat{\alpha}_i-\frac{\lambda}{T}\hat{\pi}_i\right\}>0,\hat{\tau}\leq C\varkappa_{NT}\right)+\Pr(\hat{\tau}>C\varkappa_{NT})\\
&\leq \Pr\left(\max_{i\in\mathcal{S}}\left\{\frac{1}{\hat{\delta}NT}\sum_{j\in\widehat{\mathcal{S}}}\sum_{t=1}^{T}\left(x_{jt}^{\prime}(\beta_0-\hat{\beta})-u_{0,j}+v_{jt}\right)-\frac{\lambda}{2\hat{\delta}NT}\sum_{j\in\widehat{\mathcal{S}}^{c}}\hat{\pi}_j\right.\right.\\
&\qquad\left.\left.-\frac{1}{T}\sum_{t=1}^{T}\left(x_{it}^{\prime}(\beta_0-\hat{\beta})+v_{it}\right)-\frac{\lambda}{2T}(C\varkappa_{NT})^{-\gamma}\right\}>0\right)+\Pr(\hat{\tau}>C\varkappa_{NT})
\end{aligned}$$

$$
\begin{aligned}
&\leq \Pr\left(2\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}x_{it}'(\beta_0-\hat{\beta})+v_{it}\right|-\frac{\lambda}{2T}(C\varkappa_{NT})^{-\gamma}>0\right)\\
&\quad+\Pr(\hat{\tau}>C\varkappa_{NT}) \qquad \text{(A.7)}
\end{aligned}
$$

where we use the fact that $u_{0,j}\geq 0$ and $\hat{\pi}_j\geq 0$ for all j in the last step. Then, by choosing sufficiently large $0<C<\infty$, we can easily show that first term in (A.7) is $o(1)$ due to (A.1) and $((\lambda/T)\varkappa_{NT}^{-\gamma})/\varkappa_{NT}\to\infty$ as $(N,T)\to\infty$ by Assumption 2-(3). The second term in (A.7) is also $o(1)$ because

$$
\hat{\tau}=\max_{i\in\mathcal{S}}\hat{u}_i=\max_{i\in\mathcal{S}}\{(\hat{\alpha}-\alpha_0)-(\hat{\alpha}_i-\alpha_{0,i})\}\leq 2\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}x_{it}'(\beta_0-\hat{\beta})+v_{it}\right|
$$

and (A.1), where we use the fact $u_{0,i}=0$ for $i\in\mathcal{S}$.

(ii) Next, we prove $\mathcal{S}^c\subset\hat{\mathcal{S}}^c$ w.p.a.1. Define $\mathcal{D}_i\equiv\{\hat{u}_i(\lambda)=0\}$ and then,

$$
\Pr\left(\text{there exists } i\in\mathcal{S}^c \text{ such that } \hat{u}_i(\lambda)=0\right)=\Pr\left(\bigcup_{i\in\mathcal{S}^c}\mathcal{D}_i\right).
$$

Let $|\mathcal{S}^c|=J$. We arbitrarily list the firms in $\mathcal{S}^c$ and use an auxiliary index, $[j]$ for $j=1,...,J$, to denote the j^{th} firm on the list. Then, we can partition $\bigcup_{i\in\mathcal{S}^c}\mathcal{D}_i$ into disjoint sets such that $\mathcal{D}_{[1]}\cap\left(\bigcup_{j=2}^{J}\mathcal{D}_{[j]}\right)^c$, $\mathcal{D}_{[2]}\cap\left(\bigcup_{j=3}^{J}\mathcal{D}_{[j]}\right)^c$, ..., and $\mathcal{D}_{[J]}$. Therefore, we have

$$
\begin{aligned}
&\Pr\left(\bigcup_{i\in\mathcal{S}^c}\mathcal{D}_i\right)\\
=&\sum_{j=1}^{J}\Pr\left(\mathcal{D}_{[j]}\cap\left(\bigcup_{k=j+1}^{J}\mathcal{D}_{[k]}\right)^c\right)\\
=&\sum_{j=1}^{J}\Pr\left(\hat{u}_{[j]}(\lambda)=0,\hat{u}_{[j+1]}(\lambda)>0,\hat{u}_{[j+2]}(\lambda)>0,...,\hat{u}_{[J]}(\lambda)>0\right),
\end{aligned}
$$

which is true regardless of the order of the firms on the list. So, we list the firms in $\mathcal{S}^c$ according to the size of inefficiency in ascending order so that $u_{0,[1]} \leq ... \leq u_{0,[j]} ... \leq u_{0,[J]}$. Then, we have

$$
\begin{aligned}
& \Pr\left(\text{there exists } i \in \mathcal{S}^c \text{ such that } \hat{u}_i(\lambda) = 0\right) \\
= & \sum_{j=1}^{J} \Pr\left(\hat{u}_{[j]}(\lambda) = 0, \hat{u}_{[j+1]}(\lambda) > 0, \hat{u}_{[j+2]}(\lambda) > 0, ..., \hat{u}_{[J]}(\lambda) > 0\right) \\
= & \sum_{j=1}^{J} \Pr\left(\hat{u}_{[j]}(\lambda) = 0 \,\middle|\, \hat{u}_{[j+1]}(\lambda) > 0, ..., \hat{u}_{[J]}(\lambda) > 0\right) \times \Pr\left(\hat{u}_{[j+1]}(\lambda) > 0 \,\middle|\, \hat{u}_{[j+2]}(\lambda) > 0, ...\right) ... \\
& ... \times \Pr\left(\hat{u}_{[J-1]}(\lambda) > 0 \,\middle|\, \hat{u}_{[J]}(\lambda) > 0\right) \times \Pr\left(\hat{u}_{[J]}(\lambda) > 0\right) \\
\leq & \sum_{j=1}^{J} \Pr\left(\hat{u}_{[j]}(\lambda) = 0 \,\middle|\, \hat{u}_{[j+1]}(\lambda) > 0, ..., \hat{u}_{[J]}(\lambda) > 0\right) \\
= & \sum_{j=1}^{J} \Pr\Bigg(\frac{1}{\hat{\delta}NT} \sum_{i \in \hat{\mathcal{S}}} \sum_{t=1}^{T} \left(x_{it}'(\beta_0 - \hat{\beta}) - u_{0,i} + v_{it}\right) - \frac{\lambda}{2\hat{\delta}NT} \sum_{i \in \hat{\mathcal{S}}^c} \hat{\pi}_i \\
& -\frac{1}{T} \sum_{t=1}^{T} \left(x_{[j]t}'(\beta_0 - \hat{\beta}) - u_{0,[j]} + v_{[j]t}\right) - \frac{\lambda}{2T} \hat{\pi}_{[j]} < 0 \,\Bigg|\, \hat{u}_{[j+1]}(\lambda) > 0, ..., \hat{u}_{[J]}(\lambda) > 0\Bigg) \\
= & \sum_{j=1}^{J} \Pr\Bigg(\underbrace{u_{0,[j]} - \frac{\sum_{i \in \hat{\mathcal{S}}} u_{0,i}}{\hat{\delta}N}}_{(*)} + \frac{1}{\hat{\delta}NT} \sum_{i \in \hat{\mathcal{S}}} \sum_{t=1}^{T} \left(x_{it}'(\beta_0 - \hat{\beta}) + v_{it}\right) - \frac{\lambda}{2\hat{\delta}NT} \sum_{i \in \hat{\mathcal{S}}^c} \hat{\pi}_i \\
& -\frac{1}{T} \sum_{t=1}^{T} \left(x_{[j]t}'(\beta_0 - \hat{\beta}) + v_{[j]t}\right) - \frac{\lambda}{2T} \hat{\pi}_{[j]} < 0 \,\Bigg|\, \hat{u}_{[j+1]}(\lambda) > 0, ..., \hat{u}_{[J]}(\lambda) > 0\Bigg) \qquad \text{(A.8)}
\end{aligned}
$$

We let $\hat{\mathcal{S}}^* = \mathcal{S}^c \cap \hat{\mathcal{S}}$ and $\hat{\delta}^* = |\hat{\mathcal{S}}^*|/N$. Then, $(*)$ in the j^{th} probability of (A.8) satisfies

$$
u_{0,[j]} - \frac{\sum_{i \in \hat{\mathcal{S}}} u_{0,i}}{\hat{\delta}N} \geq u_{0,[j]} - \frac{\hat{\delta}^* u_{0,[j]}}{\hat{\delta}}
$$

since $u_{0,i} = 0$ for all $i \in \mathcal{S}$ and $u_{0,[j]} = \max_{i \in \hat{\mathcal{S}}^*} u_{0,i}$ in the j^{th} event by construction, which further gives us the results

$$u_{0,[j]}-\frac{\hat{\delta}^{*}}{\hat{\delta}}u_{0,[j]}=\frac{\delta}{\hat{\delta}}u_{0,[j]}\geq\delta u_{0,[j]}\geq\delta\eta \tag{A.9}$$

since $\hat{\delta}-\hat{\delta}^{*}=\delta$ and $\delta\leq\hat{\delta}\leq 1$ as $\mathcal{S}\subset\hat{\mathcal{S}}$.

Let $\hat{\eta}$ $=\min_{i\in\mathcal{S}^c}\hat{u}_i$ and $\breve{\alpha}=\left|\frac{1}{\hat{\delta}NT}\sum_{i\in\hat{\mathcal{S}}}\sum_{t=1}^{T}\left(x_{it}'(\beta_0-\hat{\beta})+v_{it}\right)\right|$. Then, by choosing sufficiently large $0<C<\infty$, we have

$$\begin{aligned}
&\Pr\left(\text{there exists } i\in\mathcal{S}^c \text{ such that } \hat{u}_i(\lambda)=0\right)\\
\leq\ &\Pr\Big(\text{there exists } i\in\mathcal{S}^c \text{ such that } \hat{u}_i(\lambda)=0,\ ||\beta_0-\hat{\beta}||\leq\varkappa_{NT},\ \hat{\eta}\geq\eta-C\varkappa_{NT},\\
&\breve{\alpha}\leq C\varkappa_{NT},\ \mathcal{S}\subset\hat{\mathcal{S}}\Big)+\Pr\left(||\beta_0-\hat{\beta}||>\varkappa_{NT}\right)+\Pr\left(\breve{\alpha}>C\varkappa_{NT}\right)\\
&+\Pr\left(\hat{\eta}<\eta-C\varkappa_{NT}\right)+\Pr\left(\mathcal{S}\not\subset\hat{\mathcal{S}}\right)
\end{aligned} \tag{A.10}$$

where $\Pr\left(||\beta_0-\hat{\beta}||>\varkappa_{NT}\right)=o(1)$ by Assumption 2-(2)-(i), $\Pr\left(\mathcal{S}\not\subset\hat{\mathcal{S}}\right)=o(1)$ by the first part of this proof, $\Pr\left(\breve{\alpha}>C\varkappa_{NT}\right)=o(1)$ by the fact that $\breve{\alpha}\leq\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}x_{it}'(\beta_0-\hat{\beta})+v_{it}\right|$ and (A.1), and $\Pr\left(\hat{\eta}<\eta-C\varkappa_{NT}\right)=o(1)$ by the fact that

$$|\hat{\eta}-\eta|\leq|\hat{\eta}-u_{\ell}|+|\hat{u}_{\ell_0}-\eta| \tag{A.11}$$

and (A.3) where ℓ $=\operatorname{argmin}_{i\in\mathcal{S}^c}\hat{u}_i$ and ℓ_0 $=\operatorname{argmin}_{i\in\mathcal{S}^c}u_{0,i}$.[1] Furthermore, we have

$$\frac{\lambda}{2\hat{\delta}NT}\sum_{i\in\hat{\mathcal{S}}^c}\hat{\pi}_i+\frac{\lambda}{2T}\hat{\pi}_{[j]}\leq\frac{\lambda}{2\hat{\delta}NT}(1-\hat{\delta})N\hat{\eta}^{-\gamma}+\frac{\lambda}{2T}\hat{\eta}^{-\gamma}=\frac{\lambda}{2\hat{\delta}T}\hat{\eta}^{-\gamma}\leq\frac{\lambda}{\delta T}\hat{\eta}^{-\gamma}, \tag{A.12}$$

where we use the fact $\hat{\mathcal{S}}^c\subset\mathcal{S}^c$ and $\delta\leq\hat{\delta}\leq 1$ as $\mathcal{S}\subset\hat{\mathcal{S}}$. Then, for the first term in (A.10),

[1] Note that $|\hat{\eta}-\eta|\leq|\hat{u}_{\ell_0}-\eta|$ if $\hat{\eta}>\eta$ and $|\hat{\eta}-\eta|\leq|\hat{\eta}-u_{\ell}|$ if $\hat{\eta}<\eta$.

by combining (A.8), (A.9) and (A.12), we have

$$
\begin{aligned}
& \Pr\Big(\text{there exists } i \in \mathcal{S}^c \text{ such that } \hat{u}_i(\lambda) = 0,\ ||\beta_0 - \hat{\beta}|| \le \varkappa_{NT},\ \hat{\eta} \ge \eta - C\varkappa_{NT},\ \breve{\alpha} \le C\varkappa_{NT}, \\
& \mathcal{S} \subset \hat{\mathcal{S}}\Big) \\
\le\ & \sum_{j=1}^{J} \Pr\left(\delta\eta - C\varkappa_{NT} - \left|\frac{1}{T}\sum_{t=1}^{T}\left\{x_{[j]t}'(\beta_0 - \hat{\beta}) + v_{[j]t}\right\}\right| - \frac{\lambda}{\delta T}\hat{\eta}^{-\gamma} < 0,\ ||\beta_0 - \hat{\beta}|| \le \varkappa_{NT},\right. \\
& \left.\hat{\eta} \ge \eta - C\varkappa_{NT}\right) \\
\le\ & \sum_{j=1}^{J} \Pr\left(\delta\eta - C\varkappa_{NT} - \left|\frac{1}{T}\sum_{t=1}^{T}\left\{x_{[j]t}'(\beta_0 - \hat{\beta}) + v_{[j]t}\right\}\right| - \frac{\lambda}{\delta T}(\eta - C\varkappa_{NT})^{-\gamma} < 0,\right. \\
& \left.||\beta_0 - \hat{\beta}|| \le \varkappa_{NT}\right) \\
\le\ & \sum_{j=1}^{J} \Pr\left(\delta\eta - C\varkappa_{NT} - \varkappa_{NT}\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\left\{x_{[j]t} - E[x_{[j]t}]\right\}\right\| + E\left\|x_{[j]t}\right\|\right) - \left|\frac{1}{T}\sum_{t=1}^{T} v_{[j]t}\right|\right. \\
& \left.-\frac{\lambda}{\delta T}(\eta - C\varkappa_{NT})^{-\gamma} < 0\right) \\
\le\ & \sum_{j=1}^{J} \Pr\left(\delta\eta - C\varkappa_{NT} - \varkappa_{NT}\left(C\varkappa_{NT} + E\left\|x_{[j]t}\right\|\right) - \left|\frac{1}{T}\sum_{t=1}^{T} v_{[j]t}\right| - \frac{\lambda}{\delta T}(\eta - C\varkappa_{NT})^{-\gamma} < 0\right) \\
& + \sum_{j=1}^{J} \Pr\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\left\{x_{it} - E[x_{it}]\right\}\right\| > C\varkappa_{NT}\right) \\
\le\ & N \max_{1\le i\le N} \Pr\left(\left|\frac{1}{T}\sum_{t=1}^{T} v_{it}\right| > \Re_{NT}\right) + N \max_{1\le i\le N} \Pr\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\left\{x_{it} - E[x_{it}]\right\}\right\| > C\varkappa_{NT}\right) \quad \text{(A.13)}
\end{aligned}
$$

where $\Re_{NT} = \delta\eta - C\varkappa_{NT} - \varkappa_{NT}\left(C\varkappa_{NT} + E||x_{it}||\right) - \frac{\lambda}{\delta T}(\eta - C\varkappa_{NT})^{-\gamma}$. Then we can easily show that the two terms in (A.13) are $o(1)$ by an application of Lemma A.1 and the fact that $\Re_{NT}/\varkappa_{NT} = \frac{\delta\eta}{\varkappa_{NT}} - C - C\varkappa_{NT} - E||x_{it}|| - \frac{\lambda}{\delta T}\eta^{-\gamma}\varkappa_{NT}^{-1}(1 - C\varkappa_{NT}/\eta)^{-\gamma} \to \infty$ as $(N, T) \to \infty$ by Assumption 1 and 2. Thus, the proof is complete. ■

Proof Theorem 2 By Theorem 1, w.p.a 1, we have

$$\sqrt{\delta NT}(\hat{\alpha}(\lambda)-\alpha_0) = \frac{1}{\sqrt{\delta NT}}\sum_{i\in\mathcal{S}}\sum_{t=1}^{T}\left(x_{it}'(\beta_0-\hat{\beta})+v_{it}\right)-\frac{\lambda}{2\sqrt{\delta NT}}\sum_{i\in\mathcal{S}^c}\hat{\pi}_i$$

The second term is $o_p(1)$ since

$$\frac{\lambda}{\sqrt{\delta NT}}\sum_{i\in\mathcal{S}^c}\hat{\pi}_i \leq \sqrt{\frac{(1-\delta)^2}{\delta}}\lambda\sqrt{\frac{N}{T}}\eta^{-\gamma}\left(\frac{\hat{\eta}}{\eta}\right)^{-\gamma}=o_p(1) \tag{A.14}$$

by Assumption 2-(3) and the fact that

$$\frac{\hat{\eta}}{\eta}\leq 1+\frac{|\hat{\eta}-\eta|}{\eta}=1+o_p(1),$$

due to (A.11) and $\varkappa_{NT}/\eta\to 0$ as $(N,T)\to\infty$ by Assumption 2-(2)-(iii).

Since $\hat{\beta}-\beta_0=(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}')^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{v}_{it}$, and $\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{v}_{it}=\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}v_{it}$, we have

$$\begin{aligned}&\sqrt{\delta NT}(\hat{\alpha}(\lambda)-\alpha_0)\\ =\ &\frac{1}{\sqrt{\delta NT}}\sum_{i\in\mathcal{S}}\sum_{t=1}^{T}v_{it}\\ &-\sqrt{\delta}\left(\frac{1}{\delta NT}\sum_{i\in\mathcal{S}}\sum_{t=1}^{T}x_{it}'\right)\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}'\right)^{-1}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}v_{it}\right)+o_p(1).\end{aligned}$$

We define

$$\Upsilon_{\mathcal{S}} = \operatorname*{plim}_{N,T\to\infty}\frac{1}{\delta NT}\sum_{i\in\mathcal{S}}\sum_{t=1}^{T}x_{it}$$

$$H_0 = \operatorname*{plim}_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{it}\tilde{x}_{it}'$$

where $H_0 > 0$ by Assumption 3. We split the sample into $\mathcal{S}$ and $\mathcal{S}^c$ and define two statistics as

$$\begin{aligned}
\Xi_{\mathcal{S},NT} &\equiv \frac{1}{\sqrt{\delta NT}} \sum_{i\in\mathcal{S}} \sum_{t=1}^{T} \left\{ v_{it} - \delta \Upsilon_{\mathcal{S}}' H_0^{-1} \tilde{x}_{it} v_{it} \right\} \\
\Xi_{\mathcal{S}^c,NT} &\equiv \frac{1}{\sqrt{(1-\delta)NT}} \sum_{i\in\mathcal{S}^c} \sum_{t=1}^{T} \sqrt{\delta(1-\delta)} \Upsilon_{\mathcal{S}}' H_0^{-1} \tilde{x}_{it} v_{it},
\end{aligned}$$

which are independent since the observations are cross-sectionally independent. By Assumption 3, we have

$$\begin{aligned}
\Xi_{\mathcal{S},NT} &\xrightarrow{d} \mathcal{N}\left(0, \sigma_{\mathcal{S}_1}^2 + \delta^2 \sigma_{\mathcal{S}_2}^2 - 2\delta \sigma_{\mathcal{S}_1\mathcal{S}_2}\right) \\
\Xi_{\mathcal{S}^c,NT} &\xrightarrow{d} \mathcal{N}\left(0, \delta(1-\delta)\sigma_{\mathcal{S}^c}^2\right)
\end{aligned}$$

as $(N,T) \to \infty$, where

$$\begin{aligned}
\sigma_{\mathcal{S}_1}^2 &= \operatorname*{plim}_{N,T\to\infty} \frac{1}{\delta NT} \sum_{i\in\mathcal{S}} \sum_{t=1}^{T} \sum_{k=1}^{T} v_{it} v_{ik} \\
\sigma_{\mathcal{S}_2}^2 &= \Upsilon_{\mathcal{S}}' H_0^{-1} \left\{ \operatorname*{plim}_{N,T\to\infty} \frac{1}{\delta NT} \sum_{i\in\mathcal{S}} \sum_{t=1}^{T} \sum_{k=1}^{T} \tilde{x}_{it} v_{it} v_{ik} \tilde{x}_{it}' \right\} H_0^{-1} \Upsilon_{\mathcal{S}} \\
\sigma_{\mathcal{S}_1\mathcal{S}_2} &= \Upsilon_{\mathcal{S}}' H_0^{-1} \left\{ \operatorname*{plim}_{N,T\to\infty} \frac{1}{\delta NT} \sum_{i\in\mathcal{S}} \sum_{t=1}^{T} \sum_{k=1}^{T} \tilde{x}_{it} v_{it} v_{ik} \right\} \\
\sigma_{\mathcal{S}^c}^2 &= \Upsilon_{\mathcal{S}}' H_0^{-1} \left\{ \operatorname*{plim}_{N,T\to\infty} \frac{1}{(1-\delta)NT} \sum_{i\in\mathcal{S}^c} \sum_{t=1}^{T} \sum_{k=1}^{T} \tilde{x}_{it} v_{it} v_{ik} \tilde{x}_{it}' \right\} H_0^{-1} \Upsilon_{\mathcal{S}}.
\end{aligned}$$

Hence, $\sqrt{\delta NT}(\hat{\alpha}(\lambda) - \alpha_0) = \Xi_{\mathcal{S},NT} + \Xi_{\mathcal{S}^c,NT} \xrightarrow{d} \mathcal{N}\left(0, \sigma_{\mathcal{S}_1}^2 + \delta^2\sigma_{\mathcal{S}_2}^2 - 2\delta^2\sigma_{\mathcal{S}_1\mathcal{S}_2} + \delta(1-\delta)\sigma_{\mathcal{S}^c}^2\right)$

and the desired result follows.[2]

For the second result, for $i \in \mathcal{S}^c$, we have

$$\begin{aligned}\sqrt{T}(\hat{u}_i(\lambda) - u_{0,i}) &= \sqrt{T}(\hat{\alpha}(\lambda) - \alpha_0) - \frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_{it}'(\beta_0 - \hat{\beta}) - \frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_{it} - \frac{\lambda}{2\sqrt{T}}\hat{\pi}_i \\ &\equiv \Psi_{1,NT} + \Psi_{2i,NT} + \Psi_{3i,T} + \Psi_{4i,NT},\end{aligned}$$

where $\Psi_{1,NT} = O_p(1/\sqrt{\delta N}) = o_p(1)$ from the first result, $\Psi_{2i,NT} = O_p(1/\sqrt{N}) = o_p(1)$ since $\hat{\beta} - \beta_0 = O_p(1/\sqrt{NT})$, and $\Psi_{4i,NT} = o_p(1)$ by a similar argument as in (A.14). Since $\Psi_{3i,T} \xrightarrow{d} \mathcal{N}(0, \sigma_i^2)$ as $T \to \infty$ by Assumption 3, where $\sigma_i^2 = \text{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{T} v_{it}v_{ik}$ for each i, we have the desired result. ■

Proof of Theorem 3 We first define

$$\Lambda_- = \{\lambda : \Pr(\hat{\mathcal{S}}(\lambda) \supsetneq \mathcal{S}) \to 1 \text{ as } (N,T) \to \infty\}$$

$$\Lambda_0 = \{\lambda : \Pr(\hat{\mathcal{S}}(\lambda) = \mathcal{S}) \to 1 \text{ as } (N,T) \to \infty\}$$

$$\Lambda_+ = \{\lambda : \Pr(\hat{\mathcal{S}}(\lambda) \subsetneq \mathcal{S}) \to 1 \text{ as } (N,T) \to \infty\}$$

[2] When v_{it} is conditionally homoskedastic across i, we have $\sigma_{\mathcal{S}_2}^2 = \sigma_{\mathcal{S}^c}^2 = \Upsilon_{\mathcal{S}}' H_0^{-1}\left\{\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{T}\tilde{x}_{it}v_{it}v_{ik}\tilde{x}_{it}'\right\}H_0^{-1}\Upsilon_{\mathcal{S}}$ and the limiting expression simplies to $\mathcal{N}\left(0, \sigma_{\mathcal{S}_1}^2 + \delta\sigma_{\mathcal{S}_2}^2 - 2\delta\sigma_{\mathcal{S}_1\mathcal{S}_2}\right)$.

similarly as Hui, Warton and Foster (2015).[3] We denote the post-LASSO version of $\hat{\theta}(\lambda)$ by $\hat{\theta}_{\hat{\mathcal{S}}(\lambda)}$,[4] the post-LASSO version of $\hat{\sigma}^2(\lambda)$ by $\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda)}$, where

$$\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda)} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - x_{it}^{'}\hat{\beta} - \hat{\theta}_{\hat{\mathcal{S}}(\lambda)}\right)^2,$$

and the post-LASSO BIC by $\overline{\mathrm{BIC}}(\lambda)$,

$$\overline{\mathrm{BIC}}(\lambda) = \log \hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda)} + \frac{\phi_{NT}}{NT}|\hat{\mathcal{S}}^c(\lambda)|.$$

The following lemma shows that asymptotically a λ that yields an over-fitted or under-fitted model can't be selected by $\overline{\mathrm{BIC}}(\lambda)$.

Lemma A.2 *Suppose Assumptions 1 and 2 hold and there exists $\lambda_0 \in \Lambda_0$. Then,*

$$\Pr\left(\inf_{\lambda\in\Lambda_-\cup\Lambda_+} \overline{BIC}(\lambda) > \overline{BIC}(\lambda_0)\right) \to 1 \quad \textit{as } (N,T)\to\infty$$

Proof of Lemma A.2 (i) We first show $\Pr\left(\inf_{\lambda\in\Lambda_-}\overline{\mathrm{BIC}}(\lambda) > \overline{\mathrm{BIC}}(\lambda_0)\right) \to 1$ as $(N,T)\to\infty$. Let $\lambda_- \in \Lambda_-$. Since $\Pr(\hat{\mathcal{S}}(\lambda_-) \supsetneq \mathcal{S}) \to 1$ as $(N,T)\to\infty$, for sufficiently large (N,T), we have

$$\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_-)} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - x_{it}^{'}\hat{\beta} - \hat{\theta}_{\hat{\mathcal{S}}(\lambda_-)}\right)^2$$

[3] Recall Assumption 2-(3): i) $\lambda T^{-1/2}N^{1/2}\eta^{-\gamma} \to 0$; ii) $\lambda T^{(\gamma-1)/2}(\log N)^{-\gamma-1} \to \infty$ for some $\gamma > 1$. Theorem 1 implies that, for $\lambda \in \Lambda_0$, both i) and ii) must be satisfied. For $\lambda \in \Lambda_+$, Assumption i) is satisfied, but not ii), that is, λ is not large enough, so some zero inefficiencies are estimated as nonzero, resulting in over-fitted models. For $\lambda \in \Lambda_-$, ii) is satisfied, but not i), resulting in under-fitted models. In finite samples under-fitted models include the cases where some efficient firms are estimated as inefficient, while some inefficient firms are estimated as efficient. However, Theorem 1 and its proof imply that we can ignore these cases asymptotically.

[4] These post-LASSO version estimates are simply least squares estimates given the estimated set of efficient firms, $\hat{\mathcal{S}(\lambda)}$.

$$
\begin{aligned}
&= \frac{1}{NT}\sum_{t=1}^{T}\sum_{i\in\mathcal{S}}\left(x_{it}'(\beta_0-\hat{\beta})-\left(\hat{\alpha}_{\hat{\mathcal{S}}(\lambda_-)}-\alpha_0\right)+v_{it}\right)^2\\
&\quad+\frac{1}{NT}\sum_{t=1}^{T}\sum_{i\in\hat{\mathcal{S}}^*}\left(x_{it}'(\beta_0-\hat{\beta})-\left(\hat{\alpha}_{\hat{\mathcal{S}}(\lambda_-)}-\alpha_0\right)-u_{0,i}+v_{it}\right)^2\\
&\quad\quad+\frac{1}{NT}\sum_{t=1}^{T}\sum_{i\in\hat{\mathcal{S}}^{**}}\left(x_{it}'(\beta_0-\hat{\beta})-\left(\hat{\alpha}_{\hat{\mathcal{S}}(\lambda_-)}-\alpha_0\right)+\left(\hat{u}_{i,\hat{\mathcal{S}}(\lambda_-)}-u_{0,i}\right)+v_{it}\right)^2
\end{aligned}
$$

where $\hat{\mathcal{S}}^*=\mathcal{S}^c\cap\hat{\mathcal{S}}(\lambda_-)$ and $\hat{\mathcal{S}}^{**}=\mathcal{S}^c\cap\hat{\mathcal{S}}^c(\lambda_-)$. Similarly, for large (N,T),

$$
\begin{aligned}
\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_0)} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it}-x_{it}'\hat{\beta}-\hat{\theta}_{\hat{\mathcal{S}}(\lambda_0)}\right)^2\\
&= \frac{1}{NT}\sum_{t=1}^{T}\sum_{i\in\mathcal{S}}\left(x_{it}'(\beta_0-\hat{\beta})-\left(\hat{\alpha}_{\hat{\mathcal{S}}(\lambda_0)}-\alpha_0\right)+v_{it}\right)^2\\
&\quad+\frac{1}{NT}\sum_{t=1}^{T}\sum_{i\in\hat{\mathcal{S}}^*}\left(x_{it}'(\beta_0-\hat{\beta})-\left(\hat{\alpha}_{\hat{\mathcal{S}}(\lambda_0)}-\alpha_0\right)+\left(\hat{u}_{i,\hat{\mathcal{S}}(\lambda_0)}-u_{0,i}\right)+v_{it}\right)^2\\
&\quad\quad+\frac{1}{NT}\sum_{t=1}^{T}\sum_{i\in\hat{\mathcal{S}}^{**}}\left(x_{it}'(\beta_0-\hat{\beta})-\left(\hat{\alpha}_{\hat{\mathcal{S}}(\lambda_0)}-\alpha_0\right)+\left(\hat{u}_{i,\hat{\mathcal{S}}(\lambda_0)}-u_{0,i}\right)+v_{it}\right)^2
\end{aligned}
$$

Then, for large (N,T), it can be verified that

$$
\begin{aligned}
\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_-)}-\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_0)} &= \delta\left\{\hat{\alpha}_{\hat{\mathcal{S}}(\lambda_-)}-\alpha_0-\frac{1}{\delta NT}\sum_{t=1}^{T}\sum_{i\in\mathcal{S}}\left(x_{it}'(\beta_0-\hat{\beta})+v_{it}\right)\right\}^2\\
&\quad+\frac{1}{N}\sum_{i\in\hat{\mathcal{S}}^*}\left\{\hat{\alpha}_{\hat{\mathcal{S}}(\lambda_-)}-\alpha_0+u_{i,0}-\frac{1}{T}\sum_{t=1}^{T}\left(x_{it}'(\beta_0-\hat{\beta})+v_{it}\right)\right\}^2\\
&> \frac{1}{N}\sum_{i\in\hat{\mathcal{S}}^*}\left\{\hat{\alpha}_{\hat{\mathcal{S}}(\lambda_-)}-\alpha_0+u_{i,0}-\frac{1}{T}\sum_{t=1}^{T}\left(x_{it}'(\beta_0-\hat{\beta})+v_{it}\right)\right\}^2\\
&= \frac{1}{N}\sum_{i\in\hat{\mathcal{S}}^*}\left\{\frac{1}{\hat{\delta}NT}\sum_{t=1}^{T}\sum_{i\in\hat{\mathcal{S}}(\lambda_-)}\left(x_{it}'(\beta_0-\hat{\beta})+v_{it}\right)-\frac{1}{\hat{\delta}N}\sum_{i\in\hat{\mathcal{S}}^*}u_{i,0}\right.
\end{aligned}
$$

$$
+u_{i,0}-\frac{1}{T}\sum_{t=1}^{T}\Big(x_{it}'(\beta_0-\hat{\beta})+v_{it}\Big)\Bigg\}^2
$$

$$
\geq \frac{1}{N}\sum_{i\in\hat{\mathcal{S}}^*}\Bigg\{\underbrace{\left|u_{i,0}-\frac{1}{\hat{\delta}N}\sum_{i\in\hat{\mathcal{S}}^*}u_{i,0}\right|}_{(*)}
$$

$$
-\underbrace{\left|\frac{1}{\hat{\delta}NT}\sum_{t=1}^{T}\sum_{i\in\hat{\mathcal{S}}(\lambda_-)}\Big(x_{it}'(\beta_0-\hat{\beta})+v_{it}\Big)-\frac{1}{T}\sum_{t=1}^{T}\Big(x_{it}'(\beta_0-\hat{\beta})+v_{it}\Big)\right|}_{(**)}\Bigg\}^2
$$

by the reverse triangle inequality and the fact that

$$
\hat{\alpha}_{\hat{\mathcal{S}}(\lambda_-)}-\alpha_0=\frac{1}{\hat{\delta}NT}\sum_{t=1}^{T}\sum_{i\in\hat{\mathcal{S}}(\lambda_-)}\Big(x_{it}'(\beta_0-\hat{\beta})+v_{it}\Big)-\frac{1}{\hat{\delta}N}\sum_{i\in\hat{\mathcal{S}}^*}u_{i,0}
$$

where $\hat{\delta}=\left|\hat{\mathcal{S}}(\lambda_-)\right|/N$. Also note that $(*)$ is $O_p(1)$ or has the rate of η which converges to zero slower than $(**)$.

Therefore, for large (N,T), we have

$$
\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_-)}-\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_0)} > \hat{\delta}^*\left\{\Im-2\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\Big(x_{it}'(\beta_0-\hat{\beta})+v_{it}\Big)\right|\right\}^2 \tag{A.15}
$$

where $\Im=\min_{i\in\hat{\mathcal{S}}^*}\left|u_{i,0}-\frac{1}{\hat{\delta}N}\sum_{i\in\hat{\mathcal{S}}}u_{i,0}\right|$ and $\hat{\delta}^*=\left|\hat{\mathcal{S}}^*\right|/N$.

Finally, note that for any $\lambda_-\in\Lambda_-$,

$$
\overline{\mathrm{BIC}}(\lambda_-)-\overline{\mathrm{BIC}}(\lambda_0) = \log\left\{1+\frac{\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_-)}-\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_0)}}{\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_0)}}\right\}-\frac{\phi_{NT}}{T}\hat{\delta}^*
$$

$$\geq \min\left\{\log 2, \frac{\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_-)} - \hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_0)}}{2\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_0)}}\right\} - \frac{\phi_{NT}}{T}\hat{\delta}^*,$$

and $\log 2 - \frac{\phi_{NT}}{T}\hat{\delta}^* > 0$ as $(N,T) \to 0$ due to the condition that $(\phi_{NT}/T)^{1/2}\eta^{-1} \to 0$. Therefore, to prove $\Pr\left(\inf_{\lambda\in\Lambda_-}\overline{\mathrm{BIC}}(\lambda) > \overline{\mathrm{BIC}}(\lambda_0)\right) \to 1$ as $(N,T) \to \infty$, it suffice to show

$$\inf_{\lambda\in\Lambda_-}\left\{\frac{\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_-)} - \hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_0)}}{2\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_0)}}\right\} - \frac{\phi_{NT}}{T}\hat{\delta}^* \tag{A.16}$$

is positive w.p.a.1 as $(N,T) \to \infty$.

Inequality (A.15) implies that (A.16) is asymptotically greater than

$$\begin{aligned}
&\frac{\hat{\delta}^*}{2}\hat{\sigma}^{-2}_{\hat{\mathcal{S}}(\lambda_0)}\left\{\Im - 2\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^T\left(x'_{it}(\beta_0 - \hat{\beta}) + v_{it}\right)\right|\right\}^2 - \frac{\phi_{NT}}{T}\hat{\delta}^* \\
= \;& \frac{\phi_{NT}}{T}\hat{\delta}^*\left\{\frac{1}{2\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_0)}}\left(\left(\frac{T}{\phi_{NT}}\right)^{1/2}\left[\Im - 2\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^T\left(x'_{it}(\beta_0 - \hat{\beta}) + v_{it}\right)\right|\right]\right)^2 - 1\right\},
\end{aligned}$$

which is asymptotically positive since $\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_0)}$ is bounded, $\Im$ is $O_p(1)$ or $O_p(\eta)$ hence asymptotically dominates $\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^T\left(x'_{it}(\beta_0 - \hat{\beta}) + v_{it}\right)\right| = O_p\left(\frac{\log N}{\sqrt{T}}\right)$ due to Assumption 2-(2)-(iii), and $\left(\frac{T}{\phi_{NT}}\right)^{1/2}\Im \to \infty$ by the condition that $(\phi_{NT}/T)^{1/2}\eta^{-1} \to 0$.

(ii) Next, we show $\Pr\left(\inf_{\lambda\in\Lambda_+}\overline{\mathrm{BIC}}(\lambda) > \overline{\mathrm{BIC}}(\lambda_0)\right) \to 1$ as $(N,T) \to \infty$. Let $\lambda_+ \in \Lambda_+$. Similarly as in (i), for large (N,T), it can be verified that

$$\begin{aligned}
\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_+)} - \hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_0)} \;\geq\; & -\hat{\delta}^{\circ}\left\{\hat{\alpha}_{\hat{\mathcal{S}}(\lambda_0)} - \alpha_0 - \frac{1}{\hat{\delta}^{\circ}NT}\sum_{t=1}^T\sum_{i\in\hat{\mathcal{S}}^{\circ}}\left(x'_{it}(\beta_0 - \hat{\beta}) + v_{it}\right)\right\}^2 \\
& -\hat{\delta}^{\circ\circ}\max_{1\le i\le N}\left\{\left|\hat{\alpha}_{\hat{\mathcal{S}}(\lambda_0)} - \alpha_0\right| + \left|\frac{1}{T}\sum_{t=1}^T\left(x'_{it}(\beta_0 - \hat{\beta}) + v_{it}\right)\right|\right\}^2.
\end{aligned}$$

where $\hat{\delta}^{\circ} = |\hat{\mathcal{S}}^{\circ}|/N$ and $\hat{\delta}^{\circ\circ} = |\hat{\mathcal{S}}^{\circ\circ}|/N$ with $\hat{\mathcal{S}}^{\circ} = \mathcal{S} \cap \hat{\mathcal{S}}(\lambda_+)$ and $\hat{\mathcal{S}}^{\circ\circ} = \mathcal{S} \cap \hat{\mathcal{S}}^c(\lambda_+)$.

Therefore, to show $\Pr\left(\inf_{\lambda \in \Lambda_+} \overline{\text{BIC}}(\lambda) > \overline{\text{BIC}}(\lambda_0)\right) \to 1$ as $(N,T) \to \infty$, it suffices to show

$$\begin{aligned}\overline{\text{BIC}}(\lambda_+) - \overline{\text{BIC}}(\lambda_0) \;\; &\geq \;\; \frac{\phi_{NT}}{T}\hat{\delta}^{\circ\circ} - \frac{\hat{\delta}^{\circ}}{2\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_0)}} \underbrace{\left\{ \hat{\alpha}_{\hat{\mathcal{S}}(\lambda_0)} - \alpha_0 - \frac{1}{\hat{\delta}^{\circ} NT} \sum_{t=1}^{T} \sum_{i \in \hat{\mathcal{S}}^{\circ}} \left(x_{it}'(\beta_0 - \hat{\beta}) + v_{it} \right) \right\}^2}_{(*)} \\ &\quad - \frac{\hat{\delta}^{\circ\circ}}{2\hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_0)}} \underbrace{\max_{1 \leq i \leq N} \left\{ \left| \hat{\alpha}_{\hat{\mathcal{S}}(\lambda_0)} - \alpha_0 \right| + \left| \frac{1}{T} \sum_{t=1}^{T} \left(x_{it}'(\beta_0 - \hat{\beta}) + v_{it} \right) \right| \right\}^2}_{(**)}\end{aligned}$$

is positive w.p.a.1 as $(N,T) \to \infty$, which follows by the condition $\phi_{NT}/(\log N)^2 \to \infty$ since $(**)$ is greater than $(*)$, but $(**) = O_p\left(\frac{(\log N)^2}{T}\right)$ because $|\hat{\alpha}_{\hat{\mathcal{S}}(\lambda_0)} - \alpha_0| = O_p(\frac{1}{\sqrt{\delta NT}})$ due to Theorem 2 and $\max_{1 \leq i \leq N} \left| \frac{1}{T} \sum_{t=1}^{T} \left(x_{it}'(\beta_0 - \hat{\beta}) + v_{it} \right) \right| = O_p\left(\frac{\log N}{\sqrt{T}}\right)$.[5] ■

Next, to link the post-LASSO BIC and LASSO BIC, we show the following:

$$\hat{\sigma}^2(\lambda_0) - \hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_0)} = o_p\left(\frac{1}{NT}\right). \tag{A.17}$$

Due to the shrinkage effect, we have $\hat{\sigma}^2(\lambda_0) - \hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_0)} > 0$, and similarly as in the proof of Lemma A.2 above, we can show that, for large (N,T),

$$\hat{\sigma}^2(\lambda_0) - \hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_0)} \;\; = \;\; \delta \left\{ \frac{\lambda}{2\delta NT} \sum_{i \in \mathcal{S}^c} \hat{\pi}_i \right\}^2 + \frac{1}{N} \sum_{i \in \mathcal{S}^c} \left\{ \frac{\lambda}{2T} \hat{\pi}_i \right\}^2$$

where we use the fact that $\hat{\alpha}(\lambda_0) - \alpha_0 = \frac{1}{\delta NT} \sum_{t=1}^{T} \sum_{i \in \mathcal{S}} \left(x_{it}'(\beta_0 - \hat{\beta}) + v_{it} \right) - \frac{\lambda}{2\delta NT} \sum_{i \in \mathcal{S}^c} \hat{\pi}_i$

[5] Even when $|\hat{\mathcal{S}}^{\circ\circ}|$ is finite so $\hat{\delta}^{\circ\circ} = O\left(\frac{1}{N}\right)$ as $N \to \infty$, we obtain the same conclusion since $\hat{\delta}^{\circ} \to \delta$ in this case, so $(*) = O_p\left(\frac{1}{NT}\right)$.

and $(\hat{\alpha}(\lambda_0) - \alpha_0) - (\hat{u}_i(\lambda_0) - u_{0,i}) = \frac{1}{T}\sum_{t=1}^{T}\left(x_{it}'(\beta_0 - \hat{\beta}) + v_{it}\right) - \frac{\lambda}{2T}\hat{\pi}_i$ for $i \in \mathcal{S}^c$ w.p.a 1 as $(N,T) \to \infty$. Then, using the results in the proof of Theorem 2, we have

$$\hat{\sigma}^2(\lambda_0) - \hat{\sigma}^2_{\hat{\mathcal{S}}(\lambda_0)} \leq \frac{1}{NT}\left\{\sqrt{\frac{(1-\delta)^2}{4\delta}}\lambda\sqrt{\frac{N}{T}}\hat{\eta}^{-\gamma}\right\}^2 + \frac{1-\delta}{NT}\left\{\frac{\lambda}{2}\sqrt{\frac{N}{T}}\hat{\eta}^{-\gamma}\right\}^2 = o_p\left(\frac{1}{NT}\right)$$

since $\lambda\sqrt{\frac{N}{T}}\hat{\eta}^{-\gamma} = o_p(1)$.

Finally, (A.17) and the fact $\mathrm{BIC}(\lambda) > \overline{\mathrm{BIC}}(\lambda)$ for any λ due to shrinkage effect imply

$$\mathrm{BIC}(\lambda) - \mathrm{BIC}(\lambda_0) > \overline{\mathrm{BIC}}(\lambda) - \overline{\mathrm{BIC}}(\lambda_0) + o_p\left(\frac{1}{NT}\right),$$

which gives

$$\Pr\left(\inf_{\lambda\in\Lambda_-\cup\Lambda_+} \mathrm{BIC}(\lambda) > \mathrm{BIC}(\lambda_0)\right) \to 1 \quad \text{as } (N,T) \to \infty.$$

This means that asymptotically a λ which yields an over-fitted or under-fitted model can't be chosen based on the BIC criterion, so the desired result follows. ■

References

Hui, F. K. C., Warton, D. I. and Foster, S. D. (2015), 'Tuning parameter selection for the adaptive lasso using eric', *Journal of the American Statistical Association* **110**(509), 262–269.

Su, L., Shi, Z. and Phillips, P. C. B. (2016), 'Identifying latent structures in panel data', *Econometrica* **84**(6), 2215–2264.

B. Additional Simulations for $\delta \in \{0.1, 0.9\}$

Table B.1: Estimation Accuracy: $\delta = 0.1$

		RMSE		Point estimate ($\alpha_0 = 1$)		Rank correlation	
(N, T)	σ_u	$\hat{U}_{LASSO}$	$\hat{U}_{LSDV}$	$\hat{\alpha}_{LASSO}$	$\hat{\alpha}_{LSDV}$	LASSO	LSDV
(100, 10)	1	0.4537 (0.1765)	0.8630 (0.1820)	1.166 (0.272)	1.761 (0.204)	0.87 (0.041)	0.85 (0.039)
(100, 30)	1	0.2623 (0.0753)	0.4822 (0.1056)	1.059 (0.143)	1.420 (0.121)	0.94 (0.019)	0.93 (0.019)
(100, 50)	1	0.2014 (0.0576)	0.3675 (0.0830)	1.034 (0.108)	1.318 (0.095)	0.96 (0.013)	0.95 (0.014)
(100, 70)	1	0.1733 (0.0481)	0.3089 (0.0700)	1.025 (0.095)	1.266 (0.081)	0.97 (0.011)	0.96 (0.011)
(100, 10)	4	0.4987 (0.1918)	0.7802 (0.1880)	1.225 (0.294)	1.663 (0.217)	0.98 (0.006)	0.98 (0.006)
(100, 30)	4	0.2818 (0.1003)	0.4585 (0.1174)	1.103 (0.168)	1.390 (0.138)	0.99 (0.003)	0.99 (0.003)
(100, 50)	4	0.2136 (0.0711)	0.3528 (0.0914)	1.063 (0.124)	1.297 (0.107)	0.99 (0.002)	0.99 (0.002)
(100, 70)	4	0.1722 (0.0453)	0.2914 (0.0713)	1.041 (0.089)	1.245 (0.084)	1.00 (0.001)	1.00 (0.001)
(200, 10)	1	0.4011 (0.0627)	0.9625 (0.1703)	1.025 (0.153)	1.874 (0.185)	0.89 (0.029)	0.85 (0.026)
(200, 70)	1	0.1661 (0.0191)	0.3502 (0.0675)	0.985 (0.053)	1.313 (0.075)	0.97 (0.008)	0.96 (0.008)
(200, 10)	4	0.4327 (0.0743)	0.8770 (0.1721)	1.122 (0.178)	1.779 (0.193)	0.98 (0.004)	0.98 (0.004)
(200, 70)	4	0.1614 (0.0187)	0.3353 (0.0708)	1.017 (0.057)	1.295 (0.08)	1.00 (0.001)	1.00 (0.001)
(400, 10)	1	0.4168 (0.0399)	1.0597 (0.1713)	0.920 (0.086)	1.981 (0.184)	0.91 (0.021)	0.85 (0.018)
(400, 70)	1	0.1794 (0.0217)	0.3868 (0.0643)	0.952 (0.039)	1.353 (0.070)	0.97 (0.005)	0.96 (0.005)
(400, 10)	4	0.4097 (0.0376)	0.9973 (0.1728)	1.045 (0.123)	1.911 (0.188)	0.98 (0.003)	0.98 (0.003)
(400, 70)	4	0.1577 (0.0103)	0.3799 (0.0668)	0.999 (0.039)	1.346 (0.073)	1.00 (0.001)	1.00 (0.001)
(1000, 10)	1	0.4792 (0.0430)	1.1787 (0.1546)	0.822 (0.057)	2.108 (0.164)	0.93 (0.014)	0.85 (0.011)
(1000, 10)	4	0.4158 (0.0276)	1.1115 (0.1612)	0.970 (0.081)	2.037 (0.171)	0.99 (0.002)	0.98 (0.002)

Table B.2: Estimation Accuracy: $\delta = 0.9$

		RMSE		Point estimate ($\alpha_0 = 1$)		Rank correlation	
(N, T)	σ_u	$\hat{U}_{LASSO}$	$\hat{U}_{LSDV}$	$\hat{\alpha}_{LASSO}$	$\hat{\alpha}_{LSDV}$	LASSO	LSDV
(100, 10)	1	0.2772	1.0699	1.175	1.994	0.84	0.81
		(0.1068)	(0.1713)	(0.144)	(0.184)	(0.133)	(0.151)
(100, 30)	1	0.1415	0.6292	1.057	1.582	0.91	0.89
		(0.0458)	(0.0996)	(0.153)	(0.107)	(0.090)	(0.099)
(100, 50)	1	0.1046	0.4901	1.018	1.455	0.94	0.92
		(0.0368)	(0.0769)	(0.186)	(0.082)	(0.068)	(0.076)
(100, 70)	1	0.0886	0.4137	0.985	1.383	0.95	0.93
		(0.0404)	(0.0640)	(0.228)	(0.069)	(0.053)	(0.056)
(100, 10)	4	0.2744	1.0646	1.174	1.988	0.96	0.96
		(0.1062)	(0.1736)	(0.120)	(0.186)	(0.038)	(0.039)
(100, 30)	4	0.1382	0.6229	1.076	1.577	0.98	0.98
		(0.0461)	(0.0975)	(0.056)	(0.104)	(0.026)	(0.027)
(100, 50)	4	0.0980	0.4889	1.049	1.455	0.98	0.98
		(0.0323)	(0.0753)	(0.039)	(0.080)	(0.018)	(0.019)
(100, 70)	4	0.0799	0.4138	1.037	1.384	0.99	0.99
		(0.0249)	(0.0660)	(0.031)	(0.071)	(0.018)	(0.018)
(200, 10)	1	0.1991	1.1702	1.088	2.099	0.89	0.83
		(0.0439)	(0.1619)	(0.064)	(0.170)	(0.075)	(0.084)
(200, 70)	1	0.0683	0.4496	1.013	1.422	0.96	0.95
		(0.0164)	(0.0598)	(0.067)	(0.063)	(0.028)	(0.032)
(200, 10)	4	0.1992	1.1657	1.091	2.095	0.97	0.97
		(0.0441)	(0.1621)	(0.061)	(0.172)	(0.019)	(0.020)
(200, 70)	4	0.0628	0.4488	1.015	1.420	0.99	0.99
		(0.0117)	(0.0575)	(0.017)	(0.061)	(0.007)	(0.007)
(400, 10)	1	0.1718	1.2552	1.046	2.190	0.91	0.84
		(0.0205)	(0.1504)	(0.036)	(0.158)	(0.048)	(0.058)
(400, 70)	1	0.0656	0.4800	1.008	1.454	0.97	0.96
		(0.0069)	(0.0573)	(0.011)	(0.060)	(0.017)	(0.020)
(400, 10)	4	0.1727	1.2531	1.050	2.187	0.98	0.98
		(0.0221)	(0.1502)	(0.035)	(0.159)	(0.010)	(0.011)
(400, 70)	4	0.0591	0.4802	1.007	1.454	1.00	0.99
		(0.0068)	(0.0567)	(0.010)	(0.060)	(0.003)	(0.003)
(1000, 10)	1	0.1674	1.3605	1.016	2.301	0.93	0.85
		(0.0112)	(0.1436)	(0.021)	(0.150)	(0.026)	(0.035)
(1000, 10)	4	0.1641	1.3736	1.023	2.314	0.99	0.98
		(0.0112)	(0.1461)	(0.019)	(0.152)	(0.005)	(0.005)

Table B.3: Selection Accuracy

	$\sigma_u = 1$				$\sigma_u = 2$				$\sigma_u = 4$			
(N, T)	$P_{\mathcal{S}}$	$P_{\mathcal{S}^c}$	$\hat{\delta}$	Max-miss	$P_{\mathcal{S}}$	$P_{\mathcal{S}^c}$	$\hat{\delta}$	Max-miss	$P_{\mathcal{S}}$	$P_{\mathcal{S}^c}$	$\hat{\delta}$	Max-miss
						$\boldsymbol{\delta} = 0.1$						
(100, 10)	0.6822	0.7729	0.2726	0.7364	0.6438	0.8805	0.1720	0.6366	0.6557	0.9332	0.1257	0.5638
	(0.2362)	(0.1175)	(0.1241)	(0.3008)	(0.2506)	(0.0765)	(0.0885)	(0.3380)	(0.2465)	(0.0447)	(0.0592)	(0.3286)
(100, 30)	0.7612	0.8197	0.2384	0.4694	0.7320	0.9066	0.1572	0.3982	0.7323	0.9509	0.1175	0.3220
	(0.2034)	(0.0916)	(0.0975)	(0.1893)	(0.2275)	(0.0569)	(0.0681)	(0.2026)	(0.2302)	(0.0360)	(0.0495)	(0.2071)
(100, 50)	0.7945	0.8431	0.2207	0.3659	0.7625	0.9191	0.1490	0.3125	0.7761	0.9573	0.1161	0.2546
	(0.1916)	(0.0769)	(0.0821)	(0.1427)	(0.2171)	(0.0497)	(0.0606)	(0.1594)	(0.2083)	(0.0318)	(0.0438)	(0.1708)
(100, 70)	0.8066	0.8585	0.2080	0.3132	0.7962	0.9246	0.1474	0.2625	0.8087	0.9606	0.1163	0.2148
	(0.1879)	(0.0730)	(0.0791)	(0.1298)	(0.1991)	(0.0441)	(0.0534)	(0.1285)	(0.1791)	(0.0284)	(0.0376)	(0.1388)
(200, 10)	0.8186	0.6933	0.3579	0.9995	0.7713	0.8416	0.2197	0.9037	0.7444	0.9191	0.1473	0.7990
	(0.1404)	(0.1013)	(0.1019)	(0.2437)	(0.1648)	(0.0657)	(0.0720)	(0.2924)	(0.1685)	(0.0421)	(0.0509)	(0.3007)
(200, 70)	0.8913	0.8257	0.2460	0.4118	0.8690	0.9117	0.1664	0.3651	0.8630	0.9550	0.1268	0.3110
	(0.0984)	(0.0587)	(0.0595)	(0.1025)	(0.1152)	(0.0376)	(0.0420)	(0.1161)	(0.1217)	(0.0224)	(0.0288)	(0.1256)
(1000, 10)	0.9660	0.5136	0.5344	1.5388	0.9336	0.7472	0.3209	1.4418	0.8972	0.8789	0.1987	1.2904
	(0.0252)	(0.0531)	(0.0495)	(0.2005)	(0.0419)	(0.0418)	(0.0408)	(0.1961)	(0.0567)	(0.0276)	(0.0294)	(0.2220)
						$\boldsymbol{\delta} = 0.9$						
(100, 10)	0.7380	0.7468	0.6895	0.3797	0.7306	0.8555	0.6720	0.2943	0.7482	0.9222	0.6812	0.2093
	(0.1465)	(0.1741)	(0.1420)	(0.2794)	(0.1564)	(0.1270)	(0.1470)	(0.2897)	(0.1515)	(0.0887)	(0.1393)	(0.2878)
(100, 30)	0.8134	0.7934	0.7527	0.2462	0.8140	0.8867	0.7439	0.1759	0.8207	0.9383	0.7448	0.1117
	(0.1188)	(0.1756)	(0.1162)	(0.2235)	(0.1220)	(0.1062)	(0.1135)	(0.1847)	(0.1192)	(0.0770)	(0.1097)	(0.1622)
(100, 50)	0.8483	0.7935	0.7841	0.2218	0.8530	0.9048	0.7773	0.1272	0.8568	0.9467	0.7765	0.0804
	(0.1050)	(0.1933)	(0.1044)	(0.2420)	(0.1037)	(0.0991)	(0.0966)	(0.1475)	(0.1057)	(0.0739)	(0.0970)	(0.1296)
(100, 70)	0.8707	0.8062	0.8030	0.2031	0.8697	0.9092	0.7918	0.1119	0.8739	0.9535	0.7912	0.0637
	(0.0978)	(0.2192)	(0.0999)	(0.2808)	(0.0976)	(0.0975)	(0.0912)	(0.1387)	(0.0947)	(0.0682)	(0.0867)	(0.1075)
(200, 10)	0.8697	0.6501	0.8178	0.6744	0.8703	0.7940	0.8038	0.6103	0.8753	0.8882	0.7990	0.4721
	(0.0771)	(0.1375)	(0.0786)	(0.2677)	(0.0766)	(0.1035)	(0.0744)	(0.3021)	(0.0748)	(0.0754)	(0.0705)	(0.3402)
(200, 70)	0.9404	0.7892	0.8674	0.2829	0.9476	0.8841	0.8644	0.2276	0.9509	0.9378	0.8620	0.1565
	(0.0429)	(0.1144)	(0.0445)	(0.1503)	(0.0409)	(0.0738)	(0.0392)	(0.1442)	(0.0422)	(0.0568)	(0.0396)	(0.1513)
(1000, 10)	0.9675	0.5090	0.9198	1.2077	0.9670	0.7041	0.8999	1.2131	0.9697	0.8341	0.8894	1.1399
	(0.0154)	(0.0675)	(0.0189)	(0.2175)	(0.0154)	(0.0568)	(0.0174)	(0.2426)	(0.0144)	(0.0419)	(0.0152)	(0.2432)