Nets:
Network Estimation for Time Series
Appendix

Matteo Barigozzi[‡] Christian Brownlees[†]

A Proofs of main results

A.1 Notation and definitions

Estimation is conditional on a given value of $\mathbf{c} = (c_{11} \dots c_{nn})'$. We define the dimension of the parameters' space as $m = n^2 p + n(n-1)/2$. We collect the parameters of interest in (8) into the $m \times 1$ vector $\boldsymbol{\theta} = (\boldsymbol{\alpha}' \; \boldsymbol{\rho}')'$, where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}'_{11} \dots \boldsymbol{\alpha}'_{1p} \dots \boldsymbol{\alpha}'_{n1} \dots \boldsymbol{\alpha}'_{np})'$ and $\boldsymbol{\alpha}'_{ik} = (\alpha_{i1k} \dots \alpha_{ink})$ is the i-th row of the VAR matrix $\mathbf{A}_k$ with $k = 1, \dots, p$. The $n(n-1)/2 \times 1$ vector $\boldsymbol{\rho}$ contains the stacked partial correlations of the VAR innovations. Similarly the parameters in (7) are collected into the $m \times 1$ vector $\boldsymbol{\phi} = (\boldsymbol{\beta}' \; \boldsymbol{\rho}')'$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}'_{11} \dots \boldsymbol{\beta}'_{1p} \dots \boldsymbol{\beta}'_{n1} \dots \boldsymbol{\beta}'_{np})'$ and $\boldsymbol{\beta}'_{ik} = (\beta_{i1k} \dots \beta_{ink})$ for $i = 1, \dots, n$ and $k = 1, \dots, p$. Define as $\boldsymbol{\theta}_0$, $\boldsymbol{\phi}_0$, and $\mathbf{c}_0$ the true values of the parameters.

With reference to the minimisation problem in (11)-(12), recall that the adaptive LASSO weights are defined as $w_i = C_\bullet / |\widetilde{\theta}_{T\,i}|$, with $\lambda_T^{\mathbf{G}} = C_\alpha \lambda_T$ and $\lambda_T^{\mathbf{C}} = C_\rho \lambda_T$. Hereafter, for simplicity and without loss of generality we assume that $C_\alpha = C_\rho = 1$. Moreover, we define the sample score and Hessian of the unconstrained problem as

$$\mathbf{S}_T(\boldsymbol{\theta}, \mathbf{c}) = \frac{1}{T} \sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \mathbf{y}_t, \mathbf{c}), \qquad \mathbf{H}_T(\boldsymbol{\theta}, \mathbf{c}) = \frac{1}{T} \sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \mathbf{y}_t, \mathbf{c}).$$

where $\nabla_{\boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}}$ and $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$, and $\ell(\boldsymbol{\theta}, \mathbf{y}_t, \mathbf{c})$ is the unconstrained loss function defined in (9). The population counterparts of the above are defined as $\mathbf{S}_0(\boldsymbol{\theta}, \mathbf{c}) = \mathsf{E}[\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \mathbf{y}_t, \mathbf{c})]$ and $\mathbf{H}_0(\boldsymbol{\theta}, \mathbf{c}) = \mathsf{E}[\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \mathbf{y}_t, \mathbf{c})]$.

For a given symmetric matrix $\mathbf{A}$ we denote by $\mu_{\min}(\mathbf{A})$ and $\mu_{\max}(\mathbf{A})$ its smallest and largest eigenvalues respectively. For a generic matrix $\mathbf{B}$, the notation $\|\mathbf{B}\| = \sqrt{\mu_{\max}(\mathbf{B}\mathbf{B}')}$ is used for the spectral norm. For a generic vector $\mathbf{b}$, the notation $\|\mathbf{b}\| = \sqrt{\sum_i b_i^2}$ indicates the Euclidean norm and $\|\mathbf{b}\|_\infty = \max_i |b_i|$. In what follows we use the symbol K to

[‡] Department of Statistics, London School of Economics and Political Science,
Houghton Street, WC2A 2AE, London, UK, e-mail: m.barigozzi@lse.ac.uk;
[†] Department of Economics and Business, Universitat Pompeu Fabra and Barcelona GSE,
Ramon Trias Fargas 25-27, 08005, Barcelona, Spain, e-mail: christian.brownlees@upf.edu.

denote a generic positive constant. The value of K needs not to be the same from line to line. When more than one distinct constant are present in the same equation we denote them by $K_0, K_1, K_2, \ldots.$ The symbols $\kappa_0, \kappa_1, \kappa_2, \ldots$ denote universal constants that are unique throughout the paper.

A.2 Preliminary results

We start by stating a series of Lemmas and Conditions which are proved in Appendix B.

LEMMA A1. *The parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are related by means of the equations*

$$\beta_{ijk} = \alpha_{ijk} - \sum_{\substack{l=1\\l\neq i}}^{n} \rho^{il}\sqrt{\frac{c_{ll}}{c_{ii}}}\alpha_{ljk}, \qquad \gamma_{ij} = \rho^{ij}\sqrt{\frac{c_{jj}}{c_{ii}}}, \qquad i,j = 1,\ldots,n, \quad k = 1,\ldots,p.$$

Moreover, the error terms in (6) *and* (7) *coincide, that is,* $e_{it} = u_{it}$.

LEMMA A2. *For a given value of* $\mathbf{c}$*, define the* $n^2 \times n^2$ *matrix* $\mathbf{M}(\boldsymbol{\rho};\mathbf{c}) = (\mathrm{diag}\,\mathbf{C})^{-1}\mathbf{C}\otimes\mathbf{I}_n$*. Then,*

$$\underbrace{\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\rho}\end{pmatrix}}_{\boldsymbol{\phi}} = \begin{pmatrix} \mathbf{M}(\boldsymbol{\rho};\mathbf{c}) & \ldots & \mathbf{0} & \mathbf{0}\\ \vdots & \ddots & \vdots & \vdots\\ \mathbf{0} & \ldots & \mathbf{M}(\boldsymbol{\rho};\mathbf{c}) & \mathbf{0}\\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{I}_{n(n-1)/2}\end{pmatrix} \underbrace{\begin{pmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\rho}\end{pmatrix}}_{\boldsymbol{\theta}}.$$

LEMMA A3. *Consider the mapping* $\mathbf{g}_{\mathbf{c}} : \mathbb{R}^m \to \mathbb{R}^m$*, such that* $\mathbf{g}_{\mathbf{c}}(\boldsymbol{\theta}) = \boldsymbol{\phi}$*. Then, under Assumption 1, there exists a function* $\mathbf{h}_{\mathbf{c}_0} : \mathbb{R}^m \to \mathbb{R}^m$ *such that:* $\mathbf{h}_{\mathbf{c}_0}(\mathbf{g}_{\mathbf{c}_0}(\boldsymbol{\theta}_0)) = \boldsymbol{\theta}_0$*, that is* $\mathbf{g}_{\mathbf{c}_0}$ *is invertible in* $\boldsymbol{\theta}_0$*.*

LEMMA A4. *Under Assumption 1, the true value of the parameters is such that*

$$\boldsymbol{\theta}_0 = \arg\min_{\boldsymbol{\theta}} \mathsf{E}[\ell(\boldsymbol{\theta};\mathbf{y}_t,\mathbf{c}_0)].$$

CONDITION 1. *Under Assumption 1, there exist constants* $\underline{L}, \overline{L}$ *such that*

$$0 < \underline{L} \leq \mu_{\min}(\mathbf{H}_0(\boldsymbol{\theta}_0,\mathbf{c}_0)) \leq \mu_{\max}(\mathbf{H}_0(\boldsymbol{\theta}_0,\mathbf{c}_0)) \leq \overline{L} < \infty.$$

CONDITION 2. *Under Assumption 2, for* T *sufficiently large there exist constants* $C_3, C_4 > 0$ *such that, for any* $\eta > 0$*, with probability at least* $1 - O(T^{-\eta})$*, we have*

$$\max_{1\leq i\leq m}\left|\mathbf{S}_{T\,i}(\boldsymbol{\theta}_0,\mathbf{c}_0) - \mathbf{S}_{T\,i}(\boldsymbol{\theta}_0,\widehat{\mathbf{c}}_T)\right| \leq C_3\sqrt{\frac{\log T}{T}},$$
$$\max_{1\leq i,j\leq m}\left|\mathbf{H}_{T\,ij}(\boldsymbol{\theta}_0,\mathbf{c}_0) - \mathbf{H}_{T\,ij}(\boldsymbol{\theta}_0,\widehat{\mathbf{c}}_T)\right| \leq C_4\sqrt{\frac{\log T}{T}}.$$

LEMMA A5. *Under Assumptions 1 and 2 and the same conditions as in Proposition 1, for* T *sufficiently large there exist constants* $\kappa_0, \kappa_1, \kappa_2, \kappa_3 > 0$ *such that, for any* $\eta > 0$ *and any* $\mathbf{u}$

in $\mathbb{R}^{q_T}$*, with probability at least* $1 - O(T^{-\eta})$*, we have*

$$\text{(a)} \quad \left\| \mathbf{S}_{T\,\mathcal{A}}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T) \right\| \leq \kappa_0 \sqrt{q_T} \sqrt{\frac{\log T}{T}};$$

$$\text{(b)} \quad \left| \mathbf{u}' \mathbf{S}_{T\,\mathcal{A}}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T) \right| \leq \kappa_1 \left\| \mathbf{u} \right\| \sqrt{q_T} \sqrt{\frac{\log T}{T}};$$

$$\text{(c)} \quad \left\| \mathbf{H}_{T\,\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T)\mathbf{u} - \mathbf{H}_{0\,\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0)\mathbf{u} \right\| \leq \kappa_2 \left\| \mathbf{u} \right\| q_T \sqrt{\frac{\log T}{T}};$$

$$\text{(d)} \quad \left| \mathbf{u}' \mathbf{H}_{T\,\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T)\mathbf{u} - \mathbf{u}' \mathbf{H}_{0\,\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0)\mathbf{u} \right| \leq \kappa_3 \left\| \mathbf{u} \right\|^2 q_T \sqrt{\frac{\log T}{T}}.$$

Lemma A6. *For any subset* $\mathcal{S} \subseteq \mathcal{A} \cup \mathcal{A}^c$*, we have* $\widehat{\boldsymbol{\theta}}_T^{\mathcal{S}} = \operatorname{argmin}_{\boldsymbol{\theta}:\boldsymbol{\theta}_{\mathcal{S}}^c = \mathbf{0}} \mathcal{L}_T(\boldsymbol{\theta}, \widehat{\mathbf{c}}_T)$*, where* $\mathcal{L}_T(\boldsymbol{\theta}, \widehat{\mathbf{c}}_T)$ *is defined in* (12)*, if and only if the* i*-th component of the sample score satisfies*

$$\begin{aligned} S_{Ti}(\widehat{\boldsymbol{\theta}}_T^{\mathcal{S}}, \widehat{\mathbf{c}}_T) &= -\frac{\lambda_T}{|\widetilde{\theta}_{Ti}|} \operatorname{sign}(\widehat{\theta}_{Ti}^{\mathcal{S}}), & \text{if } \widehat{\theta}_{Ti}^{\mathcal{S}} \neq 0, \\ |S_{Ti}(\widehat{\boldsymbol{\theta}}_T^{\mathcal{S}}, \widehat{\mathbf{c}}_T)| &\leq \frac{\lambda_T}{|\widetilde{\theta}_{Ti}|}, & \text{if } \widehat{\theta}_{Ti}^{\mathcal{S}} = 0. \end{aligned}$$

If the solution is not unique then $|S_{Ti}(\overline{\boldsymbol{\theta}}_T^{\mathcal{S}}, \widehat{\mathbf{c}}_T)| \leq \lambda_T (|\widetilde{\theta}_i|)^{-1}$ *for some specific solution* $\overline{\boldsymbol{\theta}}_T^{\mathcal{S}}$*, then since* $S_{Ti}(\boldsymbol{\theta}, \widehat{\mathbf{c}}_T)$ *is continuous in* $\boldsymbol{\theta}$*, then* $\widehat{\theta}_i = 0$ *for all solutions* $\widehat{\boldsymbol{\theta}}$*. Hence, if* $\mathcal{S} = \mathcal{A} \cup \mathcal{A}^c$*, we have the unconstrained optimisation and* $\widehat{\boldsymbol{\theta}}_T^{\mathcal{S}} = \widehat{\boldsymbol{\theta}}_T$*, while if* $\mathcal{S} = \mathcal{A}$*, we have the restricted optimisation and* $\widehat{\boldsymbol{\theta}}_T^{\mathcal{S}} = \widehat{\boldsymbol{\theta}}_T^{\mathcal{A}}$*.*

Lemma A7. *Under Assumptions 1 and 2 and the same conditions in Proposition 1, there exists a constant* $\kappa_4 > 0$ *such that, for* T *sufficiently large and any* $\eta > 0$*,*

$$\Pr\Big(\exists\, \boldsymbol{\theta}^* = \operatorname*{argmin}_{\boldsymbol{\theta}:\boldsymbol{\theta}_{\mathcal{A}}^c = \mathbf{0}} \mathcal{L}_T(\boldsymbol{\theta}, \widehat{\mathbf{c}}_T) : \boldsymbol{\theta}^* \in D(\boldsymbol{\theta}_0) \Big) \geq 1 - O(T^{-\eta}),$$

where $D(\boldsymbol{\theta}_0) = \{ \boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \kappa_4 \sqrt{q_T} \lambda_T \}$*.*

Lemma A8. *Under Assumptions 1 and 2 and under the same conditions as in Proposition 1, there exists a constant* $\kappa_5 > 0$ *such that, for* T *sufficiently large and any* $\eta > 0$

$$\Pr\left(\|\mathbf{S}_T(\boldsymbol{\theta}, \widehat{\mathbf{c}}_T)\| > \sqrt{q_T} \frac{\lambda_T}{\min_{i \in \mathcal{A}} |\theta_{0i}|} \right) \geq 1 - O(T^{-\eta}),$$

for any $\boldsymbol{\theta} \in S(\boldsymbol{\theta}_0)$ *where* $S(\boldsymbol{\theta}_0) = \{ \boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \geq \kappa_5 \sqrt{q_T} \lambda_T,\ \boldsymbol{\theta}_{\mathcal{A}^c} = \mathbf{0} \}$*.*

Lemma A9. *Under Assumptions 1 and 2 and under the same conditions as in Proposition 1, for* T *sufficiently large and any* $\eta > 0$

$$\Pr\left(\max_{j \in \mathcal{A}^c} |S_{T\,j}(\widehat{\boldsymbol{\theta}}_T^{\mathcal{A}}, \widehat{\mathbf{c}}_T)| \leq \frac{\lambda_T}{\max_{j \in \mathcal{A}^c} |\widetilde{\theta}_{Tj}|} \right) \geq 1 - O(T^{-\eta}).$$

A.3 Consistency of our estimators

To obtain consistency we follow the same strategy as in Meinshausen and Bühlmann (2006), Peng et al. (2009), and Fan and Peng (2004). In order to show Proposition 1 first we prove consistency of the so-called "restricted problem", that is we show consistency of the estimator defined in (11) when restricted to the set of parameters $\boldsymbol{\theta}$ such that $\boldsymbol{\theta}_{\mathcal{A}^c} = \mathbf{0}$, which is denoted as $\widehat{\boldsymbol{\theta}}_T^{\mathcal{A}}$. That is, the estimator of the non-zero parameters obtained when we assume to know those that are zero.

Proposition A1. (estimation consistency). *Suppose that, as* $T \to \infty$, $q_T = o\left(\sqrt{\frac{T}{\log T}}\right)$, $\lambda_T \sqrt{\frac{T}{\log T}} \to \infty$, *and* $\sqrt{q_T}\lambda_T = o(1)$. *Then, under Assumptions 1 and 2, for any* $\eta > 0$, $\widehat{\boldsymbol{\theta}}_T^{\mathcal{A}}$ *exists with probability at least* $1 - O(T^{-\eta})$, *and there exists a constant* $\kappa_R > 0$ *such that*

$$\Pr\left(\left\|\widehat{\boldsymbol{\theta}}_{T\mathcal{A}}^{\mathcal{A}} - \boldsymbol{\theta}_{0\mathcal{A}}\right\| \le \kappa_R \sqrt{q_T}\lambda_T\right) \ge 1 - O(T^{-\eta}).$$

Moreover, if the signal sequence $\{s_T\}$ *is such that,* $\frac{s_T}{\sqrt{q_T}\lambda_T} \to \infty$, *then* $\Pr\big(\mathrm{sign}(\widehat{\theta}_{T\,i}^{\mathcal{A}}) = \mathrm{sign}(\theta_{0\,i})\big) \ge 1 - O(T^{-\eta})$, *for any* $i \in \mathcal{A}$.

Proof of proposition A1. From Lemma A6, we have

$$\|\mathbf{S}_{T\mathcal{A}}(\widehat{\boldsymbol{\theta}}_T^{\mathcal{A}})\|_\infty \le \lambda_T \max_{i\in\mathcal{A}} \frac{1}{|\widetilde{\theta}_{Ti}|}.$$

Moreover, for any $i \in \mathcal{A}$,

$$\frac{1}{|\widetilde{\theta}_{Ti}|} = \sqrt{\frac{1}{\widetilde{\theta}_{Ti}^2}} \le \frac{1}{|\theta_{0i}|} + \sqrt{\frac{2}{\theta_{0i}^3}|\widetilde{\theta}_{Ti} - \theta_{0i}| + o(|\widetilde{\theta}_{Ti} - \theta_{0i}|)}. \tag{A-1}$$

Define $\theta_0^* = \min_{i\in\mathcal{A}} |\theta_{0i}|$ and notice that $\theta_0^* > 0$ and define also $\nu_T = \sqrt{q_T}\lambda_T$, therefore $\nu_T \to 0$ as $T \to \infty$. Using Assumption 2 and (A-1), there exists a constant $K > 0$ such that for T sufficiently large and for any $\eta > 0$, we have with probability at least $1 - O(T^{-\eta})$

$$\begin{aligned}
\|\mathbf{S}_{T\mathcal{A}}(\widehat{\boldsymbol{\theta}}_T^{\mathcal{A}})\| &\le \sqrt{q_T}\ \|\mathbf{S}_{T\mathcal{A}}(\widehat{\boldsymbol{\theta}}_T^{\mathcal{A}})\|_\infty \le \nu_T \max_{i\in\mathcal{A}} \frac{1}{|\widetilde{\theta}_{Ti}|} \\
&\le \nu_T \left[\max_{i\in\mathcal{A}} \frac{1}{|\theta_{0i}|} + K\left(q_T \frac{\log T}{T}\right)^{1/4}\right] \\
&\le \frac{\nu_T}{\theta_0^*} + \nu_T K \left(q_T \frac{\log T}{T}\right)^{1/4}.
\end{aligned} \tag{A-2}$$

Notice that the last term on the rhs of (A-2) is $o(\nu_T)$, thus it can be neglected so that for T sufficiently large and for any $\eta > 0$, we have

$$\Pr\left(\|\mathbf{S}_{T\mathcal{A}}(\widehat{\boldsymbol{\theta}}_T^{\mathcal{A}})\| \le \frac{\nu_T}{\theta_0^*}\right) \ge 1 - O(T^{-\eta}). \tag{A-3}$$

From Lemma A8 we also have that for T sufficiently large and for any $\eta > 0$

$$\Pr\left(\|\mathbf{S}_{T\,\mathcal{A}}(\boldsymbol{\theta})\| \leq \frac{\nu_T}{\theta_0^*}\right) \geq 1 - O(T^{-\eta}). \tag{A-4}$$

for any $\boldsymbol{\theta}$ such that $\boldsymbol{\theta}_{\mathcal{A}^c} = \mathbf{0}$ and $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \kappa_5 \nu_T$. Therefore, (A-4) implies that inside a disc of radius $\kappa_5\nu_T$ condition (A-3) is satisfied. In particular, (A-3) is a consequence of the Karush-Kuhn-Tucker condition in Lemma A6 for $\widehat{\boldsymbol{\theta}}_T^{\mathcal{A}}$ to be a minimum. Moreover, by Lemma A7, such minimum always exists in a disc of radius $\kappa_4\nu_T$. Hence, if we define $\kappa_R = \min(\kappa_4, \kappa_5)$, for T sufficiently large and for any $\eta > 0$, we have

$$\Pr\left(\|\widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}^{\mathcal{A}} - \boldsymbol{\theta}_{0\,\mathcal{A}}\| \leq \kappa_R \nu_T\right) \geq 1 - O(T^{-\eta}). \tag{A-5}$$

Finally, for any $i \in \mathcal{A}$ and for T sufficiently large, we have $|\theta_{0i}| > s_T > 2\kappa_R\nu_T$. Moreover, for any $i \in \mathcal{A}$, in general we have

$$\Pr\left(\text{sign}(\widehat{\theta}_{T\,i}^{\mathcal{A}}) = \text{sign}(\theta_{0\,i})\right) \geq \Pr\left(\|\widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}^{\mathcal{A}} - \boldsymbol{\theta}_{0\,\mathcal{A}}\| \leq \kappa_R\nu_T, |\theta_{0i}| > 2\kappa_R\nu_T\right), \tag{A-6}$$

which by (A-5) implies sign consistency. This completes the proof. □

Proof of proposition 1. (a) By Proposition A1 and Lemma A8 the non-zero coefficients of $\widehat{\boldsymbol{\theta}}_T^{\mathcal{A}}$ satisfy the Karush-Kuhn-Tucker condition in Lemma A6. Moreover, by Lemma A9 for T sufficiently large and for any $\eta > 0$ also the zero coefficients satisfy the Karush-Kuhn-Tucker condition with probability at least $1 - O(T^{-\eta})$. Therefore, since with probability at least $1 - O(T^{-\eta})$ the restricted estimator $\widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}^{\mathcal{A}}$ is also a solution of the unrestricted problem, we proved the existence of a solution of the unrestricted problem. On the other hand, by Lemma A9 and the Karush-Kuhn-Tucker condition in Lemma A6, with probability at least $1 - O(T^{-\eta})$, any solution of the unrestricted problem is a solution of the restricted problem. That is,

$$\Pr\left(\widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}^{\mathcal{A}} = \widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}\right) \geq 1 - O(T^{-\eta}). \tag{A-7}$$

As a consequence of (A-7), given the unrestricted estimator, $\widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}$, for T sufficiently large, for any $\eta > 0$ and for all $j \in \mathcal{A}^c$ we have

$$\begin{aligned}
\Pr\left(\widehat{\theta}_{T\,j} = 0\right) &= \Pr\left(\widehat{\theta}_{T\,j} = 0 \,\middle|\, \widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}^{\mathcal{A}} = \widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}\right) \Pr\left(\widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}^{\mathcal{A}} = \widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}\right) \\
&\quad + \Pr\left(\widehat{\theta}_{T\,j} = 0 \,\middle|\, \widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}^{\mathcal{A}} \neq \widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}\right) \Pr\left(\widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}^{\mathcal{A}} \neq \widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}\right) \\
&\geq \Pr\left(\widehat{\theta}_{T\,j} = 0 \,\middle|\, \widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}^{\mathcal{A}} = \widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}\right) \Pr\left(\widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}^{\mathcal{A}} = \widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}\right) = \Pr\left(\widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}^{\mathcal{A}} = \widehat{\boldsymbol{\theta}}_{T\,\mathcal{A}}\right) \geq 1 - O(T^{-\eta}).
\end{aligned}$$

This proves part (a). Part (b) follows directly from Proposition A1 and (A-7). This completes the proof. □

B Proof of Lemmas A1-A9 and Conditions 1-2

Proof of lemma A1. The var(p) model (1) has n equations given by

$$y_{i\,t} = \sum_{k=1}^{p}\sum_{j=1}^{n} \alpha_{ijk}\, y_{j\,t-k} + \epsilon_{i\,t}, \quad i = 1,\ldots,n, \tag{B-1}$$

where ϵ_{it} is the i-the element of the vector $\boldsymbol{\epsilon}_t$. Then, by substituting (B-1) in (7), we have, for any $i = 1,\ldots,n$,

$$\begin{aligned}
y_{i\,t} &= \sum_{k=1}^{p}\sum_{j=1}^{n} \beta_{ijk}\, y_{j\,t-k} + \sum_{\substack{h=1\\h\neq i}}^{n} \gamma_{ih}\, y_{h\,t} + e_{i\,t}\\
&= \sum_{k=1}^{p}\sum_{j=1}^{n} \beta_{ijk}\, y_{j\,t-k} + \sum_{\substack{h=1\\h\neq i}}^{n} \gamma_{ih} \left(\sum_{k=1}^{p}\sum_{j=1}^{n} \alpha_{hjk}\, y_{j\,t-k} + \epsilon_{h\,t}\right) + e_{i\,t}\\
&= \sum_{k=1}^{p}\sum_{j=1}^{n} \left(\beta_{ijk} + \sum_{\substack{h=1\\h\neq i}}^{n} \gamma_{ih}\alpha_{hjk}\right) y_{j\,t-k} + \sum_{\substack{h=1\\h\neq i}}^{n} \gamma_{ih}\, \epsilon_{h\,t} + e_{i\,t}.
\end{aligned} \tag{B-2}$$

By comparing the rhs of (B-2) with (B-1) we have

$$\alpha_{ijk} = \beta_{ijk} + \sum_{\substack{h=1\\h\neq i}}^{n} \gamma_{ih}\alpha_{hjk}, \quad i,j = 1,\ldots,n, \quad k = 1,\ldots,p, \tag{B-3}$$

$$\epsilon_{i\,t} = \sum_{\substack{h=1\\h\neq i}}^{n} \gamma_{ih}\, \epsilon_{h\,t} + e_{i\,t}, \quad i = 1,\ldots,n. \tag{B-4}$$

and therefore, from (6) we also have $e_{i\,t} = u_{i\,t}$. From (B-4) using Lemma 3 in Peng et al. (2009) we have

$$\gamma_{ih} = \rho^{ih}\sqrt{\frac{c_{hh}}{c_{ii}}}, \quad i,h = 1,\ldots,n, \tag{B-5}$$

and clearly when $i = h$, $\gamma_{ih} = \rho^{ih} = 1$. By substituting (B-5) into (B-3) we complete the proof. □

Proof of lemma A2. First define the $n \times n$ matrix $\mathbf{R} = \mathbf{I}_n - (\text{diag}\ \mathbf{C})^{-1/2}\mathbf{C}\ (\text{diag}\ \mathbf{C})^{-1/2}$. Form the definition of partial correlation (4), we see that $\mathbf{R}$ is a matrix with ρ^{ij} as generic (i,j) entry whenever $i \neq j$ and zero otherwise. Now from Lemma A1 we immediately have that, for any $k = 1, \ldots, p$

$$\begin{pmatrix} \boldsymbol{\beta}_{1k} \\ \vdots \\ \boldsymbol{\beta}_{nk} \end{pmatrix} = \left\{ \mathbf{I}_{n^2} - \left[(\text{diag}\,\mathbf{C})^{-1/2} \mathbf{R} (\text{diag}\,\mathbf{C})^{1/2} \otimes \mathbf{I}_n \right] \right\} \begin{pmatrix} \boldsymbol{\alpha}_{1k} \\ \vdots \\ \boldsymbol{\alpha}_{nk} \end{pmatrix}$$

$$= \left\{ \mathbf{I}_{n^2} - \left[\left(\mathbf{I}_n - (\text{diag}\,\mathbf{C})^{-1} \mathbf{C} \right) \otimes \mathbf{I}_n \right] \right\} \begin{pmatrix} \boldsymbol{\alpha}_{1k} \\ \vdots \\ \boldsymbol{\alpha}_{nk} \end{pmatrix} = \mathbf{M}(\boldsymbol{\rho}; \mathbf{c}) \begin{pmatrix} \boldsymbol{\alpha}_{1k} \\ \vdots \\ \boldsymbol{\alpha}_{nk} \end{pmatrix}. \quad \text{(B-6)}$$

The statement of the lemma follow straightforwardly from (B-6). □

Proof of lemma A3. From Lemma A2 we have

$$\mathbf{g}_{\mathbf{c}_0}(\boldsymbol{\theta}_0) = \boldsymbol{\phi}_0 = \begin{pmatrix} \mathbf{M}(\boldsymbol{\rho}_0; \mathbf{c}_0) & \ldots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \ldots & \mathbf{M}(\boldsymbol{\rho}_0; \mathbf{c}_0) & \mathbf{0} \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{I}_{n(n-1)/2} \end{pmatrix} \boldsymbol{\theta}_0.$$

Then consider the Jacobian $\nabla_{\boldsymbol{\theta}} \mathbf{g}_{\mathbf{c}_0}(\boldsymbol{\theta}_0)$ which has (i,j)-th entry $\partial g_{\mathbf{c}_0,i}(\boldsymbol{\theta}_0)/\partial\theta_j = \partial\phi_i/\partial\theta_j$:

$$\nabla_{\boldsymbol{\theta}} \mathbf{g}_{\mathbf{c}_0}(\boldsymbol{\theta}_0) = \begin{pmatrix} \mathbf{M}(\boldsymbol{\rho}_0; \mathbf{c}_0) & \ldots & \mathbf{0} & \nabla_{\boldsymbol{\rho}} \mathbf{M}(\boldsymbol{\rho}_0; \mathbf{c}_0)(\boldsymbol{\alpha}_{0,11} \ldots \boldsymbol{\alpha}_{0,n1})' \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \ldots & \mathbf{M}(\boldsymbol{\rho}_0; \mathbf{c}_0) & \nabla_{\boldsymbol{\rho}} \mathbf{M}(\boldsymbol{\rho}_0; \mathbf{c}_0)(\boldsymbol{\alpha}_{0,1p} \ldots \boldsymbol{\alpha}_{0,np})' \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{I}_{n(n-1)/2} \end{pmatrix}. \quad \text{(B-7)}$$

Since $\mathbf{M}(\boldsymbol{\rho}_0; \mathbf{c}_0)$ is positive definite because of Assumption 1, the Jacobian in $\boldsymbol{\theta}_0$ is positive definite too and the mapping $\mathbf{g}_{\mathbf{c}_0}$ is invertible in $\boldsymbol{\theta}_0$ and this completes the proof. □

Proof of Lemma A4. Notice that the loss related to (7) is given by

$$\ell(\boldsymbol{\phi}_0; \mathbf{y}_t, \mathbf{c}_0) = \sum_{i=1}^{n} \left(y_{i\,t} - \sum_{k=1}^{p} \sum_{j=1}^{n} \beta_{ijk}\, y_{j\,t-k} - \sum_{\substack{h=1 \\ h \neq i}}^{n} \rho^{ih} \sqrt{\frac{c_{0,hh}}{c_{0,ii}}} y_{h\,t} \right)^2. \quad \text{(B-8)}$$

Clearly $\boldsymbol{\phi}_0$ is a minimizer of (B-8) (using Assumption 1 for second order conditions):

$$\boldsymbol{\phi}_0 = \arg\min_{\boldsymbol{\phi}} \mathsf{E}[\ell(\boldsymbol{\phi}; \mathbf{y}_t, \mathbf{c}_0)]. \quad \text{(B-9)}$$

In order for $\boldsymbol{\theta}_0$ to be a minimum, we need to verify that first and second order conditions hold. The first order conditions are given by[1]

$$\mathsf{E}[\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}_0; \mathbf{y}_t, \mathbf{c}_0)] = \mathsf{E}[\nabla_{\boldsymbol{\phi}} \ell(\boldsymbol{\phi}_0; \mathbf{y}_t, \mathbf{c}_0) \nabla_{\boldsymbol{\theta}} \mathbf{g}_{\mathbf{c}_0}(\boldsymbol{\theta}_0)] = \mathsf{E}[\nabla_{\boldsymbol{\phi}} \ell(\boldsymbol{\phi}_0; \mathbf{y}_t, \mathbf{c}_0)] \nabla_{\boldsymbol{\theta}} \mathbf{g}_{\mathbf{c}_0}(\boldsymbol{\theta}_0) = \mathbf{0}, \quad \text{(B-10)}$$

[1]Notice that we can exchange integral and differentiation operators as the loss function is such that $\ell \in \mathcal{C}^\infty(\mathbb{R}^m)$.

since $\mathsf{E}[\nabla_{\boldsymbol{\phi}}\ell(\boldsymbol{\phi}_0;\mathbf{y}_t,\mathbf{c}_0)] = \mathbf{0}$ because of (B-9). The second order conditions are

$$\begin{aligned}\mathsf{E}[\nabla_{\boldsymbol{\theta\theta}}\ell(\boldsymbol{\theta}_0;\mathbf{y}_t,\mathbf{c}_0)] &= \mathsf{E}[\nabla_{\boldsymbol{\theta\phi}}\ell(\boldsymbol{\phi}_0;\mathbf{y}_t,\mathbf{c}_0)]\nabla_{\boldsymbol{\theta}}\mathbf{g}_{\mathbf{c}_0}(\boldsymbol{\theta}_0) + \mathsf{E}[\nabla_{\boldsymbol{\phi}}\ell(\boldsymbol{\phi}_0;\mathbf{y}_t,\mathbf{c}_0)]\nabla_{\boldsymbol{\theta\theta}}\mathbf{g}_{\mathbf{c}_0}(\boldsymbol{\theta}_0)\\ &= \mathsf{E}[\nabla_{\boldsymbol{\phi\phi}}\ell(\boldsymbol{\phi}_0;\mathbf{y}_t,\mathbf{c}_0)]\left(\nabla_{\boldsymbol{\theta}}\mathbf{g}_{\mathbf{c}_0}(\boldsymbol{\theta}_0)\right)^2, \end{aligned} \qquad \text{(B-11)}$$

which we used (B-10). Now, (B-11) is positive definite since the first term is positive definite because of (B-9) and the second term is positive definite because of Lemma A3. □

Proof of condition 1. The inequality on the lhs is proved in the proof of Lemma A4, while the inequality on the rhs is proved in condition B1 in the supplementary appendix of Peng et al. (2009). □

Proof of condition 2. This is an immediate consequence of consistency of the pre-estimator $\widehat{\mathbf{c}}_T$ given in Assumption 2 and the continuous mapping theorem. □

Proof of lemma A5. (a) We begin by noting that the sample averages of the partial derivatives of ℓ in $(\boldsymbol{\theta}_0', \boldsymbol{c}_0')'$ satisfy a Bernstein–type exponential inequality. The partial derivatives of ℓ are

$$\begin{aligned}\frac{\partial\ell(\boldsymbol{\theta}_0,\mathbf{y}_t,\mathbf{c}_0)}{\partial\alpha_{0\,ijk}} &= -2u_{i\,t}y_{j\,t-k} + \sum_{\substack{l=1\\l\neq i}}^{n} 2\rho^{il}\sqrt{\frac{c_{0\,ii}}{c_{0\,ll}}}u_{l\,t}y_{j\,t-k},\\ \frac{\partial\ell(\boldsymbol{\theta}_0,\mathbf{y}_t,\mathbf{c}_0)}{\partial\rho_0^{ij}} &= -2\sqrt{\frac{c_{0\,ii}}{c_{0\,jj}}}u_{i\,t}\epsilon_{j\,t} - 2\sqrt{\frac{c_{0\,jj}}{c_{0\,ii}}}u_{j\,t}\epsilon_{i\,t}.\end{aligned}$$

We only show this for the partial derivatives with respect to the $\boldsymbol{\alpha}$ coefficients. The proof for the partial derivatives of the $\boldsymbol{\rho}$ coefficients follows analogous steps. In particular, we that show that the averages of the partial derivatives of the the $\boldsymbol{\alpha}$ coefficients satisfy an exponential inequality that does not depend on n. From (6) we have $\mathsf{Var}(u_{l\,t}) \leq \mathsf{Var}(\epsilon_{l\,t})$ for any $l = 1,\dots,n$, therefore, there exists a constant $K > 0$ such that

$$\mathsf{Var}\left(\sum_{\substack{l=1\\l\neq i}}^{n}\rho_0^{il}\sqrt{\frac{c_{0\,ii}}{c_{0\,ll}}}u_{l\,t}\right) \leq \mathsf{Var}\left(\sum_{\substack{l=1\\l\neq i}}^{n}\rho_0^{il}\sqrt{\frac{c_{0\,ii}}{c_{0\,ll}}}\epsilon_{l\,t}\right) = \mathsf{Var}(\epsilon_{i\,t}) \leq K, \qquad \text{(B-12)}$$

where the last equality is given in (6). Define

$$A_{T\,ijk} = -\frac{2}{T}\sum_{t=1}^{T}u_{i\,t}y_{j\,t-k}, \qquad B_{T\,ijk} = \frac{2}{T}\sum_{t=1}^{T}\left(\sum_{\substack{l=1\\l\neq i}}^{n}\rho^{il}\sqrt{\frac{c_{0\,ii}}{c_{0\,ll}}}u_{l\,t}y_{j\,t-k}\right).$$

By Assumption 1, $y_{i\,t}$ is a zero–mean strongly mixing process with moments satisfying the Cramér condition

$$\mathsf{E}[|y_{i\,t}|^k] \leq k!c^{k-2}\mathsf{E}[y_{i\,t}^2] < \infty, \qquad k = 3,4,\dots.$$

Thus, $|T^{-1}\sum_{t=1}^{T}y_{i\,t}|$ satisfy the Bernstein-type exponential inequality in Theorem 1 by Doukhan and Neumann (2007) (see also Theorem 1.4 in Bosq, 1996), i.e. for any i there exists a constant $K_0 > 0$ such that, for any $\varepsilon > 0$,

$$\Pr\left(\left|\frac{1}{T}\sum_{t=1}^{T} y_{it}\right| > \varepsilon\right) \leq \exp\left\{-K_0 T\varepsilon^2\right\}. \tag{B-13}$$

Since u_{it} is i.i.d. by construction, then as a consequence of (B-13) and Remark 2.2 in Dedecker et al. (2007), there exists also a constant $K_1 > 0$ such that

$$\Pr\left(|A_{T\,ijk}| > \varepsilon\right) \leq \exp\left\{-K_1 T\varepsilon^2\right\}. \tag{B-14}$$

Moreover, by Assumption 1, for any $i = 1, \ldots, n$ we have $0 < c_{0\,ii} < \infty$ and therefore because of (B-12) each term in parenthesis in $B_{T\,ijk}$ has finite variance and zero mean, therefore, using arguments analogous to those used for (B-14), there exists also a constant $K_2 > 0$ such that

$$\Pr\left(|B_{T\,ijk}| > \varepsilon\right) \leq \exp\left\{-K_2 T\varepsilon^2\right\}.$$

Therefore, there exists a constant $K_3 > 0$ such that

$$\begin{aligned}\Pr\left(\left|\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \ell(\boldsymbol{\theta}_0, \mathbf{y}_t, \mathbf{c}_0)}{\partial \alpha_{0\,ijk}}\right|^2 > \varepsilon^2\right) &= \Pr\left(|A_{T\,ijk} + B_{T\,ijk}| > \varepsilon\right)\\ &\leq \Pr\left(|A_{T\,ijk}| + |B_{T\,ijk}| > \varepsilon\right) \leq 2\,\exp\left\{-K_3 T\varepsilon^2\right\}.\end{aligned}$$

Note that as a consequence there exist a constant $K_4 > 0$ such that

$$\begin{aligned}\Pr(\|\mathbf{S}_{T\,\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0)\| > \varepsilon) &= \Pr(\|\mathbf{S}_{T\,\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0)\|^2 > \varepsilon^2) = \Pr\left(\sum_{i=1}^{q_T}|S_{T\,\mathcal{A}\,i}|^2 > \varepsilon^2\right),\\ &\leq q_T \Pr\left(|S_{T\,\mathcal{A}\,i}|^2 > \frac{\varepsilon^2}{q_T}\right) = q_T \Pr\left(|S_{T\,\mathcal{A}\,i}| > \frac{\varepsilon}{\sqrt{q_T}}\right),\\ &\leq 2q_T \exp\left\{-K_4 T\frac{\varepsilon^2}{q_T}\right\}.\end{aligned}$$

By setting the rhs of the last expression equal to $\delta = O(T^{-\eta})$ for $\eta > 0$ and solving with respect to ε we get that for T sufficiently large there exist a constant κ_0 such that

$$\varepsilon \leq \kappa_0 \sqrt{q_T}\sqrt{\frac{\log T}{T}}. \tag{B-15}$$

Then, for T sufficiently large there exist a constant κ_0 such that

$$\|\mathbf{S}_{T\,\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0)\| < \kappa_0 \sqrt{q_T}\sqrt{\frac{\log T}{T}},$$

with at least probability $1 - O(T^{-\eta})$. Moreover, we have

$$\|\mathbf{S}_{T\,\mathcal{A}}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T)\| \leq \|\mathbf{S}_{T\,\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0)\| + \|\mathbf{S}_{T\,\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0) - \mathbf{S}_{T\,\mathcal{A}}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T)\| \tag{B-16}$$

and for T sufficiently large the second term is $O(\sqrt{(q_T \log T)\, T^{-1}}\,) = o(1)$ by Condition 2. Part (a) follows by combining (B-15) and (B-16). Part (b) follows from (a) and the Cauchy-Schwarz inequality.

(c) We begin by noting that

$$\begin{aligned}
\|\mathbf{H}_{T\,\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0)\mathbf{u} - \mathbf{H}_{0\,\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0)\mathbf{u}\|^2 &\leq 2\|\mathbf{u}\|^2 \|\mathbf{H}_{T\,\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0) - \mathbf{H}_{0\,\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0)\|^2, \\
&\leq 2\|\mathbf{u}\|^2 \sum_{i=1}^{q_T}\sum_{j=1}^{q_T} \left[\mathbf{H}_{T\,\mathcal{A}\mathcal{A}\,ij}(\boldsymbol{\theta}_0, \mathbf{c}_0) - \mathbf{H}_{0\,\mathcal{A}\mathcal{A}\,ij}(\boldsymbol{\theta}_0, \mathbf{c}_0)\right]^2.
\end{aligned}$$

Next, we focus on showing that the differences

$$A_{T\,ij} = \mathbf{H}_{T\,\mathcal{A}\mathcal{A}\,ij}(\boldsymbol{\theta}_0, \mathbf{c}_0) - \mathbf{H}_{0\,\mathcal{A}\mathcal{A}\,ij}(\boldsymbol{\theta}_0, \mathbf{c}_0)$$

satisfy an appropriate Bernstein–type exponential inequality. We begin by noting that the first $n^2p \times n^2p$ diagonal block of the Hessian has entries

$$\frac{\partial^2 \ell(\boldsymbol{\theta}_0, \mathbf{y}_t, \mathbf{c}_0)}{\partial \alpha_{ij'k'} \partial \alpha_{ijk}} = 2 y_{j\,t-k} y_{j'\,t-k'} \left(1 + \sum_{\substack{l=1\\ l\neq i}}^{n} \rho_0^{il} \sqrt{\frac{c_{0\,ll}}{c_{0\,ii}}} \left(\rho_0^{li}\sqrt{\frac{c_{0\,ii}}{c_{0\,ll}}} - 1\right)\right)$$

for any $j, j' = 1 \ldots n$ and any $k, k' = 1 \ldots p$ and

$$\frac{\partial^2 \ell(\boldsymbol{\theta}_0, \mathbf{y}_t, \mathbf{c}_0)}{\partial \alpha_{i'j'k'} \partial \alpha_{ijk}} = 2 y_{j\,t-k} y_{j'\,t-k'} \left(\left(1 - \rho_0^{ii'}\sqrt{\frac{c_{0\,i'i'}}{c_{0\,ii}}}\right) + \sum_{\substack{l=1\\ l\neq i}}^{n} \rho_0^{il} \sqrt{\frac{c_{0\,ll}}{c_{0\,ii}}} \left(\rho_0^{li}\sqrt{\frac{c_{0\,ii}}{c_{0\,ll}}} - 1\right)\right),$$

for $i \neq i'$ and any $j, j' = 1 \ldots n$ and any $k, k' = 1 \ldots p$. The second $n(n-1)/2 \times n(n-1)/2$ diagonal block has entries

$$\frac{\partial^2 \ell(\boldsymbol{\theta}_0, \mathbf{y}_t, \mathbf{c}_0)}{\partial \rho^{ij'} \partial \rho^{ij}} = 2\sqrt{\frac{c_{0\,jj} c_{0\,ii}}{c_{0\,ii} c_{0\,j'j'}}} \epsilon_{j'\,t}\epsilon_{j\,t},$$

for any $i, j, j' = 1 \ldots n$ with $i \neq j$, $i \neq j'$ and $j \neq j'$. It is straightforward to check that the averages of the partial derivatives with respect to the $\boldsymbol{\rho}$ coefficients satisfy a Bernstein–type inequality. As far as the partial derivatives with respect to the $\boldsymbol{\alpha}$ coefficients we need to show that this term does not grow with n. Notice that by Assumption 1 and from (6), there exists a constant $K_1 > 0$ such that

$$\sum_{\substack{l=1\\ l\neq i}}^{n} |\rho_0^{il}| \leq \sum_{\substack{l=1\\ l\neq i}}^{n} (\rho_0^{il})^2 \leq \frac{\mathsf{Var}(\epsilon_{i\,t})}{\mu_{\min}(\mathbf{C}_0^{-1})} = \mathsf{Var}(\epsilon_{i\,t})\mu_{\max}(\mathbf{C}_0) < K_1. \tag{B-17}$$

Thus, given (B-17), and since by Assumption 1, we have $0 < c_{0\,ii} < \infty$ for any $i = 1, \ldots, n$, there exists a constant $K_2 > 0$ such that

$$\left| \sum_{\substack{l=1 \\ l \neq i}}^{n} \rho_0^{il} \sqrt{\frac{c_{0\,ll}}{c_{0\,ii}}} \left(\rho_0^{li} \sqrt{\frac{c_{0\,ii}}{c_{0\,ll}}} - 1 \right) \right| \leq \sum_{\substack{l=1 \\ l \neq i}}^{n} (\rho_0^{il})^2 + \sum_{\substack{l=1 \\ l \neq i}}^{n} |\rho_0^{il}| \sqrt{\frac{c_{0\,ll}}{c_{0\,ii}}} < K_2.$$

By the Cauchy–Schwartz inequality, we have that the mixed partial derivatives with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\rho}$ also not grow with n and satisfy a Bernstein–type concentration inequality. Thus, there exists a constant $K_3 > 0$ such that $|A_{1T\,ijk,i'j'k'}| \leq K_3 |y_{j\,t-k} y_{j'\,t-k'} - \mathsf{E}[y_{j\,t-k} y_{j'\,t-k'}]|$ for any (i, j, k) and (i', j', k'). Therefore, by Assumption 1 and the same arguments leading to (B-14) there exists a constant $K_4 > 0$ such that

$$\Pr\left(\sum_{i=1}^{q_T} \sum_{j=1}^{q_n} |A_{T\,ij}|^2 \geq \varepsilon^2 \right) \leq q_T^2 \Pr\left(|A_{T\,ij}| \geq \frac{\varepsilon}{q_T} \right) \leq 2 q_T^2 \exp\left\{ -K_4 T \frac{\varepsilon^2}{q_T^2} \right\}.$$

By setting the rhs of the last expression equal to $\delta = O(T^{-\eta'})$ for $\eta' > 0$ and solving with respect to ε we get that for T sufficiently large there exist a constant $\kappa_2 > 0$ such that

$$\varepsilon \leq \kappa_2 \, q_T \sqrt{\frac{\log T}{T}}.$$

Finally, for $\eta > 0$ and T sufficiently large there exists a constant κ_2 such that

$$\|\mathbf{H}_{T\,\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0)\mathbf{u} - \mathbf{H}_{0\,\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0)\mathbf{u}\| < \kappa_2 \|\mathbf{u}\|^2 q_T \sqrt{\frac{\log T}{T}},$$

with at least probability $1 - O(T^{-\eta})$. Part (c) follows as part (a) by using Condition 2 and the conditions in the statements of Propositions A1 and 1. Part (d) follows from (c) and the Cauchy-Schwarz inequality. This completes the proof. □

Proof of lemma A6. See Lemma 2.1 in Bühlmann and van de Geer (2011). □

Proof of lemma A7. Define $\nu_T = \sqrt{q_T}\lambda_T$, therefore $\nu_T \to 0$ as $T \to \infty$. Consider a generic vector $\mathbf{u} \in \mathbb{R}^m$ such that $\mathbf{u}_{\mathcal{A}^c} = \mathbf{0}$ and $\|\mathbf{u}\| = C$. Define $L_T(\boldsymbol{\theta}, \mathbf{c}) = \frac{1}{T}\sum_{t=1}^{T} \ell(\boldsymbol{\theta}, \mathbf{y}_t, \mathbf{c})$ and $\ell(\boldsymbol{\theta}, \mathbf{y}_t, \mathbf{c})$ is the unconstrained loss function defined in (9). The increment of the sample loss defined in (11)-(12) is

$$\begin{aligned} Q_T(\boldsymbol{\theta}_0 + \nu_T \mathbf{u}) &= \frac{1}{T}\Big(\mathcal{L}_T(\boldsymbol{\theta}_0 + \nu_t \mathbf{u}, \widehat{\mathbf{c}}_T) - \mathcal{L}_T(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T) \Big) = \\ &= \Big[L_T(\boldsymbol{\theta}_0 + \nu_t \mathbf{u}, \widehat{\mathbf{c}}_T) - L_T(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T) \Big] - \lambda_T \sum_{\substack{i=1 \\ i \in \mathcal{A}}}^{m} \frac{|\theta_{0i}| - |\theta_{0i} + \nu_T u_i|}{|\widetilde{\theta}_{Ti}|} \\ &\geq \Big[L_T(\boldsymbol{\theta}_0 + \nu_t \mathbf{u}, \widehat{\mathbf{c}}_T) - L_T(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T) \Big] - \lambda_T \nu_T \sum_{\substack{i=1 \\ i \in \mathcal{A}}}^{m} \frac{|u_i|}{|\widetilde{\theta}_{Ti}|}. \end{aligned} \tag{B-18}$$

Start from the first term in (B-18). By Lemma A5, for T sufficiently large and for any $\eta > 0$, we have with probability at least $1 - O(T^{-\eta})$

$$L_T(\boldsymbol{\theta}_0 + \nu_t \mathbf{u}, \widehat{\mathbf{c}}_T) - L_T(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T) = \nu_T \mathbf{u}'_{\mathcal{A}} \mathbf{S}_{T\mathcal{A}}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T) + \frac{1}{2}\nu_T^2 \mathbf{u}'_{\mathcal{A}} \mathbf{H}_{T\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T)\mathbf{u}_{\mathcal{A}} \tag{B-19}$$
$$= \nu_T \mathbf{u}'_{\mathcal{A}} \mathbf{S}_{T\mathcal{A}}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T) + \frac{1}{2}\nu_T^2 \mathbf{u}'_{\mathcal{A}} \mathbf{H}_{0\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T)\mathbf{u}_{\mathcal{A}} + \frac{1}{2}\nu_T^2 \mathbf{u}'_{\mathcal{A}}\Big(\mathbf{H}_{T\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T) - \mathbf{H}_{0\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0)\Big)\mathbf{u}_{\mathcal{A}} + o(\nu_T^2)$$
$$\geq -\kappa_1 \|\mathbf{u}_{\mathcal{A}}\| \sqrt{q_T}\sqrt{\frac{\log T}{T}}\nu_T - \kappa_3 \|\mathbf{u}_{\mathcal{A}}\|^2 q_T \sqrt{\frac{\log T}{T}}\nu_T^2 + \frac{1}{2}\nu_T^2 u'_{\mathcal{A}} \mathbf{H}_{0\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0)\mathbf{u}_{\mathcal{A}}.$$

By the conditions given in the statements of Propositions A1 and 1 and since $\|\mathbf{u}_{\mathcal{A}}\| = C$, for the first and second term on the rhs of (B-19) we have

$$-\kappa_1 C \sqrt{q_T}\sqrt{\frac{\log T}{T}}\frac{\lambda_T}{\lambda_T}\nu_T = \nu_T^2 o(1) = o(\nu_T^2), \tag{B-20}$$
$$-\kappa_3 C^2 q_T \sqrt{\frac{\log T}{T}}\nu_T^2 = \nu_T^2 o(1) = o(\nu_T^2), \tag{B-21}$$

and both terms can be neglected for T sufficiently large. Moreover, by Condition 1, we have

$$\frac{1}{2}\nu_T^2 u'_{\mathcal{A}} \mathbf{H}_{0\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0)\mathbf{u}_{\mathcal{A}} \geq \frac{1}{2}\nu_T^2 C^2 \mu_{\min}(\mathbf{H}_{0\mathcal{A}\mathcal{A}}) \geq \frac{1}{2}\nu_T^2 C^2 \underline{L} > 0. \tag{B-22}$$

Then, notice that, by Cauchy-Schwarz inequality,

$$\left(\sum_{\substack{i=1\\ i\in\mathcal{A}}}^{m} \frac{|u_i|}{|\widetilde{\theta}_{Ti}|}\right)^2 \leq C^2 \sum_{\substack{i=1\\ i\in\mathcal{A}}}^{m} \frac{1}{\widetilde{\theta}_{Ti}^2}. \tag{B-23}$$

Moreover, for any $i \in \mathcal{A}$,

$$\frac{1}{\widetilde{\theta}_{Ti}^2} = \frac{1}{\theta_{0i}^2} - \frac{2\theta_{0i}}{\theta_{0i}^4}(\widetilde{\theta}_{Ti} - \theta_{0i}) + o((\widetilde{\theta}_{Ti} - \theta_{0i})) \leq \frac{1}{\theta_{0i}^2} + \frac{2}{\theta_{0i}^3}|\widetilde{\theta}_{Ti} - \theta_{0i}| + o(|\widetilde{\theta}_{Ti} - \theta_{0i}|). \tag{B-24}$$

Define $\theta_{0\,\min}^2 = \min_{i\in\mathcal{A}} \theta_{0i}^2$ and notice that $|\theta_{0\,\min}| > 0$. Then, combining Assumption 2 and (B-24), there exists a constant $K > 0$ such that for T sufficiently large and for any $\eta > 0$, we have with probability at least $1 - O(T^{-\eta})$

$$C^2 \sum_{\substack{i=1\\ i\in\mathcal{A}}}^{m} \frac{1}{\widetilde{\theta}_{Ti}^2} \leq \frac{C^2 q_T}{\theta_{0\,\min}^2} + C^2 K q_T \sqrt{q_T \frac{\log T}{T}}. \tag{B-25}$$

Therefore, using (B-23) and (B-25), for the second term in (B-18), for T sufficiently large and for any $\eta > 0$, with probability at least $1 - O(T^{-\eta})$, we have

$$-\lambda_T \nu_T \sum_{\substack{i=1\\ i\in\mathcal{A}}}^{m} \frac{|u_i|}{|\widetilde{\theta}_{Ti}|} \geq -\lambda_T \nu_T C \sqrt{q_T}\left[\frac{1}{|\theta_{0\min}|} + \sqrt{K}\left(q_T \frac{\log T}{T}\right)^{1/4}\right]$$

$$\geq -\nu_T^2 \frac{C}{|\theta_{0\min}|} + \nu_T^2 \sqrt{K}\left(q_T \frac{\log T}{T}\right)^{1/4}, \tag{B-26}$$

and notice that the last term is $o(\nu_T^2)$, thus it can be neglected for T sufficiently large. Then, by substituting (B-19) and (B-26) in (B-18), and using (B-20), (B-21), and (B-22), we have, for T sufficiently large and for any $\eta > 0$,

$$\Pr\left(Q_T(\boldsymbol{\theta}_0 + \nu_T \mathbf{u}) \geq \frac{1}{2}\nu_T^2 C^2 \underline{L} - \frac{C}{|\theta_{0\min}|}\nu_T^2 = \nu_T^2 C\left(\frac{\underline{L}}{2}C - \frac{1}{|\theta_{0\min}|}\right)\right) \geq 1 - O(T^{-\eta}).$$

Thus, if we choose $C = 2/(\underline{L}|\theta_{0\min}|) + \epsilon$, for any $\epsilon > 0$, then for T sufficiently large and for any $\eta > 0$

$$\Pr\left(\inf_{\substack{\mathbf{u}:\mathbf{u}_{\mathcal{A}^c}=\mathbf{0}\\ \|\mathbf{u}\|=C}} Q_T(\boldsymbol{\theta}_0 + \nu_T\mathbf{u}) > 0\right) = \Pr\left(\inf_{\substack{\mathbf{u}:\mathbf{u}_{\mathcal{A}^c}=\mathbf{0}\\ \|\mathbf{u}\|=C}} \mathcal{L}_T(\boldsymbol{\theta}_0 + \nu_T\mathbf{u}, \widehat{\mathbf{c}}_T) > \mathcal{L}_T(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T)\right) \geq 1 - O(T^{-\eta}),$$

which means that there exists a local minimum for the restricted problem within the disc $D(\boldsymbol{\theta}_0) = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \nu_T C\}$, with probability at least $1 - O(T^{-\eta})$. By choosing $\kappa_4 = C$, we complete the proof. □

Proof of lemma A8. Define $\nu_T = \sqrt{q_T}\lambda_T$, therefore $\nu_T \to 0$ as $T \to \infty$. Then, any $\boldsymbol{\theta} \in S(\boldsymbol{\theta}_0)$ can be written as $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \nu_T\mathbf{u}$, where $\mathbf{u}_{\mathcal{A}^c} = \mathbf{0}$, $\|\mathbf{u}\| \geq \kappa_5$, and $\|\mathbf{u}\| \leq C < \infty$. For any $\boldsymbol{\theta} \in S(\boldsymbol{\theta}_0)$, we can write

$$\begin{aligned}\mathbf{S}_{T\mathcal{A}}(\boldsymbol{\theta}, \widehat{\mathbf{c}}_T) &= \mathbf{S}_{T\mathcal{A}}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T) + \nu_T \mathbf{H}_{T\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T)\mathbf{u}\\ &= \mathbf{S}_{T\mathcal{A}}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T) + \nu_T\Big(\mathbf{H}_{T\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T) - \mathbf{H}_{0\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0)\Big)\mathbf{u} + \nu_T\mathbf{H}_{0\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0)\mathbf{u} + o(\nu_T).\end{aligned}$$

Thus, by Lemma A5, for T sufficiently large and for any $\eta > 0$, we have, with probability at least $1 - O(T^{-\eta})$,

$$\|\mathbf{S}_{T\mathcal{A}}(\boldsymbol{\theta}, \widehat{\mathbf{c}}_T)\| \geq -\kappa_0\sqrt{q_T}\sqrt{\frac{\log T}{T}} - \kappa_2\|\mathbf{u}\|\nu_T q_T\sqrt{\frac{\log T}{T}} + \nu_T\|\mathbf{H}_{0\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0)\mathbf{u}\|.$$

The first and second term on the rhs of the last expression are both $o(\nu_T)$. Then, using Condition 1, for T sufficiently large and for any $\eta > 0$, with probability at least $1 - O(T^{-\eta})$ we have

$$\|\mathbf{S}_{T\mathcal{A}}(\boldsymbol{\theta}, \widehat{\mathbf{c}}_T)\| \geq \nu_T\|\mathbf{H}_{0\mathcal{A}\mathcal{A}}(\boldsymbol{\theta}_0, \mathbf{c}_0)\mathbf{u}\| \geq \nu_T\,\underline{L}\,\kappa_5.$$

Define $\theta_0^* = \min_{i\in\mathcal{A}}|\theta_{0i}|$ and notice that $\theta_0^* > 0$. By choosing $\kappa_5 = 1/(\underline{L}\theta_0^*) + \epsilon$ for any $\epsilon > 0$, we complete the proof. □

Proof of lemma A9. In the following define $\mathbf{v}_T = (\widehat{\boldsymbol{\theta}}_T^{\mathcal{A}} - \boldsymbol{\theta}_0)$. For any $j \in \mathcal{A}^c$ we have

$$\begin{aligned} S_{T\,j}(\widehat{\boldsymbol{\theta}}_T^{\mathcal{A}}, \widehat{\mathbf{c}}_T) &= S_{T\,j}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T) + H_{T\,j}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T)\mathbf{v}_T + o(\|\mathbf{v}_T\|) \\ &= \underbrace{S_{T\,j}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T)}_{A_{1T\,j}} + \underbrace{H_{0\,j}(\boldsymbol{\theta}_0, \mathbf{c}_0)\mathbf{v}_T}_{A_{2T\,j}} + \underbrace{\left[H_{T\,j}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T) - H_{0\,j}(\boldsymbol{\theta}_0, \mathbf{c}_0)\right]\mathbf{v}_T}_{A_{3T\,j}} + o(\|\mathbf{v}_T\|). \end{aligned} \tag{B-27}$$

Start from the first term on the rhs of (B-27). By an argument analogous to the proof of Lemma A5 (a) we have that for T sufficiently large and for any $\eta' > 0$ there exists a constant $K_1 > 0$ such that

$$\Pr\left(|A_{1T\,j}| \le K_1\sqrt{\frac{\log T}{T}}\right) \ge 1 - O(T^{-\eta'}). \tag{B-28}$$

For the second term on the rhs of (B-27) from Condition 1 we have

$$|A_{2T\,j}| \le \|H_{0\,j}(\boldsymbol{\theta}_0, \mathbf{c}_0)\|\,\|\mathbf{v}_T\| \le \mu_{\max}(\mathbf{H}_0(\boldsymbol{\theta}_0, \mathbf{c}_0))\,\|\mathbf{v}_T\| \le \overline{L}\,\|\mathbf{v}_T\|.$$

Therefore, if we define $K_2 = \kappa_R \overline{L}$, by Proposition A1 we have that for T sufficiently large and for any $\eta' > 0$

$$\Pr\left(|A_{2T\,j}| \le K_2\sqrt{q_T}\lambda_T\right) \ge 1 - O(T^{-\eta'}). \tag{B-29}$$

For the third term on the rhs of (B-27), we have

$$|A_{3T\,j}| \le \|\left[H_{T\,j}(\boldsymbol{\theta}_0, \widehat{\mathbf{c}}_T) - H_{0\,j}(\boldsymbol{\theta}_0, \mathbf{c}_0)\right]\|\,\|\mathbf{v}_T\|.$$

Then, using an argument similar to the proof of Lemma A5 (c) and by Proposition A1 and by defining $K_3 = \kappa_2\kappa_R$ we have that for T sufficiently large and for any $\eta' > 0$

$$\Pr\left(|A_{3T\,j}| \le K_3 q_T \lambda_T\sqrt{\frac{\log T}{T}}\right) \ge 1 - O(T^{-\eta'}). \tag{B-30}$$

Moreover, by Assumption 2 there exists a constant $K_4 > 0$ such that for T sufficiently large and for any $\eta' > 0$ we have

$$\Pr\left(\frac{1}{\max_{j\in\mathcal{A}^c}|\widetilde{\theta}_{Tj}|} > K_4\sqrt{\frac{T}{\log T}}\right) \ge 1 - O(T^{-\eta'}). \tag{B-31}$$

From (B-31) and (B-27) we have

$$\begin{aligned} \Pr\left(|S_{T\,j}(\widehat{\boldsymbol{\theta}}_T^{\mathcal{A}}, \widehat{\mathbf{c}}_T)| \le \frac{\lambda_T}{\max_{j\in\mathcal{A}^c}|\widetilde{\theta}_{Tj}|}\right) &\ge \Pr\left(|A_{1T\,j}| + |A_{2T\,j}| + |A_{3T\,j}| \le \frac{\lambda_T}{\max_{j\in\mathcal{A}^c}|\widetilde{\theta}_{Tj}|}\right) \\ &\ge \Pr\left(|A_{1T\,j}| + |A_{2T\,j}| + |A_{3T\,j}| \le K_4\lambda_T\sqrt{\frac{T}{\log T}}\right). \end{aligned} \tag{B-32}$$

Notice that, as $T \to \infty$, we have

$$\lambda_T \sqrt{\frac{T}{\log T}} \to \infty, \quad \sqrt{q_T}\lambda_T \to 0, \quad q_T \lambda_T \sqrt{\frac{\log T}{T}} \to 0, \quad \sqrt{\frac{\log T}{T}} \to 0. \tag{B-33}$$

where the first three conditions are assumed in Proposition 1 while the last one is trivial.

Finally, consider the complementary of (B-32), then, by combining (B-28)-(B-30) with (B-33), we have that for T sufficiently large and for any $\eta' > 0$

$$\Pr\left(|A_{1Tj}| + |A_{2Tj}| + |A_{3Tj}| \geq K_4 \lambda_T \sqrt{\frac{T}{\log T}} \right) \leq \sum_{k=1}^{3} \Pr\left(|A_{kTj}| \geq K_4 \lambda_T \sqrt{\frac{T}{\log T}} \right) = O(T^{-\eta'}),$$

which implies that

$$\Pr\left(|S_{Tj}(\widehat{\boldsymbol{\theta}}_T^{\mathcal{A}}, \widehat{\mathbf{c}}_T)| \leq \frac{\lambda_T}{\max_{j \in \mathcal{A}^c} |\widetilde{\theta}_{Tj}|} \right) \geq 1 - O(T^{-\eta'}). \tag{B-34}$$

Given $n = O(T^{\zeta})$, define $\eta' = \eta + \zeta$, then for T sufficiently large and for any $\eta > 0$, from (B-34) we have

$$\Pr\left(\max_{j \in \mathcal{A}^c} |S_{Tj}(\widehat{\boldsymbol{\theta}}_T^{\mathcal{A}}, \widehat{\mathbf{c}}_T)| \geq \frac{\lambda_T}{\max_{j \in \mathcal{A}^c} |\widetilde{\theta}_{Tj}|} \right) \leq n \Pr\left(|S_{Tj}(\widehat{\boldsymbol{\theta}}_T^{\mathcal{A}}, \widehat{\mathbf{c}}_T)| \geq \frac{\lambda_T}{\max_{j \in \mathcal{A}^c} |\widetilde{\theta}_{Tj}|} \right) = O(T^{-\eta}).$$

By considering the complementary event we complete the proof. □

C Ticker List

Table C-1 contains the list of tickers used in the empirical application.

Table C-1: U.S. BLUECHIPS

Ticker	Company Name	Sector	Ticker	Company Name	Sector
AMZN	Amazon.com	Cons. Disc.	ABT	Abbott Laboratories	Health Care
CMCSA	Comcast	Cons. Disc.	AMGN	Amgen	Health Care
DIS	Walt Disney	Cons. Disc.	BAX	Baxter International	Health Care
F	Ford Motor	Cons. Disc.	BMY	Bristol-Myers Squibb	Health Care
FOXA	Twenty-First Century Fox	Cons. Disc.	GILD	Gilead Sciences	Health Care
HD	Home Depot	Cons. Disc.	JNJ	Johnson & Johnson	Health Care
LOW	Lowes	Cons. Disc.	LLY	Lilly (Eli) & Co.	Health Care
MCD	McDonalds	Cons. Disc.	MDT	Medtronic	Health Care
NKE	NIKE	Cons. Disc.	MRK	Merck & Co.	Health Care
SBUX	Starbucks	Cons. Disc.	PFE	Pfizer	Health Care
TGT	Target	Cons. Disc.	UNH	United Health	Health Care
TWX	Time Warner	Cons. Disc.	BA	Boeing Company	Industrials
CL	Colgate-Palmolive	Cons. Stap.	CAT	Caterpillar	Industrials
COST	Costco	Cons. Stap.	EMR	Emerson Electric	Industrials
CVS	CVS Caremark	Cons. Stap.	FDX	FedEx	Industrials
KO	The Coca Cola Company	Cons. Stap.	GD	General Dynamics	Industrials
MDLZ	Mondelez International	Cons. Stap.	GE	General Electric	Industrials
MO	Altria	Cons. Stap.	HON	Honeywell Intl	Industrials
PEP	PepsiCo	Cons. Stap.	LMT	Lockheed Martin	Industrials
PG	Procter & Gamble	Cons. Stap.	MMM	3M Company	Industrials
WMT	Wal-Mart Stores	Cons. Stap.	NSC	Norfolk Southern	Industrials
APA	Apache	Energy	RTN	Raytheon	Industrials
APC	Anadarko Petroleum	Energy	UNP	Union Pacific	Industrials
COP	ConocoPhillips	Energy	UPS	United Parcel Service	Industrials
CVX	Chevron	Energy	UTX	United Technologies	Industrials
DVN	Devon Energy	Energy	AAPL	Apple	Technology
HAL	Halliburton	Energy	ACN	Accenture plc	Technology
NOV	National Oilwell Varco	Energy	CSCO	Cisco Systems	Technology
OXY	Occidental Petroleum	Energy	EBAY	eBay	Technology
SLB	Schlumberger Ltd.	Energy	EMC	EMC	Technology
XOM	Exxon Mobil	Energy	HPQ	Hewlett-Packard	Technology
AIG	AIG	Financials	IBM	IBM	Technology
ALL	Allstate	Financials	INTC	Intel	Technology
AXP	American Express Co	Financials	MSFT	Microsoft	Technology
BAC	Bank of America	Financials	ORCL	Oracle	Technology
BK	Bank of New York	Financials	QCOM	QUALCOMM	Technology
C	Citigroup	Financials	TXN	Texas Instruments	Technology
COF	Capital One Financial	Financials	T	AT&T	Technology
GS	Goldman Sachs	Financials	VZ	Verizon	Technology
JPM	JPMorgan Chase	Financials	DD	Du Pont	Materials
MET	MetLife	Financials	DOW	Dow Chemical	Materials
MS	Morgan Stanley	Financials	FCX	Freeport-McMoran	Materials
SPG	Simon Property	Financials	MON	Monsanto	Materials
USB	U.S. Bancorp	Financials	AEP	American Electric Power	Utilities
WFC	Wells Fargo	Financials	EXC	Exelon	Utilities

The table reports the list of tickers, company names and industry sectors.

References

Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction.* Springer, New York.

Bühlmann, P. and van de Geer, S. (2011). *Statistics for High–Dimensional Data: Methods, Theory and Applications.* Springer, New York.

Dedecker, J., Doukhan, P., Lang, G., León, J., Louhichi, S., and Prieur, C. (2007). *Weak Dependence: With Examples and Applications.* Springer, New York.

Doukhan, P. and Neumann, M. H. (2007). Probability and moment inequalities for sums of weakly dependent random variables, with applications. *Stochastic Processes and their Applications*, 117:878–903.

Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics*, 32:928–961.

Meinshausen, N. and Bühlmann, P. (2006). High dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34:1436–1462.

Peng, J., Wang, P., Zhou, N., and Zhu, J. (2009). Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*, 104:735–746.