Supplemental Section

A SPACE method

Algorithm 1 (SPACE pseudocode)

Input: Standardize data to have mean zero and standard deviation one
Input: Fix maximum number of iterations: r_{max}
Input: Fix initial estimate: ($\hat{\omega}_{ii}^{(0)} = 1/s_{ii}$ as suggested)
Input: Choose weights[a]: w_i ($w_i = \omega_{ii}$ or $w_i = 1$)
Set $r \leftarrow 1$
repeat

`## Update partial correlations`

Update $\hat{\boldsymbol{\rho}}^{(r)}$ by minimizing (with current estimates $\{\hat{\omega}_{ii}^{(r-1)}\}_{i=1}^{p}$ as fixed)

$$\frac{1}{2}\sum_{i=1}^{p}\left(w_i\|\mathbf{Y}_i - \sum_{j\neq i}\rho^{ij}\sqrt{\frac{\hat{\omega}_{jj}^{(r-1)}}{\hat{\omega}_{ii}^{(r-1)}}}\mathbf{Y}_j\|_2^2\right) + \lambda\sum_{1\le i<j\le p}\left|\rho^{ij}\right| \tag{18}$$

`## Update conditional variances`

Update $\{\omega_{ii}^{(r)}\}_{i=1}^{p}$ by computing (with fixed $\hat{\rho}_{ij}^{(r-1)}$ and $\hat{\omega}_{ii}^{(r-1)}$ for all i and j)

$$\frac{1}{\hat{\omega}_{ii}^{(r)}} = \frac{1}{n}\|\mathbf{Y}_i - \sum_{j\neq i}(\hat{\rho}^{ij})^{(r-1)}\sqrt{\frac{\hat{\omega}_{jj}^{(r-1)}}{\hat{\omega}_{ii}^{(r-1)}}}\mathbf{Y}_j\|_2^2 \tag{19}$$

for $i = 1, \ldots, p$.

$r \leftarrow r + 1$
Update weights: w_i
until $r == r_{max}$
Return $(\hat{\boldsymbol{\rho}}^{(r_{max})}, \{\hat{\omega}_{ii}^{(r_{max})}\}_{i=1}^{p})$

---

[a]Peng et al. (2009) suggest two natural choices of weights w_i: (1) uniform weights $w_i = 1$ for all $i = 1, 2, \ldots, p$ (ii) partial variance weights $w_i = \omega_{ii}$.

Proof of Lemma 1: Note that when fixing the diagonals $\{\omega_{ii}\}_{i=1}^{p}$, the minimization in (18) in the SPACE algorithm (with weights $w_i = \omega_{ii}$), corresponds to minimizing Q_{spc} with respect to $\boldsymbol{\rho}$. Now, let $\hat{\omega}_{ii}$ be the minimizer of Q_{spc} with respect to ω_{ii}, fixing $\{\beta_{ij}\}_{1\le i\neq j\le p}$ (where

$\beta_{ij} = \rho^{ij}\sqrt{\frac{\omega_{jj}}{\omega_{ii}}} = -\frac{\omega_{ij}}{\omega_{ii}}$). Then, it follows that

$$\hat{\omega}_{ii} = \left(\frac{1}{n}\|\mathbf{Y}_i - \sum_{j\neq i}\beta_{ij}\mathbf{Y}_j\|_2^2\right)^{-1} \tag{20}$$

The result follows by comparing (20) with the updates in (19). $\square$

B Proof of Lemma 2

Let $\mathbf{Y}$ denote the $n \times p$ matrix with j^{th} column given by $\mathbf{Y}_j$ for $j = 1, 2, \ldots, p$. Define $Q_{\text{sym}}(\boldsymbol{\alpha}, \breve{\Omega}) = \frac{1}{2}\left(\sum_{j=1}^{p}\mathcal{L}_{\text{sym},j}(\alpha_{jj}, \breve{\Omega}_j)\right) + \lambda\left(\sum_{1\leq i<j\leq p}|\omega_{ij}|\right)$ so that

$$\mathcal{L}_{\text{sym},j}(\alpha_{jj}, \breve{\Omega}_j) = n\log\alpha_{jj} + \frac{1}{\alpha_{jj}}\|\mathbf{Y}_j + \mathbf{Y}\breve{\Omega}_j\alpha_{jj}\|_2^2 \tag{21}$$

where $\boldsymbol{\alpha} = (\alpha_{11}\ \alpha_{22}\ \cdots\ \alpha_{pp})'$, $\alpha_{jj} = 1/\omega_{jj}$ and $\breve{\Omega}_j$ is the j^{th} column of $\breve{\Omega}$. Recall that $\breve{\Omega}$ is the matrix Ω with zeros in place of the diagonal entries. If follows that

$$\frac{\partial Q_{\text{sym}}(\boldsymbol{\alpha}, \breve{\Omega})}{\partial\alpha_{jj}} = \frac{n}{\alpha_{jj}} - \frac{\mathbf{Y}_j'\mathbf{Y}_j}{\alpha_{jj}^2} + \breve{\Omega}_j'\mathbf{Y}'\mathbf{Y}\breve{\Omega}_j, \quad \text{and} \quad \frac{\partial^2 Q_{\text{sym}}(\boldsymbol{\alpha}, \breve{\Omega})}{\partial\alpha_{jj}^2} = -\frac{n}{\alpha_{jj}^2} + 2\frac{\mathbf{Y}_j'\mathbf{Y}_j}{\alpha_{jj}^3} \tag{22}$$

It is clear that in general $\partial^2 Q_{\text{sym}}(\boldsymbol{\alpha}, \breve{\Omega})/\partial\alpha_{jj}^2 \not\geq 0$. Hence, $Q_{\text{sym}}(\boldsymbol{\alpha}, \breve{\Omega})$ is not convex.

C Proof of Lemma 3

Proof. i) Rewrite the SPLICE objective function $Q_{\text{spl}}(\mathbf{B}, \mathbf{D}) = \mathcal{L}_{\text{spl}}(\mathbf{B}, \mathbf{D}) + \lambda\sum_{i<j}|\beta_{ij}|$ where

$$\mathcal{L}_{\text{spl}}(\mathbf{B}, \mathbf{D}) = \frac{1}{2}\left[n\log\det(\mathbf{D}^2) + \text{tr}(\mathbf{D}^{-2}\mathbf{A})\right],$$

and $\mathbf{A} = [a_{ij}] = (\mathbf{I} - \mathbf{B})\mathbf{Y}'\mathbf{Y}(\mathbf{I} - \mathbf{B}')$. The function $\mathcal{L}_{\text{spl}}(\mathbf{B}, \mathbf{D})$ with all variables fixed except d_{jj} is given by

$$\mathcal{L}_{\text{spl},j}(\mathbf{B}, d_{jj}) = \frac{1}{2}\left[n\log d_{jj}^2 + \frac{a_{jj}}{d_{jj}^2}\right] + \text{constants}.$$

Now,

$$
\begin{aligned}
\frac{\partial Q_{\text{spl}}(\mathbf{B},\mathbf{D})}{\partial d_{jj}} &= \frac{n}{d_{jj}} - \frac{a_{jj}}{d_{jj}^3} \\
\frac{\partial^2 Q_{\text{spl}}(\mathbf{B},\mathbf{D})}{\partial d_{jj}^2} &= -\frac{n}{d_{jj}^2} + 3\frac{a_{jj}}{d_{jj}^4}
\end{aligned}
$$

It is clear in general $\partial Q_{\text{spl}}^2(\mathbf{B},\mathbf{D})/\partial d_{jj}^2 \ngeq 0$. Hence $Q_{\text{spl}}(\mathbf{B},\mathbf{D})$ is not convex.

ii) Similarly, define $Q_{\text{spl}}^*(\mathbf{B},\mathbf{C}) = \mathcal{L}_{\text{spl}}^*(\mathbf{B},\mathbf{C}) + \lambda \sum_{i<j} |\beta_{ij}|$ where

$$
\mathcal{L}_{\text{spl}}^*(\mathbf{B},\mathbf{C}) = \frac{1}{2}\left[n \log \mathbf{C}^{-2} + \operatorname{tr}(\mathbf{C}^2\mathbf{A})\right].
$$

It is clear that for a fixed $\mathbf{C}$, $\mathcal{L}_{\text{spl}}^*(\mathbf{B},\mathbf{C})$ is a convex function in $\mathbf{B}$ (Rocha et al., 2008). Now for a fixed $\mathbf{B}$ let

$$
\begin{aligned}
\mathcal{L}_{\text{spl},j}^*(\mathbf{B}, c_{jj}) &= \frac{1}{2}\left[-2n \log c_{jj} + c_{jj}^2 a_{jj}\right] + \text{constants} \\
\frac{\partial Q_{\text{spl}}^*(\mathbf{B},\mathbf{C})}{\partial c_{jj}} &= -\frac{n}{c_{jj}} + c_{jj} a_{jj} \\
\frac{\partial^2 Q_{\text{spl}}^*(\mathbf{B},\mathbf{C})}{\partial c_{jj}^2} &= \frac{n}{c_{jj}^2} + a_{jj}
\end{aligned}
$$

Now, note that $\partial (Q_{\text{spl}}^*)^2(\mathbf{B},\mathbf{C})/\partial c_{jj}^2 \geq 0$ since $a_{jj} \geq 0$.

To see that $a_{jj} \geq 0$ note that $\mathbf{A} = (\mathbf{I}-\mathbf{B})\mathbf{Y}'\mathbf{Y}(\mathbf{I}-\mathbf{B}') = \mathbf{G}'\mathbf{G}$, where $\mathbf{G} = \mathbf{Y}(\mathbf{I}-\mathbf{B}')$ Now, $a_{jj} = \mathbf{G}_{\bullet j}'\mathbf{G}_{\bullet j} = \|\mathbf{G}_{\bullet j}\|^2 \geq 0$ $\square$

D CONCORD algorithm

Algorithm 2 (CONCORD pseudocode)

Input: standardize data to have mean zero and standard deviation one
Input: Fix maximum number of iterations: r_{max}
Input: Fix initial estimate: $\hat{\Omega}^{(0)}$
Input: Fix convergence threshold: ϵ
Set $r \leftarrow 1$
converged = FALSE
Set $\hat{\Omega}^{\text{current}} \leftarrow \hat{\Omega}^{(0)}$
repeat
    $\hat{\Omega}^{\text{old}} \leftarrow \hat{\Omega}^{\text{current}}$
    `## Updates to partial covariances` ω_{ij}
    for $i \leftarrow 1, 2, \cdots, p-1$ **do**
        for $j \leftarrow i+1, \cdots, p$ **do**

$$\hat{\omega}_{ij}^{\text{current}} \leftarrow (T_{ij}(\Omega^{\text{current}}))_{ij} \tag{23}$$

        end for
    end for

    `## Updates to partial variances` ω_{ii}
    for $i \leftarrow 1, 2, \cdots, p$ **do**

$$\hat{\omega}_{ii}^{\text{current}} \leftarrow (T_{ii}(\Omega^{\text{current}}))_{ii} \tag{24}$$

    end for

    $\hat{\Omega}^{(r)} \leftarrow \hat{\Omega}^{\text{current}}$
    `## Convergence checking`
    if $\|\hat{\Omega}^{\text{current}} - \hat{\Omega}^{\text{old}}\|_{\max} < \epsilon$ **then**
        converged = TRUE
    else
        $r \leftarrow r+1$
    end if

until converged = TRUE or $r > r_{\max}$
Return final estimate: $\hat{\Omega}^{(r)}$

E Computational complexity

We now proceed to show that the computational cost of each iteration of CONCORD is $min\,(O(np^2), O(p^3))$, that is, the CONCORD algorithm is competitive with other proposed methods. The updates in Equations in (23) and (24) are implemented differently depending on whether $n \geq p$ or $n < p$.

Case 1 ($n \geq p$): Let us first consider the case when $n \geq p$. Note that both sums in (11) are inner products between a row in $\hat{\Omega}$ and a row in $\mathbf{S}$. Clearly, computing these sums require $O(p)$ operations each. Similarly, the update in (10) requires $O(p)$ operations. Since there are $O(p^2)$ entries in Ω, one complete sweep of updates over all entries in $\hat{\Omega}$ would require $O(p^3)$ operations.

Case 2 ($n < p$): Let us now consider the case when $n < p$. We show below that the updates can be performed in $O(np^2)$ operations. The main idea here is that the coordinate-wise calculations at each iteration, which involves an inner product of two $p \times 1$ vectors, can be reduced to an inner product calculation involving auxiliary variables (residual variables to be more specific) of dimension $n \times 1$. The following lemmas are essential ingredients in calculating the computational complexity in this setting. In particular, Lemma 6 expresses the inner product calculations in (10) and (11) in terms of residual vectors.

Lemma 6. *For $1 \leq i, j \leq p$,*

$$\sum_{k \neq j} \omega_{ik} s_{jk} = -\omega_{ij} s_{jj} + \omega_{ii} \mathbf{Y}_j' \mathbf{r}_i,$$

where $\mathbf{Y}_j$ is the j^{th} column of the data matrix $\mathbf{Y}$, and $\mathbf{r}_i = \mathbf{Y}_i + \sum_{k \neq i} \frac{\omega_{ik}}{\omega_{ii}} \mathbf{Y}_k$ is an n-vector of residuals of regressing $\mathbf{Y}_i$ on the rest.

The following lemma now quantifies the computational cost of updating the residual vectors during each iteration of the CONCORD algorithm.

Lemma 7. *Define the residual vector $\mathbf{r}_m$ for $m = 1, 2, \ldots, p$ as follows:*

$$\mathbf{r}_m = \mathbf{r}_m(\Omega) = \mathbf{Y}_m + \sum_{k \neq m} \frac{\omega_{mk}}{\omega_{mm}} \mathbf{Y}_k \tag{25}$$

where $\Omega = ((\Omega_{ij}))_{1 \leq i,j \leq p}$. Then,

1. *For $m \neq k, l$, the residual vector $\mathbf{r}_m$ is functionally independent of ω_{kl}. (The term ω_{kl} appears only in the expressions for the residual vectors $\mathbf{r}_k$ and $\mathbf{r}_l$.)*

2. *Fix all the elements of* Ω *except* ω_{kl}. *Suppose* ω_{kl} *is changed to* ω^*_{kl}. *Then, updating the residual vectors* $\mathbf{r}_k$ *and* $\mathbf{r}_l$ *requires* $O(n)$ *operations. (Hence, updating* $\mathbf{r}_k$ *and* $\mathbf{r}_l$ *after each update in* (23) *requires* $O(n)$ *operations.)*

3. *For* $m \neq k$, *the residual vector* $\mathbf{r}_m$ *is functionally independent of* ω_{kk}. *(The term* ω_{kk} *appears only in the expression for the residual vector* $\mathbf{r}_k$.*)*

4. *Fix all elements of* Ω *except* ω_{kk}. *Suppose* ω_{kk} *is changed to* ω^*_{kk}. *Then, updating the residual vector* $\mathbf{r}_k$ *requires* $O(n)$ *operations. (Hence, updating* $\mathbf{r}_k$ *after each update in* (24) *requires* $O(n)$ *operations).*

The proofs of Lemmas 6 and 7 are straightforward and are given in Supplemental Sections G and H. Note that the inner product between $\mathbf{y}_j$ and $\mathbf{r}_i$ takes $O(n)$ operations. Hence, by Lemma 6 the updates in (23) and (24) require $O(n)$ operations. Also, after each update in (23) and (24) the residual vectors need to be appropriately modified. By Lemma 7, this modification can also be achieved in $O(n)$ operations. As a result, one complete sweep of updates over all entries in $\hat{\Omega}$ can be performed in $O(np^2)$ operations.

Hence, we conclude that the computational complexity of the CONCORD algorithm is competitive with the SPACE and Symmetric lasso algorithms, which are also min $(O(np^2), O(p^3))$.

F Proof of Lemma 4

Note that for $1 \leq i \leq p$,

$$Q_{\text{con}}(\Omega) = -n \log \omega_{ii} + \frac{n}{2}\left(\omega_{ii}^2 s_{ii} + 2\omega_{ii}\sum_{j\neq i}\omega_{ij}s_{ij}\right) + \text{ terms independent of } \omega_{ii}. \tag{26}$$

where $s_{ij} = \mathbf{Y}_i'\mathbf{Y}_j/n$. Hence,

$$\begin{aligned}\frac{\partial}{\partial \omega_{ii}} Q_{\text{con}}(\Omega) = 0 \quad &\Leftrightarrow \quad -\frac{1}{\omega_{ii}} + \omega_{ii}s_{ii} + \sum_{j\neq i}\omega_{ij}s_{ij} = 0 \\ &\Leftrightarrow \quad \omega_{ii} = \frac{-\sum_{j\neq i}\omega_{ij}s_{ij} + \sqrt{\left(\sum_{j\neq i}\omega_{ij}s_{ij}\right)^2 + 4s_{ii}}}{2s_{ii}},\end{aligned}$$

Note that since $\omega_{ii} > 0$ the positive root has been retained as the solution.

Also, for $1 \leq i < j \leq p$,

$$Q_{\text{con}}(\Omega) = n\frac{s_{ii} + s_{jj}}{2}\omega_{ij}^2 + n\left(\sum_{j' \neq j}\omega_{ij'}s_{jj'} + \sum_{i' \neq i}\omega_{i'j}s_{ii'}\right)\omega_{ij} + \lambda|\omega_{ij}| + \text{ terms independent of } \omega_{ij}. \tag{27}$$

It follows that

$$(T_{ij}(\Omega))_{ij} = \frac{S_{\frac{\lambda}{n}}\left(-\left(\sum_{j' \neq j}\omega_{ij'}s_{jj'} + \sum_{i' \neq i}\omega_{i'j}s_{ii'}\right)\right)}{s_{ii} + s_{jj}},$$

where S_η is the soft-thresholding operator given by $S_\eta(x) = \text{sign}(x)(|x| - \eta)_+$.

G Proof of Lemma 6

Let $\mathbf{Y}_j$ denote j^{th} column of the data matrix $\mathbf{Y}$. Then, using the identity $\sum_{k=1}^{p}\omega_{ik}s_{jk} = \omega_{ij}s_{jj} + \sum_{k \neq j}\omega_{ik}s_{jk} = \omega_{ii}s_{ij} + \sum_{k \neq i}\omega_{ik}s_{jk}$,

$$\begin{aligned}
\sum_{k \neq j}\omega_{ik}s_{jk} &= -\omega_{ij}s_{jj} + \omega_{ii}\left(s_{ij} + \sum_{k \neq i}\frac{\omega_{ik}}{\omega_{ii}}s_{jk}\right) \\
&= -\omega_{ij}s_{jj} + \omega_{ii}\mathbf{Y}_j'\left(\mathbf{Y}_i + \sum_{k \neq i}\frac{\omega_{ik}}{\omega_{ii}}\mathbf{Y}_k\right) \\
&= -\omega_{ij}s_{jj} + \omega_{ii}\mathbf{Y}_j'\mathbf{r}_i,
\end{aligned}$$

where $\mathbf{r}_i = \mathbf{Y}_i + \sum_{k \neq i}\frac{\omega_{ik}}{\omega_{ii}}\mathbf{Y}_k$ is an n-vector of residuals after regressing the i^{th} variable on the rest. □

H Proof of Lemma 7

1. Result follows easily from inspecting $\mathbf{r}_k$ and $\mathbf{r}_l$.

2. If ω_{kl} is updated to ω_{kl}^*, it follows from part 1 that among all the residual vectors, only $\mathbf{r}_k$ and $\mathbf{r}_l$ change values. The residual vector $\mathbf{r}_k$ can be updated as follows:

$$\mathbf{r}_k^* = \mathbf{r}_k + \frac{(\omega_{kl}^* - \omega_{kl})}{\omega_{kk}}\mathbf{Y}_l\,.$$

   Clearly, this update requires $O(n)$ operations. The vector $\mathbf{r}_l$ can be updated similarly.

3. Result follows easily from inspecting $\mathbf{r}_k$.

4. If ω_{kk} is updated to ω_{kk}^*, it follows from part 3 that among all the residual vectors, only $\mathbf{r}_k$ changes value. The residual vector $\mathbf{r}_k$ can be updated as follows:

$$\mathbf{r}_k^* = (\mathbf{r}_k - \mathbf{Y}_k)\frac{\omega_{kk}}{\omega_{kk}^*} + \mathbf{Y}_k \, .$$

Clearly, this update requires $O(n)$ operations. $\square$

I Proof of Lemma 5

Proof. **(CONCORD)** Let $A = nS$ Expanding the ℓ_2-norm of the residual, we have

$$\|\omega_{ii}\mathbf{Y}_i + \sum_{j\neq i}\omega_{ij}\mathbf{Y}_j\|_2^2 = \|\sum_{j=1}^{p}\omega_{ij}\mathbf{Y}_j\|_2^2 = \|\mathbf{Y}\omega_{i\bullet}\|_2^2 = \omega_{i\bullet}'\mathbf{Y}'\mathbf{Y}\omega_{i\bullet} = \omega_{i\bullet}'\mathbf{A}\omega_{i\bullet}$$

Hence, (12) is equivalent to

$$\begin{aligned}\mathcal{L}_{\text{con}}(\Omega) = \frac{1}{2}\sum_{i=1}^{p}(-2n\log\omega_{ii} + \omega_{i\bullet}'\mathbf{A}\omega_{i\bullet}) &= -n\sum_{i=1}^{p}\log\omega_{ii} + \frac{1}{2}\sum_{i=1}^{p}\omega_{i\bullet}'A\omega_{i\bullet} \\ &= -n\log\left(\prod_{i=1}^{p}\omega_{ii}\right) + \frac{n}{2}\operatorname{tr}(\Omega\mathbf{S}\Omega) \\ &= \frac{n}{2}\left(-\log\det\Omega_D^2 + \operatorname{tr}(\mathbf{S}\Omega^2)\right).\end{aligned}$$

Hence, $G_{\text{con}}(\Omega) = \Omega_D$ and $H_{\text{con}}(\Omega) = \Omega^2$

(SPACE with unit weights) Reparameterizing (13) using the identity $-\rho^{ij}\sqrt{\omega_{jj}/\omega_{ii}} = \omega_{ij}/\omega_{ii}$, the ℓ_2-norm of the residual can be expressed as follows.

$$\|\mathbf{Y}_i + \sum_{j\neq i}\frac{\omega_{ij}}{\omega_{ii}}\mathbf{Y}_j\|_2^2 = \|\frac{1}{\omega_{ii}}(\omega_{ii}\mathbf{Y}_i + \sum_{j\neq i}\omega_{ij}\mathbf{Y}_j)\|_2^2 = \frac{1}{\omega_{ii}^2}\omega_{i\bullet}'\mathbf{A}\omega_{i\bullet}$$

Hence, (13) is equivalent to

$$\begin{aligned}\mathcal{L}_{\text{spc},1}(\Omega) &= -\frac{n}{2}\log\det\Omega_D + \frac{1}{2}\sum_{i=1}^{p}\frac{1}{\omega_{ii}^2}\omega_{i\bullet}'\mathbf{A}\omega_{i\bullet} \\ &= -\frac{n}{2}\log\det\Omega_D + \frac{n}{2}\sum_{i=1}^{p}\frac{\omega_{i\bullet}'}{\omega_{ii}}\mathbf{S}\frac{\omega_{i\bullet}}{\omega_{ii}}\end{aligned}$$

$$
\begin{aligned}
&= -\frac{n}{2}\log\det\Omega_D + \frac{1}{2}\operatorname{tr}(\Omega_D^{-1}\Omega\mathbf{A}\Omega\Omega_D^{-1}) \\
&= \frac{n}{2}\left(-\log\det\Omega_D + \operatorname{tr}(\mathbf{S}\Omega\Omega_D^{-2}\Omega)\right).
\end{aligned}
$$

Therefore, $G_{\text{spc},1}(\Omega) = \Omega_D$ and $H_{\text{spc},1}(\Omega) = \Omega\Omega_D^{-2}\Omega$.

(SPACE with ω_{ii} weights) Similar to the analysis for SPACE1 with unit weights, the ℓ_2-norm of the residual for the SPACE2 formulation (i.e., with weights ω_{ii}) can be expressed as follows.

$$
\begin{aligned}
\omega_{ii}\|\mathbf{Y}_i - \sum_{j\neq i}\rho^{ij}\sqrt{\frac{\omega_{jj}}{\omega_{ii}}}\mathbf{Y}_j\|_2^2 &= \omega_{ii}\left(\frac{1}{\omega_{ii}^2}\omega_{i\bullet}'\mathbf{A}\omega_{i\bullet}\right) \\
&= \frac{1}{\omega_{ii}}\omega_{i\bullet}'\mathbf{A}\omega_{i\bullet}
\end{aligned}
$$

Hence, (14) is equivalent to

$$
\begin{aligned}
\mathcal{L}_{\text{spc},2}(\Omega) &= -\frac{n}{2}\log\det\Omega_D + \frac{1}{2}\sum_{i=1}^{p}\frac{1}{\omega_{ii}}\omega_{i\bullet}'\mathbf{A}\omega_{i\bullet} \\
&= -\frac{n}{2}\log\det\Omega_D + \frac{n}{2}\sum_{i=1}^{p}\frac{\omega_{i\bullet}'}{\sqrt{\omega_{ii}}}\mathbf{S}\frac{\omega_{i\bullet}}{\sqrt{\omega_{ii}}} \\
&= -\frac{n}{2}\log\det\Omega_D + \frac{n}{2}\operatorname{tr}(\Omega_D^{-1/2}\Omega\mathbf{S}\Omega\Omega_D^{-1/2}) \\
&= \frac{n}{2}\left(-\log\det\Omega_D + \operatorname{tr}(\mathbf{S}\Omega\Omega_D^{-1}\Omega)\right)
\end{aligned}
$$

Therefore, $G_{\text{spc},2}(\Omega) = \Omega_D$ and $H_{\text{spc},2}(\Omega) = \Omega\Omega_D^{-1}\Omega$.

(SYMLASSO) Reparameterizing (15) by $\alpha_{ii} = 1/\omega_{ii}$ and $-\rho^{ij}\sqrt{\omega_{jj}/\omega_{ii}} = \omega_{ij}/\omega_{ii}$ yields (14). It follows that $G_{\text{sym}}(\Omega) = \Omega_D$, $H_{\text{sym}}(\Omega) = \Omega\Omega_D^{-1}\Omega$.

(SPLICE) Reparameterizing (16) by $d_{ii}^2 = 1/\omega_{ii}$ and $\beta_{ij} = \rho^{ij}\sqrt{\omega_{jj}/\omega_{ii}}$ yields (14). It follows that $G_{\text{spl}}(\Omega) = \Omega_D$, $H_{\text{spl}}(\Omega) = \Omega\Omega_D^{-1}\Omega$. $\square$

J Effect of correction factor

Following steps similar to proof of Lemma 4, the update formulas for $\bar{Q}_{\text{con}}(\Omega) = \mathcal{L}_{\text{con}}(\Omega) + \lambda \sum_{i<j} |\omega_{ij}|$ of (12) can be shown to be

$$(T_{kk}(\Omega))_{kk} = \frac{-\sum_{j\neq k} \omega_{kj} s_{kj} + \sqrt{\left(\sum_{j\neq k} \omega_{kj} s_{kj}\right)^2 + 2 s_{kk}}}{2 s_{kk}} \tag{28}$$

$$(T_{kl}(\Omega))_{kl} = \frac{S_{\frac{\lambda}{n}}\left(-\left(\sum_{j\neq l} \omega_{kj} s_{jl} + \sum_{j\neq k} \omega_{lj} s_{jk}\right)\right)}{s_{kk} + s_{ll}} \tag{29}$$

J.1 Numerical example

Analysis on a dataset ($n = 1000$) generated from following Ω was used for this example.

$$\Omega = \begin{pmatrix} 1.0 & 0.3 & 0.0 \\ 0.3 & 1.0 & 0.3 \\ 0.0 & 0.3 & 1.0 \end{pmatrix}$$

Without penalty, i.e. $\lambda = 0$, computed solutions Ω_{con} from using CONCORD and $\Omega_{\text{uncorrected}}$ from using update formulas (28) and (29) are

$$\Omega_{\text{uncorrected}} = \begin{pmatrix} 0.675 & 0.089 & -0.015 \\ 0.089 & 0.658 & 0.117 \\ -0.015 & 0.117 & 0.668 \end{pmatrix}, \quad \Omega_{\text{con}} = \begin{pmatrix} 0.974 & 0.257 & 0.007 \\ 0.257 & 0.983 & 0.344 \\ 0.007 & 0.344 & 0.978 \end{pmatrix}$$

It is clear that the estimate Ω_{con} with the correction factor performs better parameter estimation.

K Proof of Theorem 1

Khare and Rajaratnam (2014) establish convergence of the cyclic coordinatewise minimization algorithm for a general class of objective functions. The proof of convergence for CONCORD relies on showing that the corresponding objective function is a special case of the general class of objective functions considered in Khare and Rajaratnam (2014). A more detailed version of the following argument can be found in (Khare and Rajaratnam, 2014, Section 4.1). We provide the main steps here for convenience and completeness.

Let $\mathbf{y} = \mathbf{y}(\Omega) \in \mathbb{R}^{p^2}$ denote a vectorized version of Ω obtained by shifting the corresponding diagonal entry at the bottom of each column of Ω, and then stacking the columns on top of each other. Let P^i denote the $p \times p$ permutation matrix such that $P^i \mathbf{z} = (z_1, \cdots, z_{i-1}, z_{i+1}, \cdots, z_p, z_i)$ for every $\mathbf{z} \in \mathbb{R}^p$. It follows by the definition of $\mathbf{y}$ that

$$\mathbf{y} = \mathbf{y}(\Omega) = ((P^1 \Omega_{\cdot 1})^T, (P^2 \Omega_{\cdot 2})^T, \cdots, (P^p \Omega_{\cdot p})^T)^T.$$

Let $\mathbf{x} = \mathbf{x}(\Omega) \in \mathbb{R}^{\frac{p(p+1)}{2}}$ be the symmetric version of $\mathbf{y}$, obtained by removing all ω_{ij} with $i > j$ from $\mathbf{y}$. More precisely,

$$\mathbf{x} = \mathbf{x}(\Omega) = (\omega_{11}, \omega_{12}, \omega_{22}, \cdots, \omega_{1p}, \omega_{2p}, \cdots, \omega_{pp})^T.$$

Let $\tilde{P}$ be the $p^2 \times \frac{p(p+1)}{2}$ matrix such that every entry of $\tilde{P}$ is either 0 or 1, exactly one entry in each row of $\tilde{P}$ is equal to 1, and $\mathbf{y} = \tilde{P}\mathbf{x}$. Let $\tilde{S}$ be a $p^2 \times p^2$ block diagonal matrix with p diagonal blocks, and the i^{th} diagonal block is equal to $\tilde{S}^i := \frac{1}{2} P^i S (P^i)^T$, where $S = \frac{1}{n}\mathbf{Y}^T\mathbf{Y}$. It follows that

$$\begin{aligned} \frac{1}{2}\sum_{i=1}^p \Omega_{\cdot i}^T S \Omega_{\cdot i} = \frac{1}{2}\sum_{i=1}^p \Omega_{\cdot i}^T (P^i)^T P^i S (P^i)^T P^i \Omega_{\cdot i} &= \frac{1}{2}\sum_{i=1}^p (P^i \Omega_{\cdot i})^T (P^i S (P^i)^T)(P^i \Omega_{\cdot i}) \\ &= \mathbf{y}^T \tilde{S} \mathbf{y} \\ &= \mathbf{x}^T \tilde{P}^T \tilde{S} \tilde{P} \mathbf{x}. \end{aligned} \tag{30}$$

Note that for every $1 \leq i \leq p$, the matrix $\tilde{S}^i = \frac{1}{2} P^i S (P^i)^T$ is positive semi-definite. Let $\tilde{S}^{1/2}$ denote the $p^2 \times p^2$ block diagonal matrix with p diagonal blocks, such that the i^{th} diagonal block is given by $(\tilde{S}^i)^{1/2}$. Let $E = \tilde{S}^{1/2}\tilde{P}$. It follows by (30) that

$$\frac{1}{2}\sum_{i=1}^p \Omega_{\cdot i}^T S \Omega_{\cdot i} = (E\mathbf{x})^T (E\mathbf{x}). \tag{31}$$

By the definition of $\mathbf{x}(\Omega)$, we obtain

$$\omega_{ii} = x_{\frac{i(i+1)}{2}} \tag{32}$$

for every $1 \leq i \leq p$. Let

$$S_0 = \left\{ j : \ 1 \leq j \leq \frac{p(p+1)}{2}, \ j \neq \frac{i(i+1)}{2} \text{ for any } 1 \leq i \leq p \right\},$$

and

$$\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^{\frac{p(p+1)}{2}} : x_j \geq 0 \text{ for every } j \in S_0^c\}.$$

It follows by (8), (31) and (32) that the CONCORD algorithm can be viewed as a cyclic coordinatewise minimization algorithm for minimizing the function

$$Q_{con}(\mathbf{x}) = n \left\{ \mathbf{x}^T E^T E \mathbf{x} - \sum_{i \in S_0^c} \log x_i + \frac{\lambda}{n} \sum_{j \in S_0} |x_j| \right\}, \tag{33}$$

subject to $\mathbf{x} \in \mathcal{X}$. For every $1 \leq i \leq p(p+1)/2$, there exist $1 \leq k, l \leq p$ such that $x_i = \omega_{kl}$. Note that $\|E_{\cdot i}\|^2 = \frac{S_{kk}+S_{ll}}{2} > 0$. It also follows from (Khare and Rajaratnam, 2014, Lemma 4.1) that for every $\xi \in \mathbb{R}$, the set $R_\xi := \{\mathbf{x} \in \mathcal{X} : Q_{con}(\mathbf{x}) \leq \xi\}$ is bounded in the sense that for every $i \in S_0$, x_i is uniformly bounded above and below, and for every $i \in S_0^c$, x_i is uniformly bounded above and below (from zero). It follows by (Khare and Rajaratnam, 2014, Theorem 3.1) that the sequence of iterates produced by the CONCORD algorithm converges.

L Application to breast cancer data

Gene Symbol	CONCORD	SYMLASSO	SPACE1	SPACE2	Reference
HNF3A (FOXA1)	+	+	+	+	Koboldt and Others (2012), Albergaria et al. (2009), Davidson et al. (2011), Lacroix and Leclercq (2004), Robinson et al. (2011)
TONDU	+	+	+	+	
FZD9	+	+	+	+	Katoh (2008), Rønneberg et al. (2011)
KIAA0481	+	+	+	+	[Gene record discontinued]
KRT16	+	+	+		Glinsky et al. (2005), Joosse et al. (2012), Pellegrino et al. (1988)
KNSL6 (KIF2C)	+			+	Eschenbrenner et al. (2011), Shimo et al. (2007, 2008)
FOXC1	+	+	+	+	Du et al. (2012), Sizemore and Keri (2012), Wang et al. (2012), Ray et al. (2011), Tkocz et al. (2012)
PSA	+	+		+	Kraus et al. (2010), Mohajeri et al. (2011), Sauter et al. (2004), Yang et al. (2002)
GATA3	+	+	+	+	Koboldt and Others (2012), Davidson et al. (2011), Albergaria et al. (2009), Eeckhoute et al. (2007), Jiang et al. (2010), Licata et al. (2010), Yan et al. (2010)
C20ORF1 (TPX2)	+				Maxwell and Others (2011), Bibby et al. (2009)
E48		+	+	+	
ESR1				+	Zheng et al. (2012)

Table 6: Summary of the top hub genes identified by each of the four methods, CONCORD, SYMLASSO, SPACE1 & SPACE2: Genes indicated by '+' denote the 10 most highly connected genes for each of the methods. References are provided at the end of this supplemental section.

M Application to portfolio optimization

M.1 Constituents of Dow Jones Industrial Average

Symbol	Description	Return (%)	Risk (%)	SR
AA	Alcoa Inc.	9.593	41.970	0.109
AXP	American Express Company	18.706	38.913	0.352
BA	The Boeing Company	13.417	32.685	0.258
BAC	Bank of America Corporation	13.182	48.588	0.168
CAT	Caterpillar Inc.	19.042	35.050	0.401
CSCO	Cisco Systems, Inc.	22.650	44.565	0.396
CVX	Chevron Corporation	15.486	26.716	0.392
DD	E. I. du Pont de Nemours and Company	10.591	30.537	0.183
DIS	The Walt Disney Company	12.312	32.800	0.223
GE	General Electric Company	12.449	31.667	0.235
HD	The Home Depot, Inc.	17.266	34.422	0.356
HPQ	Hewlett-Packard Company	10.769	40.727	0.142
IBM	International Business Machines Corporation	18.715	29.944	0.458
INTC	Intel Corporation	18.325	41.543	0.321
JNJ	Johnson & Johnson	13.664	22.087	0.392
JPM	JPMorgan Chase & Co.	18.292	42.729	0.311
KO	The Coca-Cola Company	10.617	24.092	0.233
MCD	McDonald's Corp.	14.457	26.114	0.362
MMM	3M Company	12.596	25.353	0.300
MRK	Merck & Co. Inc.	12.385	29.616	0.249
MSFT	Microsoft Corporation	18.612	33.904	0.401
PFE	Pfizer Inc.	14.376	29.060	0.323
PG	Procter & Gamble Co.	13.262	24.241	0.341
T	AT&T, Inc.	11.231	28.781	0.217
TRV	The Travelers Companies, Inc.	14.726	31.706	0.307
UTX	United Technologies Corp.	18.618	28.760	0.474
VZ	Verizon Communications Inc.	11.403	27.728	0.231
WMT	Wal-Mart Stores Inc.	15.495	27.955	0.375
XOM	Exxon Mobil Corporation	15.466	25.764	0.406

Table 7: Dow Jones Industrial Average component stocks and their respective realized returns, realized risk and Sharpe ratios. The risk-free rate is set at 5%.

M.2 Investment periods

k	Date Range	k	Date Range	k	Date Range	k	Date Range
1	95/02/18-95/03/17	59	99/07/31-99/08/27	117	04/01/10-04/02/06	175	08/06/21-08/07/18
2	95/03/18-95/04/14	60	99/08/28-99/09/24	118	04/02/07-04/03/05	176	08/07/19-08/08/15
3	95/04/15-95/05/12	61	99/09/25-99/10/22	119	04/03/06-04/04/02	177	08/08/16-08/09/12
4	95/05/13-95/06/09	62	99/10/23-99/11/19	120	04/04/03-04/04/30	178	08/09/13-08/10/10
5	95/06/10-95/07/07	63	99/11/20-99/12/17	121	04/05/01-04/05/28	179	08/10/11-08/11/07
6	95/07/08-95/08/04	64	99/12/18-00/01/14	122	04/05/29-04/06/25	180	08/11/08-08/12/05
7	95/08/05-95/09/01	65	00/01/15-00/02/11	123	04/06/26-04/07/23	181	08/12/06-09/01/02
8	95/09/02-95/09/29	66	00/02/12-00/03/10	124	04/07/24-04/08/20	182	09/01/03-09/01/30
9	95/09/30-95/10/27	67	00/03/11-00/04/07	125	04/08/21-04/09/17	183	09/01/31-09/02/27
10	95/10/28-95/11/24	68	00/04/08-00/05/05	126	04/09/18-04/10/15	184	09/02/28-09/03/27
11	95/11/25-95/12/22	69	00/05/06-00/06/02	127	04/10/16-04/11/12	185	09/03/28-09/04/24
12	95/12/23-96/01/19	70	00/06/03-00/06/30	128	04/11/13-04/12/10	186	09/04/25-09/05/22
13	96/01/20-96/02/16	71	00/07/01-00/07/28	129	04/12/11-05/01/07	187	09/05/23-09/06/19
14	96/02/17-96/03/15	72	00/07/29-00/08/25	130	05/01/08-05/02/04	188	09/06/20-09/07/17
15	96/03/16-96/04/12	73	00/08/26-00/09/22	131	05/02/05-05/03/04	189	09/07/18-09/08/14
16	96/04/13-96/05/10	74	00/09/23-00/10/20	132	05/03/05-05/04/01	190	09/08/15-09/09/11
17	96/05/11-96/06/07	75	00/10/21-00/11/17	133	05/04/02-05/04/29	191	09/09/12-09/10/09
18	96/06/08-96/07/05	76	00/11/18-00/12/15	134	05/04/30-05/05/27	192	09/10/10-09/11/06
19	96/07/06-96/08/02	77	00/12/16-01/01/12	135	05/05/28-05/06/24	193	09/11/07-09/12/04
20	96/08/03-96/08/30	78	01/01/13-01/02/09	136	05/06/25-05/07/22	194	09/12/05-10/01/01
21	96/08/31-96/09/27	79	01/02/10-01/03/09	137	05/07/23-05/08/19	195	10/01/02-10/01/29
22	96/09/28-96/10/25	80	01/03/10-01/04/06	138	05/08/20-05/09/16	196	10/01/30-10/02/26
23	96/10/26-96/11/22	81	01/04/07-01/05/04	139	05/09/17-05/10/14	197	10/02/27-10/03/26
24	96/11/23-96/12/20	82	01/05/05-01/06/01	140	05/10/15-05/11/11	198	10/03/27-10/04/23
25	96/12/21-97/01/17	83	01/06/02-01/06/29	141	05/11/12-05/12/09	199	10/04/24-10/05/21
26	97/01/18-97/02/14	84	01/06/30-01/07/27	142	05/12/10-06/01/06	200	10/05/22-10/06/18
27	97/02/15-97/03/14	85	01/07/28-01/08/24	143	06/01/07-06/02/03	201	10/06/19-10/07/16
28	97/03/15-97/04/11	86	01/08/25-01/09/21	144	06/02/04-06/03/03	202	10/07/17-10/08/13
29	97/04/12-97/05/09	87	01/09/22-01/10/19	145	06/03/04-06/03/31	203	10/08/14-10/09/10
30	97/05/10-97/06/06	88	01/10/20-01/11/16	146	06/04/01-06/04/28	204	10/09/11-10/10/08
31	97/06/07-97/07/04	89	01/11/17-01/12/14	147	06/04/29-06/05/26	205	10/10/09-10/11/05
32	97/07/05-97/08/01	90	01/12/15-02/01/11	148	06/05/27-06/06/23	206	10/11/06-10/12/03
33	97/08/02-97/08/29	91	02/01/12-02/02/08	149	06/06/24-06/07/21	207	10/12/04-10/12/31
34	97/08/30-97/09/26	92	02/02/09-02/03/08	150	06/07/22-06/08/18	208	11/01/01-11/01/28
35	97/09/27-97/10/24	93	02/03/09-02/04/05	151	06/08/19-06/09/15	209	11/01/29-11/02/25
36	97/10/25-97/11/21	94	02/04/06-02/05/03	152	06/09/16-06/10/13	210	11/02/26-11/03/25
37	97/11/22-97/12/19	95	02/05/04-02/05/31	153	06/10/14-06/11/10	211	11/03/26-11/04/22
38	97/12/20-98/01/16	96	02/06/01-02/06/28	154	06/11/11-06/12/08	212	11/04/23-11/05/20
39	98/01/17-98/02/13	97	02/06/29-02/07/26	155	06/12/09-07/01/05	213	11/05/21-11/06/17
40	98/02/14-98/03/13	98	02/07/27-02/08/23	156	07/01/06-07/02/02	214	11/06/18-11/07/15
41	98/03/14-98/04/10	99	02/08/24-02/09/20	157	07/02/03-07/03/02	215	11/07/16-11/08/12
42	98/04/11-98/05/08	100	02/09/21-02/10/18	158	07/03/03-07/03/30	216	11/08/13-11/09/09
43	98/05/09-98/06/05	101	02/10/19-02/11/15	159	07/03/31-07/04/27	217	11/09/10-11/10/07
44	98/06/06-98/07/03	102	02/11/16-02/12/13	160	07/04/28-07/05/25	218	11/10/08-11/11/04
45	98/07/04-98/07/31	103	02/12/14-03/01/10	161	07/05/26-07/06/22	219	11/11/05-11/12/02
46	98/08/01-98/08/28	104	03/01/11-03/02/07	162	07/06/23-07/07/20	220	11/12/03-11/12/30
47	98/08/29-98/09/25	105	03/02/08-03/03/07	163	07/07/21-07/08/17	221	11/12/31-12/01/27
48	98/09/26-98/10/23	106	03/03/08-03/04/04	164	07/08/18-07/09/14	222	12/01/28-12/02/24
49	98/10/24-98/11/20	107	03/04/05-03/05/02	165	07/09/15-07/10/12	223	12/02/25-12/03/23
50	98/11/21-98/12/18	108	03/05/03-03/05/30	166	07/10/13-07/11/09	224	12/03/24-12/04/20
51	98/12/19-99/01/15	109	03/05/31-03/06/27	167	07/11/10-07/12/07	225	12/04/21-12/05/18
52	99/01/16-99/02/12	110	03/06/28-03/07/25	168	07/12/08-08/01/04	226	12/05/19-12/06/15
53	99/02/13-99/03/12	111	03/07/26-03/08/22	169	08/01/05-08/02/01	227	12/06/16-12/07/13
54	99/03/13-99/04/09	112	03/08/23-03/09/19	170	08/02/02-08/02/29	228	12/07/14-12/08/10
55	99/04/10-99/05/07	113	03/09/20-03/10/17	171	08/03/01-08/03/28	229	12/08/11-12/09/07
56	99/05/08-99/06/04	114	03/10/18-03/11/14	172	08/03/29-08/04/25	230	12/09/08-12/10/05
57	99/06/05-99/07/02	115	03/11/15-03/12/12	173	08/04/26-08/05/23	231	12/10/06-12/10/26
58	99/07/03-99/07/30	116	03/12/13-04/01/09	174	08/05/24-08/06/20		

Table 8: Investment periods in YY/MM/DD format

M.3 Details of minimum variance portfolio rebalancing

The investment period during which a set of portfolio weights are held constant is also referred to as the "holding period". The number of trading days in the k-th investment period, L_k, may vary if rebalancing time points are chosen to coincide with either calendar months, weeks or fiscal quarters. Let t index the number of an arbitrary day during the entire investment horizon. The number of trading days T_j in the first j investment periods is given by

$$T_j = \sum_{k=1}^{j} L_k, \tag{34}$$

where $j = 1, 2, \ldots, K$ with $T_0 = 0$. We consider holding N_{est} constant for all investment periods, $k = 1, 2, \ldots$. For convenience, denote by k_t the investment period that trading day t belongs to: i.e., $k_t = k(t) := \{k : t \in [T_{k-1}, T_k]\}$.

The algorithm for the minimum variance portfolio rebalancing strategy (MVR) can now be described as follows: At the beginning of time period k, that is after T_{k-1} days, compute an estimate of the covariance matrix $\widehat{\Sigma}_k$ for period k from N_{est} past returns: i.e., $\{r_t : t \in [T_{k-1} - N_{\text{est}} + 1, T_{k-1}]\}$. Then, compute a new set of portfolio weights $w_k = (\mathbf{1}^T \widehat{\Sigma}_k^{-1} \mathbf{1})^{-1} \widehat{\Sigma}_k^{-1} \mathbf{1}$, and hold this portfolio constant until the T_k-th trading day. The process is then repeated for the next holding period.

M.4 Details of cross-validation

Consider the matrix of returns $\mathbf{R}$ for all the stocks in the portfolio in the estimation horizon preceding the start of the investment period $(k-1)$.

$$\mathbf{R} = ((r_{ti})), \text{ where } i \in \{1, \ldots, p\}, \ t \in \{T_{k-1} - N_{\text{est}} + 1, \ldots, T_{k-1}\}.$$

Hence, $\mathbf{R}$ is an N_{est}-by-p matrix, and the column vector $\mathbf{R}_j$ is an N_{est}-vector of returns for the j-th stock.

Now denote by $\Omega(\lambda) = ((\omega_{ij}(\lambda)))_{1 \le i,j \le p}$ an estimate of Ω obtained by ℓ_1-regularization methods such as Glasso or CONCORD. The use of λ makes explicit the dependence of these estimation methods on the penalty parameter λ. The data are the over the estimation horizon is divided into m-folds. The penalty parameter is chosen so as to minimize the out of sample predictive risk (PR) given by

$$PR(\lambda) = \sum_{m=1}^{M} \left\{ \frac{1}{N_m} \sum_{i=1}^{p} \|\mathbf{R}_i^{(m)} - \sum_{j \neq i} \beta_{ij}^{(\backslash m)}(\lambda) \mathbf{R}_j^{(m)}\|_2^2 \right\},$$

where $\mathbf{R}_i^{(m)}$ is the vector of returns for stock i in fold m, and where N_m is the number of observations in the m-th fold. The regression coefficient $\beta_{ij}^{(\backslash m)}(\lambda)$ is determined as follows: $\beta_{ij}^{(\backslash m)}(\lambda) = -\frac{\omega_{ij}^{(\backslash m)}(\lambda)}{\omega_{ii}^{(\backslash m)}(\lambda)}$, with $\Omega^{(\backslash m)}(\lambda)$ based on using all the available data within a given estimation horizon except for fold m. The optimal choice of penalty parameter λ^* is then determined as follows:

$$\lambda^* = \arg \inf_{\lambda \geq 0} PR(\lambda).$$

M.5 Performance metrics

For comparison purposes with (Won et al., 2012), we use the following quantities to assess the performance of the five MVR strategies. The formulas for these metrics are given below.

- *Realized return*: The average return of the portfolio over the entire investment horizon.

$$r_p = \frac{1}{T} \sum_{t=1}^{T} r_t' w_{k_t}$$

- *Realized risk*: The risk (standard error) of the portfolio over the entire investment horizon.

$$\sigma_p = \left[\frac{1}{T} \sum_{t=1}^{T} (r_t' w_{k_t} - r_p)^2 \right]^{1/2}$$

- *Realized Sharpe ratio (SR)*: The realized excess return of the portfolio over the risk-free rate per unit realized risk for the entire investment horizon.

$$SR = \frac{r_p - r_f}{\sigma_p} \tag{35}$$

- *Turnover*: The amount of new portfolio assets purchased or sold over each trading period. The turnover for the k-th investment period when the portfolio weights w_k are held constant is given by

$$TO(k) = \sum_{i=1}^{p} \left| w_{ik} - \left(\prod_{t=T_{k-1}+1}^{T_{k-1}+L_k} (1 + r_{it}) \right) w_{i(k-1)} \right| \tag{36}$$

with $w_{i0} = 0$ for all $i = 1, \dots, p$.

- *Size of the short side* The proportion of the negative weights to the sum of the absolute weights of each portfolio. The short side for the k-th investment period is given by

$$SS(k) = \frac{\sum_{i=1}^{p} |\min(w_{ik}, 0)|}{\sum_{i=1}^{p} |w_{ik}|}$$

The average and standard error of the short sides over the all investment periods is

$$\overline{SS} = \frac{1}{K} \sum_{k=1}^{K} SS(k), \quad \hat{\sigma}_{SS} = \left[\frac{1}{K} \sum_{k=1}^{K} (SS(k) - \overline{SS})^2 \right]^{1/2}$$

- *Normalized wealth growth*: Accumulated wealth derived from the portfolio over the trading period when the initial budget is normalized to one. Note that both transaction costs and borrowing costs are taken into account. Let $W(t-1)$ denote the wealth of the portfolio after the $(t-1)$-th trading day. Then, the wealth of the portfolio after the t-th trading day is given by

$$W(t) = \begin{cases} W(t-1)\left(1 + r_t' w_{k_t} - TC(k_t) - BC(k_t)\right), & t = T_{k_t-1} + 1 \\ W(t-1)\left(1 + r_t' w_{k_t}\right), & t \neq T_{k_t-1} + 1 \end{cases},$$

where $TC(k)$ and $BC(k)$ are transaction costs (of trading stocks) and borrowing costs (of capital for taking short positions on stocks), respectively. On the first day of each trading period, we adjust the return for these trading costs. Denote the transaction cost rate by r_c, then the transaction cost incurred at the beginning of period k is given by

$$TC(k) = r_c \cdot TO(k). \tag{37}$$

The borrowing cost rate, $BC(k)$, depends on the short side of the portfolio weights during the $(k-1)$-th period. Denote the borrowing daily percentage by r_b, then the

N_{est}	Sample	Glasso	CONCORD	CondReg	LedoitWolf	DJIA
35	**17.08** (33.86)	13.10 (16.57)	13.29 (17.04)	13.62 (17.74)	12.33 (**15.58**)	8.51 (18.96)
40	**16.66** (26.52)	13.13 (16.57)	13.34 (17.02)	13.39 (17.74)	11.78 (**15.46**)	8.51 (18.96)
45	11.13 (23.19)	12.74 (16.52)	**13.05** (17.04)	**13.05** (17.77)	10.99 (**15.43**)	8.51 (18.96)
50	9.90 (20.95)	12.89 (16.39)	**13.21** (17.04)	13.08 (17.65)	11.25 (**15.36**)	8.51 (18.96)
75	11.61 (17.45)	11.28 (15.57)	**13.10** (17.04)	12.77 (17.15)	10.56 (**15.10**)	8.51 (18.96)
150	9.40 (15.41)	10.28 (14.97)	**13.20** (17.08)	12.76 (16.30)	10.63 (**14.66**)	8.51 (18.96)
225	10.49 (14.98)	10.38 (14.89)	**13.58** (17.10)	12.92 (16.04)	11.04 (**14.52**)	8.51 (18.96)
300	10.41 (14.95)	10.37 (14.95)	**13.66** (17.16)	12.85 (16.07)	10.94 (**14.52**)	8.51 (18.96)

Table 9: Realized returns of different investment strategies corresponding to different estimators with various N_{est} (realized risks are given in parentheses). The maximum annualized returns and risks are highlighted in bold.

N_{est}	Sample	Glasso	CONCORD	CondReg	LedoitWolf
35	8.42 (3.19)	0.45 (0.12)	**0.38** (**0.10**)	0.39 (0.27)	1.40 (0.38)
40	5.81 (2.28)	0.41 (0.12)	**0.34** (**0.10**)	0.37 (0.26)	1.29 (0.36)
45	4.58 (1.65)	0.39 (0.12)	**0.31** (**0.10**)	0.36 (0.23)	1.20 (0.35)
50	3.74 (1.19)	0.39 (0.13)	**0.28** (**0.09**)	0.36 (0.25)	1.11 (0.33)
75	2.03 (0.67)	0.50 (0.19)	**0.21** (**0.08**)	0.43 (0.29)	0.86 (0.29)
150	0.87 (0.32)	0.73 (0.27)	**0.14** (**0.07**)	0.40 (0.22)	0.54 (0.23)
225	0.57 (0.24)	0.56 (0.22)	**0.11** (**0.07**)	0.31 (0.13)	0.41 (0.18)
300	0.44 (0.21)	0.44 (0.23)	**0.09** (**0.07**)	0.24 (0.11)	0.33 (0.17)

Table 10: Average turnovers for various estimation horizons, N_{est} (standard errors are given in parentheses). The minimum average and standard error values for each row are highlighted in bold.

borrowing cost rate is given by

$$BC(k) = ((1 + r_b)^{L_{k-1}} - 1) \sum_{i=1}^{p} |\min(w_{i(k-1)}, 0)|. \tag{38}$$

N Proof of Theorem 2

The result follows by noting the following straightforward facts

1. The existence of a minimizer follows by the convexity of Q_{con}.

2. By assumptions (A0) and (A1), for any $\eta > 0$, $\{\widehat{\alpha}_{n,ii}\}_{1 \leq i \leq p_n}$ are uniformly bounded away from zero and infinity with probability larger than $1 - O(n^{-\eta})$.

3. When the diagonal entries are fixed at $\{\widehat{\alpha}_{n,ii}\}_{1 \leq i \leq p_n}$, then the objective function Q_{con} (reparameterized from ω^o to θ) is same as the objective function of SPACE with

N_{est}	Sample	Glasso	CONCORD	CondReg	LedoitWolf
35	41.13 (3.18)	0.66 (0.84)	**0.05** (**0.14**)	1.75 (5.00)	20.50 (6.64)
40	38.64 (3.47)	0.64 (0.75)	**0.05** (**0.14**)	1.78 (5.04)	20.45 (6.63)
45	36.89 (4.26)	0.90 (0.85)	**0.05** (**0.14**)	1.84 (4.95)	20.31 (6.61)
50	35.46 (4.38)	1.35 (1.19)	**0.04** (**0.11**)	2.17 (5.44)	20.33 (6.66)
75	30.89 (5.37)	8.67 (3.76)	**0.04** (**0.11**)	4.91 (7.38)	20.13 (6.83)
150	25.65 (6.25)	23.48 (4.68)	**0.02** (**0.07**)	9.07 (6.31)	19.60 (6.82)
225	23.68 (6.69)	23.36 (6.27)	**0.01** (**0.05**)	10.71 (3.22)	19.26 (6.91)
300	22.45 (6.90)	22.42 (6.87)	**0.00** (**0.02**)	9.95 (2.93)	18.85 (7.10)

Table 11: Average short sides for various estimation horizons, N_{est} (standard errors are given in parentheses). The minimum average and standard error values for each row are highlighted in bold.

N_{est}	Sample	Glasso	CONCORD	CondReg	LedoitWolf
35	567.958 (214.05)	22.635 (5.62)	**18.642** (**4.53**)	20.757 (17.46)	91.316 (25.19)
40	394.508 (149.90)	20.660 (5.70)	**16.858** (**4.40**)	20.013 (16.78)	85.661 (24.16)
45	315.340 (108.87)	19.899 (5.80)	**15.470** (**4.22**)	19.419 (15.27)	80.524 (23.39)
50	260.887 (81.13)	20.146 (6.39)	**14.081** (**4.06**)	19.695 (16.04)	76.154 (22.43)
75	150.242 (45.87)	30.942 (10.92)	**10.516** (**3.17**)	25.191 (19.19)	63.481 (20.94)
150	75.700 (27.88)	61.495 (18.40)	**6.596** (**2.24**)	26.788 (12.83)	46.680 (17.78)
225	56.242 (22.09)	54.117 (18.82)	**5.155** (**1.80**)	22.973 (6.08)	39.441 (15.72)
300	46.904 (20.09)	47.118 (20.72)	**4.404** (**1.67**)	18.823 (5.16)	35.065 (14.89)

Table 12: Average trading costs in basis points for various estimation horizons, N_{est} (standard errors are given in parentheses). Borrowing rate is taken to be 7% APR and transaction cost rate is taken to be 0.5% of principal for each transaction. The minimum transaction cost for each row is highlighted in bold.

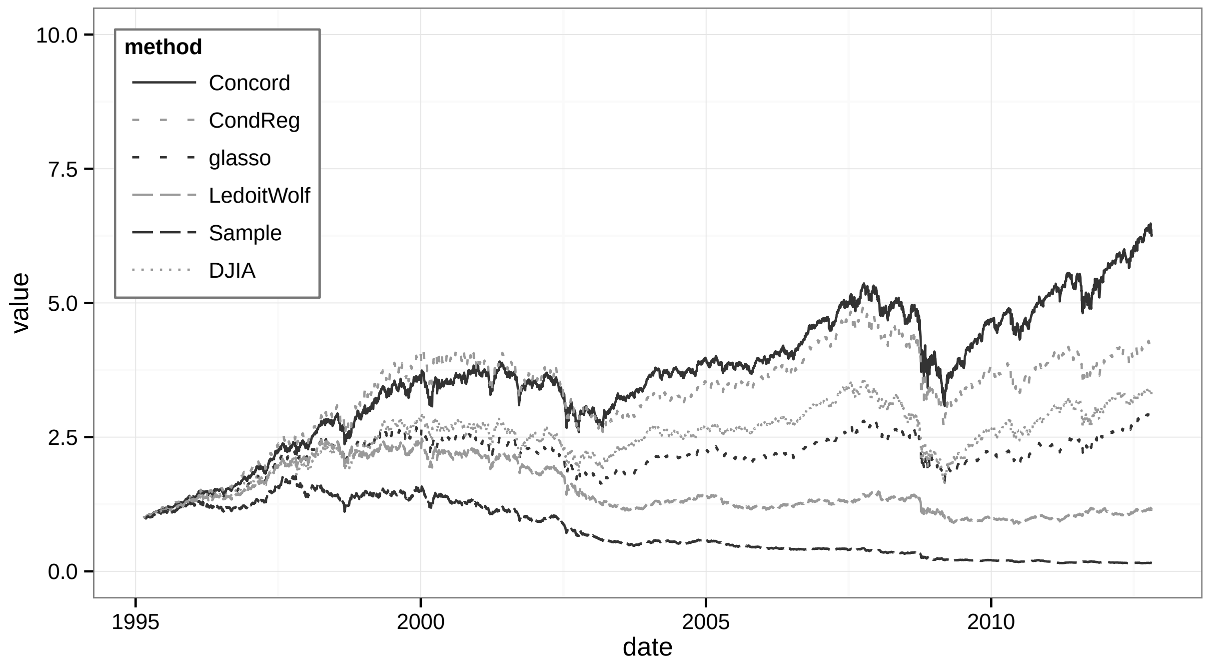


Figure 3: Normalized wealth growth after adjusting for transaction costs (0.5% of principal) and borrowing costs (interest rate of 7% APR) with $N_{\text{est}} = 75$.

weights $w_i = \widehat{\alpha}^2_{n,ii}$ (which are uniformly bounded), except that the penalty term is now $\sum_{1\le i<j\le p_n} \lambda_n \sqrt{\widehat{\alpha}_{n,ii}\widehat{\alpha}_{n,jj}}\theta_{ij}$, instead of $\sum_{1\le i<j\le p_n} \lambda_n\theta_{ij}$ as in Q_{spc}.

4. Since $\bar{\theta}_{n,ij} = \frac{\bar{\omega}_{n,ij}}{\sqrt{\widehat{\alpha}_{n,ii}\widehat{\alpha}_{n,jj}}}$, using the uniform boundedness of $\{\widehat{\alpha}_{n,ii}\}_{1\le i\le p_n}$, there exists a constant C_1 such that for any $\eta > 0$,

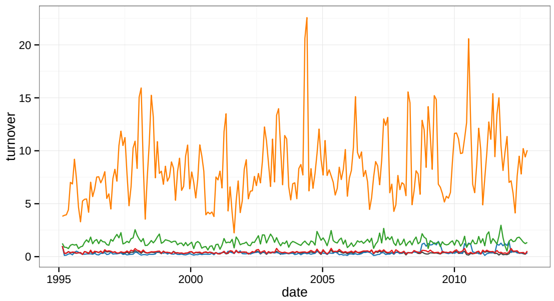
$$\|\widehat{\omega}^o_n - \bar{\omega}^o_n\|_2 \le C_1\|\widehat{\theta}^o_n - \bar{\theta}^o_n\|_2$$

holds with probability larger than $1 - O(n^{-\eta})$.

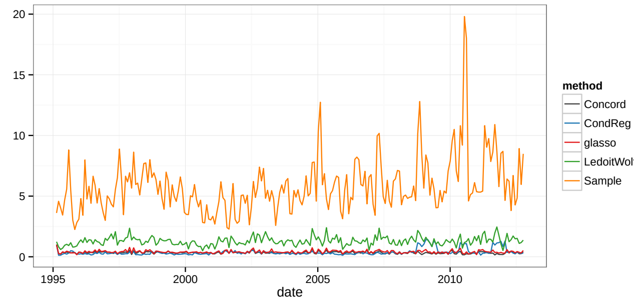
5. For $1 \le i < j \le p_n$, $\text{sign}(\hat{\omega}_{n,ij}) = \text{sign}(\hat{\theta}_{n,ij})$, since they differ by a positive multiplicative constant.

6. When the penalty term in SPACE is replaced by $\sum_{1\le i<j\le p_n} \lambda_n \sqrt{\widehat{\alpha}_{n,ii}\widehat{\alpha}_{n,jj}}\theta_{ij}$, the uniform boundedness of $\{\widehat{\alpha}_{n,ii}\}_{1\le i\le p_n}$ implies that Theorems 1, 2 and 3 of Peng et al. (2009) hold with trivial modifications at appropriate places. The result now follows immediately using these theorems along with the above assertions. $\square$

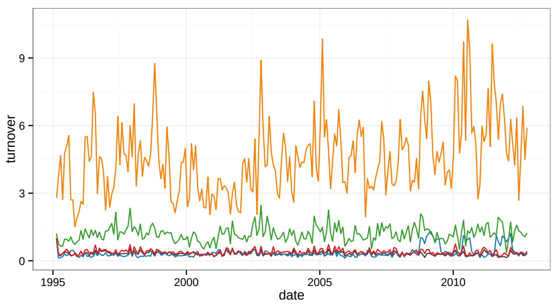
*R*emark: Note that Theorem 2 on the consistency of CONCORD has been formulated as to exactly parallel the result given for SPACE by Peng et al. (2009). An accurate estimator



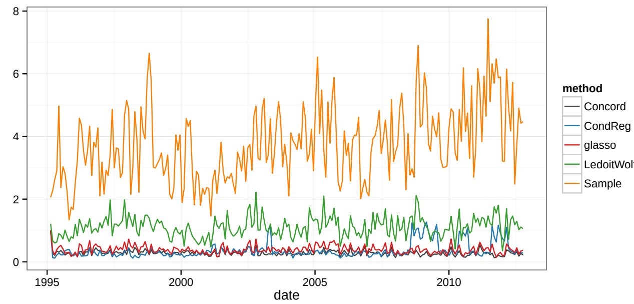
(a) $N_{\text{est}} = 35$



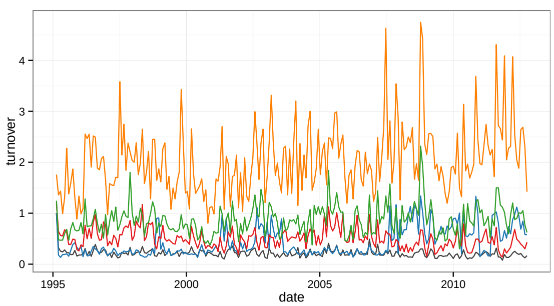
(b) $N_{\text{est}} = 40$



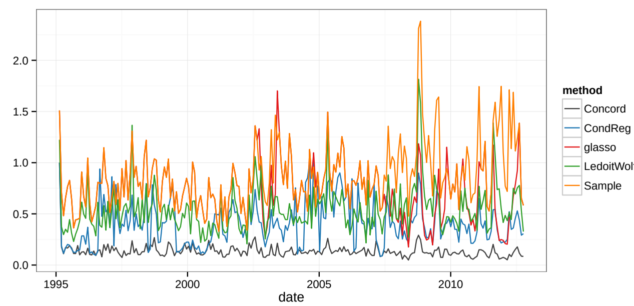
(c) $N_{\text{est}} = 45$



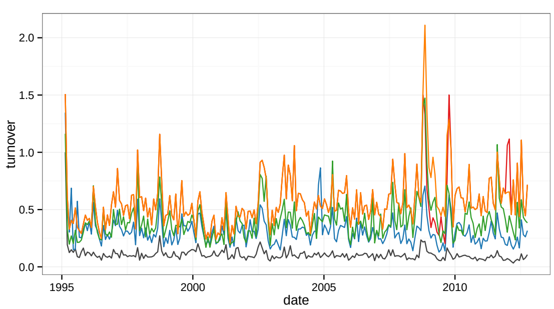
(d) $N_{\text{est}} = 50$



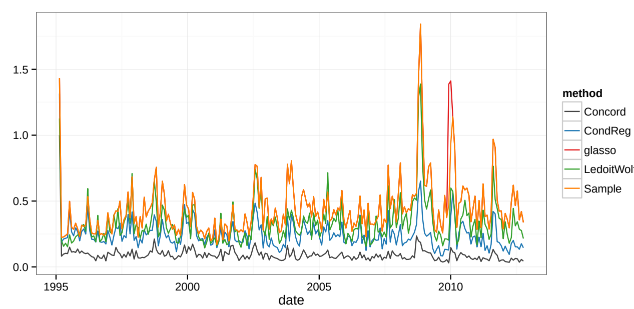
(e) $N_{\text{est}} = 75$



(f) $N_{\text{est}} = 150$

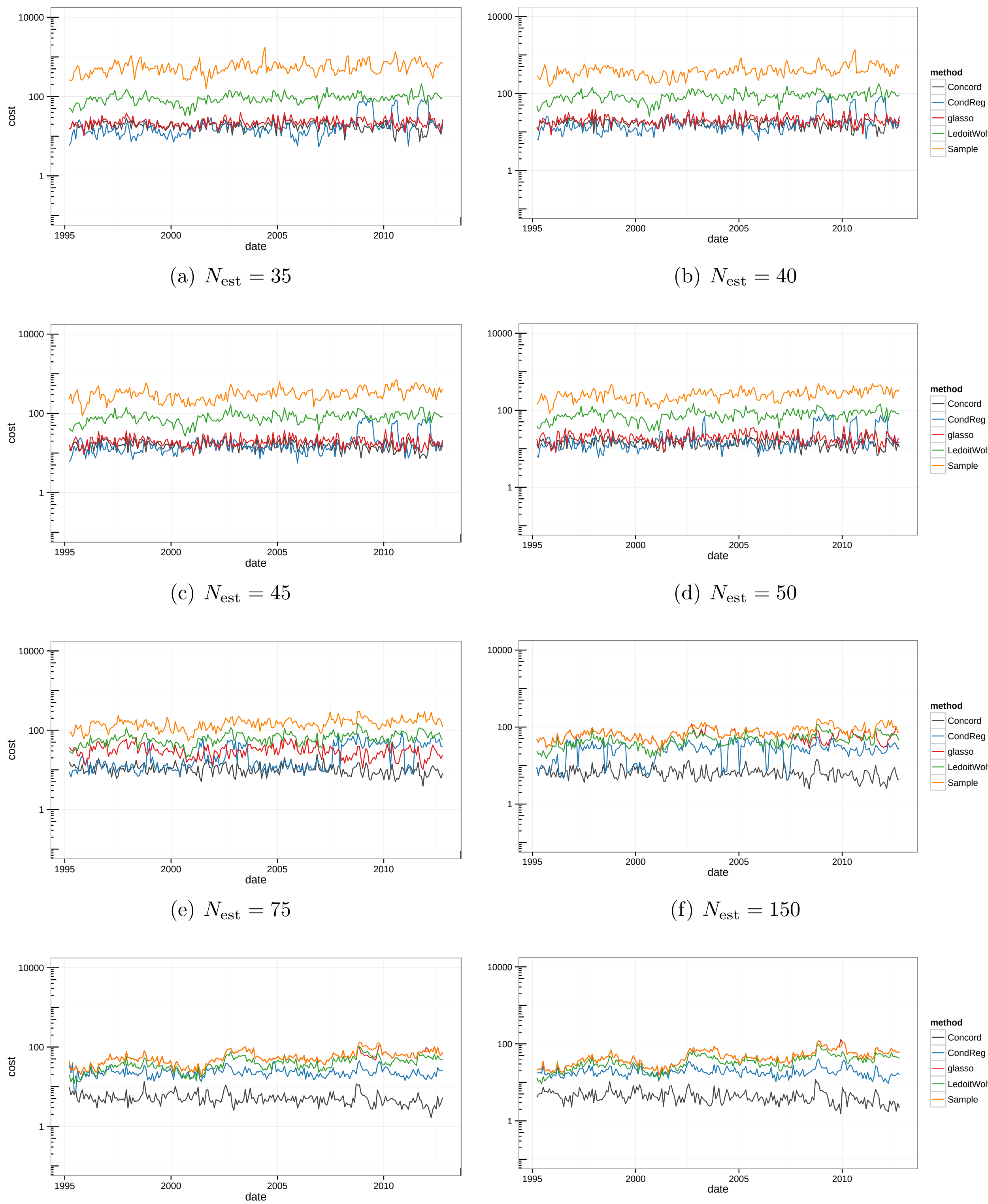


(g) $N_{\text{est}} = 225$



(h) $N_{\text{est}} = 300$

Figure 4: Turnover in percentage points.



(a) $N_{\text{est}} = 35$ (b) $N_{\text{est}} = 40$

(c) $N_{\text{est}} = 45$ (d) $N_{\text{est}} = 50$

(e) $N_{\text{est}} = 75$ (f) $N_{\text{est}} = 150$

(g) $N_{\text{est}} = 225$ (h) $N_{\text{est}} = 300$

Figure 5: Trading costs in basis points for each trading period. Borrowing rate is taken to be 7% APR and transaction cost rate is taken to be 0.5% APR. The y-axes are log-scaled.

of $\bar{\omega}_{ii}$ when $p_n > n$ can be obtained by using the inverse of the sample conditional variance of each variable. In practice, however, once can simply use the diagonal estimates given by CONCORD, and there is no need for recourse to external estimates. Note also that CONCORD estimates themselves always exist, regardless of the sample size, and with certainty will lead to estimates, even when $p_n > n$. This property follows directly from the convergence of the CONCORD algorithm.

O Joint convexity of the SYMLASSO in the Ω parameterization

We will show that the SYMLASSO objective function in (4) is jointly convex if we reparameterize in terms of Ω (see also Lee and Hastie (2014)). However, the SYMLASSO objective function is not in general strictly convex if $n < p$, and hence the convergence of the coordinatewise descent algorithm is not guaranteed. It follows from the proof of Lemma 5 that the SYMLASSO objective function (in terms of Ω) is given by

$$\begin{aligned} Q_{\text{sym}}(\Omega) &= \frac{n}{2}\left[-\log|\Omega_D| + tr(S\Omega\Omega_D^{-1}\Omega)\right] + \lambda \sum_{1\le i<j\le p} |\omega_{ij}| \\ &= \frac{n}{2}\left[-\sum_{i=1}^{p} \log\omega_{ii} + \frac{1}{\omega_{ii}}\omega_{i\bullet}^T S\omega_{i\bullet}\right] + \lambda \sum_{1\le i<j\le p} |\omega_{ij}|. \end{aligned}$$

To prove the convexity of $Q_{\text{sym}}(\Omega)$, we first prove the following lemma.

Lemma 8. *Consider the function f on $\mathbb{R}_+ \times \mathbb{R}^k$ defined by $f(\mathbf{a}) = \frac{\mathbf{a}^T A\mathbf{a}}{a_1}$. If A is positive semi-definite, then f is a convex function.*

Proof It follows by straightforward manipulations that

$$f(\mathbf{a}) = A_{11}a_1 + 2\sum_{j=2}^{k+1} A_{1j}a_j + \frac{\mathbf{a}_{-1}^T A_{-1}\mathbf{a}_{-1}}{a_1}, \tag{39}$$

where $\mathbf{a}_{-1} := (a_j)_{j=2}^{k+1}$ and A_{-1} is the principle submatrix of A obtained by excluding the first row and the first column. Since the first two terms above are clearly convex functions of $\mathbf{a}$, it suffices to prove that the third term $\frac{\mathbf{a}_{-1}^T A_{-1}\mathbf{a}_{-1}}{a_1}$ is a convex function of $\mathbf{a}$. Again, by straightforward manipulations, it follows that the Hessian matrix of this term is given by

$$H = \frac{2}{a_1^3}\begin{pmatrix} \mathbf{a}_{-1}^T A_{-1}\mathbf{a}_{-1} & -(a_1A_{-1}\mathbf{a}_{-1})^T \\ -a_1A_{-1}\mathbf{a}_{-1} & a_1^2 A_{-1} \end{pmatrix}.$$

Hence, for any $\mathbf{b} \in \mathbb{R}^{k+1}$ (with $\mathbf{b}_{-1} := (b_j)_{j=2}^{k+1}$), it follows that

$$\begin{aligned} & \mathbf{b}^T H \mathbf{b} \\ = & \frac{2}{a_1^3}\left(b_1^2 \mathbf{a}_{-1}^T A_{-1} \mathbf{a}_{-1} - 2 b_1 a_1 \mathbf{b}_{-1}^T A_{-1} \mathbf{a}_{-1} + a_1^2 \mathbf{b}_{-1}^T A_{-1} \mathbf{b}_{-1}\right). \end{aligned} \tag{40}$$

Since A_{-1} is positive semi-definite, it follows that if $\mathbf{b}_{-1}^T A_{-1} \mathbf{b}_{-1} = 0$, then $A_{-1}\mathbf{b}_{-1} = 0$. In this case

$$\mathbf{b}^T H \mathbf{b} = \frac{2}{a_1^3}\left(b_1^2 \mathbf{a}_{-1}^T A_{-1} \mathbf{a}_{-1}\right) \geq 0.$$

If $\mathbf{b}_{-1}^T A_{-1} \mathbf{b}_{-1} > 0$, then it follows by (40) that

$$\begin{aligned} & \mathbf{b}^T H \mathbf{b} \\ = & \frac{2b_1^2}{a_1^3}\left(\mathbf{a}_{-1}^T A_{-1} \mathbf{a}_{-1} - \frac{(\mathbf{b}_{-1}^T A_{-1} \mathbf{a}_{-1})^2}{\mathbf{b}_{-1}^T A_{-1} \mathbf{b}_{-1}}\right) + \frac{2}{a_1^3}\left(a_1 \sqrt{\mathbf{b}_{-1}^T A_{-1} \mathbf{b}_{-1}} - b_1 \frac{\mathbf{b}_{-1}^T A_{-1} \mathbf{a}_{-1}}{\sqrt{\mathbf{b}_{-1}^T A_{-1} \mathbf{b}_{-1}}}\right)^2 \\ \geq & \ 0. \end{aligned}$$

The last statement follows by noting that $\left(\mathbf{a}_{-1}^T A_{-1} \mathbf{a}_{-1}\right)\left(\mathbf{b}_{-1}^T A_{-1} \mathbf{b}_{-1}\right) \geq (\mathbf{b}_{-1}^T A_{-1} \mathbf{a}_{-1})^2$ (using the positive semi-definiteness of A_{-1} and the Cauchy-Schwarz inequality). Hence H is a positive semi-definite matrix, which combined with (39) implies that f is a convex function. □

It follows by the above lemma that $\frac{1}{\omega_{ii}} \omega_{i\bullet}^T S \omega_{i\bullet}$ is a convex function in $\omega_{i\bullet}$ (and hence Ω) for every $1 \leq i \leq p$. Since $-\log x$ and $|x|$ are convex functions, it follows that $Q_{\text{sym}}(\Omega)$ is a convex function.

P Examples where the Incoherence condition (A3) is satisfied

We now present two lemmas which outline settings where the Incoherence condition (A3) is satisfied. The first lemma shows that (A3) is satisfied if the true correlations are sufficiently small. This lemma can be regarded as a parallel result to (Zhao and Yu, 2006, Corollary 2), which shows that the irrepresentable condition for lasso regression is satisfied if the entries of $\frac{1}{n} X_n^T X_n$ (X_n being the regression design matrix) are bounded by $\frac{c}{2q_n - 1}$ for some $0 \leq c < 1$.

Lemma 9. *Let*

$$d_n := \max_{1 \leq i \leq p_n} |\{j : \bar{\omega}_{n,ij} \neq 0\}|.$$

The incoherence condition (A3) is satisfied if

$$\frac{|\bar{\Sigma}_{n,ij}|}{\sqrt{\bar{\Sigma}_{n,ii}\bar{\Sigma}_{n,jj}}} \leq \frac{\sqrt{2}\delta\lambda_{min}}{\sqrt{q_n d_n}\lambda_{max}},$$

for every $n \geq 1$ and $1 \leq i \neq j \leq p_n$.

Proof: It can be shown by straightforward algebraic manipulations that

$$\bar{\mathcal{L}}''_{\mathcal{A}_n,\mathcal{A}_n}(\bar{\Omega}_n) = U_n^T V_n U_n,$$

where V_n is a p_n-block diagonal matrix with the i^{th} diagonal block given by $\bar{\Sigma}_n$ without the i^{th} row and column, and U_n is an appropriate $p_n(p_n-1) \times q_n$ orthogonal matrix with 0 and 1 elements. Each column of U_n has exactly two 1's. Hence for any $\mathbf{x} \in \mathbb{R}^{q_n}$, it follows that $\mathbf{x}^T U_n^T U_n \mathbf{x} = 2\mathbf{x}^T\mathbf{x}$. It follows that the smallest eigenvalue of $U_n^T V_n U_n$ is bounded below by $\frac{2}{\lambda_{max}}$. Consequently, the largest eigenvalue of $(U_n^T V_n U_n)^{-1}$ is bounded above by $\frac{\lambda_{max}}{2}$.

Since the diagonal entries of $\bar{\Sigma}_n$ are uniformly bounded above by $\frac{1}{\lambda_{min}}$, it follows that

$$|\bar{\Sigma}_{n,kl}| \leq \frac{\sqrt{2}\delta}{\sqrt{q_n d_n}\lambda_{max}},$$

for every $n \geq 1$ and $1 \leq k \neq l \leq p_n$. Note that for every $(i,j) \notin \mathcal{A}_n$, $\bar{\mathcal{L}}''_{ij,\mathcal{A}_n}(\bar{\Omega}_n)$ has at most $2d_n$ non-zero entries. Hence, we get that

$$\left\|\bar{\mathcal{L}}''_{ij,\mathcal{A}_n}(\bar{\Omega}_n)\right\| \leq \sqrt{2d_n} \times \frac{\sqrt{2}\delta}{\sqrt{q_n d_n}\lambda_{max}} = \frac{2\delta}{\sqrt{q_n}\lambda_{max}}.$$

Finally, we note from the discussion above that

$$\begin{aligned}
& \left|\bar{\mathcal{L}}''_{ij,\mathcal{A}_n}(\bar{\Omega}_n)\left[\bar{\mathcal{L}}''_{\mathcal{A}_n,\mathcal{A}_n}(\bar{\Omega}_n)\right]^{-1} \operatorname{sign}(\bar{\omega}^o_{\mathcal{A}_n})\right| \\
\leq & \left\|\bar{\mathcal{L}}''_{ij,\mathcal{A}_n}(\bar{\Omega}_n)\right\| \left\|\left[\bar{\mathcal{L}}''_{\mathcal{A}_n,\mathcal{A}_n}(\bar{\Omega}_n)\right]^{-1}\right\| \left\|\operatorname{sign}(\bar{\omega}^o_{\mathcal{A}_n})\right\| \\
\leq & \frac{2\delta}{\sqrt{q_n}\lambda_{max}} \times \frac{\lambda_{max}}{2} \times \sqrt{q_n} \\
= & \delta.
\end{aligned}$$

Hence (A3) is satisfied. □

The next lemma shows that the Incoherence condition (A3) holds if the true $\bar{\Omega}_n$'s are tridiagonal matrices satisfying some mild conditions. This lemma can be regarded as a parallel result to (Zhao and Yu, 2006, Corollary 3).

Lemma 10. *Suppose that $\bar{\Omega}_n$ is a tridiagonal matrix with all diagonal entries equal to* 1 *and the non-zero off-diagonal entries equal to ρ_n, for every $n \geq 1$. If $\rho := \sup_n |\rho_n|$ satisfies*

$$\frac{8\rho}{(1-\rho^2)(2-\rho^4/2)} \leq \delta,$$

then (A3) is satisfied.

Proof: Using standard results for inverse of tridiagonal matrices, it follows that

$$\bar{\Sigma}_{n,ij} = \frac{\rho_n^{|i-j|}}{1-\rho_n^2},$$

for every $1 \leq i, j \leq p_n$. Note that $\mathcal{A}_n = \{(i-1, i) : \ 2 \leq i \leq p_n\}$, and $|\mathcal{A}_n| = p_n - 1$. Hence, $\bar{\mathcal{L}}''_{\mathcal{A}_n,\mathcal{A}_n}(\bar{\Omega}_n)$ is a tridiagonal matrix (with the i^{th} row corresponding to the edge $(i, i+1)$), with

$$\bar{\mathcal{L}}''_{i(i+1),i(i+1)}(\bar{\Omega}_n) = \bar{\Sigma}_{n,ii} + \bar{\Sigma}_{n,(i+1)(i+1)} = \frac{2}{1-\rho_n^2},$$

for every $1 \leq i \leq p_n - 1$, and

$$\bar{\mathcal{L}}''_{i(i+1),(i+1)(i+2)}(\bar{\Omega}_n) = \bar{\Sigma}_{n,i(i+2)} = \frac{\rho_n^2}{1-\rho_n^2},$$

for every $1 \leq i \leq p_n - 2$. Again, using standard results for inverse of tridiagonal matrices, it follows that

$$\left(\bar{\mathcal{L}}''_{\mathcal{A}_n,\mathcal{A}_n}(\bar{\Omega}_n)\right)^{-1}_{i(i+1),j(j+1)} = \frac{(1-\rho_n^2)(\rho_n^2/2)^{|i-j|}}{2-\rho_n^4/2},$$

for every $1 \leq i, j \leq p_n - 1$. Using the fact that $\sum_{i=0}^{\infty} a^i = \frac{1}{1-a}$ for $|a| < 1$, we conclude that each entry in $\left(\bar{\mathcal{L}}''_{\mathcal{A}_n,\mathcal{A}_n}\right)^{-1}(\bar{\Omega}_n)\,\mathrm{sign}(\bar{\omega}^o_{\mathcal{A}_n})$ is bounded above in absolute value by $\frac{2}{2-\rho_n^4/2}$. Moreover, if $i < j$ and $(i,j) \notin \mathcal{A}_n$, then $\bar{\mathcal{L}}''_{ij,\mathcal{A}_n}(\bar{\Omega}_n)$ has at most four non-zero entries (entries corresponding to the edges $(i-1,i), (i,i+1), (j-1,j)$ and $(j,j+1)$, if applicable). All of these non-zero entries are bounded above in absolute value by $\frac{|\rho_n|}{1-\rho_n^2}$. It follows that for every $(i,j) \notin \mathcal{A}_n$,

$$\begin{aligned}
&\left|\bar{\mathcal{L}}''_{ij,\mathcal{A}_n}(\bar{\Omega}_n)\left[\bar{\mathcal{L}}''_{\mathcal{A}_n,\mathcal{A}_n}(\bar{\Omega}_n)\right]^{-1}\mathrm{sign}(\bar{\omega}^o_{\mathcal{A}_n})\right| \\
\leq\ & \frac{4|\rho_n|}{1-\rho_n^2} \times \frac{2}{2-\rho_n^4/2} \\
=\ & \frac{8|\rho_n|}{(1-\rho_n^2)(2-\rho_n^4/2)}
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{8|\rho|}{(1-\rho^2)(2-\rho^4/2)} \\
&\leq \delta.
\end{aligned}
$$

Hence (A3) is satisfied. $\square$

Q Non-convergence of SPACE

We provide a simple example where the SPACE algorithm (with uniform weights) does not converge, and the iterates alternate between two matrices. A sample of $n = 4$ *i.i.d.* vectors was generated from the $\mathcal{N}(\mathbf{0}, \Sigma)$ distribution with Σ as in (2). The standardized data is as follows:

$$
\begin{pmatrix}
0.659253 & -0.635923 & 0.492419 \\
0.994414 & -1.015863 & 1.115863 \\
-1.150266 & 1.141668 & -1.135115 \\
-0.503401 & 0.510117 & -0.473166
\end{pmatrix}. \tag{41}
$$

The SPACE algorithm was implemented with choice of weights $w_i = 1$ and $\lambda = 0.2$. Again, after the first few iterations, it turns out that successive SPACE iterates alternate between

$$
\begin{pmatrix}
1.432570 & 1.416740 & -2.132500 \\
1.416740 & 3552.598070 & 0.000000 \\
-2.132500 & 0.000000 & 89.163310
\end{pmatrix}
\text{ and }
\begin{pmatrix}
3552.565950 & 1.416720 & 0.000000 \\
1.416720 & 1.404240 & 2.100770 \\
0.000000 & 2.100770 & 123.137260
\end{pmatrix},
$$

thereby also establishing non-convergence of the SPACE algorithm in the case when the weights $w_i = 1$. Note that some of the elements in the two matrices above are vastly different. The sparsity pattern is also different, thereby yielding two different partial correlation graphs.

References

Albergaria, A., Paredes, J., Sousa, B., Milanezi, F., Carneiro, V., Bastos, J., Costa, S., Vieira, D., Lopes, N., Lam, E. W., Lunet, N., and Schmitt, F. (2009). Expression of FOXA1 and GATA-3 in breast cancer: the prognostic significance in hormone receptor-negative tumours. *Breast Cancer Research*, 11(3):R40.

Bibby, R. A., Tang, C., Faisal, A., Drosopoulos, K., Lubbe, S., Houlston, R., Bayliss, R., and Linardopoulos, S. (2009). A cancer-associated aurora A mutant is mislocalized and misregulated due to loss of interaction with TPX2. *The Journal of Biological Chemistry*, 284(48):33177–84.

Davidson, B., Stavnes, H. T., Holth, A., Chen, X., Yang, Y., Shih, I.-M., and Wang, T.-L. (2011). Gene expression signatures differentiate ovarian/peritoneal serous carcinoma from breast carcinoma in effusions. *Journal of Cellular and Molecular Medicine*, 15(3):535–44.

Du, J., Li, L., Ou, Z., Kong, C., Zhang, Y., Dong, Z., Zhu, S., Jiang, H., Shao, Z., Huang, B., and Lu, J. (2012). FOXC1, a target of polycomb, inhibits metastasis of breast cancer cells. *Breast Cancer Research and Treatment*, 131(1):65–73.

Eeckhoute, J., Keeton, E. K., Lupien, M., Krum, S. A., Carroll, J. S., and Brown, M. (2007). Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer. *Cancer Research*, 67(13):6477–83.

Eschenbrenner, J., Winsel, S., Hammer, S., Sommer, A., Mittelstaedt, K., Drosch, M., Klar, U., Sachse, C., Hannus, M., Seidel, M., Weiss, B., Merz, C., Siemeister, G., and Hoffmann, J. (2011). Evaluation of activity and combination strategies with the microtubule-targeting drug sagopilone in breast cancer cell lines. *Frontiers in Oncology*, 1:44.

Glinsky, G. V., Berezovska, O., and Glinskii, A. B. (2005). Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. *The Journal of clinical investigation*, 115(6):1503–21.

Jiang, S., Katayama, H., Wang, J., Li, S. A., Hong, Y., Radvanyi, L., Li, J. J., and Sen, S. (2010). Estrogen-induced aurora kinase-A (AURKA) gene expression is activated by GATA-3 in estrogen receptor-positive breast cancer cells. *Hormones & Cancer*, 1(1):11–20.

Joosse, S. A., Hannemann, J., Spötter, J., Bauche, A., Andreas, A., Müller, V., and Pantel, K. (2012). Changes in Keratin Expression during Metastatic Progression of Breast Cancer: Impact on the Detection of Circulating Tumor Cells. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 18(4):993–1003.

Katoh, M. (2008). WNT signaling in stem cell biology and regenerative medicine. *Current Drug Targets*, 9(7):565–70.

Khare, K. and Rajaratnam, B. (2014). Convergence of cyclic coordinate l1 minimization. *Preprint, Department of Statistics, Stanford University (soon to be available on arxiv)*.

Koboldt, D. C. and Others (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418):61–70.

Kraus, T. S., Cohen, C., and Siddiqui, M. T. (2010). Prostate-specific antigen and hormone receptor expression in male and female breast carcinoma. *Diagnostic Pathology*, 5:63.

Lacroix, M. and Leclercq, G. (2004). About GATA3, HNF3A, and XBP1, three genes co-expressed with the oestrogen receptor-alpha gene (ESR1) in breast cancer. *Molecular and Cellular Endocrinology*, 219(1-2):1–7.

Lee, J. and Hastie, T. (2014). Learning the structure of mixed graphical models. *to appear in Journal of Computational and Graphical Statistics*.

Licata, L. A., Hostetter, C. L., Crismale, J., Sheth, A., and Keen, J. C. (2010). The RNA-binding protein HuR regulates GATA3 mRNA stability in human breast cancer cell lines. *Breast Cancer Research and Treatment*, 122(1):55–63.

Maxwell, C. A. and Others (2011). Interplay between BRCA1 and RHAMM regulates epithelial apicobasal polarization and may influence risk of breast cancer. *PLoS Biology*, 9(11):e1001199.

Mohajeri, A., Zarghami, N., Pourhasan Moghadam, M., Alani, B., Montazeri, V., Baiat, A., and Fekhrjou, A. (2011). Prostate-specific antigen gene expression and telomerase activity in breast cancer patients: possible relationship to steroid hormone receptors. *Oncology Research*, 19(8-9):375–80.

Pellegrino, M. B., Asch, B. B., Connolly, J. L., and Asch, H. L. (1988). Differential expression of keratins 13 and 16 in normal epithelium, benign lesions, and ductal carcinomas of the human breast determined by the monoclonal antibody Ks8.12. *Cancer Research*, 48(20):5831–6.

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(486):735–746.

Ray, P. S., Bagaria, S. P., Wang, J., Shamonki, J. M., Ye, X., Sim, M.-S., Steen, S., Qu, Y., Cui, X., and Giuliano, A. E. (2011). Basal-like breast cancer defined by FOXC1 expression offers superior prognostic value: a retrospective immunohistochemical study. *Annals of Surgical Oncology*, 18(13):3839–47.

Robinson, J. L. L., Macarthur, S., Ross-Innes, C. S., Tilley, W. D., Neal, D. E., Mills, I. G., and Carroll, J. S. (2011). Androgen receptor driven transcription in molecular apocrine breast cancer is mediated by FoxA1. *The EMBO Journal*, 30(15):3019–27.

Rocha, G., Zhao, P., and Yu, B. (2008). A path following algorithm for Sparse Pseudo-Likelihood Inverse Covariance Estimation (SPLICE). Technical report, Statistics Department, UC Berkeley, Berkeley, CA.

Rønneberg, J. A., Fleischer, T., Solvang, H. K., Nordgard, S. H., Edvardsen, H., Potapenko, I., Nebdal, D., Daviaud, C., Gut, I., Bukholm, I., Naume, B. r., Bø rresen Dale, A.-L., Tost, J., and Kristensen, V. (2011). Methylation profiling with a panel of cancer related genes: association with estrogen receptor, TP53 mutation status and expression subtypes in sporadic breast cancer. *Molecular Oncology*, 5(1):61–76.

Sauter, E. R., Lininger, J., Magklara, A., Hewett, J. E., and Diamandis, E. P. (2004). Association of kallikrein expression in nipple aspirate fluid with breast cancer risk. *International Journal of Cancer*, 108(4):588–91.

Shimo, A., Nishidate, T., Ohta, T., Fukuda, M., Nakamura, Y., and Katagiri, T. (2007). Elevated expression of protein regulator of cytokinesis 1, involved in the growth of breast cancer cells. *Cancer Science*, 98(2):174–81.

Shimo, A., Tanikawa, C., Nishidate, T., Lin, M.-L., Matsuda, K., Park, J.-H., Ueki, T., Ohta, T., Hirata, K., Fukuda, M., Nakamura, Y., and Katagiri, T. (2008). Involvement of kinesin family member 2C/mitotic centromere-associated kinesin overexpression in mammary carcinogenesis. *Cancer Science*, 99(1):62–70.

Sizemore, S. T. and Keri, R. A. (2012). The Forkhead Box Transcription Factor FOXC1 Promotes Breast Cancer Invasion by Inducing Matrix Metalloprotease 7 (MMP7) Expression. *The Journal of Biological Chemistry*, 287(29):24631–40.

Tkocz, D., Crawford, N. T., Buckley, N. E., Berry, F. B., Kennedy, R. D., Gorski, J. J., Harkin, D. P., and Mullan, P. B. (2012). BRCA1 and GATA3 corepress FOXC1 to inhibit the pathogenesis of basal-like breast cancers. *Oncogene*, 31(32):3667–3678.

Wang, J., Ray, P. S., Sim, M.-S., Zhou, X. Z., Lu, K. P., Lee, A. V., Lin, X., Bagaria, S. P., Giuliano, A. E., and Cui, X. (2012). FOXC1 regulates the functions of human basal-like breast cancer cells by activating NF-κB signaling. *Oncogene*.

Won, J.-H., Lim, J., Kim, S.-J., and Rajaratnam, B. (2012). Condition Number Regularized Covariance Estimation. *Journal of the Royal Statistical Society: Series B*.

Yan, W., Cao, Q. J., Arenas, R. B., Bentley, B., and Shao, R. (2010). GATA3 inhibits breast cancer metastasis through the reversal of epithelial-mesenchymal transition. *The Journal of Biological Chemistry*, 285(18):14042–14051.

Yang, Q., Nakamura, M., Nakamura, Y., Yoshimura, G., Suzuma, T., Umemura, T., Tamaki, T., Mori, I., Sakurai, T., and Kakudo, K. (2002). Correlation of prostate-specific antigen promoter polymorphisms with clinicopathological characteristics in breast cancer. *Anticancer Research*, 22(3):1825–8.

Zhao, P. and Yu, B. (2006). On Model Selection Consistency of Lasso. *Journal of Machine Learning Research*, 7:2541–2563.

Zheng, Y., Huo, D., Zhang, J., Yoshimatsu, T. F., Niu, Q., and Olopade, O. I. (2012). Microsatellites in the Estrogen Receptor (ESR1, ESR2) and Androgen Receptor (AR) Genes and Breast Cancer Risk in African American and Nigerian Women. *PLoS ONE*, 7(7):e40494.