Sparse Portfolios

Ekaterina Seregina

Department of Economics, University of California Riverside

UCR Econometrics Seminar
October 9, 2020

Markowitz Mean-Variance Optimization

- Suppose we observe $i = 1, \ldots, p$ excess returns over $t = 1, \ldots, T$ period of time: $\mathbf{r}_t = (r_{1t}, \ldots, r_{pt})' \sim \mathcal{D}(\mathbf{m}, \boldsymbol{\Sigma})$.
- Want: Find the *optimal* weights of portfolio, $\mathbf{w}$.
- The optimal portfolio may be subject to the constraints on
  - (i) a desired expected return, μ;
  - (ii) maximum risk, σ;
  - (iii) weights sum up to one.
- Need:
  - (i) expected returns, $\mathbf{m}$;
  - (ii) inverse covariance (*precision*) matrix, $\boldsymbol{\Theta} := \boldsymbol{\Sigma}^{-1}$.

Large Portfolio Optimization: Objective

- Many stocks available for investing:
  S&P500; NASDAQ ($> 3,000$); Russell ($> 1,000$)
- Break down the search for the optimal weights:
  - (i) which stocks to buy?
    - ➔ Buy all stocks – non-sparse portfolio;
    - ➔ Buy a subset of all stocks – sparse portfolio
  - (ii) how much to invest in these stocks?
- Challenges of optimizing over a large number of assets:
  - (iii) Consistent weight estimation when $p = p_T \to \infty$ as $T \to \infty$ and/or $p > T$;
  - (iv) Factor structure of returns: consistent estimation of factors and factor loadings;
  - (v) Easy to monitor, low rebalancing costs, and robust performance during recessions.

Large Portfolio Optimization: Solution # 1

Buy a subset of all available stocks: pick a few top-performing stocks:

Lyle & Yohn, 2020: *when do you stop* and *what is the guarantee that the best performers of the last month or year will still show superior performance today?*

✗(i) which stocks to buy?

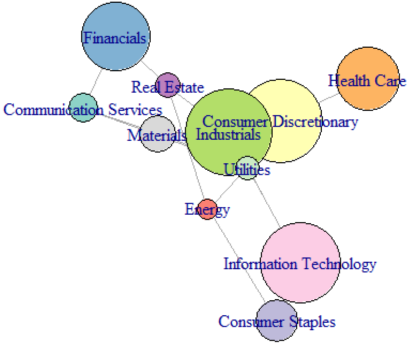
✓(ii) how much to invest in these stocks?

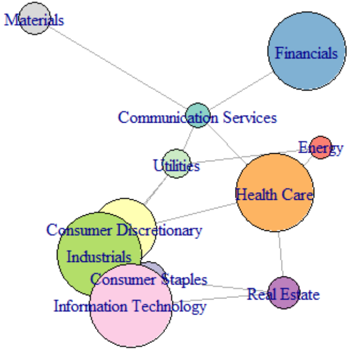
✗(iii) Consistent weight estimation when $p = p_T \to \infty$ as $T \to \infty$ and/or $p > T$;

✗(iv) Factor structure of returns: consistent estimation of factors and factor loadings;

✗(v) Easy to monitor, low rebalancing costs, and robust performance during recessions.

Partial correlation networks of S&P500 sectors in 2019(left) & 2020(right).





COVID-19 outbreak

Daily returns of 495 components of the S&P500 from May 25, 2018 – September 24, 2020 (588 obs.): training period is May 25, 2018 – October 23, 2018 (105 obs.), OOS period is October 24, 2018 – September 24, 2020 (483 obs.). Rolling window w/ monthly rebalancing.

	Total OOS Performance 10/24/19–09/24/20			Before the Pandemic 01/02/19–12/31/19		During the Pandemic 01/02/20–06/30/20	
	Return ($\times$100)	**Risk** ($\times$100)	**Sharpe Ratio**	**CER** ($\times$100)	**Risk** ($\times$100)	**CER** ($\times$100)	**Risk** ($\times$100)
Best10	-0.0293	1.5478	-0.0189	-3.3748	1.0448	-5.9645	2.1522
Best30	-0.0132	1.4744	-0.0090	19.9202	0.8204	-11.6194	2.2708
Best50	0.0181	1.6157	0.0111	28.8664	0.9305	-7.4551	2.5603
Best100	-0.0861	3.8126	-0.0226	-2.0883	3.0643	-43.0285	4.9452
Best200	0.0063	1.5541	0.0041	7.7683	0.7941	-6.8099	2.6036

Table 1: Performance of portfolios that use the stocks that exhibited the best performance in terms of the average return over the last 5 months.

COVID-19 outbreak

Daily returns of 495 components of the S&P500 from May 25, 2018 – September 24, 2020 (588 obs.): training period is May 25, 2018 – October 23, 2018 (105 obs.), OOS period is October 24, 2018 – September 24, 2020 (483 obs.). Rolling window w/ monthly rebalancing.

	Total OOS Performance 10/24/19–09/24/20			Before the Pandemic 01/02/19–12/31/19		During the Pandemic 01/02/20–06/30/20	
	Return ($\times$100)	**Risk** ($\times$100)	**Sharpe Ratio**	**CER** ($\times$100)	**Risk** ($\times$100)	**CER** ($\times$100)	**Risk** ($\times$100)
Best10	-0.0293	1.5478	-0.0189	-3.3748	1.0448	-5.9645	2.1522
Best30	-0.0132	1.4744	-0.0090	19.9202	0.8204	-11.6194	2.2708
Best50	0.0181	1.6157	0.0111	28.8664	0.9305	-7.4551	2.5603
Best100	-0.0861	3.8126	-0.0226	-2.0883	3.0643	-43.0285	4.9452
Best200	0.0063	1.5541	0.0041	7.7683	0.7941	-6.8099	2.6036

Table 1: Performance of portfolios that use the stocks that exhibited the best performance in terms of the average return over the last 5 months.

Large Portfolio Optimization: Solution # 2

Buy a subset of all available stocks: imposing constraints $\rightarrow$ set some weights to zero:

- gross-exposure/ℓ_1-constraint (Brodie et al., 2009; Li, 2015; Ao et al., 2019)
- Downward-biased weight estimates $\rightarrow$ very low or negative portfolio return $\rightarrow$ eliminate all benefits from stock selection!

✓(i) which stocks to buy?

✓(ii) how much to invest in these stocks?

✗(iii) Consistent weight estimation when $p = p_T \rightarrow \infty$ as $T \rightarrow \infty$ and/or $p > T$;

✗(iv) Factor structure of returns: consistent estimation of factors and factor loadings;

✗(v) Easy to monitor, low rebalancing costs, and robust performance during recessions.

Large Portfolio Optimization: Solution # 3

Use a high-dimensional precision matrix estimator

- Nodewise regression to handle high dimensions (Meinshausen & Bühlmann, 2006, Caner et al., 2019 & 2020)

✓(i) which stocks to buy?

✓(ii) how much to invest in these stocks?

✓(iii) Consistent weight estimation when $p = p_T \to \infty$ as $T \to \infty$ and/or $p > T$;

✗(iv) Factor structure of returns: consistent estimation of factors and factor loadings;

✗(v) Easy to monitor, low rebalancing costs, and robust performance during recessions.

COVID-19 outbreak

Daily returns of 495 components of the S&P500 from May 25, 2018 – September 24, 2020 (588 obs.): training period is May 25, 2018 – October 23, 2018 (105 obs.), OOS period is October 24, 2018 – September 24, 2020 (483 obs.). Rolling window w/ monthly rebalancing.

	Total OOS Performance 10/24/19–09/24/20			Before the Pandemic 01/02/19–12/31/19		During the Pandemic 01/02/20–06/30/20	
	Return ($\times$100)	**Risk** ($\times$100)	**Sharpe Ratio**	**CER** ($\times$100)	**Risk** ($\times$100)	**CER** ($\times$100)	**Risk** ($\times$100)
EW	0.0108	1.8781	0.0058	28.5420	0.8010	-19.7207	3.3169
Index	0.0351	1.7064	0.0206	27.8629	0.7868	-9.0802	2.9272
Nodewise Regr'n	0.0322	1.6384	0.0196	29.6292	0.6856	-11.7431	2.8939
Standard Lasso	-0.0003	0.0107	-0.0250	-0.1274	0.0148	-0.0021	0.0002
Our Post-Lasso-based	0.1247	1.7254	0.0723	45.2686	1.0386	12.4196	2.8554

Table 2: Performance of non-sparse and sparse portfolios.

Selected Stocks

COVID-19 outbreak

Daily returns of 495 components of the S&P500 from May 25, 2018 – September 24, 2020 (588 obs.): training period is May 25, 2018 – October 23, 2018 (105 obs.), OOS period is October 24, 2018 – September 24, 2020 (483 obs.). Rolling window w/ monthly rebalancing.

	Total OOS Performance 10/24/19–09/24/20			Before the Pandemic 01/02/19–12/31/19		During the Pandemic 01/02/20–06/30/20	
	Return (×100)	**Risk** (×100)	**Sharpe Ratio**	**CER** (×100)	**Risk** (×100)	**CER** (×100)	**Risk** (×100)
EW	0.0108	1.8781	0.0058	28.5420	0.8010	-19.7207	3.3169
Index	0.0351	1.7064	0.0206	27.8629	0.7868	-9.0802	2.9272
Nodewise Regr'n	0.0322	1.6384	0.0196	29.6292	0.6856	-11.7431	2.8939
Standard Lasso	-0.0003	0.0107	-0.0250	-0.1274	0.0148	-0.0021	0.0002
Our Post-Lasso-based	0.1247	1.7254	0.0723	45.2686	1.0386	12.4196	2.8554

Table 2: Performance of non-sparse and sparse portfolios.

Selected Stocks

Large Portfolio Optimization: This Paper

✓ (i) *Use ℓ_1-regularized regression representation to get portfolio weights* $\Rightarrow$ which stocks to buy?

✓ (ii) *Use de-biasing and post-Lasso to get unbiased consistent weight estimates* $\Rightarrow$ how much to invest in these stocks?

Develop a novel HD precision matrix estimator which combines the benefits of graphical models & factor models:

✓ (iii) $\Rightarrow$ Consistent weight estimation when $p = p_T \to \infty$ as $T \to \infty$ and/or $p > T$;

✓ (iv) $\Rightarrow$ Factor structure of returns: consistent estimation of factors and factor loadings;

✓ (v) *Empirical application to the components of S&P500* $\Rightarrow$ easy to monitor, low rebalancing costs, and robust performance during recessions.

Markowitz Risk-Constrained Problem (MRC)

Optimal portfolio achieves a desired expected return, μ, with minimum variance:

$$\begin{cases} \min_{\mathbf{w}} \frac{1}{2}\mathbf{w}'\mathbf{\Sigma}\mathbf{w} \\ \text{s.t. } \mathbf{m}'\mathbf{w} \geq \mu \end{cases}$$

Optimal portfolio maximizes expected return given a maximum risk-tolerance level, σ:

$$\begin{cases} \max_{\mathbf{w}} \mathbf{w}'\mathbf{m} \\ \text{s.t. } \mathbf{w}'\mathbf{\Sigma}\mathbf{w} \leq \sigma^2 \end{cases}$$

$$\max_{\mathbf{w}} \frac{\mathbf{m}'\mathbf{w}}{\sqrt{\mathbf{w}'\mathbf{\Sigma}\mathbf{w}}} \text{ s.t. } \mathbf{m}'\mathbf{w} \geq \mu \text{ or } \mathbf{w}'\mathbf{\Sigma}\mathbf{w} \leq \sigma^2 \tag{1}$$

- Let $\mu = \sigma\sqrt{\mathbf{m}'\mathbf{\Theta}\mathbf{m}}$:

$$\mathbf{w}_{\text{MRC}} = \frac{\sigma}{\sqrt{\mathbf{m}'\mathbf{\Theta}\mathbf{m}}}\mathbf{\Theta}\mathbf{m} = \frac{\sigma}{\sqrt{\theta}}\mathbf{\Theta}\mathbf{m}, \tag{2}$$

where $\theta := \mathbf{m}'\mathbf{\Theta}\mathbf{m}$ is the square of the maximum Sharpe ratio.

Sparse Portfolio

- Define

$$r_c := \frac{1+\theta}{\theta}\mu \equiv \sigma\frac{1+\theta}{\sqrt{\theta}} \tag{3}$$

- When r_c is determined by equation (3), the solution to the following unconstrained regression problem yields MRC portfolio weights:

$$\mathbf{w}_{\mathrm{MRC}} = \underset{\mathbf{w}}{\operatorname{argmin}}\, \mathbb{E}\left[r_c - \mathbf{w}'\mathbf{r}_t\right]^2 = \frac{\sigma}{\sqrt{\theta}}\boldsymbol{\Theta}\mathbf{m} \tag{4}$$

Let $\mathbf{R}$ be a $T \times p$ matrix of excess returns and $\mathbf{r}_c$ be a $T \times 1$ constant vector. Consider a high-dimensional linear model

$$\mathbf{r}_c = \mathbf{R}\mathbf{w} + \mathbf{e}, \quad \text{where} \quad \mathbf{e} \sim \mathcal{D}(\mathbf{0}, \sigma_e^2\mathbf{I}).$$

Proof

Sparse Portfolio

- Define

$$r_c := \frac{1+\theta}{\theta}\mu \equiv \sigma\frac{1+\theta}{\sqrt{\theta}} \tag{3}$$

- When r_c is determined by equation (3), the solution to the following unconstrained regression problem yields MRC portfolio weights:

$$\mathbf{w}_{\text{MRC}} = \underset{\mathbf{w}}{\operatorname{argmin}}\, \mathbb{E}\left[r_c - \mathbf{w}'\mathbf{r}_t\right]^2 = \frac{\sigma}{\sqrt{\theta}}\boldsymbol{\Theta}\mathbf{m} \tag{4}$$

Let $\mathbf{R}$ be a $T \times p$ matrix of excess returns and $\mathbf{r}_c$ be a $T \times 1$ constant vector. Consider a high-dimensional linear model

$$\mathbf{r}_c = \mathbf{R}\mathbf{w} + \mathbf{e}, \quad \text{where} \quad \mathbf{e} \sim \mathcal{D}(\mathbf{0}, \sigma_e^2\mathbf{I}).$$

▸ Proof

Sparse Portfolio

- To get sparse weights we impose standard LASSO (ℓ_1) penalty which yields the following constrained optimization problem:

$$\widehat{\mathbf{w}}_{\text{MRC,SPARSE}} := \widehat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1}{T} \sum_{t=1}^{T} (r_c - \mathbf{w}'\mathbf{r}_t)^2 + 2\lambda \|\mathbf{w}\|_1 \quad (5)$$

- Problems: (a) the estimator in (5) is biased and (b) r_c needs to be estimated.
- Solutions: (a) use de-biasing (van de Geer et al., 2014, Belloni et al., 2015, Javanmard et al., 2018).
- For now, suppose we have consistent estimators of $\mathbf{r}_c$ and $\mathbf{\Theta}$.

Sparse De-Biased Portfolio

$$\text{(KKT)}: \ -\mathbf{R}'(\widehat{\mathbf{r}}_c - \mathbf{R}\widehat{\mathbf{w}})/T + \lambda\widehat{\mathbf{g}} = 0,$$

Let $\widehat{\boldsymbol{\Sigma}} = \mathbf{R}'\mathbf{R}/T$, then we can rewrite the KKT conditions:

$$\widehat{\boldsymbol{\Sigma}}(\widehat{\mathbf{w}} - \mathbf{w}) + \lambda\widehat{\mathbf{g}} = \mathbf{R}'\mathbf{e}/T. \tag{6}$$

Multiply both sides of (6) by $\widehat{\boldsymbol{\Theta}}$, add and subtract $(\widehat{\mathbf{w}} - \mathbf{w})$:

$$\widehat{\mathbf{w}} - \mathbf{w} + \widehat{\boldsymbol{\Theta}}\lambda\widehat{\mathbf{g}} = \widehat{\boldsymbol{\Theta}}\mathbf{R}'\mathbf{e}/T - \underbrace{\sqrt{T}(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}} - \mathbf{I}_p)(\widehat{\mathbf{w}} - \mathbf{w})}_{\Delta}/\sqrt{T}$$

$$\widehat{\mathbf{w}}_{\text{MRC,DEBIASED}} = \widehat{\mathbf{w}} + \widehat{\boldsymbol{\Theta}}\lambda\widehat{\mathbf{g}} = \widehat{\mathbf{w}} + \widehat{\boldsymbol{\Theta}}\mathbf{R}'(\widehat{\mathbf{r}}_c - \mathbf{R}\widehat{\mathbf{w}})/T, \tag{7}$$

$\widehat{\mathbf{g}}$ is a $p \times 1$ vector arising from the subgradient of $\|\mathbf{w}\|_1$.

Sparse De-Biased Portfolio

$$(\text{KKT}): \ -\mathbf{R}'(\widehat{\mathbf{r}}_c - \mathbf{R}\widehat{\mathbf{w}})/T + \lambda\widehat{\mathbf{g}} = 0,$$

Let $\widehat{\mathbf{\Sigma}} = \mathbf{R}'\mathbf{R}/T$, then we can rewrite the KKT conditions:

$$\widehat{\mathbf{\Sigma}}(\widehat{\mathbf{w}} - \mathbf{w}) + \lambda\widehat{\mathbf{g}} = \mathbf{R}'\mathbf{e}/T. \tag{6}$$

Multiply both sides of (6) by $\widehat{\mathbf{\Theta}}$, add and subtract $(\widehat{\mathbf{w}} - \mathbf{w})$:

$$\widehat{\mathbf{w}} - \mathbf{w} + \widehat{\mathbf{\Theta}}\lambda\widehat{\mathbf{g}} = \widehat{\mathbf{\Theta}}\mathbf{R}'\mathbf{e}/T - \underbrace{\sqrt{T}(\widehat{\mathbf{\Theta}}\widehat{\mathbf{\Sigma}} - \mathbf{I}_p)(\widehat{\mathbf{w}} - \mathbf{w})}_{\Delta}/\sqrt{T}$$

$$\widehat{\mathbf{w}}_{\text{MRC,DEBIASED}} = \widehat{\mathbf{w}} + \widehat{\mathbf{\Theta}}\lambda\widehat{\mathbf{g}} = \widehat{\mathbf{w}} + \widehat{\mathbf{\Theta}}\mathbf{R}'(\widehat{\mathbf{r}}_c - \mathbf{R}\widehat{\mathbf{w}})/T, \tag{7}$$

$\widehat{\mathbf{g}}$ is a $p \times 1$ vector arising from the subgradient of $\|\mathbf{w}\|_1$.

Sparse De-Biased Portfolio

$$\text{(KKT)}: \; -\mathbf{R}'(\widehat{\mathbf{r}}_c - \mathbf{R}\widehat{\mathbf{w}})/T + \lambda\widehat{\mathbf{g}} = 0,$$

Let $\widehat{\boldsymbol{\Sigma}} = \mathbf{R}'\mathbf{R}/T$, then we can rewrite the KKT conditions:

$$\widehat{\boldsymbol{\Sigma}}(\widehat{\mathbf{w}} - \mathbf{w}) + \lambda\widehat{\mathbf{g}} = \mathbf{R}'\mathbf{e}/T. \tag{6}$$

Multiply both sides of (6) by $\widehat{\boldsymbol{\Theta}}$, add and subtract $(\widehat{\mathbf{w}} - \mathbf{w})$:

$$\widehat{\mathbf{w}} - \mathbf{w} + \widehat{\boldsymbol{\Theta}}\lambda\widehat{\mathbf{g}} = \widehat{\boldsymbol{\Theta}}\mathbf{R}'\mathbf{e}/T - \underbrace{\sqrt{T}(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}} - \mathbf{I}_p)(\widehat{\mathbf{w}} - \mathbf{w})}_{\Delta}/\sqrt{T}$$

$$\widehat{\mathbf{w}}_{\text{MRC,DEBIASED}} = \widehat{\mathbf{w}} + \widehat{\boldsymbol{\Theta}}\lambda\widehat{\mathbf{g}} = \widehat{\mathbf{w}} + \widehat{\boldsymbol{\Theta}}\mathbf{R}'(\widehat{\mathbf{r}}_c - \mathbf{R}\widehat{\mathbf{w}})/T, \tag{7}$$

$\widehat{\mathbf{g}}$ is a $p \times 1$ vector arising from the subgradient of $\|\mathbf{w}\|_1$.

Alternative Portfolio Formulations: Problem

- Problem with Debiased-Portfolio: specific portfolio choice (MRC).

$$\max_{\mathbf{w}} \frac{\mathbf{m}'\mathbf{w}}{\sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}} \text{ s.t. } \mathbf{m}'\mathbf{w} \geq \mu \text{ or } \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \leq \sigma^2, \tag{8}$$

Alternative Portfolio Formulations: Problem

- Problem with Debiased-Portfolio: specific portfolio choice (MRC).

$$\max_{\mathbf{w}} \frac{\mathbf{m}'\mathbf{w}}{\sqrt{\mathbf{w}'\mathbf{\Sigma}\mathbf{w}}} \text{ s.t. } \mathbf{m}'\mathbf{w} \geq \mu \text{ or } \mathbf{w}'\mathbf{\Sigma}\mathbf{w} \leq \sigma^2, \mathbf{w}'\boldsymbol{\iota} = 1 \qquad (8)$$

Why not add weight constraint to (8)?

Alternative Portfolio Formulations: Problem

- Problem with Debiased-Portfolio: specific portfolio choice (MRC).

$$\max_{\mathbf{w}} \frac{\mathbf{m}'\mathbf{w}}{\sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}} \text{ s.t. } \mathbf{m}'\mathbf{w} \geq \mu \text{ or } \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \leq \sigma^2, \mathbf{w}'\boldsymbol{\iota} = 1 \qquad (8)$$

  Why not add weight constraint to (8)?
- (8) has two solutions, when $\boldsymbol{\iota}'\boldsymbol{\Theta}\mathbf{m} < 0$ the maximum value cannot be achieved exactly (Maller & Turkington, 2003).

Alternative Portfolio Formulations: Solution

$$\min_{\mathbf{w}} \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}, \text{ s.t. } \mathbf{m}'\mathbf{w} \geq \mu, \ \mathbf{w}'\boldsymbol{\iota} = 1 \tag{9}$$

Global Minimum-Variance (GMVP): if $\mathbf{m}'\mathbf{w} > \mu$

$$\mathbf{w}_{GMV} = (\boldsymbol{\iota}'\boldsymbol{\Theta}\boldsymbol{\iota})^{-1}\boldsymbol{\Theta}\boldsymbol{\iota}$$

Markowitz Weight-Constrained (MWC): if $\mathbf{m}'\mathbf{w} = \mu$

$$\mathbf{w}_{MWC} = (1 - a_1)\mathbf{w}_{GMV} + a_1\mathbf{w}_M,$$

$$\mathbf{w}_M = (\boldsymbol{\iota}'\boldsymbol{\Theta}\mathbf{m})^{-1}\boldsymbol{\Theta}\mathbf{m},$$

$$a_1 = \frac{\mu(\mathbf{m}'\boldsymbol{\Theta}\boldsymbol{\iota})(\boldsymbol{\iota}'\boldsymbol{\Theta}\boldsymbol{\iota}) - (\mathbf{m}'\boldsymbol{\Theta}\boldsymbol{\iota})^2}{(\mathbf{m}'\boldsymbol{\Theta}\mathbf{m})(\boldsymbol{\iota}'\boldsymbol{\Theta}\boldsymbol{\iota}) - (\mathbf{m}'\boldsymbol{\Theta}\boldsymbol{\iota})^2},$$

Algorithm 1 Sparse Portfolio Using Post-Lasso

1: Use Lasso regression in (5):

$$\widehat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1}{T} \sum_{t=1}^{T} (\hat{r}_c - \mathbf{w}'\mathbf{r}_t)^2 + 2\lambda \|\mathbf{w}\|_1$$

to select the model $\widehat{\Xi} := \text{support}(\widehat{\mathbf{w}})$.

- Apply additional thresholding to remove stocks with small estimated weights:

$$\widehat{\mathbf{w}}(t) = (\widehat{w}_j \mathbb{1}\left|\widehat{w}_j\right| > t,\ j = 1, \ldots, p),$$

- The corresponding selected model is denoted as $\widehat{\Xi}(t) := \text{support}(\widehat{\mathbf{w}}(t))$.

2: Choose a desired portfolio formulation (MRC, MWC, GMV) and apply it to the selected subset of stocks $\widehat{\Xi}(t)$.

How to get consistent estimators of $\mathbf{r}_c$ and $\mathbf{\Theta}$?

- Recall, $\theta = \mathbf{m}'\mathbf{\Theta}\mathbf{m}$ and

$$r_c := \frac{1+\theta}{\theta}\mu \equiv \sigma\frac{1+\theta}{\sqrt{\theta}}$$

- If $\widehat{\mathbf{m}}$ is the sample mean, then $\|\widehat{\mathbf{m}} - \mathbf{m}\|_{\max} = \mathcal{O}_p(\sqrt{\log(p)/T})$ (Chang et. al., 2018)
- We need a consistent estimator of HD precision matrix $\mathbf{\Theta}$.

Existing Competing Approaches to Estimate HD Precision Matrix

1. Graphical Models: estimate precision matrix directly (Nodewise-Regression by Meinshausen & Bühlmann (MB), 2006).
   - Assumption: sparse precision matrix.
2. Factor Models:

$$\underbrace{\mathbf{r}_t}_{p\times 1} = \mathbf{B}\underbrace{\mathbf{f}_t}_{K\times 1} + \varepsilon_t, \quad t = 1, \ldots, T \tag{10}$$

   - $\mathbf{f}_t = (f_{1t}, \ldots, f_{Kt})'$ are the factors,
   - $\mathbf{B}$ is a $p \times K$ matrix of factor loadings,
   - ε_t is the idiosyncratic component

   Idea: estimate covariance matrix using equation (10), invert it.

Question: how to use graphical models under the factor structure?

Factor Nodewise Regression (FMB)

$$\underbrace{\mathbf{R}}_{p\times T} = \underbrace{\mathbf{B}}_{p\times K}\mathbf{F} + \mathbf{E}. \tag{11}$$

Challenge: when factors are present, the precision matrix of returns cannot be sparse.

$$\begin{aligned}
\boldsymbol{\Sigma}_{\varepsilon} &= T^{-1}\mathbf{E}\mathbf{E}'; & \boldsymbol{\Theta}_{\varepsilon} &= \boldsymbol{\Sigma}_{\varepsilon}^{-1},\\
\boldsymbol{\Sigma}_{f} &= T^{-1}\mathbf{F}\mathbf{F}'; & \boldsymbol{\Theta}_{f} &= \boldsymbol{\Sigma}_{f}^{-1},\\
\text{cov}(\mathbf{r}_t) = \boldsymbol{\Sigma} &= \mathbf{B}\boldsymbol{\Sigma}_f\mathbf{B}' + \boldsymbol{\Sigma}_{\varepsilon}; & \boldsymbol{\Theta} &= \boldsymbol{\Sigma}^{-1}.
\end{aligned}$$

Factor Nodewise Regression (FMB)

$$\underbrace{\mathbf{R}}_{p\times T} = \underbrace{\mathbf{B}}_{p\times K}\mathbf{F} + \mathbf{E}. \tag{11}$$

Challenge: when factors are present, the precision matrix of returns cannot be sparse.

$$\widehat{\boldsymbol{\Sigma}}_{\varepsilon} = T^{-1}\widehat{\mathbf{E}}\widehat{\mathbf{E}}'; \qquad \widehat{\boldsymbol{\Theta}}_{\varepsilon} \leftarrow \text{Gr.Mdl: MB},$$

$$\widehat{\boldsymbol{\Sigma}}_{f} = T^{-1}\widehat{\mathbf{F}}\widehat{\mathbf{F}}'; \qquad \widehat{\boldsymbol{\Theta}}_{f} = \widehat{\boldsymbol{\Sigma}}_{f}^{-1},$$

Solution: use Sherman-Morrison-Woodbury (SMW) formula to estimate the precision of excess returns:

$$\text{FMB} \rightarrow \widehat{\boldsymbol{\Theta}} = \underbrace{\widehat{\boldsymbol{\Theta}}_{\varepsilon}}_{\text{MB}} - \widehat{\boldsymbol{\Theta}}_{\varepsilon}\widehat{\mathbf{B}}[\underbrace{\widehat{\boldsymbol{\Theta}}_{f}}_{\text{F.Mdl}} + \widehat{\mathbf{B}}'\widehat{\boldsymbol{\Theta}}_{\varepsilon}\widehat{\mathbf{B}}]^{-1}\underbrace{\widehat{\mathbf{B}}'}_{\text{F.Mdl}}\widehat{\boldsymbol{\Theta}}_{\varepsilon}. \tag{12}$$

Factor Nodewise Regression (FMB)

$$\underbrace{\mathbf{R}}_{p\times T} = \underbrace{\mathbf{B}}_{p\times K}\mathbf{F} + \mathbf{E}. \tag{11}$$

Challenge: when factors are present, the precision matrix of returns cannot be sparse.

$$\widehat{\mathbf{\Sigma}}_{\varepsilon} = T^{-1}\widehat{\mathbf{E}}\widehat{\mathbf{E}}'; \qquad \widehat{\mathbf{\Theta}}_{\varepsilon} \leftarrow \text{Gr.Mdl: MB},$$
$$\widehat{\mathbf{\Sigma}}_{f} = T^{-1}\widehat{\mathbf{F}}\widehat{\mathbf{F}}'; \qquad \widehat{\mathbf{\Theta}}_{f} = \widehat{\mathbf{\Sigma}}_{f}^{-1},$$

Solution: use Sherman-Morrison-Woodbury (SMW) formula to estimate the precision of excess returns:

$$\text{FMB} \rightarrow \widehat{\mathbf{\Theta}} = \underbrace{\widehat{\mathbf{\Theta}}_{\varepsilon}}_{\text{MB}} - \widehat{\mathbf{\Theta}}_{\varepsilon}\widehat{\mathbf{B}}[\underbrace{\widehat{\mathbf{\Theta}}_{f}}_{\text{F.Mdl}} + \widehat{\mathbf{B}}'\widehat{\mathbf{\Theta}}_{\varepsilon}\widehat{\mathbf{B}}]^{-1}\underbrace{\widehat{\mathbf{B}}'}_{\text{F.Mdl}}\widehat{\mathbf{\Theta}}_{\varepsilon}. \tag{12}$$

Using MB to Estimate $\boldsymbol{\Theta}_{\varepsilon}$

- Let $\widehat{\boldsymbol{\varepsilon}}_j$ be a $T \times 1$ vector of estimated residuals for the j-th regressor.
- The remaining covariates are collected in a $T \times (p-1)$ matrix $\widehat{\mathbf{E}}_{-j}$.

For each $j = 1, \ldots, p$ we run the following Lasso regressions:

$$\widehat{\boldsymbol{\gamma}}_j = \arg \min_{\boldsymbol{\gamma} \in \mathbb{R}^{p-1}} \left(\left\| \widehat{\boldsymbol{\varepsilon}}_j - \widehat{\mathbf{E}}_{-j}\boldsymbol{\gamma} \right\|_2^2 / T + 2\lambda_j \|\boldsymbol{\gamma}\|_1 \right), \tag{13}$$

where $\widehat{\boldsymbol{\gamma}}_j = \{\widehat{\gamma}_{j,k}; j = 1, \ldots, p, k \neq j\}$.

- For $j = 1, \ldots, p$, define

$$\hat{\tau}_j^2 = \left\| \widehat{\boldsymbol{\varepsilon}}_j - \widehat{\mathbf{E}}_{-j}\widehat{\boldsymbol{\gamma}}_j \right\|_2^2 / T + \lambda_j \left\| \widehat{\boldsymbol{\gamma}}_j \right\|_1 \tag{14}$$

Using MB to Estimate $\mathbf{\Theta}_{\varepsilon}$

- Define

$$\widehat{\mathbf{C}} = \begin{pmatrix} 1 & -\widehat{\gamma}_{1,2} & \cdots & -\widehat{\gamma}_{1,p} \\ -\widehat{\gamma}_{2,1} & 1 & \cdots & -\widehat{\gamma}_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ -\widehat{\gamma}_{p,1} & -\widehat{\gamma}_{p,2} & \cdots & 1 \end{pmatrix}$$

and write

$$\widehat{\mathbf{T}}^2 = \text{diag}(\hat{\tau}_1^2, \ldots, \hat{\tau}_p^2)$$

- The approximate inverse is defined as

$$\widehat{\mathbf{\Theta}} = \widehat{\mathbf{T}}^{-2}\widehat{\mathbf{C}}. \tag{15}$$

Using MB to Estimate $\boldsymbol{\Theta}_\varepsilon$

1. Matrix symmetrization procedure (Fan et al., 2018):

$$\widehat{\Theta}^s_{\varepsilon,ij} = \widehat{\Theta}_{\varepsilon,ij}\mathbb{1}\left[\left|\widehat{\Theta}_{\varepsilon,ij}\right| \leq \left|\widehat{\Theta}_{\varepsilon,ji}\right|\right] + \widehat{\Theta}_{\varepsilon,ji}\mathbb{1}\left[\left|\widehat{\Theta}_{\varepsilon,ij}\right| > \left|\widehat{\Theta}_{\varepsilon,ji}\right|\right]$$

2. Eigenvalue cleaning (Callot et al., 2017) to make $\widehat{\boldsymbol{\Theta}}^s_\varepsilon$ positive definite:
   2.1 Write the spectral decomposition $\widehat{\boldsymbol{\Theta}}^s_\varepsilon = \widehat{\mathbf{V}}'_\varepsilon \widehat{\boldsymbol{\Lambda}}_\varepsilon \widehat{\mathbf{V}}_\varepsilon$
   2.2 Let $\boldsymbol{\Lambda}_{\varepsilon,m} := \min\{\widehat{\boldsymbol{\Lambda}}_{\varepsilon,i} | \widehat{\boldsymbol{\Lambda}}_{\varepsilon,i} > 0\}$. Replace all $\widehat{\boldsymbol{\Lambda}}_{\varepsilon,i} < \boldsymbol{\Lambda}_{\varepsilon,m}$ with $\boldsymbol{\Lambda}_{\varepsilon,m}$ and define the diagonal matrix with cleaned eigenvalues as $\widetilde{\boldsymbol{\Lambda}}_\varepsilon$
   2.3 Use $\widetilde{\boldsymbol{\Theta}}_\varepsilon = \widehat{\mathbf{V}}'_\varepsilon \widetilde{\boldsymbol{\Lambda}}_\varepsilon \widehat{\mathbf{V}}_\varepsilon$ which is symmetric and positive definite

Assumptions

$$\underbrace{\mathbf{r}_t}_{p\times 1} = \mathbf{B}\underbrace{\mathbf{f}_t}_{K\times 1} + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T$$

- **(A.1)** (Spiked covariance model) As $p \to \infty$, $\Lambda_1 > \Lambda_2 + \ldots > \Lambda_K \gg \Lambda_{K+1} \geq \ldots \geq \Lambda_p \geq 0$, where $\Lambda_j = \mathcal{O}(p)$ for $j \leq K$, and $\Lambda_j = o(p)$ for $j > K$.
- **(A.2)** (Pervasive factors) There exists a p.d. $K \times K$ matrix $\breve{\mathbf{B}}$ such that $\left|\left|\left| p^{-1}\mathbf{B}'\mathbf{B} - \breve{\mathbf{B}} \right|\right|\right|_2 \to 0$ and $\lambda_{\min}(\breve{\mathbf{B}})^{-1} = \mathcal{O}(1)$ as $p \to \infty$.
- **(A.3)** (Beta mixing) Let $\mathcal{F}_{-\infty}^{t}$ and $\mathcal{F}_{t+k}^{\infty}$ denote the σ-algebras generated by $\{\boldsymbol{\varepsilon}_u : u \leq t\}$ and $\{\boldsymbol{\varepsilon}_u : u \geq t+k\}$ respectively. Then $\{\boldsymbol{\varepsilon}\}_u$ is β-mixing in the sense that $\beta_k \to 0$ as $k \to \infty$:

$$\beta_k = \sup_t \mathbb{E}\left[\sup_{B \in \mathcal{F}_{t+k}^{\infty}} \left| \Pr\left(B | \mathcal{F}_{-\infty}^{t}\right) - \Pr\left(B\right) \right| \right].$$

Assumptions

- Let $\text{cov}(\mathbf{r}_t) := \mathbf{\Sigma} = \mathbf{\Gamma}_p \mathbf{\Lambda}_p \mathbf{\Gamma}_p^{'}$
- Define $\widehat{\mathbf{\Sigma}}, \widehat{\mathbf{\Lambda}}_K, \widehat{\mathbf{\Gamma}}_K$ to be the estimators of $\mathbf{\Sigma}, \mathbf{\Lambda}_p, \mathbf{\Gamma}_p$
- Let $\widehat{\mathbf{\Lambda}}_K = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_K)$ and $\widehat{\mathbf{\Gamma}}_K = (\hat{v}_1, \dots, \hat{v}_K)$

(B.1) $\left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|_{\max} = \mathcal{O}_p(\sqrt{\log p/T})$,

(B.2) $\left\|(\widehat{\mathbf{\Lambda}}_K - \mathbf{\Lambda}_p)\mathbf{\Lambda}_p^{-1}\right\|_{\max} = \mathcal{O}_p(\sqrt{\log p/T})$,

(B.3) $\left\|\widehat{\mathbf{\Gamma}}_K - \mathbf{\Gamma}_p\right\|_{\max} = \mathcal{O}_p(\sqrt{\log p/(Tp)})$.

Assumptions

Sub-Gaussian distributions: the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}^{SG}$ with $\widehat{\boldsymbol{\Lambda}}_K^{SG}$ and $\widehat{\boldsymbol{\Gamma}}_K^{SG}$ constructed with the first K leading empirical eigenvalues and eigenvectors of $\widehat{\boldsymbol{\Sigma}}^{SG}$:

(B.1) $\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right\|_{\max} = \mathcal{O}_p(\sqrt{\log p/T})$, ✓

(B.2) $\left\|(\widehat{\boldsymbol{\Lambda}}_K - \boldsymbol{\Lambda}_p)\boldsymbol{\Lambda}_p^{-1}\right\|_{\max} = \mathcal{O}_p(\sqrt{\log p/T})$, ✓

(B.3) $\left\|\widehat{\boldsymbol{\Gamma}}_K - \boldsymbol{\Gamma}_p\right\|_{\max} = \mathcal{O}_p(\sqrt{\log p/(Tp)})$. ✓

Assumptions

Elliptical distributions:

- $\widehat{\boldsymbol{\Sigma}}^{EL1} = \widehat{\mathbf{D}}\widehat{\mathbf{R}}_1\widehat{\mathbf{D}}$, where $\widehat{\mathbf{R}}_1$ is obtained using the Kendall's tau correlation coefficients and $\widehat{\mathbf{D}}$ is a robust estimator of variances constructed using the Huber loss
- $\widehat{\boldsymbol{\Sigma}}^{EL2} = \widehat{\mathbf{D}}\widehat{\mathbf{R}}_2\widehat{\mathbf{D}}$, where $\widehat{\mathbf{R}}_2$ is obtained using the spatial Kendall's tau estimator
- $\widehat{\boldsymbol{\Lambda}}_K^{EL}$ is the matrix of the first K leading empirical eigenvalues of $\widehat{\boldsymbol{\Sigma}}^{EL1}$, and $\widehat{\boldsymbol{\Gamma}}_K^{EL}$ is the matrix of the first K leading empirical eigenvectors of $\widehat{\boldsymbol{\Sigma}}^{EL2}$

(B.1) $\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right\|_{\max} = \mathcal{O}_p(\sqrt{\log p/T})$, ✓

(B.2) $\left\|(\widehat{\boldsymbol{\Lambda}}_K - \boldsymbol{\Lambda}_p)\boldsymbol{\Lambda}_p^{-1}\right\|_{\max} = \mathcal{O}_p(\sqrt{\log p/T})$, ✓

(B.3) $\left\|\widehat{\boldsymbol{\Gamma}}_K - \boldsymbol{\Gamma}_p\right\|_{\max} = \mathcal{O}_p(\sqrt{\log p/(Tp)})$. ✓

Assumptions

(B.1) $\left\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\right\|_{\max}=\mathcal{O}_p(\sqrt{\log p/T})$,

(B.2) $\left\|(\widehat{\boldsymbol{\Lambda}}_K-\boldsymbol{\Lambda}_p)\boldsymbol{\Lambda}_p^{-1}\right\|_{\max}=\mathcal{O}_p(\sqrt{\log p/T})$,

(B.3) $\left\|\widehat{\boldsymbol{\Gamma}}_K-\boldsymbol{\Gamma}_p\right\|_{\max}=\mathcal{O}_p(\sqrt{\log p/(Tp)})$.

(C.1) $\|\boldsymbol{\Sigma}\|_{\max}=\mathcal{O}(1)$ and $\|\mathbf{B}\|_{\max}=\mathcal{O}(1)$,

Notation

$$\widehat{\mathbf{\Theta}} = \widehat{\mathbf{\Theta}}_{\varepsilon} - \widehat{\mathbf{\Theta}}_{\varepsilon}\widehat{\mathbf{B}}[\widehat{\mathbf{\Theta}}_{f} + \widehat{\mathbf{B}}'\widehat{\mathbf{\Theta}}_{\varepsilon}\widehat{\mathbf{B}}]^{-1}\widehat{\mathbf{B}}'\widehat{\mathbf{\Theta}}_{\varepsilon}$$

$$\widehat{\mathbf{r}}_{c} = \mathbf{R}\mathbf{w} + \mathbf{e} \tag{16}$$

- Denote $S_0 := \{j; \mathbf{w}_j \neq 0\}$ to be the active set of variables, where $\mathbf{w}$ is a vector of true portfolio weights in equation (16). Also, let $s_0 := |S_0|$.

$$\widehat{\gamma}_j = \arg\min_{\boldsymbol{\gamma}\in\mathbb{R}^{p-1}} \left(\left\|\widehat{\boldsymbol{\varepsilon}}_j - \widehat{\mathbf{E}}_{-j}\gamma\right\|_2^2/T + 2\lambda_j\|\gamma\|_1\right), \tag{17}$$

- Let $S_j := \{k; \gamma_{j,k} \neq 0\}$ be the active set for row $\boldsymbol{\gamma}_j$ for the nodewise regression in (17), and let $s_j := \left|S_j\right|$. Define $\bar{s} := \max_{1\leq j\leq p} s_j$.
- We use $a \lesssim_{\mathrm{P}} b$ to denote $a = \mathcal{O}_P(b)$.

Asymptotic Properties of FMB

$$\widehat{\boldsymbol{\Theta}} = \widehat{\boldsymbol{\Theta}}_{\varepsilon} - \widehat{\boldsymbol{\Theta}}_{\varepsilon}\widehat{\mathbf{B}}[\widehat{\boldsymbol{\Theta}}_{f} + \widehat{\mathbf{B}}'\widehat{\boldsymbol{\Theta}}_{\varepsilon}\widehat{\mathbf{B}}]^{-1}\widehat{\mathbf{B}}'\widehat{\boldsymbol{\Theta}}_{\varepsilon}$$

Theorem 1 (Consistency of $\widehat{\boldsymbol{\Theta}}_{\varepsilon}$)

Suppose that Assumptions (A1)-(A3), (B1)-(B3) and (C1) hold. Let $\omega_T := \sqrt{\log p/T} + 1/\sqrt{p}$*. Then* $\max_{i\leq p}(1/T)\sum_{t=1}^{T}|\hat{\varepsilon}_{it} - \varepsilon_{it}|^2 \lesssim_P \omega_T^2$ *and* $\max_{i,t}|\hat{\varepsilon}_{it} - \varepsilon_{it}| \lesssim_P \omega_T = o_p(1)$*. Under the sparsity assumption* $\bar{s}^2\omega_T = o(1)$*, with* $\lambda_j \asymp \omega_T$*, we have*

$$\max_{1\leq j\leq p}\left\|\widehat{\boldsymbol{\Theta}}_{\varepsilon,j} - \boldsymbol{\Theta}_{\varepsilon,j}\right\|_1 \lesssim_P \bar{s}\omega_T,$$

$$\max_{1\leq j\leq p}\left\|\widehat{\boldsymbol{\Theta}}_{\varepsilon,j} - \boldsymbol{\Theta}_{\varepsilon,j}\right\|_2^2 \lesssim_P \bar{s}\omega_T^2$$

Asymptotic Properties of FMB

Theorem 2 (Consistency of $\widehat{\mathbf{\Theta}}$)

Under the assumptions of Theorem 1 and, in addition, assuming $\left\|\mathbf{\Theta}_{\varepsilon,j}\right\|_2 = \mathcal{O}(1)$*, we have*

$$\max_{1\leq j\leq p}\left\|\widehat{\mathbf{\Theta}}_j - \mathbf{\Theta}_j\right\|_1 \lesssim_P \bar{s}^2\omega_T,$$

$$\max_{1\leq j\leq p}\left\|\widehat{\mathbf{\Theta}}_j - \mathbf{\Theta}_j\right\|_2^2 \lesssim_P \bar{s}\omega_T^2.$$

Lemma 1

Under the assumptions of Theorem 2, we have $|\hat{r}_c - r_c| \lesssim_P \bar{s}^2\omega_T = o_p(1)$*, where* r_c *was defined in (3).*

Asymptotic Properties of De-Biased Portfolio Weights

Theorem 3 (Consistency of $\widehat{\mathbf{w}}_{\text{MRC,DEBIASED}}$)
Under the assumptions of Theorem 2, consider the linear model $\widehat{\mathbf{r}}_c = \mathbf{R}\mathbf{w} + \mathbf{e}$ with $\mathbf{e} \sim \mathcal{D}(\mathbf{0}, \sigma_e^2 \mathbf{I})$, where $\sigma_e^2 = \mathcal{O}(1)$. Consider a suitable choice of the regularization parameters $\lambda \asymp \omega_T$ for the lasso regression in (5) and $\lambda_j \asymp \omega_T$ uniformly in j for the lasso for nodewise regression in (13). Assume $(s_0 \vee \bar{s}^2) \log(p)/\sqrt{T} = o(1)$. Then

$$\sqrt{T}(\widehat{\mathbf{w}}_{MRC,DEBIASED} - \mathbf{w}) = W + \Delta,$$
$$W = \widehat{\mathbf{\Theta}}\mathbf{R}'\mathbf{e}/\sqrt{T},$$
$$\|\Delta\|_\infty \lesssim_P (s_0 \vee \bar{s}^2) \log(p)/\sqrt{T} = o_p(1).$$

Asymptotic Properties of Post-Lasso Portfolio Weights

Theorem 4 (Consistency of post-Lasso weights estimator)

Suppose the restricted eigenvalue condition and the restricted sparse eigenvalue condition on the empirical Gram matrix hold (see Condition RE($\bar{c}$) and Condition RSE(m) of Belloni & Chernozhukov, 2013, p. 529). Let $\widehat{\mathbf{w}}$ be the post-Lasso MRC weight estimator from Algorithm 1, we have

$$\|\widehat{\mathbf{w}} - \mathbf{w}\|_1 \lesssim_P \begin{cases} \sigma_e\Big((s_0\omega_T) \vee (\bar{s}^2\omega_T)\Big), & \textit{in general,} \\ \sigma_e s_0\Big(\sqrt{\frac{1}{T}} + \frac{1}{\sqrt{p}}\Big), & \textit{if } s_0 \geq \bar{s}^2 \textit{ and } \Xi = \hat{\Xi} \text{ wp} \to 1. \end{cases}$$

Gaussian Setting: DGP

$$\underbrace{\mathbf{r}_t}_{p\times 1} = \mathbf{m} + \mathbf{B}\underbrace{\mathbf{f}_t}_{K\times 1} + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T$$

- $\mathbf{m}_i \sim \mathcal{N}(1, 1)\ \forall i = 1, \ldots, p$
- $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$, with $(\boldsymbol{\Sigma}_\varepsilon)_{ij} = \rho^{|i-j|}$, $i, j \in 1, \ldots, p$
- $\mathbf{f}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_f = \mathbf{I}_K/10)$; $\mathbf{B} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_K/100)$
- **Case 1**: low-dimensional with $p < T$: $p = T^\delta$, $\delta = 0.85$ and $T = [2^h]$, for $h = 7, 7.5, 8, \ldots, 9.5$.
- **Case 2**: high-dimensional with $p > T$: $p = 3 \cdot T^\delta$, $\delta = 0.85$, all else equal.
- We set $\rho = 0.5$ and fix the number of factors $K = 3$.

Gaussian Setting: DGP

Let $\boldsymbol{\Sigma} = \mathbf{B}\boldsymbol{\Sigma}_f\mathbf{B}' + \boldsymbol{\Sigma}_\varepsilon$. Recall,

$$\mathbf{w}_{\text{MRC}} = \frac{\sigma}{\sqrt{\mathbf{m}'\boldsymbol{\Theta}\mathbf{m}}}\boldsymbol{\Theta}\mathbf{m} = \frac{\sigma}{\sqrt{\theta}}\boldsymbol{\alpha}, \text{ where } \boldsymbol{\alpha} := \boldsymbol{\Theta}\mathbf{m}$$

To create sparse MRC portfolio weights we use the following procedure:

1. Threshold the vector $\boldsymbol{\alpha} = \boldsymbol{\Theta}\mathbf{m}$ to keep the top $p/2$ entries with largest absolute values $\rightarrow$ sparse $\boldsymbol{\alpha}$.
2. Use $\boldsymbol{\Sigma}\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$ as the values for the mean and covariance matrix parameters to generate multivariate Gaussian returns

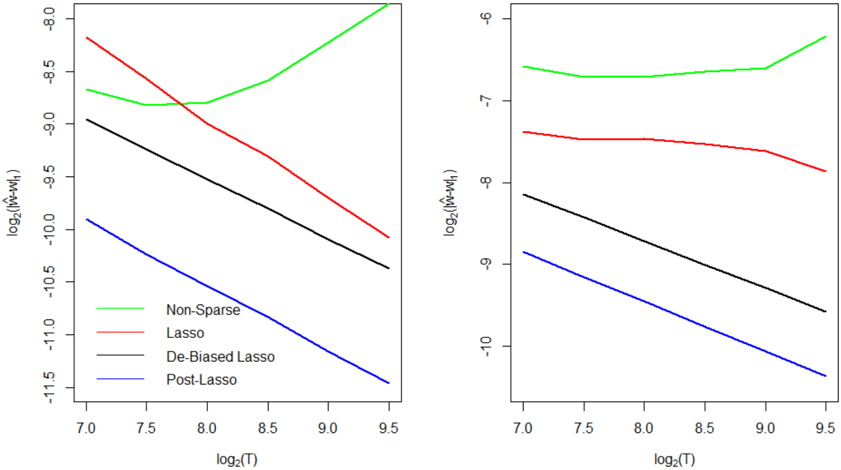


Figure 2: Averaged errors of the estimators of $\mathbf{w}_{MRC}$ for Case 1 (left) and for Case 2 (right).

Elliptical Distribution Setting: DGP

$$\underbrace{\mathbf{r}_t}_{p\times 1} = \mathbf{m} + \mathbf{B}\underbrace{\mathbf{f}_t}_{K\times 1} + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T$$

- Let $(\mathbf{f}_t, \boldsymbol{\varepsilon}_t)$ jointly follow the multivariate t-distribution with the degrees of freedom ν.
- Draw T independent samples of $(\mathbf{f}_t, \boldsymbol{\varepsilon}_t)$ from the multivariate t-distribution with zero mean and covariance matrix $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_f, \boldsymbol{\Sigma}_\varepsilon)$.
- $\boldsymbol{\Sigma}_f = \mathbf{I}_K$ and $\boldsymbol{\Sigma}_\varepsilon$ has a Toeplitz structure parameterized by $\rho = 0.5$.

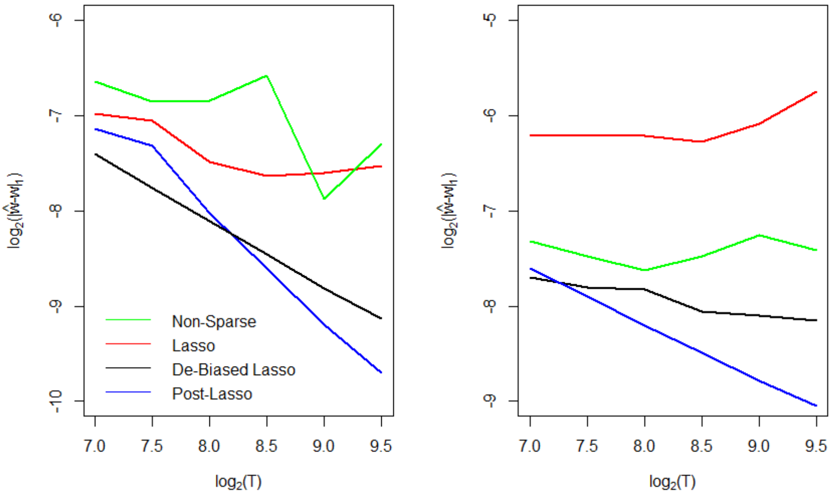


Figure 3: Elliptical Distribution ($\nu = 4.2$): Averaged errors of the estimators of $\mathbf{w}_{\text{MRC}}$ for Case 1 (left) and for Case 2 (right).

Data

- **Data:** CRSP and Compustat, monthly returns of the components of the S&P500:
  - Full sample: 480 observations on 355 stocks from January 1, 1980 - December 1, 2019.
  - Training: January 1, 1980 - December 1, 1994 (180 obs).
  - Test: January 1, 1995 - December 1, 2019 (300 obs).
  - Monthly rebalancing.
  - The composite index is reported as ^GSPC.
- **Factors:** Fama-French factors (FF), statistical factors (PC).
- **Targets:** return target $\mu \in \{0.7974\%, 0.0378\%\}$, risk target $\sigma \in \{0.05, 0.013\}$ for monthly and daily data.

Finding # 1: Relaxing the constraint that weights sum up to one leads to higher return and SR, however, the risk-constraint $\sigma = 0.05$ is violated:

	MRC			MWC			GMVP		
	Return	**Risk**	**SR**	**Return**	**Risk**	**SR**	**Return**	**Risk**	**SR**
EW	0.0081	0.0520	0.1553						
Index	0.0063	0.0458	0.1389						
MB	0.0539	0.2522	0.2138	0.0070	0.0021	0.1539	0.0082	0.0020	0.1860
FMB (PC1)	0.0324	0.1865	0.1737	0.0072	0.0447	0.1616	0.0080	0.0436	0.1818
FMB (PC2)	0.0287	0.1049	0.2743	0.0069	0.0346	0.1968	0.0076	0.0346	0.2211
FMB (PC3)	0.0228	0.0911	0.2498	0.0059	0.0332	0.1817	0.0058	0.0332	0.1766
FMB (FF1)	0.0497	0.2200	0.2258	0.0071	0.0447	0.1582	0.0083	0.0436	0.1921
FMB (FF3)	0.0384	0.1319	0.2908	0.0067	0.0387	0.1754	0.0080	0.0361	0.2223
FMB (FF5)	0.0373	0.1277	0.2921	0.0068	0.0374	0.1788	0.0081	0.0361	0.2250

Table 3: Non-Sparse portfolios: Monthly portfolio returns, risk and Sharpe ratio. Targeted risk is set at $\sigma = 0.05$

Finding # 1: Relaxing the constraint that weights sum up to one leads to higher return and SR, however, the risk-constraint $\sigma = 0.05$ is violated:

	MRC			MWC			GMVP		
	Return	**Risk**	**SR**	**Return**	**Risk**	**SR**	**Return**	**Risk**	**SR**
EW	0.0081	0.0520	0.1553						
Index	0.0063	0.0458	0.1389						
MB	0.0539	0.2522	0.2138	0.0070	0.0021	0.1539	0.0082	0.0020	0.1860
FMB (PC1)	0.0324	0.1865	0.1737	0.0072	0.0447	0.1616	0.0080	0.0436	0.1818
FMB (PC2)	0.0287	0.1049	0.2743	0.0069	0.0346	0.1968	0.0076	0.0346	0.2211
FMB (PC3)	0.0228	0.0911	0.2498	0.0059	0.0332	0.1817	0.0058	0.0332	0.1766
FMB (FF1)	0.0497	0.2200	0.2258	0.0071	0.0447	0.1582	0.0083	0.0436	0.1921
FMB (FF3)	0.0384	0.1319	0.2908	0.0067	0.0387	0.1754	0.0080	0.0361	0.2223
FMB (FF5)	0.0373	0.1277	0.2921	0.0068	0.0374	0.1788	0.0081	0.0361	0.2250

Table 3: Non-Sparse portfolios: Monthly portfolio returns, risk and Sharpe ratio. Targeted risk is set at $\sigma = 0.05$

Finding # 2: Factor Graphical Models outperform EW portfolio and Index in terms of return and SR, however, they have higher risk:

	MRC			MWC			GMVP		
	Return	**Risk**	**SR**	**Return**	**Risk**	**SR**	**Return**	**Risk**	**SR**
EW	0.0081	0.0520	0.1553						
Index	0.0063	0.0458	0.1389						
MB	0.0539	0.2522	0.2138	0.0070	0.0021	0.1539	0.0082	0.0020	0.1860
FMB (PC1)	0.0324	0.1865	0.1737	0.0072	0.0447	0.1616	0.0080	0.0436	0.1818
FMB (PC2)	0.0287	0.1049	0.2743	0.0069	0.0346	0.1968	0.0076	0.0346	0.2211
FMB (PC3)	0.0228	0.0911	0.2498	0.0059	0.0332	0.1817	0.0058	0.0332	0.1766
FMB (FF1)	0.0497	0.2200	0.2258	0.0071	0.0447	0.1582	0.0083	0.0436	0.1921
FMB (FF3)	0.0384	0.1319	0.2908	0.0067	0.0387	0.1754	0.0080	0.0361	0.2223
FMB (FF5)	0.0373	0.1277	0.2921	0.0068	0.0374	0.1788	0.0081	0.0361	0.2250

Table 3: Non-Sparse portfolios: Monthly portfolio returns, risk and Sharpe ratio. Targeted risk is set at $\sigma = 0.05$

Finding # 3: In most cases, factor-based portfolios outperform non-factor-based counterparts in terms of return and SR:

	MRC			MWC			GMVP		
	Return	**Risk**	**SR**	**Return**	**Risk**	**SR**	**Return**	**Risk**	**SR**
EW	0.0081	0.0520	0.1553						
Index	0.0063	0.0458	0.1389						
MB	0.0539	0.2522	0.2138	0.0070	0.0021	0.1539	0.0082	0.0020	0.1860
FMB (PC1)	0.0324	0.1865	0.1737	0.0072	0.0447	0.1616	0.0080	0.0436	0.1818
FMB (PC2)	0.0287	0.1049	0.2743	0.0069	0.0346	0.1968	0.0076	0.0346	0.2211
FMB (PC3)	0.0228	0.0911	0.2498	0.0059	0.0332	0.1817	0.0058	0.0332	0.1766
FMB (FF1)	0.0497	0.2200	0.2258	0.0071	0.0447	0.1582	0.0083	0.0436	0.1921
FMB (FF3)	0.0384	0.1319	0.2908	0.0067	0.0387	0.1754	0.0080	0.0361	0.2223
FMB (FF5)	0.0373	0.1277	0.2921	0.0068	0.0374	0.1788	0.0081	0.0361	0.2250

Table 3: Non-Sparse portfolios: Monthly portfolio returns, risk and Sharpe ratio. Targeted risk is set at $\sigma = 0.05$

Finding # 4: De-biasing leads to higher return and SR. Despite higher risk of de-biased portfolios, the risk constraint $\sigma = 0.05$ is satisfied:

Sparse MRC			
	Return	**Risk**	**SR**
Lasso (PC0)	0.0007	0.0048	0.1406
Debiased Lasso (PC0)	0.0023	0.0088	0.2266
Lasso (PC1)	0.0008	0.0059	0.1350
Debiased Lasso (PC1)	0.0091	0.0300	0.3117
Lasso (PC2)	0.0006	0.0052	0.1122
Debiased Lasso (PC2)	0.0067	0.0265	0.2542
Lasso (PC3)	0.0007	0.0064	0.1051
Debiased Lasso (PC3)	0.0064	0.0245	0.2603
Lasso (FF1)	0.0007	0.0039	0.1902
Debiased Lasso (FF1)	0.0109	0.0346	0.3213
Lasso (FF3)	0.0004	0.0040	0.1113
Debiased Lasso (FF3)	0.0072	0.0265	0.2721
Lasso (FF5)	0.0002	0.0042	0.0577
Debiased Lasso (FF5)	0.0073	0.0300	0.2467

Table 4: Sparse portfolio: Monthly portfolio returns, risk and Sharpe ratio. Targeted risk is set at $\sigma = 0.05$.

Define the excess portfolio return at time $t+1$ with transaction costs, $c = 50$bps, as

$$r_{t+1,\text{portfolio}} = \widehat{\mathbf{w}}_t'\mathbf{r}_{t+1} - c(1 + \widehat{\mathbf{w}}_t'\mathbf{r}_{t+1}) \sum_{j=1}^{p} \left| \hat{w}_{t+1,j} - \hat{w}_{t,j}^{+} \right|,$$

$$\text{where} \quad \hat{w}_{t,j}^{+} = \hat{w}_{t,j} \frac{1 + r_{t+1,j} + r_{t+1}^{f}}{1 + r_{t+1,\text{portfolio}} + r_{t+1}^{f}}$$

- $r_{t+1,j} + r_{t+1}^{f}$ is sum of the excess return of the j-th asset and risk-free rate,
- $r_{t+1,\text{portfolio}} + r_{t+1}^{f}$ is the sum of the excess return of the portfolio and risk-free rate

$$\text{Turnover} = \frac{1}{T-m} \sum_{t=m}^{T-1} \sum_{j=1}^{p} \left| \hat{w}_{t+1,j} - \hat{w}_{t,j}^{+} \right|, \; m - \text{training sample}$$

Finding # 5: De-biased sparse portfolios have lower risk, turnover and return compared to post-Lasso and non-sparse counterparts. The OOS SR is, overall, comparable:

	De-Biasing				Post-Lasso				Non-Sparse			
	MRC				MRC				MRC			
	Return	**Risk**	**SR**	**T/O**	**Return**	**Risk**	**SR**	**T/O**	**Return**	**Risk**	**SR**	**T/O**
PC0	0.0023	0.0100	0.2266	0.7952	0.0287	0.1217	0.2362	2.1249	0.0539	0.2522	0.2138	2.9458
PC1	0.0091	0.0300	0.3117	1.0562	0.0242	0.1187	0.2034	2.272	0.0324	0.1865	0.1737	3.6900
PC2	0.0067	0.0265	0.2542	1.2113	0.0290	0.1005	0.2882	2.1756	0.0287	0.1049	0.2743	3.7190
PC3	0.0064	0.0245	0.2603	1.2337	0.0312	0.1118	0.2784	2.2449	0.0228	0.0911	0.2498	3.8783
FF1	0.0109	0.0346	0.3213	0.8298	0.0207	0.1192	0.1738	2.1589	0.0497	0.2200	0.2258	2.7245
FF3	0.0072	0.0265	0.2721	0.9142	0.0157	0.1245	0.1263	2.2245	0.0384	0.1319	0.2908	2.4670
FF5	0.0073	0.0300	0.2467	0.9507	0.0212	0.1127	0.1879	2.2542	0.0373	0.1277	0.2921	2.4853

Table 5: Sparse vs Non-sparse portfolio: Monthly portfolio returns, risk, Sharpe ratio and turnover. Targeted risk is set at $\sigma = 0.05$.

Finding # 6: Sparse post-Lasso using PCs have, overall, lower risk, turnover and higher OOS SR compared to non-sparse counterparts:

	De-Biasing				Post-Lasso				Non-Sparse			
	MRC				MRC				MRC			
	Return	**Risk**	**SR**	**T/O**	**Return**	**Risk**	**SR**	**T/O**	**Return**	**Risk**	**SR**	**T/O**
PC0	0.0023	0.0100	0.2266	0.7952	0.0287	0.1217	0.2362	2.1249	0.0539	0.2522	0.2138	2.9458
PC1	0.0091	0.0300	0.3117	1.0562	0.0242	0.1187	0.2034	2.272	0.0324	0.1865	0.1737	3.6900
PC2	0.0067	0.0265	0.2542	1.2113	0.0290	0.1005	0.2882	2.1756	0.0287	0.1049	0.2743	3.7190
PC3	0.0064	0.0245	0.2603	1.2337	0.0312	0.1118	0.2784	2.2449	0.0228	0.0911	0.2498	3.8783
FF1	0.0109	0.0346	0.3213	0.8298	0.0207	0.1192	0.1738	2.1589	0.0497	0.2200	0.2258	2.7245
FF3	0.0072	0.0265	0.2721	0.9142	0.0157	0.1245	0.1263	2.2245	0.0384	0.1319	0.2908	2.4670
FF5	0.0073	0.0300	0.2467	0.9507	0.0212	0.1127	0.1879	2.2542	0.0373	0.1277	0.2921	2.4853

Table 5: Sparse vs Non-sparse portfolio: Monthly portfolio returns, risk, Sharpe ratio and turnover. Targeted risk is set at $\sigma = 0.05$.

Finding # 7: Imposing the constraint that portfolio weights sum up to one deteriorates performance: CER and risk are decreased. This is amplified during economic downturns.

	Asian & Rus. Fin. Crisis (1997-1998)		Argen. Great Depr. & dot-com bubble (1999-2002)		Fin. Crisis (2007-2009)		Ukr. Crisis & Chin. St Mkt Crash (2013-2015)	
	CER	**Risk**	**CER**	**Risk**	**CER**	**Risk**	**CER**	**Risk**
EW	0.2712	0.0547	-0.0322	0.0519	-0.4987	0.1203	0.2343	0.0344
Index	0.3222	0.0508	-0.1698	0.0539	-0.4924	0.0929	0.2155	0.0364
Markowitz Risk-Constrained (MRC)								
MB	2.1662	0.3381	-0.1140	0.2916	-3.0688	0.5101	0.9755	0.0847
FMB (PC1)	1.3285	0.0892	0.4241	0.1297	-3.0470	0.4735	1.2022	0.1658
FMB (PC2)	1.3153	0.0883	0.5016	0.1286	0.1312	0.1219	1.0534	0.1359
FMB (FF1)	2.0379	0.3029	0.0861	0.2660	-2.7247	0.4301	1.5366	0.1751
Global Minimum-Variance Portfolio (GMV)								
MB	0.2791	0.0496	-0.0470	0.0476	-0.4637	0.1015	0.1601	0.0340
FMB (PC1)	0.3960	0.0374	-0.1224	0.0510	-0.4588	0.0987	0.2656	0.0337
FMB (PC2)	0.4117	0.0364	-0.1227	0.0505	-0.3444	0.0393	0.3100	0.0321
FMB (FF1)	0.2784	0.0487	-0.0396	0.0468	-0.4570	0.0986	0.2640	0.0332

Finding # 8: Most non-sparse portfolios produced large negative CER during Fin. crisis 2007-09.

	Asian & Rus. Fin. Crisis (1997-1998)		Argen. Great Depr. & dot-com bubble (1999-2002)		Fin. Crisis (2007-2009)		Ukr. Crisis & Chin. St Mkt Crash (2013-2015)	
	CER	**Risk**	**CER**	**Risk**	**CER**	**Risk**	**CER**	**Risk**
EW	0.2712	0.0547	-0.0322	0.0519	-0.4987	0.1203	0.2343	0.0344
Index	0.3222	0.0508	-0.1698	0.0539	-0.4924	0.0929	0.2155	0.0364
Markowitz Risk-Constrained (MRC)								
MB	2.1662	0.3381	-0.1140	0.2916	-3.0688	0.5101	0.9755	0.0847
FMB (PC1)	1.3285	0.0892	0.4241	0.1297	-3.0470	0.4735	1.2022	0.1658
FMB (PC2)	1.3153	0.0883	0.5016	0.1286	0.1312	0.1219	1.0534	0.1359
FMB (FF1)	2.0379	0.3029	0.0861	0.2660	-2.7247	0.4301	1.5366	0.1751
Global Minimum-Variance Portfolio (GMV)								
MB	0.2791	0.0496	-0.0470	0.0476	-0.4637	0.1015	0.1601	0.0340
FMB (PC1)	0.3960	0.0374	-0.1224	0.0510	-0.4588	0.0987	0.2656	0.0337
FMB (PC2)	0.4117	0.0364	-0.1227	0.0505	-0.3444	0.0393	0.3100	0.0321
FMB (FF1)	0.2784	0.0487	-0.0396	0.0468	-0.4570	0.0986	0.2640	0.0332

Finding # 9: Risk of de-biased sparse portfolios is low even during economic downturns. However, in contrast to non-sparse portfolios, de-biased sparse portfolios that use PCs always produce positive CER.

	Asian & Rus. Fin. Crisis (1997-1998)		Argen. Great Depr. & dot-com bubble (1999-2002)		Fin. Crisis (2007-2009)		Ukr. Crisis & Chin. St Mkt Crash (2013-2015)	
	CER	**Risk**	**CER**	**Risk**	**CER**	**Risk**	**CER**	**Risk**
EW	0.2712	0.0547	-0.0322	0.0519	-0.4987	0.1203	0.2343	0.0344
Index	0.3222	0.0508	-0.1698	0.0539	-0.4924	0.0929	0.2155	0.0364
				Debiased MRC				
DL(PC1)	0.3553	0.0300	0.0674	0.0251	0.0440	0.0341	0.2066	0.0153
DL(PC2)	0.2962	0.0261	0.0540	0.0304	0.1129	0.0408	0.1567	0.0217
DL(FF1)	0.4149	0.0277	-0.0852	0.0275	-0.0258	0.0230	0.1681	0.0240
DL(FF3)	0.2123	0.0142	-0.0640	0.0300	-0.0406	0.0202	0.1782	0.0186

Finding # 10: Post-Lasso based portfolios have higher CER and higher risk compared to de-biased sparse portfolios, however, Post-Lasso portfolios have higher CER than non-sparse counterparts.

	Asian & Rus. Fin. Crisis (1997-1998)		Argen. Great Depr. & dot-com bubble (1999-2002)		Fin. Crisis (2007-2009)		Ukr. Crisis & Chin. St Mkt Crash (2013-2015)	
	CER	**Risk**	**CER**	**Risk**	**CER**	**Risk**	**CER**	**Risk**
EW	0.2712	0.0547	-0.0322	0.0519	-0.4987	0.1203	0.2343	0.0344
Index	0.3222	0.0508	-0.1698	0.0539	-0.4924	0.0929	0.2155	0.0364
Post-Lasso MRC								
PL(PC1)	3.0881	0.2211	1.8399	0.1445	6.6131	0.2962	-0.2484	0.1281
PL(PC2)	2.3196	0.1401	4.5196	0.2173	-0.8538	0.1884	1.7153	0.1253
PL(FF1)	2.3433	0.1568	-0.9059	0.1121	2.8639	0.2404	1.4470	0.1828
PL(FF3)	0.6691	0.1887	7.8165	0.2595	-0.9998	0.1410	-0.1561	0.1799
Post-Lasso GMV								
PL(PC1)	0.4467	0.0693	0.1031	0.0649	-0.0206	0.1108	-0.0934	0.0371
PL(PC2)	0.4403	0.0593	0.8150	0.0955	-0.3694	0.1243	0.0698	0.0492
PL(FF1)	0.3385	0.0616	0.8151	0.0877	-0.5545	0.1213	0.1754	0.0452
PL(FF3)	0.0711	0.0713	0.1458	0.1061	0.0295	0.0694	0.1634	0.0463

Conclusions

1. We propose the Factor Nodewise Regression (FMB) to estimate HD precision matrix, which combines the benefits of graphical models and factor structure; we prove consistency of FMB.
2. We use de-biasing and post-lasso to propose two consistent sparse estimators of portfolio weights
3. We develop a simple approach based on FMB for factor investing.
4. An empirical application to the constituents of S&P500 highlights the advantages of FMB approach and sparse portfolio.

Work in Progress: time-varying FGM; projected FGM (incorporate info on the companies' fundamentals when selecting stocks); optimal # stocks in a sparse portfolio.

thank you!

Proof of (4)

$$\mathbb{E}\left[\mathbf{r}_c - \mathbf{w}'\mathbf{r}_t\right]^2 = r_c^2 + \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} + (\mathbf{w}'\mathbf{m})^2 - 2r_c\mathbf{w}'\mathbf{m}$$

$$\text{F.O.C.} \quad \mathbf{w}\boldsymbol{\Sigma} + (\mathbf{w}'\mathbf{m})\mathbf{m} - r_c\mathbf{m} = 0$$

Left multiply both parts by $\mathbf{m}'\boldsymbol{\Theta}$:

$$\mathbf{w}'\mathbf{m} + (\mathbf{w}'\mathbf{m}) \cdot \mathbf{m}'\boldsymbol{\Theta}\mathbf{m} - r_c \cdot \mathbf{m}'\boldsymbol{\Theta}\mathbf{m} = 0, \tag{18}$$

Recall, $\theta = \mathbf{m}'\boldsymbol{\Theta}\mathbf{m}$,

$$\mathbf{w}'\mathbf{m} = \frac{\theta}{1+\theta} r_c = \zeta, \tag{19}$$

Combine (18) and (19):

$$\mathbf{w} = \frac{\xi}{\theta}\boldsymbol{\Theta}\mathbf{m}, \text{ if } r_c = \sigma\frac{1+\theta}{\sqrt{\theta}}, \text{ and } \mu = \sigma\sqrt{\mathbf{m}'\boldsymbol{\Theta}\mathbf{m}} \tag{20}$$

$$\mathbf{w}_{\text{MRC}} = \operatorname*{argmin}_{\mathbf{w}} \mathbb{E}\left[r_c - \mathbf{w}'\mathbf{r}_t\right]^2 = \frac{\sigma}{\sqrt{\theta}}\boldsymbol{\Theta}\mathbf{m} \quad \square \tag{21}$$

Go back

Figure 4: Stocks selected by Post-Lasso in August, 2019

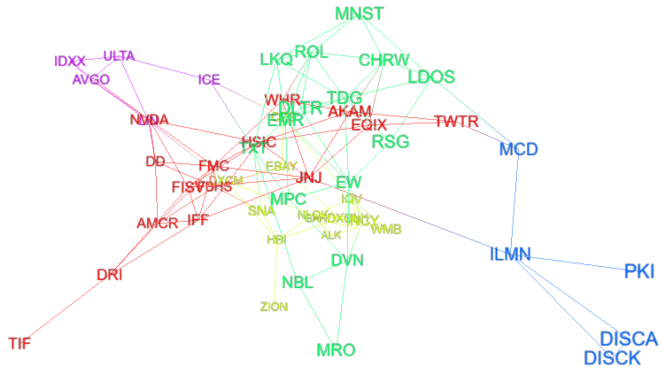


Figure 5: Stocks selected by Post-Lasso in May, 2020