A Basket Half Full: Sparse Portfolios

(https://arxiv.org/abs/2011.04278)

Ekaterina Seregina

University of California, Riverside

January, 2021

1. Motivation

"The people closest to you hearing that you study Economics might say "<u>What stock should I invest in?</u>" "

Markowitz Mean-Variance Optimization

- Agent: Suppose you are an investor with some savings which you would like to invest in the financial market.
- We observe $i = 1, \ldots, p$ excess returns over $t = 1, \ldots, T$ period of time: $\mathbf{r}_t = (r_{1t}, \ldots, r_{pt})' \sim \mathcal{D}(\mathbf{m}, \boldsymbol{\Sigma})$, where $\mathcal{D}$ is sub-Gaussian or Elliptical.
- Want: To find the *optimal* portfolio, $\mathbf{w}'\mathbf{r}_t$.
- May be subject to the constraints on
  - (i) a desired expected return, $\mu = \mathbb{E}[\mathbf{w}'\mathbf{r}_t]$;
  - (ii) maximum risk, $\sigma = \sqrt{\text{Var}(\mathbf{w}'\mathbf{r}_t)}$;
  - (iii) weights sum up to one, $\mathbf{w}'\mathbf{r}_t = 1$.
- Need: To estimate:
  - (i) expected returns, $\mathbf{m}$;
  - (ii) inverse covariance (*precision*) matrix, $\boldsymbol{\Theta} \equiv \boldsymbol{\Sigma}^{-1}$.

Large Portfolio Optimization: Challenges

- Break down the search for the optimal weights:
  1. Which stocks to buy?
     - ➔ Buy all stocks – non-sparse portfolio;
     - ➔ Buy a subset of all stocks – sparse portfolio
  2. How much to invest in these stocks?
- Many stocks available for investing:
  S&P500; NASDAQ ($> 3,000$); Russell ($> 1,000$)
- Challenges of optimizing over a large number of assets:
  (i) Consistent estimation of portfolio weights and exposure when $p = p_T \to \infty$ as $T \to \infty$ and/or $p > T$;
  (ii) Using sample covariance matrix as an estimator of $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$ is not feasible;
  (iii) Easy to monitor, low rebalancing costs, and robust performance during recessions.

Large Portfolio Optimization: Challenges

- Break down the search for the optimal weights:
  1. Which stocks to buy?
     - ➔ Buy all stocks – non-sparse portfolio;
     - ➔ Buy a subset of all stocks – sparse portfolio
  2. How much to invest in these stocks?
- Many stocks available for investing:
  S&P500; NASDAQ $(> 3,000)$; Russell $(> 1,000)$
- Challenges of optimizing over a large number of assets:
  (i) Consistent estimation of portfolio weights and exposure when $p = p_T \to \infty$ as $T \to \infty$ and/or $p > T$;
  (ii) Using sample covariance matrix as an estimator of $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$ is not feasible;
  (iii) Easy to monitor, low rebalancing costs, and robust performance during recessions.

Agenda of the Existing Research on HD Portfolio Optimization: find an improved estimator of the inverse covariance matrix.

Non-Sparse Portfolios

1. **Covariance Matrix Estimators**: develop an improved estimator of the sample covariance matrix.
   - Shrinking eigenvalues of sample covariance matrix (Ledoit & Wolf, JMA 2004; RFS 2017)
   - Estimate covariance matrix under factor structure assuming spiked covariance model (Fan et al., AoS 2011; JRSSB 2013; AoS 2016)
2. **Precision Matrix Estimators**: no need to estimate covariance, obtain an improved estimator of inverse covariance matrix directly.
   - Estimate the elements of precision matrix column by column – nodewise regression (Meinshausen and Bühlmann, AoS 2006)
   - Constrained ℓ_1-minimization for inverse matrix estimation – CLIME (Cai et al., JASA 2011)

Non-Sparse Portfolios

1. **Covariance Matrix Estimators**: develop an improved estimator of the sample covariance matrix.
   - Shrinking eigenvalues of sample covariance matrix (Ledoit & Wolf, JMA 2004; RFS 2017)
   - Estimate covariance matrix under factor structure assuming spiked covariance model (Fan et al., AoS 2011; JRSSB 2013; AoS 2016)
2. **Precision Matrix Estimators**: no need to estimate covariance, obtain an improved estimator of inverse covariance matrix directly.
   - Estimate the elements of precision matrix column by column – nodewise regression (Meinshausen and Bühlmann, AoS 2006)
   - Constrained ℓ_1-minimization for inverse matrix estimation – CLIME (Cai et al., JASA 2011)

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**.

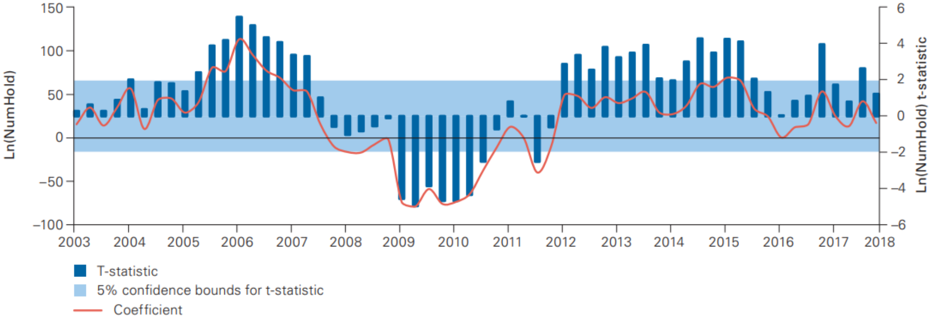
	Total OOS Performance 10/24/19–09/24/20			Before the Pandemic 01/02/19–12/31/19		During the Pandemic 01/02/20–06/30/20	
	Return ($\times 100$)	**Risk** ($\times 100$)	**Sharpe Ratio**	**CER** ($\times 100$)	**Risk** ($\times 100$)	**CER** ($\times 100$)	**Risk** ($\times 100$)
EW	0.0108	1.8781	0.0058	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
CLIME	0.0793	3.1279	0.0373	31.5294	1.0215	-25.3004	3.8972
LW	0.0317	1.7190	0.0184	29.5513	0.7924	-14.9328	3.0115

Table 1: Performance of non-sparse portfolios.

Let us Review Some Stylized Facts

Source: Tidmore et al. (The Journal of Investing, 2019)

Data: active US equity funds' quarterly data from January 2000 to December 2017 from Morningstar, Inc.
Goal: study the impact of the number of stock holdings on fund excess returns.



Additional Evidence

Sparse Portfolios

Hypothesis: holding sparse portfolios is the key to hedging risk during recessions
Previous Studies on Sparse Allocations (Brodie et al., PNAS 2009; Li, JBES 2015):

- Limited to low-dimensional setup (number of stocks ≤ 100)
- Lack theoretical guarantees of resulting sparse wealth allocations
- Are suboptimal due to the bias induced by the sparse penalty

"The Law of the Few (80/20 Principle) poses that in any situation roughly 80% of the 'work' will be done by 20% of the participants."

M. Gladwell (*The Tipping Point: How Little Things Can Make a Big Difference, 2000*)

This Paper

Develop a methodology to
construct sparse portfolio in high dimensions for several different portfolio formulations:

- **Step 1**: Reformulate standard constrained Markowitz optimization problem as an unconstrained regression problem
- **Step 2**: Apply Lasso penalty to induce sparsity in wealth allocations + correct for the bias
- **Step 3**: Graphical Model + Factor Structure = Factor Graphical model $\rightarrow$ precision matrix estimator under the factor model for HD portfolio allocation

Summary of Contributions

Theoretical Contributions:

- Establish the oracle bounds of sparse weight estimators, portfolio exposure and precision matrix estimator;
- Provide guidance regarding the distribution of portfolio weights

Empirical Contributions:

- Examine the merit of sparse portfolios during different market scenarios;
- Demonstrate that our strategy can be used as a hedging vehicle during economic downturns;
- Identify industries that serve as “safe havens” 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
CLIME	0.0793	3.1279	0.0373	31.5294	1.0215	-25.3004	3.8972
LW	0.0317	1.7190	0.0184	29.5513	0.7924	-14.9328	3.0115
Our Post-Lasso-based	0.1247	1.7254	0.0723	45.2686	1.0386	12.4196	2.8554
Our De-biased Estimator	0.0275	0.5231	0.0526	23.7629	0.4972	6.5813	0.5572

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

▸ Selected Stocks

2. Sparse Portfolios

Markowitz Risk-Constrained Problem (MRC)

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

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

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

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

$$\max_{\mathbf{w}} \frac{\mathbf{m}'\mathbf{w}}{\sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}} \text{ s.t. } \mathbf{m}'\mathbf{w} \geq \mu \text{ or } \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \leq \sigma^2 \tag{3}$$

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

$$\mathbf{w}_{\text{MRC}} = \frac{\sigma}{\sqrt{\mathbf{m}'\boldsymbol{\Theta}\mathbf{m}}}\boldsymbol{\Theta}\mathbf{m} = \frac{\sigma}{\sqrt{\theta}}\boldsymbol{\Theta}\mathbf{m}, \tag{4}$$

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

Markowitz Risk-Constrained Problem (MRC)

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

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

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

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

$$\max_{\mathbf{w}} \frac{\mathbf{m}'\mathbf{w}}{\sqrt{\mathbf{w}'\mathbf{\Sigma}\mathbf{w}}} \text{ s.t. } \mathbf{m}'\mathbf{w} \geq \mu \text{ or } \mathbf{w}'\mathbf{\Sigma}\mathbf{w} \leq \sigma^2 \tag{3}$$

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

$$\mathbf{w}_{\text{MRC}} = \frac{\sigma}{\sqrt{\mathbf{m}'\Theta\mathbf{m}}}\Theta\mathbf{m} = \frac{\sigma}{\sqrt{\theta}}\Theta\mathbf{m}, \tag{4}$$

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

Markowitz Risk-Constrained Problem (MRC)

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

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

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

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

$$\max_{\mathbf{w}} \frac{\mathbf{m}'\mathbf{w}}{\sqrt{\mathbf{w}'\mathbf{\Sigma}\mathbf{w}}} \text{ s.t. } \mathbf{m}'\mathbf{w} \geq \mu \text{ or } \mathbf{w}'\mathbf{\Sigma}\mathbf{w} \leq \sigma^2 \tag{3}$$

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

$$\mathbf{w}_{\text{MRC}} = \frac{\sigma}{\sqrt{\mathbf{m}'\mathbf{\Theta}\mathbf{m}}}\mathbf{\Theta}\mathbf{m} = \frac{\sigma}{\sqrt{\theta}}\mathbf{\Theta}\mathbf{m}, \tag{4}$$

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

Markowitz Risk-Constrained Problem (MRC)

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

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

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

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

$$\max_{\mathbf{w}} \frac{\mathbf{m}'\mathbf{w}}{\sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}} \text{ s.t. } \mathbf{m}'\mathbf{w} \geq \mu \text{ or } \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} \leq \sigma^2 \tag{3}$$

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

$$\mathbf{w}_{\text{MRC}} = \frac{\sigma}{\sqrt{\mathbf{m}'\boldsymbol{\Theta}\mathbf{m}}}\boldsymbol{\Theta}\mathbf{m} = \frac{\sigma}{\sqrt{\theta}}\boldsymbol{\Theta}\mathbf{m}, \tag{4}$$

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

How to Induce Sparsity?

Consider any linear regression model:

$$\underbrace{\mathbf{y}}_{T\times 1} = \underbrace{\mathbf{X}}_{T\times p}\underbrace{\boldsymbol{\beta}}_{p\times 1} + \boldsymbol{\varepsilon} \quad (5)$$

To solve for $\boldsymbol{\beta}$ we minimize residual sum of squares:

$$\boldsymbol{\beta} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, \mathbb{E}\left[y_t - \boldsymbol{\beta}'\mathbf{x}_t\right]^2 \quad (6)$$

Sample counterpart: $\hat{\boldsymbol{\beta}} = \operatorname{argmin}_{\boldsymbol{\beta}} \sum_{t=1}^{T}(y_t - \boldsymbol{\beta}'\mathbf{x}_t)^2$
To make it sparse we penalize elements of $\boldsymbol{\beta}$ with small weights:

$$\hat{\boldsymbol{\beta}}_{\text{Sparse}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{t=1}^{T}(y_t - \boldsymbol{\beta}'\mathbf{x}_t)^2 + 2\lambda\sum_{j=1}^{p}|\beta_j| \quad (7)$$

When $\lambda = 0 \to \boldsymbol{\beta}$ is NOT sparse, when $\lambda = \infty \to$ no element of $\boldsymbol{\beta}$ is selected.

Can We Apply Similar Idea to Find Sparse Portfolios?

Recall portfolio weight expression obtained from constrained Markowitz optimization problem (MRC):

$$\mathbf{w}_{\text{MRC}} = \frac{\sigma}{\sqrt{\mathbf{m}'\mathbf{\Theta}\mathbf{m}}}\mathbf{\Theta}\mathbf{m} = \frac{\sigma}{\sqrt{\theta}}\mathbf{\Theta}\mathbf{m}, \tag{8}$$

Question: can we obtain the same weight as in (8) using linear regression?

$$\underbrace{\mathbf{y}}_{T\times 1} = \underbrace{\mathbf{X}}_{T\times p}\underbrace{\mathbf{w}_{\text{MRC}}}_{p\times 1} + \varepsilon \tag{9}$$

$$\mathbf{w}_{\text{MRC}} = \underset{\mathbf{w}}{\operatorname{argmin}}\, \mathbb{E}\left[y_t - \mathbf{w}'\mathbf{x}_t\right]^2 \overset{?}{=} \frac{\sigma}{\sqrt{\theta}}\mathbf{\Theta}\mathbf{m} \tag{10}$$

Answer: YES, we just have to determine y_t and $\mathbf{x}_t$ such that the second equality in (10) holds.

This Paper

Develop a methodology to construct sparse portfolio in high dimensions for several different portfolio formulations:

- **Step 1**: Reformulate standard constrained Markowitz optimization problem as an unconstrained regression problem
- **Step 2**: Apply Lasso penalty to induce sparsity in wealth allocations + correct for the bias
- **Step 3**: Graphical Model + Factor Structure = Factor Graphical model $\rightarrow$ precision matrix estimator under the factor model for HD portfolio allocation

Sparse Portfolio

- Define

$$y_t = y \equiv \sigma \frac{1+\theta}{\sqrt{\theta}}, \text{ where } \theta = \mathbf{m}'\boldsymbol{\Theta}\mathbf{m} \tag{11}$$

- If y_t is determined by equation (11) and $\mathbf{x}_t = \mathbf{r}_t$, the solution to the following unconstrained regression problem yields MRC portfolio weights:

$$\mathbf{w}_{\text{MRC}} = \underset{\mathbf{w}}{\operatorname{argmin}}\, \mathbb{E}\left[y - \mathbf{w}'\mathbf{r}_t\right]^2 = \frac{\sigma}{\sqrt{\theta}}\boldsymbol{\Theta}\mathbf{m} \tag{12}$$

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

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

▸ Proof

This Paper

Develop a methodology to construct sparse portfolio in high dimensions for several different portfolio formulations:

- **Step 1**: Reformulate standard constrained Markowitz optimization problem as an unconstrained regression problem
- **Step 2**: Apply Lasso penalty to induce sparsity in wealth allocations + correct for the bias
- **Step 3**: Graphical Model + Factor Structure = Factor Graphical model $\rightarrow$ precision matrix estimator under the factor model for HD portfolio allocation

Sparse Portfolio

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

$$\widehat{\mathbf{w}}_{\text{MRC,SPARSE}} \equiv \widehat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1}{T} \sum_{t=1}^{T} (y - \mathbf{w}'\mathbf{r}_t)^2 + 2\lambda \sum_{j=1}^{p} |w_j| \quad (13)$$

- Tuning parameter λ: we use the first 2/3 of the training data (training window) to estimate weights and tune the shrinkage intensity λ in the remaining 1/3 of the training sample to yield the highest Sharpe Ratio (validation window).

Sparse Portfolio

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

$$\widehat{\mathbf{w}}_{\text{MRC,SPARSE}} \equiv \widehat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1}{T}\sum_{t=1}^{T}(y - \mathbf{w}'\mathbf{r}_t)^2 + 2\lambda \sum_{j=1}^{p} |w_j| \quad (14)$$

$$y_t = y \equiv \sigma \frac{1+\theta}{\sqrt{\theta}}, \text{ where } \theta = \mathbf{m}'\mathbf{\Theta}\mathbf{m}$$

- **Problems**: (a) the estimator in (14) is biased and (b) y needs to be estimated.
- **Solutions**: (a) use de-biasing (van de Geer et al., AoS 2014, Belloni et al., Biometrika 2015, Javanmard et al., AoS 2018).
- For now, suppose we have a consistent estimator of y, denoted as $\hat{y}$.

Sparse De-Biased Portfolio

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

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

$$\widehat{\mathbf{\Sigma}}(\widehat{\mathbf{w}} - \mathbf{w}) + \lambda\widehat{\mathbf{g}} = \mathbf{R}'\mathbf{e}/T. \tag{15}$$

Multiply both sides of (15) by $\widehat{\mathbf{\Theta}}$, add and subtract $(\widehat{\mathbf{w}} - \mathbf{w})$:

$$\widehat{\mathbf{w}} - \mathbf{w} + \widehat{\mathbf{\Theta}}\lambda\widehat{\mathbf{g}} = \widehat{\mathbf{\Theta}}\mathbf{R}'\mathbf{e}/T - \underbrace{\sqrt{T}(\widehat{\mathbf{\Theta}}\widehat{\mathbf{\Sigma}} - \mathbf{I}_p)(\widehat{\mathbf{w}} - \mathbf{w})}_{\Delta}/\sqrt{T} \tag{16}$$

$$\widehat{\mathbf{w}} = \mathbf{w} - \widehat{\mathbf{\Theta}}\lambda\widehat{\mathbf{g}} \tag{17}$$

$$\widehat{\mathbf{w}}_{\text{MRC,DEBIASED}} = \widehat{\mathbf{w}} + \widehat{\mathbf{\Theta}}\lambda\widehat{\mathbf{g}} = \widehat{\mathbf{w}} + \widehat{\mathbf{\Theta}}\mathbf{R}'(\widehat{\mathbf{y}} - \mathbf{R}\widehat{\mathbf{w}})/T, \tag{18}$$

$\widehat{\mathbf{g}}$ is a $p \times 1$ vector arising from the subgradient of $\sum_{j=1}^{p} |w_j|$.

Alternative Portfolio Formulations: Problem

- What happens when in addition to target return and risk constraints we require portfolio weights sum up to one?

$$\max_{\mathbf{w}} \frac{\mathbf{m}'\mathbf{w}}{\sqrt{\mathbf{w}'\mathbf{\Sigma}\mathbf{w}}} \text{ s.t. } \mathbf{m}'\mathbf{w} \geq \mu \text{ or } \mathbf{w}'\mathbf{\Sigma}\mathbf{w} \leq \sigma^2, \mathbf{w}'\boldsymbol{\iota} = 1 \tag{19}$$

  Why not add weight constraint to (19)?
- (19) has two solutions, when $\boldsymbol{\iota}'\mathbf{\Theta}\mathbf{m} < 0$ the maximum value cannot be achieved exactly (Maller & Turkington, 2003).

Alternative Portfolio Formulations: Solution

$$\min_{\mathbf{w}} \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}, \text{ s.t. } \mathbf{m}'\mathbf{w} \geq \mu, \ \mathbf{w}'\boldsymbol{\iota} = 1 \tag{20}$$

Global Minimum-Variance (GMV): 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 (13):

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

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

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

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 estimator of $\mathbf{y}$?

- Recall, $\theta = \mathbf{m}'\boldsymbol{\Theta}\mathbf{m}$, where $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$ is precision matrix, and

$$y \equiv \sigma \frac{1+\theta}{\sqrt{\theta}}$$

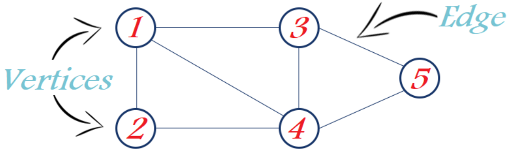
- If $\widehat{\mathbf{m}}$ is the sample mean, then $\|\widehat{\mathbf{m}} - \mathbf{m}\|_{\max} = \mathcal{O}_p(\sqrt{\log(p)/T})$ (Chang et. al., 2018)
- We need a consistent estimator of HD precision matrix $\boldsymbol{\Theta}$.

3. Factor Graphical Model

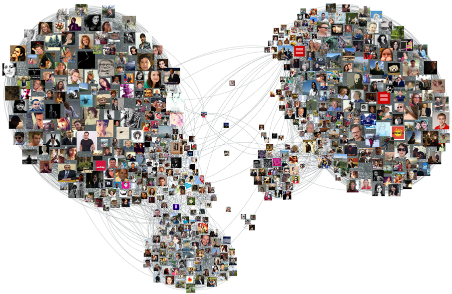
Existing Approaches to Estimate HD Precision Matrix

1. Graphical Models: estimate precision matrix directly (no need to obtain covariance matrix)

- Consistent estimation of sparse HD precision: Meinshausen & Buhlmann (Nodewise regression, AoS 2006), Friedman et al. (Graphical Lasso, Biostatistics 2008), Cai et al. (CLIME, JASA 2011)



- Absence of an edge between two vertices: conditionally independent given the rest.
- Graphical models aim at predicting agents' behavior by examining the web of relationships in which they are embedded.



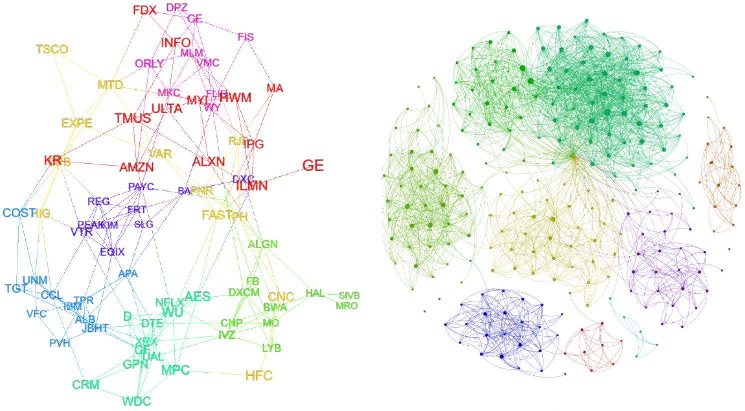
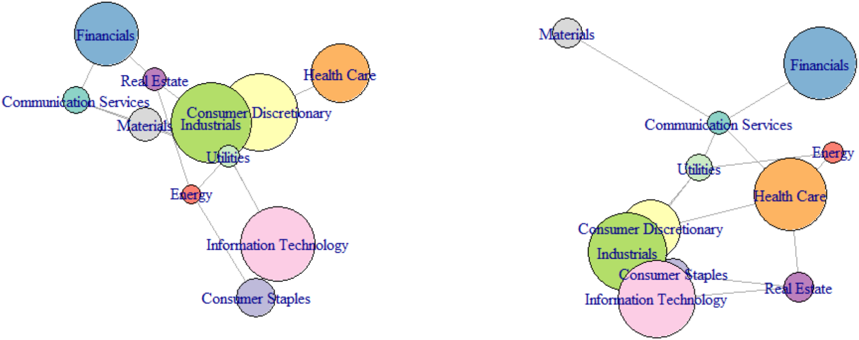


Figure 5: Stocks selected by Post-Lasso in May, 2020

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



Existing Approaches to Estimate HD Precision Matrix

1. Graphical Models: estimate precision matrix directly (no need to obtain covariance matrix)

- Consistent estimation of sparse HD precision: Meinshausen & Bühlmann (Nodewise regression, AoS 2006), Friedman et al. (Graphical Lasso, Biostatistics 2008), Cai et al. (CLIME, JASA 2011)

Strong sparsity assumption on precision matrix

When stock returns are driven by common factors, the sparsity assumption imposed by Graphical Models is too strong.

Do Stock Returns Exhibit the Factor Structure?

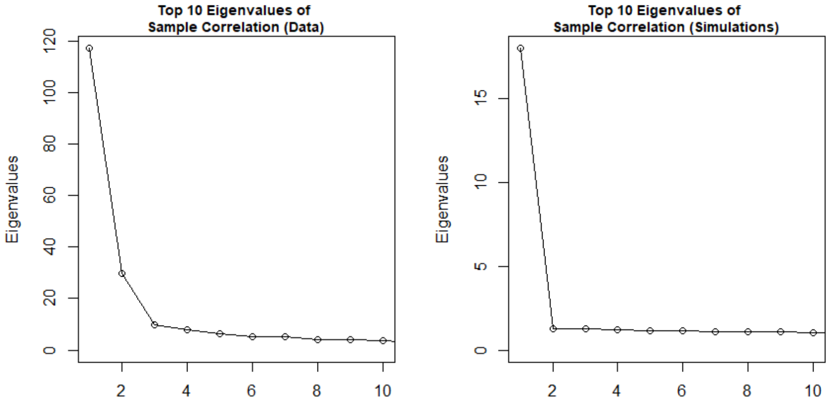


Figure 1A: Eigenvalues of the sample correlation matrix of S&P500 Index constituents daily returns during 2015–2020 (left) and the returns simulated using # factors$\equiv K = 2$ (right). [Similar pattern indicative of the factor structure was documented for S&P100 by Ding et al., JoE 2020].

Partial Correlations Under the Factor Structure

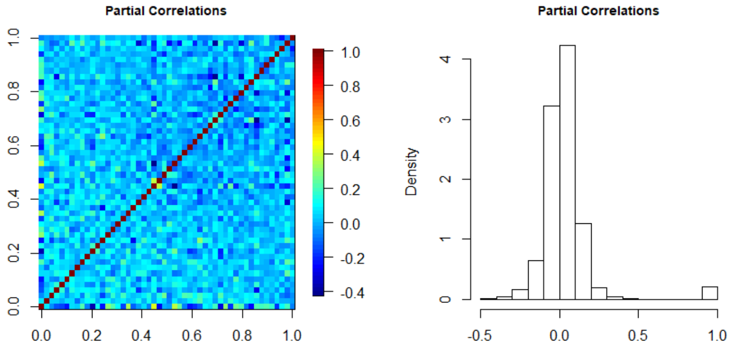


Figure 2A: Heatmap and histogram of sample partial correlations estimated using the sample correlation matrix: $T = 1000$, $p = 50$, $K = 2$.

Graphical Models Under the Factor Structure

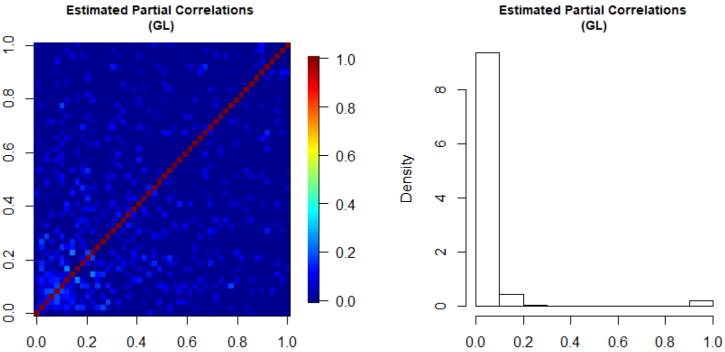


Figure 3A: Heatmap and histogram of sample partial correlations estimated using Graphical Lasso with no factors: $T = 1000$, $p = 50$, $K = 2$.

Existing Approaches to Estimate HD Precision Matrix

2. Factor Models:

$$\underbrace{\mathbf{r}_t}_{p\times 1} = \mathbf{B}\underbrace{\mathbf{f}_t}_{K\times 1} + \varepsilon_t, \quad t = 1, \ldots, T \tag{21}$$

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

1. Estimate covariance matrix using factor structure
2. Invert it to obtain precision matrix

This Paper

Develop a methodology to construct sparse portfolio in high dimensions for several different portfolio formulations:

Step 1: Reformulate standard constrained Markowitz optimization problem as an unconstrained regression problem

Step 2: Apply ℓ_1 penalty to induce sparsity in wealth allocations + correct for the bias

Step 3: Graphical Model + Factor Structure = Factor Graphical model $\rightarrow$ precision matrix estimator under the factor model for HD portfolio allocation

How to Use Graphical Models Under the Factor Structure?

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

$$\underbrace{\mathbf{R}}_{p\times T} = \underbrace{\mathbf{B}}_{p\times K}\mathbf{F} + \mathbf{E}. \tag{22}$$

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

$$\boldsymbol{\Sigma}_\varepsilon = T^{-1}\mathbf{E}\mathbf{E}'; \qquad \boldsymbol{\Theta}_\varepsilon = \boldsymbol{\Sigma}_\varepsilon^{-1},$$

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

$$\text{cov}(\mathbf{r}_t) = \boldsymbol{\Sigma} = \mathbf{B}\boldsymbol{\Sigma}_f\mathbf{B}' + \boldsymbol{\Sigma}_\varepsilon; \qquad \boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}.$$

Factor Nodewise Regression (FMB)

Solution:

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

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

$$\text{FMB} \rightarrow \widehat{\mathbf{\Theta}} = \underbrace{\widehat{\mathbf{\Theta}}_{\varepsilon}}_{\text{MB}} - \widehat{\mathbf{\Theta}}_{\varepsilon}\widehat{\mathbf{B}}[\underbrace{\widehat{\mathbf{\Theta}}_{f}}_{\text{FM}} + \widehat{\mathbf{B}}'\widehat{\mathbf{\Theta}}_{\varepsilon}\widehat{\mathbf{B}}]^{-1}\underbrace{\widehat{\mathbf{B}}'}_{\text{FM}}\widehat{\mathbf{\Theta}}_{\varepsilon} \tag{23}$$

Sketch of the FMB Algorithm (CRAN (beta version))

Partial Correlations Under the Factor Structure

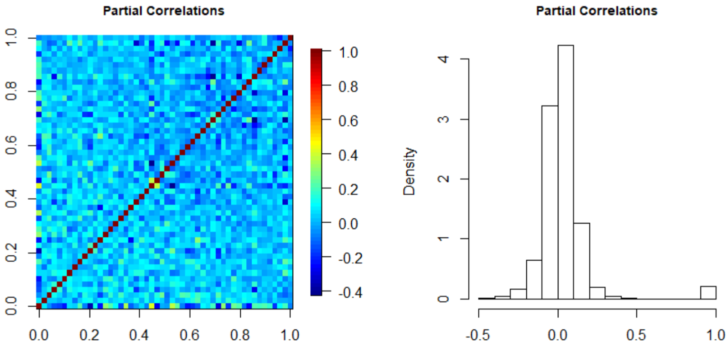


Figure 2A: Heatmap and histogram of sample partial correlations estimated using the sample correlation matrix: $T = 1000$, $p = 50$, $K = 2$.

Factor Graphical Models Under the Factor Structure

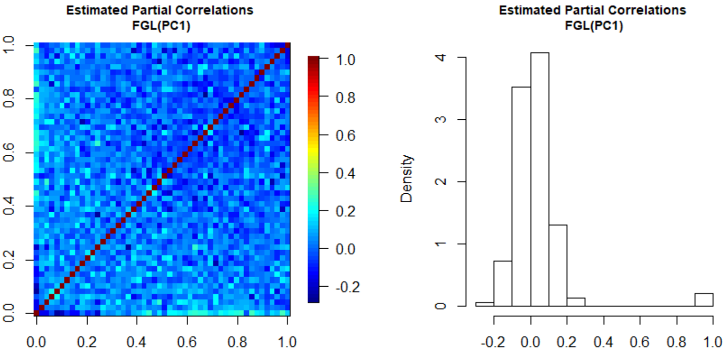


Figure 4A: Heatmap and histogram of sample partial correlations estimated using FGL with 1 estimated factor: $T = 1000$, $p = 50$, $K = 2$.

Factor Graphical Models Under the Factor Structure

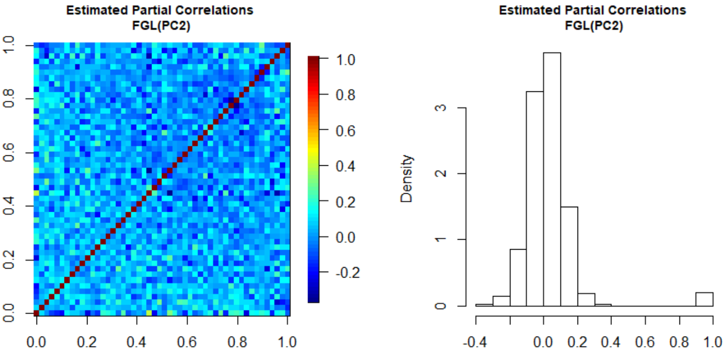


Figure 5A: Heatmap and histogram of sample partial correlations estimated using FGL with 2 estimated factors: $T = 1000$, $p = 50$, $K = 2$.

Factor Graphical Models Under the Factor Structure

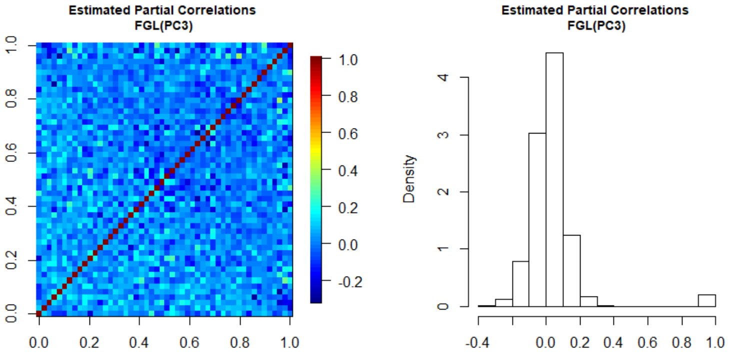


Figure 6A: Heatmap and histogram of sample partial correlations estimated using FGL with 3 estimated factors: $T = 1000$, $p = 50$, $K = 2$.

Theoretical Contributions: Summary

1. Establish consistency of HD precision estimator, $\widehat{\mathbf{\Theta}}$, using FMB:

$$\widehat{\mathbf{\Theta}} = \widehat{\mathbf{\Theta}}_{\varepsilon} - \widehat{\mathbf{\Theta}}_{\varepsilon}\widehat{\mathbf{B}}[\widehat{\mathbf{\Theta}}_{f} + \widehat{\mathbf{B}}'\widehat{\mathbf{\Theta}}_{\varepsilon}\widehat{\mathbf{B}}]^{-1}\widehat{\mathbf{B}}'\widehat{\mathbf{\Theta}}_{\varepsilon}. \text{ (Theorem 3)}$$

2. Use $\widehat{\mathbf{\Theta}}$, de-biasing and post-lasso to establish consistency of sparse $\widehat{\mathbf{w}}_{\xi}$, $\xi = \{\text{GMV, MWC, MRC}\}$ and their distribution. (Theorems 4,5)
3. Use $\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{\Theta}}^{-1}$ and $\widehat{\mathbf{w}}_{\xi}$ to establish the bounds on portfolio exposure $\widehat{\mathbf{w}}_{\xi}'\widehat{\mathbf{\Sigma}}\widehat{\mathbf{w}}_{\xi}$. (Theorem 6)

Elliptical distributions: Theorems 3-6 continue to hold. ✓

4. Empirical Application

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 31, 2019.
  - Training: January 1, 1980 - December 31, 1994 (180 obs).
  - Test: January 1, 1995 - December 31, 2019 (300 obs).
  - Monthly rebalancing.
  - The composite index is reported as ^GSPC.
- **Factors:** Fama-French factors (FF), statistical factors (PC):
  - **FF1** = excess return on the market; **FF3** = FF1 + Size Factor (SMB) + Value Factor (HML); **FF5** = FF3 + Profitability Factor (RMW) + Risk Factor (CMA)
- **Targets:**
  (return target, risk target) = $(\mu, \sigma) = (0.7974\%, 0.05)$.

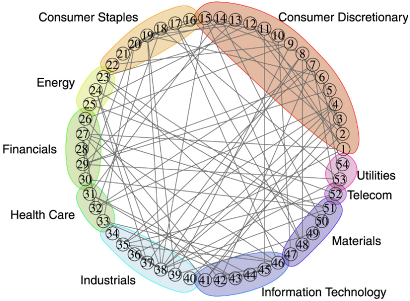
Finding # 1: Non-sparse portfolios produced large negative CER during the Global Fin. crisis 2007-09.

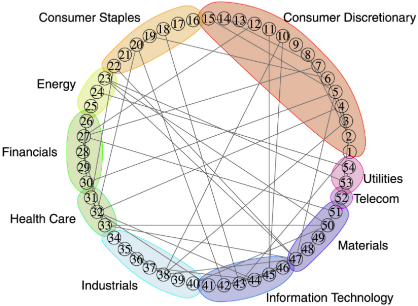
	Asian & Rus. Fin. Crisis (1997-1998)		Argen. Great Depr. & dot-com bubble (1999-2002)		Fin. Crisis (2007-2009)	
	CER	**Risk**	**CER**	**Risk**	**CER**	**Risk**
EW	0.2712	0.0547	-0.0322	0.0519	-0.4987	0.1203
Index	0.3222	0.0508	-0.1698	0.0539	-0.4924	0.0929
			Markowitz Risk-Constrained (MRC)			
MB	2.1662	0.3381	-0.1140	0.2916	-3.0688	0.5101
CLIME	1.3285	0.0892	0.4241	0.1297	-3.0470	0.4735
LW	0.9134	0.1021	0.3677	0.1412	-0.3196	0.3751
FMB (PC)	1.3153	0.0883	0.5016	0.1286	-0.1312	0.1219
FMB (FF1)	2.0379	0.3029	0.0861	0.2660	-2.7247	0.4301
			Global Minimum-Variance Portfolio (GMV)			
MB	0.2791	0.0496	-0.0470	0.0476	-0.4637	0.1015
CLIME	0.3960	0.0374	-0.1224	0.0510	-0.4588	0.0987
LW	0.3127	0.0415	-0.0952	0.0483	-0.4013	0.0693
FMB (PC)	0.4117	0.0364	-0.1227	0.0505	-0.3444	0.0393
FMB (FF1)	0.2784	0.0487	-0.0396	0.0468	-0.4570	0.0986

Finding # 2: Sparse portfolios produce positive CER during the Global Fin. crisis 2007-09.

	Asian & Rus. Fin. Crisis (1997-1998)		Argen. Great Depr. & dot-com bubble (1999-2002)		Fin. Crisis (2007-2009)	
	CER	**Risk**	**CER**	**Risk**	**CER**	**Risk**
EW	0.2712	0.0547	-0.0322	0.0519	-0.4987	0.1203
Index	0.3222	0.0508	-0.1698	0.0539	-0.4924	0.0929
			Debiased MRC			
DL(PC)	0.2962	0.0261	0.1567	0.0217	0.1129	0.0408
DL(FF1)	0.4149	0.0277	0.1681	0.0240	-0.0258	0.0230
DL(FF3)	0.2123	0.0142	0.1782	0.0186	-0.0406	0.0202
			Post-Lasso MRC			
PL(PC)	3.0881	0.2211	1.7153	0.1281	2.6131	0.1862
PL(FF1)	2.3433	0.1568	1.4470	0.1828	2.8639	0.2404
PL(FF3)	0.6691	0.1887	-0.1561	0.1799	-0.9998	0.1410
			Post-Lasso GMV			
PL(PC)	0.4403	0.0593	0.8150	0.0955	-0.3694	0.1243
PL(FF1)	0.3385	0.0616	0.8151	0.0877	-0.5545	0.1213
PL(FF3)	0.0711	0.0713	0.1458	0.1061	0.0295	0.0694

Partial Correlation Network of S&P500 Sub Industries in 2005 (left) and in 2008 (right).





Finding # 3: The market fear in the crisis broke the connections of stock sectors and sub industries.

5. Conclusions

Conclusions

Methodological Contributions:

- Develop a framework to construct sparse portfolio in high dimensions for several different portfolio formulations

Theoretical Contributions:

- Establish the oracle bounds of sparse weight estimators, portfolio exposure and precision matrix estimator;
- Provide guidance regarding the distribution of portfolio weights

Conclusions

Empirical Contributions:

- Examine the merit of sparse portfolios during different market scenarios;
- Demonstrate that our strategy can be used as a hedging vehicle during economic downturns
- Identify industries that could serve as safe havens during recessions:
  - ➔ Return-drivers during GFC and COVID-19 outbreak: **consumer staples, healthcare, retail and food**;
  - ➔ Least attractive investment: **insurance sector**

Future Work

- **Stocks and taxes**: the effect of Biden's tax proposals? "Behavioral responses" to changes in capital gains taxes: higher capital gains tax $\Rightarrow$ discourage capital gains realizations $\Rightarrow$ investors will likely respond by holding onto stocks rather than selling $\Rightarrow$ less efficient market.
- How to select **mutual funds** from the stocks they hold?
- Robo-Advisor using **Reinforcement Learning**
- **Time-Varying Networks** using autoregression to incorporate the past information (derive closed-form solutions for the ADMM subproblems to further speed up the runtime).

Thank you!

6. Appendix

$$\min_{\mathbf{w}} \ -U \equiv \frac{\gamma}{2}\mathbf{w}'\mathbf{\Sigma}\mathbf{w} - \mathbf{w}'\mathbf{m}, \text{ s.t. } \mathbf{w}'\boldsymbol{\iota} = 1, \ |\text{supp}(\mathbf{w})| \leq \bar{p}, \ \bar{p} \leq p$$

Monthly data on the constituents of the S&P100: set $\gamma = 3$, $\bar{p} = \{5, 10, \ldots, 90\}$ (Lagrangian relaxation procedure). If $\bar{p} < p \rightarrow U^{\text{Sparse}}$, if $\bar{p} = p \rightarrow U^{\text{Non-Sparse}}$:

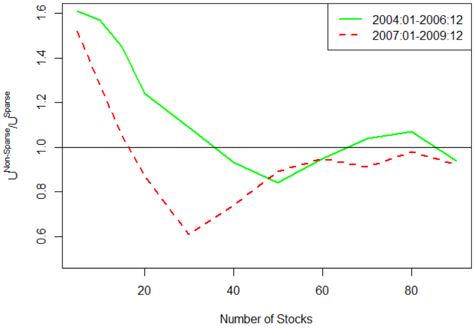


Figure 7A: $U^{\text{Non-Sparse}}/U^{\text{Sparse}}$ as a function of $\bar{p}$ averaged over the test window.

Go back

Proof of (12)

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

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

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

$$\mathbf{w}'\mathbf{m} + (\mathbf{w}'\mathbf{m}) \cdot \mathbf{m}'\boldsymbol{\Theta}\mathbf{m} - y \cdot \mathbf{m}'\boldsymbol{\Theta}\mathbf{m} = 0, \tag{24}$$

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

$$\mathbf{w}'\mathbf{m} = \frac{\theta}{1+\theta}y = \zeta, \tag{25}$$

Combine (24) and (25):

$$\mathbf{w} = \frac{\zeta}{\theta}\boldsymbol{\Theta}\mathbf{m}, \text{ if } y = \sigma\frac{1+\theta}{\sqrt{\theta}}, \text{ and } \mu = \sigma\sqrt{\mathbf{m}'\boldsymbol{\Theta}\mathbf{m}} \tag{26}$$

$$\mathbf{w}_{\text{MRC}} = \underset{\mathbf{w}}{\operatorname{argmin}}\, \mathbb{E}\left[y - \mathbf{w}'\mathbf{r}_t\right]^2 = \frac{\sigma}{\sqrt{\theta}}\boldsymbol{\Theta}\mathbf{m} \quad \square \tag{27}$$

Go back

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

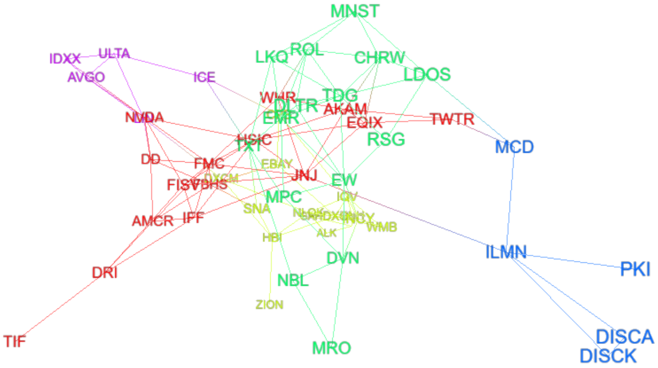
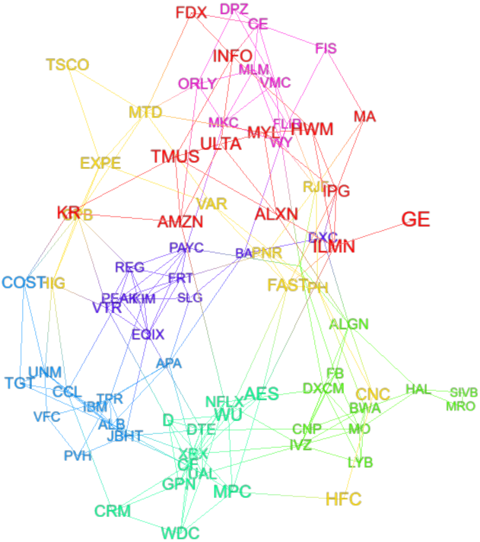


Figure 5: Stocks selected by Post-Lasso in May, 2020



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

- 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:

$$\widehat{\boldsymbol{\gamma}}_j = \arg\min_{\boldsymbol{\gamma}\in\mathbb{R}^{p-1}} \left(\left\| \widehat{\boldsymbol{\varepsilon}}_j - \widehat{\mathbf{E}}_{-j}\boldsymbol{\gamma} \right\|_2^2 / T + 2\lambda_j \|\boldsymbol{\gamma}\|_1 \right), \tag{28}$$

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

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

$$\hat{\tau}_j^2 = \left\| \widehat{\boldsymbol{\varepsilon}}_j - \widehat{\mathbf{E}}_{-j}\widehat{\boldsymbol{\gamma}}_j \right\|_2^2 / T + \lambda_j \left\| \widehat{\boldsymbol{\gamma}}_j \right\|_1 \tag{29}$$

Go back

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

- 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

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

- The approximate inverse is defined as

$$\widehat{\boldsymbol{\Theta}} = \widehat{\mathbf{T}}^{-2}\widehat{\mathbf{C}}. \tag{30}$$

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 $\boldsymbol{\Lambda}_{\varepsilon,m} \equiv \min\{\widehat{\boldsymbol{\Lambda}}_{\varepsilon,i} | \widehat{\boldsymbol{\Lambda}}_{\varepsilon,i} > 0\}$. 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

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 # 4: 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
PC	0.0091	0.0300	0.3117	1.2113	0.0290	0.1005	0.2882	2.1756	0.0287	0.1049	0.2743	3.7190
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, $(\mu, \sigma) = (0.0080, 0.05)$.

Finding # 5: 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 (PC)	0.0006	0.0052	0.1122
Debiased Lasso (PC)	0.0067	0.0265	0.2542
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, $(\mu, \sigma) = (0.0080, 0.05)$.

7. Asymptotic Theory

Assumptions

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

(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 $k \to \infty$:

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

Assumptions

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

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

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

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

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

Notation

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

$$\widehat{\mathbf{y}} = \mathbf{R}\mathbf{w} + \mathbf{e} \tag{31}$$

- Denote $S_0 \equiv \{j : \mathbf{w}_j \neq 0\}$ to be the active set of variables, where $\mathbf{w}$ is a vector of true portfolio weights in equation (31). Also, let $s_0 \equiv |S_0|$.

$$\widehat{\gamma}_j = \arg\min_{\boldsymbol{\gamma}\in\mathbb{R}^{p-1}} \left(\left\| \widehat{\varepsilon}_j - \widehat{\mathbf{E}}_{-j}\gamma \right\|_2^2 / T + 2\lambda_j \|\gamma\|_1 \right), \tag{32}$$

- Let $S_j \equiv \{k : \gamma_{j,k} \neq 0\}$ be the active set for row $\boldsymbol{\gamma}_j$ for the nodewise regression in (32), and let $s_j \equiv \left|S_j\right|$. Define $\bar{s} \equiv \max_{1\leq j\leq p} s_j$.
- 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 $\omega_T \equiv \sqrt{\log p/T} + 1/\sqrt{p}$. *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 $|\hat{y} - y| \lesssim_P \bar{s}^2\omega_T = o_p(1)$, *where* y *was defined in (11).*

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 $\widehat{\mathbf{y}} = \mathbf{R}\mathbf{w} + \mathbf{e}$ 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 (13) and $\lambda_j \asymp \omega_T$ uniformly in j for the lasso for nodewise regression in (28). Assume $(s_0 \vee \bar{s}^2)\log(p)/\sqrt{T} = o(1)$. Then

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

If $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma_e^2\mathbf{I})$, let $\widehat{\mathbf{\Omega}} \equiv \widehat{\mathbf{\Theta}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{\Theta}}'$. Then $W|\mathbf{R} \sim \mathcal{N}_p(\mathbf{0}, \sigma_e^2\widehat{\mathbf{\Omega}})$ and $\left\|\widehat{\mathbf{\Omega}} - \mathbf{\Theta}\right\|_\infty = o_p(1)$.

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

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