Mean Variance Optimization and Regularization

解读者：屠雪永

2020.10.24

1.Xing, Xin, Jinjin Hu, and Yaning Yang. “Robust Minimum Variance Portfolio with L-Infinity Constraints.” *Journal of Banking & Finance* 46 (September 2014): 107–17.

2.Kremer, Philipp J., Sangkyun Lee,et al. “Sparse Portfolio Selection via the Sorted ℓ1-Norm.” *Journal of Banking & Finance* 110 (January 2020): 105687.

3.Goto, Shingo, and Yan Xu. “Improving Mean Variance Optimization through Sparse Hedging Restrictions.” *Journal of Financial and Quantitative Analysis* 50, no. 6 (December 2015): 1415–41.

Outline

- Mean Variance Optimization
- Regularization
- Robust minimum variance portfolio with L-infinity constraints
  - Motivation
  - Research design
  - Empirical result
  - Conclusion
- Sparse portfolio selection via the sorted $\ell 1$-Norm
- Improving Mean Variance Optimization through Sparse Hedging Restrictions

Mean Variance Optimization

A Markowitz (1952) mean-variance optimal stock portfolio:

$$\max_{w} w^T \mu, \tag{5}$$

$$\text{s.t. } w^T \Sigma w = \Sigma_p, \tag{6}$$

$$w^T e = 1, \tag{7}$$

$$w = \frac{\Sigma^{-1}\mu}{e^T \Sigma^{-1} \mu}. \tag{8}$$

Optimization object: μ Σ Σ^{-1} μ and Σ

Finite sample ?

Regularization

- By regularization, we make models simpler and get more accurate prediction with **finite samples.**

$$L(W) = \frac{1}{N}\sum_{i=1}^{N} L_{\mathrm{i}}(f(x_i, W), y_i) + \lambda R(W)$$

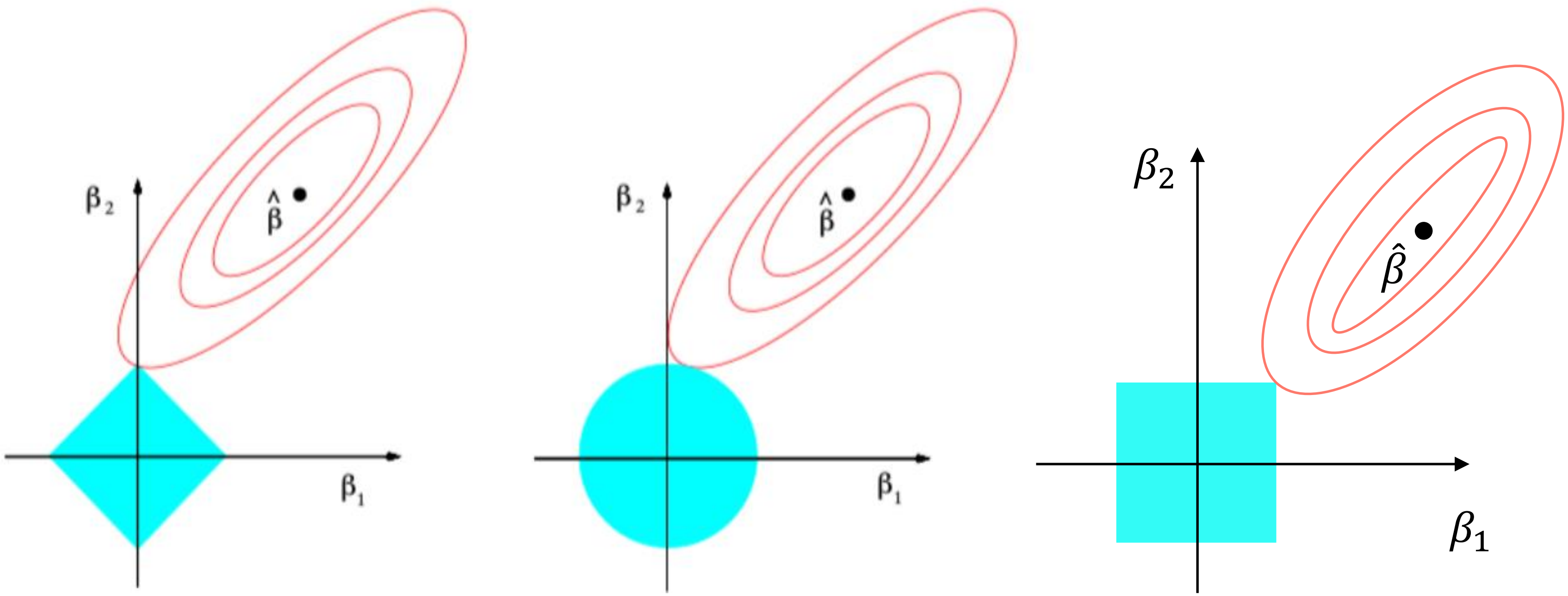
Data loss: Model predictions should match training data

Regularization: Model should be "simple", so it works on test data.

Regularization

- Linear Model:
  - $Y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \epsilon$
- OLS:
  - $RSS = \sum(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2$
- LASSO
  - $\sum(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} |\beta_j| = RSS + \lambda \sum_{j=1}^{p} |\beta_j|$
  - The error should be small, and the absolute value of beta should not be too large
- Ridge Regression
  - $\sum(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} \beta_j^2 = \mathrm{RSS} + \lambda \sum_{j=1}^{p} \beta_j^2$
  - The error should be small, and the square of beta should not be too large
- l_∞ Norm
  - $\sum(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} \beta_j^2 = \mathrm{RSS} + \lambda max_{1 \le \mathrm{i} \le \mathrm{N}} \{|\beta_i|\}$

L1 vs L2 regularization



L1 L2 $L\infty$

1.Introduction--Motivation

The l_1 constraint can produce sparse weights ,but tends to **select only one** from a group of highly correlated variables (assets) and forces the others in the group being unselected.

So we propose to add a l_∞norm constraint or to add a pairwise l_∞ norm constraint in the l_1 norm constrained minimum-variance portfolio (MVP) problem:

(1)The l_∞ constraint **controls the largest absolute component** of the weight vector .

(2)The pairwise l_∞ constraint encourages **retaining the cluster structure** of highly correlated assets in MVP optimization.

1. Robust minimum variance portfolio with L-infinity constraints

2. Research design--The l_1 $-l_\infty$ constrained MVP

Let $\boldsymbol{w} = (w_1, \dots, w_N)$ be the vector of portfolio weights, $\hat{\Sigma}$ be the sample covariance matrix which is the estimation of covariance matrix Σ of asset returns .The l_1 norm of $\boldsymbol{w}$ is defined as $||\boldsymbol{w}||_1 = \sum_{i=1}^{N} |w_i|$ and the l_∞norm as $||\boldsymbol{w}||_\infty = max_{1\leq i\leq N}\{|w_i|\}$.Let **1** be the vector of 1's . Our constrained MVP optimization strategy can be formulated as follows.

$$\min_{\mathbf{w}} \mathbf{w}^\top \widehat{\Sigma} \mathbf{w}, \tag{1}$$

$$\text{s.t. } \mathbf{w}^\top \mathbf{1} = 1, \tag{2}$$

$$\|\mathbf{w}\|_1 + \alpha \|\mathbf{w}\|_\infty \leqslant c, \tag{3}$$

$$\min_{\mathbf{w}} \mathbf{w}^\top \widehat{\Sigma} \mathbf{w} + \lambda(\|\mathbf{w}\|_1 + \alpha\|\mathbf{w}\|_\infty), \tag{4}$$

$$\text{s.t. } \mathbf{1}^\top \mathbf{w} = 1, \tag{5}$$

1. Robust minimum variance portfolio with L-infinity constraints

2. Research design--The $l_1 - l_\infty^{(p)}$ constrained MVP

As did in the octagonal shrinkage and clustering algorithm (**OSCAR**) for linear regression analysis

$$\min_{\mathbf{w}} \mathbf{w}^\top \widehat{\Sigma} \mathbf{w}, \tag{6}$$

$$\text{s.t. } \mathbf{1}^\top \mathbf{w} = 1, \tag{7}$$

$$\|\mathbf{w}\|_1 + \alpha \sum \max\{|w_i|, |w_j|\} \leqslant c, \tag{8}$$

$$\min_{\mathbf{w}} \mathbf{w}^\top \widehat{\Sigma} \mathbf{w} + \lambda \left(\|\mathbf{w}\|_1 + \alpha \sum_{i<j} \max\{|w_i|, |w_j|\} \right), \tag{9}$$

$$\text{s.t. } \mathbf{1}^\top \mathbf{w} = 1, \tag{10}$$

$$\sum_{i<j} \max\{|w_i|, |w_j|\} = \sum_{i=1}^{N} (i-1)|w|_{(i)}$$

Where $|\mathrm{w}|_{(i)}$ is the ith smallest component of $|\boldsymbol{w}|$

The constraint region is also a polygon, but has more vertices than the region of $l_1 - l_\infty$ constraint. Thus the region of OSCAR constraint is **more likely to assign the same weight to those highly correlated assets** and the region is more smaller than the region of $l_1 - l_\infty$ constraint.

1. Robust minimum variance portfolio with L-infinity constraints

2. Empirical result-Out-of-sample evaluation

$$\widehat{\mu} = \frac{1}{T-\tau}\sum_{t=\tau}^{T-1}\widehat{\mathbf{w}}_t^\top \mathbf{r}_{t+1},$$

$$\widehat{\sigma}^2 = \frac{1}{T-\tau-1}\sum_{t=\tau}^{T-1}(\widehat{\mathbf{w}}_t^\top \mathbf{r}_{t+1} - \widehat{\mu})^2.$$

Then, the out-of-sample Sharpe ratio is defined as

$$\widehat{\mathrm{SR}} = \frac{\widehat{\mu}}{\widehat{\sigma}},$$

and the out-of-sample turnover value is defined as

$$\widehat{\mathrm{TO}} = \frac{1}{T-\tau-1}\sum_{t=\tau}^{T-1}\|\widehat{\mathbf{w}}_{t+1} - \widehat{\mathbf{w}}_{t^+}\|_1.$$

3. Empirical result-Out-of-sample evaluation

Table 2
Out-of-sample performance measures for real data sets.

Source	Sharpe ratio				Variance ($\times 10^{-3}$)				Turnover			
	10Ind	30Ind	100FF	CRSP	10Ind	30Ind	100FF	CRSP	10Ind	30Ind	100FF	CRSP
l_1-l_∞	0.328	0.305	0.513	0.174	1.198	1.239	1.356	1.201	0.267	0.243	0.698	0.068
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[–]	[–]	[–]	[–]
$l_1-l_\infty^{(p)}$	0.311	0.294	0.473	0.158	1.222	1.214	1.358	1.921	0.072	0.176	0.243	0.061
	[0.365]	[0.614]	[0.178]	[0.396]	[0.467]	[0.415]	[0.800]	[0.031]	[–]	[–]	[–]	[–]
DS	0.315	0.301	0.501	0.158	1.239	1.206	1.275	1.400	0.109	0.133	0.476	0.156
	[0.590]	[0.655]	[0.173]	[0.484]	[0.829]	[0.539]	[0.095]	[0.124]	[–]	[–]	[–]	[–]
MINU	0.314	0.260	0.136	0.081	1.274	1.398	7.226	1.110	0.137	0.416	7.653	5.380
	[0.075]	[0.231]	[0.001]	[0.143]	[0.046]	[0.146]	[0.003]	[0.001]	[–]	[–]	[–]	[–]
MINC	0.294	0.269	0.314	0.164	1.296	1.319	1.749	1.223	0.046	0.064	0.226	0.132
	[0.620]	[0.035]	[0.001]	[0.731]	[0.294]	[0.342]	[0.001]	[0.130]	[–]	[–]	[–]	[–]
l_1	0.316	0.289	0.509	0.164	1.252	1.222	1.302	1.214	0.113	0.188	0.374	0.133
	[0.638]	[0.371]	[0.158]	[0.731]	[0.853]	[0.852]	[0.535]	[0.130]	[–]	[–]	[–]	[–]
l_2	0.322	0.306	0.499	0.158	1.203	1.210	1.289	1.122	0.079	0.147	0.496	0.155
	[0.300]	[0.791]	[0.174]	[0.531]	[0.054]	[0.464]	[0.214]	[0.001]	[–]	[–]	[–]	[–]
l_1-l_2	0.322	0.303	0.502	0.168	1.206	1.198	1.272	1.324	0.082	0.141	0.442	0.052
	[0.849]	[0.842]	[0.300]	[0.329]	[0.053]	[0.062]	[0.016]	[0.028]	[–]	[–]	[–]	[–]
1/N	0.256	0.228	0.256	0.155	1.816	2.298	2.357	2.314	0.023	0.028	0.023	0.006
	[0.048]	[0.084]	[0.001]	[0.539]	[0.003]	[0.001]	[0.001]	[0.001]	[–]	[–]	[–]	[–]

- The $l_1 - l_\infty$ strategy has **the largest out-of-sample Sharpe ratio** across all of the four data sets while the $l_1 - l_\infty^{(p)}$ strategy **have relative lower Sharpe ratios**.
- The $l_1 - l_\infty$ strategy performs the best among all the methods for the 10Ind data.
- In addition, the the $l_1 - l_\infty^{(p)}$ method generally **has smaller turnover values** while the $l_1 - l_\infty$ strategy has larger turnover values than other norm constrained strategies.

3. Empirical result-Out-of-sample evaluation

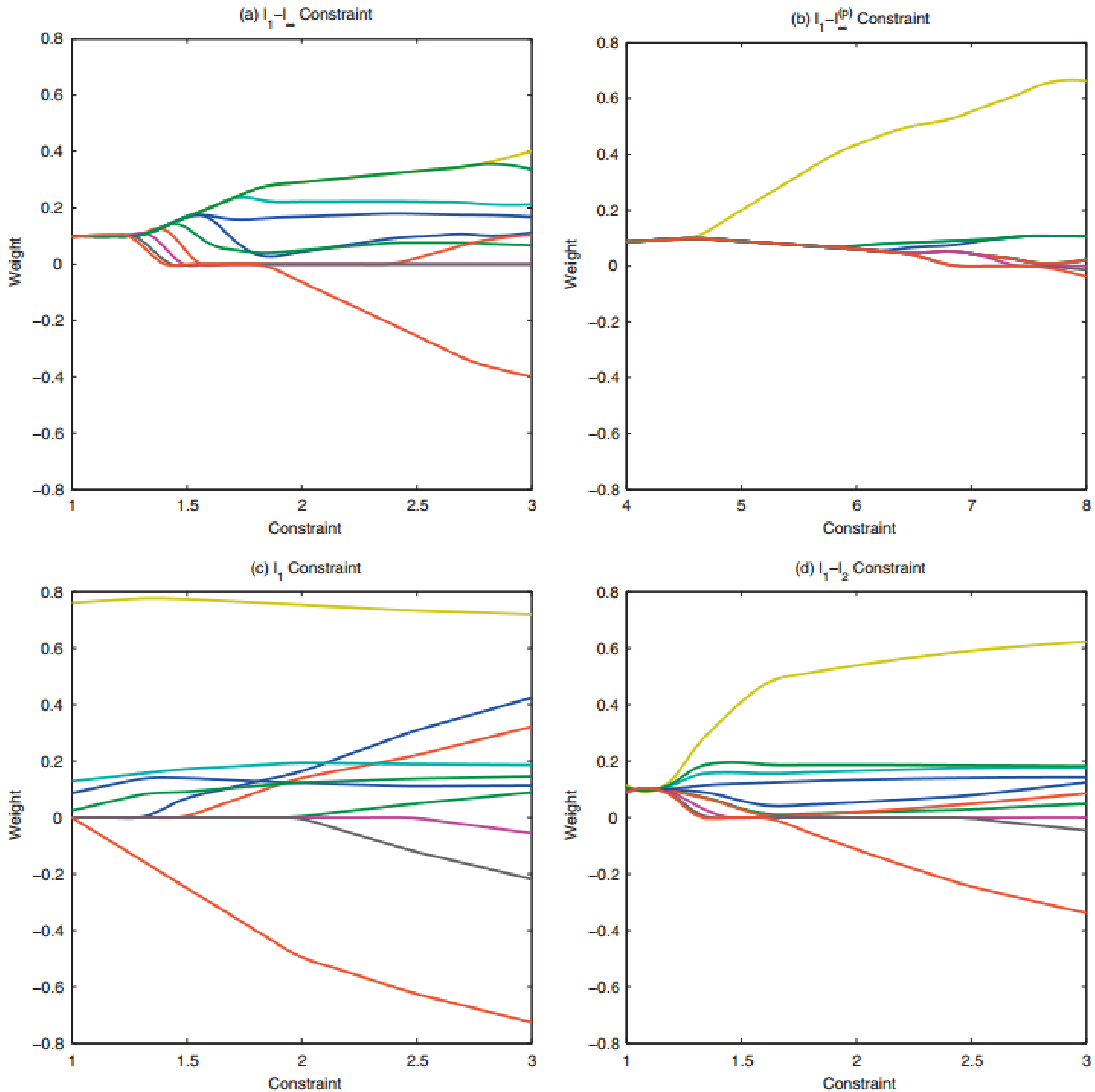


Fig. 2. Solution paths of several norm constrained portfolios.

- The $l_1 - l_\infty$ have a **good control of large absolute solutions.**
- Several lines coincide for the $l_1 - l_\infty^{(p)}$ strategy, implying that some of the estimated weights are equal and therefore the corresponding **assets are clustered**.
- In addition, the $l_1 - l_\infty^{(p)}$ strategy has similar maximum weights compared with the other constraint strategies but has much smaller negative weights.

4. Conclusion

- Firstly, the proposed methods can **produce more stable and robust portfolio solutions** in the sense of having better out-of-sample performances compared with existing approaches.
- Secondly, the $l_1 - l_\infty$ constrained MVP has **the largest Sharpe ratios** and the $l_1 - l_\infty^{(p)}$ method has slightly smaller Sharpe ratios but **has smaller turnover values**.

1.Introduction--Motivation

An ideal portfolio has:

a) conservative asset weights, which are **stable in time**, to avoid high turnover and transaction costs.

b) still promotes **the right amount of diversification**, while being able to control the total amount of shorting.

- **GMV** :the sample covariance matrix might exhibit **estimation error**.Multicollinearity and extreme observations, especially for a large number of assets
- **LASSO**: reduced recovery of sparse signals when applied to highly dependent data.**it is ineffective in the presence of no short selling** ($\text{i.e.}\, w_i \geq 0$) and an imposed budget constraint (i.e., $\sum_{i=1}^{k} w_i = 1$)
- **SLOPE** (the Sorted 1 Penalized Estimator): SLOPE continues to **shrink the active weights, even when short sales are restricted** (i.e. $w_i \geq 0, \forall\, i = 1, \ldots, k$).

2. Research design——SLOPE

Given k jointly normally distributed asset returns $R_1, \dots R_k$,with expected value vector $\mu = [\mu_1, \dots, \mu_k]$ and covariance matrix Σ,

$$\min_{\boldsymbol{w}\in\mathbb{R}^k} \frac{\phi}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} - \boldsymbol{\mu}'\boldsymbol{w}, \text{ subject to } \sum_{i=1}^{k} w_i = 1 \quad (1)$$

$$\min_{\boldsymbol{w}\in\mathbb{R}^k} \frac{\phi}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} - \boldsymbol{\mu}'\boldsymbol{w} + \rho_\lambda(\boldsymbol{w}) \text{ s.t. } \sum_{i=1}^{k} w_i = 1 \quad (2)$$

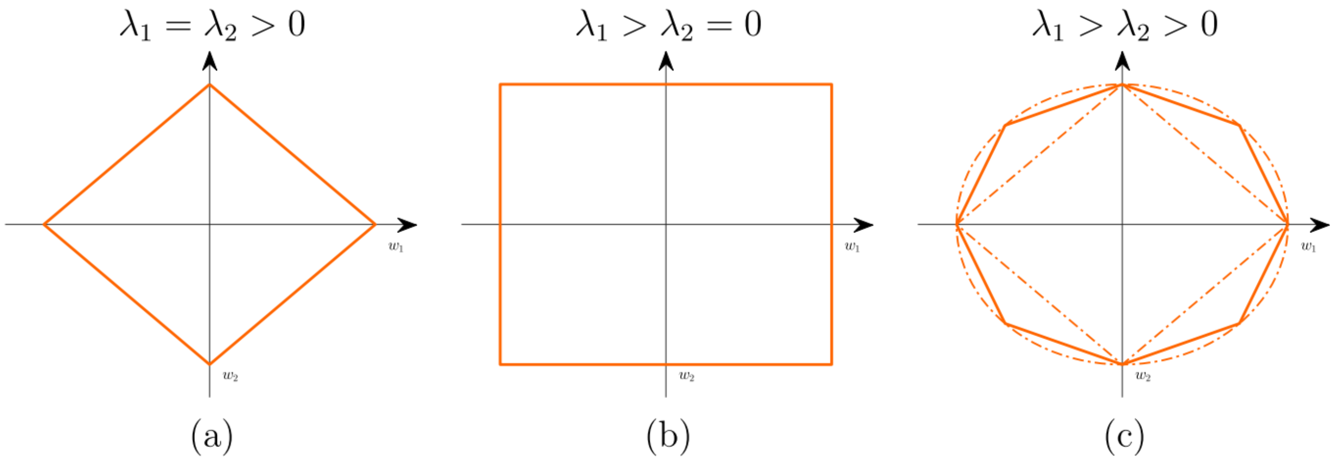
$$\rho_\lambda(\boldsymbol{w}) := \sum_{i=1}^{k} \lambda_i |w|_{(i)} = \lambda_1|w|_{(1)} + \lambda_2|w|_{(2)} + \ldots + \lambda_k|w|_{(k)}$$

$$\text{s.t. } \lambda_1 \ge \lambda_2 \ge \ldots \lambda_k \ge \mathbf{0} \text{ and } |w|_{(1)} \ge |w|_{(2)} \ge \ldots |w|_{(k)}, \quad (3)$$

If $\rho_\lambda(\boldsymbol{w}) = \lambda \times \sum_{i=1}^{k} |w_i|$,it will be the LASSO

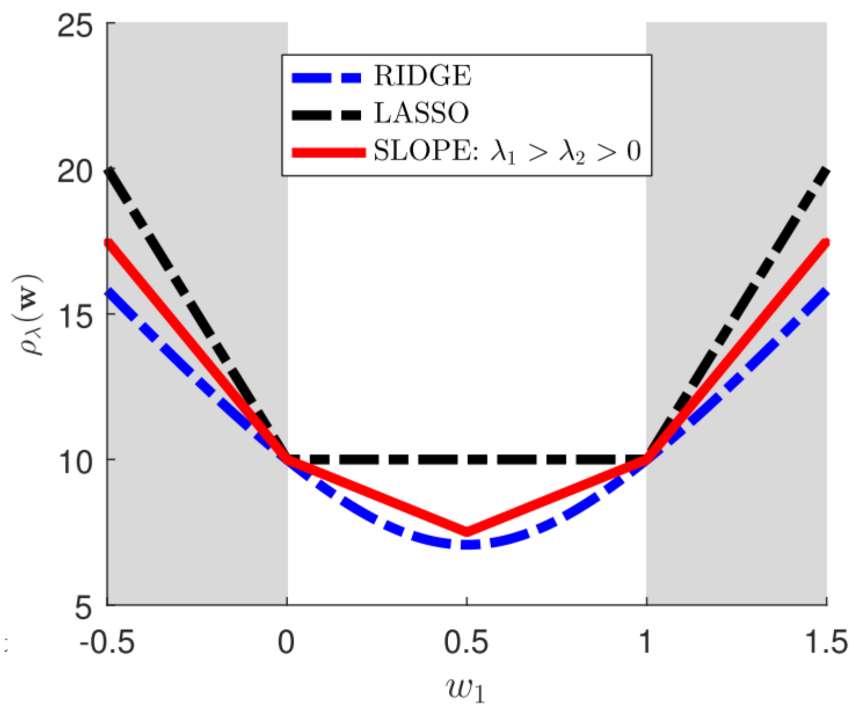
$\varphi > 0$ is the coefficient of relative risk aversion

2. Research design



(a)the SLOPE sphere coincides with the well studied diamond shape of the **LASSO** penalty.
(b) the ℓ_∞**-Norm**
(c) combining the properties of the Lasso and the ℓ_∞ penalties, even able to reach one of RIDGE's special solutions

to control for monitoring and transaction costs of financial assets, we **prefer SLOPE over the RIDGE estimator**, because **it can promote sparsity** by exploiting the singularities.



2. Research design

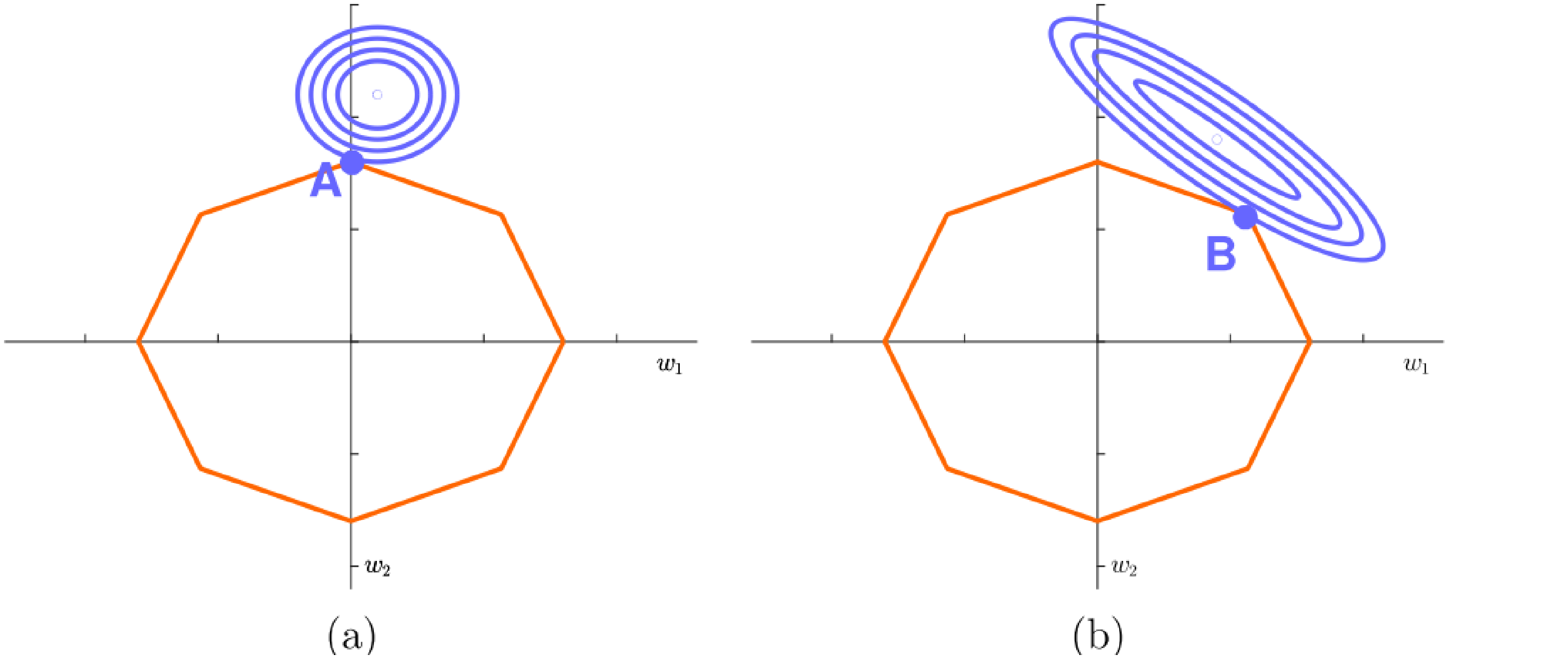
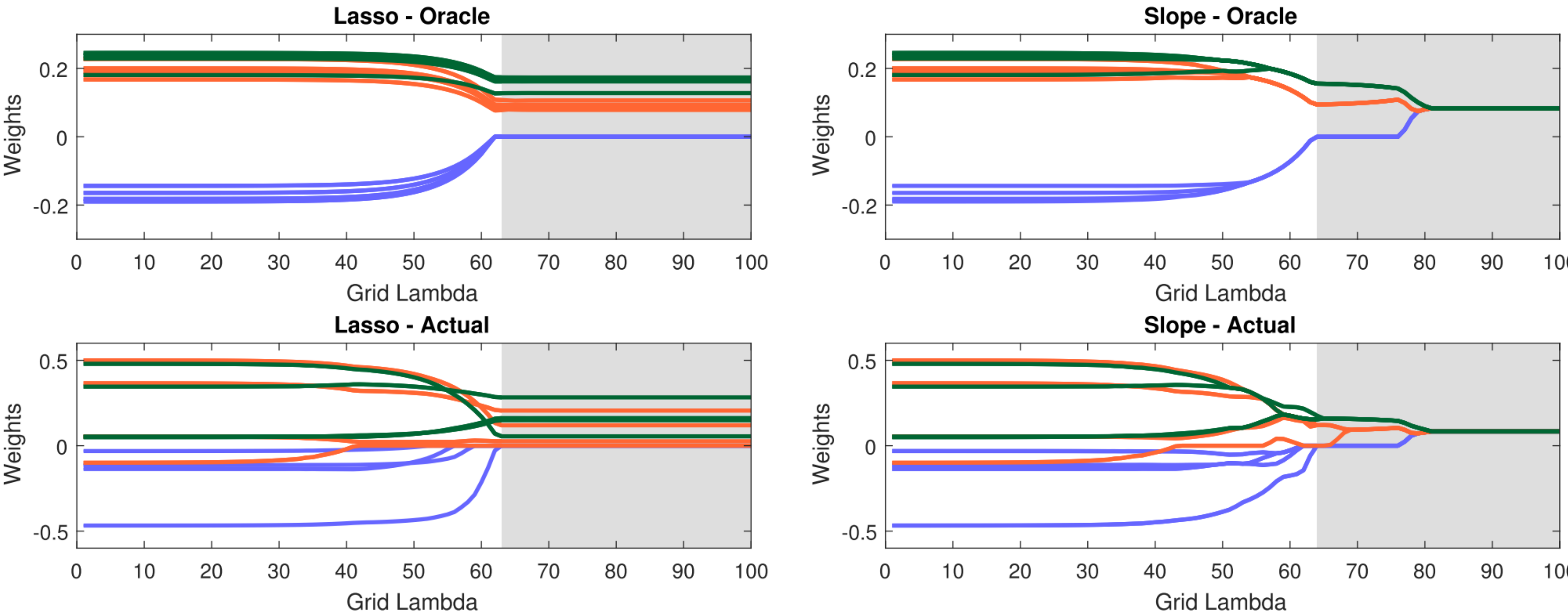


Fig. 3. Sorted 1-Norm Penalty without Budget Constraint. The figure plots in Panel (a) and (b), considering orthogonal design and correlated design, respectively.

- promote sparsity , automatically group assets with similar correlation.

3. Empirical result



(a) (b)

PS: Equally colored weights characterize assets with the same underlying factor exposure.

- The LASSO shrinks the weights up until the no short sale area.
- The octagonal shape of the penalty pushes the solution towards the equally weighted portfolio.

2.Sparse portfolio selection via the sorted ℓ1-Norm

3. Empirical result

Table 2
Risk- and return measures.

	Vol. (in %)				Sharpe Ratio			
	10Ind	30Ind	100FF	SP500	10Ind	30Ind	100FF	SP500
EW	14.491	16.257	17.509	20.238	0.776**	0.656***	0.677***	0.205
GMV	10.910	9.152	6.058	11.497	1.102*	1.283***	3.124***	0.057**
GMV-LO	11.473	11.214	13.134	10.825	1.012	0.998***	0.954***	0.389
ERC	13.578	15.029	16.906	17.948	0.840*	0.730***	0.714***	0.233
RIDGE	11.907	12.241	13.219	11.393	0.978	0.955	1.069***	0.490
LASSO	11.364	10.781	10.853	9.505	1.028	1.046	1.421***	0.572
SLOPE	11.352	10.822	10.977	9.643	1.024	1.047	1.382	0.534
SLOPE - LO	11.689	11.865	13.709	11.981	0.950**	0.946**	0.917***	0.344
SLOPE - EW	12.539	12.904	14.390	11.017	0.908**	0.929*	0.858***	0.645

	AP				Turnover			
	10Ind	30Ind	100FF	SP500	10Ind	30Ind	100FF	SP500
EW	10.000	30.000	100.000	443.000	0.049	0.057	0.056	0.077
GMV	9.982	29.885	99.469	434.377	0.125	0.273	0.852	2.748
GMV-LO	5.371	8.562	9.220	27.553	0.064	0.074	0.101	0.238
ERC	10.000	30.000	100.000	443.000	0.048	0.054	0.055	0.076
RIDGE	9.989	29.824	98.171	408.474	0.061	0.078	0.109	0.212
LASSO	6.755	12.301	18.371	130.211	0.079	0.104	0.184	0.434
SLOPE	7.027	13.231	22.598	145.552	0.078	0.101	0.172	0.409
SLOPE - LO	7.299	18.465	34.616	129.632	0.119	0.217	0.409	0.590
SLOPE - EW	6.330	6.128	17.022	76.386	0.052	0.055	0.057	0.284

- **SLOPE yields consistently lower variance ;**
- SLOPE has a good performance of the **out-of-sample SR**.
- SLOPE is able to promote **sparse solutions** and to **reduce the overall portfolio turnover.**

3. Empirical result

$$\widehat{\mathrm{WDiv}} = \frac{1}{k \times \sum_{i=1}^{k} \hat{w}_i^2}$$

Table 3
Diversification measures.

	DR				WDiv				RDiv			
	10Ind	30Ind	100FF	SP500	10Ind	30Ind	100FF	SP500	10Ind	30Ind	100FF	SP500
EW	1.270	1.343	1.212	1.675	1.000	1.000	1.000	1.000	0.933	0.935	0.958	0.894
GMV	1.255	1.362	0.958	3.147	0.197	0.078	0.013	0.012	0.197	0.078	0.013	0.012
GMV-LO	1.289	1.414	1.299	1.944	0.320	0.150	0.062	0.032	0.320	0.150	0.062	0.032
ERC	1.300	1.382	1.225	1.728	0.935	0.914	0.963	0.880	1.000	1.000	1.000	1.000
RIDGE	1.330	1.457	1.256	1.920	0.540	0.430	0.262	0.248	0.577	0.440	0.188	0.120
LASSO	1.289	1.415	1.237	2.221	0.309	0.143	0.044	0.062	0.301	0.132	0.030	0.029
SLOPE	1.295	1.426	1.247	2.213	0.319	0.155	0.054	0.071	0.312	0.144	0.056	0.032
SLOPE - LO	1.315	1.457	1.295	1.936	0.417	0.287	0.209	0.206	0.437	0.319	0.221	0.219
SLOPE - EW	1.289	1.314	1.294	1.808	0.403	0.181	0.118	0.170	0.408	0.182	0.120	0.166

- SLOPE-LO and SLOPE consistently outperform the LASSO across all datasets for the WDiv and the RDiv. Except for the SP500, this is also true for the DR.

4. Conclusion

- Firstly, SLOPE has the advantage of still **being active in the no short sales area** and given an imposed budget constraint.
- Furthermore, SLOPE can identify assets with **the same underlying risk factor exposure** and group them together.
- SLOPE can create sparse portfolios with **a reduced turnover rate, improve risk- and weight diversification**, and have a high degree of flexibility in the portfolio construction process.

1.Introduction--Motivation

To minimize portfolio risk, we propose a “sparse” estimator of **the inverse covariance matrix** to reduce estimation errors:

- In practice with **finite samples**, however, multicollinearity makes the hedge trades too unstable and unreliable.
- The inverse covariance matrix prescribes the hedge trades in which a stock is hedged by all the other stocks in the portfolio. (Stevens,1998)
- **By shrinking trade sizes and reducing the number of stocks in each hedge trade**, we propose a “sparse” estimator of the inverse covariance matrix.

2. Research design

Stevens (1998) shows that the inverse covariance matrix Σ^{-1} reveals the optimal hedging relations among stocks.

$$r_{i,t} = \alpha_i + \sum_{k=1,k\neq i}^{N} \beta_{i|k} r_{k,t} + \varepsilon_{i,t}, \tag{1}$$

$$\psi_{ij} = \begin{cases} -\dfrac{\beta_{i|j}}{\upsilon_i} & \text{if } i \neq j \\ \dfrac{1}{\upsilon_i} & \text{if } i = j \end{cases}. \tag{2}$$

$$\Psi(i,\cdot) = \frac{1}{\upsilon_i}\left[-\beta_{i|1}, \ldots, -\beta_{i|i-1}, 1, -\beta_{i|i+1}, \ldots, -\beta_{i|N}\right]. \tag{3}$$

where $r_{i,t}$ denotes the ith stock return in period t; $\varepsilon_{i,t}$ is the unhedgeable component of $r_{i,t}$,whose variance is denoted by $\upsilon_i = var(\varepsilon_{i,t})$; and υ_i is a measure of the ith stock's unhedgeable risk.

The inverse of the $N \times N$ covariance matrix Σ by Σ^{-1} =Ψ =[ψ_{ij}], where ψ_{ij} denotes the (i,j)th element of Ψ.

2. Research design

We conquer the multicollinearity by penalizing the L1 norm of the parameters that need to be estimated:

$$(4) \qquad \hat{\beta}_{i|k}^{\text{lasso}} = \arg\min_{\beta} \left\{ \sum_{t=1}^{T} \left(r_{i,t} - \sum_{k=1,k\neq i}^{N} \beta_{i|k} r_{k,t} \right)^2 + \gamma \sum_{k=1,k\neq i}^{N} \left| \beta_{i|k} \right| \right\}.$$

The row-by-row (or column-by-column) lasso estimation does **not restrict the inverse covariance matrix to be positive definite or symmetric**.

To avoid this problem, the N hedge regressions should be estimated jointly as a group, rather than separately. By quasi-maximum likelihood (QML) once for all

$$(6) \qquad \max_{\Psi=[\psi_{ij}]} \frac{T}{2} \ln(\det(\Psi)) - \frac{T}{2} \operatorname{trace}(\hat{S}\Psi) - \rho \sum_{i=1,i\neq j}^{N} \sum_{j=1,j\neq i}^{N} |\psi_{ij}|,$$

$\rho \geq 0$, $\hat{S}$ is the sample covariance matrix.

QML problem (6) is equivalent to a *N*-coupled lasso problem. We solve QML estimation problem (6) using their graphical lasso (glasso) algorithm.

3. Empirical result-Data

Data Set	Descriptor 1	Description 2	Market 3	N 4	N/T 5	Time Period 6
1	SZBM100	100 (10 × 10) portfolios formed on size and BM	U.S.	100	0.833	July 1963–Dec. 2010
2	IND48	48 industry portfolios	U.S.	48	0.400	July 1963–Dec. 2010
3	SZBM100 + IND48	Combination of SZBM100 and IND48	U.S.	148	1.233	July 1963–Dec. 2010
4	NON_US	Combination of INT_MKT and INT_VAL_GRO	Int'l	133	1.108	Jan. 1975–Dec. 2010
5	Individuals	100 stocks from NYSE and AMEX	U.S.	100	0.833	Jan. 1973–Dec. 2010

		Out-of-Sample Analysis Period			Proposed Optimizer: $\hat{\Psi}_\rho$	
Data Set	Descriptor	Training Period 1	Testing Period 2	T_f 3	ρ 4	Sparsity 5
1	SZBM100	July 1973–June 1983	July 1983–Dec. 2010	330	1.7	45.0%
2	IND48	July 1973–June 1983	July 1983–Dec. 2010	330	1.3	32.2%
3	SZBM100 + IND48	July 1973–June 1983	July 1983–Dec. 2010	330	1.9	47.1%
4	NON_US	Jan. 1985–Dec. 1994	Jan. 1995–Dec. 2010	192	0.8	44.0%
5	Individuals	Jan. 1983–Dec. 1992	Jan. 1993–Dec. 2010	216	5.9	32.4%

PS:Degree of sparsity is the average percentage of 0 off-diagonal elements in the estimated inverse covariance matrix $\hat{\Psi}_\rho$.

3. Empirical result-Out-of-sample Portfolio Risk Minimization

TABLE 4 Out-of-Sample Portfolio Risk (monthly)

Data Set	Descriptor	*Panel A. Return Variance* ($\%^2$)						*Panel B.* $\sigma_{ALT} - \sigma$ *for* σ_{ALT} $\hat{\Psi}_\rho$ (%)				
		$\sigma^2_{\hat{\Psi}_\rho}$	$\sigma^2_{\hat{S}-1}$	σ^2_{EW}	σ^2_{JM}	σ^2_{LW}	σ^2_{IND}	$\sigma_{\hat{S}-1}$	σ_{EW}	σ_{JM}	σ_{LW}	σ_{IND}
1	SZBM100	13.30 (1)	65.30 (5)	25.43 (3)	58.57 (4)	21.98 (2)	NA	4.43***	1.40***	4.01*	1.04***	NA
2	IND48	12.45 (1)	17.59 (4)	22.88 (5)	16.22 (3)	13.15 (2)	NA	0.66***	1.25***	0.50*	0.10	NA
3	SZBM100 + IND48	10.70 (1)	$T < N$ (5)	24.12 (4)	16.43 (3)	12.74 (2)	NA	$T < N$	1.64***	0.78***	0.30	NA
4	NON_US	15.25 (1)	$T < N$ (5)	29.56 (4)	23.64 (3)	15.92 (2)	NA	$T < N$	1.53***	0.96*	0.08	NA
5	Individuals	10.10 (1)	19.44 (5)	23.79 (6)	13.40 (4)	11.36 (3)	11.01 (2)	1.21***	1.72***	0.45***	0.19***	0.14***

PS:Numbers below the variances (in parentheses) are the ranking of portfolio variance among the six portfolios .

- Our proposed portfolio GMV- $\widehat{\Psi}_\rho$ generates **lower out-of-sample portfolio variance** than alternative portfolios in all data sets .

3. Empirical result-Out-of-sample Sharpe Ratio

TABLE 5 Out-of-Sample Sharpe Ratio (monthly)

Data Set	Descriptor	*Panel A. Monthly SRs* $\hat{\Psi}_\rho$	$\hat{S}^{-1}$	$1/N$	JM	LW	IND	*Panel B. Differences in SRs between GMV-$\hat{\Psi}_\rho$ and the Other Five Portfolios* $\hat{S}^{-1}$	$1/N$	JM	LW	IND
1	SZBM100	$\boxed{0.260}$ (1)	0.112 (4)	0.113 (3)	0.065 (5)	0.215 (2)	NA	0.148**	0.147***	0.195***	0.045	NA
2	IND48	0.126 (3)	0.070 (5)	0.129 (2)	0.151 (1)	0.119	NA	0.056**	−0.003	−0.025	0.007	NA
3	SZBM100 + IND48	0.267 (2)	$T < N$ (5)	0.119 (4)	0.123 (3)	0.315	NA	$T < N$	0.148***	0.144***	−0.048	NA
4	NON_US	$\boxed{0.273}$ (1)	$T < N$ (5)	0.138 (4)	0.142 (3)	0.227	NA	$T < N$	0.136**	0.132**	0.046**	NA
5	Individuals	0.147 (2)	0.100 (6)	0.181 (1)	0.101 (5)	0.105 (4)	0.138 (3)	0.047***	−0.034***	0.046***	0.042***	0.010**

PS:Numbers below the variances (in parentheses) are the ranking of portfolio variance among the six portfolios .

- GMV- $\hat{\Psi}_\rho$ generates the highest or second-highest **out-of-sample Sharpe ratios**, except for the 48-industry portfolio.

4. Conclusion

- By mitigating estimation errors in the hedge portfolios, we can enhance the ability of the mean variance optimizer in **reducing the out-of-sample portfolio risk**.

5. Consideration

- With limited samples, it is efficient to utilize L1 Norm to estimation many parameters.
- Combine different methods to deal with different scenarios.