Journal of Banking & Finance 46 (2014) 107–117

Contents lists available at ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

Robust minimum variance portfolio with L-infinity constraints

Xin Xing [a], Jinjin Hu [b,*], Yaning Yang [a]

[a] Department of Statistics and Finance, University of Science and Technology of China, China
[b] Department of Statistics, School of Management, Fudan University, China

ARTICLE INFO

Article history:
Received 16 March 2013
Accepted 6 May 2014
Available online 15 May 2014

JEL classification:
G11

Keywords:
Minimum variance portfolio
Norm constraints
Sparsity
Clustering
Robust

ABSTRACT

Portfolios selected based on the sample covariance estimates may not be stable or robust, particularly so in situations with a large number of assets. The l_1 or l_2 norm constrained portfolio optimization method has been used as a robust method to control the sparsity or to shrink the estimated weights of assets. In this paper, we propose to add an additional l_∞ norm constraint or to add a pairwise l_∞ norm constraint in the l_1 norm constrained minimum-variance portfolio (MVP) problem. The l_∞ constraint controls the largest absolute component of the weight vector and the pairwise l_∞ constraint encourages retaining the cluster structure of highly correlated assets in MVP optimization. By simulation study and analysis of empirical data, we find that the proposed portfolios often have better out-of-sample performance in terms of Sharpe ratios, variances and turn-overs than existing popular portfolio strategies including the l_1 norm constrained MVP, l_2 norm constrained MVP and the $1/N$ portfolio. In addition, we provide moment shrinkage interpretations of the new strategies and an upper bound of errors in the approximation of the empirical optimal portfolio risk based on the theoretical optimal portfolio risk.

1. Introduction

Markowitz (1952) laid on the theoretical foundation for the optimal portfolio choice problem. To apply the Markowitz theory in practice, one needs to estimate the mean and covariances of the asset returns and it is reported that the Markowitz portfolio is very sensitive to the estimation error of the mean and covariance matrix (DeMiguel et al., 2009a,b; Brandt, 2009; Green and Hollofield, 1992). In other words, the weights of the Markowitz portfolio may have large errors and hence unstable if the estimated means and covariances are inaccurate, and consequently the selected portfolios have poor out-of-sample performance. Merton (1980) pointed out that it is more difficult to estimate means than estimating covariances of asset returns in the mean–variance portfolio problem. Jagannathan and Ma (2003) reported that the portfolios are often efficient if one minimizes portfolio variance by ignoring the constraint on means of the assets. We therefore focus only on the problem of minimum variance portfolio (MVP) in this paper as in Jagannathan and Ma (2003) and DeMiguel et al. (2009a). But our method can be easily extended to the mean–variance portfolios.

The error of sample covariance matrix may have large effect on the MVP solution. This problem gets more pronounced when the size of the portfolio is big. Fan et al. (2008) studied the relationship between the error and the size of portfolio. As exemplified in Fan et al. (2008), for N assets, there are $N(N+1)/2$ parameters in the covariance matrix that need to be estimated. Even if the estimation error of each entry of the covariance matrix is at the order of $O(n^{-0.5})$ which is 0.05 if sample size $n = 400$, the aggregated error over the $N(N+1)/2$ estimates could be very large if N is relatively large. Therefore, robust portfolio selection methods that are less sensitive to the estimation error of the covariance matrix is plausible. For the low frequency data such as monthly return data, two types of robust portfolio selection methods have been received substantial investigations in the literature. The first is to shrink the covariance matrix estimator by assuming structural models on the covariance matrix such as the factor model or Bayesian model (Ledoit and Wolf, 2003; Ledoit and Wolf, 2004). The second is to impose norm constraints, such as the l_1 or l_2 norm constraint, on the portfolio weights. Ledoit and Wolf (2003) proposed the l_2 constrained method in optimizing Markowitz portfolio. Ledoit and Wolf (2003), Brodie et al. (2009), and Fan et al. (2008) investigated the l_1 constrained method. DeMiguel et al. (2009a) provided a covariance shrinkage interpretation of the constrained portfolio.

The l_2 constrained portfolio is closely related to the ridge regression method in statistical literature which shrinks the parameter estimates towards zeroes. The l_1 constrained portfolio is related to the least absolute shrinkage selection operator (LASSO) method (Tibshirani, 1996) for variable selection in linear regression model

* Corresponding author. Tel.: +86 2125011230.
E-mail address: hujj@fudan.edu.cn (J. Hu).

http://dx.doi.org/10.1016/j.jbankfin.2014.05.004
0378-4266/

and it can strictly shrink small parameters to zeroes and therefore the corresponding variables would not be selected. In the framework of constrained MVP problem, the l_1 constraint can effectively produce sparse weights with some assets being assigned precisely zero weights. However, this method tends to select only one from a group of highly correlated variables (assets) and forces the others in the group being unselected. The elastic net method (Zou and Hastie, 2005) combines these two approaches by using the linear combination of the l_1 and l_2 norm constraints which tends to select the highly correlated variables together as a group. Yen (2010) studied the elastic net method (l_1 and l_2 constrained MVP) and proposed coordinate-wise descent algorithms for solving the constrained optimization problem. Bondell and Reich (2008) proposed the octagonal shrinkage and clustering algorithm for regression (OSCAR) to simultaneously select variables and cluster variables in the framework of linear regression by imposing pairwise l_∞ norm constraint in addition to the l_1 norm constraint. In this paper, we propose to impose additional l_∞-related norm constraints besides the l_1 constraint in the MVP problem. Specifically, we propose to impose the combination of l_1 and l_∞ constraint or the combination of l_1 and pairwise l_∞ constraint to the MVP optimization problem. For short, these two strategies will be denoted as l_1–l_∞ strategy and l_1–$l_\infty^{(p)}$ strategy, respectively.

Our first approach, the l_1–l_∞ strategy, is to add l_∞ (maximum) norm constraint in addition to the l_1 constraint in minimizing portfolio variances, aiming at shrinking the absolute values of the largest weights. With the l_1 norm constraint only, one can get an optimal portfolio with sparsity (Brodie et al., 2009). This means that the weights of some assets will be forced to be 0, but weights of other assets will possibly be assigned with large absolute values. For example, an asset tends to receive a negative weight in the global MVP if it has higher variance and higher covariances compared with other assets (Jagannathan and Ma, 2003). Since large absolute weights, especially large negative weights are unfavorable in practice, we therefore add an additional l_∞ norm constraint to the l_1 norm constraint in order to alleviate this problem. Simulations and analysis of empirical data show that this method can provide sparse solution and can properly control the largest absolute value of the weights, and the resultant portfolio has relatively larger out-of-sample Sharpe ratios compared with existing norm-constrained portfolios. We focus on the l_1–l_∞ constraint instead of l_2–l_∞ constraint, because the l_2–l_∞ constraint does not lead to sparse solution with zero weights while l_1–l_∞ does.

Tola et al. (2008) investigated the clustering method in estimating correlation matrix and showed that the resulted MVP is robust to estimating error of the correlation matrix. Our second approach, the l_1–$l_\infty^{(p)}$ strategy, is to add a pairwise l_∞ constraint in addition to the l_1 constraint to retain the clustering structure of assets. This approach is the OSCAR (Bondell and Reich, 2008) counterpart in MVP problem. The pairwise l_∞ or maximum norm forces the weights of highly correlated assets to be equal and thus the assets are clustered. Therefore, the l_1–$l_\infty^{(p)}$ method can produce robust portfolios with both sparsity and clustering being properly incorporated. Results of simulation studies and empirical data analysis confirm these properties. For example, the out-of-sample turnover values of l_1 and pairwise l_∞ constrained portfolio are usually smaller than those of other methods. We also show that imposing l_1–$l_\infty^{(p)}$ constraints can be regarded as a shrinkage of sample covariance matrix in an unconstrained minimization problem.

Candelon et al. (2012) proposed a double shrinkage (DS) method to improve the stability of the MVP. The DS method shrinks the sample covariance matrix by regularizing it before optimization, then the estimators of portfolio weights are shrunken towards the equally-weighted portfolio. Empirical data results illustrate that DS method often decreases the portfolio turnovers which is quite similar to the l_1–$l_\infty^{(p)}$ strategy. On the other hand, in terms of out-of-sample Sharpe ratios and return-loss relative to the $1/N$ strategy measure, results show that l_1–l_∞ outperforms the DS approach in these respects.

The rest of this article is organized as follows. In Section 2, we propose the l_1–l_∞ constrained MVP and the l_1–$l_\infty^{(p)}$ constrained MVP. In Section 3, we prove some properties of the proposed portfolios. In Section 4, we compare the out-of-sample performance of our method to some other existing methods by simulation study and empirical data analysis. Section 5 is the conclusion. Proofs of the propositions and details of the algorithm are available in Appendix A.

2. Methods

2.1. The l_1–l_∞ constrained MVP

In order to construct robust portfolio in the presence of estimation error of sample covariance matrices, our strategy is to impose constraints of l_1 and l_∞ norms of portfolio weights in minimizing the portfolio risk. The effect of adding l_1 constraint is to shrink the estimates of the weights, specifically weights that are less than a pre-specified threshold will be shrunken to zero, resulting in a sparse solution. However, some of the non-zero estimated portfolio weights may still be large even after penalizing on the l_1 constraint. We therefore impose an additional l_∞ constraint to preclude such a situation.

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

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

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

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

where $\alpha \geq 0$ and $c > 0$ are tuning parameters with α controlling the trade-off between the two norms and c being the threshold of the constraint. The constrained optimization problem can be re-formulated as a penalized optimization problem

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

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

where λ is the penalty parameter and there is a one to one correspondence between c and λ.

When $w_i \geqslant 0, i = 1 \ldots N$ and c goes to $\frac{\alpha}{N} + 1$, the solution converges to the $1/N$ portfolios. [1]

The norm constraints restrict the original plane, $\mathbf{1}^\top\mathbf{w} = 1$, to a smaller region by bounding the magnitude of the weights. The geometric interpretation of the constrained region is showed in Fig. 1 for the two-dimension situation. In the left panel, the polygon corresponds to the l_1–l_∞ constraint in (3) and the line corresponds to the constraint in (2). The intersection of these two regions (the two points on the line) is the constrained region of (2) and (3). The right panel shows the optimization solution with or without the l_1–l_∞ constraint. The elliptic curves are the contour plot of the portfolio risks. Without the l_1–l_∞ constraint, the optimal short-sale constrained portfolio is achieved at the tangent point of the curve with

[1] For convenience, we assume that $w_1 \geqslant w_2 \geqslant \ldots \geqslant w_N \geqslant 0$, if $c = \alpha/N + 1$, from $\|\mathbf{w}\|_1 + \alpha\|\mathbf{w}\|_\infty \leqslant c$, we can get that $\sum_{i=1}^N w_i + \alpha w_1 \leqslant \frac{\alpha}{N} + 1$, so that $w_1 \leqslant 1/N$, $\sum_{i=2}^N w_i \geqslant \frac{N-1}{N}$, while for each $2 \leqslant i \leqslant N, w_i \leqslant w_1 \leqslant 1/N$, so only the equality can satisfy those inequalities, so for each $1 \leqslant i \leqslant N, w_i = 1/N$.

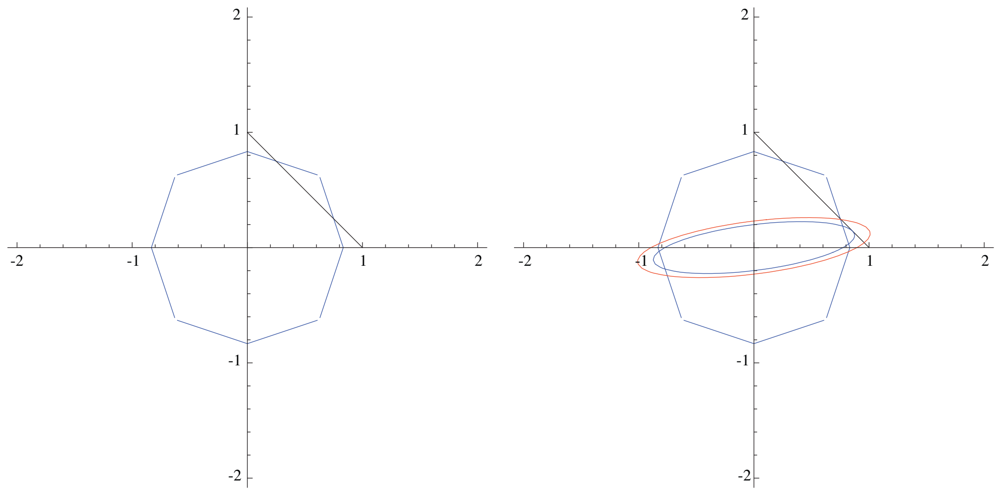

Fig. 1. Geometric interpretation of the l_1–l_∞ constrained portfolio and the short-sale constrained portfolio.

the line. For the l_1–l_∞ constraint, the optimal portfolio corresponds to the intersection point of the line with the polygon. Properties of this portfolio will be discussed in Section 3.

2.2. The l_1–$l_\infty^{(p)}$ constrained MVP

The l_∞ constraint in (3) effectively level off the weights with large absolute values, in which the weights are treated equally or symmetrically. A generalization of this approach is to put more penalties on the larger weights, as did in the octagonal shrinkage and clustering algorithm (OSCAR) for linear regression analysis (Bondell and Reich, 2008). In this case, we replace the standard l_∞ norm by the sum of the maximum of each two components of the weight vector (i.e. pairwise l_∞ norm). Specifically, the constrained MVP problem can be written in the following form.

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

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

$$\|\mathbf{w}\|_1 + \alpha \sum_{i<j} \max\{|w_i|, |w_j|\} \leqslant c, \tag{8}$$

where $\alpha \geq 0$ and $c > 0$ are tuning parameters with c being the threshold of the constraint and α controlling the relative weights between the l_1 norm constraint and the sum of pairwise l_∞ norm constraint. Similar as before, the constrained optimization problem is equivalent to the following penalized optimization problem

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

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

where λ is the penalty parameter and there is a one to one correspondence between c and λ.

Since $\sum_{i<j} \max\{|w_i|, |w_j|\} = \sum_{i=1}^{N} (i-1)|w|_{(i)}$, where $|w|_{(i)}$ is the i-th smallest component of $|\mathbf{w}|$, then the constraint in (8) effectively puts weights $i-1$ on the i-th smallest value of $|\mathbf{w}|$. The 2-dimension region of the OSCAR constraint is the same as that for the l_1–l_∞ constraint. The constraint region for higher dimension case is also a polygon, but has more vertices than the region of l_1–l_∞ 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_∞ constraint. If we classify the assets with highly correlated coefficients as a group, the optimal portfolio with only a l_1 norm constraint tends to arbitrarily select only one from the group and forces the other weights at the same group to be zero. That will increase the specification error and hard to be explained in practice. With the pairwise l_∞ constraints, we can allocate the same weights to those highly correlated assets, so that to decrease the specification error.

Observe that under the condition $\mathbf{1}^\top \mathbf{w} = 1$ the minimal value of $\|\mathbf{w}\|_1 + \alpha \sum_{i<j} \max\{|w_i|, |w_j|\}$ in (8) is $1 + \frac{\alpha(N-1)}{2}$ which is attained when all weights are equal $(w_i = 1/N, i = 1, 2, \ldots, N)$. Therefore, when the threshold c is taken to be close to this minimal value (i.e., the constraint is very stringent), then the solution to the l_1–$l_\infty^{(p)}$ optimization problem will be the $1/N$ policy. On the other hand, if c is much larger than $1 + \frac{\alpha(N-1)}{2}$, the solution tends to have grouped structure and within each group the weights are equal. In this sense, the l_1–$l_\infty^{(p)}$ strategy can be regarded as a generalization of the $1/N$ strategy.

3. Properties

We will show in this section that the l_1–l_∞ norm-constrained optimization problem can be regarded as being equivalent to an optimization problem with no norm constraints in which the sample covariance matrix is properly modified. The modification on the covariance matrix is determined by the norm constraint imposed on the asset weights.

Proposition 1. *The l_1–l_∞ constrained minimization problem* (4)–(5) *is equivalent to the following minimum-variance problem*

$$\min_{\mathbf{w}} \left\{\mathbf{w}^\top \widetilde{\Sigma}\mathbf{w}\right\}, \quad \text{s.t. } \mathbf{w}^\top \mathbf{1} = 1,$$

where

$$\widetilde{\Sigma} = \widehat{\Sigma} - \lambda(\mathbf{v}\mathbf{1}^\top + \mathbf{1}\mathbf{v}^\top) + \frac{1}{2}\lambda\alpha(\mathbf{s}\mathbf{1}^\top + \mathbf{1}\mathbf{s}^\top) \tag{11}$$

where $\mathbf{v}$ *is an* $N \times 1$ *vector whose i-th component is 1 if* w_i *is negative and is 0 otherwise, and* $\mathbf{s}$*is an* $N \times 1$ *vector whose i-th component is* $\text{sign}(w_i)$ *if* w_i *has the largest absolute value, and is 0 otherwise.*

This proposition shows that the l_1–l_∞ norm constrained portfolios can be interpreted as the solution to the conventional minimum variance problem with the sample covariance matrix properly modified by as in (11). When $\alpha = 0$, our procedure reduces to the l_1 constrained problem and the modified covariance matrix $\widetilde{\Sigma} = \widehat{\Sigma} - \lambda(\mathbf{v}\mathbf{1}' + \mathbf{1}\mathbf{v}')$ was obtained by DeMiguel et al. (2009a). This modification implies that the (i,j) element, $\widehat{\sigma}_{ij}$ of $\widehat{\Sigma}$

is shrunk to $\widetilde{\sigma}_{ij} = \widehat{\sigma}_{ij} - \lambda(v_i + v_j)$, where v_i is the indicator of the event $w_i < 0$, i.e., the covariances of two assets are reduced by λ or 2λ, depending upon one or both of their weights are negative. From empirical observations, investors tend to sell short those assets with large estimated variances and high estimated covariances with other assets. Therefore the l_1 constraint method slightly reduces these large variance estimates, and shrinking some small negative weights to zero. The third term, $\frac{1}{2}\lambda\alpha(\mathbf{s1}^\top + \mathbf{1s}^\top)$, in (11) is a result of the l_∞ constraint. In the empirical observation, the asset often has large positive weight, if it has a lower variance and has lower covariances to other assets; and the asset often has large negative weight, if it has a larger variance and has higher covariances to other assets, which is documented in Jagannathan and Ma (2003). But large variance estimates tend to have large estimation error and tend to be overly large, especially when the number of assets is near the size of the data set. Thus the third term can be regarded as a penalty to the asset which has the largest absolute value of the weights. For the asset with largest absolute weight, if the weight is negative (positive), then the l_∞ constraint reduces (raises) the variance further by a quantity of $\lambda\alpha$, and reduces (raises) the covariance estimates of this asset with others by $\lambda\alpha/2$.

Proposition 2. *The* l_1–$l_\infty^{(p)}$ *strategy* (9)–(10) *is equivalent to the minimum-variance problem*

$$\min_{\mathbf{w}}\left\{\mathbf{w}^\top \widetilde{\Sigma}\mathbf{w}\right\}, \quad \text{s.t. } \mathbf{w}^\top\mathbf{1} = 1,$$

where

$$\widetilde{\Sigma} = \widehat{\Sigma} - \lambda(\mathbf{v1}^\top + \mathbf{1v}^\top) + \frac{1}{2}\lambda\alpha(\mathbf{o1}^\top + \mathbf{1o}^\top), \tag{12}$$

where $\mathbf{o}$ *is an* $N \times 1$ *vector with the i-th component being* $o_i = (\text{Rank}(|w_i|) - 1)\text{sign}(w_i)$, *and* $\text{Rank}(|w_i|)$ *is the rank of* $|w_i|$ *in* $\{|w_1|, \ldots, |w_N|\}$.

Similar as before, the second term in (12) is a result of the l_1 constraint which effectively reduces the variances and covariances of the assets with negative weights. The third term corresponds to the pairwise l_∞ constraint, it does not penalize only on the asset with largest absolute variance, but also on other assets with relative large variance.

Fan et al. (2008) documented that, for l_1 constrained MVP problem, the difference between the minimum risk obtained from using sample covariance matrix and that from the true covariance matrix is dominated by the estimation error of the covariance matrix. For the l_1–l_∞ constrained problem, we have similar result.

Proposition 3. *Let T be sample size and let* $\widehat{\mathbf{w}}_{\text{opt}}$ *be the solution to the* l_1–l_∞ *constrained problem* (1)–(3) *and denote the minimum variance of the optimal portfolio by* $R_T(\widehat{\mathbf{w}}_{\text{opt}}) = \widehat{\mathbf{w}}_{\text{opt}}^\top \widehat{\Sigma}\widehat{\mathbf{w}}_{\text{opt}}$. *Let* $\mathbf{w}_{\text{opt}}$ *be the optimizer of problem* (1)–(3) *when the true covariance matrix* Σ *is used in* (1) *instead of* $\widehat{\Sigma}$, *and* $R(\mathbf{w}_{\text{opt}}) = \mathbf{w}_{\text{opt}}^\top\Sigma\mathbf{w}_{\text{opt}}$ *be the minimum variance when* Σ *is used. Then we have*

$$|R(\mathbf{w}_{\text{opt}}) - R_T(\widehat{\mathbf{w}}_{\text{opt}})| < 3a_T\left(c - \frac{\alpha}{N}\right)^2 \tag{13}$$

where $a_T = \|\widehat{\Sigma} - \Sigma\|_\infty = \max\{|\widehat{\sigma}_{ij} - \sigma_{ij}|, 1 \leqslant i,j \leqslant N\}$.

It was shown in Fan et al. (2008) that under mild conditions the estimation error $a_T = \|\widehat{\Sigma} - \Sigma\| = O_p\left(\sqrt{\frac{\log N}{T}}\right)$. Therefore, as far as $\log N/T = o_p(1), a_T$ and $|R(\mathbf{w}_{opt}) - R_T(\widehat{\mathbf{w}}_{\text{opt}})|$ converges to 0 in probability.

4. Out-of-sample evaluation

In this section, we analyze four empirical data sets and investigate our approach by simulation studies. We compare the out-of-sample performance of our strategies with several existing norm-constrained portfolios and the methods without norm constraints, namely, the MVP with shortsale constraint (MINC) and the MVP without shortsale constraints (MINU). Descriptions of the four real data sets are given in Table 1.

4.1. Methods for evaluating out-of-sample performances

As did in DeMiguel et al. (2009a), DeMiguel et al. (2009b), the out-of-sample performances of a portfolio can be measured by the portfolio variance, the Sharpe ratio, the portfolio turnover (trading volume) and return-loss relative to the $1/N$ strategy.These quantities are computed by the "rolling window" method with a specific window width τ. In our implementations, the window width is chosen to be 120 months and there are $\tau = 120$ data points in each window. Let the total number of data points be T. For $t = \tau, \tau + 1, \ldots, T - 1$, the "rolling window" procedure applies the optimization procedure to data points in the window $[t - \tau + 1, t]$ to calculate the optimal portfolio weight $\widehat{\mathbf{w}}_t$, and predict the portfolio return at time $t + 1$ by $\widehat{\mathbf{w}}_t^\top\mathbf{r}_{t+1}$, where $\mathbf{r}_{t+1}$ denotes the vector of all N asset returns at time $t + 1$. The resulted predictions will be used to compute the out-of-sample mean and variance of the optimal portfolio as given below.

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

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

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

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

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

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

where $\hat{w}_{t+1}$ is the desired portfolio weight at $t + 1$ after rebalancing, and $\hat{w}_{t^+}$ is the portfolio weight before rebalancing at $t + 1$. For each strategy denoted by k, the return-loss relative to $1/N$ strategy is defined as

$$\text{return} - \text{loss}_k = \frac{\mu_{ew}}{\sigma_{ew}} \times \sigma_k - \mu_k,$$

where μ_k and σ_k are the out-of-sample mean and variance of each strategy k, μ_{ew} and σ_{ew} are the out-of-sample mean and variance of $1/N$ strategy. The return-loss measures the additional return needed for strategy k to perform as well as the $1/N$ strategy in terms of the Sharpe ratio.

Note that these quantities are dependent on the specific values of tuning parameters (α, c) in the constrained MVP problem. The tuning parameters are chosen to be the ones that correspond to the best out-of-sample performance. We use cross validation method to achieve this goal. Specifically, if our objective is to maximize the out-of-sample sharpe ratio, based on the τ sample returns in the estimation window, we firstly delete the ith $(1 \leqslant i \leqslant \tau)$ sample return from the estimation window, then calculate the optimal portfolio weights $(\widehat{\mathbf{w}}_{(\alpha,c)})_{(i)}$ based on the sample returns without the ith sample return, finally calculate the out-of-sample return $(\widehat{\mathbf{w}}_{(\alpha,c)})_{(i)}^\top\mathbf{r}_i$. Repeat the above process for each i ranging from 1 to τ, thus we have τ out-of-sample portfolio returns $(r_{(\alpha,c)})_i$ which are dependent on the tuning parameters

Table 1
The data sets of monthly asset returns analyzed in this study.

No.	Data set	Abbreviation	N	Time period	Source
1	Ten industry portfolios	10Ind	10	01/1970–07/2011	K.French
2	Thirty industry portfolios	30Ind	30	01/1970–07/2011	K.French
3	One hundred Portfolios	100FF	100	01/1970–07/2011	K.French
4	One hundred Random Portfolio	CRSP	100	12/1971–05/2011	CRSP

[a]This table lists four data sets of monthly asset returns. The first and second sets are industry portfolios. The third set is formed on size and book-to-market. The forth set selects 100 stocks randomly from CRSP (The Center for Research in Security Prices). [b]Source: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

(α, c), so that we can calculate the out-of-sample Sharpe ratio $\widehat{\mathrm{SR}}_{(\alpha,c)} = \frac{\widehat{\mu}}{\widehat{\sigma}}$, where $\widehat{\mu} = \frac{1}{\tau}\sum_{i=1}^{\tau}(r_{(\alpha,c)})_i$, and $\widehat{\sigma}^2 = \frac{1}{\tau-1}\sum_{i=1}^{\tau-1}((r_{(\alpha,c)})_i - \widehat{\mu})^2$. $\widehat{\sigma}$ is the square root of $\widehat{\sigma}^2$. We select the optimal (α, c) that maximizes the out-of-sample Sharpe ratio; that is, $(\alpha^{\star}, c^{\star}) = argmax_{(\alpha,c)}\widehat{\mathrm{SR}}_{(\alpha,c)}$. Similar procedure is applied in the case of minimizing the out-of-sample variance or turnover.

4.2. Real data analysis

Table 2 shows the out-of-sample Sharpe ratios, variances and turnovers for the four data sets 10Ind, 30Ind, 100FF, CRSP under different portfolio strategies. MINU is the minimum-variance portfolio with short sales unconstrained and MINC is the minimum-variance portfolio with shortsales constrained. DS is the double shrinkage method proposed by Candelon et al. (2012). We can see that the l_1–l_∞ 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 other three norm constrained strategies including DS approach have similar Sharpe ratios and are generally smaller than those of l_1–l_∞ constraint strategy and the l_1–$l_\infty^{(p)}$ strategy, though the differences are not always statistically significant. As for out-of-sample variance, the l_1–l_∞ strategy performs the best among all the methods for the 10Ind data, and the differences are usually statistically significant. But for the other three data sets with larger number of assets, other approaches may have smaller variances. In addition, the l_1–$l_\infty^{(p)}$ method generally has smaller turnover values while the l_1–l_∞ strategy has larger turnover values than other norm constrained strategies, which is not unexpected since the l_1–$l_\infty^{(p)}$ approach tends to cluster assets and assign equal weight to assets in each cluster. Therefore, the l_1–$l_\infty^{(p)}$ strategy resembles the $1/N$-portfolio most in sense of having stable weights.

Table 3
Return-loss relative to $1/N$ strategy for real data sets.

Source	10Ind	30Ind	100FF	CRSP
l_1–l_∞	−0.00249	−0.00271	−0.00946	−0.00066
l_1–$l_\infty^{(p)}$	−0.00192	−0.00230	−0.00800	−0.00013
DS	−0.00208	−0.00253	−0.00875	−0.00011
MINU	−0.00207	−0.00142	0.01020	0.00247
MINC	−0.00137	−0.00149	−0.00243	−0.00031
l_1	−0.00212	−0.00213	−0.00913	−0.00031
l_2	−0.00229	−0.00271	−0.00872	−0.0001
l_1–l_2	−0.00229	−0.00260	−0.00877	−0.00047
1/N	0.00000	0.00000	0.00000	0.00000

Table 3 presents the return-loss for each strategy k for the four data sets 10Ind, 30Ind, 100FF, CRSP. From the perspective of return-loss measure, l_1–l_∞ performs best across the four data sets.

Fig. 2 illustrates the solution paths of several constrained optimization methods applied to the 10Ind data set presented in 1. Each line corresponds to one asset, representing the weights as a function of the threshold c with a fixed value of α which is chosen according to the best out-of-sample performance in Table 2. All lines are around zero for small value of c, which means that the solutions are sparse when the constraints are stringent. This figure indicates that, compared with the l_1 or the l_1–l_2 method, the l_1–l_∞ and the l_1–$l_\infty^{(p)}$ strategies can have more sparse solutions. From the first panel of Fig. 2, the l_1–l_∞ have a good control of large absolute solutions. From the second panel one can see that 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.

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]	[–]	[–]	[–]	[–]

[a]The numbers in square brackets are p-values of the portfolio Sharpe ratios and variances for a strategy is different from that for l_1–l_∞ strategy. The p-values are computed using the stationary bootstrap method proposed by Ledoit and Wolf (2008).

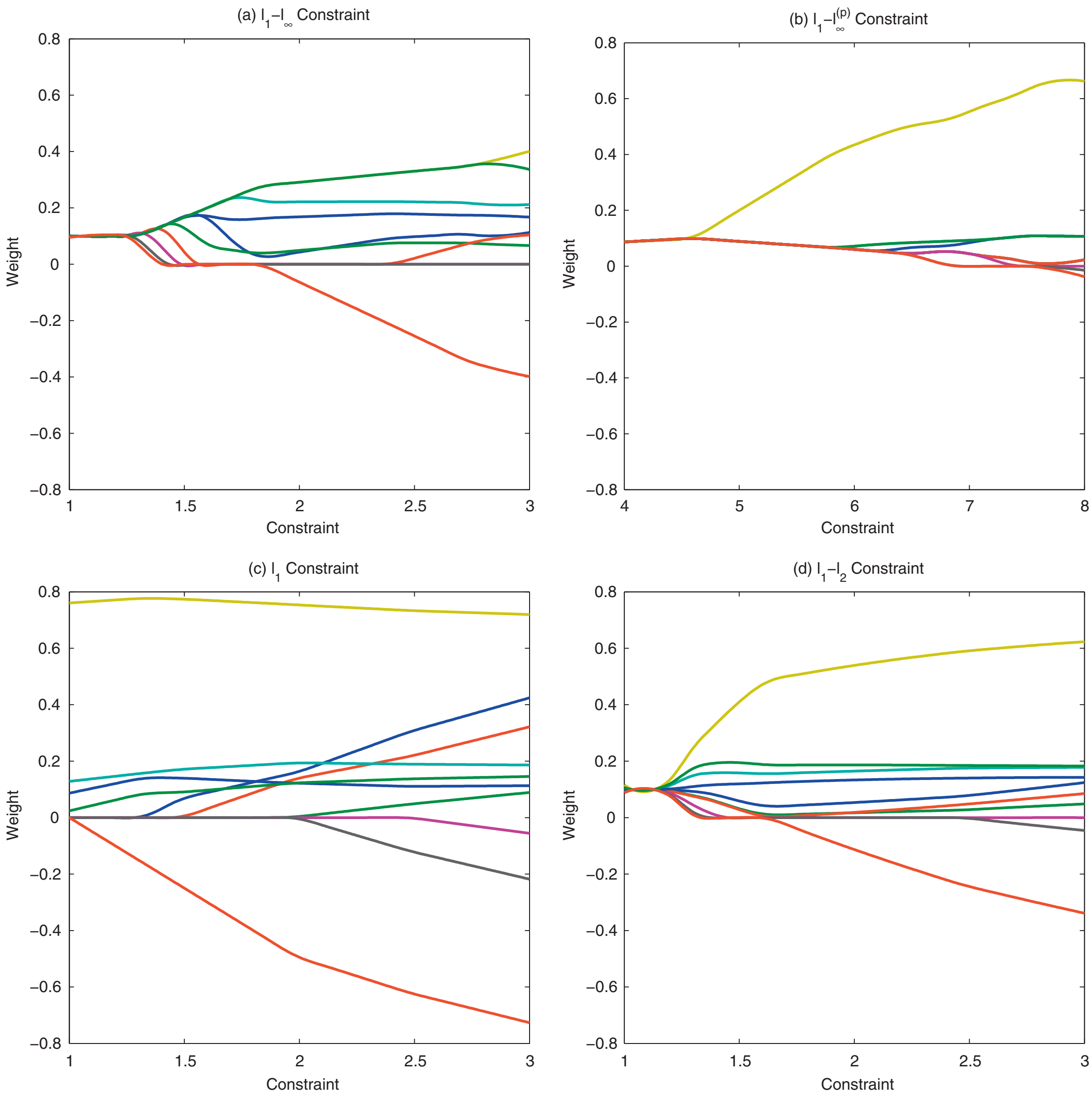

Fig. 2. Solution paths of several norm constrained portfolios.

4.3. Simulation study

In this section, we investigate our methods by simulation studies. We use the three-factor model in Fama and French (1993) to generalize data. Let $\mathbf{r} = (r_1, \ldots, r_N)^\top$, where r_i is the return rate of asset i; $\mathbf{f} = (K_m - r_f, SMB, HML)^\top$, where K_m is the return of the whole stock market, $K_m - r_f$ is the first factor which is similar to the *CAPM* (capital asset pricing model), *SMB* and *HML* stand for "small (market capitalization) minus big" factor and "high (book-to-market ratio) minus low" factor respectively. These factors represent the historic excess returns of small caps over big caps and of value stocks over growth stocks. Let r_f be the risk-free return rate, $\epsilon = (\epsilon_1, \ldots, \epsilon_N)^\top$, and

$$B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ \cdots & \cdots & \cdots \\ b_{N1} & b_{N2} & b_{N3} \end{pmatrix},$$

where $b_{ij}, 1 \leqslant i \leqslant N, 1 \leqslant j \leqslant 3$ are the factor loadings. The three factor model can be represented as follows,

$$\mathbf{r} = \mathbf{1}r_f + B\mathbf{f} + \epsilon. \tag{14}$$

We assume that $E(\epsilon|f) = 0$ and $\text{cov}(\epsilon|f) = \text{diag}(\sigma_1^2, \ldots, \sigma_N^2)$, then the covariance matrix of $\mathbf{r}$ is shown to be

$$\Sigma = B\text{cov}(f)B^\top + \text{cov}(\epsilon) = B\Sigma_f B^\top + \text{diag}(\sigma_1^2, \ldots, \sigma_N^2). \tag{15}$$

We generate $T = 10, 120$ monthly returns for $N = 30$ assets for two different sets of factor loadings (denoted as scenario I and II) and two different sets of error distributions (denoted as $h = 0$ and $h = 0.05$, see below for details). First, two sets of factor loading coefficients are generated from the multi-normal distributions $N_{\text{I}}(\mu_b, \Sigma_b)$ and $N_{\text{II}}(\mu_b, \Sigma_b)$ using parameter values specified in Table 4. The factor loading coefficients are fixed once being generated. Second, the three factors are generated from the multi-normal distribution $N(\mu_f, \Sigma_f)$ with parameters as shown in Table 4. The parameter values in Table 4 are estimated from real data. Third, similar to Fan et al. (2008), $\sigma_1, \ldots, \sigma_N$ are generated from Gamma distribution $Gamma_{\text{I}}(1, 3)$ and $Gamma_{\text{II}}(9, 0.5)$ for scenarios I and II, respectively. These two distributions are obtained by empirically fitting the 100Ind, 30Ind data sets respectively. Finally, we generate $\epsilon_1, \ldots, \epsilon_N$ independently and ϵ_i is generated from a contaminated normal distribution as follows. With probability $1 - h, \epsilon_i$ is generated from normal distribution $N(0, \sigma_i^2)$ and with

Table 4
The data sets of monthly asset returns analyzed in this study.

Parameters for factor loadings				Parameters for factors			
I-μ_b	I-Σ_b			μ_f	Σ_f		
1.086	0.010	−0.001	0.002	0.446	22.128	4.429	−4.522
0.490	−0.001	0.250	0.011	0.185	4.429	10.246	−2.397
0.365	0.002	0.011	0.204	0.405	−4.522	-2.397	9.300
II-μ_b	II-Σ_b			μ_f	Σ_f		
1.055	0.031	0.030	0.010	0.446	22.128	4.429	−4.522
0.101	0.030	0.061	0.012	0.185	4.429	10.246	−2.397
0.241	0.010	0.012	0.082	0.405	−4.522	-2.397	9.300

[a]The parameters of the Factor Loading Coefficients of I and II are calculated from monthly return data set 100Ind and 30Ind presented in Table 1. The mean and covariance matrix of the factor is calculated from the Fama/French Factors in Kenneth R. French Data Library. Through Using the least-square regression, we get the coefficients then calculate the sample mean and covariance matrix as the parameters for the factors.

Table 5
Simulation results of performance measures.

Scenario h	Sharpe ratio				Variance ($\times 10^{-3}$)				Turnover			
	I 0	I 0.05	II 0	II 0.05	I 0	I 0.05	II 0	II 0.05	I 0	I 0.05	II 0	II 0.05
l_1–l_∞	0.161	0.155	0.125	0.120	1.576	1.912	1.880	2.168	0.062	0.047	0.041	0.038
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[–]	[–]	[–]	[–]
l_1–$l_\infty^{(p)}$	0.156	0.150	0.121	0.118	1.460	1.588	1.916	2.505	0.109	0.107	0.002	0.003
	[0.499]	[0.321]	[0.047]	[0.015]	[0.001]	[0.001]	[0.005]	[0.001]	[–]	[–]	[–]	[–]
DS	0.145	0.150	0.120	0.118	1.788	2.126	1.613	1.980	0.038	0.028	0.080	0.026
	[0.636]	[0.749]	[0.139]	[0.001]	[0.001]	[0.001]	[0.001]	[0.804]	[–]	[–]	[–]	[–]
MINU	0.046	0.060	0.065	0.067	0.441	0.821	1.141	1.721	0.393	0.360	0.254	0.205
	[0.015]	[0.017]	[0.052]	[0.009]	[0.001]	[0.001]	[0.004]	[0.001]	[–]	[–]	[–]	[–]
MINC	0.151	0.145	0.091	0.090	1.689	1.785	1.171	1.546	0.061	0.055	0.041	0.070
	[0.507]	[0.733]	[0.146]	[0.558]	[0.001]	[0.001]	[0.001]	[0.151]	[–]	[–]	[–]	[–]
l_1	0.156	0.146	0.091	0.090	1.328	1.298	1.171	1.564	0.078	0.109	0.041	0.070
	[0.738]	[0.742]	[0.140]	[0.531]	[0.001]	[0.001]	[0.001]	[0.154]	[–]	[–]	[–]	[–]
l_2	0.159	0.149	0.123	0.119	1.387	1.550	1.971	2.111	0.066	0.025	0.029	0.046
	[0.568]	[0.827]	[0.100]	[0.004]	[0.001]	[0.001]	[0.006]	[0.679]	[–]	[–]	[–]	[–]
l_1–l_2	0.159	0.150	0.122	0.118	1.534	1.679	2.509	2.277	0.056	0.065	0.025	0.018
	[0.609]	[0.600]	[0.035]	[0.016]	[0.001]	[0.011]	[0.311]	[0.002]	[–]	[–]	[–]	[–]
1/N	0.133	0.134	0.120	0.115	2.798	2.828	2.551	2.653	0.000	0.000	0.000	0.000
	[0.446]	[0.319]	[0.017]	[0.082]	[0.001]	[0.001]	[0.001]	[0.001]	[–]	[–]	[–]	[–]

[a]The numbers in square brackets are p-values of the portfolio Sharpe ratios and variances for a strategy is different from that for l_1–l_∞ strategy. The p-values are computed using the stationary bootstrap method proposed by Ledoit and Wolf (2008).

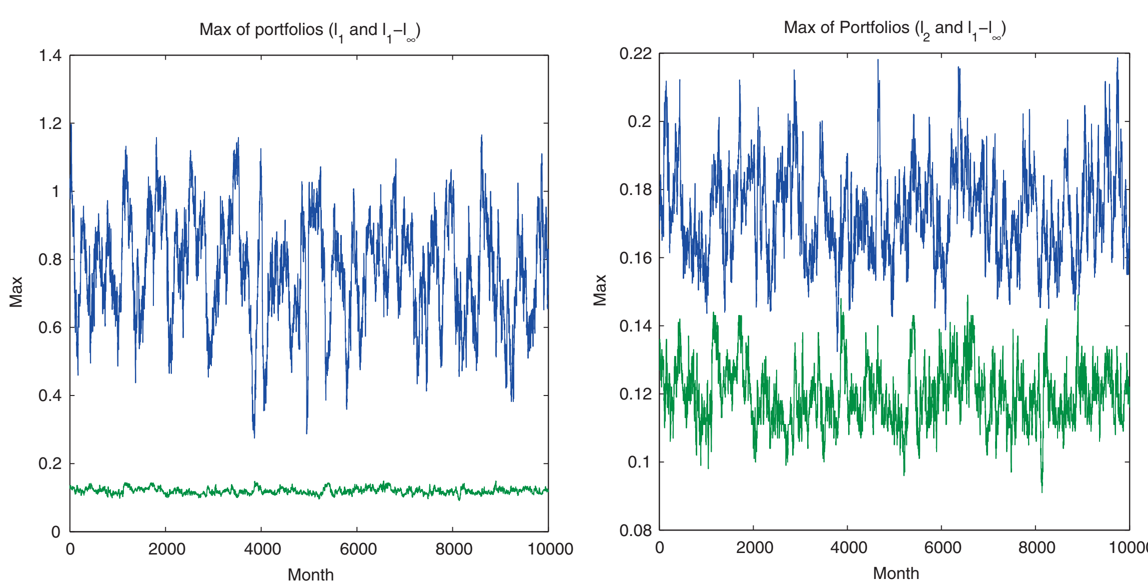

Fig. 3. The left panel shows the maximum absolute weights of l_1 and l_1–l_∞ methods, the blue line is for l_1 strategy and the green line is for l_1–l_∞ strategy. The right panel shows the maximum absolute weights of l_2 and l_1–l_∞ methods, the blue line is for l_2 strategy and the green line is for l_1–l_∞ strategy.

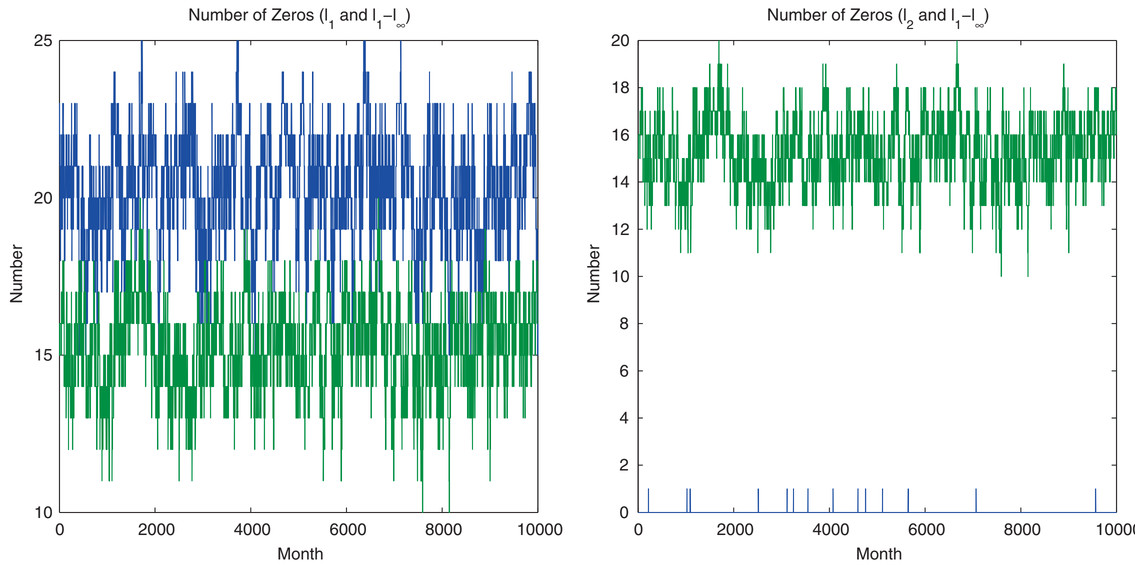

Fig. 4. The left panel shows the number of zeros of the weights of the l_1 and l_1–l_∞ methods, the blue line is for l_1 strategy and the green line is for l_1–l_∞ strategy. The right panel shows the number of zeros of the weights of l_2 and l_1–l_∞ methods, the blue line is for l_2 strategy and the green line is for l_1–l_∞ strategy.

probability h, ϵ_i is generated as $5\sigma_i^2$ or $-5\sigma_i^2$ with equal probabilities. The contamination proportion h is taken to be small ($h = 0$ and $h = 0.05$) in our simulations. When $h = 0$, the error ϵ's are from normal distribution and there are no outliers.

We generate monthly return data using the method described above for $T = 10$, 120 months and use a window width $\tau = 120$ months, which matches our choice when analyzing the real data sets. Then we use the last 10,000 months ($T - \tau = 10,000$ months) for the out-of-sample performance evaluation. The out-of-sample Sharpe Ratio, Variance and Turnover using the method are calculated as described in Section 4.1. We generate four data sets with the same factor loading coefficients. Results are listed in 4.

Table 5 shows the out-of-sample Sharpe ratios, variances and turnovers for the four simulated data sets. We can see that the l_1–l_∞ strategy has the largest out-of-sample Sharpe ratio across all of the four data sets for different h and different scenarios I and II, though the differences are not always statistically significant,but the l_1–$l_\infty^{(p)}$ strategy have relative lower Sharpe ratios. Though the l_2 and l_1–l_2 strategies have large Sharpe ratios, the weights are very similar to the $1/N$ strategy. The weights estimated by the l_1–l_∞ strategy have good sparse property. As for out-of-sample variance, the l_1–l_∞ method performs better than the l_1–l_2 method, and the differences are usually statistically significant.

In addition, the weights produced from the l_1–$l_\infty^{(p)}$ strategy have the lowest turnover values. Fig. 3 shows that the weights of assets obtained in the 10,000 rolling windows under scenario I. It can be seen that the l_1–l_∞ method is not only stable but also produces smaller weights than both the l_1 and l_2 strategies. Fig. 4 shows that the weights of assets obtained from the l_1–l_∞ strategy are more sparse than the l_2 method but not as sparse as those of the l_1 method.

5. Conclusion

Many robust approaches to the MVP problem had been proposed in the literature. Among them, the l_1 or l_2 norm constrained optimization method have been shown to be robust to estimation error of the sample covariance estimate. Optimization methods using the combination of two norm constraints have been widely studied in statistical literature. Especially, the elastic net (l_1–l_2) method and the OSCAR (l_1–$l_\infty^{(p)}$) method have been proposed to manage highly correlated or clustered variables while controlling sparsity in regression problem. In this study, we apply the similar approaches in the framework of optimizing MVP.

Since exceptionally large absolute weights may be unstable and sensitive to estimation error of sample covariance matrix, in this study we propose an l_1–l_∞ norm constrained method for MVP optimization. We also consider imposing a pairwise l_∞ constraint to encourage the highly correlated assets being clustered by forcing the corresponding weights to be equal. An extremal case is the $1/N$ portfolio, in which all the weights are equal and all assets is thought to form one cluster and there are no subclusters. In this sense, the pairwise l_1–$l_\infty^{(p)}$ method is parallel to the $1/N$ strategy, but allowing cluster structure among various assets.

We show that the proposed norm constrained optimization method can be thought as unconstrained optimization problems with the estimated covariance matrix being modified by factors related to the norm constraints. Empirical data analysis and simulation studies show that 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. In real data analysis and simulation studies, the l_1–l_∞ 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.

Acknowledgments

This work is partially supported by China NSF grant. (Hu and Yang).

Appendix A. Proof of Propositions

A.1. Proof of Proposition 1

Proof. Proof of Proposition 1. Through a simple algebra, constraint (3) can be written as

$$1-2\sum_{i\in\mathcal{N}}w_i+\alpha|\mathbf{w}_{\mathcal{M}}|\leqslant c,$$

where
$\mathcal{N}=\{i:w_i<0\},\mathcal{M}=\{i,|w_i|\ \textit{is the maximum of}\ |w_1|\ldots,|w_N|\}$.
Then using the penalty method, the optimization problem (1)–(3) can be presented as

$$\min_w \mathbf{w}^\top\widehat{\Sigma}\mathbf{w}+\lambda\left(1-2\sum_{i\in\mathcal{N}}w_i+\alpha|\mathbf{w}_{\mathcal{M}}|\right),$$
$$\text{s.t.}\ \sum_{i=1}^N w_i=1.$$

With the constraint $\mathbf{w}^\top\mathbf{1}=1$, the problem can be rewritten as

$$\min_{\mathbf{w}} \mathbf{w}^\top\{\widehat{\Sigma}-\lambda[\mathbf{v}\mathbf{1}^\top+\mathbf{1}\mathbf{v}^\top-\frac{1}{2}\alpha(\mathbf{s}\mathbf{1}^\top+\mathbf{1}\mathbf{s}^\top)]\}\mathbf{w},$$
$$\text{s.t.}\ \sum_{i=1}^N w_i=1,$$

where $\mathbf{v}\in\mathcal{R}^n$ is a vector whose i-th component is one if the i-th weight is negative, and zero otherwise, and $\mathbf{s}\in\mathcal{R}^n$ is a vector whose i-th component is $sign(w_i)$ if the i-th weight has the largest absolute value. This completes the proof.

Proof. Proof of Proposition 2. Note that constraint (8) can be rewritten as

$$\|\mathbf{w}\|_1+\alpha\sum_{i=1}^N(i-1)|w|_{(i)}\leqslant c,$$

where $|w|_{(1)}\leqslant|w|_{(2)}\leqslant\cdots\leqslant|w|_{(n)}$. The rest of the proof is completely analogous to those for Proposition 1. □

A.2. Proof of Proposition 3: Note that

$$\begin{aligned}|R(\mathbf{w}_{\rm opt})-R_T(\widehat{\mathbf{w}}_{\rm opt})|&\leqslant|R(\mathbf{w}_{\rm opt})-R(\widehat{\mathbf{w}}_{\rm opt})|+|R(\widehat{\mathbf{w}}_{\rm opt})-R_T(\widehat{\mathbf{w}}_{\rm opt})|\\&=R(\widehat{\mathbf{w}}_{\rm opt})-R_T(\widehat{\mathbf{w}}_{\rm opt})+R_T(\widehat{\mathbf{w}}_{\rm opt})-R_T(\mathbf{w}_{opt})\\&\quad+R_T(\mathbf{w}_{opt})-R(\mathbf{w}_{\rm opt})+|R(\widehat{\mathbf{w}}_{\rm opt})-R_T(\widehat{\mathbf{w}}_{\rm opt})|\\&\leqslant|R(\widehat{\mathbf{w}}_{\rm opt})-R_T(\widehat{\mathbf{w}}_{\rm opt})|+|R(\mathbf{w}_{\rm opt})-R_T(\mathbf{w}_{\rm opt})|\\&\quad+|R(\widehat{\mathbf{w}}_{\rm opt})-R_T(\widehat{\mathbf{w}}_{\rm opt})|\\&\leqslant 3\sup_{\|\mathbf{w}\|_{l_1}+\|\mathbf{w}_\infty\|\leqslant c}|R(\mathbf{w})-R_T(\mathbf{w})|,\end{aligned}$$

where the second equality is derived by the fact that $R(\mathbf{w}_{\rm opt})-R(\widehat{\mathbf{w}}_{\rm opt})\leqslant 0$, since $\mathbf{w}_{\rm opt}$ is the minimizer of $R(w)$. The third inequality is derived from the fact that $R_T(\widehat{\mathbf{w}}_{\rm opt})-R_T(\mathbf{w}_{\rm opt})\leqslant 0$ since $\widehat{\mathbf{w}}_{\rm opt}$ is the optimal solution that minimizes $R_T(\mathbf{w})$.

With simple algebra, it is easy to obtain:

$$|R(\mathbf{w})-R_T(\mathbf{w})|=|\mathbf{w}^\top(\widehat{\Sigma}-\Sigma)\mathbf{w}|\leqslant a_T\|\mathbf{w}\|_1^2.$$

Using the l_1–l_∞ constraint $\|\mathbf{w}\|_1+\alpha\|\mathbf{w}\|_\infty\leqslant c$, we can easily get

$$\|\mathbf{w}\|_1+\alpha\|\mathbf{w}\|_\infty-\frac{\alpha}{N}\leqslant c-\frac{\alpha}{N}.$$

Using the fact, $\alpha\|\mathbf{w}\|_\infty-\frac{\alpha}{N}\geqslant 0$, then we have:

$$\|\mathbf{w}\|_1\leqslant\left(c-\frac{\alpha}{N}\right).$$

Combining the above facts, we have

$$|R(\mathbf{w}_{\rm opt})-R_n(\widehat{\mathbf{w}}_{\rm opt})|\leqslant 3a_T\left(c-\frac{\alpha}{N}\right)^2.$$

Appendix B. The algorithm

Here we describe the algorithm in detail to calculate the two proposed constrained MVP optimization problem. We can write $w_i=w_i^+-w_i^-,|w_i|=w_i^++w_i^-$, where $w_i^+=w_i$ if $w_i\geqslant 0,w_i^+=0$, otherwise; $w_i^-=-w_i$ if $w_i\leqslant 0$, and $w_i^-=0$, otherwise. Both w_i^+ and w_i^- are non-negative and only one is nonzero, if $w_i\neq 0$. Without loss of generality, suppose that $|w_N|=max_{1\leqslant i\leqslant N}|w_i|$, The proposed l_1–l_∞ constrained MVP optimization problem can be written as follows.

$$\begin{aligned}&\min\ \mathbf{w}^\top\widehat{\Sigma}\mathbf{w},\\&\text{s.t.}\ |w_i|\leqslant|w_N|,\ 1\leqslant i\leqslant N-1,\\&\quad\sum_{i=1}^N(w_i^++w_i^-)+\alpha(w_N^++w_N^-)\leqslant c,\\&\quad\sum_{i=1}^N(w_i^+-w_i^-)=1,\\&\quad w_i^+\geqslant 0,\ w_i^-\geqslant 0,\ 1\leqslant i\leqslant N.\end{aligned}$$

There are N possible maximum of the weights, so the above quadratic program has $3N+1$ linear constraints.

Without loss of generality, suppose further that $|w_1|\leqslant|w_2|\leqslant\ldots\leqslant|w_N|$ the l_1–$l_\infty^{(p)}$ constrained MVP optimization problems can be written as follows.

$$\begin{aligned}&\min_w\ \mathbf{w}^\top\widehat{\Sigma}\mathbf{w},\\&\text{s.t.}\ |w_1|\leqslant|w_2|\leqslant\ldots\leqslant|w_N|,\\&\quad\sum_{i=1}^N(\alpha(i-1)+1)(w_i^++w_i^-)\leqslant c,\\&\quad\sum_{i=1}^N(w_i^+-w_i^-)=1,\\&\quad w_i^+\geqslant 0,\ w_i^-\geqslant 0,\ 1\leqslant i\leqslant N.\end{aligned}$$

The number of possible ordering of the weights is $N!$, results $N!$ possible weighted linear combinations. So the total number of linear constraints will be $N!+2N+1$. Instead of directly solving this large-scale quadratic programming problem, we adopt the sequential algorithm which proceeds as follows.

(i) Solve the quadratic program with 2N + 2 constraints using the ordering of weights obtained from solving the quadratic program with only one constraint ($\sum_i w_i=1$).
(ii) If the solution has different ordering, then add the linear constraint corresponding to the new ordering and solve the more restrictive quadratic programming.
(iii) Repeat (i), (ii) until the ordering remains constant, i.e, any additional constraint will no longer affect the solution.

The algorithm is based on the fact that, if the ordering of the solution does not change in the iterations, this solution automatically satisfies the remaining constraints. And the feasible region at each step contains the region in the previous steps, the algorithm is guaranteed to converge to the final fully constrained solution. And usually, it takes N steps or less to get the final solution.

Appendix C. Out-of-sample performance for different window length

Table C.1 and C.2 present the out-of-sample performance measures for the four real data sets with window length 60 and 180 respectively. The results are quite similar to the results in Table 2, this illustrate that the empirical results are robust to the different window length.

Table C.1
Out-of-sample performance measures for real data sets (τ is 60).

Source	Sharpe ratio				Variance ($\times 10^{-3}$)				Turnover			
	10Ind	30Ind	100FF	CRSP	10Ind	30Ind	100FF	CRSP	10Ind	30Ind	100FF	CRSP
l_1–l_∞	0.325	0.302	0.499	0.179	1.226	1.307	1.311	1.869	0.163	0.386	0.617	0.089
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[–]	[–]	[–]	[–]
l_1–$l_\infty^{(p)}$	0.266	0.284	0.431	0.165	1.246	1.327	1.569	2.023	0.055	0.146	0.223	0.051
	[0.493]	[0.120]	[0.158]	[0.233]	[0.001]	[0.047]	[0.105]	[0.033]	[–]	[–]	[–]	[–]
DS	0.312	0.297	0.440	0.162	1.241	1.214	1.271	2.084	0.244	0.181	0.643	0.040
	[0.405]	[0.279]	[0.180]	[0.395]	[0.001]	[0.311]	[0.150]	[0.147]	[–]	[–]	[–]	[–]
MINU	0.301	0.193	0.253	0.164	1.299	2.085	2.385	2.613	0.170	1.305	3.543	1.162
	[0.518]	[0.085]	[0.001]	[0.153]	[0.001]	[0.004]	[0.001]	[0.001]	[–]	[–]	[–]	[–]
MINC	0.302	0.282	0.282	0.174	1.245	1.263	1.780	1.336	0.092	0.117	0.191	0.261
	[0.226]	[0.854]	[0.001]	[0.024]	[0.001]	[0.029]	[0.004]	[0.198]	[–]	[–]	[–]	[–]
l_1	0.310	0.271	0.425	0.171	1.187	1.236	1.393	1.484	0.205	0.349	0.444	0.266
	[0.443]	[0.472]	[0.303]	[0.080]	[0.002]	[0.340]	[0.641]	[0.001]	[–]	[–]	[–]	[–]
l_2	0.321	0.297	0.450	0.167	1.227	1.196	1.264	1.855	0.085	0.231	0.510	0.872
	[0.348]	[0.323]	[0.428]	[0.196]	[0.001]	[0.831]	[0.410]	[0.001]	[–]	[–]	[–]	[–]
l_1–l_2	0.315	0.285	0.450	0.174	1.159	1.216	1.277	1.940	0.148	0.307	0.508	0.069
	[0.057]	[0.732]	[0.141]	[0.239]	[0.001]	[0.001]	[0.076]	[0.001]	[–]	[–]	[–]	[–]
1/N	0.266	0.243	0.253	0.164	1.825	2.331	2.385	2.241	0.023	0.028	0.026	0.005
	[0.499]	[0.053]	[0.001]	[0.191]	[0.540]	[0.001]	[0.001]	[0.001]	[–]	[–]	[–]	[–]

[a]The numbers in square brackets are p-values of the portfolio Sharpe ratios and variances for a strategy is different from that for l_1–l_∞ strategy. The p-values are computed using the stationary bootstrap method proposed by Ledoit and Wolf (2008).

Table C.2
Out-of-sample performance measures for real data sets (τ is 180).

Source	Sharpe ratio				Variance ($\times 10^{-3}$)				Turnover			
	10Ind	30Ind	100FF	CRSP	10Ind	30Ind	100FF	CRSP	10Ind	30Ind	100FF	CRSP
l_1–l_∞	0.300	0.302	0.477	0.183	1.280	1.243	1.229	1.761	0.062	0.135	0.254	0.044
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[–]	[–]	[–]	[–]
l_1–$l_\infty^{(p)}$	0.285	0.294	0.463	0.148	1.746	1.214	1.358	1.921	0.072	0.066	0.224	0.031
	[0.294]	[0.837]	[0.153]	[0.106]	[0.014]	[0.317]	[0.125]	[0.001]	[–]	[–]	[–]	[–]
DS	0.293	0.287	0.460	0.148	1.316	1.245	1.234	1.911	0.076	0.067	0.353	0.038
	[0.254]	[0.960]	[0.481]	[0.222]	[0.007]	[0.196]	[0.168]	[0.001]	[–]	[–]	[–]	[–]
MINU	0.293	0.257	0.328	0.050	1.341	1.330	2.441	2.965	0.106	0.299	2.098	0.986
	[0.251]	[0.846]	[0.654]	[0.223]	[0.009]	[0.336]	[0.001]	[0.001]	[–]	[–]	[–]	[–]
MINC	0.269	0.255	0.305	0.118	1.362	1.352	1.613	1.299	0.039	0.050	0.086	0.107
	[0.423]	[0.697]	[0.093]	[0.916]	[0.003]	[0.243]	[0.157]	[0.548]	[–]	[–]	[–]	[–]
l_1	0.291	0.276	0.440	0.149	1.322	1.229	1.322	1.756	0.096	0.136	0.189	0.236
	[0.247]	[0.973]	[0.002]	[0.213]	[0.009]	[0.125]	[0.040]	[0.001]	[–]	[–]	[–]	[–]
l_2	0.288	0.289	0.442	0.137	1.296	1.197	1.243	1.973	0.041	0.109	0.273	0.498
	[0.254]	[0.942]	[0.062]	[0.213]	[0.011]	[0.221]	[0.560]	[0.001]	[–]	[–]	[–]	[–]
l_1–l_2	0.293	0.287	0.455	0.157	1.289	1.195	1.240	1.997	0.056	0.122	0.233	0.511
	[0.144]	[0.516]	[0.002]	[0.230]	[0.002]	[0.009]	[0.015]	[0.004]	[–]	[–]	[–]	[–]
1/N	0.240	0.217	0.201	0.147	1.843	2.318	2.461	2.137	0.021	0.025	0.024	0.005
	[0.546]	[0.026]	[0.001]	[0.441]	[0.001]	[0.001]	[0.001]	[0.002]	[–]	[–]	[–]	[–]

[a]The numbers in square brackets are p-values of the portfolio Sharpe ratios and variances for a strategy is different from that for l_1–l_∞ strategy. The p-values are computed using the stationary bootstrap method proposed by Ledoit and Wolf (2008).

References

Bondell, H.D., Reich, B.J., 2008. Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR. Biometrics 64, 115–123.

Brandt, M.W., 2009. Portfolio choice problems. In: Ait-Sahalia, Y., Hansen, L.P. (Eds.), Handbook of Financial Econometrics. Elsevier, St. Louis, MO.

Brodie, J.I., Daubechies, C.D., Giannone, M.D., Loris, I., 2009. Sparse and stable Markowitz portfolios. Proceedings of the National Academy of Sciences of the United States of America 106, 12267–12272.

Candelon, B., Hurlin, C., Tokpavi, S., 2012. Sampling error and double shrinkage estimation of minimum variance portfolios. Journal of Empirical Finance 19, 511–527.

DeMiguel, V., Garlappi, L., Nogales, F.J., Uppal, R., 2009a. Generalized approach to portfolio optimization: improving performance by constraining portfolio norms. Management Science 22, 1915–1953.

DeMiguel, V., Garlappi, L., Uppal, R., 2009b. Optimal versus naive diversification: how inefficient is the $1/N$ portfolio strategy? The Review of Financial Studies 22, 1915–1953.

Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fan, J., Zhang, J., Yu, K., 2008. Asset allocation and risk assessment with gross exposure constraints for vast portfolios. The Annals of Statistics 25, 1425–1432.

Green, R., Hollifield, B., 1992. When will mean–variance efficient portfolios be well diversified? Journal of Finance 47, 1785–1809.

Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: why imposing the wrong constraints helps. Journal of Finance 58, 1651–1684.

Ledoit, O., Wolf, M., 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. Journal of Empirical Finance 10, 603–621.

Ledoit, O., Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis 88, 365–411.

Ledoit, O., Wolf, M., 2008. Robust performance hypothesis testing with the Sharpe ratio. Journal of Empirical Finance 15, 850–859.

Markowitz, H., 1952. Portfolio selection. The Journal of Finance 7, 77–91.

Merton, R.C., 1980. On estimation the expected return on the market. Journal Financial Economics 8, 323–361.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of Royal Statistical Society Series B (Statistical Methodology) 58, 267–288.

Tola, V., Lillo, F., Gallegati, M., Mantegna, R.N., 2008. Cluster analysis for portfolio optimization. Journal of Economic Dynamics 32, 235–258.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67, 301–320.

Yen, Y.M., 2010. A Note on Sparse Minimum Variance Portfolios and Coordinate-Wise Descent Algorithms. Quantitative Finance Papers 1005.5082, arXiv.org.