Published in final edited form as:
J Am Stat Assoc. 2012 ; 107(498): 592–606. doi:10.1080/01621459.2012.682825.

Vast Portfolio Selection with Gross-exposure Constraints*

Jianqing Fan,
Frederick L. Moore Professor, Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08540, USA and honorary professor, Department of Statistics, Shanghai University of Economics and Finance, Shanghai, China

Jingjin Zhang, and
McKinsey & Company, 233 Taicang Road, Shanghai, China

Ke Yu
J.P. Morgan Chase, 168 Robinson Road, Singapore

Jianqing Fan: jqfan@princeton.edu; Jingjin Zhang: jingjin.zhang@gmail.com; Ke Yu: keyu0205@gmail.com

Abstract

We introduce the large portfolio selection using gross-exposure constraints. We show that with gross-exposure constraint the empirically selected optimal portfolios based on estimated covariance matrices have similar performance to the theoretical optimal ones and there is no error accumulation effect from estimation of vast covariance matrices. This gives theoretical justification to the empirical results in Jagannathan and Ma (2003). We also show that the no-short-sale portfolio can be improved by allowing some short positions. The applications to portfolio selection, tracking, and improvements are also addressed. The utility of our new approach is illustrated by simulation and empirical studies on the 100 Fama-French industrial portfolios and the 600 stocks randomly selected from Russell 3000.

Keywords

Short-sale constraint; mean-variance efficiency; portfolio selection; risk assessment; risk optimization; portfolio improvement

1 Introduction

Portfolio selection and optimization have been a fundamental problem in finance ever since Markowitz (1952, 1959) laid down the ground-breaking work on the mean-variance analysis. Markowitz posed the mean-variance analysis by solving a quadratic optimization problem. This approach has had a profound impact on the financial economics and is a milestone of modern finance. It leads to the celebrated Capital Asset Pricing Model (CAPM), developed by Sharpe (1964), Lintner (1965) and Black (1972). However, there are documented facts that the Markowitz portfolio is very sensitive to errors in the estimates of the inputs, namely the expected return and the covariance matrix. The problem gets more severe when the portfolio size is large.

To appreciate the challenge of dimensionality, suppose that we have a pool of 2,000 candidate assets and wish to select some for investment. The covariance matrix alone

*This project was supported by the NSF grant DMS-070433 and National Institute of General Medical Sciences of NIH through Grant Number R01-GM072611. The authors are truly grateful to the editor, Xuming He, the Associate Editor and three referees for the insightful comments and constructive suggestions, which help improve significantly the manuscript.

involves over 2,000,000 unknown parameters. Yet, the sample size *n* is usually no more than 400 (about 1.5 years daily data). Now, each element in the covariance matrix is estimated with the accuracy of order $O(n^{-\frac{1}{2}})$ or 0.05. Aggregating them over millions of estimates can lead to devastating effects, resulting in adverse performance in the selected portfolio. As a result, the allocation vector that we get based on the empirical data can be very different from the allocation vector we want based on the theoretical inputs. Hence, the optimal portfolio does not perform well in empirical applications, and it is very important to find a robust portfolio that does not depend on the aggregation of estimation errors.

Several techniques have been suggested to reduce the sensitivity of the Markowitz optimal portfolios to input uncertainty. Chopra and Ziemba (1993) proposed a James-Stein estimator for the means and Ledoit and Wolf (2003) proposed to shrink the sample covariance matrix. Fan *et al.* (2008) studied the covariance matrix estimated based on the factor model and demonstrated that the resulting allocation vector significantly outperforms the allocation vector based on the sample covariance. Pesaran and Zaffaroni (2008) investigated how the optimal allocation vector depends on the covariance matrix with a factor structure when portfolio size is large. However, these techniques, while reducing the sensitivity of input vectors in the mean-variance allocation, are inadequate to address the adverse effect due to the accumulation of estimation errors, particularly when portfolio size is large.

Various efforts have been made to modify the Markowitz mean-variance optimization problem to make the resulting allocation depend less sensitively on the input vectors. De Roon *et al.* (2001) considered testing-variance spanning with the no-short-sale constraint. Goldfarb and Iyengar (2003) studied some robust portfolio optimization problems. Jagannathan and Ma (2003) imposed the no-short-sale constraint on the Markowitz mean-variance optimization problem and gave insightful explanation and demonstration of why the constraints help. However, as to be shown in this paper, the optimal no-short-sale portfolio is not diversified enough. The constraint on gross exposure needs relaxing in order to enlarge the pools of admissible portfolios. We will provide statistical insights to the question why the constraint prevents the risks or utilities of selected portfolios from accumulation of statistical estimation errors. This is a prominent contribution of this paper in addition to the utilities of our formulation in portfolio selection, tracking, and improvement. Our result provides a theoretical insight to the phenomenon, observed by Jagannathan and Ma (2003), why the wrong constraint helps on risk reduction for large portfolios.

We approach the utility optimization problem by introducing a gross-exposure constraint on the allocation vector. A sketch of the idea appeared in Fan (2007). This makes not only the Markowitz problem more practical, but also bridges the gap between the no-short-sale utility optimization problem of Jagannathan and Ma (2003) and the unconstrained utility optimization problem of Markowitz (1952, 1959). As the gross exposure parameter increases from 1 to infinity, our utility optimization progressively ranges from no-short-sales constraint to no constraint on short sales. We will demonstrate that for a wide range of the constraint parameters, the optimal portfolio does not sensitively depend on the estimation errors of the input vectors. The oracle and empirical optimal portfolios, based respectively on the true and estimated parameters, have approximately the same utility. In addition, the empirical and theoretical risks are also approximately the same for any allocation vector satisfying the gross-exposure constraint. The extent to which the gross-exposure constraint impacts on utility approximations is explicitly unveiled. These theoretical results are demonstrated by several simulation and empirical studies. They lend further support to the conclusions made by Jagannathan and Ma (2003) in their empirical studies.

Our approach has important implications in practical portfolio selection and allocation. Monitoring and managing a portfolio of many stocks is not only time consuming but also expensive. Therefore, it is ideal to pick a reasonable number of assets to mitigate these two problems. Ideally, we would like to construct a robust portfolio of a reasonably small size to reduce trading, re-balancing, monitoring, and research costs. We also want to control the gross exposure of the portfolio to avoid too extreme long and short positions. As demonstrated later, the exposure constrained optimization problem (2.1) provides a good solution to the problem.

The paper is organized as follows. Section 2 introduces the constrained utility optimization and demonstrates that the estimation error has limited impact on the utility optimization. Its applications to portfolio tracking and selection are discussed in Section 3. The proposed techniques are illustrated by simulation studies in Section 4 and by real data in Section 5. Section 6 concludes and all technical proofs and conditions are relegated to the appendix.

2 Portfolio optimization with gross-exposure constraints

Suppose we have p assets with returns $R_1, \cdots, R_p$ to be managed. Let $\mathbf{R}$ be the return vector, $\mathbf{\Sigma}$ be its associated covariance matrix, and $\mathbf{w}$ be its portfolio allocation vector, satisfying $\mathbf{w}^T \mathbf{1} = 1$. Then the variance of the portfolio return $\mathbf{w}^T \mathbf{R}$ is given by $\mathbf{w}^T \mathbf{\Sigma} \mathbf{w}$.

2.1 Constraints on gross exposure

For a given portfolio with allocation $\mathbf{w}$, the total proportions of long and short positions are

$$w^+ = (\|\mathbf{w}\|_1 + 1)/2 \quad \text{and} \quad w^- = (\|\mathbf{w}\|_1 - 1)/2,$$

respectively, since $w^+ + w^- = \|\mathbf{w}\|_1$ and $w^+ - w^- = 1$. The constraint $\|\mathbf{w}\|_1 \le c$ prevents extreme positions in the portfolio. When $c = 1$, this means that no short sales are allowed. When $c = \infty$, there is no constraint on short sales. As a generalization to the work by Markowitz (1952), Jagannathan and Ma (2003) and Fan (2007), our utility optimization problem with gross-exposure constraint is

$$\begin{array}{ll} \max_{\mathbf{w}} & E[U(\mathbf{w}^T \mathbf{R})] \\ s.t. & \mathbf{w}^T \mathbf{1} = 1, \ \|\mathbf{w}\|_1 \le c, \ \mathbf{A} \ \mathbf{w} = \mathbf{a}, \end{array} \qquad (2.1)$$

The utility function $U(\cdot)$ can also be replaced by any risk measures such as those in Artzner *et al.* (1999), and in this case the utility maximization should be risk minimization.

The extra constraints $\mathbf{A}\mathbf{w} = \mathbf{a}$ are related to the constraints on percentage of allocations on each sector or industry. It can also be the constraint on the expected return of the portfolio or factor exposures. For example, the portfolios can be constrained without exposure (market-neutral) to the market risk such as the returns of SP500. The problem (2.1) has independently investigated in several fields from different angles. See Fan (1997), Fan, Zhang and Yu (2008), which is an earlier draft of this paper, DeMiguel *et al.* (2008), and Bordie *et al.* (2009).

2.2 Utility and risk approximations

It is well known that when the return vector $\mathbf{R} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma})$ and $U(x) = 1 - \exp(-Bx)$, with B being the absolute risk aversion parameter, the utility optimization is equivalent to maximizing the Markowitz mean-variance function:

$$M(\mu, \sum) = \mathbf{w}^T \mu - \lambda \mathbf{w}^T \sum \mathbf{w}, \quad (2.2)$$

where $\lambda = B/2$. The solution to (2.2) is $\mathbf{w}^{\text{opt}} = c_1\Sigma^{-1}\boldsymbol{\mu} + c_2\Sigma^{-1}\mathbf{1}$ with c_1 and c_2 depending on $\boldsymbol{\mu}$ and Σ as well. The solution depends sensitively on the input vectors $\boldsymbol{\mu}$ and Σ, and their accumulated estimation errors. The problem can result in extreme positions, which make it impractical.

These two problems disappear when the gross-exposure constraint $\|\mathbf{w}\|_1 \le c$ is imposed for a moderate c. The sensitivity of utility function to the estimation errors can easily be bounded as follows:

$$|M(\widehat{\mu}, \widehat{\sum}) - M(\mu, \sum)| \le \|\widehat{\mu} - \mu\|_\infty \|\mathbf{w}\|_1 + \lambda \left\|\widehat{\sum} - \sum\right\|_\infty \|\mathbf{w}\|_1^2, \quad (2.3)$$

where $\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_\infty$ and $\|\hat{\Sigma} - \Sigma\|_\infty$ are the maximum componentwise estimation errors. Therefore, as long as each element is estimated well, the overall utility is approximated well without accumulation of estimation errors. The story is very different in the case that there is no constraint on the short-sale in which $c = \infty$. In this case, the estimation error does accumulate and they are negligible only for a portfolio with a moderate size, as demonstrated in Fan *et al.* (2008).

Specifically, if we consider the risk minimization with no short-sale constraint, then analogously to (2.3), we have

$$|R(\mathbf{w}, \widehat{\sum}) - R(\mathbf{w}, \sum)| \le \left\|\widehat{\sum} - \sum\right\|_\infty \|\mathbf{w}\|_1^2, \quad (2.4)$$

where the risk is defined by $R(\mathbf{w}, \Sigma) = \mathbf{w}^T\Sigma\mathbf{w}$. The right hand of (2.4) obtains its minimum when $\|\mathbf{w}\|_1 = 1$, the no-short-sale portfolio. In this case,

$$|R(\mathbf{w}, \widehat{\sum}) - R(\mathbf{w}, \sum)| \le \left\|\widehat{\sum} - \sum\right\|_\infty. \quad (2.5)$$

The inequality (2.5) is the mathematics behind the conclusions in Jagannathan and Ma (2003). In particular, we see easily that estimation errors from (2.5) do not accumulate in the risk. Even when the constraint is wrong (excluding the optimal portfolio), we lose somewhat the theoretical optimal risk due to the limited space of portfolios, yet we gain substantially the reduction of the error accumulation of statistical estimation. As a result, the actual risks of the empirical optimal portfolios selected based on wrong constraints can outperform the Markowitz portfolio.

2.3 Risk optimization: some theory

As it is very hard to estimate accurately the expected returns $\boldsymbol{\mu}$, the focus is shifted to the risk minimization in empirical finance. From now on, we consider the risk minimization problem:

$$\min_{\mathbf{w}^T\mathbf{1}=1,\ \|\mathbf{w}\|_{1\le c}} \mathbf{w}^T \sum \mathbf{w}. \quad (2.6)$$

This is a simple quadratic programming problem and can be computed easily for each given c. The problem with linear constraints can be solved similarly.

To simplify the notation, we let

$$R(\mathbf{w})=\mathbf{w}^T\sum\mathbf{w},\quad R_n(\mathbf{w})=\mathbf{w}^T\widehat{\sum}\mathbf{w},\quad (2.7)$$

be respectively the theoretical and empirical portfolio risks with allocation $\mathbf{w}$, where $\hat{\Sigma}$ is an estimated covariance matrix based on the data with sample size n. Let

$$\mathbf{w}_{\mathrm{opt}}=\mathrm{argmin}_{\mathbf{w}^T\mathbf{1}=1,\ \|\mathbf{w}\|_1\le c} R(\mathbf{w}),\quad \widehat{\mathbf{w}}_{\mathrm{opt}}=\mathrm{argmin}_{\mathbf{w}^T\mathbf{1}=1,\ \|\mathbf{w}\|_1\le c} R_n(\mathbf{w})\quad (2.8)$$

be respectively the theoretical optimal allocation vector we want and empirical optimal allocation vector we get.

Theorem 1—Let $a_n = \|\hat{\Sigma} - \Sigma\|_\infty$ Then, we have (without any conditions)

$$\begin{aligned}|R(\mathbf{w}_{\mathrm{opt}})-R_n(\widehat{\mathbf{w}}_{\mathrm{opt}})| &\le a_n c^2\\ |R(\widehat{\mathbf{w}}_{\mathrm{opt}})-R_n(\widehat{\mathbf{w}}_{\mathrm{opt}})| &\le a_n c^2\\ |R(\widehat{\mathbf{w}}_{\mathrm{opt}})-R(\mathbf{w}_{\mathrm{opt}})| &\le 2a_n c^2,\end{aligned}$$

and $E\{R_n(\hat{\mathbf{w}}_{\mathrm{opt}})\} \le R(\mathbf{w}_{\mathrm{opt}}) \le R(\hat{\mathbf{w}}_{\mathrm{opt}})$.

Theorem 1 shows the theoretical minimum risk $R(\mathbf{w}_{\mathrm{opt}})$ (also called the oracle risk) and the actual risk $R(\hat{\mathbf{w}}_{\mathrm{opt}})$ of the invested portfolio are approximately the same as long as the c is not too large and the accuracy of estimated covariance matrix is not too poor. Both of these risks are unknown. The empirical minimum risk $R_n(\hat{\mathbf{w}}_{\mathrm{opt}})$ is known, and can be overly optimistic (too small). But, it is close to both the theoretical risk and the actual risk. The results hold without any conditions on $\hat{\Sigma}$. In particular, elementwise estimation of covariance matrix is allowed. The concept of risk approximation is similar to persistent(Greenshtein and Ritov, 2004).

In Theorem 1, we do not specify the rate a_n. This depends on the model assumption and method of estimation. For example, one can use the factor model to estimate the covariance matrix as in Jagannathan and Ma (2003), Ledoit and Wolf (2003), and Fan *et al.* (2008). One can also estimate the covariance via the dynamic equi-correlation model of Engle and Kelly (2011) or more generally dynamically equi-factor loading models. One can also aggregate the large covariance matrix estimation based on the high frequency data (Barndorff-Nielsen and Shephard, 2002; Zhang, *et al.*, 2005; Barndorff-Nielsen *et al.*, 2011).

To understand the impact of the portfolio size p on the accuracy a_n, let us consider the sample covariance matrix $\mathbf{S}_n$ based on a sample $\{\mathbf{R}_t\}_{t=1}^n$ over n periods. We assume herewith that p is large relative to sample size to reflect the size of the portfolio, i.e., $p = p_n \to \infty$. When p is fixed, the following results hold trivially.

Theorem 2—Under Condition 1 in the Appendix, we have

$$\left\|\widehat{\mathbf{S}}_n - \sum\right\|_\infty = O_p\left(\sqrt{\frac{\log p}{n}}\right).$$

This theorem shows that the portfolio size enters into the maximum estimation error only at the logarithmic order. Hence, the portfolio size does not play a significant role in risk minimization as long as the constraint on gross exposure is in place.

In general, the uniform convergence result in Theorem 2 typically holds as long as the estimator of each element of the covariance matrix is root-n consistent with exponential tails.

Theorem 3—Let σ_{ij} and $\hat{\sigma}_{ij}$ be the (i, j)th element of the matrices Σ and $\hat{\Sigma}$, respectively. If for a sufficiently large x,

$$\max_{i,j} P\{\sqrt{n}|\sigma_{ij} - \widehat{\sigma}_{ij}| > x\} < \exp(-Cx^{1/a}),$$

for two positive constants a and C, then

$$\left\|\sum - \widehat{\sum}\right\|_\infty = O_P\left(\frac{(\log p)^a}{\sqrt{n}}\right). \quad (2.9)$$

In addition, if Condition 2 in Appendix holds, then (2.9) holds for sample covariance matrix, and if Condition 3 holds, then (2.9) holds for $a = 1/2$.

2.4 Relation with Covariance Regularization

By the Lagrange multiplier method, problem (2.6) is to minimize

$$\mathbf{w}^T \widehat{\sum} \mathbf{w}/2 + \lambda_1(\|\mathbf{w}\|_1 - c) + \lambda_2(1 - \mathbf{w}^T \mathbf{1}).$$

Let $\mathbf{g}$ be the subgradient vector of the function $\|\mathbf{w}\|_1$, whose i^{th} element being − 1, 1 or any values in [− 1, 1] depending on whether w_i is positive, negative or zero. Then, the Karush-Kuhn-Tucker conditions for the constrained optimization (2.6) are

$$\widehat{\sum}\mathbf{w} + \lambda_1\mathbf{g} - \lambda_2\mathbf{1} = 0, \quad (2.10)$$

$$\lambda_1(c - \|\mathbf{w}\|_1) = 0, \quad \lambda_1 \geq 0, \quad (2.11)$$

in addition to the constraints $\mathbf{w}^T\mathbf{1} = 1$ and $\|\mathbf{w}\|_1 \leq c$. Let $\tilde{\mathbf{w}}$ be the solution to (2.10) and (2.11).

Theorem 4—The constrained portfolio optimization (2.6) is equivalent to the mean-variance problem

$$\min_{\mathbf{w}^T\mathbf{1}=1} \mathbf{w}^T \tilde{\sum}_c \mathbf{w}, \quad (2.12)$$

with the regularized covariance matrix

$$\tilde{\sum}_c = \widehat{\sum} + \lambda_1(\tilde{\mathbf{g}}\mathbf{1}^T + \mathbf{1}\tilde{\mathbf{g}}^T), \quad (2.13)$$

where $\tilde{\mathbf{g}}$ is the subgradient evaluated at $\tilde{\mathbf{w}}$ and &*lambda;*$_1$ is the Language multiplier defined by (2.10) and (2.11).

The result is of a similar spirit of Jagannathan and Ma (2003) and DeMiguel et al. (2008).

2.5 Choice of gross exposure parameter

The gross exposure parameter c is typically given by investors. As referees pointed out, a data-driven choice will be helpful for investors. Let m be the number of data used in the testing period. A natural estimate of the risk profile with gross exposure c is to use the first $n - m$ data points to get a sample covariance matrix and hence the allocation vectors $\{\hat{\mathbf{w}}_c\}$. Then, compute the risk profile

$$R(c; n-m, m) = m^{-1} \sum_{t=n-m+1}^{n} (\widehat{\mathbf{w}}_c^T \mathbf{R}_t)^2, \quad (2.14)$$

in which the learning period up to time $n - m$ and the testing period of length m are stressed. The function $R(c; n - m, m)$ is useful for investors to choose c. The optimal data-driven choice is the one that minimizes the function.

In practice, one is not willing to use a large m. This leaves the risk evaluation (2.14) unstable. To increase the reliability of risk evaluations, we can use

$$R(c) = K^{-1} \sum_{j=1}^{k} R(c; n - jm, m). \quad (2.15)$$

This increases the testing period by a factor of K times and was advocated in Fan and Yao (2003). For example, if $K = 4$ and $m = n/10$, 40% of data are used in the evaluation of the risks in (2.15), while only 10% are used in (2.14).

3 Portfolio tracking and asset selection

The risk minimization problem (2.6) has important applications in portfolio tracking and asset selection. It also allows one to improve the utility of existing portfolios. We first illustrate its connection to a penalized least-squares problem, upon which the whole solution path can easily be found (Efron, *et al.* 2004) and then outline its applications in finance.

3.1 Connection with regression problem

Markowitz's risk minimization problem can be recast as a regression problem. By using the fact that the sum of total weights is one, we have

$$
\begin{aligned}
\mathrm{var}(\mathbf{w}^T\mathbf{R}) &= \min_b E(\mathbf{w}^T\mathbf{R}-b)^2 \\
&= \min_b E(Y-w_1X_1-\cdots-w_{p-1}X_{p-1}-b)^2,
\end{aligned} \quad (3.1)
$$

where $Y = R_p$ and $X_j = R_p - R_j$ ($j = 1, \cdots, p - 1$). Finding the optimal weight $\mathbf{w}$ is the same as finding the regression coefficient $\mathbf{w}^* = (w_1, \cdots, w_{p-1})^T$.

The gross-exposure constraint $\|\mathbf{w}\|_1 \le c$ can now be expressed as $\|\mathbf{w}^*\|_1 \le c - |1 - \mathbf{1}^T\mathbf{w}^*|$. The latter cannot be expressed as

$$
\left\|\mathbf{w}^*\right\|_1 \le d, \quad (3.2)
$$

for a given constant d. Thus, problem (2.6) is similar to

$$
\min_{b, \left\|\mathbf{w}^*\right\|_1 \le d} E(Y-\mathbf{w}^{*T}\mathbf{X}-b)^2, \quad (3.3)
$$

where $\mathbf{X} = (X_1, \cdots, X_{p-1})^T$. But, they are not equivalent. The latter depends on the choice of asset Y, while the former does not.

Efron *et al.* (2004) developed an efficient algorithm by using the least-angle regression (LARS), called the LARS-LASSO algorithm, to efficiently find the whole solution path $\mathbf{w}^*_{\mathrm{opt}}(d)$, for all $d \ge 0$, to (3.3). The number of non-vanishing weights varies as d ranges from 0 to ∞. It recruits successively more assets and gradually all assets.

3.2 Portfolio tracking and asset selection

If the variable Y is the portfolio to be tracked, problem (3.3) can be interpreted as finding a limited number of stocks with a gross-exposure constraint to minimize the expected tracking error. As d increases, the number of selected stocks increases, the tracking error decreases, but the short percentage increases. With the LARS-LASSO algorithm, we can plot the expected tracking error and the number of selected stocks, against d. This enables us to make an optimal decision on how many stocks to pick to manage the trade-off among the expected tracking errors, the number of selected stocks and short positions.

3.3 An approximate solution path to risk minimization

The solution path to (3.3) also provides a nearly optimal solution path to the problem (2.6). For example, the allocation with $-\mathbf{w}^*_{\mathrm{opt}}(d)$ on the first $p - 1$ stocks and the rest on the last stock is a feasible allocation vector to the problem (2.6) with

$$
c = d + |1 - \mathbf{1}^T\mathbf{w}^*_{\mathrm{opt}}(d)|. \quad (3.4)
$$

This will not be the optimal solution to the problem (2.6) as it depends on the choice of Y. However, when Y is properly chosen, the solution is nearly optimal, as to be demonstrated. For example, by taking Y to be the no-short-sale portfolio, then problem (3.3) with $d = 0$ is the same as the solution to problem (2.6) with $c = 1$. We can then use (3.3) to provide a nearly optimal solution.

In summary, to compute (2.6) for all c, we first find the solution with $c = 1$ using a quadratic programming. This yields the optimal no-short-sale portfolio. We then take this portfolio as Y in problem (3.3) and apply the LARS-LASSO algorithm to obtain the solution path $-\mathbf{w}^*_{\text{opt}}(d)$. Finally, compute the gross-exposure of the portfolio with $\mathbf{w}^*_{\text{opt}}(d)$ on the first $p - 1$ stocks and the rest on the optimal no-short-sale portfolio, called it c, namely, regard the aforementioned portfolio as an approximate solution to the problem (2.6). This yields the whole solution path to the problem (2.6). As shown in Figure 1(a) below and the empirical studies, the approximation is indeed quite accurate.

In the above algorithm, one can also take a tradable index or an ETF (electronically traded fund) in the set of stocks under consideration as Y and applies the same technique to obtain a nearly optimal solution. We have experimented this and obtained good approximations, too.

4 Simulation studies

In this section, we use simulations to verify our theoretical results and to quantify the finite sample behaviors. In particular, we would like to demonstrate that the risk profile of the optimal no-short-sale portfolio can be improved substantially and that the LARS algorithm yields a good approximate solution to the risk minimization with gross-exposure constraint. In addition, we would like to demonstrate that when the covariance matrix is estimated reasonably accurately, the risk that we want (the oracle risk, $\sqrt{R(\mathbf{w}_{\text{opt}})}$) and the risk that we get (the actual risk of the empirical optimal portfolio, $\sqrt{R(\widehat{\mathbf{w}}_{\text{opt}})}$) are approximately the same for a wide range of the exposure coefficient. However, the story is very different when the constraint is loose. All the simulation studies are out-of-sample studies.

4.1 A simulated Fama-French three-factor model

Let R_i be the excessive return over the risk free interest rate of the i^{th} stock. Fama and French (1993) identified three key factors that capture the cross-sectional risk in the US equity market. The first factor is the excess return of the proxy of the market portfolio and the other two factors are related to the market capitalizations and book-to-market ratios. More specifically, we assume that the excess return follows the following three-factor model:

$$R_i=b_{i1}\, f_1+b_{i2}\, f_2+b_{i3}\, f_3+\varepsilon_i, \quad i=1,\cdots,p, \quad (4.1)$$

where $\{b_{ij}\}$ are the factor loadings of the i^{th} stock on the factor f_j, and ε_i is the idiosyncratic noise, independent of the three factors. We assume further that the idiosyncratic noises are independent of each other, whose marginal distributions are the Student-t with degree of freedom 6 and standard deviation σ_i.

To facilitate the presentation, we write the factor model (4.1) in the matrix form:

$$\mathbf{R}=\mathbf{B}\mathbf{f}+\boldsymbol{\varepsilon}, \quad (4.2)$$

where $\mathbf{B}$ is the factor loading coefficient matrix. Throughout this simulation, we assume that $E(\boldsymbol{\varepsilon}|\mathbf{f}) = \mathbf{0}$ and $\text{cov}(\boldsymbol{\varepsilon}|\mathbf{f})=\text{diag}(\sigma_1^2,\cdots,\sigma_p^2)$. Then, model (4.2) implies that $\mathbf{R}$ has the covariance

$$\sum = \mathrm{cov}(\mathbf{Bf}) + \mathrm{cov}(\boldsymbol{\varepsilon}) = \mathbf{B}\mathrm{cov}(\mathbf{f})\mathbf{B}^T + \mathrm{diag}(\sigma_1^2, \cdots, \sigma_p^2). \quad (4.3)$$

We simulate the n-period returns of p assets as follows. See Fan *et al.* (2008) for additional details. First of all, the factor loadings are generated from the trivariate normal distribution $N(\boldsymbol{\mu}_b, \mathbf{cov}_b)$, where the parameters are given in Table 1. Once the factor loadings are generated, they are taken as the parameters and thus kept fixed throughout simulations. The levels of idiosyncratic noises are generated from a gamma distribution with the shape parameter 3.3586 and the scale parameter 0.1876, conditioned on the noise level of at least 0.1950. The returns of the three factors over n periods are drawn from the trivariate normal distribution $N(\boldsymbol{\mu}_f, \mathbf{cov}_f)$, with the parameters given in Table 1. The parameters are calibrated to the market data of 30 industrial portfolios from May 1, 2002 to August 29, 2005 and are taken from Fan *et al.* (2008). Thus, our numerical experiments simulate the portfolio returns, which are less volatile than individual stocks. They allow a larger choice of c.

4.2 Accuracy of LARS approximation

As mentioned in §3.3, the LARS algorithm provides an approximate solution. The first question is then the accuracy of the approximation. As a byproduct, we also demonstrate that the optimal no-short-sale portfolio is not diversified enough and can be significantly improved.

To demonstrate this, we took 100 stocks with covariance matrix given by (4.3). For each given c in the interval [1, 3], we applied a quadratic programming algorithm to solve problem (2.6) and obtained its associated minimum portfolio risk, depicted in Figure 1(a). We also employed the LARS algorithm using the optimal no-short-sale portfolio as Y, with d ranging from 0 to 3. This yields a solution path along with its associated portfolio risk path. The risk profiles are in fact very close. The number of stocks for the optimal no-short-sale portfolio is 9. As c increases, the number of stocks picked by (2.6) also increases, as demonstrated in Figure 1(b) and the portfolio gets more diversified.

The approximated and exact solutions have very similar risk functions. Figure 1(a) shows that the optimal no-short-sale portfolio is very conservative and can be improved drastically as the constraint relaxes. At $c = 2$ (corresponding to 18 stocks with 50% short positions and 150% long positions), the risk decreases from 8.1% to 4.9%. The decrease of risks slows down drastically after that point. This shows that the optimal no-short-sale constraint portfolio can be improved substantially by using our method.

The next question is whether the improvement can be realized with the estimated covariance matrix. To illustrate this, we simulated $n = 252$ from the three-factor model (4.1) and estimated the covariance matrix by the sample covariance matrix. The actual and empirical risks of the selected portfolio for a typical simulated data set are depicted in Figure 1(c). For a range up to $c = 1.7$, they are approximately the same. The range widens when the covariance matrix is estimated with a better accuracy. To demonstrate this effect, we show in Figure 1(d) the case with sample size $n = 756$. However, when the gross exposure parameter is large, they can differ substantially.

4.3 Portfolio improvement

To demonstrate further how much our method can be used to improve the existing portfolio, we assume that the current portfolio is an equally weighted portfolio among 200 stocks. This is the portfolio Y. The returns of these 200 stocks are simulated from model (4.1) over a period of $n = 252$. The theoretical risk of this equally weighted portfolio is 13.58%, while the empirical risk of this portfolio is 13.50% for a typical realization. Here, the typical sample refers to the one that has the median value of the empirical risks among 200 simulations. This particular simulated data set is used for the further analysis.

We now pretend that this equally weighted portfolio is the one that an investor holds and the investor seeks possible improvement of the efficiency by modifying some of the weights. The investor employs the LARS-LASSO technique (3.3), taking the equally weighted portfolio as Y and the 200 stocks as potential X. Figure 2 depicts the empirical and actual risks, and the number of stocks whose weights are modified in order to improve the risk profile of equally weighted portfolio.

The risk profile of the equally weighted portfolio can be improved substantially. When the sample covariance is used, at $d = 1$, Figure 2(a) reveals the empirical risk is only about one half of the equally weighted portfolio, while Figure 2(b) shows that the number of stocks whose weights have been modified is only 4. The total percentage of short positions is only about 48%. The actual risk of this portfolio is very close to the empirical one, giving an actual risk reduction of nearly 50%. At $d = 2$, corresponding to about 130% of short positions, the empirical risk is reduced by a factor of about 3. Increasing the gross exposure parameter only slightly reduces the empirical risk, but quickly increases the actual risk. Applying our criterion to the empirical risk, which is known at the time of decision making, one would have chosen a gross exposure parameter somewhat less than 1.5 (the place where the empirical fails to decrease quickly), realizing a sizable reduction of the actual risk.

Similar conclusions can be made for the covariance matrix based on the factor model. In this case, the covariance matrix is estimated more accurately and hence the empirical and actual risks are closer for a wider range of the gross exposure parameter d. This is consistent with our theory. The substantial gain in this case is due to the fact that the factor model is correct and hence incurs no modeling biases in estimating covariance matrices. For the real financial data, however, the accuracy of the factor model is unknown. As soon as $d \geq 3$ the empirical reduction of risk is not significant.

4.4 Risk approximations

We now use simulations to demonstrate the closeness of the risk approximations with the gross-exposure constraints. The factor model (4.1) is used to generate the returns of p stocks over a period of $n = 252$ days. The number of simulations is 101. The covariance matrix is estimated by either the sample covariance or the factor model (4.3) whose coefficients are estimated from the sample. We examined two cases: $p = 200$ and $p = 500$. The first case corresponds to a non-degenerate sample covariance matrix whereas the second case corresponds to a degenerate one.

We first examine the case $p = 200$ with a sample of size 252. Figure 3(a) summarizes the 10th, 50th and 90th percentiles of the actual risks of empirically selected portfolios among 101 simulations. They summarize the distributions of these actual risks. The sampling variation is indeed small. It is clear from Figure 3(c) that the theoretical risk fails to decrease noticeably when $c = 3$ and increasing the gross-exposure will not improve very much the theoretical optimal risk profile. In fact, for the true covariance, the global minimum variance portfolio has $c = 4.22$, which involves 161% of short positions, and minimum risk 2.68%. For any $c \geq 4.22$, the theoretical risk is constant. On the other hand, increasing gross exposure c makes it harder to estimate theoretical allocation vector. As a result, the actual risk increases when c gets larger. When c gets larger (beyond the scale plotted here), the actual risk increases steadily, while empirical risk decreases.

Combining the results in both top and bottom panels, Figure 3 gives a comprehensive overview of the risk approximations. The top panel shows the sampling variability of the actual risks, whereas the bottom panel depicts the approximation errors of the actual risks (biases). For example, when c is small, both approximation errors and sampling variabilities are small, whereas the approximation errors dominate the sampling variability when the global exposure parameter is large. It is clear that the risk approximations are much more accurate for the covariance matrix estimation based on the factor model. This is somewhat expected as the data generating process is a factor model.

We now consider the case where there are 500 potential stocks with only a year of data ($n = 252$). In this case, the sample covariance matrix is always degenerate. Therefore, the global minimum portfolio based on the sample covariance always has zero empirical risk. On the other hand, with the gross-exposure constraint, the actual and empirical risks approximate quite well for a wide range of gross exposure parameters (Figure 4). To gauge the relative scale of the range, we note that for the given theoretical covariance, the global minimum portfolio has $c = 4.01$, which involves 150% of short positions with the minimal risk 1.69%. The optimal no-short-sale portfolio, selected from 500 stocks, has actual risk 6.47%, which is not much smaller than 7.35% selected from 200 stocks.

5 Empirical Studies

5.1 Fama-French 100 Portfolios

We use the daily returns of 100 industrial portfolios formed on size and book to market ratio from the website of Kenneth French from Jan 2, 1998 to December 31, 2007. These 100 portfolios are formed by the two-way sort of the stocks in the CRSP database, according to the market equity and the ratio of book equity to market equity, 10 categories in each variable. At the end of each month from 1998 to 2007, the covariance matrix of the 100 assets is estimated according to three estimators, the sample covariance, Fama-French 3-factor model, and the RiskMetrics with $\lambda = 0.97$, using the past 12 months' daily return data. These covariance matrices are then used to construct optimal portfolios under various exposure constraints. The portfolios are then held for one month and rebalanced at the beginning of the next month. The means, standard deviations and other characteristics of these portfolios are recorded and presented in Table 2. They represent the actual returns and actual risks. The optimal portfolios with the gross-exposure constraints pick certain numbers of assets each month. The average numbers of assets over the study period are presented in Table 2.

First of all, the optimal no-short-sale portfolios, while selecting about 6 assets from 100 portfolios, are not diversified enough. Their risks can easily be improved by relaxing the gross-exposure constraint with $c = 2$. This is shown in Table 2 and Figure 5(a), no matter which method is used to estimate the covariance matrix. The risk stops decreasing drastically when $c = 2.5$. Interestingly, the Sharpe ratios peak around $c = 2.5$ stocks too. After that point, the Sharpe Ratio actually falls for the covariance estimation based on the sample covariance and the factor model.

For $c < 2.5$, the portfolios selected using the RiskMetrics have lower risks. In comparison with the sample covariance matrix, the RiskMetrics estimates the covariance matrix using a much smaller effective time window. As a result, the biases are usually smaller than the sample covariance matrix. Since each asset is a portfolio in this study, its risk is smaller than stocks. Hence, the covariance matrix can be estimated more accurately with the RiskMetrics in this study. This explains why the resulting selected portfolios by using RiskMetrics have smaller risks. However, their associated returns tend to be smaller too. As a result, their Sharpe ratios are actually smaller. The Sharpe ratios actually peak at around $c = 1.5$ assets.

It is surprising to see that the unmanaged equally weighted portfolio, which invests 1 percent on each of the 100 industrial portfolios, is far from optimal in terms of the risk during the study period. The value-weighted index CRSP does not fare much better. They are all outperformed by the optimal portfolios with gross-exposure constraints during the study period. This is in line with our theory. Indeed, the equally weighted portfolio and CRSP index are two specific members of the no-short-sale portfolio, and should be outperformed by the optimal no-short-sale portfolio, if the covariance matrix is estimated with reasonable accuracy.

5.2 Russell 3000 Stocks

We now apply our techniques to study the portfolio behavior using Russell 3000 stocks. The study period is from January 2, 2003 to December 31, 2007. To avoid computation burden and the issues of missing data, we picked 600 stocks randomly from 1000 stocks in the 3000 stocks that constitute Russell 3000 on December 31, 2007. Those 1000 stocks have least percents of missing data in the five-year study period. This forms the universe of the stocks under our study.

At the end of each month from 2003 to 2007, the covariance of the 600 stocks is estimated according to various estimators using the past 24 months' daily returns. As a result, the sample covariance matrix is degenerate. We use these covariance matrices to construct optimal portfolios under various gross-exposure constraints and hold these portfolios for one month. The daily returns of these portfolios are recorded and hence the standard deviations are computed.

Table 3 summarizes the risks of the optimal portfolios constructed using 3 different methods of estimating covariance matrix and using 6 different gross-exposure constraints. The global minimum portfolio, which does not exist empirically but is approximated by $c = 8$, is not efficient for vast portfolios due to accumulation of errors in the estimated covariance matrix. This can be seen easily from Figure 6. The ex-post annualized volatilities of constructed portfolios using the sample covariance and RiskMetrics shoot up quickly (beyond $c = 2$). The risk continues to grow if we relax further the gross-exposure constraint, which is beyond the range of our pictures. This provides further evidence to support the claim of Jagannathan and Ma (2003).

The optimal no-short-sale portfolios are not efficient in terms of ex-post risk calculation. They can be improved, when portfolios are allowed to have 50% short positions, say, corresponding to $c = 2$. This is due to the fact that the no-short-sale portfolios are not diversified enough. The risk approximations are still accurate when $c \leq 2$. On the other hand, the optimal no-short-sale portfolios outperform substantially the global minimum portfolio (proxied by $c = 8$), which is consistent with the conclusion drawn in Jagannathan and Ma (2003) and with our risk approximation theory.

The risks of optimal portfolios tend to be smaller and stable, when the covariance matrix is estimated from the factor model. For vast portfolios, such an estimation of covariance matrix tends to be most stable among the three methods that we considered here. As a result, its associated portfolio risks tend to be the smallest among the three methods. As the covariance matrix estimated by RiskMetrics uses a shorter time window than that based on the sample covariance matrix, the resulting estimates tend to be even more unstable. As a result, its associated optimal portfolios tend to have the highest risks.

6 Conclusion

The portfolio optimization with the gross-exposure constraint bridges the gap between the optimal no-short-sale portfolio and no constraint on short-sale in the Markowitz's framework. The gross-exposure constraint helps control the discrepancies between the empirical risk which can be overly optimistic, oracle risk which is not obtainable, and the actual risk of the selected portfolio which is unknown. We demonstrate that for a range of gross exposure parameters, these three risks are actually very close. The approximation errors are controlled by the worst elementwise estimation error of the covariance matrix. There is no accumulation of estimation errors.

We provided theoretical insights into the observation made by Jagannathan and Ma (2003) that the optimal no-short-sale portfolio has smaller actual risk than that for the global minimum portfolio and offered empirical evidence to strengthen the conclusion. We demonstrated that the optimal no-short-sale portfolio is not diversified enough. It is still a conservative portfolio that can be improved by allowing some short positions. This is demonstrated by our empirical studies and supported by our risk approximation theory: Increasing gross exposure somewhat does not excessively increase the risk approximation errors, but increases significantly the space of allowable portfolios and hence decreases drastically the oracle and actual risks.

Practical portfolio choices always involve constraints on individual assets. This is commonly understood as an effort of reducing the risks of the selected portfolios. Our theoretical result provides further mathematical insights to support such a statement. The constraints on individual assets also put a constraint on the gross exposure and hence control the risk approximation errors, which makes the empirical risk and actual risk closer.

References

1. Artzner P, Delbaen F, Eber J, Heath D. Coherent measures of risk. Mathematical Finance. 1999; 9:203–228.
2. Barndorff-Nielsen O, Hansen P, Lunde A, Shephard N. Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading, Journal of Econometrics. Journal of Econometrics. 2011; 162:149–169.
3. Barndorff-Neilsen OE, Shephard N. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal Royal Statistics Society B. 2002; 64:253–280.
4. Black F. Capital market equilibrium with restricted borrowing. Journal of Business. 1972; 45:444–454.
5. Brodie J, Daubechies I, De Mol C, Giannoned D, Loris I. Sparse and stable Markowitz portfolios. Proc Natl Acad Sci USA. 2009; 106:12267–12272. [PubMed: 19617537]
6. Bosq, D. Nonparametric Statistics for Stochastic Processes: Estimation and Prediction. 2. Springer-Verlag; Berlin: 1998. p. 110Lecture Notes in Statistics
7. Chopra VK, Ziemba WT. The effect of errors in means, variance and covariances on optimal portfolio choice. Journal of Portfolio Management. 1993 winter;:6–11.
8. DeMiguel V, Garlappi L, Nogales FJ, Uppal R. A generalized approach to portofolio optimization: Improving performance by constraining portfolio norms. Management Science. 2008; 55:798–812.
9. De Roon FA, Nijman TE, Werker BJM. Testing for mean-variance spanning with short sales constraints and transaction costs: The case of emerging markets. Journal of Finance. 2001; 54:721–741.
10. Doukhan P, Neumann MH. Probability and moment inequalities for sums of weakly dependent random variables, with applications. Stochastic Processes and their Applications. 2007; 117:878–903.
11. Dumbgen L, van der Geer SA, Veraar MC, Wellner JA. Nemirovskis Inequalities Revisited. American Mathematical Monthly. 2010; 117:138–160. [PubMed: 20224805]

12. Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression (with discussions). The Annals of Statistics. 2004; 32:409–499.
13. Emery, M.; Nemirovski, A.; Voiculescu, D. Lectures on Probability Theory and Statistics. Springer; 2000. p. 106-107.
14. Engle, RF.; Kelly, BT. Journal of Business and Economic Statistics. 2011. Dynamic equicorrelation. to appear
15. Engle RF, Shephard N, Shepphard E. Fitting and testing vast dimensional time-varying covariance models. 2008 Manuscript.
16. Fama E, French K. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics. 1993; 33:3–56.
17. Fan, J. Variable screening in high-dimensional feature space. In: Ji, L.; Liu, K.; Yang, L.; Yau, ST., editors. Proceedings of the 4th International Congress of Chinese Mathematicians. Vol. II. 2007. p. 735-747.
18. Fan J, Fan Y, Lv J. Large dimensional covariance matrix estimation via a factor model. Journal of Econometrics. 2008; 147:186–197.
19. Fan, J.; Yao, Q. Nonlinear Time Series: Nonparametric and Parametric Methods. Springerverlag; New York: 2003.
20. Fan, J.; Zhang, J.; Yu, K. Asset allocation and risk sssessment with gross exposure constraints for vast portfolios. 2008. www.ssrn.com
21. Goldfarb D, Iyengar G. Robust portfolio selection problems. Mathematics of Operations Research. 2003; 28:1–38.
22. Greenshtein E, Ritov Y. Persistence in high-dimensional predictor selection and the virtue of overparametrization. Bernoulli. 2004; 10:971–988.
23. Jagannathan R, Ma T. Risk reduction in large portfolios: Why imposing the wrong constraints helps. Journal of Finance. 2003; 58:1651–1683.
24. Ledoit O, Wolf M. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. Journal of Empirical Finance. 2003; 10:603–621.
25. Lintner J. The valuation of risky assets and the selectin of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics. 1965; 47:13–37.
26. Markowitz HM. Portfolio selection. Journal of Finance. 1952; 7:77–91.
27. Markowitz, H. Portfolio Selection: Efficient Diversification of Investments. John Wiley & Sons; New York: 1959.
28. Neumann MH, Paparoditis E. Goodness-of-fit tests for Markovian time series models: Central limit theory and bootstrap applcations. Bernoulli. 2008; 14:14–46.
29. Pesaran MH, Zaffaroni P. Optimal asset allocation with factor models for large portfolios. 2008 Manuscript.
30. Sharpe W. Capital asset prices: A theory of market equilibrium under conditons of risks. Journal of Finance. 1964; 19:425–442.
31. Zhang L, Mykland PA, Aït-Sahalia Y. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association. 2005; 100:1394–1411.

Appendix: Conditions and Proofs

Throughout this appendix, we will assume that $\boldsymbol{\mu} = E\mathbf{R}_t$ and $\mathbf{S}=E(\mathbf{R}_t\mathbf{R}_t^T)$ are independent of t. Let $\mathcal{F}_t$ be the filtration generated by the process $\{\mathbf{R}_t\}$.

Condition 1

Let $\mathbf{Y}_t$ be the $p(p+1)/2$-dimensional vector constructed from the symmetric matrix $\mathbf{R}_t\mathbf{R}_t^T - \mathbf{S}$. Assume that $\boldsymbol{\mu} = 0$ and $\mathbf{Y}_t$ follows the vector autoregressive model:

$$\boldsymbol{Y}_t = \mathbf{A}_1 \boldsymbol{Y}_{t-1} + \cdots + \mathbf{A}_k \boldsymbol{Y}_{t-k} + \boldsymbol{\varepsilon}_t,$$

for coefficient matrices $\mathbf{A}_1, \cdots, \mathbf{A}_k$ with $E\{\boldsymbol{\varepsilon}_t | \mathcal{F}_t\} = 0$ and $\sup_t E\|\boldsymbol{\varepsilon}_t\|_\infty < \infty$. Assume in addition that $\sup_t E\|\mathbf{A}_j \mathbf{Y}_t\|_\infty = O_p(n^{1/2})$ for all $1 \le j \le k$ and $\|\mathbf{b}_{(j)}\|_1 < \infty$ where $\mathbf{b}_{(j)}$ is the j-th row of matrix $\mathbf{B}^{-1}$, with $\mathbf{B} = \mathbf{I} - \mathbf{A}_1 - \cdots - \mathbf{A}_k$.

The conditions are imposed to facilitate the technical proof. They are not weakest possible. In particular, the condition such as $\max_t E\|\boldsymbol{\varepsilon}_t\|_\infty < \infty$ can be relaxed by replacing an upper bound depending on p such as $\log p$, and the conclusion continues to hold with some simple modifications. The assumptions on matrices $\{\mathbf{A}_j\}$ can easily be checked when they are diagonal. In particular, the assumption holds when $\{\mathbf{R}_t\}$ are a sequence of independently identically distributed random vectors. Since we assume $\boldsymbol{\mu} = 0$, the sample covariance matrix refers to the second moment.

Before introducing Condition 2, let us introduce the strong mixing coefficient $\alpha(k)$ of the process $\{\mathbf{R}_t\}$, which is given by

$$\alpha(k) = \sup_t \sup \{|P(AB) - P(A)P(B)| : A \in \sigma(\mathbf{R}_s, s \le t), \ B \in \sigma(\mathbf{R}_s, s \ge t+k)\},$$

where $\sigma(\mathbf{R}_s, s \le t)$ is the sigma-algebra generated by $\{\mathbf{R}_s, s \le t\}$.

Condition 2

Suppose that $\|\mathbf{R}_t\|_\infty < B$ for a constant $B > 0$ and that as $q \to \infty$, $\alpha(q) = O(\exp(-Cq^{1/b}))$ and $a > (b+1)/2$ in Theorem 3. In addition, $\log n = O(\log p)$.

Condition 3

Let η_t be $R_{ti}R_{tj} - ER_{ti}R_{tj}$ or $R_{ti} - ER_{ti}$ (we suppress its dependence on i and j). Assume that there exist nonnegative constants a, b, and B and a function $\rho(\cdot)$ such that

$$|\mathrm{cov}(\eta_{s_1} \cdots \eta_{s_u}, \eta_{t_1} \cdots \eta_{t_v})| \le B^{u+v} [(u+v)!]^b v \rho(t_1 - s_u),$$

for all i and j and any $1 \le s_1 \le \cdots \le s_u \le t_1 \le \cdots \le t_v \le n$ where

$$\sum_{s=0}^{\infty} (s+1)^k \rho(s) \le B^k (k!)^a \quad \text{for all } k > 0.$$

and

$$E|\eta_t|^k \le (k!)^v B^k, \quad \text{for all } k > 0.$$

In addition, we assume that $\log p = o(n^{1/(2a+2b+3)})$.

For AR and ARCH processes, Neumann and Paraporodidis (2008) show that this covariance weak dependence condition holds with $a = 1$, $b = 0$ and $\rho(s) = h^s$ for some $h < 1$.

Proof of Theorem 1

First of all, it is easy to see that

$$|R_n(\mathbf{w})-R(\mathbf{w})|=|\mathbf{w}^T(\widehat{\sum}-\sum)\mathbf{w}| \;\leq\; a_n\|\mathbf{w}\|_1^2, \quad \text{(A.1)}$$

which is bounded by $a_n c^2$. This proves the second inequality.

To prove the first inequality, by using $R(\mathbf{w}_{\text{opt}}) - R(\hat{\mathbf{w}}_{\text{opt}}) \leq 0$, we have that

$$\begin{aligned} R(\mathbf{w}_{\text{opt}})-R_n(\widehat{\mathbf{w}}_{\text{opt}}) &= R(\mathbf{w}_{\text{opt}})-R(\widehat{\mathbf{w}}_{\text{opt}})+R(\widehat{\mathbf{w}}_{\text{opt}})-R_n(\widehat{\mathbf{w}}_{\text{opt}}) \\ &\leq R(\widehat{\mathbf{w}}_{\text{opt}})-R_n(\widehat{\mathbf{w}}_{\text{opt}}) \\ &\leq a_n c^2, \end{aligned}$$

where the last inequality follows from (A.1). Similarly, it follows from $R_n(\mathbf{w}_{\text{opt}}) - R_n(\hat{\mathbf{w}}_{\text{opt}}) \geq 0$ that

$$\begin{aligned} R(\mathbf{w}_{\text{opt}})-R_n(\widehat{\mathbf{w}}_{\text{opt}}) &= R(\mathbf{w}_{\text{opt}})-R(\mathbf{w}_{\text{opt}})+R(\mathbf{w}_{\text{opt}})-R_n(\widehat{\mathbf{w}}_{\text{opt}}) \\ &\leq R(\mathbf{w}_{\text{opt}})-R_n(\mathbf{w}_{\text{opt}}) \\ &\leq -a_n c^2. \end{aligned}$$

Combining the last two results, the third inequality follows.

Now, let us prove the third inequality. Using again $R(\hat{\mathbf{w}}_{\text{opt}}) - R(\mathbf{w}_{\text{opt}}) \geq 0$ and $R_n(\hat{\mathbf{w}}_{\text{opt}}) - R_n(\mathbf{w}_{\text{opt}}) \leq 0$, we have

$$\begin{aligned} R(\widehat{\mathbf{w}}_{\text{opt}})-R(\mathbf{w}_{\text{opt}}) &= R(\widehat{\mathbf{w}}_{\text{opt}})-R_n(\widehat{\mathbf{w}}_{\text{opt}})+R_n(\widehat{\mathbf{w}}_{\text{opt}})-R_n(\mathbf{w}_{\text{opt}})+R_n(\mathbf{w}_{\text{opt}})-R(\mathbf{w}_{\text{opt}}) \\ &\leq R(\widehat{\mathbf{w}}_{\text{opt}})-R_n(\widehat{\mathbf{w}}_{\text{opt}})+R_n(\mathbf{w}_{\text{opt}})-R(\mathbf{w}_{\text{opt}}) \\ &\leq 2\sup_{\|\mathbf{w}\|\leq c}|R_n(\mathbf{w})-R(\mathbf{w})|. \end{aligned} \quad \text{(A.2)}$$

This together with (A.1) proves the third inequality. The last inequality follows from the fact that

$$E\{R_n(\widehat{\mathbf{w}}_{\text{opt}})\} \leq E\{R_n(\mathbf{w}_{\text{opt}})\}=R(\mathbf{w}_{\text{opt}}).$$

This completes the proof.

We need the following lemma to prove Theorem 2.

Lemma 1

Let $\xi_1, \cdots, \xi_n$ be a series of p-dimensional random vectors. Assume that ξ_t is $\mathcal{F}_t$-adaptive and each component is a martingale difference: $E(\xi_{t+1}|\mathcal{F}_t) = 0$. Then, for any $p \geq 3$ and $r \in [2, \infty]$, we have for some universal constant C

$$E\left|\left|\sum_{t=1}^{n}\xi_t\right|\right|_1^2 \leq C\min[r, \log p]\sum_{t=1}^{n}E\|\xi_t\|_r^2, \quad \text{(A.3)}$$

where $\|\xi_t\|_r$ is the l_r-norm of the vector ξ_t in R^p.

This is an extension of the Nemirovski's inequality to the marginale difference sequence. The proof follows similar arguments on page 188 of Emery el al (2000) and Dumbgen *et al.* (2010).

Proof of lemma 1

Let $V(\mathbf{x})=\|\mathbf{x}\|_r^2$. Then, there exists a universal constant C such that

$$V(\mathbf{x}+\mathbf{y}) \leq V(\mathbf{x})+\mathbf{y}^T V'(\mathbf{x})+CrV(\mathbf{y}),$$

where $V'(\mathbf{x})$ is the gradient vector of $V(\mathbf{x})$. Using this, we have

$$V(\sum_{t=1}^{n}\xi_t) \leq V(\sum_{t=1}^{n-1}\xi_t)+\xi_n^T V'(\sum_{t=1}^{n-1}\xi_t)+CrV(\xi_n). \quad \text{(A.4)}$$

Since ξ_n is a martingale difference and $V'(\sum_{t=1}^{n-1}\xi_t)$ is $\mathcal{F}_{n-1}$ adaptive, we have

$$E\xi_n^T V'(\sum_{t=1}^{n-1}\xi_t)=0.$$

By taking the expectation on both sides of (A.4), we have

$$EV(\sum_{t=1}^{n}\xi_t) \leq EV(\sum_{t=1}^{n-1}\xi_t)+CrEV(\xi_n).$$

Iteratively applying the above formula, we have

$$E\left|\left|\sum_{t=1}^{n}\xi_t\right|\right|_r^2 \leq Cr\sum_{t=1}^{n}\|\xi_t\|_r^2. \quad \text{(A.5)}$$

This proves the first half of the inequality (A.3).

To prove the inequality (A.3), without loss of generality, assume that $r \geq \log p$. Let $r' = \log p > 1$. Then, for any $\mathbf{x}$ in the p-dimensional space,

$$\|\mathbf{x}\|_r \leq \|\mathbf{x}\|_{r'} \leq p^{\frac{1}{r'}-\frac{1}{r}} \|\mathbf{x}\|_r.$$

Hence, by (A.5)

$$\begin{aligned} E\left\|\sum_{t=1}^{n}\xi_t\right\|_r^2 &\leq E\left\|\sum_{t=1}^{n}\boldsymbol{\xi}_t\right\|_{r'}^2 \\ &\leq C\log p \sum_{t=1}^{n} E\|\boldsymbol{\xi}_t\|_{r'}^2 \\ &\leq C\log p \sum_{t=1}^{n} p^{2(\frac{1}{r'}-\frac{1}{r})} E\|\boldsymbol{\xi}_t\|_r^2. \end{aligned}$$

Using the simple fact $p^{\frac{2}{r'}} = e^2$, we complete the proof of the inequality (A.3).

Proof of Theorem 2

Applying lemma 1, with $r = \infty$, we have

$$E\left\|n^{-1}\sum_{t=1}^{n}\boldsymbol{\xi}_t\right\|_\infty^2 \leq \frac{C\log p}{n}\max_t E\|\boldsymbol{\xi}_t\|_\infty^2, \quad \text{(A.6)}$$

for all t, where $E\|\boldsymbol{\xi}_t\|_\infty^2 = E(\max_{1\leq j\leq p}\xi_{tj}^2)$. As a result, by Condition 1, an application of (A.6) to $p(p+1)/2$-element of $\boldsymbol{\varepsilon}_t$ yields

$$E\left\|(n-k)^{-1}\sum_{t=k+1}^{n}(\boldsymbol{Y}_t - \mathbf{A}_1\boldsymbol{Y}_{t-1} - \cdots - \mathbf{A}_k\boldsymbol{Y}_{t-k})\right\|_\infty^2 \leq \frac{C\log p^2}{n-k}\max_t E\|\boldsymbol{\varepsilon}_t\|_\infty^2.$$

Note that each of the summation $(n-k)^{-1}\sum_{t=k+1}^{n}\boldsymbol{Y}_{t-j}(\text{for } j \leq k)$ (for $j \leq k$) is approximately the same as $n^{-1}\sum_{t=1}^{n}\boldsymbol{Y}_t$ since k is finite, by appealing to Condition 1. Hence, we can easily show that

$$\left\|\mathbf{B}n^{-1}\sum_{t=1}^{n}\boldsymbol{Y}_t\right\|_\infty = O_p\left(\sqrt{\frac{\log p}{n}}\right).$$

By the assumption on the matrix $\mathbf{B}$, we can easily deduce that

$$\left|\left|n^{-1}\sum_{t=1}^{n}\mathbf{Y}_t\right|\right|_\infty = O_p\left(\sqrt{\frac{\log p}{n}}\right).$$

Rearranging this into matrix form, we conclude that

$$\left|\left|n^{-1}\sum_{t=1}^{n}\mathbf{R}_t\mathbf{R}_t^T - \mathbf{S}\right|\right|_\infty = O_p\left(\sqrt{\frac{\log p}{n}}\right).$$

Proof of Theorem 3

Note that by the union bound of probability, we have for any $D > 0$,

$$P\{\sqrt{n}\left|\left|\sum - \widehat{\sum}\right|\right|_\infty > D(\log p)^a\} \leq p^2 \max_{i,j} P\{\sqrt{n}|\sigma_{ij} - \widehat{\sigma}_{ij}| > D(\log p)^a\}.$$

By the assumption of the theorem, the above probability is bounded by

$$p^2 \exp\left(-C[D(\log p)^a]^{1/a}\right) = p^2 p^{-CD1/a},$$

which tends to zero when D is large enough. This proves the first part of the theorem.

We now prove the second part of the α-mixing process. Let ξ_t be an $\mathcal{F}_t$ adaptive random variable with $E\xi_t = 0$ and assume that $|\xi_t| \leq B$ for all t. Then, by Theorem 1.3 of Bosq (1998), for any integer $q \leq n/2$, we have

$$P(|\bar{\xi}_n| > \varepsilon) \leq 4\exp\left(-\frac{q\varepsilon^2}{8B^2}\right) + 22(1 + 4B/\varepsilon)^{1/2} q\alpha([n/(2q)]), \quad \text{(A.7)}$$

where $\bar{\xi}_n = n^{-1}\sum_{t=1}^{n}\xi_t$. Taking $\varepsilon_n = 4BD(\log p)^a/\sqrt{n}$ and $q = n(\log p)^{1-2a}/2$, we obtain from (A.7) that

$$P(|\bar{\xi}_n| > \varepsilon_n) = 4p^{-D^2} + o(n^{3/2})\alpha((\log p)^{2a-1}).$$

Now, the assumption on the mixing coefficient $\alpha(\cdot)$, we conclude that for sufficiently large D,

$$P(|\bar{\xi}_n| > \varepsilon_n) = o(p^{-2}), \quad \text{(A.8)}$$

for $a > (b + 1)/2$.

Applying (A.8) to $\xi_t = R_{ti}R_{tj} - ER_{ti}R_{tj}$ with a sufficiently large D, we have

$$P(n^{-1}\sum_{t=1}^{n}|R_{ti}R_{tj}-ER_iR_j|>\varepsilon_n)=o(p^{-2}).$$

This together with the first part of the proof of Theorem 2 yield that

$$\left|\left|n^{-1}\sum_{t=1}^{n}\mathbf{R}_t\mathbf{R}_t^T-\mathbf{S}\right|\right|_\infty=O_p(\varepsilon_n),$$

where we borrow the notation from the proof of Theorem 2. Similarly, by an application of (A.8), we obtain

$$\|\overline{\mathbf{R}}_n-\mu\|_\infty=O_p(\varepsilon_n).$$

Combining the last two results, we prove the second part of the theorem.

The proof of the third part of the theorem follows similar steps. By Theorem 1 of Doukhan and Neumann (2007), under Condition 3, we have

$$P(|\sum_{t=1}^{n}\eta_t|>\sqrt{n}x)\le\exp(-C\ \min\{x^2,(\sqrt{n}x)^c\})$$

for some $C > 0$, where $c = 1/(a + b + 2)$. Now, taking $x = D(\log p)^{1/2}$, we have

$$x^2/(\sqrt{n}x)^c=O((\log p)^{1-c/2}/n^{c/2})=o(1),$$

since $\log p = o(n^{1/(2\mu+2\nu+3)})$. Thus, the exponent is as large as

$$C\min\{x^2,(\sqrt{n}x)^c\}\ge CD^2\log p,$$

for sufficiently large n. Consequently,

$$P(|\sum_{t=1}^{n}\eta_t|>D\sqrt{n\log p})\le\exp(-CD^2\log p)=o(p^{-2})$$

for sufficiently large D. Now, substituting the definition of η_t, we have

$$P(n^{-1}\sum_{t=1}^{n}|R_{ti}R_{tj}-ER_iR_j|>D\sqrt{(\log p)/n})=o(p^{-2}). \quad \text{(A.9)}$$

$$P(n^{-1}\sum_{t=1}^{n}|R_{ti}-ER_i|>D\sqrt{(\log p)/n})=o(p^{-2}). \quad \text{(A.10)}$$

Combining the results in (A.9) and (A.10) and using the same argument as proving the first part of Theorem 2, we have

$$\left|\left|n^{-1}\sum_{t=1}^{n}\mathbf{R}_t\mathbf{R}_t^T-\mathbf{S}\right|\right|_\infty=O_P\left(\sqrt{(\log p)/n}\right).$$

and

$$\left|\left|n^{-1}\sum_{t=1}^{n}\mathbf{R}_t-\mu\right|\right|_\infty=O_P\left(\sqrt{(\log p)/n}\right).$$

The conclusion follows from these two results.

Proof of Theorem 4

First of all, note that the solution to problem (2.12) is given by

$$\mathbf{w}_{\text{opt}}=\tilde{\Sigma}_c^{-1}\mathbf{1}/\mathbf{1}^T\tilde{\Sigma}_c^{-1}\mathbf{1}.$$

By $\tilde{\mathbf{w}}^T\mathbf{1}=1$ and $\tilde{\mathbf{g}}^T\tilde{\mathbf{w}}=\|\tilde{\mathbf{w}}\|_1$, we have

$$\begin{aligned}\tilde{\Sigma}_c\tilde{\mathbf{w}}&=\widehat{\Sigma}\tilde{\mathbf{w}}+\lambda_1\tilde{\mathbf{g}}+\lambda_1\left|\left|\tilde{\mathbf{w}}\right|\right|_1\mathbf{1}\\&=(\lambda_2+\lambda_1c)\mathbf{1},\end{aligned}$$

in which the last equality utilizes (2.10) and (2.11). Thus, $\tilde{\mathbf{w}}=(\lambda_2+\lambda_1c)\tilde{\Sigma}_c^{-1}\mathbf{1}$, which has the same direction $\mathbf{w}_{\text{opt}}$. Since $\mathbf{1}^T\tilde{\mathbf{w}}=1$, they must be equal. This completes the proof.

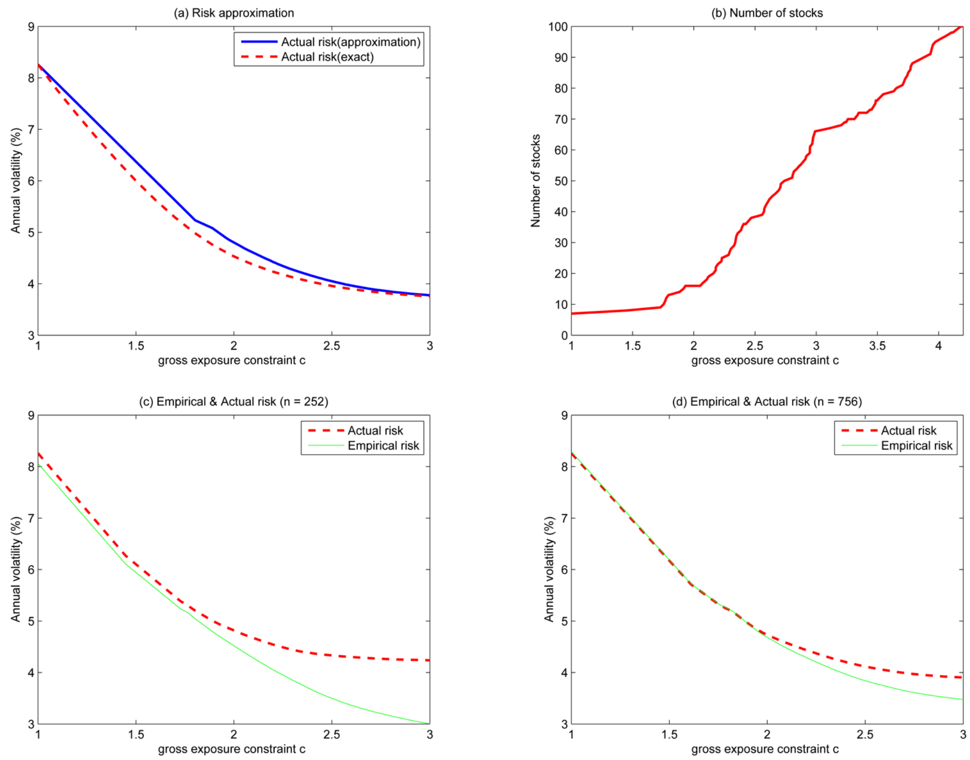

Figure 1.
Comparisons of optimal portfolios selected by the exact and approximate algorithms with a known covariance matrix. (a) The risks for the exact algorithm (dashed line) and the LARS (approximate) algorithm. (b) The number of stocks picked by the optimization problem (2.6) as a function of the gross exposure coefficient *c*. (c) The actual risk (dashed line) and empirical risk (solid) of the portfolio selected based on the sample covariance matrix (n = 252). (d) The same as (c) except n = 756.

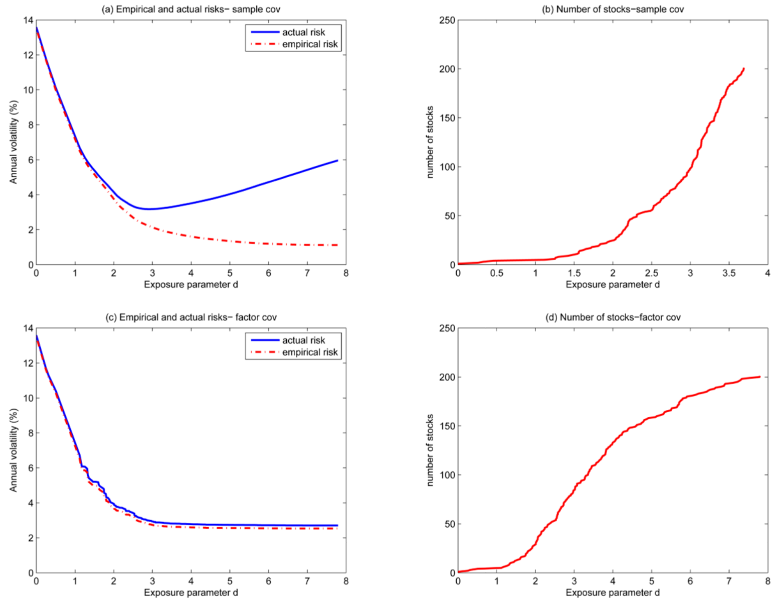

Figure 2.
Risk improvement of the 200 equally weighted portfolio by modifying the weights of the portfolio using (3.3). As the exposure parameter d increases, more weights are modified and the risks of new portfolios decrease. (a) The empirical and actual risks of the modified portfolios are plotted against exposure parameter d, based on the sample covariance matrix. (b) The number of stocks whose weights are modified as a function of d. (c) and (d) are the same as (a) and (b) except that the covariance matrix is estimated based on the factor model.

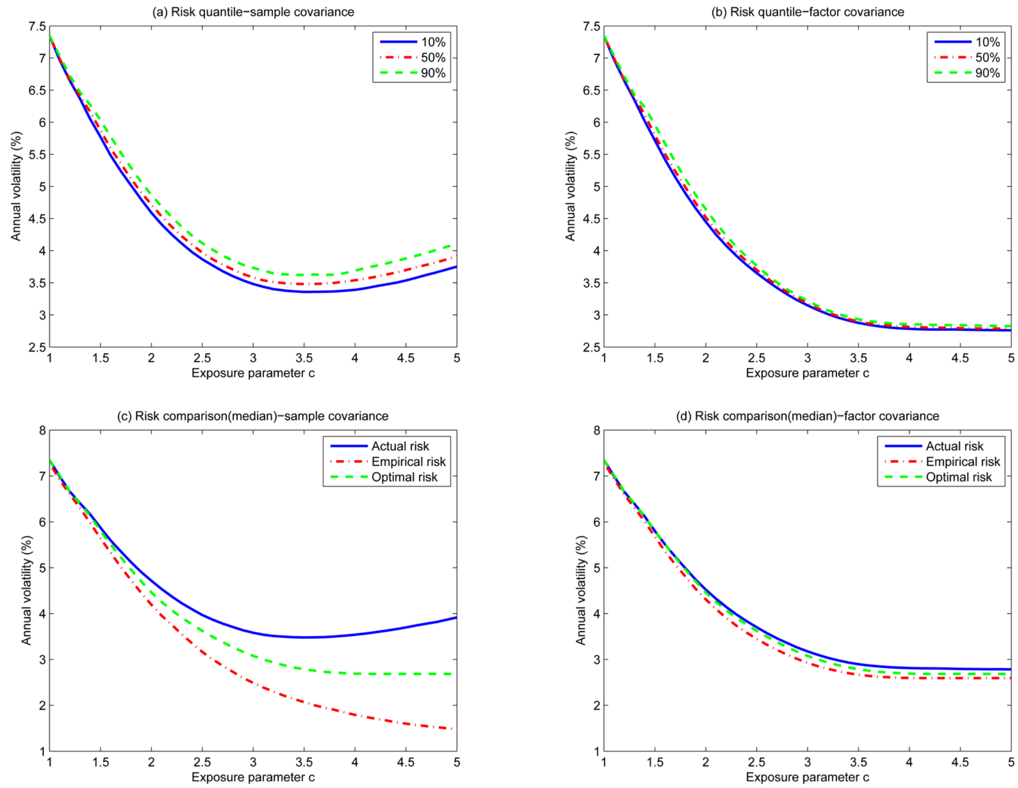

Figure 3.

The 10%, 50% and 90% quantiles of the actual risks of the 101 empirically chosen portfolios for each given gross exposure parameter *c* are shown in (a) (sample covariance matrix) and (b) (factor model) for the case with 200 stocks. They indicate the sampling variability among 101 simulations. The theoretical optimal risk, the median of the actual risks and the median of the empirical risks of 101 empirically selected portfolios are also summarized in (c) (based on the sample covariance) and (d) (based on the factor model).

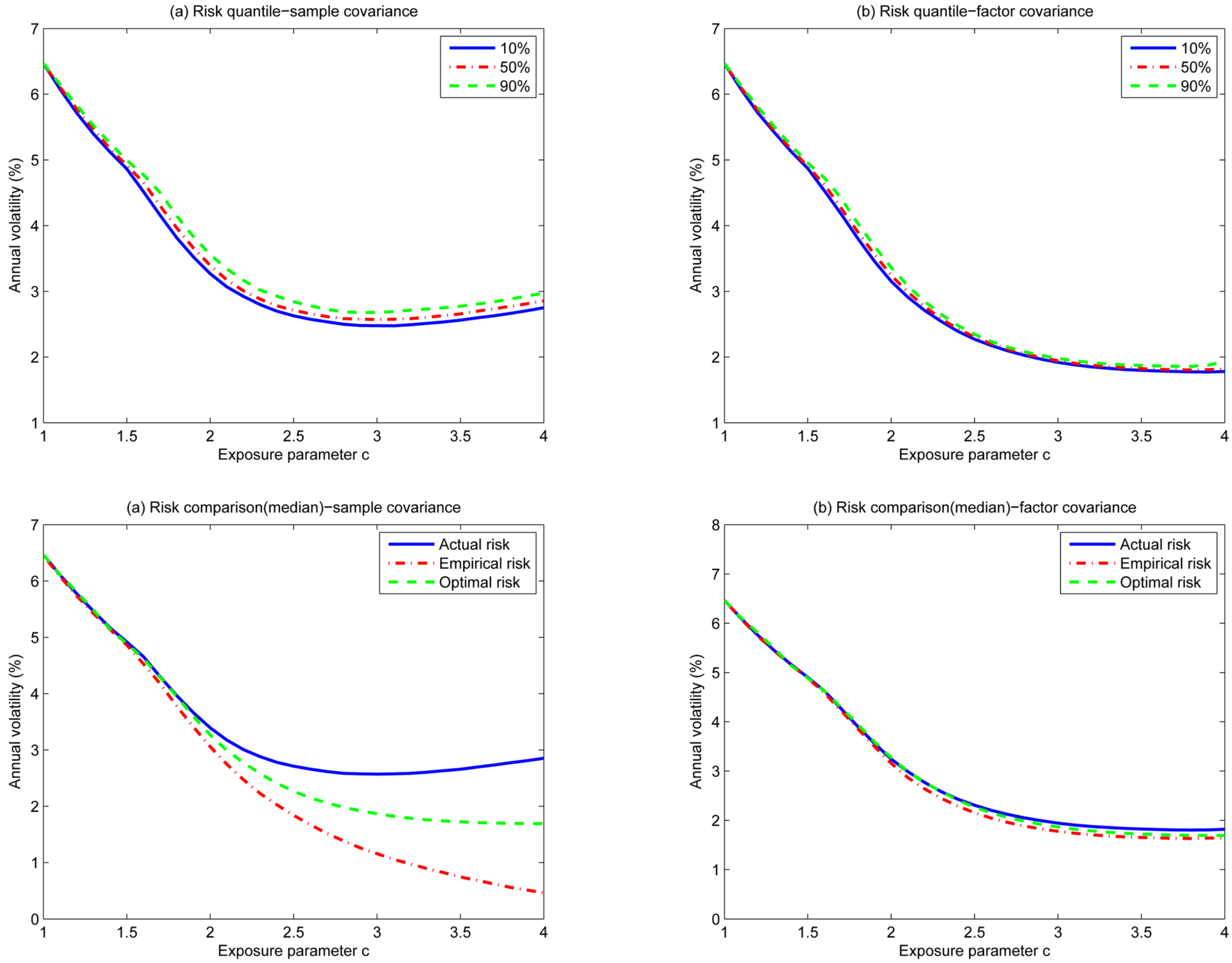

Figure 4.

This is similar to Figure 3 except $p = 500$. The sample covariance matrix is always degenerate under this setting ($n = 252$). Nevertheless, for the given range of c, the gross-constrained portfolio performs normally. The same captions as Figure 3 are used.

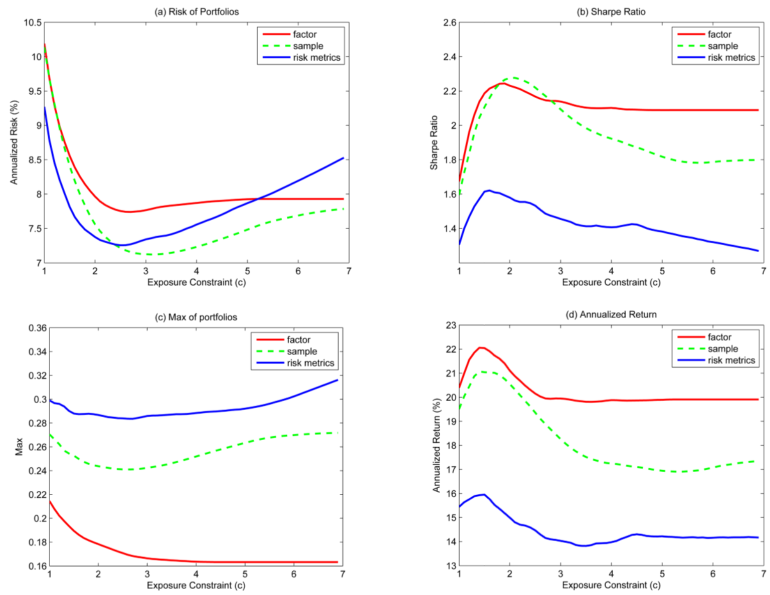

Figure 5.
Characteristics of invested portfolios as a function of exposure constraints(c) from the Fama-French 100 industrial portfolios formed by the size and book to market from Jan 2, 1998 to December 31, 2007. (a) Annualized risk of portfolios. (b) Sharpe ratio of portfolios. (c) Max weight of allocations. (d) Annualized return of portfolios

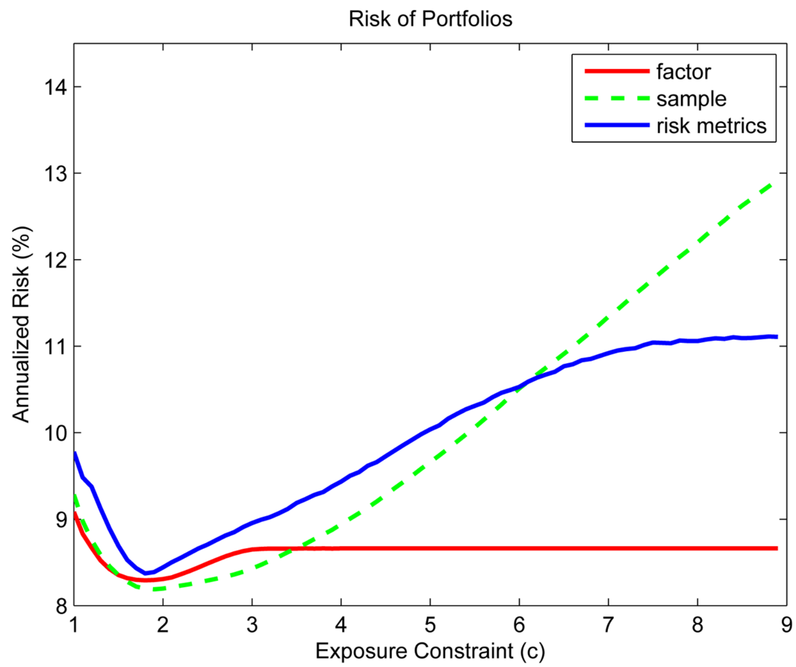

Figure 6.
Risks of the optimal portfolios as a function of the gross exposure constraint. They are the annualized volatilities as a function of the gross exposure parameter c.

Table 1

Parameters used in the simulation

Parameters for factor loadings				Parameters for factor returns			
μ_b	$\mathbf{cov}_b$			μ_f	$\mathbf{cov}_f$		
0.7828	0.02914	0.02387	0.010184	0.02355	1.2507	−0.0350	−0.2042
0.5180	0.02387	0.05395	−0.006967	0.01298	−0.0350	0.3156	−0.0023
0.4100	0.01018	−0.00696	0.086856	0.02071	−0.2042	−0.0023	0.1930

Table 2

Returns and Risks based on 100 Fama-French Industrial

Methods	Mean %	Std Dev %	Sharpe Ratio	Max Weight	Min Weight	No. of Long Positions	No. of Short Positions
			Sample Covariance Matrix Estimator				
No short(c = 1)	19.51	10.14	1.60	0.27	−0.00	6	0
Exact(c = 1.5)	21.04	8.41	2.11	0.25	−0.07	9	6
Exact(c = 2)	20.55	7.56	2.28	0.24	−0.09	15	12
Exact(c = 3)	18.26	7.13	2.09	0.24	−0.11	27	25
Approx. (c = 2, Y=NS)	21.16	7.89	2.26	0.32	−0.08	9	13
Approx. (c = 3, Y=NS)	19.28	7.08	2.25	0.28	−0.11	23	24
GMV Portfolio	17.55	7.82	1.82	0.66	−0.32	52	48
			Factor-Based Covariance Matrix Estimator				
No short(c = 1)	20.40	10.19	1.67	0.21	−0.00	7	0
Exact(c = 1.5)	22.05	8.56	2.19	0.19	−0.05	11	8
Exact(c = 2)	21.11	7.96	2.23	0.18	−0.05	17	18
Exact(c = 3)	19.95	7.77	2.14	0.17	−0.05	35	41
Approx. (c=2, Y=NS)	21.71	8.07	2.28	0.24	−0.04	10	19
Approx. (c=3, Y=NS)	20.12	7.84	2.14	0.18	−0.05	33	43
GMV Portfolio	19.90	7.93	2.09	0.43	−0.14	45	55
			Covariance Estimation from Risk Metrics				
No short(c = 1)	15.45	9.27	1.31	0.30	−0.00	6	0
Exact(c = 1.5)	15.96	7.81	1.61	0.29	−0.07	9	5
Exact(c = 2)	14.99	7.38	1.58	0.29	−0.10	13	9
Exact(c = 3)	14.03	7.34	1.46	0.29	−0.13	21	18
Approx. (c=2, Y=NS)	15.56	7.33	1.67	0.34	−0.08	9	11
Approx. (c=3, Y=NS)	15.73	6.95	1.78	0.30	−0.11	20	20
GMV Portfolio	13.99	9.47	1.12	0.78	−0.54	53	47
			Unmanaged Index				

Methods	Mean %	Std Dev %	Sharpe Ratio	Max Weight	Min Weight	No. of Long Positions	No. of Short Positions
Equal weighted	10.86	16.33	0.46	0.01	0.01	100	0
CRSP	8.2	17.9	0.26				

Table 3

Returns and Risks based on 600 stock portfolio

Methods	Std Dev %	Max Weight	Min Weight	No. of Long Positions	No. of Short Positions
Sample Covariance Matrix Estimator					
No short	9.28	0.14	0.00	53	0
c = 2	8.20	0.11	−0.06	123	67
c = 3	8.43	0.09	−0.07	169	117
c = 4	8.94	0.10	−0.08	201	154
c = 5	9.66	0.12	−0.10	225	181
c = 6	10.51	0.13	−0.10	242	201
c = 7	11.34	0.14	−0.11	255	219
c = 8	12.20	0.17	−0.12	267	235
Factor-based Covariance Matrix Estimator					
No short	9.08	0.12	0.00	54	0
c = 2	8.31	0.06	−0.03	188	120
c = 3	8.65	0.05	−0.03	314	272
c = 4	8.66	0.05	−0.03	315	273
c = 5	8.66	0.05	−0.03	315	273
c = 6	8.66	0.05	−0.03	315	273
c = 7	8.66	0.05	−0.03	315	273
c = 8	8.66	0.05	−0.03	315	273
Covariance Estimation from Risk Metrics					
No short	9.78	0.40	0.00	31	0
c = 2	8.44	0.12	−0.06	119	63
c = 3	8.95	0.11	−0.07	191	133
c = 4	9.43	0.12	−0.09	246	192
c = 5	10.04	0.12	−0.10	279	233
c = 6	10.53	0.12	−0.11	300	258
c = 7	10.92	0.13	−0.11	311	272

Methods	Std Dev %	Max Weight	Min Weight	No. of Long Positions	No. of Short Positions
c = 8	11.06	0.13	−0.10	315	277