Journal of Econometrics 186 (2015) 367–387

Contents lists available at ScienceDirect

Journal of Econometrics

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

Risks of large portfolios

Jianqing Fan [a,b,c,*], Yuan Liao [d], Xiaofeng Shi [a]

[a] Department of Operations Research and Financial Engineering, Princeton University, United States
[b] International School of Economics and Management, Capital Univeristy of Economics and Business, Beijing, China
[c] Bendheim Center for Finance, Princeton University, United States
[d] Department of Mathematics, University of Maryland, United States

ARTICLE INFO

Article history:
Available online 7 March 2015

JEL classification:
C58
C38

Keywords:
High dimensionality
Factor models
Principal components
Sparse matrix
Volatility

ABSTRACT

The risk of a large portfolio is often estimated by substituting a good estimator of the volatility matrix. However, the accuracy of such a risk estimator is largely unknown. We study factor-based risk estimators under a large amount of assets, and introduce a high-confidence level upper bound (H-CLUB) to assess the estimation. The H-CLUB is constructed using the confidence interval of risk estimators with either known or unknown factors. We derive the limiting distribution of the estimated risks in high dimensionality. We find that when the dimension is large, the factor-based risk estimators have the same asymptotic variance no matter whether the factors are known or not, which is slightly smaller than that of the sample covariance-based estimator. Numerically, H-CLUB outperforms the traditional crude bounds, and provides an insightful risk assessment. In addition, our simulated results quantify the relative error in the risk estimation, which is usually negligible using 3-month daily data.

1. Introduction

The potential of a portfolio's loss is termed as the portfolio risk. There are two types of portfolio risks. The *systematic risk* is the risk inherent to the entire market, such as risk associated with interest rates, currencies, recession, war and political instability, etc. The systematic risk cannot be diversified away, even with a well-diversified portfolio. In contrast, *specific risk* (or *idiosyncratic risk*) refers to the risk that affects a very specific group of securities or even an individual security. For example, it can be the risk of price changes due to the unique circumstances of a specific stock. Unlike systematic risk, specific risk can be reduced through diversification.

Estimating and assessing the risk of a large portfolio is an important topic in financial econometrics and risk management. The risk of a given portfolio allocation vector $\mathbf{w}_T$ is conveniently measured by $(\mathbf{w}_T'\boldsymbol{\Sigma}\mathbf{w}_T)^{1/2}$, in which $\boldsymbol{\Sigma}$ is a volatility (covariance) matrix of the assets' returns. Often multiple portfolio risks are at interests and hence it is essential to estimate the volatility matrix $\boldsymbol{\Sigma}$. The problem becomes challenging when the portfolio size is large. Suppose we have created a portfolio from two thousand assets and invested in a part of selected assets. The covariance matrix $\boldsymbol{\Sigma}$ involved then contains over two million unknown parameters. Yet, the sample size based on one year's daily data is around 252. It is hard to assess the estimation accuracy when the estimation errors from more than two million parameters are aggregated. Hence some regularization method is recommended to estimate and assess risks.

We estimate and assess the risks of a given portfolio vector $\mathbf{w}_T$ based on factor models. Two factor-based methods are compared, previously proposed by Fan et al. (2011, 2013). The first estimator assumes the factors to be known and observable. The second method deals with the case of unknown factors. In both cases, the factor model imposes a *conditionally sparse* structure, in that the idiosyncratic covariance is a large sparse matrix. This yields to an *approximate factor model* as in Chamberlain and Rothschild (1983), with a non-diagonal error covariance matrix.

We provide a new and practical method to assess the accuracy of risk estimation $\mathbf{w}_T'(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\mathbf{w}_T$. In the literature (e.g. Fan et al., 2012), this term has been bounded by

$$\widehat{\xi}_T = \|\mathbf{w}_T\|_1^2\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\|_{\max}$$

where $\|\mathbf{w}_T\|_1$ is the gross exposure of the portfolio, and is bounded when there are no extreme positions in the portfolio. However, this upper bound depends on the unknown $\boldsymbol{\Sigma}$, hence is not applicable in practice. In addition, numerical studies in this paper demonstrate

* Correspondence to: Department of Operations Research and Financial Engineering, Sherrerd Hall, Princeton University, Princeton, NJ 08544, United States. Tel.: +1 609 258 3255.
E-mail address: jqfan@princeton.edu (J. Fan).

http://dx.doi.org/10.1016/j.jeconom.2015.02.015
0304-4076/

that this upper bound is too crude: it is often of the same or even larger scale than the estimated risk. In contrast, we provide a high-confidence level upper bound (H-CLUB) for $\mathbf{w}_T'(\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma})\mathbf{w}_T$, which is of much smaller scale and easy to compute in practice. H-CLUB is constructed based on the confidence interval for the true risk. For the risk estimator $\mathbf{w}_T'\widehat{\mathbf{\Sigma}}\mathbf{w}_T$ and a given $\epsilon\in(0,1)$, we find an H-CLUB $\widehat{U}(\epsilon)$ such that

$$P(|\mathbf{w}_T'(\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma})\mathbf{w}_T|\leq\widehat{U}(\epsilon))\to 1-\epsilon.$$

In contrast, $P(|\mathbf{w}_T'(\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma})\mathbf{w}_T|\leq\widehat{\xi}_T)=1$. Hence H-CLUB is an upper bound for the risk estimation error with high confidence while the traditional bound $\widehat{\xi}_T$ is of full confidence.

For the inferential theory of the risk estimators with diversified portfolios, we prove that the effects of estimating the factor loadings and unknown factors are asymptotically negligible. Interestingly, it is found that when the dimensionality is larger than the sample size, the factor-based risk estimators have the same asymptotic variances no matter whether the factors are known or not. Hence the high dimensionality is in fact a bless for risk estimation instead of a curse from this point of view. In addition, the asymptotic variance of factor-based estimators is slightly smaller than that of the sample covariance-based estimator, but the difference is small. This demonstrates that the benefit of using a factor model is not in terms of a much smaller asymptotic variance, because the systematic risk cannot be diversified. Rather, factor models give a strictly positive definite covariance estimator, which is essential to estimate the optimal portfolio allocation vector, and also interprets the structure of the portfolio risks.

Using our simulated results based on the model calibrated from the US equity market data, we are able to quantify the relative error of the estimation error or coefficient of variation, defined as $\text{STD}(\mathbf{w}_T'\widehat{\mathbf{\Sigma}}\mathbf{w}_T)/\mathbf{w}_T'\widehat{\mathbf{\Sigma}}\mathbf{w}_T$, where $\text{STD}(\cdot)$ denotes the standard error of the estimated risk. Interestingly, this ratio is just a few percent and is approximately independent of the gross exposure $\|\mathbf{w}_T\|_1$ but sensitive to the length of the time series. On the other hand, we also quantify the relation between the crude bound and the practical H-CLUB. We find that $\widehat{\xi}_T$ is many times larger than $\widehat{U}(\epsilon)$, and the ratio $\widehat{\xi}_T/\widehat{U}(\epsilon)$ increases as the gross exposure increases. A sampling technique that picks a random portfolio with a given gross exposure level is introduced, which can be useful for portfolio optimization and understanding the overall risks within a given level of gross exposure.

The interest on large portfolios surges recently. Pesaran and Zaffaroni (2008) examined the asymptotic behavior of the portfolio weights. Brodie et al. (2009) and Fan et al. (2012) addressed the problem of portfolio selection using a regularization penalty. Gomez and Gallon (2011) numerically compared several methods of covariance matrix estimation for portfolio management. In particular, the optimal portfolio selection involves inverting an estimated $\mathbf{\Sigma}$, which is a challenging problem under a large number of assets. Gagliardini et al. (2010) considered a random coefficient model for an unbalanced panel, and focused on the observable factors, while we also study the inferential theory of the unobservable factor case. The recent works by Fan et al. (2011, 2013) are only concerned about covariance estimations and no inferential theories were studied. The literature is also found in Jacquier and Polson (2010), Antoine (2011), Chang and Tsay (2010), DeMiguel et al. (2009a,b), Ledoit and Wolf (2003), El Karoui (2010), Lai et al. (2011), Bannouh et al. (2012), Gandy and Veraart (2012), Bianchi and Carvalho (2011), among others.

The rest of the paper is organized as follows. Section 2 introduces risk estimators based on factor models under both known and unknown factors. Section 3 constructs the H-CLUB for each risk estimator based on the confidence interval for risks. Section 4 derives the limiting distributions of the risk estimators and compares their asymptotic variances. Section 5 presents simulation results. An empirical study is considered in Section 6. Finally, Section 7 concludes. All the proofs are given in the Appendix.

Throughout the paper, $\|\mathbf{w}_T\|_1=\sum_{i=1}^N|w_i|$ is used to denote the gross exposure of a given portfolio allocation vector. For a square matrix $\mathbf{A}$, $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ represent its minimum and maximum eigenvalues; $\|\mathbf{A}\|_1=\max_i\sum_j|A_{ij}|$. Let $\|\mathbf{A}\|_{\max}$ and $\|\mathbf{A}\|$ denote its element-wise sup-norm and operator norm, given by $\|\mathbf{A}\|_{\max}=\max_{i,j}|A_{ij}|$ and $\|\mathbf{A}\|=\lambda_{\max}^{1/2}(\mathbf{A}'\mathbf{A})$ respectively.

2. Estimation of portfolio's risks

Let $\{\mathbf{R}_t\}_{t=1}^T$ be a strictly stationary time series of an $N\times 1$ vector of observed excess returns and $\mathbf{\Sigma}=\text{cov}(\mathbf{R}_t)$, often known as the volatility matrix. The portfolio risk of a given allocation vector $\mathbf{w}_T$ is given by $\sqrt{\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T}$. With a covariance estimator $\widehat{\mathbf{\Sigma}}$, a straightforward estimator of the portfolio risk is $\sqrt{\mathbf{w}_T'\widehat{\mathbf{\Sigma}}\mathbf{w}_T}$. But how good such a substitution estimator is and how to assess its estimation accuracy when the dimension N is large relative to T are the questions addressed here.

The problem of estimating the risk of a given portfolio is challenging due to the high dimensionality of $\mathbf{\Sigma}$. In most cases N can be much larger than T. We assume $\mathbf{\Sigma}$ to be time-invariant within a short period, which holds approximately for locally stationary time series. Recently, Chang and Tsay (2010) proposed a Cholesky decomposition approach to estimating the large covariance matrix, and used simulation to assess its performance. A natural alternative approach is through the factor model (e.g. Stock and Watson, 2002; Bai, 2003), because the assets' returns are usually driven by a few market factors. Estimating $\mathbf{\Sigma}$ is possible when both the factor and the idiosyncratic components can be estimated well. We thus consider three estimators of $\mathbf{w}_T'\mathbf{\Sigma}\mathbf{w}_T$ for a given $\mathbf{w}_T$, based on three different estimators $\widehat{\mathbf{\Sigma}}$: sample covariance estimator, and factor model estimators with either observed or unobserved factors.

2.1. Sample-covariance-based estimator

The first estimator $\widehat{\mathbf{\Sigma}}=\mathbf{S}$ is the conventional sample covariance matrix based on $\{\mathbf{R}_t\}_{t=1}^T$. Because we are mainly concerned about the variance, for simplicity and exposition, let us assume that the returns have mean zero and $\mathbf{S}=T^{-1}\sum_{t=1}^T\mathbf{R}_t\mathbf{R}_t'$. The asymptotic impact of using $\mathbf{S}$ on the risk management has been studied by Fan et al. (2008, 2012) when N is much larger than T. The sample covariance estimator does not require any structural assumption on the assets' returns. It was shown by the aforementioned authors that for a given portfolio $\mathbf{w}_T$ with a bounded gross exposure (that is, $\|\mathbf{w}_T\|_1$ is bounded),

$$\mathbf{w}_T'(\mathbf{S}-\mathbf{\Sigma})\mathbf{w}_T\leq\|\mathbf{w}_T\|_1^2\|\mathbf{S}-\mathbf{\Sigma}\|_{\max}=O_p\left(\sqrt{\frac{\log N}{T}}\right).$$

However, when $N>T$, it is well known that $\mathbf{S}$ is singular, and therefore may result in an estimated risk close to zero for certain portfolios.

2.2. Estimating risks based on factor models

We assume the true data generating process (DGP) of $\mathbf{R}_t$ to be an "approximate factor model" (Chamberlain and Rothschild, 1983):

$$\mathbf{R}_t=\mathbf{B}\mathbf{f}_t+\mathbf{u}_t,\quad t\leq T,\tag{2.1}$$

where $\mathbf{B}$ is an $N\times K$ matrix of factor loadings; $\mathbf{f}_t$ is a $K\times 1$ vector of common factors, and $\mathbf{u}_t$ is an $N\times 1$ vector of idiosyncratic error components. In contrast to N and T, here K is assumed to be fixed. The common factors may or may not be observable. For