Finance Research Letters 9 (2012) 121–134

Contents lists available at SciVerse ScienceDirect

Finance Research Letters

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

Robust estimation of covariance and its application to portfolio optimization

Lijuan Huo [a,b], Tae-Hwan Kim [a,*], Yunmi Kim [c]

[a] School of Economics, Yonsei University, Republic of Korea
[b] Department of Applied Mathematics, School of Science, Changchun University of Science and Technology, China
[c] Department of Economics, Kookmin University, Republic of Korea

ARTICLE INFO

Article history:
Received 16 December 2011
Accepted 8 June 2012
Available online 28 June 2012

JEL classification:
C13
C18
C46
G11

Keywords:
Covariance
Robust estimation
Median

ABSTRACT

Outliers can have a considerable influence on the conventional measure of covariance, which may lead to a misleading understanding of the comovement between two variables. Both an analytical derivation and Monte Carlo simulations show that the conventional measure of covariance can be heavily influenced in the presence of outliers. This paper proposes an intuitively appealing and easily computable robust measure of covariance based on the median and compares it with some existing robust covariance estimators in the statistics literature. It is demonstrated by simulations that all of the robust measures are fairly stable and insensitive to outliers. We apply robust covariance measures to construct two well-known portfolios, the minimum-variance portfolio and the optimal risky portfolio. The results of an out-of-sample experiment indicate that a potentially large investment gain can be realized using robust measures in place of the conventional measure.

1. Introduction

Covariance, a common measure of the linear dependence of two random variables, has played an important role in many academic fields, including finance (particularly for the construction of optimal portfolios). Since the seminal work of Markowitz (1952), a number of studies have proposed a wide range of portfolios. The fundamental reason to construct a portfolio is to achieve diversification and one can obtain more benefit from diversification when the assets included in the portfolio are less

* Corresponding author. Address: Yonsei University, School of Economics, 134 Shinchon-dong, Seodaemun-gu, Seoul 120-749, Republic of Korea. Fax: +82 2 2123 8638.

E-mail address: tae-hwan.kim@yonsei.ac.kr (T.-H. Kim).

1544-6123/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.frl.2012.06.001

correlated; that is, the lower the correlation, the greater benefit from diversification. Therefore, it is vital to carefully measure the correlation between asset returns before constructing portfolios. Usually, the correlation between asset returns is measured using the conventional sample covariance – apart from normalization by standard deviations.

Kim and White (2004) demonstrate that the conventional measures of skewness and kurtosis, which are based on sample averages, are highly sensitive to outliers. Hence, it can be conjectured that the conventional measure of covariance is also sensitive to outliers because it is computed as a sample average. In this paper, we will demonstrate both analytically and by Monte Carlo simulations that this conjecture indeed holds in the presence of outliers. Financial data, particularly data on asset returns, are widely known to include outliers, which typically result from financial crises such as the 9/11 attack and the 2008 global financial crisis. Therefore, simply applying the conventional measure of covariance to asset returns can lead to a misunderstanding of the comovement between assets. Since the covariance measure is one of the important inputs to the optimization of Markowitz portfolios, the use of the conventional measure can potentially cause a reduction in the performance of the resulting portfolios.

This paper proposes an intuitively appealing and easily computable robust measure of covariance, and we compare it with several existing robust covariance estimators from the statistics literature. To do this, we closely follow the main idea in Kim and White (2004), Bonato (2011), Ergun (2011) and White et al. (2010) by constructing the proposed measure to be based on the median rather than on averages. After comparing the conventional and robust measures through Monte Carlo simulations, we employ the robust measures to construct two well-known portfolios, the minimum-variance portfolio and the optimal risky portfolio, using return data obtained from Professor Kenneth R. French's website. Previous studies have noted the instability of portfolio optimization (e.g. Jobson and Korkie, 1980, 1981; Michaud, 1989), but they typically focused on the sensitivity and uncertainty in the mean and variance measures used in the optimization process. As a result, such studies have proposed some techniques for stabilizing the mean and variance measures (e.g. Jobson et al., 1979; Adrian and Brunnermeier, 2008, Kane et al., 2011). The present paper focuses mainly on the role of the covariance measure in the construction of optimal portfolios. The results of an out-of-sample experiment indicate that a large investment gain can be realized by using robust measures in place of the conventional measure.

2. Conventional measure of covariance

We consider two stochastic processes $\{x_t\}_{t=1,\dots,T}$ and $\{y_t\}_{t=1,\dots,T}$ where x_t are assumed to be IID (independent and identically distributed) with the CDF (cumulative distribution function) F_x and y_t are also assumed to be IID with the CDF F_y. The conventional measure of covariance (denoted by C) is given by

$$C = E[(x_t - \mu_x)(y_t - \mu_y)],$$

where the expectation E is taken with respect to the joint CDF of x_t and y_t, and μ_x and μ_y are the population means of x_t and y_t, i.e., $\mu_x = E(x_t)$, $\mu_y = E(y_t)$. The conventional measure C is, of course, a population parameter and thus must be estimated. Its usual estimation is achieved by replacing the population expectation E with its corresponding sample mean:

$$\widehat{C} = \widehat{E}\left[(x_t - \hat{\mu}_x)(y_t - \hat{\mu}_y)\right],$$

where $\widehat{E}$ is the sample mean operator, i.e., $\widehat{E} = \frac{1}{T}\sum_{t=1}^{T}$, and $\hat{\mu}_x$ and $\hat{\mu}_y$ are the sample means of x_t and y_t, i.e., $\hat{\mu}_x = \widehat{E}(x_t),\ \hat{\mu}_y = \widehat{E}(y_t)$.

The above sample covariance $\widehat{C}$ is based on the sample average and thus may be influenced by outliers in either x_t or y_t. To determine the influence of outliers, we assume without loss of generality that a single outlier occurs at time $[\tau T]$ with $\tau \in (0, 1)$ only in x_t. The size of the outlier is denoted by m_x; i.e., we replace $x_{[\tau T]}$ with $x_{[\tau T]} + m_x$ to inject the outlier into the sample.[1] With this single outlier, the sample covariance becomes

$$\widehat{C} = \widehat{C}_0 + \frac{m_x}{T}\left(y_{[\tau T]} - \hat{\mu}_y\right), \tag{1}$$

[1] Note that the function $[a]$ is the usual integer function taking the integer part of the real number a.