Computational Statistics and Data Analysis 56 (2012) 88–99

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

A Bayesian information criterion for portfolio selection

Wei Lan [a], Hansheng Wang [a,*], Chih-Ling Tsai [b]

[a] *Guanghua School of Management, Peking University, Beijing, 100871, PR China*
[b] *Graduate School of Management, University of California-Davis, CA, 95616-8609, USA*

ARTICLE INFO

Article history:
Received 6 April 2010
Received in revised form 8 May 2011
Accepted 11 June 2011
Available online 26 June 2011

Keywords:
Bayesian information criterion
Minimal variance portfolio
Portfolio selection
Risk diversification
Selection consistency

ABSTRACT

The mean–variance theory of Markowitz (1952) indicates that large investment portfolios naturally provide better risk diversification than small ones. However, due to parameter estimation errors, one may find ambiguous results in practice. Hence, it is essential to identify relevant stocks to alleviate the impact of estimation error in portfolio selection. To this end, we propose a linkage condition to link the relevant and irrelevant stock returns via their conditional regression relationship. Subsequently, we obtain a BIC selection criterion that enables us to identify relevant stocks consistently. Numerical studies indicate that BIC outperforms commonly used portfolio strategies in the literature.

1. Introduction

In financial risk analysis, Markowitz (1952) proposed mean–variance portfolio selection, and this landmark study earned him the 1990 Nobel Prize in Economic Sciences shared with Merton Miller and William Sharpe. Since then, risk diversification has become an increasingly important tool for analyzing investments; see for example, Jagannathan and Ma (2003) and DeMiguel et al. (2009a,b). This topic is particularly relevant given the current financial market turbulence, which motivates us to study portfolio selection.

In his seminal mean–variance theory, Markowitz (1952) uses two parameters to characterize a portfolio's performance: expected return and variance (i.e., risk). The investor, therefore, commonly optimizes his/her investment portfolio with an appropriate trade-off between expected return and risk. Because investors have different risk attitudes, the number of potentially optimal portfolios could be large. Among various optimal portfolios, the one with minimal variance is of particular interest for two reasons. First, the minimum-risk portfolio could be attractive to those investors with strong risk aversion characteristics (e.g., governments, pension funds). Second, although the criterion is minimal risk, the actual return remains competitive; see Jagannathan and Ma (2003) and DeMiguel et al. (2009a,b). This is because it is considerably more difficult to accurately estimate the mean of a stock's return than its variance; see Jorion (1986) and Jagannathan and Ma (2003). Testing a portfolio's mean–variance spanning is, therefore, important (Huberman and Kandel, 1987), and many researchers advocate for the minimal risk criterion in portfolio selection (Jagannathan and Ma, 2003).

To effectively employ the minimum-risk criterion in practice, one needs to accurately estimate covariance matrices. Following Markowitz (1952), diversification can reduce the overall risk of an investment portfolio, and this strategy naturally leads us to favor larger portfolios. However, this notion induces high-dimension covariance matrices, which are difficult to estimate accurately; see Bickel and Levina (2008) and Rothman et al. (2009). Furthermore, past empirical evidence suggests that the sampling error in the covariance estimation process can significantly deteriorate a portfolio's out-of-sample performance. Moreover, estimation errors also lead to considerable portfolio instability. Accordingly, portfolio weights need

* Corresponding author. Tel.: +86 10 6275 7915.
E-mail address: hansheng@gsm.pku.edu.cn (H. Wang).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2011.06.012

to be adjusted frequently and appreciably, which yields non-negligible transaction costs. Hence, a good strategy should take into account estimation errors.

In the past decade, some empirical researchers (Goetzmann and Kumar, 2001; Polkovnichenko, 2003; Statman, 2004) have found that investors tend not to hold many stocks in their portfolios; average portfolio size is 3 or 4 stocks. From statistical consideration, this finding is sensible since a smaller portfolio size requires a fewer number of unknown parameters to be estimated. This in turn reduces the estimation instability and subsequently brings down the transaction or holding cost (Statman, 2004). These findings motivate us to consider alleviating the estimation error effect by controlling the size of the portfolio. To this end, we define relevant stocks (i.e., stocks that must be included in the portfolio) and irrelevant stocks (i.e., stocks that cannot provide any additional risk reduction given the existing relevant stocks). The optimal portfolio is established by choosing relevant stocks that balance diversification and estimation error. Therefore, the aim of this paper is to develop a selection criterion that enables us to consistently differentiate relevant and irrelevant stocks.

The rest of the article is organized as follows. Section 2 defines relevant and irrelevant stocks and proposes a linkage condition to link the relevant and irrelevant stock returns via their conditional regression relationship. Accordingly, we obtain the Bayesian information criterion (BIC) and demonstrate its consistency (i.e., the capability to consistently differentiate relevant and irrelevant stocks). Section 3 presents numerical examples including Monte Carlo studies and an empirical analysis. The article concludes with a brief discussion in Section 4. All technical details are left to Appendix.

2. Bayesian information criterion

2.1. Relevant and irrelevant stocks

Let X_{tj} ($1 \le j \le d$) be the return of the jth stock observed at time t and $X_t = (X_{t1}, \ldots, X_{td})^\top \in \mathbb{R}^d$, where d is the number of candidate stocks. We further assume that the X_t's are independent and identically distributed random variables with $E(X_t) = 0$ and $\text{cov}(X_t) = \Sigma$ for $t = 1, \ldots, n$. To minimize the portfolio variance, one needs to find an optimal weight vector $\omega = (\omega_1, \ldots, \omega_d)^\top \in \mathbb{R}^d$, such that the variance $\text{var}(\omega^\top X_t) = \omega^\top \Sigma \omega$ can be minimized under the constraint $\omega^\top \mathbf{1} = 1$, where $\mathbf{1} = (1, 1, \ldots, 1)^\top \in \mathbb{R}^d$. It has been shown that the optimal solution to this minimization problem is $\omega_0 = (\omega_{01}, \ldots, \omega_{0d})^\top = \Sigma^{-1}\mathbf{1}\{\mathbf{1}^\top \Sigma^{-1}\mathbf{1}\}^{-1}$; see for example, Ledoit and Wolf (2003). To assess the out-of-sample performance, we consider $X_0 \in \mathbb{R}^d$ to be an independent copy of X_t. Then, the resulting portfolio's out-of-sample variance is $\text{var}(\omega_0^\top X_0) = (\mathbf{1}^\top \Sigma^{-1}\mathbf{1})^{-1}$.

For the sake of convenience, we introduce generic notation $\mathcal{S} = \{j_1, \ldots, j_{\tilde{d}}\}$ to represent the portfolio that includes the j_1th, j_2th, $\ldots$, $j_{\tilde{d}}$th stocks. We denote its size as $|\mathcal{S}| = \tilde{d}$. Let $\mathcal{S}_F = \{1, 2, \ldots, d\}$ be the full-size portfolio that contains all candidate stocks. In addition, for any d-dimensional vector $\beta \in \mathbb{R}^d$ and $d \times d$ matrix $\Omega \in \mathbb{R}^{d\times d}$, let $\beta_{(\mathcal{S})}$ and $\Omega_{(\mathcal{S})}$ represent their corresponding sub-vector and sub-matrix. Accordingly, the return vector of the portfolio $\mathcal{S}$ at time t and its covariance matrix are given by $X_{t(\mathcal{S})} = (X_{tj} : j \in \mathcal{S})^\top \in \mathbb{R}^{|\mathcal{S}|}$ and $\Sigma_{(\mathcal{S})} = (\sigma_{j_1 j_2})_{j_1, j_2 \in \mathcal{S}} \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}$, respectively. Moreover, for any two portfolios $\mathcal{S}_a$ and $\mathcal{S}_b$, we use the notation $\Sigma_{(\mathcal{S}_a, \mathcal{S}_b)}$ to represent the sub-matrix of Σ, where its rows and columns are determined by $\mathcal{S}_a$ and $\mathcal{S}_b$, correspondingly. For example, let $\hat{\Sigma} = n^{-1}\sum X_t X_t^\top = n^{-1}(\mathbb{X}^\top \mathbb{X})$ be the sample covariance matrix of the full-size portfolio $\mathcal{S}_F$, where $\mathbb{X} = (X_1, \ldots, X_n)^\top$. Then, $\hat{\Sigma}_{(\mathcal{S}_a, \mathcal{S}_b)} = (\hat{\sigma}_{j_1 j_2} : j_1 \in \mathcal{S}_a, j_2 \in \mathcal{S}_b) \in \mathbb{R}^{|\mathcal{S}_a|\times|\mathcal{S}_b|}$ is the sample covariance between $\mathcal{S}_a$ and $\mathcal{S}_b$. The difference between the subscript with parentheses and without parentheses is noteworthy. For example, $\omega_{0(\mathcal{S})}$ denotes a sub-vector of $\omega_0 \in \mathbb{R}^d$, where ω_0 is the optimal weight vector associated with the full-size portfolio $\mathcal{S}_F$. On the other hand, $\omega_{0\mathcal{S}} = \{\Sigma_{(\mathcal{S})}^{-1}\mathbf{1}_{(\mathcal{S})}\}\{\mathbf{1}_{(\mathcal{S})}^\top \Sigma_{(\mathcal{S})}^{-1}\mathbf{1}_{(\mathcal{S})}\}^{-1}$ is the optimal weight vector computed via the portfolio $\mathcal{S}$ only, which leads to $\omega_0 = \omega_{0\mathcal{S}_F}$.

Inspired by mean–variance spanning theory (Huberman and Kandel, 1987; Gibbons et al., 1989; Kan and Zhou, 2001), we next define a stock to be relevant (irrelevant) if its corresponding weight in ω_0 is non-zero (zero). Then the optimal portfolio is $\mathcal{S}_0 = \{j : \omega_{0j} \neq 0\}$ with size $d_0 = |\mathcal{S}_0|$, while its complement is $\mathcal{S}_0^c = \mathcal{S}_F \setminus \mathcal{S}_0$ with size $|\mathcal{S}_0^c| = d - d_0$. Although the relevant and irrelevant stocks are clearly defined, they are not directly useful for constructing the likelihood function of the portfolio. This is because the conditional regression relationship between the relevant and irrelevant stocks is not explicitly specified. To this end, we obtain the following theorem, whose detailed technical proof can be found in Appendix A.1.

Theorem 1. *Assume that X_t follows a multivariate normal distribution for $t = 1, \ldots, n$. Then, a necessary and sufficient condition for $\mathcal{S} \supset \mathcal{S}_0$ is that, for any $k \notin \mathcal{S}$, we have $\sum_{j\in\mathcal{S}} \beta_{kj} = 1$, where β_{kj} are regression coefficients of X_{tk} on $\{X_{tj}, j \in \mathcal{S}\}$.*

The above theorem indicates that the condition $\sum_{j\in\mathcal{S}} \beta_{kj} = 1$ is crucial in determining whether $\mathcal{S} \supset \mathcal{S}_0$. To further understand this condition, we present an insightful discussion below. For an arbitrary portfolio $\mathcal{S}$ and any given stock $k \notin \mathcal{S}$, we have

$$X_{tk} = \sum_{j\in\mathcal{S}} X_{tj}\beta_{kj} + \varepsilon_{tk}, \qquad (2.1)$$

where (2.1) is stated in Appendix A.1 for the proof of Theorem 1. It is noteworthy that the error term ε_{tk} is assumed to be independent of X_{tj} for $j \in \mathcal{S}$, and that such an assumption is crucial for the implementation of our proposed method. Similar assumption has been used in the mean–variance spanning literature; see for example Huberman and Kandel (1987).