Testing linear factor pricing models with individual securities in Japan: Application of shrinkage estimation

Ryohei Oishi*
Hitotsubashi University

Abstract

This study proposes a multivariate test for linear factor asset pricing models when the number of assets, N, is larger than the time dimension of returns, T. We extend the exact test proposed by Gibbons et al. (1989) to obtain a non-singular covariance matrix with fewer estimation errors in the case of $T < N$. We apply the shrinkage estimation to the covariance matrix of the idiosyncratic error and obtain the sampling distribution of the proposed test statistics by using the fixed-design wild bootstrap to address the conditional heteroskedasticity and cross-sectional correlation. Our Monte-Carlo evidence shows that the proposed test statistics have a satisfying size and power. By using the proposed test statistics, we find that the Tokyo Stock Price Index is not mean-variance efficient most of the time from 1985 to 2015 compared with the tangency portfolio constructed from the individual securities of the Tokyo Stock Exchange first section, while efficiency hypothesis of the market portfolio in the second section is not rejected during periods of Asian financial crisis and the global financial crisis.

Keywords: CAPM, multivariate test, market efficiency, large dimensional panel, wild bootstrap, shrinkage estimation

JEL Codes: C12, C15, C23, G12

*Hitotsubashi University, Graduate School of Economics, 2-1 Naka, Kunitachi, Tokyo, Japan 186-8601 (em171008@g.hit-u.ac.jp)

1 Introduction

In modern portfolio theory, which started with the seminal work of Markowitz (1952), many academic researchers have examined the relationships between the return and risk, or volatility, of financial assets. The most utilized and widely known asset pricing models in academia and practice are the capital asset pricing model (CAPM) developed by Sharpe (1964), Lintner (1965), termed the SL-CAPM hereafter, as well as arbitrage pricing theory (APT) proposed by Ross (1976). From an academic point of view, these models are fundamental to asset pricing theory, as the CAPM and APT are purely derived from the partial equilibrium analysis and no-arbitrage conditions, respectively and are used in the valuation of enterprises, evaluation of fund managers, and prediction of stock returns in practice.

Despite the thorough theoretical conclusions presented thus far, they should always be tested empirically. Among the statistical tests of the above asset pricing models, one of the most successful is the exact test proposed by Gibbons et al. (1989), the GRS test hereafter. These authors propose an exact test based on an F-distribution under the assumption of the i.i.d. Gaussian error and $T > N - k - 1$, where T and N are the number of time dimensions and assets, respectively and k is the number of factors. They have also shown that the test is equivalent to testing if a given portfolio is mean-variance efficient. The crucial assumption of the GRS test is that the number of time-series data must be larger than the number of assets included in the test. If this assumption is violated, the estimated covariance matrix of the error becomes singular and the GRS test statistics can no longer be calculated. The assumption of $T > N - k - 1$ heavily restricts the test of linear asset pricing models. Most conventional research overcomes this problem by sorting individual securities into portfolios by using certain criteria. Despite its popularity in financial research, however, dealing with the problem of $T < N$ by sorting portfolios poses several problems. Firstly, by sorting securities based on a criterion that might be correlated with the return, the type-1 error becomes large (see Lo and MacKinlay (1990) and Berk (2000)). Secondly, considering the identification of what causes the rejection of the asset pricing model between the market inefficiency and instability of the coefficients, it is thus desirable to test with shorter T, which opposes the assumption $T > N-k-1$ in the case of relatively large N. Finally, including more assets may improve the power of the statistics, which also faces limitation due to the assumption of $T > N - k - 1$.

In this study, we propose a robust test of linear factor pricing models under $T < N$. We extend the GRS test statistics by applying the shrinkage estimation method to have a non-singular covariance matrix estimation. We call our method the shrinkage-GRS (S-GRS) test. The shrinkage estimation method proposed by Ledoit and Wolf (2003), Ledoit and Wolf (2004a), Ledoit and Wolf (2004b), and Bodnar et al. (2014) provides an estimator of the covariance matrix with no amplification of the error in inverting it, and a non-singular covariance matrix even in the case of $T < N$. If one just needs an invertible covariance matrix in case of $T < N$, putting several restrictions from theory or intuition on estimation is enough. On the contrary, by applying the shrinkage estimation to our proposed test statistics, two types of estimators of covariance matrix can be incorporated: one is the sample covariance matrix estimated purely from the data, and another is the covariance matrix estimated with several theoretical or intuitive restrictions. Therefore, the shrinkage estimate can optimally balance between a researcher's intuition and the data. However, because the sampling distribution of the proposed test statistics no longer exactly follows the F-distribution, we apply the fixed-design wild bootstrap presented by Gonçalves and Kilian (2004) to address the conditional heteroskedasticity and cross-sectional correlation in the idiosyncratic error. As a result, the S-GRS test can be calculated in the case of $T < N$ and is robust to the non-Gaussian error distribution with a wide range of conditional heteroskedasticity.

This study makes two main contributions. First, the proposed S-GRS test is applicable in the case of $T < N$, where the standard GRS test cannot be calculated. Second, to our best knowledge, this is the first study to analyze the efficiency of the Tokyo Stock Price Index (TOPIX) with data on individual securities in the Tokyo Stock Exchange (TSE) first section and second section, not with the portfolio data.

The remainder of this article is organized as follows. Section 2 briefly shows the SL-CAPM as well as the GRS test and the econometric assumption behind it. Section 3 describes the shrinkage estimation method and the procedure of the proposed S-GRS test. We analyze the empirical size and power of the S-GRS test in a Monte-Carlo simulation in Section 4. In Section 5, we apply the S-GRS test to analyze the efficiency of TOPIX with data on individual securities from TSE first and second section. Section 6 concludes.

2 The GRS Test

In this section, we formally derive the GRS test and review its assumptions and extension. We briefly discuss the motivation and difference of the proposed S-GRS test.

2.1 Derivation of the GRS Test

For simplicity, we discuss single factor models in this study; however, the same discussion holds in the case of multi-factor linear pricing models at the cost of more complex notation. Suppose there exists a risk-free asset whose return at time t is $r_{f,t}$. In the investment universe, there exist N assets whose return is denoted by $r_{i,t}$ for all i individual assets and the return of the market portfolio at time t is $r_{m,t}$. If the SL-CAPM exactly holds or the market portfolio is mean-variance efficient,[1]

$$\mathbb{E}[\tilde{r}_{i,t}] = \beta_i \tilde{r}_{m,t} \tag{1}$$

where $\tilde{r}_{i,t} = r_{i,t} - r_{f,t}$ and $\tilde{r}_{m,t} = r_{m,t} - r_{f,t}$ are the excess returns of security i and the market portfolio, respectively. Here, beta, the return sensitivity of asset i toward the market portfolio, can be calculated as $\beta_i = \frac{Cov(\tilde{r}_{i,t}, \tilde{r}_{m,t})}{Var(\tilde{r}_{m,t})}$. To test equation (1), we set the econometric model as

$$\tilde{r}_{i,t} = \alpha_i + \beta_i \tilde{r}_{m,t} + \tilde{\varepsilon}_{i,t} \tag{2}$$

or by stacking the cross-section into a vector,

$$\tilde{\boldsymbol{r}}_{\boldsymbol{t}} = \boldsymbol{\alpha} + \tilde{r}_{m,t}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}} \tag{3}$$

where $\tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}} \overset{iid}{\sim} \boldsymbol{N}(0, \boldsymbol{\Sigma})$ with symmetric positive definite covariance matrix $\boldsymbol{\Sigma}$, and the null hypothesis to test equation (1) is

$$H_0 : \alpha_i = 0 \; \forall i = 1, ..., N \tag{4}$$

Note that this is the joint hypothesis of all the mis-pricing of the factor asset pricing models. Thus, if there exists at least one individual asset that is not correctly priced, this null hypothesis will be rejected. Under the assumption of the i.i.d. Gaussian error and $T > N - 2$, Gibbons et al. (1989) show that the Wald-type statistics for the hypothesis (4)

$$J = \hat{\boldsymbol{\alpha}}' Var(\hat{\boldsymbol{\alpha}})^{-1} \hat{\boldsymbol{\alpha}} \tag{5}$$

[1] For explicit proof that the test of the SL-CAPM is equivalent to testing the mean-variance efficiency of the market portfolio, see Appendix A or Gibbons et al. (1989). This test can be basically understood as comparing the sharpe ratio of a given portfolio on the right-hand side with that of the ex-post tangency portfolio constructed from the assets on the left-hand side.

exactly follow the F-distribution with N degrees of freedom in the numerator and $(T-N-2)$ degrees of freedom in the denominator, which is known as the GRS test:

$$J = \frac{T-N-1}{N}\left[1+\frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2}\right]^{-1}\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\alpha}} \sim F_{N,T-N-1} \tag{6}$$

where $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ are the maximum likelihood estimates from equation (3), $\hat{\mu}_m = \frac{1}{T}\sum_{t=1}^{T}\tilde{r}_{m,t}$, $\hat{\sigma}_m^2 = \frac{1}{T}\sum_{t=1}^{T}(\tilde{r}_{m,t}-\hat{\mu}_{m,t})^2$, and $\hat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}(\tilde{\boldsymbol{r}}_{\boldsymbol{t}}-\hat{\boldsymbol{\alpha}}-\hat{\boldsymbol{\beta}}\tilde{r}_{m,t})(\tilde{\boldsymbol{r}}_{\boldsymbol{t}}-\hat{\boldsymbol{\alpha}}-\hat{\boldsymbol{\beta}}\tilde{r}_{m,t})'$.

The simplest example in which the GRS test is employed is to check the validity of asset pricing models. Fama and French (1993) establish the notable Fama-French three-factor model (FF3 model hereafter) and check its explanatory power and model fit with the GRS test. Further, Fama and French (2015) analyze the Fama–French five-factor model. Several researchers focus on the characteristics of the GRS test as a tool to discuss the mean-variance efficiency of a portfolio. Detzler and Wiggins (1997) extend the GRS statistics when short-selling is restricted and evaluate the performance of global investment funds. Grinold (1992) analyzes the efficiency of several indices (ALLORDS, DAX, TOPIX, FTA, S&P500) and finds that none the indices except for DAX satisfy mean-variance efficiency. He concludes that it is possible to construct actively managed portfolios that are more efficient than indices.

2.2 Assumptions of the GRS Test

Although empirical research has adopted the GRS test, several statistical assumptions behind it exist. In this section, we explicitly discuss those assumptions and compare them with the empirical and intuitive facts; thereafter, we review how we can avoid problems caused by the violation of such assumptions. The first assumption is that the idiosyncratic error follows the i.i.d. Gaussian error and the second is $T > N - k - 1$, where k is the number of factors in the model.

In contrast to the assumption of the Gaussian i.i.d. error, it is widely known that stock returns follow a fat-tailed distribution, which implies that the probability of a crash is higher than that predicted by the Gaussian distribution.

Concerning the problem of the non-Gaussian error, Affleck-Graves and McDonald (1989) show that the GRS test is robust to the typical non-normality observed in financial markets by adopting simulation methods. Zhou (1993) expands the GRS test in the case of elliptical distributions (e.g., Gaussian distribution, t-distribution, or mixed Gaussian distribution). The author shows that the standard GRS test rejects the null hypothesis in the United States, while the proposed GRS test does not. Chou and Zhou (2006) apply bootstrapping to the standard GRS test and generalized method of moments-based GRS test to avoid the parametric assumptions on the idiosyncratic error.

Another inconvenient fact is the existence of volatility clustering or conditional heteroskedasticity of the idiosyncratic error, which violates the i.i.d. assumption. Since the development of the autoregressive conditional heteroskedasticity (ARCH) and generalized ARCH (GARCH) models, empirical evidence has shown that high volatility periods and low volatility periods exist in financial markets, which can even be asymmetric.

Regarding the conditional heteroskedasticity in the idiosyncratic error, MacKinlay and Richardson (1991) derive the GRS test with the generalized method of moments framework under weaker assumptions. Beaulieu et al. (2007) propose a simulation-based GRS test when the distribution of the idiosyncratic error is completely known or known up to the latent parameters. Gungor and Luger (2009) propose distribution-free test statistics based on the sign statistics and Wilcoxon signed rank statistics for the single factor model. They take the maximum or squared sum of the test statistics for each individual equation to evaluate the joint hypothesis. They show that the proposed statistics have higher power than the GRS test or the methods proposed in Beaulieu et al. (2007) by simulation.

Another problematic assumption of the GRS test is $T > N-k-1$. This assumption is required to calculate the non-singular covariance matrix of the idiosyncratic error, which usually has a cross-sectional correlation. On the contrary, in standard financial markets, the number of individual securities is larger than the time-series data.[2] One may use intra-daily data to have large T; however, this might provide an unstable coefficient estimate. Many empirical researchers have overcome this problem by sorting individual securities based on certain criteria and constructing a portfolio. However, grouping individual securities into portfolios may offset the positive and negative α_i of the individual security and decrease the power of the test statistics. Berk (2000) warns researchers about using a variable that may be correlated to the expected asset returns when sorting assets into portfolios to test asset pricing models. The author theoretically shows that if researchers use a sorting variable assumed to be correlated with the expected return, tests over-reject the asset pricing model even if it is true. Lo and MacKinlay (1990) claim that data-snooping bias in the test of asset pricing seriously distorts inferences. Thus, even if the test rejects the null hypothesis that all the securities are priced correctly, researchers cannot identify if it is caused by the mis-pricing of the model or just due to the larger probability of type-1 error than that set by the researcher. From the viewpoint of portfolio managers investing in a given investment universe, it would be preferable to judge their performance with a comparison of individual securities, not with a portfolio constructed from ad-hoc criteria. As the SL-CAPM is used to calculate the required rate of return of a firm in corporate finance,[3] it is worth checking if the linear factor pricing model is empirically supported by individual-level data. We would also like to identify the inefficiency of the bench-

[2] In the case of the TSE first section, there were 2055 individual securities on December 15, 2017, according to the Japan Exchange Group. If the data are monthly, one has to obtain more than 170 years.

[3] See, for instance, Firer et al. (2012, chap. 14).

mark portfolio from the rejection of the model caused by the time-invariant coefficients in the estimation. In this case, testing with shorter T and large N is desirable. It is shown that, however, the power of the standard GRS test does not increase uniformly as N increases in Gungor and Luger (2016). Thus, from the viewpoint of the standard GRS test, short T is not desired at all.

Affleck-Graves and McDonald (1990) apply the maximum entropy estimation for the covariance matrix of the idiosyncratic error, which can provide a non-singular covariance matrix estimation even in the case of $T < N$, and obtain the distributions of the test statistics with the bootstrap method. The approach of Gungor and Luger (2009) can also be used for the single factor model. Gungor and Luger (2013) expand the analysis in Gungor and Luger (2009) to multi-factor pricing models. They split the data set into two subsamples and estimate the pricing model in the first half. They then determine the weights for each security according to the sign of $\hat{\alpha}_i$ and combine all securities into one portfolio to avoid the $T < N$ problem. Moreover, they use the sign statistics to have distribution-free test statistics. Pesaran and Yamagata (2012) propose asymptotic test statistics if the error is Gaussian or not and if the cross-sectional correlation is zero or not by applying the central limit theorem in Kelejian and Prucha (2001). Hwang and Satchell (2014) propose averaging the squared t-statistics of α_i, the intercept of equation (3). These average F-test statistics follow the F-distribution provided that the idiosyncratic error is not correlated in a cross-section, which is a strong assumption. By using simulation and theory, they find that the size and power of the proposed test statistics are distorted if a cross-sectional correlation of the idiosyncratic error exists. Gungor and Luger (2016) expand the average F-test statistics in Hwang and Satchell (2014) by taking several norms of the F-statistics from each equation (3). The proposed test statistics have higher power than those of Pesaran and Yamagata (2012) when the cross-sectional correlation is high.

There are several differences in the approach of this study to solving the singular covariance matrix problem when $T < N$. Firstly, we apply the shrinkage estimator proposed by Ledoit and Wolf (2003), Ledoit and Wolf (2004a), Ledoit and Wolf (2004b), and Bodnar et al. (2014), which can provide a non-singular covariance matrix with no amplification of the estimation error in inverting. We also apply the fixed-design wild bootstrap in Gonçalves and Kilian (2004), meaning that the S-GRS test is robust to non-normality and a wide range of conditional heteroskedasticity, and does not ignore the cross-sectional correlation. Thanks to this condition, the S-GRS test may increase the power of the statistics even when $T > N$. Secondly, there is no crucial change in the framework of the GRS test except for the application of the shrinkage estimator. Thus, the estimation procedure and interpretation of the result are standard. Thirdly, researchers can incorporate their prior information from theory or intuition about the structure of the covariance matrix for the S-GRS test because of the shrinkage estimator, which makes the S-GRS test take a balance between restrictions, which can be ad-hoc in some cases just for the non-singular estimation of the covariance matrix, and the data-sets at your hand.

3 S-GRS Test Procedure

In this section, we provide the algorithm of the proposed S-GRS test. Specifically, we briefly discuss the shrinkage estimation and how to apply it to the S-GRS test.

3.1 Shrinkage Estimation Method

The sample covariance matrix has huge estimation errors[4] when N is close to T and it even becomes singular in the case of $T < N$. The idea of shrinkage estimation is

[4] See, for example, Frankfurter et al. (1971)

classical in the field of statistics from the work of Stein et al. (1956). By using the shrinkage estimation method for the covariance matrix, one can obtain a non-singular estimate of a covariance matrix with no amplification of the estimation error when inverting it even when $T < N$.

The idea of shrinkage estimation is intuitive: combine two extreme estimates and obtain something better. More rigorously, consider two types of covariance matrix estimators. One is $\boldsymbol{S}$, a consistent estimator of a covariance matrix with zero bias and large variance (i.e., the sample covariance matrix), and the other is $\boldsymbol{\Sigma}_T$, an estimator of a covariance matrix with bias and small variance, which can be obtained by placing certain restrictions on the estimator. We take a weighted average of these two and obtain the shrinkage estimator of the covariance matrix ($\hat{\Sigma}^s$) as follows:

$$\boldsymbol{\Sigma}_s = \delta_T \boldsymbol{\Sigma}_T + \delta_S \boldsymbol{S} \tag{7}$$

Here δ_S and δ_T are the weights, called the shrinkage intensity, and $\boldsymbol{\Sigma}_T$ is the shrinkage target.

The next problem is how to choose the optimal intensity.[5] The derivation of the optimal shrinkage intensity generally takes two steps. First, one obtains the theoretical shrinkage intensity from the MSE minimization problem, which contains the unknown true parameters. Second, one shows that each element of the theoretically optimal shrinkage intensity can be estimated consistently under general asymptotics where both T and N can go to infinity but N is slower than T. Note that Ledoit and Wolf (2004b) point out that those results are well approximated in finite samples through simulation analysis. By considering the MSE, we can explicitly discuss the bias–variance trade-off. Suppose we are interested in a scalar parameter θ and estimate it as $\hat{\theta}$. The MSE of

[5]The choice of $\boldsymbol{\Sigma}_T$ is discussed later in this subsection and in the next subsection in more detail.

the estimator $\hat{\theta}$ can be written as

$$MSE(\hat{\theta}) \equiv \mathbb{E}[(\hat{\theta} - \theta)] = \mathbb{E}[(\hat{\theta} - \mathbb{E}_{\hat{\theta}}(\theta))^2] + (\mathbb{E}(\theta) - \theta)^2 \equiv Bias(\hat{\theta})^2 + Var(\hat{\theta}) \quad (8)$$

Think about two types of estimators. One is the unbiased estimator, $\hat{\theta}^u$, and the other is not unbiased but has the least variance $\hat{\theta}^{nu}$. In the case of $\hat{\theta}^u$, the first term of equation (8) is zero, while it may have a large second term. While an unbiased estimator is obtainable, this might lead to an over-fitting problem. In the case of $\hat{\theta}^{nu}$, it may be unable to completely explain the statistical characteristics behind the data while it has lower estimation error by definition. Thus, considering only minimizing either the bias or the variance may pose a practical problem. However, by minimizing the MSE, as shown in equation (8), one can have a better estimator balancing the bias and the variance.

Ledoit and Wolf (2004a), Ledoit and Wolf (2004b), and Ledoit and Wolf (2003) analytically derive the optimal shrinkage intensity for the shrinkage target as the identity matrix, the SL-CAPM, and constant correlation model,[6] respectively. Bodnar et al. (2014) derive the shrinkage intensity under a weaker assumption as well as specify the shrinkage target, which does not have to be identity but rather any positive definite matrix provided there exists $sup_n \|\mathbf{\Sigma}_T\|_{\text{Tr}} \leq M$ with finite M, where $\|\cdot\|_{\text{Tr}}$ denotes the trace norm calculated as $\|A\|_{\text{Tr}} \equiv \text{Tr}(\sqrt{(AA')})$. Here, $\text{Tr}(\cdot)$ denotes the trace of a matrix. Through the simulation analysis, they point out that setting the shrinkage target as an identity matrix following Ledoit and Wolf (2004b) may be too conservative and that replacing it with a more realistic target can improve the estimation. They also show that their general shrinkage intensity minimizes the loss function almost surely, while Ledoit and Wolf (2004b) achieve that with quadratic mean convergence.

[6]This specification is discussed in the next subsection.

In this section, we follow the discussion in Ledoit and Wolf (2004b) to offer a general idea of the shrinkage estimation. In the application of S-GRS in the next section, we follow the method proposed by Ledoit and Wolf (2004b) and Schäfer and Strimmer (2005) as their shrinkage target is the same as ours.

The MSE minimization problem we solve here is

$$\min_{\delta_T,\delta_S} \mathbb{E}[||\boldsymbol{\Sigma}_s - \boldsymbol{\Sigma}_0||_F^2] \ s.t. \ \boldsymbol{\Sigma}_s = \delta_T \boldsymbol{I} + \delta_S \boldsymbol{S} \tag{9}$$

where $\|\cdot\|_F$ denotes the Frobenius norm calculated as $\|\boldsymbol{A}\| \equiv \sqrt{\mathrm{Tr}(\boldsymbol{A}\boldsymbol{A}')/N}$, $\boldsymbol{I}$ is the conformable identity matrix, and $\boldsymbol{\Sigma}_0$ is the true covariance matrix. Note that the shrinkage target in the analysis of Ledoit and Wolf (2004b) is the identity matrix. Solving the equation (9) provides the theoretical optimal shrinkage intensity, δ_I^* and δ_S^* as follows:

$$\delta_T^* = \frac{b^2}{d^2} m \tag{10}$$

$$\delta_S^* = \frac{a^2}{d^2} \tag{11}$$

where $a^2 + b^2 = d^2$ is guaranteed and $m =< \boldsymbol{\Sigma}_0, \boldsymbol{I} >, a^2 = \|\boldsymbol{\Sigma}_0 - m\boldsymbol{I}\|^2, b^2 = E\|\boldsymbol{S} - \boldsymbol{\Sigma}_0\|^2, d^2 = E\|\boldsymbol{S} - m\boldsymbol{I}\|^2$ and $< \cdot >$ is the inner product calculated as $< \boldsymbol{A}_1, \boldsymbol{A}_2 > \equiv \mathrm{Tr}(\boldsymbol{A}_1\boldsymbol{A}_2')/N$. The scalars in equation (10) and (11) can be interpreted as follows: b^2 is the MSE of the sample covariance matrix. If this is large, estimation error of the sample covariance matrix is high, which leads to decrease the weight on the sample covariance matrix. a^2 is the difference between the true covariance matrix and the identity matrix. If this is small, the restriction put on the shrinkage target seems to be appropriate, which leads to increase the weight on the shrinkage target.

The author shows that the scalars, a^2, b^2, d^2, and m, in equation (10) and (11) can be consistently estimated under general asymptotics. Thus, we can obtain the consistent optimal shrinkage estimates, $\hat{\delta}_T^*$ and $\hat{\delta}_S^*$. Therefore, the optimal shrinkage estimator of the covariance matrix is

$$\hat{\mathbf{\Sigma}}_s^* = \hat{\delta}_T^* \boldsymbol{I} + \hat{\delta}_S^* \boldsymbol{S} \tag{12}$$

Ledoit and Wolf (2003) calculate the shrinkage covariance matrix of asset returns by using a theoretical covariance matrix structure from the CAPM as a shrinkage target. They assume that asset returns follow equation (2). Then, the covariance matrix with restriction $\mathbf{\Phi}$ can be calculated as

$$\mathbf{\Phi} = \sigma_m^2 \boldsymbol{\beta}\boldsymbol{\beta}' + \mathbf{\Delta} \tag{13}$$

where σ_m^2 is the population variance of the excess return of the market portfolio, $\boldsymbol{\beta}$ is the vector of the population beta, and $\mathbf{\Delta}$ is the diagonal matrix with $Var(\tilde{\epsilon}_{i,t})$ for each i. This can be estimated by

$$\mathbf{\Sigma}_T^{CAPM} = \hat{\sigma}_m^2 \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}' + \boldsymbol{D} \tag{14}$$

where $\hat{\sigma}_m^2$ is the same as in Section 2, the consistent estimator of the variance of the market excess return, $\hat{\boldsymbol{\beta}}$ is the OLS estimate of equation (3), and $\boldsymbol{D}$ is the diagonal matrix of the unbiased sample estimate of $Var(\tilde{\epsilon}_{i,t})$, which can be calculated from the residuals. They show that the out-of-sample volatility of the global minimum variance portfolio (GMVP) optimized by using the shrinkage covariance estimation with target $\mathbf{\Sigma}_T^{CAPM}$ from equation (14) is lower than the other GMVP whose covariance matrix is estimated by using the industry factor model, the principal component model, and so forth. Note that the GMVP only uses information on the covariance matrix to determine the optimal weight of each asset. Thus, comparing the out-of-sample volatility of several

GMVPs can shed light on which estimation method is suitable practically.

Ledoit and Wolf (2004a) set the shrinkage target from the constant correlation model. They optimize portfolios by employing several estimates of the covariance matrix and show that the information ratio[7] of a portfolio with a shrinkage covariance matrix is at least as good as the shrinkage covariance matrix presented by Ledoit and Wolf (2003).

If one just would like to obtain an invertible estimate of a covariance matrix, putting several restrictions on the estimator such as sparsity is enough. However, by applying shrinkage estimator, researchers can incorporate estimate of a covariance matrix with theoretical or intuitive restrictions with a sample covariance matrix, which is estimated purely from the data. These characteristics, taking an optimal balance between restriction and data, motivate us to apply the shrinkage estimator on our proposed test statistics.

In the next subsection, we present several possible shrinkage targets and choose two appropriate targets for the S-GRS test statistics.

3.2 Choice of Shrinkage Targets

Schäfer and Strimmer (2005) provide the following examples of shrinkage targets:

(A) Identity matrix (0)

(B) Homoskedastic matrix (1)

$\mathbf{\Sigma}_{T(i,j)} = \overline{s_{ii}}$ if $i = j$, $\mathbf{\Sigma}_{T(i,j)} = 0$ if $i \neq j$

(C) Homoskedastic and common-covariance matrix (2)

$\mathbf{\Sigma}_{T(i,j)} = \overline{s_{i,i}}$ if $i = j$, $\mathbf{\Sigma}_{T(i,j)} = \overline{s_{i,j}}$ if $i \neq j$

[7]The information ratio is defined as the ratio of active returns (expected excess returns from benchmark returns) divided by active risk (the standard deviation of active returns).

(D) Heteroskedastic matrix (N)

$\mathbf{\Sigma}_{T(i,j)} = s_{i,i}$ if $i = j$, $\mathbf{\Sigma}_{T(i,j)} = 0$ if $i \neq j$

(E) Perfect positive correlation (N)

$\mathbf{\Sigma}_{T(i,j)} = s_{i,i}$ if $i = j$, $\mathbf{\Sigma}_{T(i,j)} = \sqrt{s_{i,i}s_{j,j}}$ if $i \neq j$

(F) Constant correlation (N+1)

$\mathbf{\Sigma}_{T(i,j)} = s_{i,i}$ if $i = j$, $\mathbf{\Sigma}_{T(i,j)} = \bar{r}\sqrt{s_{i,i}s_{j,j}}$ if $i \neq j$

where $\mathbf{\Sigma}_{T(i,j)}$ denotes the (i, j) element of the shrinkage target, $s_{i,j}$ is the (i, j) element of the sample covariance matrix, $\overline{s_{i,j}}$ is the average of $s_{i,j}$ for all (i, j), and $\bar{r}$ is the average of sample correlations. Above, the number of parameters to be estimated is in the parentheses.

The choice of shrinkage target depends on the researcher, but it only affects the extent to which the MSE will be reduced, not the problem of the singular covariance matrix in $T < N$. As Schäfer and Strimmer (2005) point out, (A), (B), and (D) always provide a positive definite shrinkage estimator of the covariance matrix, while the others do not. The simplest and most complicated targets applicable to the S-GRS test seem to be (A) and (D), respectively.

In the S-GRS test, we apply this shrinkage estimation to the covariance matrix of the idiosyncratic error. It may be too conservative to assume (A); it is more realistic that we allow each security to have a different volatility of the idiosyncratic error. Thus, we apply (A) and (D) in the S-GRS test following the analytical solution of the shrinkage intensity proposed by Ledoit and Wolf (2004b) and Schäfer and Strimmer (2005).

3.3 S-GRS Test Algorithm

We test the SL-CAPM described in equation (3):

$$\tilde{\boldsymbol{r}}_{\boldsymbol{t}} = \boldsymbol{\alpha} + \tilde{r}_{m,t}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}} \tag{3}$$

and the null hypothesis is (4):

$$H_0 : \alpha_i = 0 \ \forall i = 1, ..., N \tag{4}$$

As discussed in Section 2, the estimate of the covariance matrix of the idiosyncratic error becomes singular when $T < N$ and the estimation accuracy worsens as N rises. We apply the shrinkage estimator of the covariance matrix of the idiosyncratic error to generate a singular covariance matrix estimate with no amplification of the error in inverting. The S-GRS test can be implemented in the following algorithm:

(J^s.1) Estimate $\tilde{\boldsymbol{r}}_{\boldsymbol{t}} = \boldsymbol{\alpha} + \tilde{r}_{m,t}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}}$ and obtain the OLS estimate of α_i and β_i for each i, and construct the residual matrix ($\boldsymbol{X}_r$).

(J^s.2) Calculate the sample variance covariance matrix ($\boldsymbol{S} = \frac{1}{T}\sum_{t=1}^{T}(\tilde{\boldsymbol{r}}_{\boldsymbol{t}} - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\beta}}\tilde{r}_{m,t})(\tilde{\boldsymbol{r}}_{\boldsymbol{t}} - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\beta}}\tilde{r}_{m,t})'$).

(J^s.3) Calculate the optimal shrinkage intensity and obtain $\hat{\boldsymbol{\Sigma}}_s^* = \hat{\delta}_T^*\boldsymbol{\Sigma}_T + \hat{\delta}_S^*\boldsymbol{S}$. In this research, we follow Ledoit and Wolf (2004b) and Schäfer and Strimmer (2005) for the case of the identity target (A) and heteroskedastic target (D), respectively.

(J^s.4) The S-GRS test statistics can be calculated as follows:

$$J^s = \frac{T-N-1}{N}\left[1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2}\right]^{-1}\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\Sigma}}_s^{*-1}\hat{\boldsymbol{\alpha}} \tag{15}$$

where $\hat{\mu}_m = \frac{1}{T}\sum_{t=1}^{T}\tilde{r}_{m,t}$, $\hat{\sigma}_m^2 = \frac{1}{T}\sum_{t=1}^{T}(\tilde{r}_{m,t} - \hat{\mu}_{m,t})^2$

(J^s.5) Apply the fixed-design wild bootstrap of Gonçalves and Kilian (2004) to mimic the sampling distribution, J^s, as the asymptotic distribution of J^s is unknown.

(J^s.6) Use it to find the critical value and test the null hypothesis.

In this algorithm, researchers have to choose the shrinkage target $\mathbf{\Sigma}_T$. While this choice can be ad-hoc, the choice of the shrinkage target does not affect the problem of the singular covariance matrix in the GRS test. It only affects the reduction of the MSE. Schäfer and Strimmer (2005) note that "in fact any target will lead to a reduction in MSE, albeit only a minor one in case of a strongly misspecified target." On the contrary, as Bodnar et al. (2014) suggest, the identity matrix as a shrinkage target may be too conservative. Thus, in this study, we try both the identity matrix (A) and the heteroskedastic covariance matrix with zero off-diagonals (D) as a shrinkage target. In the empirical analysis, we present the result of the former shrinkage target and use the latter as a robustness check. We find no contradictory evidence about the inefficiency of the market portfolio in either case.

Once the shrinkage estimate of the covariance matrix is used in equation (15), the test statistics no longer follow an F-distribution. To obtain the unknown distribution of the S-GRS test, in the (J^s.5) we apply the fixed-design wild bootstrap in Gonçalves and Kilian (2004). This bootstrap method is theoretically valid under a wide range of conditional heteroskedasticities including asymmetric GARCH errors. This bootstrap method also takes into account the non-normality and cross-sectional correlation. The following explains the implementation of the fixed-design wild bootstrap in (J^s.5) for the S-GRS test setting.

(WB. 1) The obtained residual matrix ($\boldsymbol{X}_r$) is as follows:

$$\begin{pmatrix} \hat{\varepsilon}_{1,1} & \hat{\varepsilon}_{2,1} & \hat{\varepsilon}_{3,1} & \cdots & \hat{\varepsilon}_{N,1} \\ \hat{\varepsilon}_{1,2} & \hat{\varepsilon}_{2,2} & \hat{\varepsilon}_{3,2} & \cdots & \hat{\varepsilon}_{N,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{\varepsilon}_{1,T} & \hat{\varepsilon}_{2,T} & \hat{\varepsilon}_{3,T} & \cdots & \hat{\varepsilon}_{N,T} \end{pmatrix} \tag{16}$$

where

$$\hat{\varepsilon}_{i,t} = \tilde{r}_{i,t} - \hat{\alpha}_i - \hat{\beta}_i \tilde{r}_{m,t} \tag{17}$$

(WB. 2) Subtract the mean of each column from the residual matrix to retain the assumption of zero unconditional mean of the idiosyncratic error in the bootstrap space.

(WB. 3) Multiply the elements of the t-th row of (16) by $\zeta_t \overset{iid}{\sim} N(0,1)$ and construct the bootstrap residual matrix.

(WB. 4) The (i,t) element of the bootstrap residual matrix is denoted $\varepsilon^*_{i,t} \equiv \hat{\varepsilon}_{i,t}\zeta_t$. The bootstrap residual matrix ($\boldsymbol{X}^*_r$) is

$$\boldsymbol{X}^*_r \equiv \begin{pmatrix} \hat{\varepsilon}^*_{1,1} & \hat{\varepsilon}^*_{2,1} & \hat{\varepsilon}^*_{3,1} & \cdots & \hat{\varepsilon}^*_{N,1} \\ \hat{\varepsilon}^*_{1,2} & \hat{\varepsilon}^*_{2,2} & \hat{\varepsilon}^*_{3,2} & \cdots & \hat{\varepsilon}^*_{N,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{\varepsilon}^*_{1,T} & \hat{\varepsilon}^*_{2,T} & \hat{\varepsilon}^*_{3,T} & \cdots & \hat{\varepsilon}^*_{N,T} \end{pmatrix} = \begin{pmatrix} \hat{\varepsilon}_{1,1}\zeta_1 & \hat{\varepsilon}_{2,1}\zeta_1 & \hat{\varepsilon}_{3,1}\zeta_1 & \cdots & \hat{\varepsilon}_{N,1}\zeta_1 \\ \hat{\varepsilon}_{1,2}\zeta_2 & \hat{\varepsilon}_{2,2}\zeta_2 & \hat{\varepsilon}_{3,2}\zeta_2 & \cdots & \hat{\varepsilon}_{N,2}\zeta_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{\varepsilon}_{1,T}\zeta_T & \hat{\varepsilon}_{2,T}\zeta_T & \hat{\varepsilon}_{3,T}\zeta_T & \cdots & \hat{\varepsilon}_{N,T}\zeta_T \end{pmatrix} \tag{18}$$

Obtain $\tilde{r}^*_{i,t}$ from

$$\tilde{r}^*_{i,t} = 0 + \hat{\beta}_i \tilde{r}_{m,t} + \varepsilon^*_{i,t} \tag{19}$$

(WB. 5) Estimate $\tilde{r}^*_{i,t} = \alpha^*_i + \beta^*_i \tilde{r}_{m,t} + \tilde{\varepsilon}^*_{i,t}$ and calculate bootstrap S-GRS statistics J^*_1 in the same way as (J^s.4). Note that one should use the same shrinkage parameter when calculating the S-GRS statistics as (WB. 1)

(WB. 6) Repeat (WB. 3), (WB. 4), and (WB. 5) B times and obtain $\{J_i^*\}_{i=1}^B$

(WB. 7) Use $\{J_i^*\}_{i=1}^B$ as the sampling distribution of the S-GRS statistics from which one can calculate the critical value.

(WB. 8) Compare the critical value with J^s and determine if the null hypothesis is rejected.

4 Simulation Evidence

In this section, we provide the empirical size and power of the S-GRS test. We show that

1. In the case of $T > N$, the S-GRS test performed similarly with the standard GRS test, and it may even improve the power compared with the GRS test.
2. In the case of $T < N$, the S-GRS test has a satisfying size and power.
3. Researchers should basically use the identity matrix in the empirical analysis.

We design three types of data generating process for the idiosyncratic error: the i.i.d. error, the factor error, and the heteroskedastic factor error. Our model is the same as that in equation (3):

$$\tilde{\boldsymbol{r}}_{\boldsymbol{t}} = \boldsymbol{\alpha} + \tilde{r}_{m,t}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}} \tag{20}$$

In the simulation, we generate the market excess return data from the standard Gaussian distribution, i.e. $\tilde{r}_{m,t} \overset{iid}{\sim} N(0,1)$ and all beta are set to one, $\beta_i = 1$. In the i.i.d. error case, the error term is generated from the multivariate Gaussian distribution with mean 0 and identity covariance matrix, i.e. $\tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}} \overset{iid}{\sim} \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{I})$. In the case of the error with a factor structure, it is generated from $\tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}} = f_t \boldsymbol{\lambda} + \boldsymbol{e}_t$, where $\boldsymbol{e}_t \overset{iid}{\sim} \boldsymbol{N}(0, \boldsymbol{I}), \boldsymbol{\lambda} \overset{iid}{\sim} \boldsymbol{N}(0, \boldsymbol{I}), f_t \overset{iid}{\sim} N(0, 1)$. Here, f_t denotes the common factor and $\boldsymbol{\lambda}$ is a vector stacked with the factor loading of each asset. In case of the heteroskedastic factor error, the diagonal elements of the covariance matrix of $\boldsymbol{e}_t$ are drawn from an uniform distribution with support from 0 to 15. In this simulation, we move α_i from 0 to 0.25 in increments of 0.05 and calculate the size and power of S-GRS. The number of simulations and bootstraps is 1000 and 500, respectively. In the case of $T < N$, only S-GRS is calculated; in the case of $T > N$, both the GRS and the S-GRS are calculated to allow us to compare the results of the two tests. All the tests are conducted at the 5% significance level. We employ two types of shrinkage targets: one is the identity matrix following Ledoit and Wolf (2004b) and the other is the heteroskedastic matrix from Schäfer and Strimmer (2005).

[Table 1 here]

Table 1 shows the empirical size and power of the S-GRS test as well as the standard GRS test. In Panels A and B, both the GRS test and the S-GRS test can be calculated as $T > N$. Note that the GRS test cannot be calculated in Panels C and D because of the problem of the singular covariance matrix. For all the cases, the size of the S-GRS test fits the 5% significance level.

In the case of the i.i.d. error with $T > N$, the power of the S-GRS test is higher than that of the GRS test in Panel A because of the shrinkage effect of the covariance matrix, which reduces the upward and downward bias of the largest and smallest eigenvalue estimations of the covariance matrix, respectively, as shown in Figure 1. Figure 1 plots the eigenvalues of the true covariance matrix, sample covariance matrix, and shrinkage covariance matrix with the identity target for the cases of $T = 50,\ N = 20$.

The idiosyncratic error is i.i.d. Gaussian. The largest and smallest eigenvalues of the sample covariance matrix have huge positive and negative deviations compared with the true covariance matrix, respectively. On the contrary, the eigenvalues of the shrinkage estimate have shrunk toward 1, which is the true eigenvalue.

[Figure 1 here]

The difference in the power between the two tests narrows as T rises as shown in Panel B. This is because the sample covariance matrix approaches the true covariance matrix and there is less gain from using the shrinkage estimation. No significant difference between the shrinkage target is found in this case.

In case of the factor error with $T > N$, there is a little size distortion in the case of the heteroskedastic target while the identity target works well. We also observe higher power in the S-GRS with the identity target compared to the standard GRS test. On the contrary, there exists a slight improvement in the heteroskedastic target in comparison to the identity target in case of heteroskedastic factor error. This is because the true covariance matrix is well captured with the heteroskedastic shrinkage target. Note that the S-GRS test with the identity matrix and heteroskedastic matrix can obtain higher power than the standard GRS test in this case as well. There exists significant improvement in power compared to the standard GRS test when T is small, which is in line with the case of the i.i.d. error.

In the case of $T < N$, the S-GRS test has a satisfying size and power. While the power decreases in the case of the factor and heteroskedastic factor error compared with the i.i.d. Gaussian error, this is observed in the case of $T > N$ as well.

When the true covariance matrix is identity, the S-GRS test with the heteroskedastic target has a size distortion and lower power while the identity target works better. This is because the true covariance matrix is the identity, thus the identity matrix target is, fortunately, equivalent to the true covariance matrix in this case, which provides an appropriate size and higher power. However, even though the true covariance matrix is heteroskedastic, the S-GRS test with the identity target works better. Thus, we suggest the identity shrinkage target should be used in empirical research.

This Monte-Carlo simulation shows that the S-GRS test statistics can work in the case of $T < N$. In addition, they can improve the power in the case of $T > N$ for small samples. This result also suggests that researchers should use the identity target except for the case when they believe that there exists large heteroskedasticity in the covariance matrix of interest. This result leads this study to use the identity target in the empirical analysis and employ the heteroskedastic target as a robustness check.

5 Empirical Illustration

In this section, we apply the S-GRS test for two linear asset pricing models in the Japanese stock market. We first analyze the SL-CAPM in which the excess return of a given asset is explained by its beta multiplied by the excess return of the market portfolio as in equation (1). We also analyze the FF3 model, which includes two additional factors to explain the variation in the excess return on assets. Fama and French (1993) formulate their FF3 model as

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_i(r_{m,t} - r_{f,t}) + \beta_i^{SMB} SMB_t + \beta_i^{HML} HML_t + \varepsilon_{i,t} \qquad (21)$$

where SMB_t and HML_t are the size factor and value factor, respectively. The SMB factor is calculated as the return of a portfolio consisting of small market capitalization firms minus that of large market capitalization firms. The HML factor is calculated as the return of a portfolio consisting of high book-to-market ratio firms minus that of low book-to-market ratio firms. Note that those returns are all from zero-investment portfolios.

We use TOPIX, a representative index of the Japanese stock market, and TSE Second Section Stock Price Index, which we call TOPIX2 in this study, when analyzing TSE first section and second section, respectively. According to the Japan Exchange Group website,[8] "TOPIX is a free-float adjusted market capitalization-weighted index that is calculated based on all the domestic common stocks listed on the TSE First Section. This is a measure of the overall trend in the stock market, and is used as a benchmark for investment in Japan stocks." Thus, it is worth analyzing TOPIX academically and practically for the following two main reasons. Firstly, as TOPIX is a capitalization-weighted index, it can be used to check the theoretical conclusion of the CAPM. Secondly, as it is widely used as a benchmark, analyzing efficiency can explain whether active fund managers can create higher efficiency than the benchmark index in Japan.

Our data set consists of monthly return data for all the individual securities on the TSE first section and second section constructed by Financial Data Solutions, Inc. The data range from November 1982 to September 2015 and from August 1993 to September 2015 for the TSE first section and second section, respectively. We use TOPIX and TOPIX2 as a market return and the return of 10-year Japanese government bonds as a risk-free rate.

We estimate the SL-CAPM and FF3 model by using 60-month subsamples with a rolling window. Assuming that the parameters are stable for at least 60 months, we can separately identify the inefficiency of TOPIX from the problem of time-invariant coefficients in the estimation. We then iterate this procedure with all the data by using a rolling window and observe a variation in the p-value of the S-GRS test. For instance, the p-value we obtain in September 2000 is the p-value of the S-GRS test

[8] `http://www.jpx.co.jp/english/markets/indices/topix/`.

from the sample between October 1995 and September 2000.

We apply two different procedures of the data construction of the individual securities: One is the method used in Pesaran and Yamagata (2012) and the other is used in Gungor and Luger (2013) and Gungor and Luger (2016). In the former case, we exclude securities that have at least one missing data point during the 60-month study period in each estimation. In the latter case, we remove securities that have at least one missing data point from the overall estimation. We call the data-set constructed following the latter the complete data. The number of securities in our universe for the TSE first and second sections in the complete data is now 1433 and 154, almost half of the maximum number of the data following Pesaran and Yamagata (2012), respectively. In the latter settings, one may intentionally drop securities that tend to move anomalously from the sample because one excludes several securities whose returns are not observed during the full sample (e.g., the company went bankrupt or was demoted from the TSE first or second sections or IPOs). Thus, we expect that both the models are more likely to be accepted in this settings.

The overall number of securities on the data-sets is 1812 and 1434 for TSE first section and second section respectively. In case of the complete data, we have 615 and 131 securities respectively.

The results are similar in the cases of the identity and the heteroskedastic target. Thus, we present the results with the identity target here and report the case of the heteroskedastic target as a robustness check in the appendix. We report four main findings from this empirical illustration. First, the CAPM and FF3 model are mostly rejected in both sections, which means the market portfolio or market portfolio plus factor portfolios are not mean-variance efficient. This may imply the possibility of the active fund management. This result is in line with Grinold (1992) and Kubota and Takehara (2017) while they conducted the GRS test with portfolios. Their result may suffer from larger type-1 error caused by the sorting problems as they use several portfolios, not individual securities. This paper adds the evidence of the inefficiency of the benchmark market portfolio from another point of view; the inefficiency compared to the ex-post tangency portfolio constructed from individual securities. Second, while CAPM and FF3 models are rejected in most periods, they tend to be accepted[9] in the TSE second section and marginally in the TSE first section during periods of huge market shocks. This implies that positive or negative external large shocks affect all the individual stocks in the market and the market return can well explain the movement of individual stocks in Japan. Hamao et al. (2007) show that after the asset price bubble in Japan, market volatility increases while idiosyncratic volatility decreases ,which is opposite to the case in the U.S. This can be the logical explanation of the result in this analysis while it requires further analysis. Third, we also observe some periods when the CAPM has higher p-values than the FF3 model. This may sound counter-intuitive, as the FF3 model nests CAPM. This seems to be, however, because the factor structures captured by the FF3 models are not explicitly modeled in the CAPM, which makes the CAPM underspecified. Thus, the error term of the CAPM incorporates factor structures and the variance-covariance matrix becomes large, which leads to a smaller test statistics. Finally but not less importantly, as expected, CAPM and FF3 models are more likely to be accepted in the case of ,probably due to the aforementioned reasons.

[Figure 2 here]

We first analyze the efficiency of TOPIX compared with the TSE first section. Figure 2 presents the p-value of the mean-variance efficiency tests of TOPIX with

[9] In comparison between the TSE first section and second section, both models are more likely to be accepted in the latter. This should be interpreted with caution. The acceptance of the null hypothesis does not guarantee that the null hypothesis is true and one should interpret it that there were no sufficient evidence to reject it. In this setting, we have quite smaller number of securities for the TSE second section. The acceptance of the model is more likely due to the lack of power of the test.

the individual securities in the TSE first section by using the S-GRS test with the identity target. The CAPM and FF3 model are rejected in all the periods, which is in line with several previous research with a data on portfolios. Take Kubota and Takehara (2017) for instance, who analyze the validity of multi-factor pricing models in the Japanese stock market by using monthly data from 1978 to 2014 with 45 portfolios. They conclude that all models are rejected by the GRS test, while the multi-factor models have a higher p-value than the single factor model. Therefore, we confirm that the multi-factor pricing model cannot fully explain the variation in the stock return not only for the portfolios but also for the individual stocks. This also implies that the TOPIX is not mean-variance efficient compared with the tangency portfolio consisting all the individual securities on the TSE first section. This result, as well as Grinold (1992), encourage active fund managers that they have a chance to achieve higher efficiency than their benchmark.

[Figure 3 here]

Figure 3 presents the p-value of the mean-variance efficiency tests of TOPIX with the complete data of the individual securities in the TSE first section by using the S-GRS test with the identity target. The data structure follows Gungor and Luger (2013) and Gungor and Luger (2016), thus securities that have at least one missing data point during the full sample are excluded. The CAPM and FF3 model are rejected in most, but not all, periods. In contrast to the result above, considering the 99% significance level, either the CAPM or the FF3 model are not rejected around 1992, 1999, and 2008. We use "around" because we are estimating with a rolling window. Those periods correspond to the financial shocks such as the crash after the Japanese asset price bubble in 1991, Asia financial crisis, Russian financial crisis, and Dot-com bubble from 1997 to 2000, and the global financial crisis in 2008.

From those observations, we conclude that the CAPM and the FF3 models are mostly rejected even though all the anomalistic stocks are excluded, but they are accepted during the periods of huge external shocks. This can also be interpreted as the market portfolio is not always mean-variance efficient in the TSE first section, which encourages the active portfolio management or smart-beta strategy.

[Figure 4 here]

We secondly analyze the efficiency of TOPIX2 compared with individual stocks in the TSE second section. Figure 4 presents the p-value of the mean-variance efficiency tests of TOPIX2 with the individual securities in the TSE second section by using the S-GRS test with the identity target. In contrast to the TSE first section, the CAPM and FF3 model are more frequently accepted while the acceptance periods are basically same as the TSE first section with complete data. In addition, periods of high p-value persists longer time compared to the results in the TSE first section. This implies that the TSE second section is influenced by the external shocks more strongly and persistently. Note that the FF3 factor model can capture the size-effect, which is prevalent in the TSE second section. This also leads to the acceptance of the model as well.

[Figure 5 here]

Figure 5 presents the p-value of the mean-variance efficiency tests of TOPIX2 with the complete data of the individual securities in the TSE second section by using the S-GRS test with the identity target. This result presents that neither the CAPM nor FF3 model are mostly rejected. In addition to this observation, the peaks of the p-values correspond to the acceptance periods in Figure 4. This result provides the robustness of our explanations: we cannot reject the efficiency of the market portfolio during the periods of huge external shocks.

6 Conclusion

In this study, we proposed a test of linear factor pricing models that is valid even in the case of $T < N$ where the standard GRS test cannot be calculated because of the problem of a singular covariance matrix estimation. We employed the shrinkage estimator of the covariance matrix to have a non-singular covariance matrix estimation and obtained the sampling distribution of the S-GRS test statistics through the fixed-design wild bootstrap. Thanks to this procedure, the S-GRS test is calculable even in $T < N$ and robust to non-normality and conditional heteroskedasticity in the idiosyncratic error. The Monte-Carlo evidence showed that the S-GRS test has a satisfying power and size, and may improve the power even when $T > N$ because of the improvement in the estimation of the covariance matrix. To our best knowledge, this study is the first to analyze the efficiency of the Japanese stock market with individual securities. We found strong evidence of the inefficiency of TOPIX compared with individual securities in the TSE first section. On the contrary, we cannot reject the efficiency of the market portfolio in the TSE second section during periods of huge financial shocks in the market such as Asia financial crisis or the global financial crisis. This research suggests several future research. First, by using the proposed test statistics, fund managers or index providers can analyze if the benchmark portfolio or proposing index is mean-variance efficient in a given investment universe beforehand. Second, the empirical illustration pointed out that the pricing models cannot be rejected during the time of financial shocks. It is worth analyzing this phenomenon more. For instance, it is worth analyzing if this is truly due to the increased volatility of market index shown by Hamao et al. (2007) and if this phenomenon can be observed in other financial markets. This result may reflect some of the characteristics of the overall market[10].

[10]This may be because, for instance, the probability of default is higher in TSE second section thatn TSE first section. If this is true, one may make a subsample of the individual securities in the TSE first section sorted by the leverage ratio and observe similar results for the high leverage ratio group.

References

Affleck-Graves, J. and B. McDonald (1989). Nonnormalities and tests of asset pricing theories. *The Journal of Finance 44*(4), 889–908.

Affleck-Graves, J. and B. McDonald (1990). Multivariate tests of asset pricing: The comparative power of alternative statistics. *Journal of Financial and Quantitative Analysis 25*(2), 163–185.

Beaulieu, M.-C., J.-M. Dufour, and L. Khalaf (2007). Multivariate tests of mean–variance efficiency with possibly non-gaussian errors: An exact simulation-based approach. *Journal of Business & Economic Statistics 25*(4), 398–410.

Berk, J. B. (2000). Sorting out sorts. *The Journal of Finance 55*(1), 407–427.

Bodnar, T., A. K. Gupta, and N. Parolya (2014). On the strong convergence of the optimal linear shrinkage estimator for large dimensional covariance matrix. *Journal of Multivariate Analysis 132*, 215–228.

Chou, P.-H. and G. Zhou (2006). Using bootstrap to test portfolio efficiency. *Annals of Economics & Finance 7*(2).

Detzler, M. and J. Wiggins (1997). The performance of actively managed international mutual funds. *Review of Quantitative Finance and Accounting 8*(3), 291–313.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics 33*(1), 3–56.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of financial economics 116*(1), 1–22.

Firer, C., S. A. Ross, R. Westerfield, and B. D. Jordan (2012). *Fundamentals of corporate finance*. McGraw-Hill Higher Education.

Frankfurter, G. M., H. E. Phillips, and J. P. Seagle (1971). Portfolio selection: the effects of uncertain means, variances, and covariances. *Journal of Financial and Quantitative Analysis 6*(5), 1251–1262.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989). A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, 1121–1152.

Gonçalves, S. and L. Kilian (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics 123*(1), 89–120.

Grinold, R. C. (1992). Are benchmark portfolios efficient? *The Journal of Portfolio Management 19*(1), 34–40.

Gungor, S. and R. Luger (2009). Exact distribution-free tests of mean-variance efficiency. *Journal of Empirical Finance 16*(5), 816–829.

Gungor, S. and R. Luger (2013). Testing linear factor pricing models with large cross sections: A distribution-free approach. *Journal of Business & Economic Statistics 31*(1), 66–77.

Gungor, S. and R. Luger (2016). Multivariate tests of mean-variance efficiency and spanning with a large number of assets and time-varying covariances. *Journal of Business & Economic Statistics 34*(2), 161–175.

Hamao, Y., J. Mei, and Y. Xu (2007). Unique symptoms of japanese stagnation: an equity market perspective. *Journal of Money, Credit and Banking 39*(4), 901–923.

Hwang, S. and S. E. Satchell (2014). Testing linear factor models on individual stocks using the average f-test. *The European Journal of Finance 20*(5), 463–498.

Kelejian, H. H. and I. R. Prucha (2001). On the asymptotic distribution of the moran i test statistic with applications. *Journal of Econometrics 104*(2), 219–257.

Kubota, K. and H. Takehara (2017). Does the fama and french five-factor model work well in japan? *International Review of Finance*.

Ledoit, O. and M. Wolf (2003). Improved estimation of the covaraince matrix of stock returns with an application to portfolio selection. *Journal of empirical finance 10*(5), 603–621.

Ledoit, O. and M. Wolf (2004a). Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management 30*(4), 110–119.

Ledoit, O. and M. Wolf (2004b). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis 88*(2), 365–411.

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *The journal of finance 20*(4), 587–615.

Lo, A. W. and A. C. MacKinlay (1990). Data-snooping biases in tests of financial asset pricing models. *The Review of Financial Studies 3*(3), 431–467.

MacKinlay, A. C. and M. P. Richardson (1991). Using generalized method of moments to test mean-variance efficiency. *The Journal of Finance 46*(2), 511–527.

Markowitz, H. (1952). Portfolio selection. *The journal of finance 7*(1), 77–91.

Pesaran, M. H. and T. Yamagata (2012). Testing capm with a large number of assets. SSRN Working Paper and The Institute for the Study of Labor (IZA) Discussion Paper.

Ross, S. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*(3), 341–360.

Schäfer, J. and K. Strimmer (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology 4*(1).

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance 19*(3), 425–442.

Stein, C. et al. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley symposium on mathematical statistics and probability*, Volume 1, pp. 197–206.

Zhou, G. (1993). Asset-pricing tests under alternative distributions. *The Journal of Finance 48*(5), 1927–1942.

A Tests of Mean-variance Efficiency

In this section, we interpret the GRS test as a test of the mean-variance efficiency of the portfolio on the right-hand side of the linear factor asset pricing models. For simplicity, we only discuss the case of the single factor model; however, the same discussion can be carried out for multi-factor pricing models. Suppose that our investment universe is N assets and we have a risk-free asset whose return at time t is $r_{f,t}$. The weight vector of a given portfolio p is denoted as $\boldsymbol{w}$, and $\boldsymbol{\mu}$ is the expected return vector of N assets, and $\boldsymbol{\Omega}$ is the covariance matrix. The return and variance of a given portfolio can be written as $r_p = \boldsymbol{w}'\boldsymbol{\mu}$ and $\sigma_p^2 = \boldsymbol{w}'\boldsymbol{\Omega}\boldsymbol{w}$, respectively. A given portfolio is mean-variance efficient provided the following holds:

$$\frac{\mathbb{E}(r_1) - r_f}{MRC_1} = \frac{\mathbb{E}(r_2) - r_f}{MRC_2} = ... = \frac{\mathbb{E}(r_n) - r_f}{MRC_n} \tag{22}$$

where $MRC_i = \frac{\partial \sigma_p}{\partial w_i}$, the marginal risk contribution of the individual asset i.

If there is an inequality in equation (22), it implies that the given portfolio can be improved in the meaning of mean-variance efficiency by placing more weight on one asset and decreasing the other until the inequality becomes an equality.

By stacking the marginal risk contribution of each asset into a vector, simple algebra shows

$$\begin{aligned} MRC \equiv \frac{\partial \sigma_p}{\partial \boldsymbol{w}} &= \frac{\partial}{\partial \boldsymbol{w}} \sqrt{\boldsymbol{w}'\boldsymbol{\Omega}\boldsymbol{w}} = \frac{\boldsymbol{\Omega}\boldsymbol{w}}{\sqrt{\boldsymbol{w}'\boldsymbol{\Omega}\boldsymbol{w}}} \\ &= \begin{pmatrix} \frac{cov(r_1, r_p)}{\sigma_p} \\ \frac{cov(r_2, r_p)}{\sigma_p} \\ \vdots \\ \frac{cov(r_n, r_p)}{\sigma_p} \end{pmatrix} = \sigma_p \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} \end{aligned} \tag{23}$$

where $\beta_i = \frac{Cov(r_i, r_p)}{\sigma_p^2}$

Thus, $\beta_i = \frac{1}{\sigma_p} MRC_i$ holds.

From equation (22), a given portfolio p is mean-variance efficient

$\Leftrightarrow \mathbb{E}(r_i) - r_f = MRC_i \times \frac{\mathbb{E}(r_p) - r_f}{\sigma_p}$

$\Leftrightarrow \mathbb{E}(r_i) - r_f = \beta_i(\mathbb{E}(r_p) - r_f)$

$\Leftrightarrow \mathbb{E}(\tilde{r}_i) = \beta_i(\mathbb{E}(\tilde{r}_p))$

Thus, the test of mean-variance efficiency is equivalent to the test of the CAPM:

$$H_0 : \alpha_i = 0 \ \forall i = 1, ..., N$$

B Other Empirical Illustrations

In this section, we present the other result of the mean-variance tests of the market portfolios. In particular, we present the case when the shrinkage target is the heteroskedastic matrix as a robustness check. The basic findings from the analysis in this section support our results in Section 5. The CAPM and FF3 models are rejected mostly while they tend to be accepted during the periods of huge external shocks. In comparison between the TSE first section and second section, both the models are more likely to be accepted in the TSE second section. The two additional factors in the FF3 model can explain the variation in the excess return of the individual assets better than the CAPM.

[Figure 6 here]

Figure 6 presents the results of the mean-variance test of the TOPIX of the TSE first section with the heteroskedastic shrinkage target. The data structure follows the setting of Pesaran and Yamagata (2012). We exclude any security that has at least one missing data point during the subsample in each rolling window estimation. The shrinkage target is the heteroskedastic matrix. Neither the CAPM nor the FF3 model are accepted for all the periods, which supports the results in Section 5.

[Figure 7 here]

Figure 7 presents the results of the mean-variance test of TOPIX with the complete data set of the TSE first section and the heteroskedastic shrinkage target. The data structure follows the setting of Gungor and Luger (2013) and Gungor and Luger (2016). We exclude securities that have at least one missing data point during the full sample. The CAPM is accepted after the crash of the bubble in 1991and during the global financial crisis, in line with the results in Section 5.

[Figure 8 here]

Figure 8 presents the results of the mean-variance test of the TOPIX of the TSE second section with the heteroskedastic shrinkage target. The data set follows the setting of Pesaran and Yamagata (2012). As before, we exclude any security that has at least one missing data point during the subsample in each rolling window estimation. There are no outstanding differences from the case of the identity shrinkage target. The only difference is that both the FF3 model and the CAPM are rejected for several years after the global financial crisis, while they are still accepted after 2014 in the case of the identity target. However, the acceptance of these two models are marginal, allowing us to conclude that there is no huge difference between the choice of shrinkage target.

[Figure 9 here]

Figure 9 presents the results of the mean-variance test of TOPIX with the complete data set of the TSE second section and the heteroskedastic shrinkage target. The data structure follows the setting of Gungor and Luger (2013) and Gungor and Luger (2016).

We exclude securities that have at least one missing data point during the full sample. The FF3 model is accepted most of the time, while the CAPM is occasionally rejected when the FF3 model is accepted. This finding implies that the two additional variables in the FF3 model can explain the variation in the return of the TSE second section. In addition, compared with Figure 8, the CAPM is accepted more often, which commonly occurs when we exclude securities that have at least one missing data point in full sample periods.

C Tables

Panel A	i.i.d			Factor			Heteroskedastic Factor		
N20, T50	S-GRS (I)	S-GRS (H)	GRS	S-GRS (I)	S-GRS (H)	GRS	S-GRS (I)	S-GRS (H)	GRS
0	0.059	0.075	0.047	0.059	0.106	0.046	0.06	0.063	0.046
0.05	0.126	0.133	0.062	0.1	0.178	0.068	0.087	0.101	0.049
0.1	0.423	0.436	0.242	0.23	0.250	0.152	0.089	0.107	0.061
0.15	0.824	0.844	0.582	0.501	0.412	0.308	0.142	0.157	0.087
0.2	0.985	0.990	0.877	0.772	0.626	0.602	0.204	0.222	0.137
0.25	0.999	0.999	0.987	0.954	0.756	0.817	0.326	0.360	0.215

Panel B	i.i.d			Factor			Heteroskedastic Factor		
N20, T100	S-GRS (I)	S-GRS (H)	GRS	S-GRS (I)	S-GRS (H)	GRS	S-GRS (I)	S-GRS (H)	GRS
0	0.057	0.062	0.039	0.056	0.111	0.052	0.052	0.068	0.046
0.05	0.213	0.226	0.156	0.116	0.167	0.088	0.065	0.074	0.056
0.1	0.767	0.772	0.663	0.396	0.271	0.369	0.087	0.103	0.103
0.15	0.994	0.995	0.982	0.810	0.470	0.769	0.219	0.242	0.194
0.2	1.000	1.000	1.000	0.983	0.702	0.979	0.399	0.411	0.309
0.25	1.000	1.000	1.000	0.999	0.899	0.998	0.586	0.616	0.493

Panel C	i.i.d			Factor			Heteroskedastic Factor		
N120, T100	S-GRS (I)	S-GRS (H)	GRS	S-GRS (I)	S-GRS (H)	GRS	S-GRS (I)	S-GRS (H)	GRS
0	0.050	0.057	NA	0.023	0.126	NA	0.045	0.092	NA
0.05	0.573	0.213	NA	0.159	0.406	NA	0.099	0.158	NA
0.1	1.000	0.767	NA	0.918	0.973	NA	0.223	0.318	NA
0.15	1.000	0.994	NA	0.999	1.000	NA	0.601	0.703	NA
0.2	1.000	1.000	NA	1.000	1.000	NA	0.919	0.955	NA
0.25	1.000	1.000	NA	1.000	1.000	NA	0.996	0.999	NA

Panel D	i.i.d			Factor			Heteroskedastic Factor		
N150, T100	S-GRS (I)	S-GRS (H)	GRS	S-GRS (I)	S-GRS (H)	GRS	S-GRS (I)	S-GRS (H)	GRS
0	0.084	0.133	NA	0.098	0.168	NA	0.063	0.131	NA
0.05	0.655	0.793	NA	0.363	0.510	NA	0.094	0.178	NA
0.1	1.000	1.000	NA	0.970	0.995	NA	0.306	0.436	NA
0.15	1.000	1.000	NA	0.999	1.000	NA	0.684	0.801	NA
0.2	1.000	1.000	NA	1.000	1.000	NA	0.958	0.983	NA
0.25	1.000	1.000	NA	1.000	1.000	NA	1.000	1.000	NA

Note: This table reports the empirical size and power of the GRS and S-GRS tests in the case of the i.i.d. error, factor error, and heteroskedastic factor error. In the case of i.i.d. error, the idiosyncratic error follows the i.i.d. standard normal distribution. The factor error is the sum of one common factor and the i.i.d. Gaussian error. In case of the heteroskedastic factor error, the diagonal elements of the covariance matrix of the idiosyncratic error are drawn from an uniform distribution with support from 0 to 15. The intercept α_i moves from 0 (size) to 0.25 (power) in increments of 0.05. The S-GRS test is based on 500 bootstraps. The parentheses next to S-GRS show the shrinkage target. (I) and (H) mean the identity target and heteroskedastic shrinkage target, respectively. Panels A and B are for $T > N$ and Panels C and D are for $T < N$. The results are based on 1000 replications. NA is used when a certain test is not computable.

Table 1: Empirical size and power of the GRS and S-GRS test

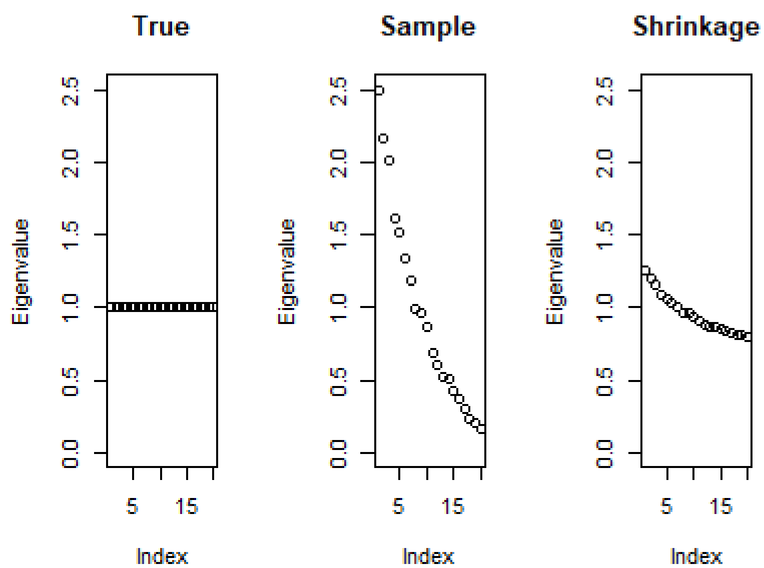

Figure 1: Eigenvalues of the covariance matrix

Note: This figure shows the eigenvalues of the true covariance matrix, sample covariance matrix, and shrinkage covariance matrix estimated from the residual of equation (3), the Sharpe–Lintner-type CAPM, in the case of $N = 20, T = 50$ with the i.i.d. Gaussian error.

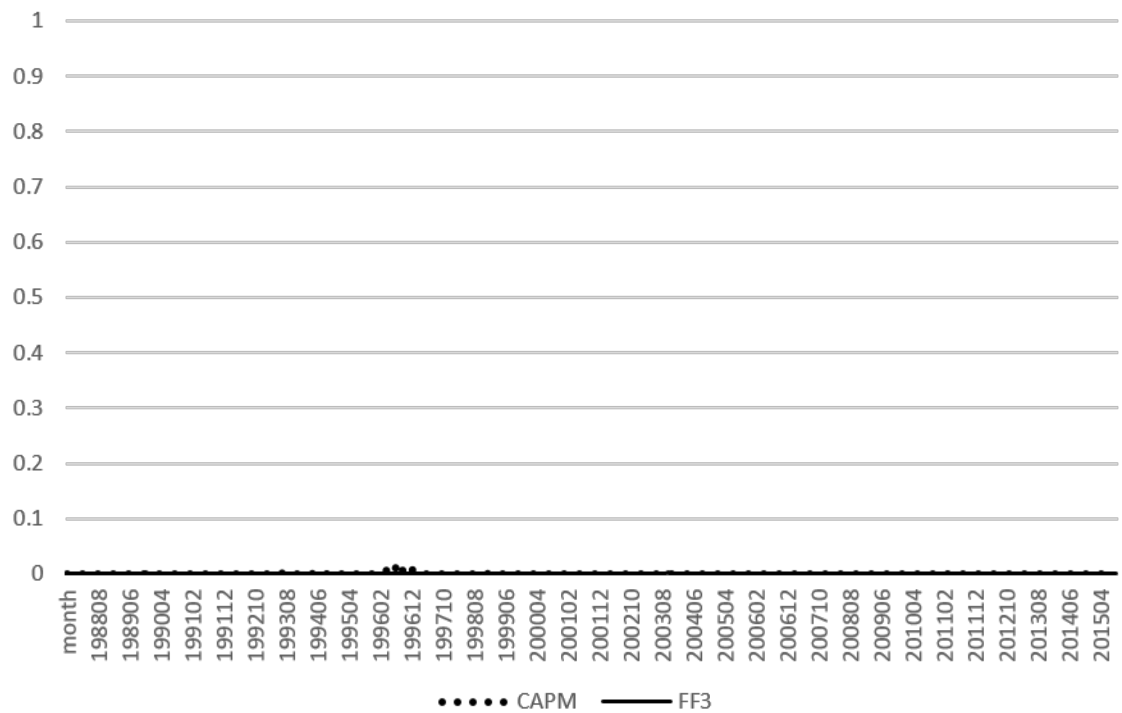

Figure 2: Efficiency of TOPIX compared with the TSE first section individual securities (identity target)

Note: This figure plots the time variations in the p-values for the S-GRS test estimated with a rolling window using the 60-month data from November 1982 to September 2015 in the TSE first section. The shrinkage target is the identity covariance matrix. The dashed line and solid line are the p-values of the mean-variance efficiency tests of the CAPM and FF3 model, respectively. Securities that have at least one missing data point during the 60-month window in each estimation are excluded.

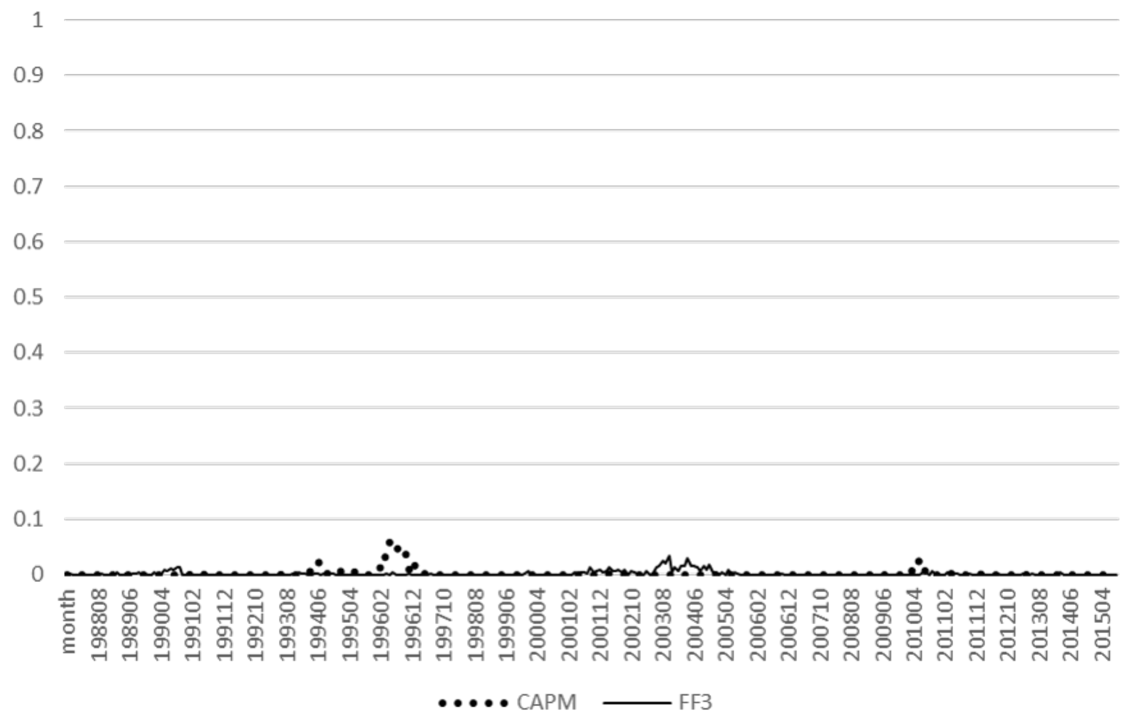

Figure 3: Efficiency of TOPIX compared with the TSE first section individual securities with complete data (identity target)

Note: This figure plots the time variations in the p-values for the S-GRS test estimated with a rolling window using the 60-month data from November 1982 to September 2015 in the TSE first section. The shrinkage target is the identity covariance matrix. The dashed line and solid line are the p-values of the mean-variance efficiency tests of the CAPM and FF3 model, respectively. Securities that have at least one missing data point during the full sample are excluded.

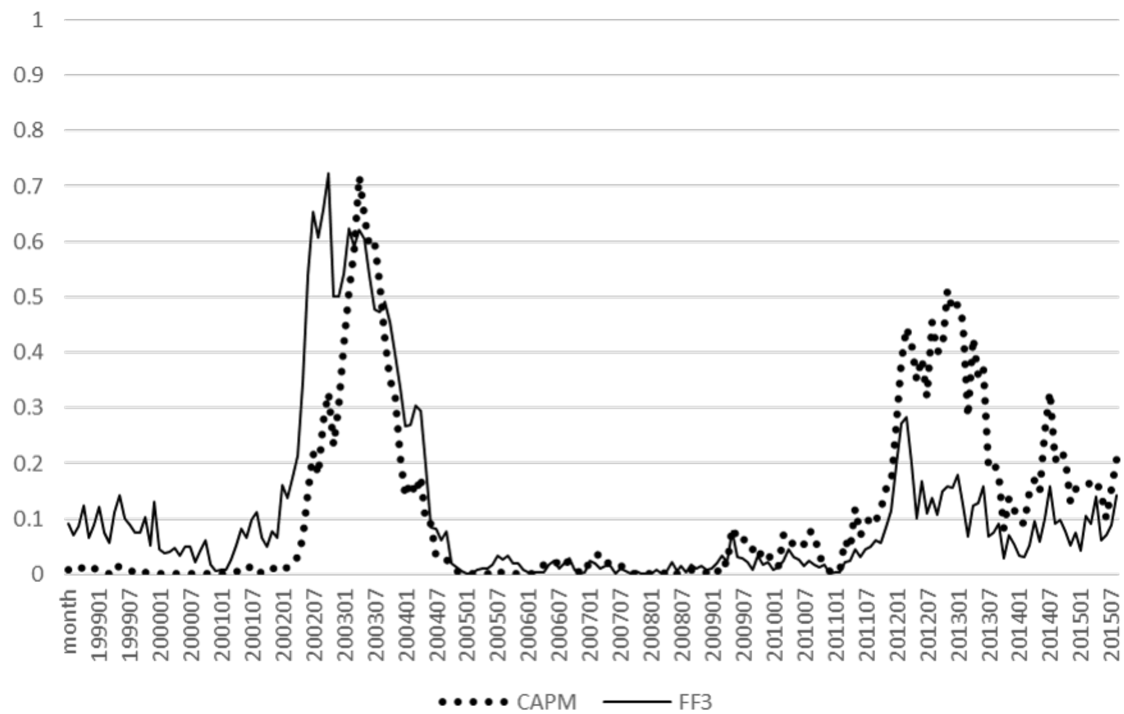

Figure 4: Efficiency of TOPIX2 compared with the TSE second section individual securities (identity target)

Note: This figure plots the time variations in the p-values for the S-GRS test estimated with a rolling window using the 60-month data from August 1993 to September 2015 in the TSE second section. The shrinkage target is the identity covariance matrix. The dashed line and solid line are the p-values of the mean-variance efficiency tests of the CAPM and FF3 model, respectively. Securities that have at least one missing data point during the 60-month window in each estimation are excluded.

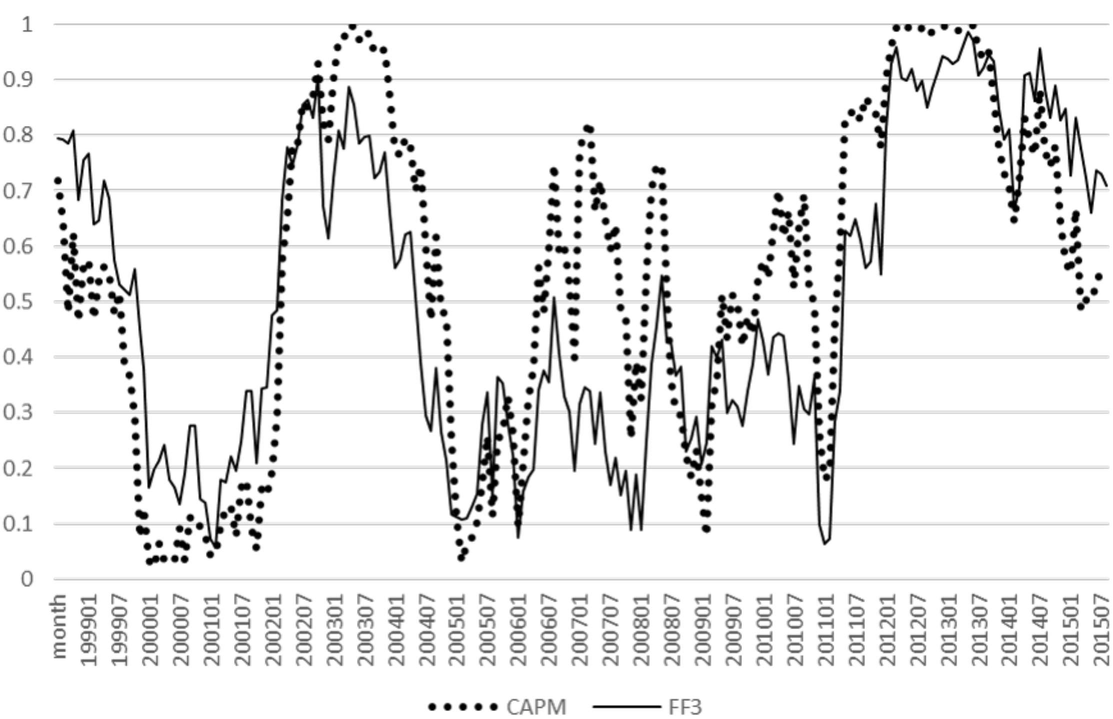

Figure 5: Efficiency of TOPIX2 compared with the TSE second section individual securities with complete data (identity target)

Note: This figure plots the time variations in the p-values for the S-GRS test estimated with a rolling window using the 60-month data from August 1993 to September 2015 in the TSE second section. The dashed line and solid line are the p-values of the mean-variance efficiency tests of the CAPM and FF3 model, respectively. Securities that have at least one missing data point during the full sample are excluded.

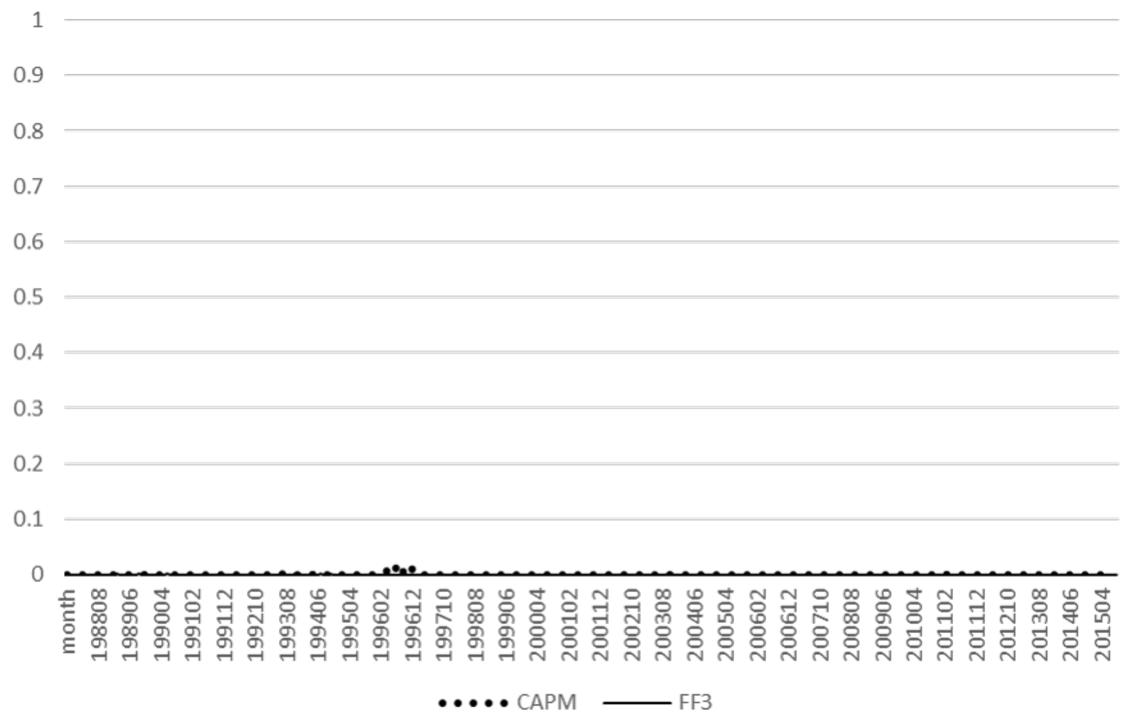

Figure 6: Efficiency of TOPIX compared with the TSE first section individual securities (heteroskedastic target)

Note: This figure plots the time variations in the p-values for the S-GRS test estimated with a rolling window using the 60-month data from November 1982 to September 2015 in the TSE first section. The shrinkage target is the heteroskedastic covariance matrix. The dashed line and solid line are the p-values of the mean-variance efficiency tests of the CAPM and FF3 model, respectively. Securities that have at least one missing data point during the 60-month window in each estimation are excluded.

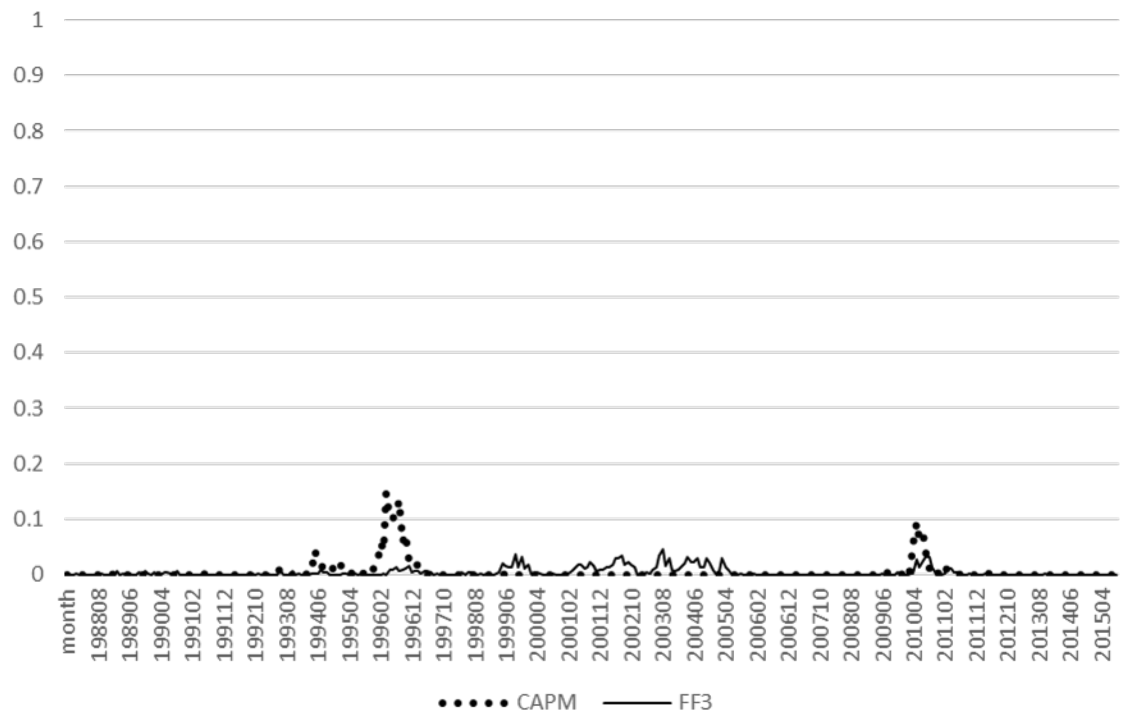

Figure 7: Efficiency of TOPIX compared with the TSE first section individual securities with complete data (heteroskedastic target)

Note: This figure plots the time variations in the p-values for the S-GRS test estimated with a rolling window using the 60-month data from November 1982 to September 2015 in the TSE first section. The shrinkage target is the heteroskedastic covariance matrix. The dashed line and solid line are the p-values of the mean-variance efficiency tests of the CAPM and FF3 model, respectively. Securities that have at least one missing data point during the full sample are excluded.

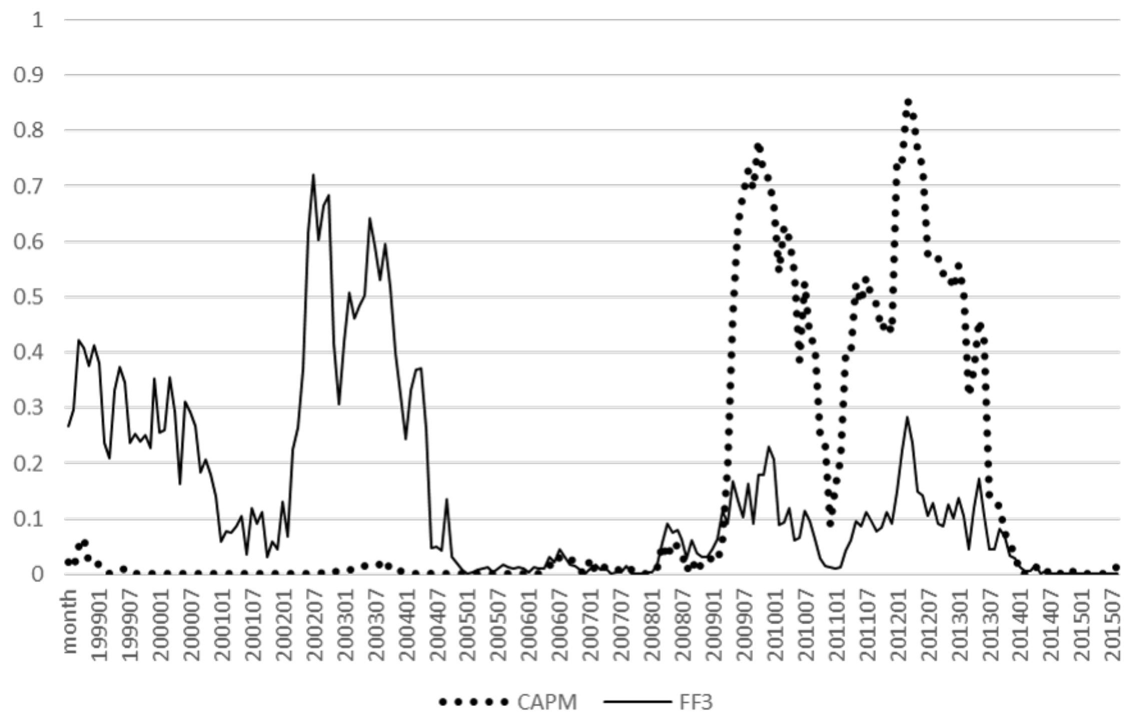

Figure 8: Efficiency of TOPIX2 compared with the TSE second section individual securities (heteroskedastic target)

Note: This figure plots the time variations in the p-values for the S-GRS test estimated with a rolling window using the 60-month data from August 1993 to September 2015 in the TSE second section. The shrinkage target is the heteroskedastic covariance matrix. The dashed line and solid line are the p-values of the mean-variance efficiency tests of the CAPM and FF3 model, respectively. Securities that have at least one missing data point during the 60-month window in each estimation are excluded.

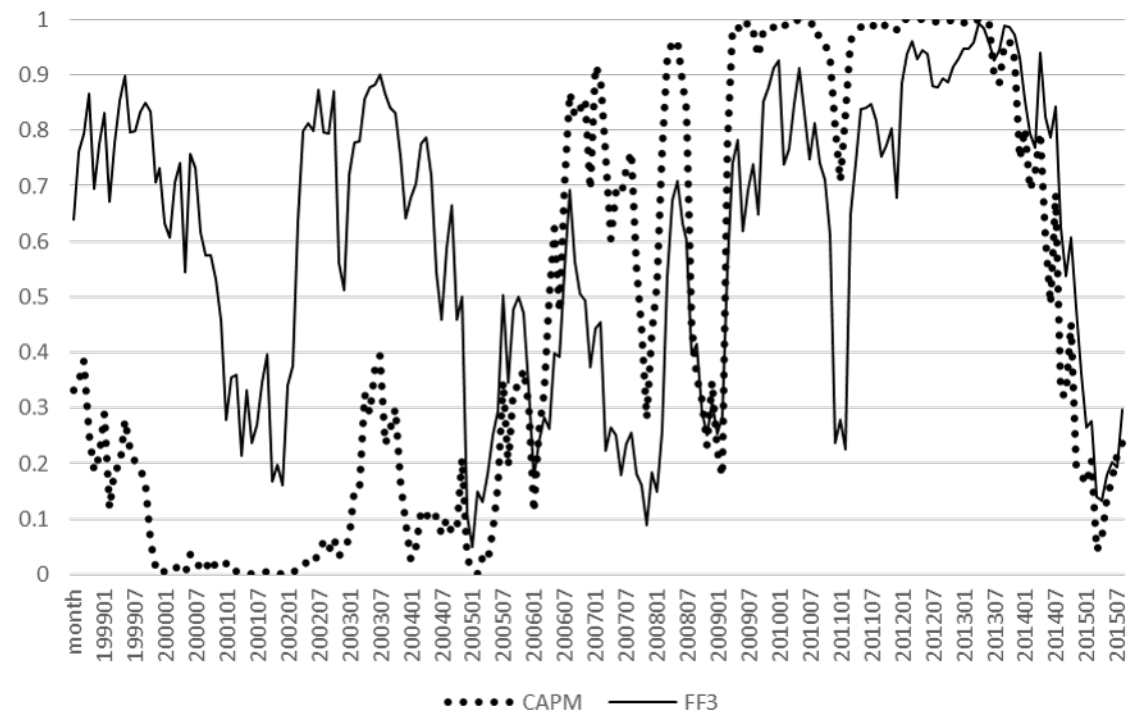

Figure 9: Efficiency of TOPIX2 compared with the TSE second section individual securities with complete data (heteroskedastic target)

Note: This figure plots the time variations in the p-values for the S-GRS test estimated with a rolling window using the 60-month data from August 1993 to September 2015 in the TSE second section. The shrinkage target is the heteroskedastic covariance matrix. The dashed line and solid line are the p-values of the mean-variance efficiency tests of the CAPM and FF3 model, respectively. Securities that have at least one missing data point during the full sample are excluded.