In-sample and Out-of-sample Sharpe Ratios of Multi-factor Asset Pricing Models⋆

RAYMOND KAN[a,∗], XIAOLU WANG[b], XINGHUA ZHENG[c]

[a] *University of Toronto, 105 St. George Street, Toronto, ON, M5S 3E6, Canada*

[b] *Iowa State University, 2167 Union Drive, Ames, IA, 50011, USA*

[c] *Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong*

Abstract

Using available return data, many multi-factor asset pricing models present impressive in-sample Sharpe ratios, significantly surpassing that of the market portfolio. Such a performance, however, contradicts the conventional wisdom in finance. Investors cannot realistically attain the in-sample Sharpe ratios. They obtain the out-of-sample Sharpe ratios, which are significantly lower. Estimation risk is one reason for this performance deterioration. We theoretically study the effect of estimation risk by obtaining the exact distributions of in-sample and out-of-sample Sharpe ratios, and argue that such effect needs to be considered in model comparisons.

Keywords: asset pricing model, Sharpe ratio, estimation risk, model comparison, exact and asymptotic distributions, stochastic representation
JEL: C10, C11, G11, G12

⋆Nikolai Roussanov was the editor for this article. We thank the editor, an anonymous reviewer, Svetlana Bryzgalova, Peter Christoffersen, Victor DeMiguel, Andrew Detzel, Wayne Ferson, Junbo Wang, Guofu Zhou, seminar participants at Bates White Economic Consulting, Carleton University, Chinese University of Hong Kong, Indiana University, Laval University, London Business School, Louisiana State University, Queensland University of Technology, Stockholm School of Business, UCLouvain, University of Toronto, and conference participants at the 2019 CFIRM Conference, 2021 Finance Conference in Honour of Phelim Boyle, 2021 SoFiE Conference, 2022 FMA annual conference, and 2022 CMStatistics conference for helpful comments and suggestions. Zheng's research is partially supported by the Research Grant Council of Hong Kong SAR grants GRF16304019, GRF15302321, GRF16304521, and T31-604/18-N.

∗Corresponding author.

Email addresses: `raymond.kan@rotman.utoronto.ca` (RAYMOND KAN), `xiaoluw@iastate.edu` (XIAOLU WANG), `xhzheng@ust.hk` (XINGHUA ZHENG)

1. Introduction

Many multi-factor asset pricing models have been proposed since the capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965). Relative to the CAPM, the multi-factor models, especially the more recently proposed ones, tend to perform better in terms of explaining the cross-section of expected asset returns. However, none of these models can completely explain the cross-section, and it is likely that additional asset pricing models will be proposed.

In this paper, we consider the CAPM that uses the value-weighted market portfolio as the benchmark portfolio, and seven popular multi-factor asset pricing models. The multi-factor models that we consider are: (1) Fama and French's (1993) 3-factor model (FF-3), which adds size and book-to-market factor to the CAPM; (2) Carhart's (1997) 4-factor model (Carhart-4), which adds the momentum factor to FF-3; (3) the betting against beta (BAB) model of Frazzini and Pedersen (2014), which adds the return difference between low and high beta portfolios as an additional factor to the CAPM; (4) Fama and French's (2015) 5-factor model (FF-5), which adds profitability and investment factors to FF-3; (5) Hou, Xue, and Zhang's (2015) q-factor model (HXZ q), which adds size, investment and profitability factors to the CAPM; (6) Barillas and Shanken's (2018) 6-factor model (BS-6), which combines the market factor, size factor, momentum factor, HXZ's profitability and investment factors, and a monthly updated book-to-market factor from Asness and Frazzini (2013); and (7) Hou, Mo, Xue, and Zhang's (2021) q^5 model (HMXZ q^5), which adds an expected growth factor to HXZ q.

Using monthly data over the 1967/1–2021/12 period, we empirically examine the performance of the eight asset pricing models in terms of sample Sharpe ratios.[1] The value-weighted market portfolio has a monthly sample Sharpe ratio of 0.133. The seven multi-factor asset pricing models all produce significantly higher sample Sharpe ratios than that of the CAPM, and the sample Sharpe ratios steadily increase with newer models, ranging from 0.184 for FF-3 to 0.599 for HMXZ q^5. Given the work of Barillas and Shanken (2017), this trend is not entirely surprising. Superior asset pricing models that are uncovered over time should have Sharpe ratios that are higher than those of past asset pricing models.

However, the conventional wisdom in finance suggests that high Sharpe ratios are not

[1] Barillas and Shanken (2017) suggest that comparing models with traded factors can be reduced to a comparison of their Sharpe ratios, and such comparison is independent of the choice of test assets.

easy to find and are unlikely to survive. A Sharpe ratio that is twice that of the market portfolio is considered to be very high (e.g., Ross, 1976; MacKinlay, 1995; Cochrane and Saá-Requejo, 2000).[2] Therefore, the high Sharpe ratios of the multi-factor models that we observe, especially the more recent ones, do not seem to be consistent with this widely accepted view.

There are several possible reasons why investors cannot obtain the high Sharpe ratios we observe for the multi-factor asset pricing models. First, market frictions, such as short-selling constraints, transaction costs, and taxes, can prevent investors from realizing the observed returns of the factor portfolios, especially those that involve long-short portfolios (e.g., Novy-Marx and Velikov, 2016; Patton and Weller, 2020; Detzel, Novy-Marx, and Velikov, 2023). Second, investors may not have confidence that the high sample Sharpe ratios of the multi-factor models can be sustained. The underlying true performance of the model may change over time. For example, publishing the model (and making investors aware of the new findings) can potentially influence the model performance in the post-publication period (e.g., Welch and Goyal, 2008; McLean and Pontiff, 2016). In addition, the published multi-factor asset pricing models may be subject to the repeated testing problem (e.g., Lo and MacKinlay, 1990; Harvey, Liu, and Zhu, 2016), and the surviving models may have unusually high sample Sharpe ratios compared to their population Sharpe ratios. This is particularly true for models that are motivated by anomalies, yet models rooted in theories are not exempt from this issue.

There is another reason that investors cannot realistically attain the sample Sharpe ratios of the multi-factor models. At the time of portfolio construction, the observed Sharpe ratio of a multi-factor model is computed based on the historical returns of the *ex post* optimal portfolio. We refer to this Sharpe ratio as the in-sample Sharpe ratio, which is unattainable for investors. When the out-of-sample performance is considered, investors would like to hold the true optimal portfolio (and obtain the population Sharpe ratio), which is constructed using the true mean and covariance matrix of the factors. In practice, these parameter values are unknown to investors. When investors have only historical data to work with, they need to estimate the optimal portfolio, which we call the sample optimal portfolio. What investors can get is what we term as the out-of-sample Sharpe ratio, that is, the Sharpe ratio of the

[2] Such a view is also supported by evidence from the investing world. For example, the Sharpe ratio of Berkshire Hathaway Inc. (ticker symbol: BRK.A), which was managed by Warren Buffett, arguably the most illustrious investor of our generation, is 0.227 for the period from November 1976 to December 2021.

sample optimal portfolio in the out-of-sample period. The out-of-sample Sharpe ratio clearly deviates from the in-sample Sharpe ratio, and it is subject to estimation risk. In this paper, we focus on understanding the effect of estimation risk.

We first empirically examine the gap between the in-sample and the out-of-sample Sharpe ratios of the multi-factor asset pricing models. Specifically, we divide our sample period into the estimation window and the out-of-sample period, and then compute and compare the in-sample and the out-of-sample Sharpe ratios of the models. The results for all the multi-factor models show that the out-of-sample Sharpe ratios are significantly lower than the in-sample ones. We also find that all the multi-factor models significantly outperform the CAPM based on the in-sample Sharpe ratio, but most of the significance disappears when we consider the out-of-sample Sharpe ratio.

One possible reason for the observed gap between the in-sample and the out-of-sample Sharpe ratios of the multi-factor models is that the model performance is different in the estimation window and the out-of-sample period due to e.g., the publication decay, the repeated testing problem, or simply random sampling. To control for this potential effect, we conduct a bootstrap exercise, and compute the in-sample and the out-of-sample Sharpe ratios based on the simulated data. Across 10,000 simulations, the average out-of-sample Sharpe ratio of a given multi-factor model continues to present a performance deterioration relative to the corresponding average in-sample Sharpe ratio. While the magnitude of this deterioration is smaller than what we observe empirically, it remains substantial. Estimation risk is the likely explanation for this remaining gap between the in-sample and the out-of-sample Sharpe ratios of the multi-factor models.

We next theoretically examine the effect of estimation risk on the performance of a multi-factor asset pricing model. Specifically, we assume that returns of the traded factors are identically and independently distributed (i.i.d.) with a multivariate normal distribution, and then study the properties of the in-sample and the out-of-sample Sharpe ratios of a multi-factor model. In this framework, the opportunity set is constant, and estimation risk is the only reason for the gap between the in-sample and the out-of-sample Sharpe ratios.

We obtain a simple stochastic representation of the in-sample and the out-of-sample Sharpe ratios of a multi-factor asset pricing model, and derive the finite sample distributions using the stochastic representation. We show that the out-of-sample Sharpe ratio is always lower than the population Sharpe ratio, whereas the in-sample Sharpe ratio is an upward biased estimator of the population Sharpe ratio. In addition, we find that the difference between the out-of-sample Sharpe ratio and the in-sample Sharpe ratio is negatively correlated with the in-sample Sharpe ratio. Therefore, the observed in-sample Sharpe ratio is not a reliable indicator of what an investor can obtain out-of-sample.

Due to estimation risk, the out-of-sample Sharpe ratio of a multi-factor model is lower than its population Sharpe ratio. Since the out-of-sample Sharpe ratio is the relevant one for investors, we argue that the effect of estimation risk needs to be considered when comparing asset pricing models. Because the out-of-sample Sharpe ratio of the CAPM is free from estimation risk, we select the CAPM as the benchmark model to illustrate the comparison. When the estimation risk is taken into account, the equality in the population Sharpe ratio does not necessarily suggest that two models are equally good; a multi-factor model needs a higher population Sharpe ratio (i.e., the break-even Sharpe ratio) to be comparable to the CAPM.

We define the break-even Sharpe ratio based on the theoretically derived finite sample distribution of the out-of-sample Sharpe ratio. As this distribution is significantly left-skewed, we propose using an expected shortfall measure to obtain the break-even Sharpe ratio. In both the classical frequentist framework and the Bayesian framework, we illustrate how to incorporate the break-even Sharpe ratio in model comparisons. With the break-even Sharpe ratio, we find that it is less likely to reject the null hypothesis that a multi-factor model is as good as the CAPM, and the effect is stronger when the estimation window is short.

Other than the finite sample distributions, the stochastic representation also enables us to obtain the limiting distributions of the in-sample and the out-of-sample Sharpe ratios. We consider two different asymptotic distributions: i) the number of traded factors (N) is fixed when the length of the estimation window (T) goes to infinity, and ii) $N \to \infty$, $T \to \infty$, and $N/T \to \rho \in (0, 1)$. When N is fixed and $T \to \infty$, the limiting distribution of the in-sample Sharpe ratio is well known, but that of the out-of-sample Sharpe ratio is new. We show that the limiting distribution of the out-of-sample Sharpe ratio is proportional to a chi-squared distribution and it converges to the population Sharpe ratio at a rate of $1/T$. For the case $N \to \infty$, $T \to \infty$, and $N/T \to \rho \in (0, 1)$, the limiting distributions are currently unavailable in the literature. We provide such distributions and show that neither the in-sample Sharpe ratio nor the out-of-sample Sharpe ratio converge to the population Sharpe ratio in this case.

Researchers may opt to use the asymptotic distribution instead of the finite sample distribution due to its simplicity. This is particularly appealing when the approximation error of the limiting distribution is small. We evaluate the accuracy of the two asymptotic distributions using our finite sample results. In approximating the exact distribution of the in-sample Sharpe ratio, the traditional fixed N asymptotic does not perform well but the fixed N/T asymptotic works well, even when N is small. In approximating the exact distribution of the out-of-sample Sharpe ratio, the fixed N asymptotic works well only for small N. As N increases, the fixed N/T asymptotic starts to do a better job than the fixed N asymptotic, but both approximations significantly deviate from the exact distribution. Therefore, unless N is small, the exact distribution is a better choice to draw inferences.

Our focus on the Sharpe ratios of the asset pricing models is built on Barillas and Shanken (2017), who argue that with traded factors, the comparison of asset pricing models can be reduced to a comparison of the Sharpe ratios. Based on this argument, model comparison in the Bayesian framework is developed (e.g., Barillas and Shanken, 2018; Chib, Zeng, and Zhao, 2020; Chib, Zhao, and Zhou, 2024). Barillas, Kan, Robotti, and Shanken (2020) show how to conduct asymptotically valid model comparisons. Fama and French (2018) use out-of-sample Sharpe ratios from bootstrap simulations to compare models. All of these studies focus on testing the equality of the population Sharpe ratios of different models. We argue that given estimation risk, the equality in the population Sharpe ratio does not necessarily suggest that two models are equally good. We recommend the effect of estimation risk be included when comparing asset pricing models.

Previous studies propose taking into account the effect of market frictions when evaluating and comparing asset pricing models. For example, Detzel, Novy-Marx, and Velikov (2023) present model comparison results after incorporating transaction costs. Li, DeMiguel, and Martin-Utrera (2023) provide a formal statistical test to compare factor models with price impact. Our paper adds to this strand of literature by recommending that the effect of estimation risk be included in model comparisons.

Our empirical results that the out-of-sample Sharpe ratios of the multi-factor asset pricing models are significantly lower than the corresponding in-sample Sharpe ratios are consistent with the findings from other papers (e.g., Welch and Goyal, 2008; McLean and Pontiff, 2016). McLean and Pontiff (2016) show that academic publications can contribute to a decrease in performance, suggesting that the underlying model parameters are likely to be time varying. In addition, poor out-of-sample Sharpe ratios can be the result of the repeated testing problem (e.g., Lo and MacKinlay, 1990; Harvey, Liu, and Zhu, 2016). We present evidence that estimation risk is another substantial contributor to the documented performance deterioration.

Finally, the effect of estimation risk has been studied extensively in the portfolio choice literature. The theoretical analyses in this paper are built and developed from the analytical tools provided in e.g., Kan and Zhou (2007), Kan, Wang, and Zhou (2022), and Kan and Wang (2023).

The remainder of the paper is organized as follows. In Section 2, we report some empirical results to motivate the theoretical analysis. In Section 3, we outline the problem and present our theoretical results. In Section 4, we illustrate how to compare a multi-factor model against the CAPM, taking into account the estimation risk. We conclude in Section 5. The Appendix contains all the proofs.

2. Some empirical results

In this section, we examine the empirical performance of the eight asset pricing models: (1) CAPM, (2) FF-3, (3) Carhart-4, (4) BAB, (5) FF-5, (6) HXZ q, (7) BS-6, and (8) HMXZ q^5. The sample period is 1967/1–2021/12. We obtain the monthly factor returns of the CAPM, FF-3, Carhart-4, and FF-5 from Ken French's website. Monthly returns of the q and q^5 factors are downloaded from `global-q.org`. Monthly returns of the betting-against-beta factor in BAB and the monthly updated value factor in BS-6 are available from AQR's website.

In Table 1, we report the maximum sample Sharpe ratios of the eight asset pricing models using data for the full sample period (1967/1–2021/12), as well as two subperiods (1967/1–1994/6 and 1994/7–2021/12). The year in which the model was first published is also presented in the table. In addition, we compare the performance of a given multi-factor model with that of the CAPM using the Gibbons-Ross-Shanken (1989) F-test (i.e., the GRS test), and report the corresponding significance in the table.

Table 1 about here

Over the 1967/1–2021/12 period, the CAPM has a monthly sample Sharpe ratio of 0.133, and all the multi-factor asset pricing models produce statistically significantly higher (at the 1% level) sample Sharpe ratios than that of the CAPM. Note that the sample Sharpe ratios for the asset pricing models steadily increase with their dates of publication. It starts with

0.133 for the CAPM in 1964, increases to 0.184 for FF-3 in 1993, and reaches 0.599 for HMXZ q^5 in 2021. A similar pattern holds for both subperiods (1967/1–1994/6 and 1994/7–2021/12). Except for the CAPM, the sample Sharpe ratios of the asset pricing models are lower in the second subperiod.[3]

The sample Sharpe ratios of the multi-factor models in Table 1 are computed based on the *ex post* optimal portfolios, and we call such Sharpe ratio the in-sample Sharpe ratio, as discussed above. The in-sample Sharpe ratios are not attainable for investors in practice. In addition, investors do not know the true mean and covariance matrix of the traded factors and have only historical data to work with. Therefore, they are not able to construct the true optimal portfolio, and the population Sharpe ratio is also not attainable. What investors can construct and hold out-of-sample is the optimal portfolio based on parameters estimated using historical data (i.e., the sample optimal portfolio). As discussed above, we call the Sharpe ratio of the sample optimal portfolio computed using the out-of-sample returns the out-of-sample Sharpe ratio.

Table 2 about here

In Table 2, we present the out-of-sample Sharpe ratios of the asset pricing models, and compare them to the in-sample Sharpe ratios. Specifically, we divide the sample period into halves, and treat the first half as the estimation window and the second half as the out-of-sample period. We assume an investor estimates the sample optimal portfolio using data in the estimation window and holds it throughout the out-of-sample period. The out-of-sample Sharpe ratio (OS-SR) is computed using the returns in the out-of-sample period of the sample optimal portfolio; the in-sample Sharpe ratio (IS-SR) is computed using the returns in the estimation window of the sample optimal portfolio.[4] In the left (right) panel of the table, the sample period is 1967/1–2021/12 (1994/12–2021/12). The in-sample Sharpe ratio of a multi-factor model is compared to that of the CAPM using the GRS test. The corresponding out-of-sample Sharpe ratio comparison is based on a one-sided test using the asymptotic distribution.[5]

[3]The exact mechanism underlying this dynamic pattern of the sample Sharpe ratio is out of the scope of this paper, and we leave it for future study.

[4]A similar empirical exercise is performed by Fama and French (2018). They use the out-of-sample Sharpe ratio to infer which asset pricing model has a higher population Sharpe ratio based on a bootstrap experiment.

[5]This approach, suggested by Jobson and Korkie (1981), tests whether the population Sharpe ratios of

Table 2 shows that for the multi-factor models, the out-of-sample Sharpe ratios (i.e., OS-SR) are all inferior to their in-sample Sharpe ratios (i.e., IS-SR), often by a substantial amount. For the 1967/1–2021/12 sample period (left panel), the performance deterioration of OS-SR relative to IS-SR (i.e., (OS-SR − IS-SR)/IS-SR) ranges from 46.97% for HMXZ q^5 to 71.28% for BS-6. For the 1994/7–2021/12 sample period (right panel), the range is from 13.60% for BAB to 68% for Carhart-4. In both panels, all of the multi-factor models have an IS-SR that is significantly higher than that of the CAPM. The significance, however, disappears in most cases when the OS-SR is considered. Some of the multi-factor models even have an OS-SR that is smaller than that of the CAPM.

At the time of portfolio construction, only the IS-SR is observable to the investor. The results in Table 2 suggest that the observed IS-SR is not a reliable indicator of what the investor can obtain out-of-sample.

Table 3 about here

One potential reason for the gap between OS-SR and IS-SR documented in Table 2 is that the model performance is poorer in the out-of-sample period than that in the estimation window. To control for such effect, we conduct a bootstrap simulation exercise. Specifically, we apply the stationary block bootstrap procedure proposed in Politis and Romano (1994) to our dataset to generate T monthly data, with an expected block length of 10 months. The T simulated monthly data are divided into halves, with the first half treated as the estimation window and the second half as the out-of-sample period. IS-SR and OS-SR, similar to those in Table 2, are computed using the simulated data. We run such simulation 10,000 times, and report the average values of IS-SR and OS-SR in Table 3. The standard deviations of IS-SR and OS-SR across 10,000 simulations are shown in brackets. We examine both $T = 660$ (i.e., the full length of our sample period) and $T = 330$.

Table 3 also shows that there is a substantial performance deterioration for OS-SR relative to IS-SR for all the multi-factor models. The magnitude of the deterioration is, however, smaller in Table 3 than that in Table 2 for all the cases except for FF-5 with $T = 330$. When $T = 660$, the magnitude of the deterioration ranges from 5.38% for BAB to 22.75% for FF-3. When $T = 330$, the magnitude increases, ranging from 10.65% for BAB to 35.68% for FF-3. These findings suggest that time-varying model performance indeed contributes to

two portfolios are equal to each other.

the gap between IS-SR and OS-SR (in Table 2),[6] but it clearly does not explain the entire gap. The remaining gap is still sizable, and the estimation risk is the likely explanation for the remaining gap.

3. Theoretical distributions of in-sample and out-of-sample Sharpe ratios

In this section, we conduct a theoretical analysis and obtain the distributions of the in-sample and the out-of-sample Sharpe ratios of a multi-factor asset pricing model to gain a better understanding of the effect of estimation risk.

3.1. The setup

We consider a multi-factor asset pricing model with N traded factors and $N \geq 2$. Let r_t be the excess returns of the N traded factors at time t. The elements of r_t can be returns of risky assets in excess of the risk-free rate, or they can be return differences of two risky assets. We define the mean and covariance matrix of r_t as $\mu = \mathbb{E}[r_t]$ and $\Sigma = \mathrm{Var}[r_t]$, respectively. We assume μ is a nonzero vector and Σ is positive definite. For a mean-variance investor who wants to hold a portfolio with a target standard deviation of σ, it is easy to show that his optimal portfolio has weights of

$$w^* = \frac{\sigma}{\theta}\Sigma^{-1}\mu$$

in the N traded factors, where $\theta = \sqrt{\mu'\Sigma^{-1}\mu}$ is the maximum Sharpe ratio that one can obtain from the N factors (i.e., the population Sharpe ratio). Obviously, the Sharpe ratio of portfolio w^* is θ:

$$\frac{w^{*\prime}\mu}{\sqrt{w^{*\prime}\Sigma w^*}} = \frac{\mu'\Sigma^{-1}\mu}{\sqrt{\mu'\Sigma^{-1}\mu}} = \sqrt{\mu'\Sigma^{-1}\mu} = \theta. \tag{1}$$

In practice, the investor does not know the mean and covariance matrix of the factors, and therefore, the optimal portfolio w^* and θ are unattainable. Suppose the investor estimates μ and Σ using historical data on r_t for $t = 1, \ldots, T$. The sample estimators of μ and Σ are

[6] There are different potential explanations for the time-varying model performance. One possibility is that the parameter values (i.e., the expected returns and the covariance matrix of the traded factors) are time varying. Explicitly modeling the time-varying parameters and exploring the corresponding implications on the out-of-sample Sharpe ratios of the asset pricing models is an interesting problem but beyond the scope of this paper.

given by

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t, \tag{2}$$

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(r_t - \hat{\mu})(r_t - \hat{\mu})', \tag{3}$$

and the sample estimator of θ is

$$\hat{\theta} = \sqrt{\hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}}. \tag{4}$$

A natural estimator of w^* is

$$\hat{w} = \frac{\sigma}{\hat{\theta}}\hat{\Sigma}^{-1}\hat{\mu}, \tag{5}$$

which we call the sample optimal portfolio. It is easy to see that $\hat{\theta}$ is the Sharpe ratio of the sample optimal portfolio in the estimation window, i.e., $\hat{w}'\hat{\mu}/\sqrt{\hat{w}'\hat{\Sigma}\hat{w}} = \sqrt{\hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}} = \hat{\theta}$, which is an *ex post* measure of performance and it is unattainable for investors. We call $\hat{\theta}$ the in-sample Sharpe ratio of the factors (or asset pricing model).

The out-of-sample mean and variance of the sample optimal portfolio are

$$\hat{w}'\mu = \frac{\sigma}{\hat{\theta}}\hat{\mu}'\hat{\Sigma}^{-1}\mu, \tag{6}$$

$$\hat{w}'\Sigma\hat{w} = \frac{\sigma^2}{\hat{\theta}^2}\hat{\mu}'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\mu}. \tag{7}$$

The corresponding Sharpe ratio is then given by

$$\tilde{\theta} = \frac{\hat{w}'\mu}{\sqrt{\hat{w}'\Sigma\hat{w}}} = \frac{\hat{\mu}'\hat{\Sigma}^{-1}\mu}{(\hat{\mu}'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\mu})^{\frac{1}{2}}}. \tag{8}$$

We call $\tilde{\theta}$ the out-of-sample Sharpe ratio of the factors (or asset pricing model). Unlike $\hat{\theta}$ or θ, $\tilde{\theta}$ is what an investor can obtain out-of-sample by holding the sample optimal portfolio $\hat{w}$.

Note that both $\hat{\theta}$ and $\tilde{\theta}$ are random variables because they depend on the realizations of $\hat{\mu}$ and $\hat{\Sigma}$. Next, we derive the distribution of $\hat{\theta}$ and $\tilde{\theta}$, assuming that r_t is i.i.d. with a multivariate normal distribution that has mean μ and covariance Σ.

3.2. Stochastic representation

In Proposition 1, we first provide a stochastic representation of $(\hat{\theta}, \tilde{\theta})$ that only depends on four univariate random variables (instead of $\hat{\mu}$ and $\hat{\Sigma}$). Such a representation greatly simplifies the problem and facilitates the derivation of the distributions of $(\hat{\theta}, \tilde{\theta})$.

Proposition 1. *Suppose* $r_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \Sigma)$ *and* $N \geq 2$. *Let* $u_1 \sim \chi^2_{T-N}$, $b \sim \text{Beta}((T-N+1)/2, (N-1)/2)$, *and they are independent of each other. Conditional on* b, *let* $\tilde{z} \sim \mathcal{N}(\sqrt{b}\sqrt{T}\theta, 1)$ *and* $\tilde{u} \sim \chi^2_{N-1}((1-b)T\theta^2)$, *and they are independent of each other and* u_1, *where* $\chi^2_\nu(\delta)$ *stands for a noncentral chi-squared random variable with* ν *degrees of freedom and a noncentrality parameter of* δ. *We have*

$$\hat{\theta} \overset{d}{=} \frac{\sqrt{\tilde{z}^2 + \tilde{u}}}{\sqrt{u_1}}, \tag{9}$$

$$\tilde{\theta} \overset{d}{=} \frac{\theta \tilde{z}}{\sqrt{\tilde{z}^2 + \tilde{u}}}. \tag{10}$$

Proposition 1 reveals that instead of the individual elements in μ and Σ, the joint distribution of $(\hat{\theta}, \tilde{\theta})$ depends only on N, T, and θ. Therefore, for asset pricing models with the same number of factors (i.e., N), as long as they have the same θ, their joint distributions of $(\hat{\theta}, \tilde{\theta})$ are identical regardless of the values of μ and Σ.

With the stochastic representation of $(\hat{\theta}, \tilde{\theta})$ in Proposition 1, we can obtain the exact moments and joint moments of $\hat{\theta}$ and $\tilde{\theta}$. Lemma 1 presents some low order moments of $\hat{\theta}$ and $\tilde{\theta}$.

Lemma 1. *Suppose* $r_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \Sigma)$ *and* $N \geq 2$. *We have*

$$\mathbb{E}[\hat{\theta}] = \frac{\Gamma\left(\frac{N+1}{2}\right)\Gamma\left(\frac{T-N-1}{2}\right)}{\Gamma\left(\frac{N}{2}\right)\Gamma\left(\frac{T-N}{2}\right)} {}_1F_1\left(-\frac{1}{2}; \frac{N}{2}; -\frac{T\theta^2}{2}\right) \quad \text{for } T \geq N+2, \tag{11}$$

$$\mathbb{E}[\tilde{\theta}] = \frac{\theta^2\sqrt{T}\Gamma\left(\frac{N+1}{2}\right)\Gamma\left(\frac{T-N+2}{2}\right)\Gamma\left(\frac{T}{2}\right)}{\sqrt{2}\Gamma\left(\frac{N+2}{2}\right)\Gamma\left(\frac{T-N+1}{2}\right)\Gamma\left(\frac{T+1}{2}\right)} {}_1F_1\left(\frac{1}{2}; \frac{N+2}{2}; -\frac{T\theta^2}{2}\right) \quad \text{for } T \geq N+1, \tag{12}$$

$$\mathbb{E}[\hat{\theta}^2] = \frac{T\theta^2 + N}{T-N-2} \quad \text{for } T \geq N+3, \tag{13}$$

$$\mathbb{E}[\tilde{\theta}^2] = \theta^2\left[\frac{T-N+1}{T} - \frac{(N-1)(T-N)}{NT} {}_1F_1\left(1; \frac{N+2}{2}; -\frac{T\theta^2}{2}\right)\right] \quad \text{for } T \geq N+1, \tag{14}$$

$$\mathbb{E}[\hat{\theta}\tilde{\theta}] = \frac{\theta^2\sqrt{T}(T-N)\Gamma\left(\frac{T}{2}\right)}{\sqrt{2}(T-N-1)\Gamma\left(\frac{T+1}{2}\right)} \quad \text{for } T \geq N+2, \tag{15}$$

where $\Gamma(a)$ *is the gamma function and* ${}_1F_1(a; b; x)$ *is the confluent hypergeometric function.*

The stochastic representation in Proposition 1 and the expressions in Lemma 1 enable us to derive some important inequalities, which are presented in Lemma 2.

Lemma 2. *Suppose* $r_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \Sigma)$ *and* $N \geq 2$. *We have*

$$\mathbb{E}[\tilde{\theta} - \hat{\theta}] < 0, \tag{16}$$

$$\text{Cov}[\tilde{\theta} - \hat{\theta}, \hat{\theta}] < 0. \tag{17}$$

From (10), it is easy to see that $\tilde{\theta} < \theta$ because $\left|\tilde{z}/\sqrt{\tilde{z}^2 + \tilde{u}}\right| < 1$. This finding is not surprising: since $\hat{w}$ is estimated with errors, the out-of-sample Sharpe ratio of $\hat{w}$ is not as good as the Sharpe ratio of the true optimal portfolio. In addition, we know that $\hat{\theta}$ is an upward biased estimator of θ, i.e., $\mathbb{E}[\hat{\theta}] > \theta$. This is because the in-sample Sharpe ratio is computed after $\hat{\mu}$ and $\hat{\Sigma}$ are observed, and the look-ahead bias allows the sample optimal portfolio to have a better in-sample performance than the population Sharpe ratio on average. Thus, the out-of-sample Sharpe ratio is lower than the in-sample Sharpe ratio on average (i.e., $\mathbb{E}[\tilde{\theta} - \hat{\theta}] < 0$). The second inequality in Lemma 2 suggests that an investor would experience a larger disappointment (in terms of $\tilde{\theta} - \hat{\theta}$) when the asset pricing model realizes a higher in-sample Sharpe ratio. This confirms our finding from the empirical results that the observable in-sample Sharpe ratio is not a reliable indicator of what an investor can get out-of-sample.

Fig. 1 about here

In Fig. 1, we plot $\mathbb{E}[\tilde{\theta} - \hat{\theta}]$ as a function of the population Sharpe ratio θ for two different values of N (3 and 6) and T (120 and 240). It is evident that estimation errors have a substantial negative impact on the out-of-sample performance of a multi-factor asset pricing model, and the influence of estimation risk is more pronounced for a larger model and a shorter estimation window. For example, when $\theta = 0.1$, the out-of-sample Sharpe ratio is, on average, 0.204 lower than the in-sample Sharpe ratio for $T = 120$ and $N = 6$. Conversely, for $T = 240$ and $N = 3$, this reduction is 0.075, which still represents 75% of the population Sharpe ratio. In addition, Fig. 1 shows that the magnitude of the reduction is decreasing in θ, which is due to a higher signal-to-noise ratio associated with larger θ.

3.3. Theoretical distributions

With the stochastic representation in Proposition 1, we derive both the finite sample distributions and the asymptotic distributions of $\hat{\theta}$ and $\tilde{\theta}$ in this subsection.

3.3.1. Finite sample distributions

For the marginal distribution of $\hat{\theta}$, we can write $\hat{\theta} \stackrel{d}{=} \sqrt{u_3}/\sqrt{u_1}$, where $u_3 = \tilde{z}^2 + \tilde{u} \sim \chi^2_N(T\theta^2)$. This implies that $\hat{\theta}^2 \stackrel{d}{=} u_3/u_1$ is proportional to a noncentral F-distribution with degrees of freedom N and $T - N$ and a noncentrality parameter $T\theta^2$, a well-known result in the literature (e.g., Gibbons, Ross, and Shanken, 1989; Kan and Robotti, 2016). Let $F^{\delta}_{m,n}(\cdot)$ and $f^{\delta}_{m,n}(\cdot)$ denote the cumulative distribution function and the density function of a noncentral F random variable with m and n degrees of freedom and a noncentrality parameter δ. The distribution and density functions of $\hat{\theta}$ are given by

$$\mathbb{P}[\hat{\theta} < c] = \mathbb{P}[\hat{\theta}^2 < c^2] = F^{T\theta^2}_{N,T-N}\left(\frac{(T-N)c^2}{N}\right), \tag{18}$$

$$f_{\hat{\theta}}(c) = \frac{2(T-N)c}{N} f^{T\theta^2}_{N,T-N}\left(\frac{(T-N)c^2}{N}\right). \tag{19}$$

In Fig. 2, we plot the density function of $\hat{\theta}/\theta$ for two different values of N (3 and 6) and θ (0.2 and 0.4) with $T = 120$. As pointed out in the previous subsection, $\hat{\theta}$ is an upward biased estimator of θ. Fig. 2 shows that the bias increases with N and decreases with θ. Thus, an asset pricing model with more factors tends to have a higher in-sample Sharpe ratio on average, even though the population Sharpe ratio may not improve from having more factors.

Fig. 2 about here

To derive the marginal distribution of $\tilde{\theta}$, we define $q = \tilde{z}/\sqrt{\tilde{u}}$ and obtain

$$\tilde{\theta} \stackrel{d}{=} \frac{\theta \tilde{z}}{\sqrt{\tilde{z}^2 + \tilde{u}}} = \frac{\theta q}{\sqrt{1+q^2}}. \tag{20}$$

Expression (20) suggests that $\tilde{\theta}$ is a monotonically increasing function of q and $-\theta \leq \tilde{\theta} \leq \theta$. Conditional on b, q is proportional to a doubly noncentral t-distribution, and we can compute the cumulative distribution function of $\tilde{\theta}$ using the following:

$$\begin{aligned}\mathbb{P}[\tilde{\theta} < c] &= \mathbb{P}\left[q < \frac{c}{\sqrt{\theta^2 - c^2}}\right] \\ &= \mathbb{P}\left[\tilde{z} < \frac{c\sqrt{\tilde{u}}}{\sqrt{\theta^2 - c^2}}\right] \\ &= \int_0^1 \int_0^\infty \Phi\left(\frac{c\sqrt{\tilde{u}}}{\sqrt{\theta^2 - c^2}} - \sqrt{b}\sqrt{T}\theta\right) f_{\tilde{u}}(\tilde{u}) f_b(b) \mathrm{d}\tilde{u}\mathrm{d}b \quad \text{for } -\theta < c < \theta,\end{aligned} \tag{21}$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal, $f_{\tilde{u}}(\tilde{u})$ is the density function of $\tilde{u} \sim \chi^2_{N-1}((1-b)T\theta^2)$, and $f_b(b)$ is the density function of $b \sim \text{Beta}((T-N+1)/2, (N-1)/2)$. Taking the derivative yields the density function of $\tilde{\theta}$ as follows:

$$f_{\tilde{\theta}}(c) = \frac{\theta^2}{(\theta^2 - c^2)^{\frac{3}{2}}} \int_0^1 \int_0^\infty \phi\left(\frac{c\sqrt{\tilde{u}}}{\sqrt{\theta^2 - c^2}} - \sqrt{b}\sqrt{T}\theta\right)\sqrt{\tilde{u}} f_{\tilde{u}}(\tilde{u}) f_b(b) \mathrm{d}\tilde{u}\mathrm{d}b \quad \text{for } -\theta < c < \theta, \tag{22}$$

where $\phi(\cdot)$ is the density function of a standard normal.

Fig. 3 about here

In Fig. 3, we plot the density function of $\tilde{\theta}/\theta$ for two different values of N (3 and 6) and θ (0.2 and 0.4) with $T = 120$.[7] The figure shows that $\tilde{\theta}$ is quite volatile and highly left-skewed. When N is large and θ is small, there is a substantial deterioration in the out-of-sample performance when an investor holds the sample optimal portfolio. For example, when $N = 6$ and $\theta = 0.2$, we have $\mathbb{P}[\tilde{\theta}/\theta < 0.8] = 0.7027$, i.e., there is more than 70% probability that an investor will lose more than 20% of the Sharpe ratio due to estimation risk. And this probability drops down to 19% when $\theta = 0.4$.

Next, we derive the joint distribution of $(\hat{\theta}, \tilde{\theta})$. The joint cumulative distribution of $(\hat{\theta}, \tilde{\theta})$ can be written as a triple integral, whereas the joint density of $(\hat{\theta}, \tilde{\theta})$ can be written as a double integral. These expressions are summarized in Proposition 2.

Proposition 2. *Suppose* $r_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \Sigma)$ *and* $N \geq 2$. *When* $c_1 > 0$ *and* $0 < c_2 \leq \theta$, *we have*

$$\begin{aligned} &\mathbb{P}[\hat{\theta} < c_1, \tilde{\theta} < c_2] \\ &= \int_0^\infty \int_0^1 \int_0^{c_1^2 v} \left[\Phi\left(\min\left[\frac{c_2\sqrt{\tilde{u}}}{\sqrt{\theta^2 - c_2^2}}, \sqrt{c_1^2 v - \tilde{u}}\right] - \sqrt{T}\theta\sqrt{b}\right) - \Phi\left(-\sqrt{c_1^2 v - \tilde{u}} - \sqrt{T}\theta\sqrt{b}\right)\right] \\ &\quad \times f_{\tilde{u}}(\tilde{u}) f_b(b) f_{u_1}(v) \mathrm{d}\tilde{u}\mathrm{d}b\mathrm{d}v, \end{aligned} \tag{23}$$

where $f_{\tilde{u}}(\tilde{u})$ *is the density function of* $\chi^2_{N-1}((1-b)T\theta^2)$, $f_b(b)$ *is the density function of* $\text{Beta}((T-N+1)/2, (N-1)/2)$, *and* $f_{u_1}(v)$ *is the density function of* χ^2_{T-N}. *When* $c_1 > 0$ *and* $-\theta \leq c_2 \leq 0$, *we have*

$$\mathbb{P}[\hat{\theta} < c_1, \tilde{\theta} < c_2]$$

[7] It can be shown that when $N > 3$, $\lim_{c \to \theta^-} f_{\tilde{\theta}}(c) = 0$ and $\lim_{c \to (-\theta)^+} f_{\tilde{\theta}}(c) = 0$. When $N = 3$, these two limits are finite, and when $N = 2$, these two limits are infinity. Proof of these results as well as the explicit expressions of the two limits for the case of $N = 3$ are available upon request.

$$= \int_0^\infty \int_0^1 \int_0^{\frac{c_1^2(\theta^2-c_2^2)v}{\theta^2}} \left[\Phi\left(\frac{c_2\sqrt{\tilde{u}}}{\sqrt{\theta^2-c_2^2}} - \sqrt{T}\theta\sqrt{b}\right) - \Phi\left(-\sqrt{c_1^2 v - \tilde{u}} - \sqrt{T}\theta\sqrt{b}\right)\right]$$
$$\times f_{\tilde{u}}(\tilde{u}) f_b(b) f_{u_1}(v) \mathrm{d}\tilde{u}\mathrm{d}b\mathrm{d}v. \tag{24}$$

The joint density of $(\hat{\theta}, \tilde{\theta})$ is given by[8]

$$f_{\hat{\theta},\tilde{\theta}}(c_1, c_2) = \int_0^1 \int_0^\infty f_{\tilde{u}}\left(\frac{c_1^2(\theta^2 - c_2^2)v}{\theta^2}\right) \phi\left(\frac{c_1 c_2 \sqrt{v}}{\theta} - \sqrt{T}\theta\sqrt{b}\right) \frac{2c_1^2 v^{\frac{3}{2}}}{\theta} f_{u_1}(v) f_b(b) \mathrm{d}v\mathrm{d}b \tag{25}$$

for $c_1 > 0$ and $-\theta < c_2 < \theta$.

With the expression of the joint density of $(\hat{\theta}, \tilde{\theta})$ and the marginal density of $\hat{\theta}$, we can compute the density of $\tilde{\theta}$ conditional on $\hat{\theta}$ using the following:

$$f_{\tilde{\theta}|\hat{\theta}}(c_2|c_1) = \frac{f_{\hat{\theta},\tilde{\theta}}(c_1, c_2)}{f_{\hat{\theta}}(c_1)}. \tag{26}$$

In Fig. 4, we plot the conditional density of $\tilde{\theta}/\theta$ for two different values of N (3 and 6) and θ (0.2 and 0.4) with $T = 120$. The plots present the conditional density of $\tilde{\theta}/\theta$ for three different values of $\hat{\theta}$: the first one is at the 10th percentile of $\hat{\theta}$ (solid line), the second one is at the 50th percentile of $\hat{\theta}$ (dotted line), and the last one is at the 90th percentile of $\hat{\theta}$ (dashed line). The figure shows that the conditional density of $\tilde{\theta}$ can be quite sensitive to the value of $\hat{\theta}$.

Fig. 4 about here

3.3.2. Asymptotic distributions

Instead of the exact distribution, researchers may opt to use the asymptotic distribution because the asymptotic distribution is often simpler than the finite sample distribution. The simplicity is particularly appealing when the approximation error of the asymptotic distribution is small. In this subsection, we present the asymptotic distributions of $(\hat{\theta}, \tilde{\theta})$ and evaluate their accuracy using the finite sample results obtained previously.

There are two different limiting distributions for $(\hat{\theta}, \tilde{\theta})$, depending on whether N is fixed or $N \to \infty$ as $T \to \infty$. For both cases, we can use the stochastic representation in Proposition 1

[8] When N is even, the inner integrals of $f_{\tilde{\theta}}(c)$ and $f_{\hat{\theta},\tilde{\theta}}(c_1, c_2)$ can be solved analytically, so $f_{\tilde{\theta}}(c)$ and $f_{\hat{\theta},\tilde{\theta}}(c_1, c_2)$ can be evaluated using a single rather than a double integral. These results are available upon request.

to derive the limiting distributions of $\hat{\theta}$ and $\tilde{\theta}$. Proposition 3 presents the limiting distribution under the assumption that N is fixed when $T \to \infty$. This assumption is used for the traditional asymptotic analysis.

Proposition 3. *Suppose $r_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \Sigma)$ and $N \geq 2$. When N is fixed and $T \to \infty$, we have*

$$\begin{bmatrix} \sqrt{T}(\hat{\theta} - \theta) \\ T(\tilde{\theta} - \theta) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} X \\ Y \end{bmatrix}, \tag{27}$$

where $X \sim \mathcal{N}\left(0, 1 + \frac{\theta^2}{2}\right)$ and $Y \sim -(1+\theta^2)/(2\theta)\chi^2_{N-1}$, and they are independent of each other.

The limiting distribution of $\sqrt{T}(\hat{\theta} - \theta)$ is well known (e.g., Barillas, Kan, Robotti, and Shanken, 2020). Note that the asymptotic distribution of $\hat{\theta}$ is the same as the one for the single asset case as discussed in Lo (2002), suggesting that using $\hat{w}$ instead of w^* has no impact on the asymptotic distribution of $\hat{\theta}$.

Our result of the limiting distribution of $T(\tilde{\theta} - \theta)$ is new. There are two points to note here. First, unlike $\hat{\theta}$, which converges to θ at a rate of $1/\sqrt{T}$, $\tilde{\theta}$ converges to θ at a rate of $1/T$. Second, the limiting distribution of $T(\tilde{\theta} - \theta)$ is not normal; instead, it is distributed as a negative random variable that is proportional to χ^2_{N-1}. This is because while $\tilde{\theta}$ converges to θ, it is always less than θ in a finite sample, so the limiting distribution of $T(\tilde{\theta} - \theta)$ is a negative random variable.

Instead of the traditional asymptotic analysis, which assumes fixed N, Proposition 4 presents the limiting distribution assuming both N and T go to infinity, but $N/T \to \rho \in (0, 1)$.

Proposition 4. *Suppose $r_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \Sigma)$ and $N \geq 2$. Let $\theta(N)$ denote the population Sharpe ratio when N is a finite number and θ be the limit of $\theta(N)$ as $N \to \infty$ with $\theta(N) - \theta = o(1/\sqrt{N})$. When $N \to \infty$, $T \to \infty$, $N/T \to \rho \in (0, 1)$, we have*

$$\begin{bmatrix} \sqrt{T}(\hat{\theta} - \bar{\theta}) \\ \sqrt{T}(\tilde{\theta} - \underline{\theta}) \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0_2, \begin{bmatrix} \frac{\theta^4 + 2\theta^2 + \rho}{2(1-\rho)^2(\theta^2+\rho)} & \frac{\rho\theta^2}{2(\theta^2+\rho)^2} \\ \frac{\rho\theta^2}{2(\theta^2+\rho)^2} & \frac{\rho\theta^2}{2(\theta^2+\rho)}\left[\frac{(1-\rho)(2\rho+\theta^2)}{(\theta^2+\rho)^2} + 2 + \theta^2\right] \end{bmatrix}\right), \tag{28}$$

where

$$\bar{\theta} = \frac{\sqrt{\theta^2 + \rho}}{\sqrt{1-\rho}} > \theta, \tag{29}$$

$$\underline{\theta} = \frac{\theta^2\sqrt{1-\rho}}{\sqrt{\theta^2+\rho}} < \theta. \quad (30)$$

Unlike the fixed N asymptotic results, Proposition 4 shows that $\hat{\theta}$ and $\tilde{\theta}$ no longer converge to θ for the case of $N/T \to \rho$. This is because for fixed N, the effect of estimation risk diminishes to nil as $T \to \infty$, but it is not the case when $N \to \infty$ and $T \to \infty$.

The result that $\hat{\theta} \xrightarrow{p} \bar{\theta}$ under the assumption that $N/T \to \rho$ can be obtained from Theorem 4.6 of El Karoui (2010). The result that $\tilde{\theta} \xrightarrow{p} \underline{\theta}$ when $N/T \to \rho$ is shown in Ao, Li, and Zheng (2019). However, the limiting distribution of $(\hat{\theta}, \tilde{\theta})$ under the assumption that $N/T \to \rho$ is not available in the literature. The stochastic representation in Proposition 1 enables us to obtain this limiting distribution by taking the appropriate limit.

Next, we evaluate the accuracy of the two asymptotic distributions of $\hat{\theta}$ and $\tilde{\theta}$ by comparing them to the exact distributions obtained previously. In Fig. 5, we plot the exact marginal density of $\hat{\theta}/\theta$ (solid line) versus its two approximations: the fixed N asymptotic (dashed line) and the $N/T \to \rho$ asymptotic (dotted line). We consider two different values of N (3 and 6) and θ (0.2 and 0.4) with $T = 120$. The figure shows that the approximate distribution based on the traditional fixed N asymptotic does not perform well, especially for the cases with large N and small θ. By contrast, the approximation based on the fixed N/T asymptotic works quite well in all cases. Thus, Fig. 5 provides supporting evidence to use the fixed N/T asymptotic distribution to approximate the exact distribution of $\hat{\theta}$, if needed.

Fig. 5 about here

In Fig. 6, we plot the exact marginal density of $\tilde{\theta}/\theta$ (solid line) versus its two approximations: the fixed N asymptotic (dashed line) and the $N/T \to \rho$ asymptotic (dotted line). Similarly, we consider two different values of N (3 and 6) and θ (0.2 and 0.4) with $T = 120$. The figure shows that when $N = 3$, the fixed N asymptotic approximation works well, especially when $\theta = 0.4$. By contrast, the fixed N/T asymptotic approximation of $\tilde{\theta}$, which is a normal distribution, does a poor job in approximating the exact distribution of $\tilde{\theta}$. When $N = 6$, the fixed N/T asymptotic approximation continues to perform poorly; and the fixed N asymptotic approximation starts to deviate significantly from the exact distribution of $\tilde{\theta}$, especially for the case with $\theta = 0.2$. Therefore, unless N is very small, none of the two asymptotic approximations provides a reliable approximation of the exact distribution of $\tilde{\theta}$.

Fig. 6 about here

In terms of approximating the exact distribution of $\tilde{\theta}$ conditional on $\hat{\theta}$, Proposition 3 shows that when N is fixed and $T \to \infty$, the limiting distributions of $\hat{\theta}$ and $\tilde{\theta}$ are independent of each other, and therefore, the limiting conditional distribution of $\tilde{\theta}$ is the same as the marginal one. From Proposition 4, it is obvious that when both N and $T \to \infty$, the limiting distribution of $\tilde{\theta}$ conditional on $\hat{\theta}$ is a normal distribution, which clearly deviates from the exact conditional distribution of $\tilde{\theta}$, as shown in Fig. 4. Thus, none of the asymptotic approximations accurately approximates the exact conditional distribution of $\tilde{\theta}$.

4. Model comparison with estimation risk

The theoretical results in the previous section suggest that estimation errors in the sample optimal portfolio result in a lower out-of-sample Sharpe ratio. In this section, we examine its implication when comparing across asset pricing models. Because relative to the population Sharpe ratio, the out-of-sample Sharpe ratio is more relevant to investors, we argue that the effect of estimation risk needs to be considered when comparing across models.

In this section, we use the CAPM as the benchmark model because its out-of-sample Sharpe ratio is free from estimation risk. We then assess the performance of a multi-factor model (with the market factor as one of its factors) against this benchmark, taking into account the effect of estimation risk. When estimation risk is taken into account, the multi-factor model needs a higher population Sharpe ratio to be seen as equally good as the CAPM. In Section 4.1, we propose a way to obtain this break-even Sharpe ratio. In Sections 4.2 and 4.3, we illustrate how to incorporate the break-even Sharpe ratio in model comparisons, in both the classical frequentist framework and the Bayesian framework.

4.1. Break-even Sharpe ratio

When comparing asset pricing models, most of the tests focus on the equality of the population Sharpe ratios across models. However, when estimation risk is taken into account, the equality in the population Sharpe ratio does not necessarily suggest that two models are equally good. A larger model (in terms of number of factors) contains a higher level of estimation risk, and it needs a higher population Sharpe ratio (i.e., the break-even Sharpe ratio) to be seen as equally good as a smaller model.[9]

We define the break-even Sharpe ratio based on the finite sample distribution of the out-of-sample Sharpe ratio of a multi-factor model that is presented in (22). Let θ_1 and θ denote the population Sharpe ratio of the CAPM and an N-factor model (with $N \geq 2$), respectively. Because the out-of-sample Sharpe ratio of the CAPM is free from estimation risk, we have $\tilde{\theta}_1 = \theta_1$. The out-of-sample Sharpe ratio of the multi-factor model is a random variable that is smaller than its population Sharpe ratio, $\tilde{\theta} < \theta$, and it is severely left-skewed, as shown in Fig. 3. Investors are concerned about the randomness, and in particular, the long left tail in the distribution of $\tilde{\theta}$. We propose using an expected shortfall measure to obtain the break-even Sharpe ratio because such a measure aggregates the information in the left tail of a distribution.

Let $\tilde{\theta}_c$ be the c-percentile value of $\tilde{\theta}$. The expected shortfall of $\tilde{\theta}$ at c-percentile is $\mathbb{E}[\tilde{\theta}|\tilde{\theta} \leq \tilde{\theta}_c]$. At a selected value of c with $0 < c < 100$, we define the break-even Sharpe ratio as the θ such that $\mathbb{E}[\tilde{\theta}|\tilde{\theta} \leq \tilde{\theta}_c] = \theta_1$, and denote it as θ_b.[10] The choice of c, to some extent, reflects the risk preference of the investor: a lower value of c is associated with a higher break-even Sharpe ratio. We set $c = 50$ in our base case. From (22), it can be seen that at a given value of c, θ_b is a function of N, T, and θ_1. When estimation risk is taken into account, instead of testing $H_0 : \theta = \theta_1$ vs. $H_1 : \theta > \theta_1$, we test $H_0 : \theta \leq \theta_b$ vs. $H_1 : \theta > \theta_b$ to compare the multi-factor model to the CAPM.

Table 4 about here

Table 4 presents the break-even Sharpe ratio for different values of θ_1 ($0.05 \leq \theta_1 \leq 0.4$), N ($2 \leq N \leq 6$), and T (120 and 240).[11] Panels A and B report the results for $c = 50$

[9]The parsimony principle suggests that if two models have equal population Sharpe ratios, the smaller model is preferred. We argue that even if the larger model has a population Sharpe ratio that is slightly higher, it is not necessarily a better model. In order for the larger model to be a better model, the improvement in the population Sharpe ratio must outweigh the effect of estimation risk. Our proposed break-even Sharpe ratio is a measure to incorporate such a trade-off.

[10]When a multi-factor model, instead of the CAPM, is used as the benchmark model, the out-of-sample Sharpe ratio of the benchmark model is also random. Then the break-even Sharpe ratio will be defined based on the joint distribution of the out-of-sample Sharpe ratio of the benchmark model and that of the test model. When both nested and non-nested models are considered, the derivation of this joint distribution is not straightforward and is beyond the scope of this paper. We leave it as well as the design of a comprehensive methodology of model comparison with estimation risk for future study.

[11]Based on (22), the explicit expression of $\mathbb{E}[\tilde{\theta}|\tilde{\theta} \leq \tilde{\theta}_c]$ contains a triple integral, which is difficult and

and 25, respectively. The table shows that when the estimation risk is accounted for, the population Sharpe ratio of a multi-factor model needs to substantially surpass that of the CAPM for the two models to be deemed equivalent. For example, when $c = 50$, $T = 120$, and $\theta_1 = 0.10$ (Panel A1), even for a two-factor model, it needs a population Sharpe ratio at least 62% higher than that of the CAPM (i.e., $\theta_b = 0.162$). Not surprisingly, the break-even Sharpe ratio increases with N and decreases with T due to a higher level of estimation risk that is associated with a larger model and a shorter estimation window. In addition, the table shows that the required improvement in the population Sharpe ratio of the multi-factor model is larger when θ_1 is smaller due to a low signal-to-noise ratio.

4.2. GRS test and the break-even Sharpe ratio

Gibbons, Ross, and Shanken (1989) provide a well-known spanning test (i.e., the GRS test) for the *ex ante* efficiency of a given portfolio. To compare the performance of a multi-factor model with that of the CAPM, the GRS test statistic is given by

$$W = \frac{(T-N)\hat{\delta}^2}{(N-1)(1+\hat{\theta}_1^2)}, \tag{31}$$

where $\hat{\delta}^2 = \hat{\theta}^2 - \hat{\theta}_1^2$, and $\hat{\theta}$ and $\hat{\theta}_1$ are the in-sample Sharpe ratios of the multi-factor model and the CAPM, respectively. Under the i.i.d. normality assumption, this test statistic follows a noncentral F-distribution with $N-1$ and $T-N$ degrees of freedom and a noncentrality parameter of $T\delta^2/(1+\hat{\theta}_1^2)$ with $\delta^2 = \theta^2 - \theta_1^2$,

$$W \sim \mathcal{F}_{N-1,T-N}^{\frac{T\delta^2}{1+\hat{\theta}_1^2}}. \tag{32}$$

If the CAPM is *ex ante* efficient (i.e., $\delta^2 = 0$), then W follows an F-distribution with $N-1$ and $T-N$ degrees of freedom.

When the estimation risk is considered, instead of testing $H_0 : \delta^2 = 0$ vs. $H_1 : \delta^2 > 0$, we test $H_0 : \delta^2 \leq \delta_b^2$ vs. $H_1 : \delta^2 > \delta_b^2$, where $\delta_b^2 = \theta_b^2 - \theta_1^2$. And the null distribution of W is

$$W \sim \mathcal{F}_{N-1,T-N}^{\frac{T\delta_b^2}{1+\hat{\theta}_1^2}}. \tag{33}$$

Because θ_b is a function of θ_1, the null distribution now depends on the nuisance parameter θ_1. We examine the test results for various values of θ_1. Specifically, we assume that the

time consuming to obtain. Instead of the explicit expression, we compute the break-even Sharpe ratio using simulation based on the stochastic representation in Proposition 1.

observed in-sample Sharpe ratio of the CAPM ($\hat{\theta}_1$) is at a certain percentile of its distribution, and obtain the corresponding population Sharpe ratio (θ_1). Under the i.i.d. multivariate normality assumption, we know that the distribution of $\hat{\theta}_1^2$ is proportional to a noncentral F-distribution with 1 and $T-1$ degrees of freedom and a noncentrality parameter of $T\theta_1^2$,

$$(T-1)\hat{\theta}_1^2 \sim \mathcal{F}_{1,T-1}^{T\theta_1^2}. \tag{34}$$

For a given value of α with $0 < \alpha < 1$, the value of θ_1, such that $\hat{\theta}_1$ is at $100(1-\alpha)$-percentile of the distribution, can be determined from the following equation

$$F_{1,T-1}^{T\theta_1^2}\left((T-1)\hat{\theta}_1^2\right) = 1 - \alpha. \tag{35}$$

Since the cdf of the noncentral F-distribution is decreasing in its noncentrality parameter, θ_1 increases with α for a given $\hat{\theta}_1$. In Table 5, we consider five different values of α: 10%, 25%, 50%, 75%, and 90%. Panels A, B, and C report the results based on the most recent 25% (i.e., 2008/4–2021/12, $T = 165$), 50% (i.e., 1994/7–2021/12, $T = 330$), and 100% (i.e., 1967/1–2021/12, $T = 660$) of the sample, respectively. In all the panels, the column "GRS" reports the in-sample Sharpe ratios ($\hat{\theta}_1$ and $\hat{\theta}$) of the models using all T observations, together with the p-values associated with the GRS test. The remaining five columns report θ_1 (in the row of the CAPM) and θ_b (in the rows of the multi-factor models) for different values of α, as well as the p-values from the null distribution as specified in (33).

Table 5 about here

Because the cdf of the noncentral F-distribution is decreasing in its noncentrality parameter, and given that $\delta_b^2 > 0$, the p-values based on the null in (33) are larger than that from the GRS test, as shown in Table 5. In addition, the p-values increase with α. This is because δ_b^2 increases with α. Note that even though $\theta_b - \theta_1$ decreases with α, both θ_b and θ_1 increase with α, and as a result, $\delta_b^2 = \theta_b^2 - \theta_1^2$ increases with α. Across the three panels, the shortest estimation window (i.e., $T = 165$) is used in Panel A, and the effect due to estimation risk is more evident. In the 2008/4–2012/12 period, the GRS test results suggest that BAB outperforms the CAPM at the 5% significance level, and FF-5 and HMXZ q^5 outperform the CAPM at the 1% level. When the estimation risk is taken into account, BAB and FF-5 are no longer significant for any of the cases; and HMXZ q^5 becomes significant at the 5% level for most of the cases. In Panel B, whether to consider the effect of estimation risk leads to different conclusions for the performance of Carhart-4. The GRS test results suggest that

Carhart-4 outperforms the CAPM at the 5% level, and the significance disappears when the effect of estimation risk is considered. Even in Panel C, when we use the full sample, we continue to find some different conclusions. For FF-3, the GRS test results suggest that it outperforms the CAPM at the 1% level, but it is no longer the case when the estimation risk is taken into account.

4.3. The Bayesian framework

Other than the classical frequentist approach (e.g., the GRS test), the Bayesian approach has also been used in model comparisons. In this subsection, we investigate the difference in the test results when the estimation risk is taken into account in the Bayesian framework.

The classical approach ignores the prior information about the distribution of the parameter values, which is specified in the Bayesian approach. Following many previous studies (e.g., Harvey and Zhou, 1990; Pástor and Stambaugh, 2000; Barillas and Shanken, 2018), we assume the prior distribution of δ^2 is proportional to a chi-squared distribution with $N-1$ degrees of freedom[12]

$$\delta^2 \sim k\chi^2_{N-1}. \tag{36}$$

Following the base case specification in Barillas and Shanken (2018), we set $k = 0.25\theta_1^2$, which corresponds to a potential 50% increase in the Sharpe ratio of a six-factor model relative to that of the CAPM.

In addition to δ^2, we also need to specify the prior distribution of θ_1. We choose a bounded support for θ_1 (i.e., $0 \leq \theta_1 \leq 0.4$), and assume that the prior of θ_1 is a beta distribution

$$\theta_1 \sim 0.4 \times \text{Beta}(a, b), \tag{37}$$

where a and b are parameters to be specified.[13] The prior mean of θ_1 is given by

$$\mathbb{E}[\theta_1] = \frac{0.4a}{a+b}. \tag{38}$$

[12]This assumption is built on the positive link between the magnitude of alpha to the residual variance, and asset pricing theory provides some motivations for such link (e.g., Harvey and Zhou, 1990; MacKinlay, 1995; Pástor and Stambaugh, 2000; Barillas and Shanken, 2018).

[13]We assume that the prior distribution of δ^2 depends on θ_1 only through k in (36) with $k = 0.25\theta_1^2$, and the prior distribution of θ_1 does not depend on δ^2. Thus, the joint prior distribution of δ^2 and θ_1 is the product of the prior of δ^2 in (36) and the prior of θ_1 in (37), i.e., $f(\delta^2, \theta_1) = f(\delta^2|\theta_1)f(\theta_1)$.

We consider three different combinations of a and b. With $a = 3$ and $b = 3$, the prior distribution of θ_1 is symmetric with a mean of 0.20. With $a = 3$ and $b = 5$, the prior distribution of θ_1 is right-skewed with a mean of 0.15 and a mode of 0.133. With $a = 5$ and $b = 3$, the prior distribution of θ_1 is left-skewed with a mean of 0.25 and a mode of 0.267.

Given the prior distribution of δ^2 and θ_1 in (36) and (37), together with $\hat{\delta}^2$ and $\hat{\theta}_1^2$ from the data, the marginal likelihood under the null $H_0 : \delta^2 \leq \delta_b^2$ and that under the alternative $H_1 : \delta^2 > \delta_b^2$ can be computed as

$$ML_0(\delta_b^2) \equiv ML(H_0 : \delta^2 \leq \delta_b^2) = \int \int_0^{\delta_b^2} f\left(\hat{\delta}^2 \middle| \hat{\theta}_1^2, \delta^2\right) f\left(\hat{\theta}_1^2 \middle| \theta_1\right) f(\delta^2, \theta_1) \mathrm{d}\delta^2 \mathrm{d}\theta_1, \tag{39}$$

$$ML_1(\delta_b^2) \equiv ML(H_1 : \delta^2 > \delta_b^2) = \int \int_{\delta_b^2}^{\infty} f\left(\hat{\delta}^2 \middle| \hat{\theta}_1^2, \delta^2\right) f\left(\hat{\theta}_1^2 \middle| \theta_1\right) f(\delta^2, \theta_1) \mathrm{d}\delta^2 \mathrm{d}\theta_1, \tag{40}$$

where $f(\hat{\delta}^2 | \hat{\theta}_1^2, \delta^2)$ is the density function of $\hat{\delta}^2$ conditional on $\hat{\theta}_1^2$ and δ^2 in (32), $f(\hat{\theta}_1^2 | \theta_1)$ is the density function of $\hat{\theta}_1^2$ conditional on θ_1 in (34), and $f(\delta^2, \theta_1)$ is the joint prior distribution of δ^2 and θ_1. The posterior null probability can be computed as

$$p_1(\delta_b^2) = \frac{ML_0(\delta_b^2)}{ML_0(\delta_b^2) + ML_1(\delta_b^2)}, \tag{41}$$

which measures the support for the null hypothesis using both the prior knowledge on the parameters and the evidence from the data.[14]

For $H_0 : \delta^2 = 0$ vs. $H_1 : \delta^2 > 0$, the distribution of δ^2 in (36) is the prior distribution under the alternative, and the marginal likelihood under the alternative is given by,

$$ML_1(0) \equiv ML(H_1 : \delta^2 > 0) = \int \int f\left(\hat{\delta}^2 \middle| \hat{\theta}_1^2, \delta^2\right) f\left(\hat{\theta}_1^2 \middle| \theta_1\right) f(\delta^2, \theta_1) \mathrm{d}\delta^2 \mathrm{d}\theta_1. \tag{43}$$

To obtain the marginal likelihood under the null, we set $\delta^2 = 0$, thus the integration over δ^2 is no longer needed,

$$ML_0(0) \equiv ML(H_0 : \delta^2 = 0) = \int f\left(\hat{\delta}^2 \middle| \hat{\theta}_1^2, \delta^2 = 0\right) f\left(\hat{\theta}_1^2 \middle| \theta_1\right) f(\theta_1) \mathrm{d}\theta_1. \tag{44}$$

[14] The prior probability for the null hypothesis $H_0 : \delta^2 \leq \delta_b^2$ can be computed as

$$p_0(\delta_b^2) = \int \int_0^{\delta_b^2} f(\delta^2, \theta_1) \mathrm{d}\delta^2 \mathrm{d}\theta_1. \tag{42}$$

For a given multi-factor model, because δ_b^2 decreases with the length of the estimation window T, the prior null probability decreases with T. The relation between $p_0(\delta_b^2)$ and N is less straightforward: δ_b^2 increases with N, but at the same time, the prior distribution of δ^2 in (36) also depends on N. The results in Table 6 suggest that $p_0(\delta_b^2)$ decreases with N.

Because of the sharp null hypothesis (i.e., $\delta^2 = 0$), we also need to separately specify the prior null probability to compute the posterior null probability. We assume equal prior odds for the null and the alternative (i.e., the prior null probability is set to 0.5, $p_0(0) = 0.5$), then the posterior null probability can be computed as

$$p_1(0) = \frac{ML_0(0)}{ML_0(0) + ML_1(0)}. \tag{45}$$

Note that with the separately specified $p_0(0)$, the comparable posterior null probability for $H_0 : \delta^2 \leq \delta_b^2$ needs to be adjusted as follows[15]

$$\tilde{p}_1(\delta_b^2) = p_1(0) + (1 - p_1(0))\, p_1(\delta_b^2), \tag{47}$$

where $p_1(\delta_b^2)$ is from (41). Thus, when the estimation risk is taken into account, the probability for the null hypothesis increases.

Table 6 reports the test results based on the Bayesian approach. Panels A, B, and C present the results for different combinations of a and b in (37). Similar to Table 5, we conduct the test using data over three different periods: 2008/4–2021/12 ($T = 165$), 1994/7–2021/12 ($T = 330$), and 1967/1–2021/12 ($T = 660$). The results of $p_1(0)$, $p_1(\delta_b^2)$, $p_0(\delta_b^2)$, and $\tilde{p}_1(\delta_b^2)$ in (45), (41), (42), and (47) are shown in the table. For brevity, we omit $p_0(0) = 0.5$ from the table.

Table 6 about here

As expected, the results in Table 6 suggest that the probability for the null hypothesis increases when the estimation risk is taken into account, and the impact is larger for the shorter estimation window. For example, for $a = 3$ and $b = 3$ (Panel A) and the 2008/4–2021/12 period, without considering the estimation risk, the probability for the null $H_0 : \delta^2 = 0$ varies between 1.6% and 60.6% across the multi-factor models when both the prior information and the data are taken into account (i.e., $p_1(0)$). When the estimation risk is considered (i.e., $H_0 : \delta^2 \leq \delta_b^2$), all the null probabilities (i.e., $\tilde{p}_1(\delta_b^2)$) increase, and the magnitude of the increase (i.e., $\tilde{p}_1(\delta_b^2) - p_1(0)$) ranges from 9.8% for HMXZ q^5 to 36.7%

[15] The corresponding prior null probability is then

$$\tilde{p}_0(\delta_b^2) = p_0(0) + (1 - p_0(0))\, p_0(\delta_b^2), \tag{46}$$

where $p_0(\delta_b^2)$ is from (42).

for BAB. The magnitude is smaller for the longer estimation window. For the 1994/7–2021/12 period, the increase in the null probability ranges from 0% to 26.2%; and for the 1967/1–2021/12 period, the range is 0% to 16.6%.

In addition, Table 6 shows that whether the estimation risk is considered or not, the role of data can change for a given model. For example, to test FF-3 against the CAPM using data over 2008/4–2021/12 in Panel A, without considering the estimation risk, the posterior null probability is lower than the prior null probability (i.e., $p_1(0) = 0.496 < p_0(0) = 0.50$), suggesting that the data does not provide further support for the null. However, note that $p_1(\delta_b^2) = 0.715 > p_0(\delta_b^2) = 0.650$, and therefore, $\tilde{p}_1(\delta_b^2) = 0.856 > \tilde{p}_0(\delta_b^2) = 0.5 + 0.5 \times 0.650 = 0.825$. That is, when the estimation risk is considered, the data now provide further support for the null hypothesis.

Finally, comparing across the three panels in Table 6, we find that the results are not very sensitive to different values of a and b.

5. Conclusion

Academic asset pricing models have produced increasingly large sample Sharpe ratios over time. Starting with the value-weighted market portfolio of the CAPM, which produced a sample Sharpe ratio of 0.133, there are now multi-factor model that produces a sample Sharpe ratio of 0.599, more than four times larger than that of the market portfolio. Such a good performance is not seen in the investing world, and the high sample Sharpe ratios of the popular multi-factor asset pricing models are also at odd with a long-standing belief in finance that high Sharpe ratios are good deals and are unlikely to survive. For example, Ross (1976) assumes that no portfolio can have Sharpe ratio that is twice as large as that of the market portfolio. MacKinlay (1995) thinks that Fama-French 3-factor model has unreasonably high sample Sharpe ratio, even after taking into account of sampling variability. Cochrane and Saá-Requejo (2000) believe that no asset should have a Sharpe ratio that is twice that of the S&P 500 (which they assume to have an annual value of 0.5, or a monthly value of 0.1443) and use this assumption to derive bounds on option prices.

While there are a number of possible reasons why the recent asset pricing models produce high sample Sharpe ratios that are unattainable for real-world investors, we focus on one possible explanation: estimation risk. In practice, because the mean and covariance matrix of the factors are unknown, investors are not able to construct and hold the true optimal portfolio of the multi-factor models. Instead, they need to estimate the optimal weights, which will lead to deteriorated out-of-sample performance.

We show empirically that the out-of-sample Sharpe ratios of the multi-factor models are significantly lower than their in-sample Sharpe ratios. Even after using a bootstrap simulation to control for potential time-varying model performance, the performance deterioration of the out-of-sample Sharpe ratio remains substantial. Estimation risk is the likely reason for this remaining gap between the in-sample and the out-of-sample Sharpe ratios.

We theoretically analyze the effect of estimation risk by obtaining the finite sample distributions of the in-sample and the out-of-sample Sharpe ratios of a multi-factor asset pricing model. The results show that the out-of-sample Sharpe ratio is always lower than the population Sharpe ratio, whereas the in-sample Sharpe ratio is an upward biased estimator of the population Sharpe ratio. The results also show that the performance deterioration of the out-of-sample Sharpe ratio is negatively correlated with the observed in-sample Sharpe ratio, suggesting that the in-sample Sharpe ratio is not a reliable indicator of what investors can obtain out-of-sample.

Given that the out-of-sample Sharpe ratio is more relevant for investors and it is subject to estimation risk, the effect of estimation risk needs to be considered when comparing asset pricing models. We recommend the use of a break-even Sharpe ratio to incorporate the effect of estimation risk, and illustrate how to include the break-even Sharpe ratio in the test to compare a multi-factor asset pricing model against the CAPM. In both the classical frequentist framework and the Bayesian framework, we show that the use of the break-even Sharpe ratio makes the null hypothesis that a multi-factor model is as good as the CAPM less likely to be rejected.

One of the limitations of our theoretical analysis is that it is based on the i.i.d. multivariate normality assumption for the returns of the traded factors. With fat-tailed distributions, it is conceivable that the problem of estimation risk is more severe than in the normality case.[16] So one should take our results as a lower bound on the impact of estimation risk on the out-of-sample performance of the multi-factor asset pricing models. In addition, if there is a concern that parameters in these models are not constant over time, then there is an additional source of risk that hampers the out-of-sample performance of the sample optimal

[16] In a recent study, Kan and Lassance (2024) examine optimal portfolio choice when the return distribution exhibits fat tails.

portfolio based on a multi-factor model.

Appendix: Proofs

Proof of Proposition 1: Under the multivariate normality assumption, it is well known that $\hat{\mu}$ and $\hat{\Sigma}$ are independent of each other and have the following distributions:

$$\hat{\mu} \sim \mathcal{N}(\mu, \Sigma/T), \tag{A1}$$
$$\hat{\Sigma} \sim \mathcal{W}_N(T-1, \Sigma/T), \tag{A2}$$

where $\mathcal{W}_N(T-1, \Sigma/T)$ is a Wishart distribution with $T-1$ degrees of freedom and covariance matrix Σ/T. Define $\eta = \Sigma^{-\frac{1}{2}}\mu/\theta$, we have $\eta'\eta = 1$. Let P be an $N \times N$ orthonormal matrix with its first column to be η. By defining

$$z = \sqrt{T}P'\Sigma^{-\frac{1}{2}}\hat{\mu} \sim \mathcal{N}\left(\begin{bmatrix} \sqrt{T}\theta \\ 0_{N-1} \end{bmatrix}, I_N\right), \tag{A3}$$
$$W = TP'\Sigma^{-\frac{1}{2}}\hat{\Sigma}\Sigma^{-\frac{1}{2}}P \sim \mathcal{W}_N(T-1, I_N), \tag{A4}$$

we can write

$$\hat{\theta} = (\hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu})^{\frac{1}{2}} = (z'W^{-1}z)^{\frac{1}{2}}, \tag{A5}$$
$$\tilde{\theta} = \frac{\mu'\hat{\Sigma}^{-1}\hat{\mu}}{(\hat{\mu}'\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\hat{\mu})^{\frac{1}{2}}} = \frac{\sqrt{T}\theta e_1'W^{-1}z}{(Tz'W^{-2}z)^{\frac{1}{2}}} = \frac{\theta e_1'W^{-1}z}{(z'W^{-2}z)^{\frac{1}{2}}}, \tag{A6}$$

where $e_1 = [1,\ 0_{N-1}']'$. Define an $N \times N$ orthonormal matrix $Q = [\tilde{z},\ Q_1]$ with its first column to be $\tilde{z} \equiv z/(z'z)^{\frac{1}{2}}$. Let

$$A = (Q'W^{-1}Q)^{-1} = \begin{bmatrix} \tilde{z}'W^{-1}\tilde{z} & \tilde{z}'W^{-1}Q_1 \\ Q_1'W^{-1}\tilde{z} & Q_1'W^{-1}Q_1 \end{bmatrix}^{-1} \equiv \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \sim \mathcal{W}_N(T-1, I_N), \tag{A7}$$

where A_{11} is the $(1,1)$ element of A. Theorem 3.2.10 of Muirhead (1982) suggests that

$$u_1 \equiv A_{11\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21} \sim \chi^2_{T-N}, \tag{A8}$$

and it is independent of A_{12} and A_{22}. In addition, using the result of Dickey (1967), we can show that

$$-A_{22}^{-1}A_{21} \sim \frac{x}{\sqrt{u_2}}, \tag{A9}$$

where $x \sim \mathcal{N}(0_{N-1}, I_{N-1})$, $u_2 \sim \chi^2_{T-N+1}$, and they are independent of each other and u_1. Since the distribution of A is independent of z, x, u_1 and u_2 are also independent of z. Using the formula for the inverse of a partitioned matrix, we can easily verify that

$$\tilde{z}'W^{-1}\tilde{z} = A^{-1}_{11\cdot 2} = \frac{1}{u_1}, \tag{A10}$$

$$Q_1'W^{-1}\tilde{z} = -A^{-1}_{22}A_{21}A^{-1}_{11\cdot 2} = \frac{x}{u_1\sqrt{u_2}}. \tag{A11}$$

With these identities, we can write

$$z'W^{-2}z = z'W^{-1}\left(\tilde{z}\tilde{z}' + Q_1Q_1'\right)W^{-1}z = (z'z)\left(\frac{1}{u_1^2} + \frac{x'x}{u_1^2u_2}\right). \tag{A12}$$

Let $z_1 \sim \mathcal{N}(\sqrt{T}\theta, 1)$ and $x_1 \sim \mathcal{N}(0,1)$ be the first element of z and x, respectively. We can write $z'z = z_1^2 + u$ and $x'x = x_1^2 + u_3$, where $u \sim \chi^2_{N-1}$ and $u_3 \sim \chi^2_{N-2}$. It follows that

$$z'W^{-2}z = \frac{(z_1^2 + u)}{u_1^2}\left(1 + \frac{x_1^2 + u_3}{u_2}\right). \tag{A13}$$

Without loss of generality, let the first column of Q_1 be

$$\xi = \frac{(I_N - \tilde{z}\tilde{z}')e_1}{[e_1'(I_N - \tilde{z}\tilde{z}')e_1]^{\frac{1}{2}}} = \frac{(I_N - \tilde{z}\tilde{z}')e_1}{\sqrt{1 - \frac{z_1^2}{z'z}}}. \tag{A14}$$

From (A11), we know that

$$\frac{x_1}{u_1\sqrt{u_2}} = \xi'W^{-1}\tilde{z} = \frac{e_1'W^{-1}\tilde{z} - \frac{e_1'\tilde{z}}{u_1}}{\sqrt{1 - \frac{z_1^2}{z'z}}} = \frac{e_1'W^{-1}z - \frac{z_1}{u_1}}{\sqrt{u}}, \tag{A15}$$

and hence

$$e_1'W^{-1}z = \frac{z_1}{u_1} + \frac{x_1\sqrt{u}}{u_1\sqrt{u_2}} = \frac{1}{u_1}\left(z_1 + \frac{x_1\sqrt{u}}{\sqrt{u_2}}\right). \tag{A16}$$

Define $q_1 = x_1/\sqrt{x'x}$ and $q_2 = z_2/\sqrt{u}$, where $z_2 \sim \mathcal{N}(0,1)$ is the second element of z. It is well known that q_1 is independent of $x'x$ and q_2 is independent of u [e.g., Theorem 1.5.6 of Muirhead (1982)]. Since x is independent of u, q_1 and q_2 are independent of both u and $x'x$. In addition, q_1 and q_2 have the same distribution, so we can replace q_1 with q_2 and write

$$x_1\sqrt{u} = \frac{x_1}{\sqrt{x'x}}\sqrt{x'x}\sqrt{u} \overset{d}{=} \frac{z_2}{\sqrt{u}}\sqrt{x'x}\sqrt{u} = z_2\sqrt{x'x}. \tag{A17}$$

In addition, let

$$b = \frac{u_2}{x'x + u_2} \sim \text{Beta}\left(\frac{T-N+1}{2}, \frac{N-1}{2}\right). \tag{A18}$$

We can write

$$e_1'W^{-1}z = \frac{1}{u_1}\left(z_1 + \frac{z_2\sqrt{x'x}}{\sqrt{u_2}}\right) = \frac{1}{u_1}\left(\frac{z_1\sqrt{b} + z_2\sqrt{1-b}}{\sqrt{b}}\right), \tag{A19}$$

$$z'W^{-2}z = \frac{z'z}{u_1^2 b} = \frac{z_1^2 + z_2^2 + u_0}{u_1^2 b}, \tag{A20}$$

where $u_0 \sim \chi^2_{N-2}$ and it is independent of z_1 and z_2. We have

$$\hat{\theta} = (z'W^{-1}z)^{\frac{1}{2}} = (z'z)^{\frac{1}{2}}(\tilde{z}'W^{-1}\tilde{z})^{\frac{1}{2}} \stackrel{d}{=} \frac{\sqrt{z_1^2 + z_2^2 + u_0}}{\sqrt{u_1}}, \tag{A21}$$

$$\tilde{\theta} = \frac{\theta e_1'W^{-1}z}{\sqrt{z'W^{-2}z}} = \frac{\theta(\sqrt{b}z_1 + \sqrt{1-b}z_2)}{(z'z)^{\frac{1}{2}}} \stackrel{d}{=} \frac{\theta(\sqrt{b}z_1 + \sqrt{1-b}z_2)}{\sqrt{z_1^2 + z_2^2 + u_0}}. \tag{A22}$$

Finally, let $\tilde{z} = \sqrt{b}z_1 + \sqrt{1-b}z_2 \sim \mathcal{N}(\sqrt{b}\sqrt{T}\theta, 1)$ and $\tilde{u} = z_1^2 + z_2^2 + u_0 - \tilde{z}^2 = z'z - \tilde{z}^2 \sim \chi^2_{N-1}((1-b)T\theta^2)$, and conditional on b, $\tilde{z}$ and $\tilde{u}$ are independent of each other. Therefore, we can write

$$\hat{\theta} \stackrel{d}{=} \frac{\sqrt{\tilde{z}^2 + \tilde{u}}}{\sqrt{u_1}}, \tag{A23}$$

$$\tilde{\theta} \stackrel{d}{=} \frac{\theta\tilde{z}}{\sqrt{\tilde{z}^2 + \tilde{u}}}. \tag{A24}$$

This completes the proof.

Proof of Lemma 1: We first cite some explicit expressions of moments of noncentral chi-squared and beta random variables. Suppose $X \sim \chi^2_\nu(\lambda)$ and $B \sim \text{Beta}(\nu_1, \nu_2)$. We have

$$\mathbb{E}[X^r] = \frac{2^r\Gamma\left(\frac{\nu}{2} + r\right)}{\Gamma\left(\frac{\nu}{2}\right)} {}_1F_1\left(-r; \frac{\nu}{2}; -\frac{\lambda}{2}\right) \quad \text{for } r > -\tfrac{\nu}{2}, \tag{A25}$$

$$\mathbb{E}[B^r] = \frac{\mathrm{B}(\nu_1 + r, \nu_2)}{\mathrm{B}(\nu_1, \nu_2)} \quad \text{for } r > -\nu_1, \tag{A26}$$

where $\mathrm{B}(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the beta function. (A25) is given in Krishnan (1967), and (A26) is obtained by direct integration. Using (A25) and the fact that $\tilde{z}^2 + \tilde{u} \sim \chi^2_N(T\theta^2)$ and it is independent of u_1, we can obtain $\mathbb{E}[\hat{\theta}]$ and $\mathbb{E}[\hat{\theta}^2]$ as

$$\begin{aligned} E[\hat{\theta}] &= E[(\tilde{z}^2 + \tilde{u})^{\frac{1}{2}}]E[u_1^{-\frac{1}{2}}] \\ &= \frac{\Gamma\left(\frac{N+1}{2}\right)\Gamma\left(\frac{T-N-1}{2}\right)}{\Gamma\left(\frac{N}{2}\right)\Gamma\left(\frac{T-N}{2}\right)} {}_1F_1\left(-\frac{1}{2}; \frac{N}{2}; -\frac{T\theta^2}{2}\right) \quad \text{for } T \geq N+2, \end{aligned} \tag{A27}$$

$$\begin{aligned}\mathbb{E}[\hat{\theta}^2] &= E[\tilde{z}^2+\tilde{u}]E[u_1^{-1}] \\ &= \frac{N+T\theta^2}{T-N-2} \quad \text{for } T \geq N+3, \end{aligned} \tag{A28}$$

For $\tilde{\theta}$, we use independence between b and (z_1, z_2, u_0) in (A22) and apply (A26) to obtain

$$\mathbb{E}[\tilde{\theta}] = \theta\mathbb{E}[b^{\frac{1}{2}}]\mathbb{E}\left[\frac{z_1}{\sqrt{z_1^2+z_2^2+u_0}}\right] = \theta\frac{\mathrm{B}\left(\frac{T-N+2}{2}, \frac{N-1}{2}\right)}{\mathrm{B}\left(\frac{T-N+1}{2}, \frac{N-1}{2}\right)}\mathbb{E}\left[\frac{z_1}{\sqrt{z_1^2+z_2^2+u_0}}\right]. \tag{A29}$$

By using the symmetry argument, the term $\sqrt{1-b}z_2$ drops out because $z_2 \sim \mathcal{N}(0,1)$. For the last expectation, we use a lemma in Kan, Wang, and Zhou (2022) to show that

$$\mathbb{E}\left[\frac{z_1}{\sqrt{z_1^2+z_2^2+u_0}}\right] = \sqrt{T}\theta\mathbb{E}\left[\frac{1}{\sqrt{y}}\right], \tag{A30}$$

where $y \sim \chi^2_{N+2}(T\theta^2)$. Then using (A25), we obtain

$$\mathbb{E}[\tilde{\theta}] = \theta\frac{\Gamma\left(\frac{T-N+2}{2}\right)\Gamma\left(\frac{T}{2}\right)}{\Gamma\left(\frac{T-N+1}{2}\right)\Gamma\left(\frac{T+1}{2}\right)}\frac{\sqrt{T}\theta\Gamma\left(\frac{N+1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{N+2}{2}\right)}{}_1F_1\left(\frac{1}{2}; \frac{N+2}{2}; -\frac{T\theta^2}{2}\right).$$

For $\mathbb{E}[\tilde{\theta}^2]$, we use (A22) and apply (A26) to obtan

$$\begin{aligned}\mathbb{E}[\tilde{\theta}^2] &= \theta^2\mathbb{E}\left[\frac{bz_1^2+(1-b)z_2^2+2\sqrt{b(1-b)}z_1z_2}{z_1^2+z_2^2+u_0}\right] \\ &= \theta^2\mathbb{E}\left[\frac{bz_1^2+(1-b)z_2^2}{z_1^2+z_2^2+u_0}\right] \\ &= \frac{\theta^2}{T}\mathbb{E}\left[\frac{(T-N+1)z_1^2+(N-1)z_2^2}{z_1^2+z_2^2+u_0}\right]. \end{aligned} \tag{A31}$$

Note that the term $2\sqrt{b(1-b)}z_1z_2/(z_1^2+z_2^2+u_0)$ vanishes in the above expectation because of symmetry. The last expectation term can be written as $\mathbb{E}[(z'Az)/(z'z)]$, where $A = \mathrm{Diag}(T-N+1,\ N-1,\ 0'_{N-2})$. Using Theorem 4 of Hillier, Kan, and Wang (2014), we obtain the expectation of the ratio of quadratic form in z as

$$\begin{aligned}\mathbb{E}\left[\frac{z'Az}{z'z}\right] &= \frac{T}{N}{}_1F_1\left(1; \frac{N+2}{2}; -\frac{T\theta^2}{2}\right) + \frac{T\theta^2(T-N+1)}{N+2}{}_1F_1\left(1; \frac{N+4}{2}; -\frac{T\theta^2}{2}\right) \\ &= T-N+1-\frac{(N-1)(T-N)}{N}{}_1F_1\left(1; \frac{N+2}{2}; -\frac{T\theta^2}{2}\right), \end{aligned} \tag{A32}$$

where the last equality follows from a recurrence relation of confluent hypergeometric function.[17] It follows that

$$\mathbb{E}[\tilde{\theta}^2] = \theta^2 \left[\frac{T-N+1}{T} - \frac{(N-1)(T-N)}{NT} {}_1F_1\left(1; \frac{N+2}{2}; -\frac{T\theta^2}{2}\right)\right] \quad \text{for } T \geq N+1. \tag{A34}$$

Finally, using (9) and (10), $\mathbb{E}[\hat{\theta}\tilde{\theta}]$ is given by

$$\begin{aligned}
\mathbb{E}[\hat{\theta}\tilde{\theta}] &= \theta\mathbb{E}\left[\frac{\tilde{z}}{\sqrt{u_1}}\right] \\
&= \theta\mathbb{E}[b^{\frac{1}{2}}]\sqrt{T}\theta\mathbb{E}[u_1^{-\frac{1}{2}}] \\
&= \theta\frac{\mathrm{B}\left(\frac{T-N+2}{2}, \frac{N-1}{2}\right)}{\mathrm{B}\left(\frac{T-N+1}{2}, \frac{N-1}{2}\right)}\sqrt{T}\theta\frac{\Gamma\left(\frac{T-N-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{T-N}{2}\right)} \\
&= \frac{\theta^2\sqrt{T}(T-N)\Gamma\left(\frac{T}{2}\right)}{\sqrt{2}(T-N-1)\Gamma\left(\frac{T+1}{2}\right)} \quad \text{for } T \geq N+2.
\end{aligned} \tag{A35}$$

This completes the proof.

Proof of Lemma 2: To prove $\mathbb{E}[\tilde{\theta} - \hat{\theta}] < 0$, we show that $\mathbb{E}[\tilde{\theta}] < \theta$ and $\mathbb{E}[\hat{\theta}] > \theta$. From (10), it is easy to see that $\tilde{\theta} < \theta$ because $|\tilde{z}/\sqrt{\tilde{z}^2 + \tilde{u}}| < 1$. Therefore, $\mathbb{E}[\tilde{\theta}] < \theta$. To prove $\mathbb{E}[\hat{\theta}] > \theta$, note that

$$\hat{\theta} = \frac{\hat{w}'\hat{\mu}}{(\hat{w}'\hat{\Sigma}\hat{w})^{\frac{1}{2}}} \geq \frac{w^{*\prime}\hat{\mu}}{(w^{*\prime}\hat{\Sigma}w^*)^{\frac{1}{2}}}, \tag{A36}$$

with the equality holds if and only if $\hat{w}$ is proportional to w^*. Since this event has probability zero, we can write the above with a strict inequality. Taking expectation on both sides and using the fact that $\hat{\mu}$ is independent of $\hat{\Sigma}$, we have

$$\mathbb{E}[\hat{\theta}] > \mathbb{E}\left[\frac{w^{*\prime}\hat{\mu}}{(w^{*\prime}\hat{\Sigma}w^*)^{\frac{1}{2}}}\right] = w^{*\prime}\mu\mathbb{E}\left[\frac{1}{(w^{*\prime}\hat{\Sigma}w^*)^{\frac{1}{2}}}\right] = \theta\mathbb{E}\left[\left(\frac{w^{*\prime}\Sigma w^*}{w^{*\prime}\hat{\Sigma}w^*}\right)^{\frac{1}{2}}\right]. \tag{A37}$$

Using 3.2.5 of Muirhead (1982), we know $y \equiv Tw^{*\prime}\hat{\Sigma}w^*/(w^{*\prime}\Sigma w^*) \sim \chi^2_{T-1}$, so we have

$$\mathbb{E}\left[\left(\frac{w^{*\prime}\Sigma w^*}{w^{*\prime}\hat{\Sigma}w^*}\right)^{\frac{1}{2}}\right] = T^{\frac{1}{2}}\mathbb{E}\left[y^{-\frac{1}{2}}\right] \geq \frac{T^{\frac{1}{2}}}{(E[y])^{\frac{1}{2}}} = \frac{T^{\frac{1}{2}}}{(T-1)^{\frac{1}{2}}} > 1, \tag{A38}$$

[17] The recurrence relation is

$$b_1F_1(a;b;z) - b_1F_1(a-1;b;z) = z_1F_1(a;b+1;z). \tag{A33}$$

The equality is obtained by setting $a = 1$, $b = (N+2)/2$ and $z = -T\theta^2/2$.

where the inequality follows because of Jensen's inequality. Therefore, we have $\mathbb{E}[\hat{\theta}] > \theta$.

Next, we prove $\text{Var}[\hat{\theta}] > \text{Cov}[\tilde{\theta}, \hat{\theta}]$. In our proof, we assume $\text{Cov}[\tilde{\theta}, \hat{\theta}] > 0$.[18] Denote $\lambda = T\theta^2$. Using Lemma 1 and applying a transformation of the confluent hypergeometric function,

$$e^z {}_1F_1(a; b; -z) = {}_1F_1(b-a; b; z), \tag{A39}$$

we can write

$$e^{\lambda}\text{Var}[\hat{\theta}] = \left(\frac{N+\lambda}{T-N-2}\right) e^{\lambda} - c_1 \frac{\Gamma\left(\frac{N+1}{2}\right)^2}{\Gamma\left(\frac{N}{2}\right)^2} {}_1F_1\left(\frac{N+1}{2}; \frac{N}{2}; \frac{\lambda}{2}\right)^2, \tag{A40}$$

$$e^{\lambda}\text{Cov}[\tilde{\theta}, \hat{\theta}] = \lambda c_2 \left[e^{\lambda} - c_3 \times {}_1F_1\left(\frac{N+1}{2}; \frac{N+2}{2}; \frac{\lambda}{2}\right) {}_1F_1\left(\frac{N+1}{2}; \frac{N}{2}; \frac{\lambda}{2}\right)\right], \tag{A41}$$

where

$$c_1 = \frac{\Gamma\left(\frac{T-N-1}{2}\right)^2}{\Gamma\left(\frac{T-N}{2}\right)^2}, \tag{A42}$$

$$c_2 = \frac{(T-N)\Gamma\left(\frac{T}{2}\right)}{\sqrt{2T}(T-N-1)\Gamma\left(\frac{T+1}{2}\right)}, \tag{A43}$$

$$c_3 = \frac{\Gamma\left(\frac{N+1}{2}\right)^2}{\Gamma\left(\frac{N}{2}\right)\Gamma\left(\frac{N+2}{2}\right)}. \tag{A44}$$

Using the inequality (see Kazarinoff, 1956)

$$x - \frac{1}{2} < x - \frac{1}{4} < \frac{\Gamma\left(x+\frac{1}{2}\right)^2}{\Gamma(x)^2} < x \tag{A45}$$

for $x > 0$, we can show that

$$c_1 < \frac{1}{\frac{T-N-1}{2} - \frac{1}{2}} = \frac{2}{T-N-2}, \tag{A46}$$

$$c_2 < \frac{(T-N)}{(T-N-1)\sqrt{2T}}\sqrt{\frac{1}{\frac{T}{2}-\frac{1}{4}}} = \frac{T-N}{(T-N-1)\sqrt{T(T-\frac{1}{2})}} < \frac{1}{T-N-2}, \tag{A47}$$

where the last inequality holds because $\sqrt{T(T-\frac{1}{2})} > T-N$ and $1/(T-N-1) < 1/(T-N-2)$. Then we have

$$e^{\lambda}\text{Var}[\hat{\theta}] > \left(\frac{N+\lambda}{T-N-2}\right) e^{\lambda} - \frac{2}{T-N-2} \frac{\Gamma\left(\frac{N+1}{2}\right)^2}{\Gamma\left(\frac{N}{2}\right)^2} {}_1F_1\left(\frac{N+1}{2}; \frac{N}{2}; \frac{\lambda}{2}\right)^2$$

[18]When $\text{Cov}[\tilde{\theta}, \hat{\theta}] < 0$, it is trivial that $\text{Var}[\hat{\theta}] > \text{Cov}[\tilde{\theta}, \hat{\theta}]$. We can prove that $\text{Cov}[\tilde{\theta}, \hat{\theta}] > 0$. For brevity, we skip the proof, but it is available upon request.

$$= \frac{1}{T-N-2} \sum_{k=0}^{\infty} \frac{\tilde{a}_k}{k!} \lambda^k, \tag{A48}$$

$$e^{\lambda} \mathrm{Cov}[\tilde{\theta}, \hat{\theta}] < \frac{\lambda}{T-N-2} \left[e^{\lambda} - c_3 \times {}_1F_1\left(\frac{N+1}{2}; \frac{N+2}{2}; \frac{\lambda}{2}\right) {}_1F_1\left(\frac{N+1}{2}; \frac{N}{2}; \frac{\lambda}{2}\right)\right]$$

$$= \frac{\lambda}{T-N-2} \sum_{k=0}^{\infty} \frac{\tilde{b}_k}{k!} \lambda^k = \frac{1}{T-N-2} \sum_{k=1}^{\infty} \frac{k\tilde{b}_{k-1}}{k!} \lambda^k, \tag{A49}$$

with

$$\tilde{a}_k = (N+k) - \frac{1}{2^{k-1}} \sum_{i=0}^{k} \binom{k}{i} \frac{\Gamma\left(\frac{N+1}{2}+i\right)\Gamma\left(\frac{N+1}{2}+k-i\right)}{\Gamma\left(\frac{N}{2}+i\right)\Gamma\left(\frac{N}{2}+k-i\right)}, \tag{A50}$$

$$\tilde{b}_k = 1 - \frac{1}{2^k} \sum_{i=0}^{k} \binom{k}{i} \frac{\Gamma\left(\frac{N+1}{2}+i\right)\Gamma\left(\frac{N+1}{2}+k-i\right)}{\Gamma\left(\frac{N}{2}+i\right)\Gamma\left(\frac{N+2}{2}+k-i\right)}. \tag{A51}$$

If $\tilde{a}_0 > 0$ and $\tilde{a}_k > k\tilde{b}_{k-1}$ for $k \geq 1$, then $\mathrm{Var}[\hat{\theta}] > \mathrm{Cov}[\tilde{\theta}, \hat{\theta}]$. Using (A45), we can show

$$\tilde{a}_0 = N - \frac{2\Gamma\left(\frac{N+1}{2}\right)^2}{\Gamma\left(\frac{N}{2}\right)^2} > N - 2 \times \frac{N}{2} = 0. \tag{A52}$$

Let $g_N(k) = \tilde{a}_k - k\tilde{b}_{k-1}$ for $k \geq 1$, and we will show

$$g_N(k) = N - \frac{1}{2^{k-1}} \left[\sum_{i=0}^{k} \binom{k}{i} \frac{\Gamma\left(\frac{N+1}{2}+i\right)\Gamma\left(\frac{N+1}{2}+k-i\right)}{\Gamma\left(\frac{N}{2}+i\right)\Gamma\left(\frac{N}{2}+k-i\right)} \right.$$

$$\left. - k \sum_{i=0}^{k-1} \binom{k-1}{i} \frac{\Gamma\left(\frac{N+1}{2}+i\right)\Gamma\left(\frac{N+1}{2}+k-1-i\right)}{\Gamma\left(\frac{N}{2}+i\right)\Gamma\left(\frac{N}{2}+k-i\right)} \right]$$

$$= N - \frac{(N-1)}{2^k} \sum_{i=0}^{k} \binom{k}{i} \frac{\Gamma\left(\frac{N+1}{2}+i\right)\Gamma\left(\frac{N-1}{2}+k-i\right)}{\Gamma\left(\frac{N}{2}+i\right)\Gamma\left(\frac{N}{2}+k-i\right)} > 0. \tag{A53}$$

Note that

$$g_N(1) = N - \left(\frac{N-1}{2}\right) \frac{\Gamma\left(\frac{N+1}{2}\right)\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N}{2}\right)\Gamma\left(\frac{N+2}{2}\right)} \times N = N\left[1 - \frac{\Gamma\left(\frac{N+1}{2}\right)^2}{\Gamma\left(\frac{N}{2}\right)\Gamma\left(\frac{N+2}{2}\right)}\right] > 0 \tag{A54}$$

due to log-convexity of $\Gamma(x)$. Using

$$\binom{k+1}{i} = \binom{k}{i} + \binom{k}{i-1}, \tag{A55}$$

we can write

$$g_N(k+1) = N - \frac{(N-1)}{2^{k+1}} \left[\sum_{i=0}^{k} \binom{k}{i} \frac{\Gamma\left(\frac{N+1}{2}+i\right)\Gamma\left(\frac{N-1}{2}+k+1-i\right)}{\Gamma\left(\frac{N}{2}+i\right)\Gamma\left(\frac{N}{2}+k+1-i\right)} \right.$$

$$+\sum_{i=0}^{k}\binom{k}{i}\frac{\Gamma\left(\frac{N+1}{2}+i+1\right)\Gamma\left(\frac{N-1}{2}+k-i\right)}{\Gamma\left(\frac{N}{2}+i+1\right)\Gamma\left(\frac{N}{2}+k-i\right)}\Bigg]. \tag{A56}$$

Then for $k \geq 1$,

$$\begin{aligned}
g_N(k+1)-g_N(k) &= \frac{(N-1)}{2^k}\sum_{i=0}^{k}\binom{k}{i}\Bigg[\frac{\Gamma\left(\frac{N+1}{2}+i\right)\Gamma\left(\frac{N-1}{2}+k-i\right)}{\Gamma\left(\frac{N}{2}+i\right)\Gamma\left(\frac{N}{2}+k-i\right)} \\
&\quad -\frac{\Gamma\left(\frac{N+1}{2}+i+1\right)\Gamma\left(\frac{N-1}{2}+k-i\right)}{2\Gamma\left(\frac{N}{2}+i+1\right)\Gamma\left(\frac{N}{2}+k-i\right)}-\frac{\Gamma\left(\frac{N+1}{2}+i\right)\Gamma\left(\frac{N-1}{2}+k+1-i\right)}{2\Gamma\left(\frac{N}{2}+i\right)\Gamma\left(\frac{N}{2}+k+1-i\right)}\Bigg] \\
&= \frac{(N-1)}{2^k}\sum_{i=0}^{k}\binom{k}{i}\frac{\Gamma\left(\frac{N+1}{2}+i\right)\Gamma\left(\frac{N-1}{2}+k-i\right)(2i-k)}{4\Gamma\left(\frac{N+2}{2}+i\right)\Gamma\left(\frac{N+2}{2}+k-i\right)} \\
&= \frac{(N-1)}{2^{k+2}}\sum_{i=0}^{k}\binom{k}{i}\frac{\Gamma\left(\frac{N-1}{2}+i\right)\Gamma\left(\frac{N-1}{2}+k-i\right)}{\Gamma\left(\frac{N+2}{2}+i\right)\Gamma\left(\frac{N+2}{2}+k-i\right)} \\
&\quad \times\frac{1}{2}\left[\left(\frac{N-1}{2}+i\right)(2i-k)+\left(\frac{N-1}{2}+k-i\right)(k-2i)\right] \\
&= \frac{(N-1)}{2^{k+3}}\sum_{i=0}^{k}\binom{k}{i}\frac{\Gamma\left(\frac{N-1}{2}+i\right)\Gamma\left(\frac{N-1}{2}+k-i\right)(k-2i)^2}{\Gamma\left(\frac{N+2}{2}+i\right)\Gamma\left(\frac{N+2}{2}+k-i\right)} > 0,
\end{aligned} \tag{A57}$$

where the second last equality is obtained by summing the series forward and backward and taking the average of the two sums. This completes the proof.

Proof of Proposition 2: From Proposition 1, we have

$$\hat{\theta} < c_1 \Rightarrow -\sqrt{c_1^2 u_1 - \tilde{u}} < \tilde{z} < \sqrt{c_1^2 u_1 - \tilde{u}}, \tag{A58}$$

$$\tilde{\theta} < c_2 \Rightarrow \tilde{z} < \frac{c_2\sqrt{\tilde{u}}}{\sqrt{\theta^2 - c_2^2}}. \tag{A59}$$

When $c_2 \leq 0$, the range of $\tilde{z}$ is $-\sqrt{c_1^2 u_1 - \tilde{u}} < \tilde{z} < \frac{c_2\sqrt{\tilde{u}}}{\sqrt{\theta^2-c_2^2}}$, when

$$\sqrt{c_1^2 u_1 - \tilde{u}} > \left|\frac{c_2\sqrt{\tilde{u}}}{\sqrt{\theta^2 - c_2^2}}\right| \Rightarrow 0 < \tilde{u} < \frac{c_1^2 u_1(\theta^2 - c_2^2)}{\theta^2}. \tag{A60}$$

Therefore,

$$\begin{aligned}
&\mathbb{P}[\hat{\theta} < c_1, \tilde{\theta} < c_2] \\
&= \int_0^\infty \int_0^1 \int_0^{\frac{c_1^2(\theta^2-c_2^2)v}{\theta^2}} \left[\Phi\left(\frac{c_2\sqrt{u}}{\sqrt{\theta^2-c_2^2}} - \sqrt{T}\theta\sqrt{b}\right) - \Phi\left(-\sqrt{c_1^2 v - u} - \sqrt{T}\theta\sqrt{b}\right)\right]
\end{aligned}$$

$$\times f_{\tilde{u}}(u) f_b(b) f_{u_1}(v) \mathrm{d}u \mathrm{d}b \mathrm{d}v. \tag{A61}$$

When $c_2 > 0$, the range of $\tilde{z}$ is

$$-\sqrt{c_1^2 u_1 - \tilde{u}} < \tilde{z} < \frac{c_2 \sqrt{\tilde{u}}}{\sqrt{\theta^2 - c_2^2}} \quad \text{if } 0 < \tilde{u} < \frac{c_1^2 u_1 (\theta^2 - c_2^2)}{\theta^2}, \tag{A62}$$

$$-\sqrt{c_1^2 u_1 - \tilde{u}} < \tilde{z} < \sqrt{c_1^2 u_1 - \tilde{u}} \quad \text{if } \frac{c_1^2 u_1 (\theta^2 - c_2^2)}{\theta^2} < \tilde{u} < c_1^2 u_1. \tag{A63}$$

Therefore,

$$\begin{aligned} &\mathbb{P}[\hat{\theta} < c_1, \tilde{\theta} < c_2] \\ &= \int_0^\infty \int_0^1 \int_0^{c_1^2 v} \left[\Phi\left(\min\left[\frac{c_2 \sqrt{u}}{\sqrt{\theta^2 - c_2^2}}, \sqrt{c_1^2 v - u} \right] - \sqrt{T}\theta\sqrt{b} \right) - \Phi\left(-\sqrt{c_1^2 v - u} - \sqrt{T}\theta\sqrt{b} \right) \right] \\ &\times f_{\tilde{u}}(u) f_b(b) f_{u_1}(v) \mathrm{d}u \mathrm{d}b \mathrm{d}v. \end{aligned} \tag{A64}$$

Taking derivative of $\mathbb{P}[\hat{\theta} < c_1, \tilde{\theta} < c_2]$ with respect to c_1 and c_2 and using the Leibniz integral rule, we obtain the joint density of $(\hat{\theta}, \tilde{\theta})$ for both cases as

$$f_{\hat{\theta},\tilde{\theta}}(c_1, c_2) = \int_0^\infty \int_0^1 f_{\tilde{u}}\left(\frac{c_1^2(\theta^2 - c_2^2) v}{\theta^2} \right) \phi\left(\frac{c_1 c_2 \sqrt{v}}{\theta} - \sqrt{T}\theta\sqrt{b} \right) \frac{2 c_1^2 v^{\frac{3}{2}}}{\theta} f_b(b) f_{u_1}(v) \mathrm{d}b \mathrm{d}v. \tag{A65}$$

This completes the proof.

Proof of Proposition 3: Using the representation of $\hat{\theta}$ in (A21) and defining w_1 and w_2 as

$$z_1 - \sqrt{T}\theta = w_1 \sim \mathcal{N}(0, 1), \tag{A66}$$

$$\frac{T - u_1}{\sqrt{2T}} \xrightarrow{d} w_2 \sim \mathcal{N}(0, 1), \tag{A67}$$

where the limiting distribution of $(T - u_1)/\sqrt{2T}$ is obtained by using the central limit theorem, we can write

$$\begin{aligned} \sqrt{T}(\hat{\theta} - \theta) &= \sqrt{T} \left(\frac{\sqrt{z_1^2 + z_2^2 + u_0}}{\sqrt{u_1}} - \theta \right) \\ &= \sqrt{T} \left(\frac{1}{\sqrt{u_1}} - \frac{1}{\sqrt{T}} \right) \sqrt{z_1^2 + z_2^2 + u_0} + \left(\sqrt{z_1^2 + z_2^2 + u_0} - \sqrt{T}\theta \right) \\ &= \frac{\sqrt{T}(T - u_1)}{\sqrt{T}\sqrt{u_1}(\sqrt{T} + \sqrt{u_1})} \sqrt{z_1^2 + z_2^2 + u_0} + \frac{z_1^2 + z_2^2 + u_0 - T\theta^2}{\sqrt{z_1^2 + z_2^2 + u_0} + \sqrt{T}\theta} \end{aligned}$$

$$
\begin{aligned}
&= \frac{\sqrt{2}(T-u_1)}{\sqrt{2T}}\left[\frac{1}{\frac{\sqrt{u_1}}{\sqrt{T}}\left(1+\frac{\sqrt{u_1}}{\sqrt{T}}\right)}\right]\left(\frac{z_1^2+z_2^2+u_0}{T}\right)^{\frac{1}{2}} + \frac{2\theta w_1 + \frac{w_1^2+z_2^2+u_0}{\sqrt{T}}}{\left(\frac{z_1^2+z_2^2+u_0}{T}\right)^{\frac{1}{2}}+\theta} \\
&\xrightarrow{d} \frac{\theta w_2}{\sqrt{2}} + w_1 \equiv X \sim \mathcal{N}\left(0, 1+\frac{\theta^2}{2}\right). \qquad \text{(A68)}
\end{aligned}
$$

The last equality is obtained by using the fact that $u_1/T \xrightarrow{p} 1$, $(z_1^2+z_2^2+u_0)/T \xrightarrow{p} \theta^2$, and $(w_1^2+z_2^2+u_0)/\sqrt{T} \xrightarrow{p} 0$.

Using the representation of $\tilde{\theta}$ in (A22) and defining v as

$$
T(1-b) \xrightarrow{d} v \sim \chi^2_{N-1}, \qquad \text{(A69)}
$$

we can write

$$
\begin{aligned}
T(\tilde{\theta}-\theta) &= T\theta\left(\frac{\sqrt{b}z_1+\sqrt{1-b}z_2-\sqrt{z_1^2+z_2^2+u_0}}{\sqrt{z_1^2+z_2^2+u_0}}\right) \\
&= -T\theta\left[\frac{(1-b)z_1^2+bz_2^2+u_0-2\sqrt{b(1-b)}z_1z_2}{\sqrt{z_1^2+z_2^2+u_0}(\sqrt{b}z_1+\sqrt{1-b}z_2+\sqrt{z_1^2+z_2^2+u_0})}\right] \\
&= -\theta\left[\frac{(1-b)(T\theta^2+2\sqrt{T}\theta w_1+w_1^2)+bz_2^2+u_0-2\sqrt{b(1-b)}(\sqrt{T}\theta+w_1)z_2}{\left(\frac{z_1^2+z_2^2+u_0}{T}\right)^{\frac{1}{2}}\left[\frac{\sqrt{b}z_1}{\sqrt{T}}+\frac{\sqrt{1-b}z_2}{\sqrt{T}}+\left(\frac{z_1^2+z_2^2+u_0}{T}\right)^{\frac{1}{2}}\right]}\right] \\
&\xrightarrow{d} -\theta\left[\frac{\theta^2 v+z_2^2+u_0-2\theta\sqrt{v}z_2}{\theta(\theta+\theta)}\right] \\
&= -\frac{\theta^2 v+z_2^2+u_0-2\theta\sqrt{v}z_2}{2\theta} \equiv Y. \qquad \text{(A70)}
\end{aligned}
$$

The second last equality follows because $b \xrightarrow{p} 1$, $(1-b)\sqrt{T}\theta w_1 \xrightarrow{p} 0$, $(1-b)w_1^2 \xrightarrow{p} 0$, $\sqrt{b(1-b)}w_1z_2 \xrightarrow{p} 0$, $(z_1^2+z_2^2+u_0)/T \xrightarrow{p} \theta^2$, $\sqrt{b}z_1/\sqrt{T} \xrightarrow{p} \theta$, $\sqrt{1-b}z_2/\sqrt{T} \xrightarrow{p} 0$. It remains to show that $Y \sim -(1+\theta^2)/(2\theta)\chi^2_{N-1}$. In order to show that, we let

$$
W = \begin{bmatrix} \sqrt{v} & 0 \\ z_2 & \sqrt{u_0} \end{bmatrix}\begin{bmatrix} \sqrt{v} & z_2 \\ 0 & \sqrt{u_0} \end{bmatrix}. \qquad \text{(A71)}
$$

From the Bartlett decomposition of Wishart distribution, we know $W \sim \mathcal{W}_2(N-1, I_2)$. Then using 3.2.8 of Muirhead (1982), we have

$$
Y = -\frac{\theta^2 v+z_2^2+u_0-2\theta\sqrt{v}z_2}{2\theta} = -\frac{[\theta,\ -1]W[\theta,\ -1]'}{2\theta} \sim -\frac{(1+\theta^2)\chi^2_{N-1}}{2\theta}. \qquad \text{(A72)}
$$

Finally, X is independent of Y because X is a function z_1 and u_1, and Y is a function of z_2, b, u_0, and (z_1, u_1) are independent of (z_2, b, u_0) from the proof of Proposition 1. This completes the proof.

Proof of Proposition 4: Based on the definition of random variables in Proposition 1, we let

$$\tilde{z}_1 = \frac{z_1}{\sqrt{T}}, \tag{A73}$$

$$\tilde{z}_2 = \frac{z_2}{\sqrt{T}}, \tag{A74}$$

$$w_1 = \frac{u_0}{T}, \tag{A75}$$

$$w_2 = \frac{u_1}{T}. \tag{A76}$$

Using the central limit theorem, we can easily show that when $N \to \infty$, $T \to \infty$, and $N/T \to \rho$, we have

$$\sqrt{T}(\tilde{z}_1 - \theta) \sim \mathcal{N}(0, 1), \tag{A77}$$

$$\sqrt{T}\tilde{z}_2 \sim \mathcal{N}(0, 1), \tag{A78}$$

$$\sqrt{T}(w_1 - \rho) \xrightarrow{d} \mathcal{N}(0, 2\rho), \tag{A79}$$

$$\sqrt{T}(w_2 - (1 - \rho)) \xrightarrow{d} \mathcal{N}(0, 2(1 - \rho)), \tag{A80}$$

$$\sqrt{T}(b - (1 - \rho)) \xrightarrow{d} \mathcal{N}(0, 2\rho(1 - \rho)), \tag{A81}$$

and these five random variables are independent of each other. From (A21) and (A22), we can write $\hat{\theta}$ and $\tilde{\theta}$ as

$$\hat{\theta} = \frac{(\tilde{z}_1^2 + \tilde{z}_2^2 + w_1)^{\frac{1}{2}}}{w_2^{\frac{1}{2}}}, \tag{A82}$$

$$\tilde{\theta} = \frac{\theta(\sqrt{b}\tilde{z}_1 + \sqrt{1 - b}\tilde{z}_2)}{(\tilde{z}_1^2 + \tilde{z}_2^2 + w_1)^{\frac{1}{2}}}, \tag{A83}$$

and both of them are functions of $(\tilde{z}_1, \tilde{z}_2, w_1, w_2, b)$. Then using the delta method and upon simplification, we obtain

$$\sqrt{T}\left(\begin{bmatrix} \hat{\theta} \\ \tilde{\theta} \end{bmatrix} - \begin{bmatrix} \frac{\sqrt{\theta^2+\rho}}{\sqrt{1-\rho}} \\ \frac{\sqrt{1-\rho}\theta^2}{\sqrt{\theta^2+\rho}} \end{bmatrix}\right) \xrightarrow{d} N\left(0_2, \begin{bmatrix} \frac{\theta^4+2\theta^2+\rho}{2(1-\rho)^2(\theta^2+\rho)} & \frac{\rho\theta^2}{2(\theta^2+\rho)^2} \\ \frac{\rho\theta^2}{2(\theta^2+\rho)^2} & \frac{\rho\theta^2}{2(\theta^2+\rho)}\left[\frac{(1-\rho)(2\rho+\theta^2)}{(\theta^2+\rho)^2} + 2 + \theta^2\right] \end{bmatrix}\right). \tag{A84}$$

This completes the proof.

References

Ao, M., Li, Y., Zheng, X., 2019. Approaching mean-variance efficiency for large portfolios. Review of Financial Studies 32, 2890–2919.

Asness, C., Frazzini, A., 2013. The devil in HML's details. Journal of Portfolio Management 39, 49–68.

Barillas, F., Kan, R., Robotti, C., Shanken, J., 2020. Model comparison with Sharpe ratios. Journal of Financial and Quantitative Analysis 55, 1840–1874.

Barillas, F., Shanken, J., 2017. Which alpha? Review of Financial Studies 30, 1316–1338.

Barillas, F., Shanken, J., 2018. Comparing asset pricing models. Journal of Finance 73, 715–754.

Carhart, M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57–82.

Chib, S., Zeng, X., Zhao, L., 2020. On comparing asset pricing models. Journal of Finance 75, 551–577.

Chib, S., Zhao, L., Zhou, G., 2024. Winners from winners: a take of risk factors. Management Science 70, 396–414.

Cochrane, J. H., Saá-Requejo., J., 2000. Beyond arbitrage: good-deal asset price bounds in incomplete markets. Journal of Political Economy 108, 79–119.

Detzel, A., Novy-Marx, R., Velikov, M., 2023. Model comparison with transaction costs. Journal of Finance 78, 1743–1775.

Dickey, J. M., 1967. Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution. Annals of Mathematical Statistics 38, 511–518.

El Karoui, N., 2010. High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: risk underestimation. Annals of Statistics 38, 3487–3566.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.

Fama, E. F., French, K. R., 2018. Choosing factors. Journal of Financial Economics 128, 234–252.

Frazzini, A., Pedersen, L. H., 2014. Betting against beta. Journal of Financial Economics 111, 1–25.

Gibbons, M., Ross, S., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121–1152.

Harvey, C., Liu, Y., Zhu, H., 2016. ... and the cross-section of expected returns. Review of Financial Studies 29, 5–68.

Harvey, C., Zhou, G., 1990. Bayesian inference in asset pricing tests. Journal of Financial Economics 26, 221–254.

Hillier, G., Kan, R., Wang, X., 2014. Generating functions and short recursions, with applications to the moments of quadratic Forms in noncentral normal vectors. Econometric Theory 30, 436–473.

Hou, K., Mo, H., Xue, C., Zhang, L., 2021. An augmented q-factor model with expected growth. Review of Finance 25, 1–41.

Hou, K., Xue, C., Zhang, L., 2015. Digesting anomalies: an investment approach. Review of Financial Studies 28, 650–705.

Jobson, J. D., Korkie, R., 1981. Performance hypothesis testing with the Sharpe and Treynor Measures. Journal of Finance 36, 889–908.

Kan, R., Lassance, N., 2024. Optimal portfolio choice with fat tails and parameter uncertainty. Unpublished working paper. University of Toronto.

Kan, R., Robotti, C., 2016. The exact distribution of the Hansen-Jagannathan bound. Management Science 62, 1915–1943.

Kan, R., Wang, X., 2023. Optimal portfolio choice with unknown benchmark efficiency. Management Science, forthcoming.

Kan, R., Wang, X., Zhou, G., 2022. Optimal portfolio choice with estimation risk: no risk-free asset case. Management Science 68, 2047–2068.

Kan, R., Zhou, G., 2007. Optimal portfolio choice with parameter uncertainty. Journal of Financial and Quantitative Analysis 42, 621–656.

Kazarinoff, D. K., 1956. On Wallis' formula. Edinburgh Mathematical Notes 40, 19–21.

Krishnan, M., 1967. The moments of a doubly noncentral t-distribution. Journal of the American Statistical Association 62, 278–287.

Li, S., DeMiguel, V., Martin-Utrera, A., 2023. Comparing factor models with price-impact costs. Unpublished working paper. London Business School.

Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–37.

Lo, A. W., 2002. The statistics of Sharpe ratios. Financial Analyst Journal 58, 36–52.

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

MacKinlay, A. C., 1995. Multi-factor models do not explain deviations from the CAPM. Journal of Financial Economics 38, 3–28.

McLean, R. D., Pontiff, J., 2016. Does academic research destroy stock return predictability? Journal of Finance 71, 5–32.

Muirhead, R. J., 1982. Aspects of Multivariate Statistical Theory. Wiley, New York.

Novy-Marx, R., Velikov, M., 2016. A taxonomy of anomalies and their trading costs. Review of Financial Studies 29, 104–147.

Pastor, L., Stambaugh, R. F., 2000. Comparing asset pricing models: an investment perspective. Journal of Financial Economics 56, 335–381.

Patton, A. J., Weller, B. M., 2020. What you see is not what you get: the costs of trading market anomalies. Journal of Financial Economics 137, 515–549.

Politis, D. N., Romano, J. P., 1994. The stationary bootstrap. Journal of the American Statistical Association 89, 1303–1313.

Ross, S. A., 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341–360.

Sharpe, W. F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. Journal of Finance 19, 425–442.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455–1508.

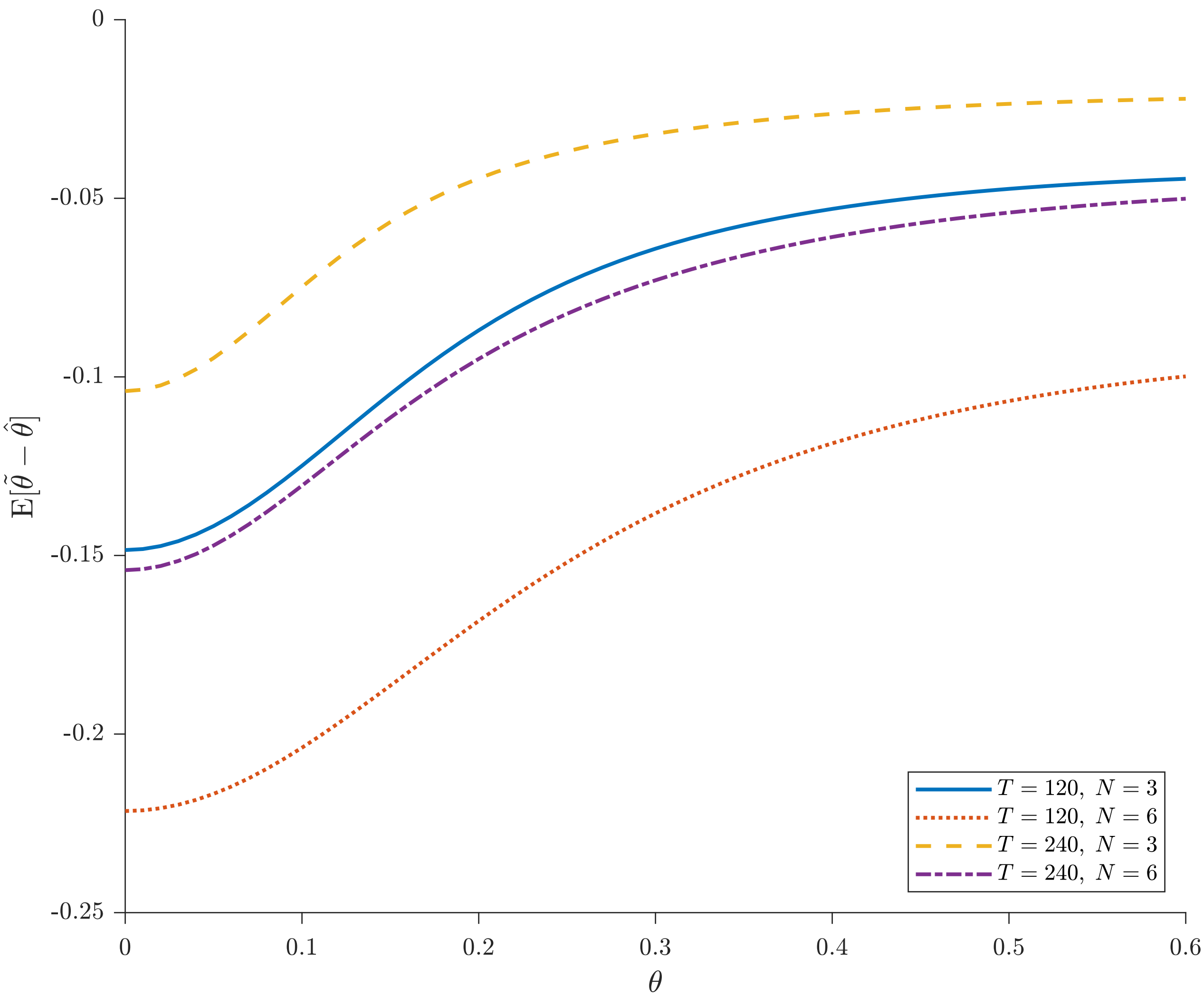

Fig. 1. Expected Difference in Out-of-sample and In-sample Sharpe Ratio of an Asset Pricing Model

The figure plots the expected difference between the out-of-sample Sharpe ratio and the in-sample Sharpe ratio, $\mathbb{E}[\tilde{\theta} - \hat{\theta}]$, as a function of the population Sharpe ratio, θ, for an asset pricing model with N traded factors and an estimation window with T periods. Plots for two different values of N (3 and 6) and T (120 and 240) are presented.

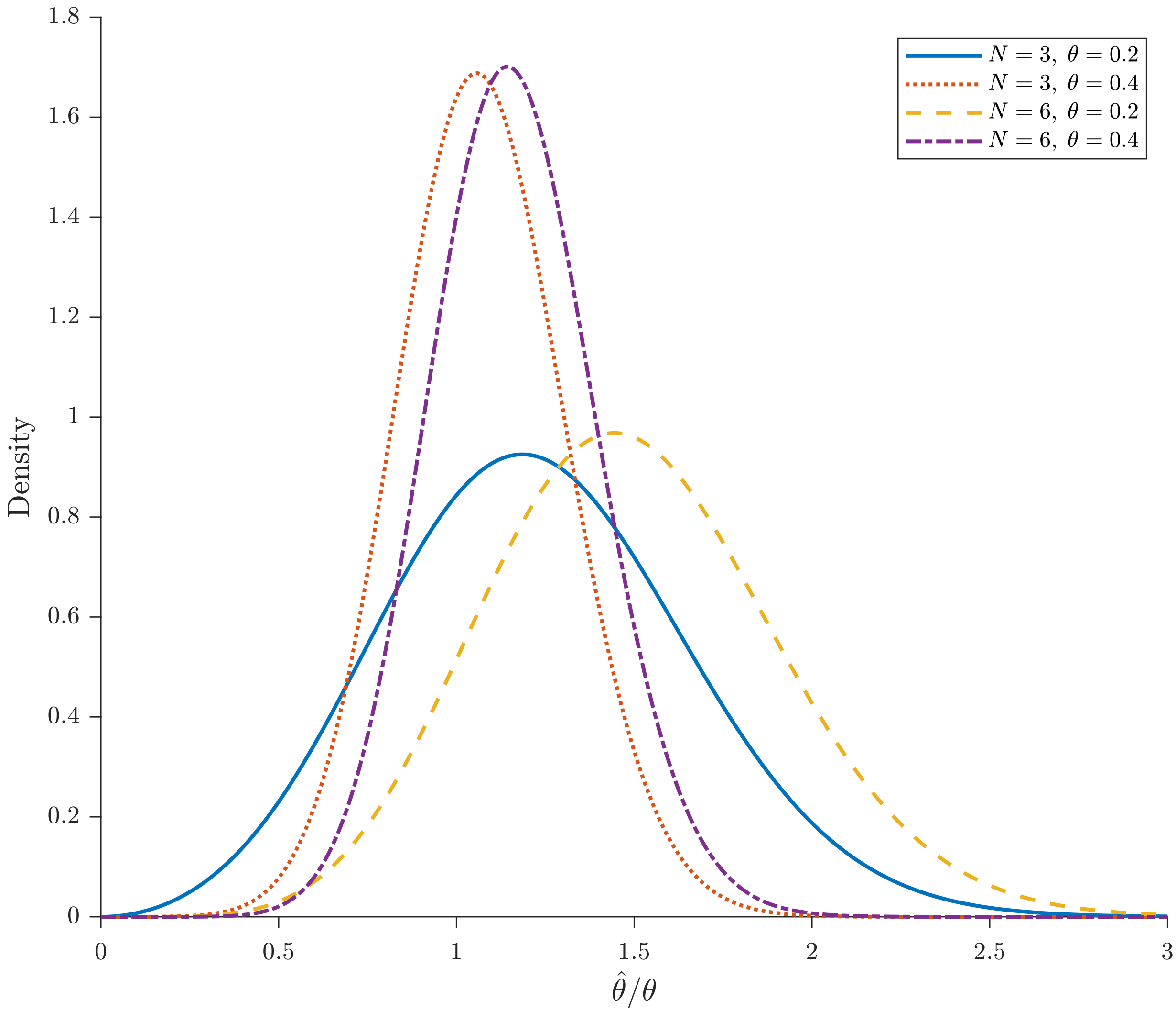

Fig. 2. Density of In-sample Sharpe Ratio of an Asset Pricing Model ($T = 120$)

The figure plots the density of $\hat{\theta}/\theta$ of an asset pricing model with N traded factors when the length of estimation window is $T = 120$, where $\hat{\theta}$ is the in-sample Sharpe ratio and θ is the population Sharpe ratio. Plots for two different values of number of traded factors ($N = 3$ and 6) and population Sharpe ratio ($\theta = 0.2$ and 0.4) are presented.

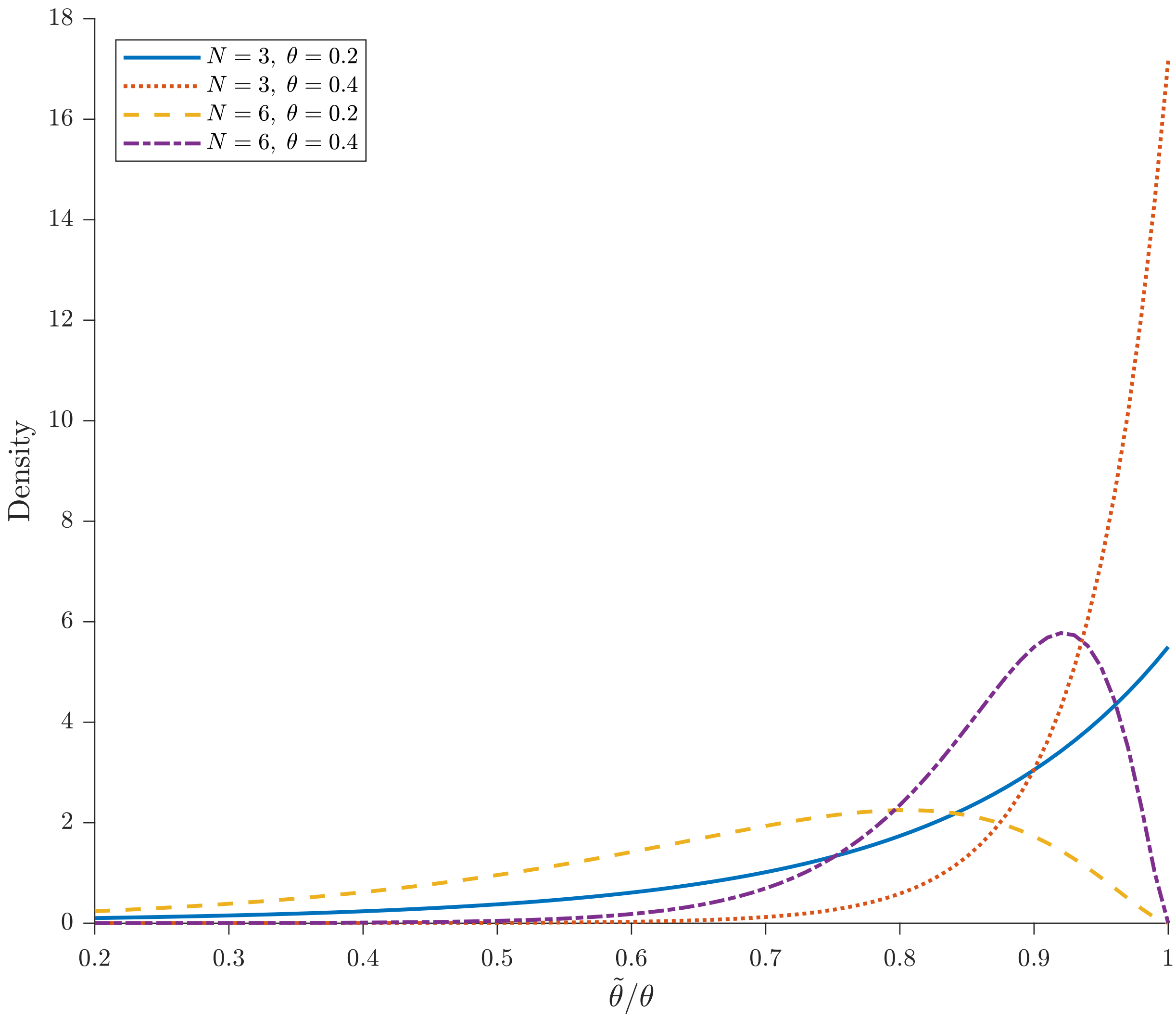

Fig. 3. Density of Out-of-sample Sharpe Ratio of an Asset Pricing Model ($T = 120$)

The figure plots the density of $\tilde{\theta}/\theta$ of an asset pricing model with N traded factors when the length of estimation window is $T = 120$, where $\tilde{\theta}$ is the out-of-sample Sharpe ratio and θ is the population Sharpe ratio. Plots for two different values of number of traded factors ($N = 3$ and 6) and population Sharpe ratio ($\theta = 0.2$ and 0.4) are presented.

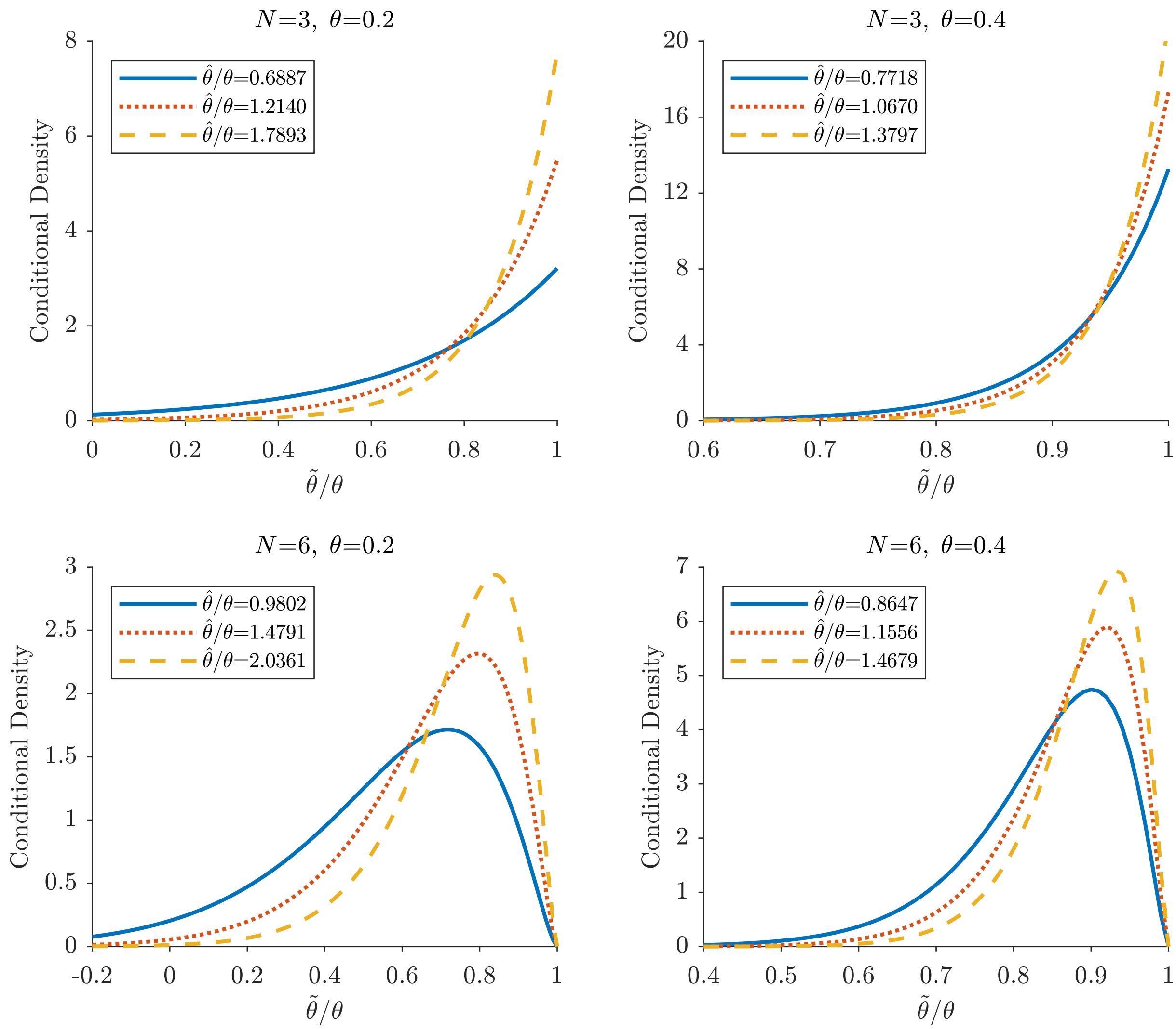

Fig. 4. Conditional Density of Out-of-sample Sharpe Ratio of an Asset Pricing Model ($T = 120$)

The figure plots the conditional density of the normalized out-of-sample Sharpe ratio ($\tilde{\theta}/\theta$) of an asset pricing model when conditional on the in-sample Sharpe ratio ($\hat{\theta}$) is at its 10th (solid line), 50th (dotted line), and 90th (dashed line) percentiles. The length of estimation window is $T = 120$ and plots for two different values of number of traded factors ($N = 3$ and 6) and population Sharpe ratio ($\theta = 0.2$ and 0.4) are presented.

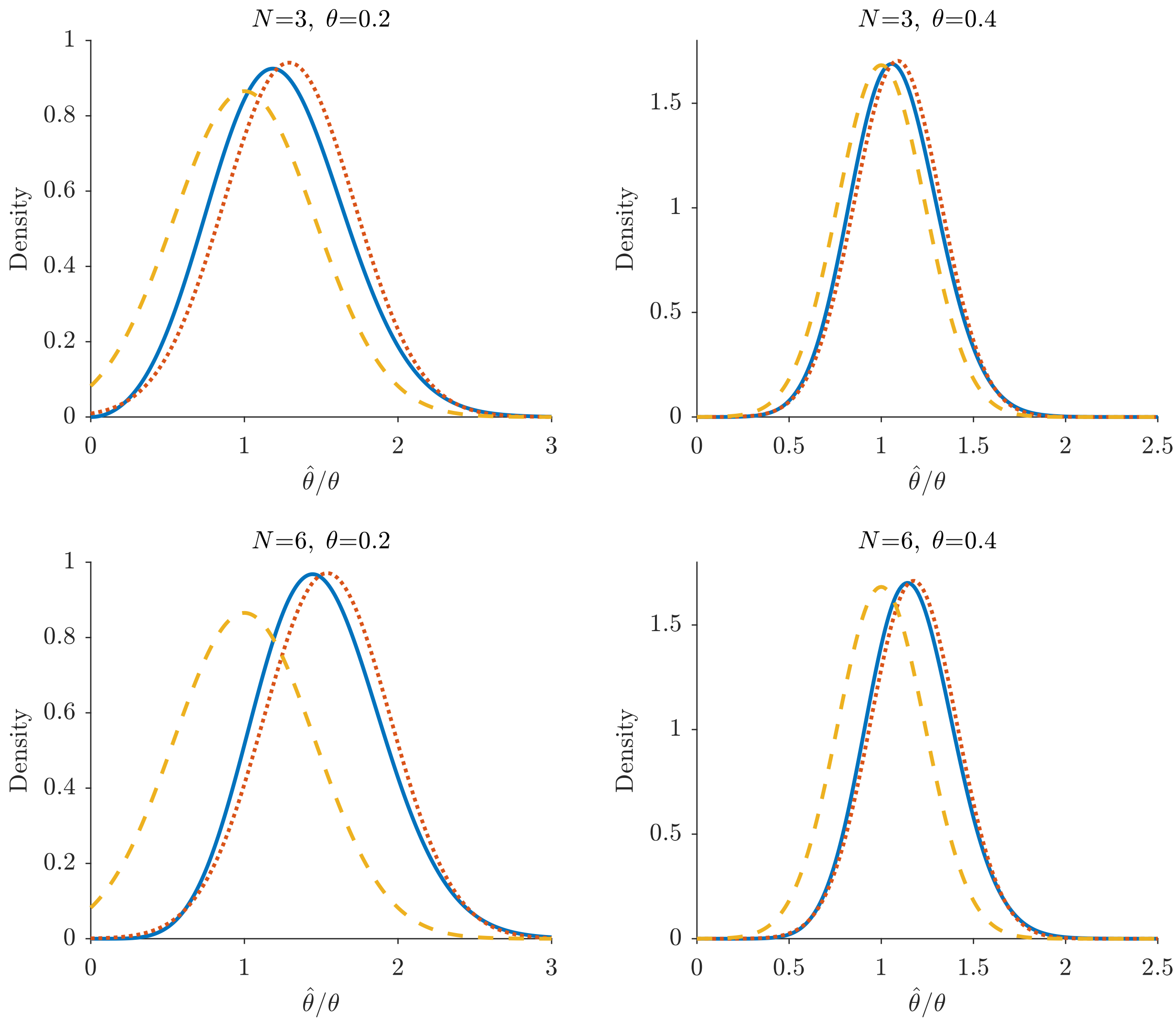

Fig. 5. Exact and Approximate Densities of In-sample Sharpe Ratio of an Asset Pricing Model ($T = 120$)

The figure plots the exact density (solid line) of the normalized in-sample Sharpe ratio ($\hat{\theta}/\theta$) and two different approximated density, the first one assumes N is fixed and $T \to \infty$ (dashed line), and the second one assumes both N and $T \to \infty$ with $N/T \to \rho$ (dotted line). The length of time series is $T = 120$ and plots for two different values of number of traded factors ($N = 3$ and 6) and population Sharpe ratio ($\theta = 0.2$ and 0.4) are presented.

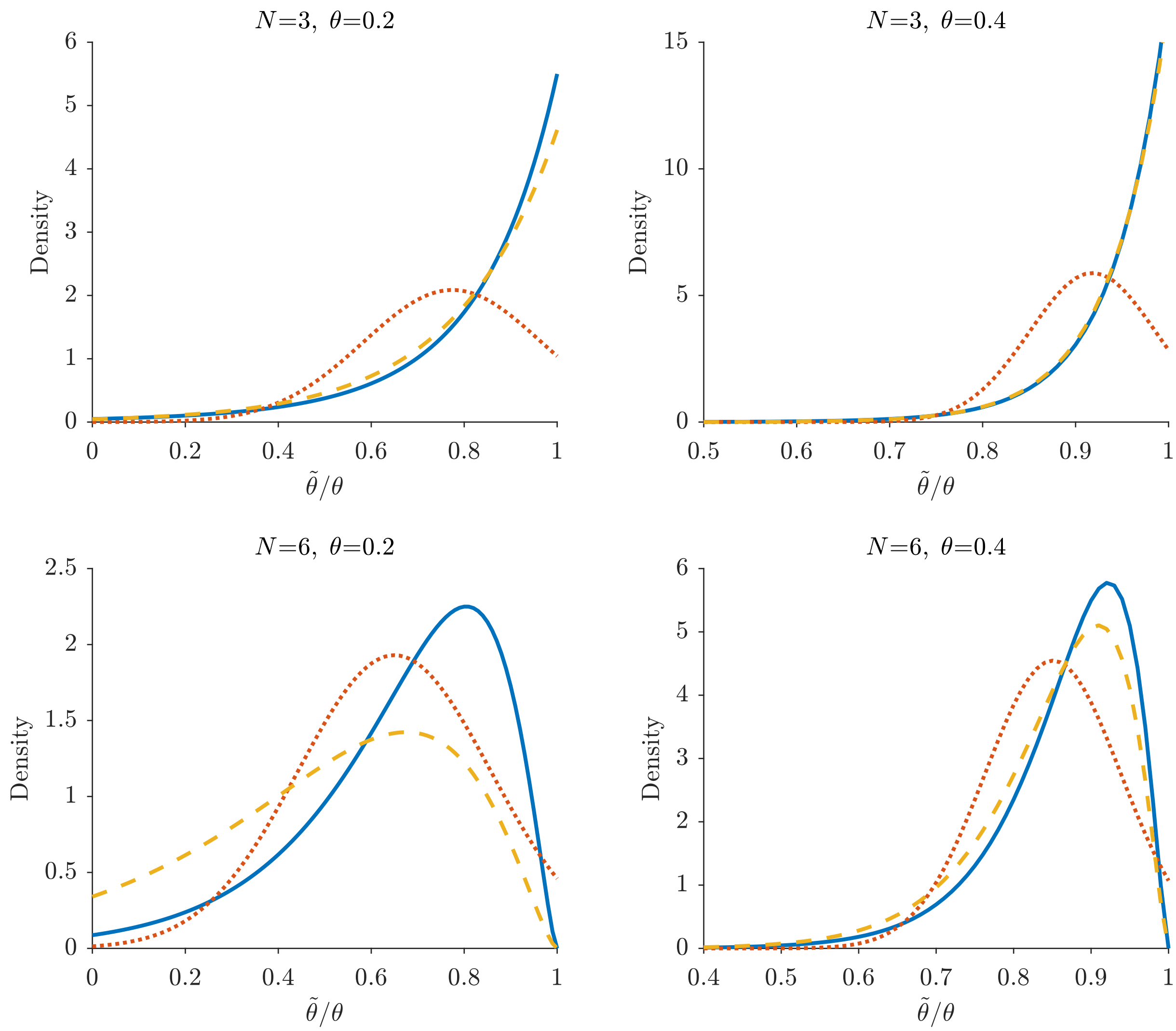

Fig. 6. Exact and Approximate Densities of Out-of-sample Sharpe Ratio of an Asset Pricing Model ($T = 120$)

The figure plots the exact density (solid line) of the normalized out-of-sample Sharpe ratio ($\tilde{\theta}/\theta$) and two different approximated density, the first one assumes N is fixed and $T \to \infty$ (dashed line), and the second one assumes both N and $T \to \infty$ with $N/T \to \rho$ (dotted line). The length of time series is $T = 120$ and plots for two different values of number of traded factors ($N = 3$ and 6) and population Sharpe ratio ($\theta = 0.2$ and 0.4) are presented.

Table 1
In-sample Sharpe Ratios of Asset Pricing Models

This table reports the in-sample Sharpe ratios of eight asset pricing models for the full sample period (1967/1–2021/12) as well as the two subperiods (1967/1–1994/6 and 1994/7–2021/12). The asset pricing models considered are: CAPM, Fama-French 3-factor model (FF-3), Carhart 4-factor model (Carhart-4), Betting-against-beta 2-factor model (BAB), Fama-French 5-factor model (FF-5), HXZ's q-factor model, Barillas and Shanken 6-factor model (BS-6), and HMXZ's q^5 model. The column "Publication Year" presents the year that the model was first published. The Gibbons-Ross-Shanken F-test is conducted to compare the in-sample Sharpe ratio of a given multi-factor model with that of the CAPM. ***, **, and * denote that the in-sample Sharpe ratio of the given model is higher than that of the CAPM at the 1%, 5%, and 10% significance levels.

	Publication Year	Full Sample 1967/1–2021/12	First Half 1967/1–1994/6	Second Half 1994/7–2021/12
CAPM	1964	0.133	0.087	0.182
FF-3	1993	0.184***	0.248***	0.185
Carhart-4	1997	0.278***	0.386***	0.244**
BAB	2014	0.305***	0.392***	0.307***
FF-5	2015	0.325***	0.496***	0.368***
HXZ q	2015	0.399***	0.611***	0.350***
BS-6	2018	0.464***	0.787***	0.367***
HMXZ q^5	2021	0.599***	0.775***	0.510***

Table 2
In-sample versus Out-of-sample Sharpe Ratios of Asset Pricing Models

In this table, we divide the sample period into halves, and treat the first half as the estimation window and the second half as the out-of-sample period. Using data in the estimation window, the sample optimal portfolio is constructed, and the out-of-sample Sharpe ratio (OS-SR) is computed using the returns in the out-of-sample period of the sample optimal portfolio. For comparison, the in-sample Sharpe ratio (IS-SR) based on the return data in the estimation window is also reported. Gibbons-Ross-Shanken F-test is conducted to compare IS-SR of a given multi-factor model with that of the CAPM. One-sided test based on asymptotic distribution is conducted to compare the OS-SR of a given multi-factor model with that of the CAPM. ***, **, * denote that the Sharpe ratio of a given multi-factor model is higher than that of the CAPM at the 1%, 5%, and 10% significance levels. The sample period for the left and the right panel are 1967/1–2021/12 and 1994/7–2021/12, respectively.

	1967/1–2021/12		1994/7–2021/12	
	IS-SR	OS-SR	IS-SR	OS-SR
CAPM	0.087	0.182	0.137	0.223
FF-3	0.248***	0.111	0.262**	0.094
Carhart-4	0.386***	0.154	0.353***	0.113
BAB	0.392***	0.189	0.331***	0.286*
FF-5	0.496***	0.192	0.393***	0.317**
HXZ q	0.611***	0.214	0.439***	0.269
BS-6	0.787***	0.226	0.500***	0.238
HMXZ q^5	0.775***	0.411***	0.601***	0.419***

Table 3
In-sample versus Out-of-sample Sharpe Ratios – Simulation Results

We apply the stationary block bootstrap procedure of Politis and Romano (1994) to our empirical dataset (1967–2021) to generate T monthly data, with an expected block length of 10 months. The T monthly data are divided into halves, with the first half treated as the estimation window and the second half as the out-of-sample period. IS-SR and OS-SR are computed using the simulated data, following the same procedure as in Table 2,. We run the simulation 10,000 times. Cross-simulation average values of IS-SR and OS-SR, and the corresponding standard deviations (in brackets), are reported in the table. In the left panel, $T = 660$, and in the right panel, $T = 330$.

	$T = 660$		$T = 330$	
	IS-SR	OS-SR	IS-SR	OS-SR
CAPM	0.137	0.136	0.139	0.140
	(0.062)	(0.062)	(0.088)	(0.088)
FF-3	0.211	0.163	0.227	0.146
	(0.074)	(0.076)	(0.123)	(0.113)
Carhart-4	0.303	0.260	0.320	0.239
	(0.073)	(0.075)	(0.127)	(0.119)
BAB	0.316	0.299	0.310	0.277
	(0.115)	(0.115)	(0.196)	(0.189)
FF-5	0.365	0.300	0.400	0.284
	(0.061)	(0.067)	(0.094)	(0.102)
HXZ q	0.424	0.386	0.442	0.373
	(0.066)	(0.067)	(0.115)	(0.108)
BS-6	0.506	0.441	0.537	0.424
	(0.084)	(0.085)	(0.143)	(0.133)
HMXZ q^5	0.624	0.585	0.647	0.574
	(0.069)	(0.071)	(0.103)	(0.105)

Table 4
Break-even Sharpe Ratios

The break-even Sharpe ratio is defined as the value of θ such that $\mathbb{E}[\tilde{\theta}|\tilde{\theta} \leq \tilde{\theta}_c] = \theta_1$ where θ_1 is the population Sharpe ratio of the CAPM, θ is the population Sharpe ratio of a multi-factor model, $\tilde{\theta}$ is the out-of-sample Sharpe ratio of the multi-factor model, and $\tilde{\theta}_c$ is the c-percentile value of $\tilde{\theta}$. The finite sample distribution of $\tilde{\theta}$ is available in (22). This table reports the break-even Sharpe ratios for different values of number of traded factors in the multi-factor model (N), the length of estimation window (T), and the population Sharpe ratio of the CAPM (θ_1). Panels A and B present the results for $c = 50$ and $c = 25$ respectively, where c is the percentile specified in $\mathbb{E}[\tilde{\theta}|\tilde{\theta} \leq \tilde{\theta}_c]$.

A. $c = 50$

θ_1	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
				A1. $T = 120$				
$N = 2$	0.123	0.162	0.200	0.240	0.283	0.328	0.375	0.422
$N = 3$	0.133	0.176	0.217	0.258	0.301	0.346	0.392	0.438
$N = 4$	0.141	0.187	0.229	0.272	0.315	0.360	0.405	0.452
$N = 5$	0.147	0.195	0.239	0.283	0.327	0.371	0.417	0.463
$N = 6$	0.152	0.203	0.248	0.293	0.337	0.382	0.427	0.474
				A2. $T = 240$				
$N = 2$	0.098	0.136	0.176	0.220	0.266	0.314	0.362	0.411
$N = 3$	0.107	0.148	0.189	0.232	0.277	0.324	0.372	0.420
$N = 4$	0.113	0.156	0.198	0.241	0.286	0.332	0.379	0.427
$N = 5$	0.118	0.163	0.206	0.249	0.294	0.340	0.386	0.434
$N = 6$	0.122	0.169	0.212	0.256	0.301	0.346	0.393	0.440

Table 4
Break-even Sharpe Ratios (Cont'd)

B. $c = 25$

θ_1	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
				B1. $T = 120$				
$N = 2$	0.154	0.187	0.221	0.258	0.299	0.342	0.388	0.434
$N = 3$	0.164	0.202	0.240	0.279	0.320	0.363	0.407	0.453
$N = 4$	0.171	0.213	0.253	0.293	0.335	0.378	0.422	0.468
$N = 5$	0.177	0.222	0.263	0.305	0.347	0.391	0.435	0.481
$N = 6$	0.182	0.229	0.273	0.315	0.358	0.402	0.447	0.492
				B2. $T = 240$				
$N = 2$	0.118	0.152	0.189	0.230	0.275	0.321	0.369	0.417
$N = 3$	0.127	0.164	0.203	0.244	0.288	0.333	0.380	0.428
$N = 4$	0.133	0.173	0.213	0.254	0.298	0.343	0.389	0.436
$N = 5$	0.137	0.180	0.221	0.263	0.306	0.351	0.397	0.443
$N = 6$	0.142	0.185	0.227	0.270	0.313	0.358	0.404	0.450

Table 5
GRS Test and the Break-even Sharpe Ratios

In this table, we compare the multi-factor asset pricing models against the CAPM using the GRS test as well as the test that also incorporates the break-even Sharpe ratio. The in-sample Sharpe ratios of the models, together with the corresponding p-values (in italics) from the GRS test, are reported in the column "GRS". Assuming that the observed in-sample Sharpe ratio of the CAPM ($\hat{\theta}_1$) is at $100(1-\alpha)$-percentile of its distribution, we obtain the corresponding population Sharpe ratio (θ_1) and report it in the row of CAPM. Given θ_1, the break-even Sharpe ratio is derived for each multi-factor model and reported in the corresponding row. The p-values from the null distribution in (33) are presented in italics, below the break-even Sharpe ratios. Five different values of α are examined: 10%, 25%, 50%, 75%, and 90%. Panels A, B, and C use data in 2008/4–2021/12, 1994/7–2021/12, and 1967/1–2021/12, respectively.

A. 2008/4–2021/12 ($T = 165$)

		$\alpha = 1 - F_{\theta_1}(\hat{\theta}_1)$				
	GRS	10%	25%	50%	75%	90%
CAPM	0.223	0.121	0.168	0.222	0.275	0.323
FF-3	0.267	0.178	0.218	0.263	0.311	0.355
	0.193	*0.554*	*0.583*	*0.602*	*0.616*	*0.628*
Carhart-4	0.268	0.189	0.229	0.275	0.322	0.366
	0.340	*0.711*	*0.746*	*0.771*	*0.789*	*0.802*
BAB	0.288	0.164	0.203	0.248	0.297	0.342
	0.024	*0.192*	*0.194*	*0.192*	*0.192*	*0.195*
FF-5	0.376	0.197	0.238	0.285	0.332	0.375
	0.009	*0.114*	*0.143*	*0.169*	*0.191*	*0.209*
HXZ q	0.294	0.189	0.229	0.275	0.322	0.366
	0.135	*0.479*	*0.522*	*0.555*	*0.579*	*0.598*
BS-6	0.315	0.204	0.246	0.293	0.340	0.384
	0.194	*0.575*	*0.633*	*0.679*	*0.713*	*0.738*
HMXZ q^5	0.430	0.197	0.238	0.285	0.332	0.375
	0.001	*0.022*	*0.030*	*0.038*	*0.046*	*0.053*

Table 5
GRS Test and the Break-even Sharpe Ratios (Cont'd)

B. 1994/7–2021/12 ($T = 330$)

		$\alpha = 1 - F_{\theta_1}(\hat{\theta}_1)$				
	GRS	10%	25%	50%	75%	90%
CAPM	0.182	0.110	0.144	0.181	0.219	0.252
FF-3	0.185	0.146	0.174	0.207	0.241	0.273
	0.828	*0.955*	*0.958*	*0.960*	*0.962*	*0.963*
Carhart-4	0.244	0.154	0.182	0.215	0.249	0.280
	0.041	*0.298*	*0.325*	*0.346*	*0.362*	*0.375*
BAB	0.307	0.136	0.164	0.197	0.232	0.264
	0.000	*0.002*	*0.002*	*0.001*	*0.001*	*0.002*
FF-5	0.368	0.160	0.189	0.221	0.255	0.286
	0.000	*0.001*	*0.001*	*0.002*	*0.002*	*0.003*
HXZ q	0.350	0.154	0.182	0.215	0.249	0.280
	0.000	*0.001*	*0.002*	*0.002*	*0.002*	*0.002*
BS-6	0.367	0.165	0.194	0.227	0.261	0.291
	0.000	*0.002*	*0.004*	*0.005*	*0.006*	*0.007*
HMSZ q^5	0.510	0.160	0.189	0.221	0.255	0.286
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*

Table 5
GRS Test and the Break-even Sharpe Ratios (Cont'd)

C. 1967/1–2021/12 ($T = 660$)

		$\alpha = 1 - F_{\theta_1}(\hat{\theta}_1)$				
	GRS	10%	25%	50%	75%	90%
CAPM	0.133	0.083	0.106	0.133	0.159	0.183
FF-3	0.184	0.107	0.127	0.150	0.174	0.197
	0.006	*0.100*	*0.106*	*0.111*	*0.114*	*0.117*
Carhart-4	0.278	0.112	0.133	0.156	0.179	0.201
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
BAB	0.305	0.100	0.120	0.144	0.168	0.191
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
FF-5	0.325	0.117	0.137	0.160	0.184	0.205
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
HXZ q	0.399	0.112	0.133	0.156	0.179	0.201
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
BS-6	0.464	0.120	0.141	0.164	0.188	0.209
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
HMXZ q^5	0.599	0.117	0.137	0.160	0.184	0.205
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*

Table 6
Test in the Bayesian Framework

In this table, the multi-factor asset pricing models are compared to the CAPM in a Bayesian framework. The prior distributions are specified in (36) and (37). For the sharp null ($H_0 : \delta^2 = 0$), we assume a prior null probability of 0.5 (i.e., $p_0(0) = 0.5$); and report the posterior null probability (i.e., $p_1(0)$) in the columns "$\delta^2 = 0$". In the columns "$\delta^2 \le \delta_b^2$", $p_1(\delta_b^2)$, $p_0(\delta_b^2)$, and $\tilde{p}_1(\delta_b^2)$ in (41), (42) and (47) are reported. Three different sample periods, i.e., 2008/4–202112, 1994/7–2021/12, and 1967/1–2021/12, are examined in the table. Panels A, B, and C present the results for different values of a and b in (37).

A. $a = 3$, $b = 3$

		2008/4–2021/12		1994/7–2021/12		1967/1–2021/12	
		$\delta^2 = 0$	$\delta^2 \le \delta_b^2$	$\delta^2 = 0$	$\delta^2 \le \delta_b^2$	$\delta^2 = 0$	$\delta^2 \le \delta_b^2$
FF-3	(p_1)	0.496	0.715	0.753	0.877	0.079	0.180
	(p_0)		0.650		0.464		0.305
	$(\tilde{p}_1)$		0.856		0.970		0.245
Carhart-4	(p_1)	0.606	0.750	0.272	0.360	0.000	0.000
	(p_0)		0.573		0.371		0.216
	$(\tilde{p}_1)$		0.901		0.534		0.000
BAB	(p_1)	0.253	0.491	0.002	0.016	0.000	0.000
	(p_0)		0.748		0.604		0.466
	$(\tilde{p}_1)$		0.620		0.018		0.000
FF-5	(p_1)	0.101	0.256	0.000	0.005	0.000	0.000
	(p_0)		0.507		0.303		0.161
	$(\tilde{p}_1)$		0.331		0.005		0.000
HXZ q	(p_1)	0.445	0.613	0.000	0.008	0.000	0.000
	(p_0)		0.573		0.371		0.216
	$(\tilde{p}_1)$		0.785		0.008		0.000
BS-6	(p_1)	0.542	0.577	0.000	0.008	0.000	0.000
	(p_0)		0.452		0.252		0.125
	$(\tilde{p}_1)$		0.806		0.008		0.000
HMXZ q^5	(p_1)	0.016	0.099	0.000	0.000	0.000	0.000
	(p_0)		0.507		0.303		0.161
	$(\tilde{p}_1)$		0.114		0.000		0.000

Table 6
Test in the Bayesian Framework (Cont'd)

B. $a = 3$, $b = 5$

		2008/4–2021/12		1994/7–2021/12		1967/1–2021/12	
		$\delta^2 = 0$	$\delta^2 \leq \delta_b^2$	$\delta^2 = 0$	$\delta^2 \leq \delta_b^2$	$\delta^2 = 0$	$\delta^2 \leq \delta_b^2$
FF-3	(p_1)	0.481	0.773	0.725	0.893	0.082	0.198
	(p_0)		0.788		0.619		0.444
	$(\tilde{p}_1)$		0.882		0.970		0.264
Carhart-4	(p_1)	0.572	0.802	0.266	0.403	0.000	0.001
	(p_0)		0.735		0.539		0.353
	$(\tilde{p}_1)$		0.915		0.562		0.001
BAB	(p_1)	0.272	0.572	0.002	0.023	0.000	0.000
	(p_0)		0.849		0.724		0.585
	$(\tilde{p}_1)$		0.688		0.025		0.000
FF-5	(p_1)	0.113	0.327	0.000	0.008	0.000	0.000
	(p_0)		0.686		0.474		0.289
	$(\tilde{p}_1)$		0.403		0.008		0.000
HXZ q	(p_1)	0.430	0.683	0.000	0.011	0.000	0.000
	(p_0)		0.735		0.539		0.353
	$(\tilde{p}_1)$		0.820		0.012		0.000
BS-6	(p_1)	0.503	0.650	0.001	0.011	0.000	0.000
	(p_0)		0.641		0.420		0.241
	$(\tilde{p}_1)$		0.826		0.012		0.000
HMXZ q^5	(p_1)	0.021	0.143	0.000	0.000	0.000	0.000
	(p_0)		0.686		0.474		0.289
	$(\tilde{p}_1)$		0.161		0.000		0.000

Table 6
Test in the Bayesian Framework (Cont'd)

C. $a = 5$, $b = 3$

		2008/4–2021/12		1994/7–2021/12		1967/1–2021/12	
		$\delta^2 = 0$	$\delta^2 \leq \delta_b^2$	$\delta^2 = 0$	$\delta^2 \leq \delta_b^2$	$\delta^2 = 0$	$\delta^2 \leq \delta_b^2$
FF-3	(p_1)	0.511	0.664	0.784	0.858	0.077	0.153
	(p_0)		0.514		0.325		0.189
	$(\tilde{p}_1)$		0.836		0.969		0.218
Carhart-4	(p_1)	0.640	0.698	0.283	0.314	0.000	0.000
	(p_0)		0.416		0.225		0.107
	$(\tilde{p}_1)$		0.892		0.508		0.000
BAB	(p_1)	0.239	0.432	0.001	0.012	0.000	0.000
	(p_0)		0.647		0.493		0.363
	$(\tilde{p}_1)$		0.568		0.013		0.000
FF-5	(p_1)	0.094	0.206	0.000	0.004	0.000	0.000
	(p_0)		0.338		0.158		0.064
	$(\tilde{p}_1)$		0.280		0.004		0.000
HXZ q	(p_1)	0.462	0.552	0.000	0.006	0.000	0.000
	(p_0)		0.416		0.225		0.107
	$(\tilde{p}_1)$		0.759		0.006		0.000
BS-6	(p_1)	0.585	0.506	0.000	0.006	0.000	0.000
	(p_0)		0.277		0.114		0.040
	$(\tilde{p}_1)$		0.795		0.006		0.000
HMXZ q^5	(p_1)	0.013	0.075	0.000	0.000	0.000	0.000
	(p_0)		0.338		0.158		0.064
	$(\tilde{p}_1)$		0.087		0.000		0.000