A Test of the Efficiency of a Given Portfolio in High Dimensions

Mikhail Chernov, Bryan Kelly, Semyon Malamud, and Johannes Schwab*

This version: March 3, 2025

Abstract

We generalize the seminal Gibbons-Ross-Shanken test to the empirically relevant case where the number of test assets far exceeds the number of observations. In such a setting, one needs to use a regularized estimator of the covariance matrix of test assets, which leads to biases in the original test statistic. Random Matrix Theory allows us to account for these biases and to evaluate the test's power. Power increases with the number of test assets and reaches the maximum for a broad range of local alternatives. These conclusions are supported by an extensive simulation study. We implement the test empirically for state-of-the-art candidate efficient portfolios and test assets.

Keywords: efficient portfolio, cross-section of stock returns, testing, regularization, random matrix theory

JEL: C12, C40, C55, C57, G12

*Mikhail Chernov is at UCLA Anderson, NBER, and CEPR. Bryan Kelly is at AQR Capital Management, Yale School of Management, and NBER. Semyon Malamud is at Swiss Finance Institute, EPFL, and CEPR, and is a consultant to AQR. Johannes Schwab is at Swiss Finance Institute, EPFL. Semyon Malamud gratefully acknowledges the financial support of the Swiss Finance Institute and the Swiss National Science Foundation, Grant 100018_192692. We are grateful to Yuan Liao, Andreas Neuhierl, and Olivier Scaillet for their excellent comments and suggestions. AQR Capital Management is a global investment management firm that may or may not apply similar investment techniques or methods of analysis as described herein. The views expressed here are those of the authors and not necessarily those of AQR.

1 Introduction

There is a long-standing tradition in cross-sectional asset pricing of testing candidate tradable factors by regressing the excess returns of test assets on these factors, assuming constant betas. The seminal Gibbons, Ross and Shanken (1989) (GRS) test determines whether the alphas in such regressions are jointly equal to zero. If the alphas are zero, the factors price the test assets unconditionally; that is, the portfolio unconditionally spanned by the candidate factors is efficient. Furthermore, the Hansen and Richard (1987) important contribution shows that if alphas are jointly zero when all feasible portfolios are used as test assets, the factors would also price assets conditionally.

However, using an effectively infinite number of test assets poses a challenge for the GRS test. As GRS demonstrate, the test's power can decline significantly as the number of test assets increases. They recommend limiting the number of assets to one-third to one-half of the sample size, T, treated as fixed. Extant research addresses this issue by considering sparse alternative hypotheses (either non-zero alphas or the covariance matrix of returns are sparse), with the work by Fan, Liao and Yao (2015) being the most prominent example. As we show in the paper, such settings are not rotation-invariant. For instance, if the alternative hypothesis posits sparse alphas or the covariance matrix for individual stocks, forming portfolios from these stocks will not preserve sparsity. Since economically interesting portfolios are commonly used as test assets, and Hansen and Richard (1987) call for the consideration of all feasible portfolios, the sparsity assumption may be questionable for our purpose.

Our paper addresses these challenges by developing a new test that is applicable when the number of test assets approaches infinity allowing for both sparse and non-sparse environments. This test enables the implementation of the Hansen and Richard (1987) insight, allowing for testing a given linear factor model's conditional pricing abilities. We also evaluate the power of the new test.

The GRS statistic is well known for incorporating the information ratio squared of test assets. However, when the number of test assets, P, exceeds the number of time periods, $P/T \to c > 1$, the covariance matrix of the regression residuals — whose inverse is crucial in this ratio — becomes singular and requires regularization. This regularization introduces a bias in the test statistic.[1] An additional complication is that, in contrast to the GRS setting, the bias depends on the unknown covariance matrix of test assets, i.e., a naive extension of GRS leads to a non-pivotal statistic. To address these issues, we apply Random Matrix Theory (RMT) to de-bias the regularized GRS statistic, obtain a pivotal test statistic, and derive its asymptotic distribution. In particular, the debiased statistic converges to zero at the rate $T^{-1/2}$.

Despite the changes, our test statistic has an economic interpretation that is similar to that of GRS. That is, it captures the distance between the Sharpe ratio squared of test assets and the candidate factor combined together and the Sharpe ratio squared of the candidate efficient portfolio alone. Thus, the distance measures the economic gain from investing in assets other than the candidate factor. If the factor is efficient, there should be no gain.

GRS is already grappling with its test's power, and adding more assets than observations appears to create an even larger problem. As we demonstrate next, the large dimension of test assets actually improves the test's power compared to GRS despite complicating the statistical inference.

First, we consider asymptotic power under a local alternative. We select this alternative to be at the "border of detectability", where the true information ratio squared is of order $T^{-1/2}$ (which is selected on the basis of the statistic's convergence rate). The corresponding power depends on the unknown covariance matrix of test asset returns, so we engage RMT again to obtain the distribution of our test statistic under the alternative and to derive the

[1] In fact, when the number of assets exceeds the number of time periods, the bias is there even in the limit of vanishing regularization. Basically, in the over-parametrized regime, the test statistic behaves as if there is a regularization; this is the so-called "implicit regularization of machine learning". See, Didisheim et al. (2024).

test's power. Next, we derive the optimal regularization parameter, which is used to estimate the covariance matrix of test assets to maximize the test's power. Lastly, in the spirit of the Hansen and Richard (1987) prescription, we evaluate the case when the asymptotic ratio of test assets to sample size, c, approaches infinity. In this case, the power does not depend on the regularization parameter, and we can fully characterize the set of alternative hypotheses that allow attaining the full power.

Second, we assess the power of our test via a series of simulation exercises where we consider various sample sizes and alternatives. First, we verify the GRS result that when $c < 1$, the test's power depends on whether the true alphas are sparse. If that is the case, power peaks at a few test assets and then drops monotonically to zero as the number of test assets increases. If alphas are not sparse, then power is inversely U-shaped: low with few and a lot of test assets. Furthermore, the sample size has to be above 300 to achieve good power properties of the test at the intermediate amounts of test assets. As we move to our test statistic and consider $c > 1$, the test's power exhibits a pattern similar to GRS under the sparse-alpha alternative. The key difference is that power does not drop to zero as $c \to 1$, but rather asymptotes to zero as $c \to \infty$. The situation changes dramatically when the alternative is not sparse. The power of the test continues to improve as c grows to infinity. We show this using the asymptotically derived power and considering finite samples via simulation. As we approach sufficiently large sample sizes, with $T > 300$, power is equal to one for all but very small values of c. This is happening without sacrificing the test's size (at the optimal value of the regularization parameter), as our simulations show.

The last component of our analysis is applying the test using actual data on asset returns. We rely on two different approaches to generate many test assets. The first one is based on 153 stock characteristics compiled by Jensen, Kelly and Pedersen (2023) (JKL). The second one is based on Didisheim, Ke, Kelly and Malamud (2024), who use a shallow neural network based on 25,000 random features to generate test portfolios. We consider three candidates

for the efficient portfolio: an unconditional combination of the Fama-French 5 factors (Fama and French, 2015) and momentum (FF6), the Brandt, Santa-Clara and Valkanov (2009) (BSV) characteristic-based approach to approximation of efficient portfolio weights, and the Didisheim, Ke, Kelly and Malamud (2024) (DKKM) complexity-based approximation of the efficient portfolio. The FF6 factors cannot price any set of test assets. When we consider pricing within a moving-window 240-month period, we frequently fail to reject BSV and DKKM factors when pricing JKL test assets. However, these factors are rejected in the full sample. BSV and DKKM cannot price the 25,000 "network" portfolios. We interpret this evidence as limits-to-learning in high-complexity environments.

Scores of empirical papers are dedicated to finding and testing an efficient portfolio. Our contribution is primarily theoretical and builds on the work of Gibbons, Ross and Shanken (1989) and Hansen and Richard (1987). While our analysis focuses on tradable factors, our test statistic can also be used for testing non-tradable ones. Extending our analysis to the case of omitted factors (such as the three-pass method of Giglio and Xiu, 2021) is an interesting direction for future research.

Like us, Pesaran and Yamagata (2023) are interested in extending GRS-type testing to many test assets. They rely on the covariance estimator via thresholding-based shrinkage of Fan, Liao and Mincheva (2011), which requires sparsity of the covariance matrix. Fan, Liao and Yao (2015) (FLY) propose a novel technique for boosting the power of high-dimensional alpha tests. Specifically, FLY develop two test statistics of zero alpha. The first one is a feasible Wald statistic in high dimensions developed under the assumption of the sparseness of the true covariance matrix (Feng, Lan, Liu and Ma, 2022 and Yu, Yao and Xue, 2024 introduce test statistics which enhance the power of the FLY test in sparse environments). The second FLY statistic is an alpha-thresholding test that achieves maximal power when alpha is sparse (only a small number of test assets have non-zero alpha). Kock

and Preinerstorfer (2019) analyze the limits of power enhancement associated with the FLY approach. Threshold-based tests lack rotation invariance.

Lan, Feng and Luo (2018) expand the applicability of the GRS test to a non-sparse covariance matrix by repeatedly projecting the set of test asset returns into lower-dimensional space where the GRS test is applicable and then aggregating the test statistics across the projections. This aggregation of randomly sampled low-rank projections leads to information loss associated with the low eigenvalues of the covariance matrix of residuals that are almost never sampled by the procedure. As a result, the asymptotic power of the test is independent of the distribution of eigenvalues of the covariance matrix and is inversely proportional to its trace. Our test optimally leverages the information about the joint distribution of eigenvalues and alphas across the various principal components, making our test more powerful. This power difference is particularly large when the low-variance components exhibit a significant alpha, as in the modern factor zoo frameworks of Dello-Preite, Uppal, Zaffaroni and Zviadadze (2022), Didisheim, Ke, Kelly and Malamud (2024), Giglio, Xiu and Zhang (2025), Lettau and Pelger (2020), among others.

Our results apply to both sparse-covariance-matrix environments and empirically relevant cases when the covariance matrix is dense. We are able to achieve a broader applicability of our results by considering the shrinkage of the covariance matrix and accounting for the associated biases via RMT. One of the implications of our approach is that it achieves maximal power under a broader set of alternatives as compared to the FLY feasible Wald statistic. However, our test loses its power when alpha is sparse: we show theoretically that the test benefits from the complexity (i.e., the power increases with the number P of test assets) only when the amount of non-zero alphas grows not too slowly with P. To re-iterate, this weakness is not necessarily an issue in practice as alphas of test portfolios do not inherit the possible sparseness of alphas of individual assets.

Carrasco and Nokho (2022) extend the Hansen and Jagannathan (1991) distance by

incorporating ridge and Tikhonov regularizations to stabilize the inversion of nearly singular return covariance matrices when many test assets are used. Gagliardini, Ossola and Scaillet (2016) also study the problem of estimating and testing asset pricing models in the $P, T \to \infty$ limit and develop a novel approach to studying conditional asset pricing models with multiple factors, instruments, and with both conditional betas and risk premia being linear in these instruments. As a part of their model validation, they introduce statistics for testing alpha in a fully specified conditional asset pricing model where a candidate conditionally efficient portfolio is estimated based on the observed instruments. Ma, Lan, Su and Tsai (2020) study testing of conditional beta models in high dimensions as well. By contrast, our methodology, as well as the GRS or the FLY tests, evaluates a candidate unconditionally efficient portfolio, which does not require conditional betas under the null hypothesis.

Finally, we note that, just like GRS, our test statistic is designed to test a single hypothesis of the absence of alpha. Extending our methods to address a much more complex problem of multiple hypothesis testing, such as the recent paper by Giglio, Liao and Xiu (2021), is an important direction for future research.

2 The Test Statistic

2.1 The Setup

We consider an asset-pricing environment with P asset *excess* returns $R_t = (R_{i,t})_{i=1}^P$ (it could be actual stocks or managed portfolios, or single stock returns interacted with some instruments). We also consider a portfolio with excess return R_t^M. Consider the following regression:

$$R_{i,t} = \alpha_i + \beta_i R_t^M + \varepsilon_{i,t}, \ \ i = 1, ..., P, \ \ t = 1, \cdots, T\,, \tag{1}$$

or, in vector form,

$$\underbrace{R_t}_{P\times 1} \;=\; \underbrace{\alpha}_{P\times 1} \;+\; \underbrace{\beta}_{P\times 1} R_t^M \;+\; \underbrace{\varepsilon_t}_{P\times 1}, \varepsilon_t = \Sigma_P^{1/2} u_t \;,\; t = 1, \cdots, T\,, \tag{2}$$

where $\Sigma_P = E[\varepsilon_t \varepsilon_t']$ is the covariance matrix of residuals, and u_t are the (uncorrelated) innovations. If the portfolio is unconditionally mean-variance efficient, that is, its return R_t^M maximizes both unconditional and conditional Sharpe ratios, then it satisfies both unconditional and conditional linear beta pricing models:

Theorem 1 *Let $\{\mathcal{F}_t\}$ be the filtration describing investors' information set, and $E_t[\cdot] = E[\cdot|\mathcal{F}_t]$ the corresponding conditional expectation. Then, the following are equivalent:*

- *R_{t+1}^M is the return on the conditionally efficient portfolio:*

$$R_{t+1}^M \;=\; a_t^{-1} \pi_t' R_{t+1},\; \pi_t \;=\; \mathrm{Cov}_t(R_{t+1})^{-1} E_t[R_{t+1}] \tag{3}$$

  for some $a_t \in \mathbb{R}$.

- *we have*

$$E_t[R_{i,t+1}] \;=\; \beta_{i,t} E_t[R_{t+1}^M] \tag{4}$$

  for all i, where $\beta_{i,t} = \frac{\mathrm{Cov}_t(R_{i,t+1}, R_{t+1}^M)}{\mathrm{Var}_t[R_{t+1}^M]}$.

- *Furthermore, R_t^M satisfies (3) with $a_t = 1 + E_t[R_{t+1}]' \mathrm{Cov}_t(R_{t+1})^{-1} E_t[R_{t+1}]$ if and only if it prices returns on any feasible portfolio unconditionally:*

$$E[R_{i,t}] \;=\; \beta_i E[R_t^M], \tag{5}$$

  where $\beta_i \;=\; \frac{\mathrm{Cov}(R_{i,t}, R_t^M)}{\mathrm{Var}[R_t^M]}$, is the unconditional beta.

See Appendix A.

Theorem 1 is a reformulation of a classic result from Hansen and Richard (1987). Namely, one can evaluate all the conditional implications of the equation (4) by testing the equation (5) using all admissible trading strategies as test assets. This result motivates us to consider the regression (2) as $P \to \infty$. Under the null hypothesis of correct unconditional pricing, $\alpha_i = 0$ for all i.

FLY develop powerful techniques for testing asset pricing models as $P \to \infty$ when the covariance matrix of test assets and/or the α vector is sparse. In Appendix B, we provide a simple argument that such sparsity is not rotation-invariant: even if individual stocks exhibit sparse covariance matrices or sparse α, portfolios constructed from a large number of nonlinear characteristics lose sparsity in both dimensions. This observation motivates us to seek an alternative testing approach.

We make the following standard technical assumptions to establish the inference results.

Assumption 1 *Returns satisfy (2) with some vector $\alpha \in \mathbb{R}^P$. Furthermore,*

- *The innovation u_t in (2) consists of i.i.d. elements, $u_t = (u_{i,t})$, and $E[u_{i,t}] = 0$, $E[u_{i,t}^2] = 1$, $E[u_{i,t}^4] < \infty$.*
- *We have $P/T \to c$ and $T^{1/2}(P/T - c) \to 0$ as $P \to \infty$.*
- *$\|\Sigma_P\| \leq K$ for some $K > 0$ as $P \to \infty$, and the eigenvalue distribution of Σ_P is asymptotically stable (see Definition 1 in Appendix C).*
- *The variance of R_t^M is strictly positive and uniformly bounded.*

Note that we do not need to assume Gaussianity and only require four finite moments for innovations. The assumption of asymptotic stability is just a technical condition allowing us to rigorously study the limit as $P \to \infty$. It is important to note that the boundedness condition for the residual covariance matrix does not contradict the classic arbitrage pricing

theory (APT) assumptions. Indeed, Chamberlain and Rothschild (1982) show that APT implies that several largest eigenvalues of the *test-asset covariance matrix* explode when $P \to \infty$. However, once these top principal components are optimally combined into an SDF, the residuals have a bounded covariance matrix. Empirically, in all of our experiments with managed portfolios, we find that the residual covariance matrix is always bounded.

Our goal is to test the hypothesis $\alpha = 0$ in the case when P is large. To this end, we estimate the β vector,

$$\hat{\beta} \;=\; \frac{\widehat{\mathrm{Cov}}(R, R^M)}{\hat{\sigma}^2_M} \;\in\; \mathbb{R}^P, \tag{6}$$

where

$$\begin{aligned}
\bar{R} \;&=\; \frac{1}{T}\sum_{t=1}^{T} R_t \\
\bar{R}^M \;&=\; \frac{1}{T}\sum_{t=1}^{T} R^M_t \\
\widehat{\mathrm{Cov}}(R, R^M) \;&=\; \frac{1}{T}\sum_{t=1}^{T} R_t R^M_t \;-\; \bar{R}\,\bar{R}^M \\
\hat{\sigma}^2_M \;&=\; \widehat{\mathrm{Var}}[R^M] \;=\; \frac{1}{T}\sum_{t=1}^{T} (R^M_t)^2 \;-\; (\bar{R}^M)^2\,.
\end{aligned} \tag{7}$$

We then define the estimated residuals,

$$\hat{\alpha}_t \;=\; R_t \;-\; R^M_t \hat{\beta}, \tag{8}$$

as well as their sample mean vector

$$\bar{\alpha} \;=\; \frac{1}{T}\sum_{t=1}^{T} \hat{\alpha}_t \tag{9}$$

and their covariance

$$\hat{\Sigma}_P \;=\; \frac{1}{T}\sum_{t=1}^{T}\hat{\alpha}_t\hat{\alpha}_t' \;-\; \bar{\alpha}\bar{\alpha}'\,. \tag{10}$$

2.2 The GRS test statistic

The classic Gibbons, Ross and Shanken (1989) test statistic is given by

$$\hat{W} \;=\; \frac{\hat{\theta}_\alpha^2}{1 \;+\; \hat{\theta}_M^2}\,, \tag{11}$$

where

$$\begin{aligned}\hat{\theta}_\alpha^2 \;&=\; \bar{\alpha}'\,\hat{\Sigma}_P^{-1}\bar{\alpha}\\ \hat{\theta}_M^2 \;&=\; \frac{(\bar{R}^M)^2}{\hat{\sigma}_M^2}\end{aligned} \tag{12}$$

are the sample squared Sharpe ratios of alphas and factors, respectively, with the former often referred to as the information ratio. Note that the test statistic is only well-defined when $P < T$. When ε_t are normally distributed,

$$\frac{T(T-P-1)}{P(T-2)}\hat{W} \sim F_{P,T-P-1}(\lambda) \tag{13}$$

where $F_{P,T-P-1}(\lambda)$ is Fisher distribution with non-centrality parameter

$$\lambda = \frac{T}{1+\hat{\theta}_M^2}\alpha'\Sigma_P^{-1}\alpha. \tag{14}$$

Under the null hypothesis of $\alpha = 0$, we have $\lambda = 0$. Importantly, the test statistic is independent of Σ_P. This feature makes the test attractive. By the definition of the $F_{P,T-P-1}(0)$

distribution, (13) can be written as

$$\frac{T(T-P-1)}{P(T-2)}\hat{W} \;\sim\; \frac{P^{-1}\sum_{k=1}^{P} Z_k^2}{(T-P-1)^{-1}\sum_{k=1}^{T-P-1}\tilde{Z}_k^2}, \tag{15}$$

where $Z_k, \tilde{Z}_k$ are independent Gaussian variables. In the asymptotic case, $T\hat{W}$ is a version of the Hansen (1982) J-statistic, whose distribution is established on the basis of the Generalized Method of Moments (GMM) framework. In the "classic" asymptotic regime when $T \to \infty$ (keeping P fixed, i.e, $P/T \to 0$), (15) implies convergence to a χ^2 distribution,

$$J \;=\; T\hat{W} \;\sim P^{-1}\sum_{k=1}^{P} Z_k^2\,. \tag{16}$$

As we now show, this standard GMM logic breaks down when $P/T \to c > 0$ even if $c < 1$. Indeed, in the limit when $T, P \to \infty, P/T \to c < 1$, we have $\frac{T(T-P-1)}{P(T-2)} \to \frac{1-c}{c}$, and the GRS statistic converges to a constant almost surely by the Law of Large Numbers, $\frac{1-c}{c}\hat{W} \to 1$. In particular, $J = T\hat{W}$ explodes as $T \to \infty$, making the GMM test statistic inapplicable. This phenomenon is known as "concentration of measure." The Central Limit Theorem implies that deviations of $\hat{W}$ from its limit are normally distributed, and we can "correct" the test statistic as follows.

Theorem 2 *In the limit as $P, T \to \infty, P/T \to c \in (0,1)$, $\hat{W}$ converges in distribution to a normal random variable:*

$$\frac{T^{1/2}\left(\hat{W} - \xi(c)\right)}{\Delta(c)^{1/2}} \xrightarrow{d} N(0,1)\,,$$

where we have defined

$$\begin{aligned} \xi(c) \;&=\; \frac{c}{1-c}, \\ \Delta(c) \;&=\; \frac{2c}{(1-c)^3}\,. \end{aligned} \tag{17}$$

See Appendix A.

The goal of this paper is to show how these arguments can be extended to the case $P > T$. A key implication of Theorem 2 is that the variance of $\hat{W}$ explodes very quickly, as $1/(1-c)^3$, when $c \uparrow 1$. As a result, the power of the test vanishes in this limit.

This inconvenient behavior can be corrected through regularization. One approach is explored by Fan, Liao and Yao (2015) in the case of a sparse covariance matrix of returns. They essentially shrink the off-diagonal elements of the covariance matrix towards zero, either entirely (keeping only the diagonal of the matrix) or partially (using a thresholding-based regularization). The proposed technique allows the authors to control the asymptotic behavior of the pivotal Wald statistic when $P, T \to \infty$.

We present another approach where the covariance matrix regularization shrinks the matrix towards zI (ridge-regularization). This allows us to avoid the sparsity requirement but introduces additional challenges in the form of asymptotic biases in the test statistic. We address these challenges in what follows.

2.3 High-Dimensional Case

Our goal is to extend (11) to the case $P \geq T$ and eliminate the singular behavior for $P < T$ as $P \uparrow T$. Such an extension is by no means straightforward. Indeed, in the case when $P \geq T$, the matrix $\hat{\Sigma}_P$ in (11) is degenerate, and $\hat{W}$ is almost surely infinite. By Theorem 2, $\hat{W}$ explodes when $c = P/T \uparrow 1$. We introduce a ridge-regularized version of the statistic:

$$\begin{aligned} \hat{\theta}^2_\alpha(z) &= \bar{\alpha}'(zI + \hat{\Sigma}_P)^{-1}\bar{\alpha} \\ \hat{W}(z) &= \frac{\hat{\theta}^2_\alpha(z)}{1 + \hat{\theta}^2_M}, \end{aligned} \tag{18}$$

where $z > 0$ is a ridge penalty. $\hat{W}(z)$ is well defined for both $P < T$ and $P \geq T$. As we now show, in complete analogy with Theorem 2, $\hat{W}(z)$ converges to a constant for any $z > 0$,

any $c > 0$, and any ε_t satisfying Assumption 1.[2] Such *universality* is a striking property of our test statistic and relies on deep results from the RMT that deal with high-dimensional random matrices such as $\hat{\Sigma}_P$ (see Appendix C).

Proposition 3 *For any $z > 0$, in the limit as $T, P \to \infty, P/T \to c$, for any $c > 0$, we have*

$$\hat{W}(z) \;-\; \tilde{\xi}(z; \Sigma_P; c) \;\to\; 0 \tag{19}$$

in probability, where the constant $\tilde{\xi}(z; \Sigma_P; c)$ is given by

$$\tilde{\xi}(z; \Sigma_P; c) \;=\; \lim_{T,P\to\infty, P/T\to c} T^{-1} E[\mathrm{tr}((zI + \hat{\Sigma}_P)^{-1}\Sigma_P)]\,. \tag{20}$$

To understand why the limit (20) is non-zero, note that $A_P \equiv (zI + \hat{\Sigma}_P)^{-1}\Sigma_P$ is a bounded $P \times P$ matrix, with positive, bounded eigenvalues $\lambda_i(A_P)$, $\varepsilon_1 \leq \lambda_i(A_P) \leq \varepsilon_2$, for some $\varepsilon_1, \varepsilon_2 > 0$,[3] so that

$$\mathrm{tr}(A_P) \;=\; \lambda_1(A_P) \;+\; \cdots \;+ \lambda_P(A_P)$$

satisfies $P\varepsilon_1 \;\leq\; \mathrm{tr}(A_P) \;\leq\; P\varepsilon_2$. Thus, $\tilde{\xi}(z; \Sigma_P; c) \in [\varepsilon_1, \varepsilon_2]$.

The key difference from the asymptotic GRS result — when $c < 1$ — is that now the convergence constant depends on the unobservable Σ_P. This is a crucial difference that seemingly prevents one from developing a test for the case of a high-dimensional cross-section of test assets. RMT allows us to express $\tilde{\xi}(z; \Sigma_P; c)$ in a way that does not depend on Σ_P.

[2] The constant coincides with the one in Theorem 2 for $c < 1$ and $z = 0$.

[3] In the limit as $P, T \to \infty$, $\hat{\Sigma}_P$ is almost surely bounded when P/T is bounded. See, e.g., Bai and Yin (2008).

Let

$$\hat{m}(-z) \;=\; P^{-1}\,\mathrm{tr}((zI+\hat{\Sigma}_P)^{-1}) \tag{21}$$

be the *empirical Stieltjes transform* of the sample covariance matrix. A classical result of RMT implies that, although $\hat{\Sigma}_P$ is random, $\hat{m}(-z)$ converges to a deterministic limit. See Appendix C for details. This property implies the following result.

Proposition 4 *Let*

$$\hat{\xi}(z;c) \;=\; (1-c+cz\hat{m}(-z))^{-1} \;-\; 1 \tag{22}$$

and let

$$\hat{\xi}'(z;c) \;=\; \left(\frac{1}{1-c(1-z\hat{m}(-z))}\right)^2 c(z\hat{m}'(-z)-\hat{m}(-z)) \tag{23}$$

be its derivative. For any $z>0$, in the limit as $T,P\to\infty, P/T\to c$, we have that $\hat{\xi}(z;c)$ satisfies

$$\begin{aligned}\tilde{\xi}(z;\Sigma_P;c) \;-\; \hat{\xi}(z;c) \;&\to\; 0\\ \tilde{\xi}'(z;\Sigma_P;c) \;-\; \hat{\xi}'(z;c) \;&\to\; 0\end{aligned} \tag{24}$$

in probability.

A key implication of Proposition 4 is that we can compute the demeaned GRS statistic, $\hat{W}(z)-\hat{\xi}(z;c)$, without knowing Σ_P. It is also possible to show that $\hat{W}(z)-\hat{\xi}(z;c)$ converges to zero at the rate $T^{-1/2}$. While the proof is non-trivial and requires tools from RMT, it can be conjectured on the basis of the Gaussian case with $z=0$ in (16) — the convergence rate

in the law of large numbers is $T^{-1/2}$. Thus, we should work with $T^{1/2}(\hat{W}(z) - \hat{\xi}(z;c))$ (and when $c < 1$ and $z = 0$, the same applies to the GRS statistic).

One could conjecture that after suitable standardization, this quantity should converge to standard normal. Proving this is non-trivial because it requires developing a Central Limit Theorem in the $P, T \to \infty$ limit. We apply the powerful results from Li, Aue and Paul (2020a) to show in Appendix D that the following holds.

Theorem 5 *Under the null hypothesis $H_0 : \alpha = 0$, we have that, for any $z > 0$ and any $c > 0$,*

$$\frac{T^{1/2}(\hat{W}(z) - \hat{\xi}(z;c))}{\hat{\Delta}(z;c)^{1/2}} \;\to\; N(0,1) \tag{25}$$

in distribution, where

$$\hat{\Delta}(z;c) \;=\; 2(\hat{\xi}(z;c) + z\hat{\xi}'(z;c))(1 + \hat{\xi}(z;c))^2\,. \tag{26}$$

Recall that, according to Assumption 1, this result does not require the normality of returns; it only requires that their first four moments are finite.

The complex expressions for our test statistic arise because, to make it pivotal, we have to address the fact that the inverse of the sample covariance matrix is a severely biased estimate of the inverse of the true covariance matrix. As we show in Appendix C (formula (C.15)), there exists an implicit shrinkage parameter $Z_* > z$ such that

$$\lim_{P,T\to\infty, P/T\to c} (zI + \underbrace{\hat{\Sigma}_P}_{random})^{-1} \;=\; \lim_{P\to\infty} \frac{Z_*}{z}(Z_* I + \underbrace{\Sigma_P}_{deterministic})^{-1}\,. \tag{27}$$

A similar bias is present in our test statistic:

$$\begin{aligned} \lim_{P,T\to\infty,P/T\to c} \underbrace{\bar{\alpha}'(zI+\hat{\Sigma}_P)^{-1}\bar{\alpha}}_{test\ with\ sample\ \hat{\Sigma}_P} &= \lim_{P,T\to\infty,P/T\to c} \frac{Z_*}{z}\bar{\alpha}'(Z_*I+\Sigma_P)^{-1}\bar{\alpha} \\ > \lim_{P,T\to\infty,P/T\to c} &\underbrace{\bar{\alpha}'(zI+\Sigma_P)^{-1}\bar{\alpha}}_{test\ with\ true\ \Sigma_P} \end{aligned} \tag{28}$$

Thus, this bias creates a gap between the pivotal and the "classic" test statistics, and the size of this bias is controlled by $Z_*/z > 1$. In our data experiments, we find that Z_*/z is often very large, and hence, the bias is significant. Hence, addressing the randomness of $\hat{\Sigma}_P$ is very important. Note also that the presence of this bias serves as another piece of evidence against the sparsity of Σ_P.

2.4 An Economic Interpretation of the Test Statistic

We start with the important observation that $\hat{\alpha}$ are excess returns on tradable portfolios. We estimate the mean of $\hat{\alpha}$ by $\bar{\alpha}$ and their covariance matrix by $\hat{\Sigma}_P(z) = \hat{\Sigma}_P + zI$. Such an estimate could be justified by shrinkage using both the frequentist and Bayesian approaches. In the former case, z could be viewed as shrinkage intensity (e.g., Ledoit and Wolf, 2004). In the latter case, z could be related to the prior about $\hat{\alpha}$'s mean (e.g., Kozak, Nagel and Santosh, 2020).

Given these estimates, the squared Sharpe ratio of $\hat{\alpha}$ is given by the classic identity

$$\hat{\theta}^2_\alpha(z) = \bar{\alpha}'\,\hat{\Sigma}_P(z)^{-1}\bar{\alpha}\,. \tag{29}$$

If $E[\varepsilon|R^M] = 0$, then $\hat{\alpha}$ and R^M are uncorrelated. Then the joint covariance matrix of $(\hat{\alpha}', R^M)'$ is block diagonal and, hence, the total Sharpe ratio of the tangency portfolio

combining $\hat{\alpha}$ and R^M is given by

$$\hat{\theta}^2_*(z) = \hat{\theta}^2_\alpha(z) + \hat{\theta}^2_M \,. \tag{30}$$

Thus, our test statistic can be represented as

$$\hat{W}(z) \;=\; \frac{\hat{\theta}^2_\alpha(z)}{1+\hat{\theta}^2_M} \;=\; \frac{\hat{\theta}^2_*(z)-\hat{\theta}^2_M}{1+\hat{\theta}^2_M} \tag{31}$$

and, therefore, be interpreted as the distance between the Sharpe ratios squared of test assets and the candidate factor combined together and the Sharpe ratio of the candidate efficient portfolio alone. This interpretation is consistent with that of the GRS statistic.

3 Power of the Test

3.1 Asymptotics Under a Local Alternative

When α in (2) is non-zero, the estimated residuals admit the representation

$$\begin{aligned} \hat{\alpha}_t \;&=\; \alpha \;+\; \hat{\varepsilon}_t, \\ \hat{\varepsilon}_t \;&=\; (\beta-\hat{\beta})R^M_t \;+\; \varepsilon_t \,, \end{aligned} \tag{32}$$

so that the average residuals are given by

$$\bar{\alpha} \;=\; \alpha \;+\; \bar{\varepsilon},\; \bar{\varepsilon} \;=\; \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t, \tag{33}$$

and the test statistic (18) can be decomposed as

$$\begin{aligned}\hat{W}(z) &= \frac{\bar{\alpha}'\,\hat{\Sigma}_P(z)^{-1}\bar{\alpha}}{1+\hat{\theta}_M^2} \\ &= (1+\hat{\theta}_M^2)^{-1}\left(\underbrace{\alpha'\,\hat{\Sigma}_P(z)^{-1}\,\alpha}_{true\ alpha} + \underbrace{\bar{\varepsilon}'\,\hat{\Sigma}_P(z)^{-1}\bar{\varepsilon} + 2\bar{\varepsilon}'\,\hat{\Sigma}_P(z)^{-1}\,\alpha}_{noise}\right).\end{aligned} \tag{34}$$

Since $E[\bar{\varepsilon}] = 0$, it is possible to show that the interaction term, $\bar{\varepsilon}'\,\hat{\Sigma}_P(z)^{-1}\,\alpha$, is negligible for large T (see Li, Aue and Paul, 2020a for a detailed argument, and Proposition H.1 in Appendix H for a simple proof in the Gaussian case). Further, $\bar{\varepsilon}'\,\hat{\Sigma}_P(z)^{-1}\bar{\varepsilon}$ is the $\hat{W}(z)$ test statistic under null. As a result, a small modification of Theorem 5 implies that $\hat{W}(z)$ is still asymptotically normal but with a non-zero mean. By (25), we have, in the sense of convergence in distribution, that

$$\lim_{P,T\to\infty,P/T\to c}\frac{T^{1/2}(\hat{W}(z)-\hat{\xi}(z))}{\hat{\Delta}(z;c)^{1/2}} \sim N\left(\lim_{P,T\to\infty,P/T\to c}\frac{T^{1/2}\,\alpha'\,\hat{\Sigma}_P(z)^{-1}\,\alpha}{\hat{\Delta}(z;c)^{1/2}(1+\hat{\theta}_M^2)},\ 1\right). \tag{35}$$

If $\alpha'\,\hat{\Sigma}_P(z)^{-1}\,\alpha >> T^{-1/2}$, then α is easily detectable because $T^{1/2}\alpha'\,\hat{\Sigma}_P(z)^{-1}\,\alpha$ explodes for large T. If $\alpha'\,\hat{\Sigma}_P(z)^{-1}\,\alpha << T^{-1/2}$, the alpha is not detectable — our test is not powerful enough. These observations prompt us to focus on the boundary case when $\alpha'\,\hat{\Sigma}_P(z)^{-1}\,\alpha = O(T^{-1/2})$.

Intuitively, the difficulty in detecting alpha is determined by the way it is distributed among the test assets. We model this distribution by assuming that α itself is randomly sampled from a distribution so that $E[\alpha_i\alpha_j] = T^{-1/2}P^{-1}\,\Omega_\alpha(i,j)$ for some matrix Ω_α, so that $E[\alpha'\,\hat{\Sigma}_P(z)^{-1}\,\alpha] = T^{-1/2}P^{-1}\,\mathrm{tr}\,E[\hat{\Sigma}_P(z)^{-1}\,\Omega_\alpha] = O(T^{-1/2})$. We formalize this with the following assumption:

Assumption 2 *$\alpha = T^{-1/4}P^{-1/2}\Omega_\alpha^{1/2}\gamma$, where $\gamma = (\gamma_i)_{i=1}^P$ with γ_i being i.i.d., with $E[\gamma_i] =$*

$0,\ E[\gamma_i^2] = 1,\ E[\gamma_i^4] < \infty$. *Furthermore, the limit*[4]

$$m_{\Sigma_P}(-z;\Omega_\alpha) \;=\; \lim_{P\to\infty} T^{1/2}\alpha'(zI+\Sigma_P)^{-1}\alpha \;=\; \lim_{P\to\infty} P^{-1}\,\mathrm{tr}(\Omega_\alpha(zI+\Sigma_P)^{-1})\,. \tag{36}$$

exists.

Under Assumption 2, the asymptotic power of the test satisfies

$$\begin{aligned}
&\lim_{P,T\to\infty,P/T\to c} \text{Probability of rejection} \\
&= \lim_{P,T\to\infty,P/T\to c} \text{Prob}\left(\frac{T^{1/2}(\hat{W}(z)-\hat{\xi}(z))}{\hat{\Delta}(z;c)^{1/2}} > \zeta\right) \\
&= \Phi\left(-\zeta + \lim_{P,T\to\infty,P/T\to c} \frac{T^{1/2}\,\alpha'\,\hat{\Sigma}_P(z)^{-1}\,\alpha}{\hat{\Delta}(z;c)^{1/2}(1+\hat{\theta}_M^2)}\right),
\end{aligned} \tag{37}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard Normal distribution, and ζ is a quantile of this distribution. The rest of this section is devoted to computing the limit in (37) and studying its properties that are established using RMT. Readers who are not interested in technical details can skip it and proceed directly to the simulation and empirical results in Sections 4 and 5.

3.2 Asympotic Power

Using power concentration results from Random Matrix Theory (see Appendix C), it is possible to prove the following result.

Theorem 6 *The following limit exists in probability:*

$$\mathcal{H}(z;c) \;=\; \lim_{T,P\to\infty,P/T\to c} \frac{T^{1/2}\alpha'(zI+\hat{\Sigma}_P)^{-1}\,\alpha}{1+\hat{\theta}_M^2}\,. \tag{38}$$

[4]The equality of the last two limits follows from Lemma C.1 in the Appendix.

Theorem 6 implies that the true alpha component $T^{1/2}\alpha'(zI + \hat{\Sigma}_P)^{-1}\,\alpha$ in the test statistic (34) converges to a constant when P, T are large. This constant is independent of the sample and depends only on c, Σ_P, and Ω_α. Appendix C / Theorem C.3 derives this constant explicitly. As a result, the following modification of Theorem 5 can be established (see Li, Aue and Paul, 2020a for details).

Theorem 7 *Under Assumption 2, for any $z > 0$, we have*

$$\frac{T^{1/2}(\hat{W}(z) - \hat{\xi}(z))}{\hat{\Delta}(z;c)^{1/2}} \;\to\; N\left(\frac{\mathcal{H}(z;c)}{\Delta(z;c)^{1/2}},\; 1\right) \tag{39}$$

in distribution, where $\mathcal{H}(z;c)$ is defined in (38). In particular, the power of the test satisfies

$$\begin{aligned} \textit{Probability of rejection} = \;& \text{Prob}\left(\frac{T^{1/2}(\hat{W}(z) - \hat{\xi}(z))}{\hat{\Delta}(z;c)^{1/2}} > \zeta\right) \\ \to \Gamma_{asym}(z;c) \;=\; & \Phi\left(-\zeta + \frac{\mathcal{H}(z;c)}{\Delta(z;c)^{1/2}}\right), \end{aligned} \tag{40}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard Normal distribution, and ζ is a quantile of this distribution.

The quantity $\mathcal{H}(z;c)/\Delta(z;c)^{1/2}$ defines the asymptotic power of the test. The larger the quantity, the more powerful the test is. In Appendix H, Proposition H.2, we derive the analog of Theorem 7 for the case of a known covariance matrix. We show that the behavior of the test statistic is very different in this case.. For example, the test statistic is well defined even for $z \to 0$.

The next proposition shows $\mathcal{H}(z;c)$ admits an intuitive economic interpretation. It measures the expected out-of-sample performance of the efficient alpha portfolio.

Proposition 8 (Test Power and the OOS Markowitz Performance) *Out-of-sample ex-*

pected return on the portfolio $(zI + \hat{\Sigma}_P)^{-1}\bar{\alpha}$ *satisfies*

$$\lim_{T,P\to\infty,P/T\to c} T^{1/2} E_T[\bar{\alpha}'(zI + \hat{\Sigma}_P)^{-1}\hat{\alpha}_{T+1}] \;=\; \mathcal{H}(z;c)\,, \tag{41}$$

where convergence is in probability.

In our simulations, we assume for simplicity that $\Omega_\alpha = aI + b\Sigma_P$ for some coefficients a, b. When $b > 0$, high-variance principal components have high alpha, and the magnitude of b defines the strength of the risk-return tradeoff. In this case, the limit (38) can be computed analytically because of the following lemma.

Lemma 1 *Suppose that* $\Omega_\alpha = aI + b\Sigma_P$. *Let* $m(-z;c) = \lim_{P,T\to\infty,P/T\to c} \hat{m}(-z)$; $\xi(z;c) = \lim_{P,T\to\infty,P/T\to c} \hat{\xi}(z;c)$. *Then, these limits are non-random and we have*

$$\mathcal{H}(z;c) \;=\; \frac{am(-z;c) + c^{-1}b\xi(z;c)}{1+\hat{\theta}_M^2}\,. \tag{42}$$

In Appendix E, we show how to compute $m(-z;c)$ and $\xi(z;c)$ analytically by solving a fixed-point equation.

3.3 Optimal Ridge Penalty

Theoretically, we would like to solve the problem

$$\max_z \frac{\mathcal{H}(z;c)}{\Delta(z;c)^{1/2}}\,. \tag{43}$$

While $\Delta(z;c) = \lim_{P,T\to\infty,P/T\to c} \hat{\Delta}(z;c)$ we need a way to estimate $\mathcal{H}(z;c)$. Here, we will use the calculations based on the leave-one-out logic introduced in Kelly, Malamud, Pourmohammadi and Trojani (2024b). In the Appendix F, we report a detailed (but somewhat tedious) calculation that leads us to the following algorithm for computing the candidate optimal ridge penalty.

- compute $A = T^{-1}\hat{\alpha}'\hat{\alpha} \in \mathbb{R}^{T\times T}$. Note that, by the properties of the trace, $\text{tr}(A) = \text{tr}(T^{-1}\hat{\alpha}\hat{\alpha}') = \text{tr}(\hat{\Sigma}_P)$.

- Since the impact of z on $(zI + \hat{\Sigma}_P)^{-1}$ depends on the size of z relative to $\hat{\Sigma}_P$, we optimize over z using a grid that scales with $P^{-1}\,\text{tr}(A)$. Namely, we let z takes values $z = \tilde{z}\cdot P^{-1}\,\text{tr}(A),\ \tilde{z} \in grid = [10^{-5}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 20, 10^2]$.

- compute $G(z) = (zI + A)^{-1}A$ and define $w = (w_t)_{t=1}^T = \text{diag}(G(z))$, $\bar{W} = (\bar{W}_t)_{t=1}^T = G(z)\mathbf{1}$ and

$$\tilde{\mathcal{H}}(z;c) \;=\; \frac{1}{1 - T^{-1}\mathbf{1}'\bar{W}} \sum_{t=1}^{T} \frac{\bar{W}_t - w_t}{1 - w_t} \tag{44}$$

  As we show in the Appendix F, $\tilde{\mathcal{H}}(z;c)$ is an unbiased estimator of $\mathcal{H}(z;c)$.

- Define

$$\Pi(z) \;=\; \frac{\tilde{\mathcal{H}}(z;c)}{\hat{\Delta}(z;c)^{1/2}} \tag{45}$$

  and find the optimal ridge penalty, solving

$$\tilde{z}_*(c) \;=\; \arg\max_{\tilde{z}\in grid} \Pi(\tilde{z}P^{-1}\,\text{tr}(A))\,. \tag{46}$$

So, the optimal ridge penalty amounts to scaling $P^{-1}\,\text{tr}(T^{-1}\hat{\alpha}'\hat{\alpha})$ by the optimal $\tilde{z}$.

3.4 Infinite Complexity

In the setting of Theorem 1, testing the conditional efficiency of a given candidate portfolio R^M formally requires considering an infinite number of test asset returns, given by interactions of security returns with all possible instruments. It is thus important to investigate the behavior of our test in the limit of infinite complexity, as $c \to \infty$. As we show in Appendix

G, this behavior is controlled by the speed with which the size of alpha, $\|\alpha\|^2$, grows with P. This growth of $\|\alpha\|^2$ controls how much additional alpha we get as we keep adding new test assets. The following result holds.

Proposition 9 *Suppose that* $\|\alpha\|^2 = \eta T^{-1} P^{1/2}$ *for some* $\eta > 0$. *Then, we have*

$$\Gamma_{approx} \;=\; \lim_{c\to\infty} \Gamma_{asym}(z;c) \;=\; \Phi\left(-\zeta \;+\; \eta \frac{1}{1+\hat{\theta}_M^2}\kappa,\right). \tag{47}$$

where

$$\kappa \;=\; \frac{1}{(2\lim_{P\to\infty} P^{-1}\operatorname{tr}(\Sigma_P^2))^{1/2}} \tag{48}$$

If $\|\alpha\|^2 = \eta T^{-1} P^{\delta}$, $\delta > 0.5$, *then the test power exhibits the virtue of complexity, so that* $\Gamma_{approx} = 1$.

Strikingly, the asymptotic power of the test for high complexity is independent of the ridge penalty. Furthermore, the power of the test does not vanish in the infinite complexity limit as long as $\|\alpha\|^2$ grows like $P^{1/2}$ (or faster). This is in stark contrast to the behavior of the GRS statistic, for which the impact of complexity on the power of the test can be highly detrimental for $c \to 1$ because the variance of $\hat{W}(0)$ explodes in Theorem 2. By contrast, a combination of ridge regularization with $c > 1$ makes complexity a virtue for our test as long as alpha grows sufficiently fast with P. This is a very strong property of our test. For example, the sparse-covariance-matrix-based Wald test statistics J_{wald} and J_{thr} of FLY attain full power when

$$\|\alpha\|^2 \;> \frac{P}{T}\left(\frac{\log\log T}{\sqrt{\log P}}\right)^2. \tag{49}$$

Thus, our test exhibits a much stronger virtue of complexity in the case when α is not sparse.

By contrast, their α-sparsity-based test statistic J_0 attains full power even when $\|\alpha\|^2$ grows much more slowly, as $\log P$.

4 Analysis of power of the portfolio efficiency test

4.1 The power of the GRS test

We start this section by reproducing the GRS analysis. As we explain above, when returns are normally distributed and $P < T$, the GRS test statistic has non-central Fisher distribution as summarized in Equations (13) and (14). As a result, the power of the GRS test is given by a closed-form monotone transformation of the non-centrality parameter — the larger λ, the easier it is to reject the null of no alpha. A key question studied by GRS is the dependence of the test power (that is, the λ-parameter) on the number of test assets, P. This dependence is naturally linked to the scarcity of α among test assets. For example, if only one asset k has alpha, we have $\alpha_k \neq 0$ and $\alpha_i = 0, \forall i \neq k$, so that

$$\lambda \;=\; \frac{T}{1+\hat{\theta}_M^2}\alpha_k^2(\Sigma^{-1})_{k,k}. \tag{50}$$

That is λ scales linearly with T and does not grow with P. More generally, the same is true when only a few assets $K < P$ are mispriced so that K/P is negligible. By contrast, if α_i are sampled randomly from $N(0,1)$ and $\Sigma_P = I$, we have by the law of large numbers that

$$\lambda \;=\; \frac{T}{1+\hat{\theta}_M^2} P \frac{1}{P}\sum_{i=1}^{P}\alpha_i^2\,. \tag{51}$$

Thus, λ increases linearly with PT, and, in principle, choosing a larger P should make it easier to reject the null. As we now show, this intuition is incorrect. Figure 1 reproduces the experiment from GRS and depicts the power of the GRS as a function of $c = P/T$ for different sample sizes T, for $\lambda = 0.1T$ (sparse α) and $\lambda = \bar{\lambda}PT$, with $\bar{\lambda} \in \{0.002, 0.0002\}$.

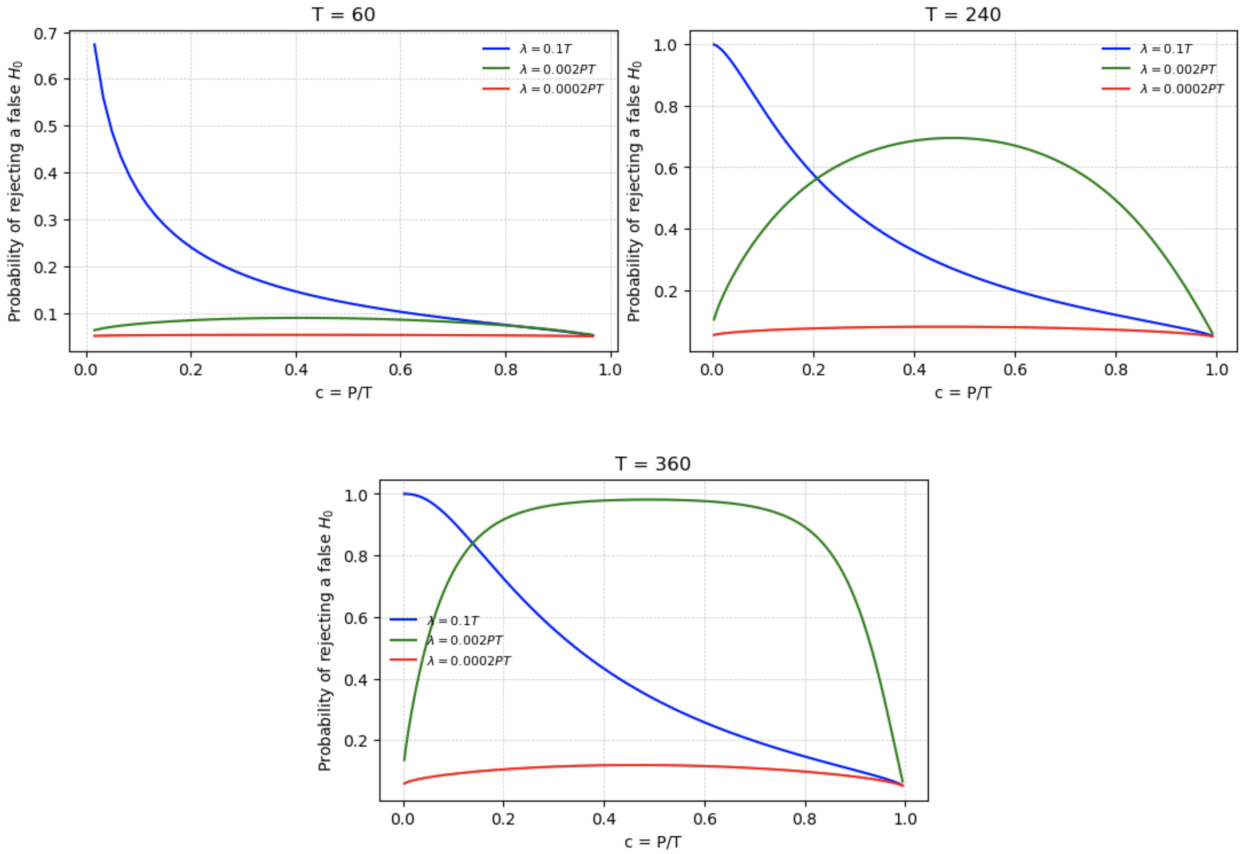


Figure 1: GRS test's power under various hypotheses for the true λ from (14). The sample size is fixed to $T = 60, 240, 360$, and we let P vary between 1 and $T - 2$.

The first key observation is that when α is sparse (so that λ does not increase with P), the power of the GRS test dramatically decreases with P. It is extremely difficult to detect scarce α in a large cross-section. By contrast, when α is not sparse, and λ grows linearly with P, the power of the GRS test is hump-shaped with respect to $c = P/T$, achieving its maximum around $P/T = 0.5$. This behavior of the test statistic is closely related to the classic bias-variance tradeoff in statistical estimation: Initially, a larger P leads to a larger λ, facilitating α detection. However, as P increases further, the GRS test suffers from overfitting, which becomes particularly intuitive through its economic interpretation from Section 2.4. An investor computing the realized squared Sharpe ratio of residuals will always overfit when $P/T > 0$. As $P/T \to 1$, this investor can, in fact, achieve a perfect,

zero-risk portfolio with an infinite in-sample Sharpe ratio. Formally, both the test statistic and its standard deviation explode when $P \to T$. This extreme overfitting makes the GRS test uninformative, and its power vanishes as $P \to T$. To avoid this catastrophic overfitting, GRS recommend picking P with $P/T \in [1/2, 2/3]$.

Importantly, the test statistic is undefined for $P > T$, and an empiricist with a dataset of $P > T$ test assets needs to make discretionary choices about how to pick a subset of test assets with $P << T$. Below, we show how our test is able to overcome these problems through an efficient use of complexity $c = P/T$.

4.2 Finite-sample power of the test

Our objective is to assess our test's power in finite samples and compare it to the asymptotic approximation derived in Theorem 7 as well as to the power of the GRS test. To this end, we use a Monte-Carlo simulation to generate synthetic datasets, compute our test statistics, and assess the validity of the null hypothesis.

4.2.1 Simulation setup in the high-dimensional case

We generate synthetic data satisfying equation (2). We sample β, R^M, Σ_P as follows:

- $\beta = (\beta_i)_{i=1}^P,\ \beta_i \sim \mathcal{U}(-3, 3)$
- $R^M \sim \mathcal{N}(\mu_M, \sigma_M^2)$ with $\mu_M = 0.07$ and $\sigma_M = 0.14$, approximately matching mean and standard deviation of SP500 over the last 50 years.
- $u_t \sim \mathcal{N}(0, I_P)$

Next, we need to generate the covariance matrix Σ_P with a non-trivial eigenvalue distribution. We simulate $\Sigma_P = U \operatorname{diag}(1 + \Lambda)U'$, where

- $\Lambda \sim \mathcal{U}(0, 1) \in [0, 1]^P$, so that the eigenvalues of Σ_P are uniformly distributed on $[1, 2]$.

- $A_{i,j} \sim \mathcal{U}(0,1)$, $A = (A_{i,j}) \in [0,1]^{P \times P}$, and $A'A = UDU'$; that is U is computed as the matrix of eigenvectors of $A'A$.

To facilitate the comparison with the GRS test, we parametrize the simulations with the non-centrality parameter λ of the GRS test (see Equation (14)). Specifically, for a chosen $\tilde{\Omega}_\alpha$ matrix, we follow these steps:

- We simulate $\gamma = (\gamma_i)_{i=1}^P \sim N(0, I)$;
- Following Assumption 2, we define

$$\tilde{\alpha} \;=\; P^{-1/2} T^{-1/4} \tilde{\Omega}_\alpha^{1/2} \gamma \,. \tag{52}$$

- Given the chosen λ parameter, we define

$$\Omega_\alpha \;=\; \eta \tilde{\Omega}_\alpha, \tag{53}$$

where

$$\eta \;=\; \frac{\lambda}{\frac{T}{1+\hat{\theta}_M^2} \tilde{\alpha}' \Sigma_P^{-1} \tilde{\alpha}}. \tag{54}$$

- We then define

$$\alpha \;=\; P^{-1/2} T^{-1/4} \Omega_\alpha^{1/2} \gamma \,. \tag{55}$$

4.2.2 Estimating the power

To estimate the power of the test in finite samples, we simulate $i = 1, ..., N$ dataset of returns of size $P \times T$, for which we can compute the test statistic and reject the null hypothesis H_0

Parameter	Values
T	$60, 240, 360$
P	$[10, 10T]$
z	$[10^{-5}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 20, 10^2]$
λ	$[0.01T, 0.002PT]$
N	500
a	1
b	0.5

Table 1: Parameters table for the simulation procedure.

if

$$\frac{T^{1/2}(\hat{W}^{(i)}(z) - \hat{\xi}^{(i)}(z;c))}{\hat{\Delta}^{(i)}(z)^{1/2}} > \zeta_{0.95} \tag{56}$$

where $\zeta_{0.95}$ denotes the 95^{th} percentile of the standard normal distribution. We get our final power estimation by counting the average number of rejections over the N datasets

$$\Gamma_{sim}(z;c) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}_{\frac{T^{1/2}(\hat{W}^{(i)}(z) - \hat{\xi}^{(i)}(z;c))}{\hat{\Delta}^{(i)}(z)^{1/2}} > \zeta_{0.95}} \tag{57}$$

For a fixed T, we let P vary between 1 and $10T$. We compute $\Gamma_{sim}(z;c)$ for a grid of ridge penalties z and for various values of the non-centrality parameter λ (see Equation (14)). The parameters used for the simulation are defined in Table 1.

4.3 Asymptotic power of the test

To illustrate our theory, we also compute the asymptotic power of the test. From Theorem 7, we know that this computation requires knowledge of $\mathcal{H}(z;c)$ and $\Delta(z;c)$. We use $\Omega_\alpha =$

$aI + b\Sigma_P$ with $a = \eta,\ b = 0.5\eta$, where η is the scale parameter from Equation (54). We use Lemma 1 to compute $\mathcal{H}(z;c)$ in terms of m and utilize Corollary C.2 in the Appendix to deal with $m(-z;c)$ and $\Delta(z;c)$.

Following GRS, we study the behavior of the test as a function of P, allowing the non-centrality parameter λ (see Equation (14)) to grow with P and T. Namely, we consider settings with $\lambda(T,P) = \bar{\lambda}TP^{\delta}$, where $\delta = 0$ corresponds to the sparse α environment (see Equation (50)), while $\delta = 1$ corresponds to the "abundant" α environment. As in GRS, we use the value $\bar{\lambda} = 0.1$ in the case of $\delta = 0$ and $\bar{\lambda} = 0.002$ in the case of $\delta = 1$.

For any target level of $\lambda(T,P)$, with $P = cT$ compute α using the formula (55) with the corresponding scale $\eta(T,P)$:

$$\eta(T,cT) \;=\; \frac{\bar{\lambda}T(cT)^{\delta}}{\frac{T}{1+\hat{\theta}_M^2}\tilde{\alpha}'\Sigma_P^{-1}\tilde{\alpha}} \tag{58}$$

We approximate the asymptotic power of the test by

$$\Gamma_{asym}(z;c;T) = \Phi\left(-\zeta_{0.95} + \eta(T;cT)\frac{\mathcal{H}(z;c)}{\Delta(z;c)^{1/2}}\right). \tag{59}$$

Note that, by Proposition 9, we obtain a virtue of complexity for $\delta = 1$ (in fact, any $\delta > 0.5$ is sufficient for the test power to grow with P).

4.4 Size of the test

If the distribution of the test statistic in a finite sample is close to the asymptotic one, then our test has the correct size,

$$\text{Probability of incorrectly rejecting } H_0 = \Phi^{-1}(\zeta) \tag{60}$$

However, in finite samples, significant size distortions may occur. They are driven by two effects: non-normality of the test statistic and slow convergence of $\hat{\xi}(z;c)$ to its limit $\xi(z;c)$. These effects are particularly strong for small z due to the near-degeneracy of the sample covariance matrix.[5] We compute the simulated size $size_{sim}(z;c)$ of our test statistic following the same approach as for the test power, but setting $\alpha = (\alpha)_{i=1}^{P} = 0$ in the simulation procedure. Namely, in each simulated sample $i = 1, ..., N$, we reject H_0 if

$$\frac{T^{1/2}(\hat{W}^{(i)}(z) - \hat{\xi}^{(i)}(z;c))}{\hat{\Delta}^{(i)}(z)^{1/2}} > \zeta_{0.95}\,. \tag{61}$$

We can then estimate the size of our test in finite samples by computing

$$Size_{sim}(z;c) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{I}_{\frac{T^{1/2}(\hat{W}^{(i,j)}(z) - \hat{\xi}^{(i,j)}(z;c))}{\hat{\Delta}^{(i,j)}(z)^{1/2}} > \zeta_{0.95}}\,. \tag{62}$$

4.5 Simulation Results

Figure 2 reports the results of our simulations for $T = 60,\ 240,\ 360$. The plots show the simulated test power Γ_{sim} from (57), its RMT-based approximation Γ_{asym} from (59), the realized test size (62) (computed in a simulation with $\alpha = 0$), as well as the power Γ_{GRS} of the GRS test, as we increase the number of test assets and the corresponding complexity $c = P/T$.

All results are computed with the theoretically optimal shrinkage penalty $\tilde{z}_*(c)$ from (46), maximizing the asymptotic analog of (45) for a given c. We find that the test power saturates approximately around $\tilde{z} = 1$. For this reason, in all our results below we use $\tilde{z}_*(c) = 1$. This means that the ridge penalty is of the same order of magnitude as the covariance matrix itself.

We see several striking patterns. First, our RMT-based theoretical approximation agrees

[5] $\hat{\Sigma}_P$ is near-degenerate for $P \uparrow T$, and is degenerate ($\operatorname{rank}\hat{\Sigma}_P \leq T$) for $P > T$.

remarkably well with the actual realized test power, even for the very small sample size $T = 60$. In agreement with the virtue of complexity principle (Proposition 9), the power of the test is monotonically increasing in c when α is dense (corresponding to $\lambda = 0.002PT$). Second, the size of the test is small and approximately independent of complexity.[6] Third, the power of the test in the dense α regime saturates very quickly even as $P \uparrow T$, while the power of the GRS test collapses to zero at the boundary $T = P$ corresponding to $c = 1$. One interesting special case is $T = 60$, where the power of the GRS test is essentially zero, while the power of our test reaches 70% for $c \approx 10$ (ten times more test assets than observations).

As expected, the power of our test is low in sparse-α environments — it is monotonically decreasing in c when α is sparse, and $\|\alpha\|^2$ grows too slowly with P. This is precisely the situation where the FLY thresholding-based test should be doing well (FLY refer to this situation as a weak-α alternative hypothesis). FLY document that, in this situation, the power of their test statistic J_{thr} is high when applied to single stocks. However, the test is not rotation-invariant and alphas of portfolios no longer satisfy the alternative hypothesis. See Appendix B for further details on rotation-invariance.

5 Empirics

In this section, we illustrate how our test works with real data through variations in the candidate efficient portfolio and test assets.

5.1 Data and Test Assets

We use monthly data for 153 stock characteristics for US stocks over the period 1963 to 2022 (from Jensen, Kelly and Pedersen, 2023, JKP henceforth). This open-source data set

[6] This size behavior depends crucially on the ridge penalty z being uniformly away from zero, which is the case for the optimal penalty $\tilde{z}_*(c)$.

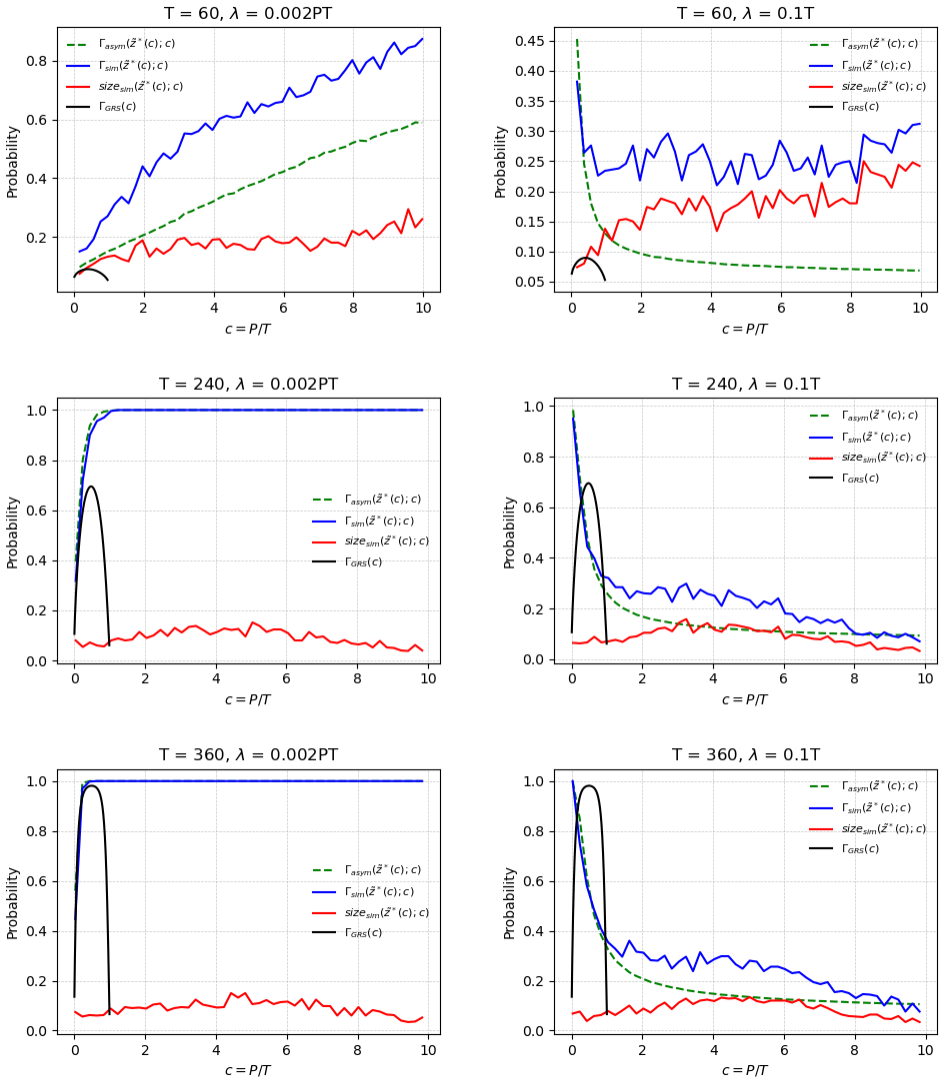


Figure 2: This figure plots the estimated test's power using Monte-Carlo simulation, the asymptotic test's power, the estimated test's size using Monte-Carlo simulation, and the GRS test's power as a function of the complexity $c = P/T$ and the optimal ridge penalty, Z_*, that maximizes the asymptotic power for each c. The non-centrality parameter used are $\lambda = 0.1T$ and $\lambda = 0.002PT$ (columns). We pick $T = 60, 240, 360$ (rows), $a = 1$, $b = 0.5$, $N = 500$ and we let P vary between 10 and $10T$.

compiles an extensive collection of stock characteristics from the finance literature.[7] The

[7]JKP characteristics for individual stocks can be downloaded at https://wrds-www.wharton.upenn.edu/pages/get-data/contributed-data-forms/global-factor-data/. Detailed documentation and further factor portfolio data are available at jkpfactors.com.

universe includes NYSE/AMEX/NASDAQ stocks with CRSP share codes 10–12. As in JKP, we define mega, large, small, and micro stocks using the 80th, 50th, 20th, and 1st percentile of NYSE stocks market caps. Nano caps are the remaining stocks.

To have a stock-month panel with a consistent set of available characteristics over time, we follow the approach of Didisheim et al. (2024). First, we restrict the raw dataset to a subset of 132 characteristics that have fewer than 30% missing values over the full sample. Next, we drop "nano" stocks whose market capitalization is below the 1^{st} percentile of the NYSE sample. Then, we exclude stock-month observations with missing values for more than a third of the characteristics. Finally, we cross-sectionally rank-standardize each characteristic into the $[-0.5, 0.5]$ interval and replace any missing characteristic values with the cross-sectional median value of zero. After this filtering procedure, the resulting $N_t \times 132$ matrix of characteristics each month constitutes the conditioning set X_t used in our empirical analysis. We also use $R_{t+1} = (R_{i,t+1})_{i=1}^{N_t}$ to denote the vector of next-month stock returns in excess of the risk-free rate.

As Theorem 1 shows, the unique unconditionally efficient portfolio that prices all assets both unconditionally and conditionally is given by

$$R^M_{t+1} = \pi_t' R_{t+1}, \ \pi_t = E_t[R_{t+1}R_{t+1}']^{-1} E_t[R_{t+1}], \tag{63}$$

where the last identity follows from the Sherman-Morrison formula (Lemma A.1 in the Appendix). As Didisheim et al. (2024) show, if $\pi_t = w(X_t)$ for some sufficiently well-behaved function w, then R^M_{t+1} can be approximated as follows. First, we generate a large number of stock characteristics $S_t = (S_{i,t})_{i=1}^{N_t} \in \mathbb{R}^{N_t \times P}$ through transformations of $X_{i,t}$, and build the corresponding managed portfolios, or, factors, with factor returns R^F_{t+1} given by

$$R^F_{t+1} = S_t' R_{t+1} \in \mathbb{R}^P. \tag{64}$$

Second, we build an *unconditionally* efficient portfolio of R^F :

$$\lambda \;=\; E[R^F_{t+1}(R^F_{t+1})']^{-1}E[R^F_{t+1}] \;\in\; \mathbb{R}^P\,. \tag{65}$$

Then,

$$R^M_{t+1} \;=\; \pi_t' R_{t+1} \;=\; \lambda' R^F_{t+1} \;+\; O(1/P). \tag{66}$$

To implement this methodology with real data, we follow Didisheim et al. (2024). First, we choose a rolling window T and a ridge penalty $z_M \geq 0$. Then, we construct an approximation $\hat{\lambda}_t(z_M)$ to infeasible λ from (65) via

$$R^M_{t+1}(z_M) \;=\; \hat{\lambda}_t(z_M)' R^F_{t+1}, \tag{67}$$

where

$$\hat{\lambda}_t(z_M) \;=\; \left(z_M I + \frac{1}{T}\sum_{\tau=1}^{T} R^F_{t+1-\tau}(R^F_{t+1-\tau})'\right)^{-1} \frac{1}{T}\sum_{\tau=1}^{T} R^F_{t+1-\tau}\,. \tag{68}$$

Note that the returns $R^M_{t+1}(z_M)$ are always out-of-sample because $\hat{\lambda}_t(z_M)$ is based on past data.

As is common in the literature, we use characteristic-managed portfolios (64) both as our test assets and as inputs R^F_{t+1} for constructing R^M_{t+1} using (67). We consider two choices for S_t :

- $S_t = X_t \in \mathbb{R}^{N_t \times 132}$. We refer to the corresponding managed portfolios

$$X_t' R_{t+1} \tag{69}$$

as *JKP factors.* We also refer to the corresponding portfolio returns R^M constructed

in (67) as BSV because this model uses portfolio weights $\pi_t = X_t\hat{\lambda}_t(z_M)$ that are linear in stock characteristics, as in Brandt, Santa-Clara and Valkanov (2009) (BSV).

- $S_t = \text{ReLU}(X_t\gamma) \in \mathbb{R}^{N_t \times P}$, where $\gamma = (\gamma_{j,k}) \in \mathbb{R}^{d\times P}$, $P = 25,000$, and $\gamma_{j,k} \sim N(0,1)$. These are non-linear *random features* (RF) studied in Didisheim et al. (2024) (DKKM). We refer to the corresponding managed portfolios

$$\text{ReLU}(X_t\gamma)'R_{t+1} \tag{70}$$

as 25,000 RF and to the efficient portfolio constructed from them using (67) as DKKM.[8]

To understand the dynamics of alpha over time, we estimate our test statistic over rolling windows of $60, 120$, and 250 months. Computing the test statistic when P is large is computationally tricky because it requires inverting a high-dimensional covariance matrix of alpha. We use non-trivial algebraic identities (see Appendix D.1, Corollary D.3) to compute it efficiently, even for very large P.

5.2 Testing Alpha Out-of-Sample

Proposition 8 implies that the power of our test is directly related to the actual out-of-sample performance of the alpha portfolio, consistent with the economic intuition behind the GRS test (see Section 2.4). To quantify the economic significance of the alpha, measured by our test statistic, we proceed as follows:

- Given the returns R_τ, $\tau \in [t-T, t]$ of test assets in the rolling window, we estimate

$$\hat{\beta}_{t-T,t} = \frac{1}{T}\sum_{\tau\in[t-T,t]} R_\tau(R^M_\tau - \bar{R}^M_{[t-T,t]}) \tag{71}$$

[8] Didisheim et al. (2024) use sin and cos activation functions in their construction of random features, while we use the more common ReLU activation. Our results are insensitive to the particular choice of non-linear activation.

and compute the realized, out-of-sample α,

$$\hat{\alpha}_{t+1}^{OOS} \;=\; R_{t+1} \;-\; R_{t+1}^{M}\hat{\beta}_{t-T,t} \tag{72}$$

as well as the in-sample α used in our test statistic,

$$\hat{\alpha}_{\tau} \;=\; R_{\tau} \;-\; R_{\tau}^{M}\hat{\beta}_{t-T,t},\ \tau \in [t-T,t]\,, \tag{73}$$

and its realized in-sample covariance matrix $\hat{\Sigma}_{P,[t-T,t]}$ and mean $\bar{\alpha}_{[t-T,t]}$. Then, we compute the out-of-sample alpha portfolio return

$$R_{\alpha,t+1}^{OOS}(z) \;=\; (\hat{\alpha}_{t+1}^{OOS})'\hat{\Sigma}_{P,[t-T,t]}(z)^{-1}\bar{\alpha}_{[t-T,t]}\,, \tag{74}$$

and compute the realized Sharpe ratio of $R_{\alpha,t+1}^{OOS}(z)$ over the period $t \geq T$. We compute this statistic efficiently without inverting the high-dimensional matrix $\hat{\Sigma}_{P,[t-T,t]}(z)$, using Lemma D.3 in the Appendix D.1.

5.3 Time-Series of the Test Statistic

We investigate the ability of our candidate SDFs, FF6, BSV based on (69), and DKKM based on (70) to price the two sets of test assets over time by computing the test statistic over a rolling window of $T = 240$ months. We always use the optimal ridge penalty (46), which is close to the value of 1, just like in simulations. To address the potential issues with transaction costs, we report the results separately for the universe of all stocks (Figure 3) and mega stocks (Figure 4).

Consistent with the literature on the factor zoo, these Figures clearly show that FF6 is unable to price even the JKP factors. By contrast, both the BSV and the DKKM candidate SDFs are often able to price the JKP factors built from the universe of all stocks. Perhaps

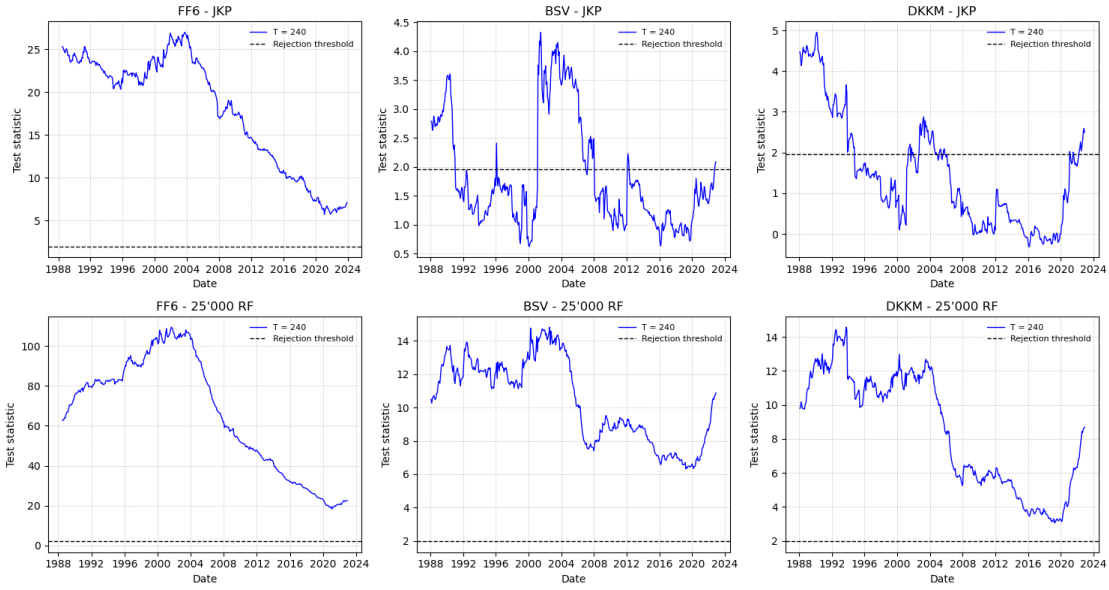


Figure 3: The plot shows the time series dynamics of $\frac{T^{1/2}(\hat{W}(z)-\hat{\xi}(z))}{\hat{\Delta}(z;c)^{1/2}}$ from Theorem 5, computed using $T = 240$. We estimate $\hat{\alpha}$ using (8)–(10). The test assets are non-linear ReLu DKKM factors (70) and linear JKP factors (69) (rows), and we use FF6, BSV, and DKKM (columns) as R_t^M. We consider all stocks.

surprisingly, the picture is different for factors built from mega stocks, where even DKKM is unable to price them.

The picture for the test assets based on 25000 non-linear characteristics (70) is more severe and more consistent. In all plots, for all time periods, neither JKP nor DKKM are able to price them. This is particularly surprising for the DKKM-SDF because it is constructed from these 25000 assets. The reason behind this phenomenon is what Didisheim et al. (2024) refer to as "limits-to-learning" — in high complexity settings (in particular, when $P > T$), there is always a gap between the performance of the estimated model and the "true," infeasible model due to lack of data. Indeed, the true optimal portfolio of the 25000 test assets is a 25,000-dimensional vector, and we can only estimate an extremely noisy version of it from the 240 observations we have access to. Table 2 reports the test statistics

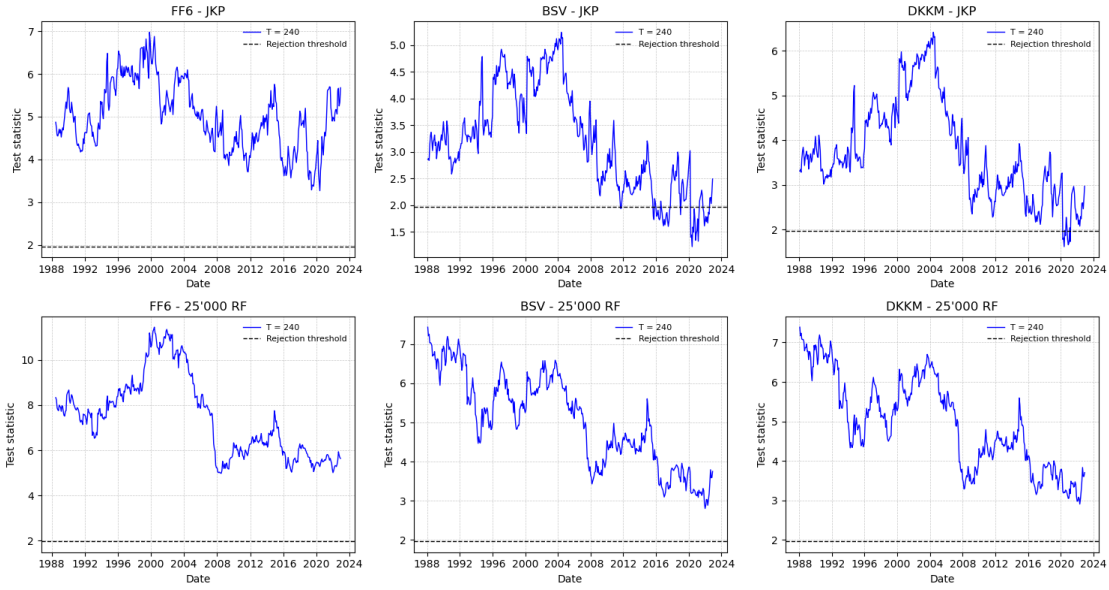


Figure 4: The plot shows the time series dynamics of $\frac{T^{1/2}(\hat{W}(z)-\hat{\xi}(z))}{\hat{\Delta}(z;c)^{1/2}}$ from Theorem 5, computed using $T = 240$. We estimate $\hat{\alpha}$ using (8)–(10). The test assets are non-linear ReLu DKKM factors (70) and linear JKP factors (69) (rows), and we use FF6, BSV, and DKKM (columns) as R_t^M. We consider mega stocks.

	FF6	BSV	DKKM
JKP	35.86	2.51	3.87
25000 RF	81.39	14.99	13.50
JKP (mega)	10.05	7.28	8.38
25000 RF (mega)	10.52	7.74	7.96

Table 2: Test statistic for each portfolio and test assets set computed on the whole sample. We consider the full stock universe as well as mega stocks only.

over the whole sample and shows that none of the candidate SDFs can price the 25000 test assets, with full-sample test statistics close to 10.

The test statistics exhibit interesting time series dynamics, suggesting that some of the α has been monotonically decreasing over the 2004-2020 period and started increasing again

after COVID-19. This increase is quite large in sample plots, possibly originating with a decrease in market efficiency due to the growth of retail trading in the post-COVID-19 period.

Figures 5 and 6 report the cumulative performance of alpha-portfolios constructed using (74). We conclude that the alpha detected by the test is indeed tradable out-of-sample and generates non-trivial performance, with the same alpha decay during the 2004-2020 period and a partial rebound in recent years. Interestingly enough, this alpha decay post-2004 is much smaller for the JKP test assets based on mega stocks, suggesting that some of the inefficiencies are persistent even in this highly liquid universe.

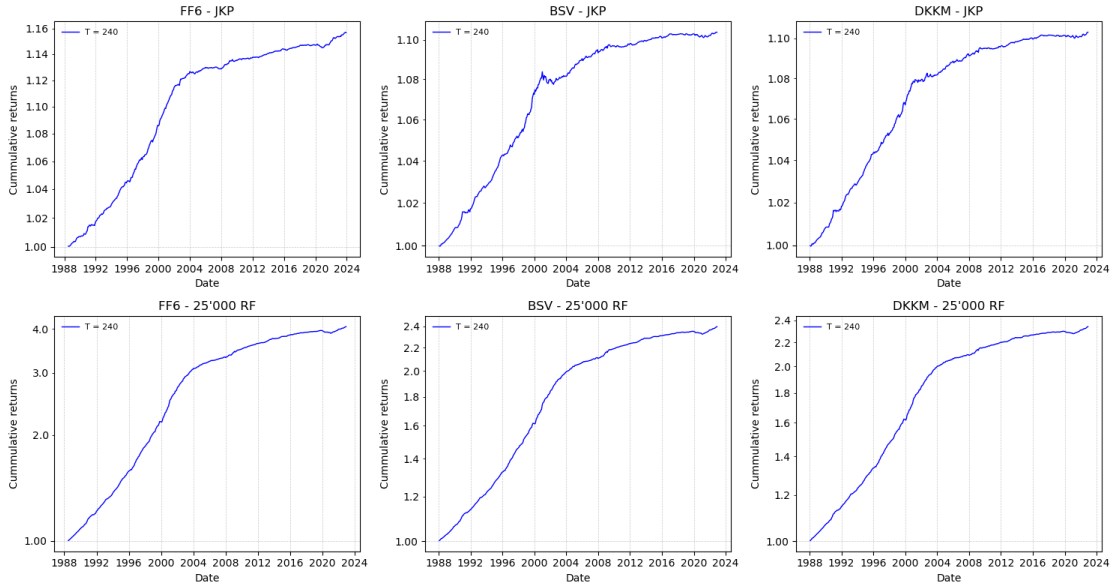


Figure 5: The plot shows the cumulative returns of the α portfolio constructed using (72), computed using $T = 240$. We estimate $\hat{\alpha}$ using (8)–(10). The test assets are non-linear ReLu DKKM factors (70) and linear JKP factors (69) (rows). We use FF6, BSV, and DKKM (columns) as R_t^M. We consider all stocks.

Table 3 shows that the realized Sharpe ratios of these alpha portfolios are quite high, exceeding 3 for all stocks and approximately 1 for the mega-universe. However, Table 4 indicates that the alpha is not "pure." Due to misestimation of β coefficients, the out-of-

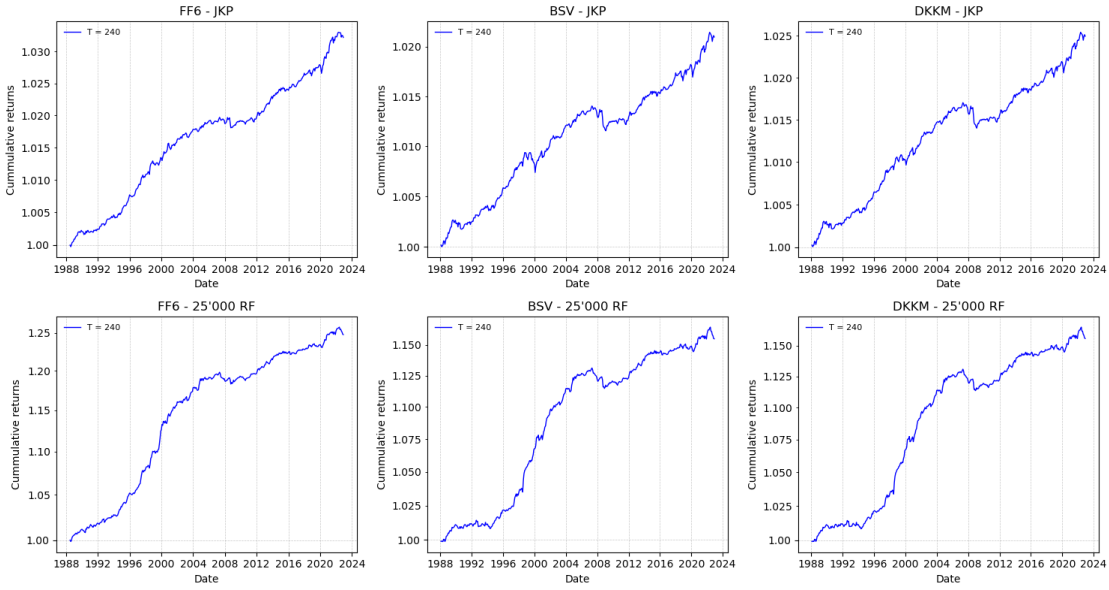


Figure 6: The plot shows the cumulative returns of the α portfolio constructed using (72), computed using $T = 240$. We estimate $\hat{\alpha}$ using (8)–(10). The test assets are non-linear ReLu DKKM factors (70) and linear JKP factors (69) (rows). We use FF6, BSV, and DKKM (columns) as R_t^M. We consider mega stocks.

Test Assets	FF6	BSV	DKKM
JKP	2.03	1.43	1.67
25000 RF	3.22	3.07	3.05
JKP (mega)	1.07	0.81	0.87
25000 RF (mega)	1.20	0.93	0.94

Table 3: $\alpha's$ portfolio Sharpe ratio with T = 240 and test assets being either non-linear ReLu DKKM factors or linear JKP factors. We consider the full stock universe as well as mega stocks only.

sample returns (74) actually preserve a non-trivial correlation with the SDF to which they were orthogonalized. The reason is that β's are overfit in-sample and hence do not work well out-of-sample. Mitigating this β-overfit problem in high complexity settings and constructing the "pure alpha" portfolios is an important direction for future research.

Test Assets	FF6	BSV	DKKM
JKP	0.09	0.30	0.40
25000 RF	0.12	0.55	0.57
JKP (mega)	0.33	0.21	0.19
25000 RF (mega)	0.35	0.39	0.36

Table 4: Correlation of the optimal portfolio of α's and R^M, with T = 240 and test assets being either non-linear ReLu DKKM factors or linear JKP factors. We consider the full stock universe as well as mega stocks only.

6 Conclusion

The seminal Gibbons, Ross and Shanken (1989) test of portfolio efficiency aims to establish whether alphas in a regression of test asset returns on the candidate portfolio are significantly different from zero. We extend this test to a setting with many test assets. The motivation for this is two-fold. First, the original test does not apply to cases when the number of test assets exceeds the number of time-series observations. Yet, as Hansen and Richard (1987) point out, one has to consider all feasible portfolios as test assets in order to be able to conclude whether the candidate efficient portfolio prices assets conditionally. Second, even if one limits themselves to a smaller set of assets, the test may suffer from weak power. Earlier studies, such as Fan, Liao and Yao (2015), address power issues using thresholding-based tools that rely on the sparseness of alphas. Such approach is not rotation-invariant, that is, if alphas are sparse for individual assets, they will not be sparse for asset portfolios hindering the implementation of the Hansen and Richard (1987) dictum.

We manage to address all these issues using the tools from Random Matrix Theory. The GRS statistic becomes biased in high dimensions because of the degeneracy in the covariance matrix of test assets. We de-bias and rescale it to obtain its asymptotic distribution. Further, we characterize the power of this statistic both analytically and via simulations. This power

grows very quickly with the number of test assets unless alphas are sparse. In that case, thresholding-based tests have better power with the caveat that they might not be empirically appropriate due to dependence on a specific rotation.

We apply the test empirically using state-of-the-art test assets and candidate efficient portfolios. Despite the progress in estimating efficient portfolios, they are rejected by our test. We interpret some of these rejections as "limits-to-learning" in high-dimensional environments.

References

Bai, Zhi-Dong and Yong-Qua Yin, "Limit of the smallest eigenvalue of a large dimensional sample covariance matrix," in "Advances In Statistics," World Scientific, 2008, pp. 108–127.

Bai, Zhidong and Hewa Saranadasa, "Effect of high dimension: by an example of a two sample problem," *Statistica Sinica*, 1996, pp. 311–329.

___ **and Wang Zhou**, "Large sample covariance matrices without independence structures in columns," *Statistica Sinica*, 2008, pp. 425–442.

Brandt, Michael W, Pedro Santa-Clara, and Rossen Valkanov, "Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns," *The Review of Financial Studies*, 2009, *22* (9), 3411–3447.

Carrasco, Marine and Cheikh Nokho, "Hansen-Jagannathan distance with many assets," *Available at SSRN 5078000*, 2022.

Chamberlain, Gary and Michael Rothschild, "Arbitrage, factor structure, and mean-variance analysis on large asset markets," 1982.

Dello-Preite, Massimo, Raman Uppal, Paolo Zaffaroni, and Irina Zviadadze, "What is Missing in Asset-Pricing Factor Models?," 2022.

Didisheim, Antoine, Shikun Barry Ke, Bryan T Kelly, and Semyon Malamud, "APT or "AIPT"? The Surprising Dominance of Large Factor Models," Technical Report, National Bureau of Economic Research 2024.

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

Fan, Jianqing, Yuan Liao, and Jiawei Yao, "Power enhancement in high-dimensional cross-sectional tests," *Econometrica*, 2015, *83* (4), 1497–1541.

___ , ___ , **and Martina Mincheva**, "High dimensional covariance matrix estimation in approximate factor models," *Annals of statistics*, 2011, *39* (6), 3320.

Feng, Long, Wei Lan, Binghui Liu, and Yanyuan Ma, "High-dimensional test for alpha in linear factor pricing models with sparse alternatives," *Journal of Econometrics*, 2022, *229* (1), 152–175.

Gagliardini, Patrick, Elisa Ossola, and Olivier Scaillet, "Time-varying risk premium in large cross-sectional equity data sets," *Econometrica*, 2016, *84* (3), 985–1046.

Gibbons, Michael R, Stephen A Ross, and Jay Shanken, "A test of the efficiency of a given portfolio," *Econometrica: Journal of the Econometric Society*, 1989, pp. 1121–1152.

Giglio, Stefano and Dacheng Xiu, "Asset pricing with omitted factors," *Journal of Political Economy*, 2021, *129* (7), 1947–1990.

___ , ___ , **and Dake Zhang**, "Test assets and weak factors," *The Journal of Finance*, 2025, *80* (1), 259–319.

___ , **Yuan Liao, and Dacheng Xiu**, "Thousands of alpha tests," *The Review of Financial Studies*, 2021, *34* (7), 3456–3496.

Hansen, Lars Peter, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 1982, *50* (4), 1029–1054.

___ **and Ravi Jagannathan**, "Implications of security market data for models of dynamic economies," *Journal of political economy*, 1991, *99* (2), 225–262.

___ **and Scott F Richard**, "The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models," *Econometrica: Journal of the Econometric Society*, 1987, pp. 587–613.

Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen, "Is there a replication crisis in finance?," *The Journal of Finance*, 2023, *78* (5), 2465–2518.

Kelly, Bryan, Semyon Malamud, and Kangying Zhou, "The virtue of complexity in return prediction," *The Journal of Finance*, 2024, *79* (1), 459–503.

Kelly, Bryan T, Semyon Malamud, Mohammad Pourmohammadi, and Fabio Trojani, "Universal Portfolio Shrinkage," Technical Report, National Bureau of Economic Research 2024.

Kock, Anders Bredahl and David Preinerstorfer, "Power in High-Dimensional Testing Problems," *Econometrica*, 2019, *87* (3), 1055–1069.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, "Shrinking the cross-section," *Journal of Financial Economics*, 2020, *135* (2), 271–292.

Lan, Wei, Long Feng, and Ronghua Luo, "Testing High-Dimensional Linear Asset Pricing Models," *Journal of Financial Econometrics*, 2018, *16* (2), 191–210.

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

Lettau, Martin and Markus Pelger, "Factors that fit the time series and cross-section of stock returns," *The Review of Financial Studies*, 2020, *33* (5), 2274–2325.

Li, Haoran, Alexander Aue, and Debashis Paul, "High-dimensional general linear hypothesis tests via non-linear spectral shrinkage," *Bernoulli*, 2020, *26* (4), 2541–2571.

—— , —— , —— , Jie Peng, and Pei Wang, "An Adaptable Generalization of Hotelling's T^2 Test in High Dimension," *The Annals of Statistics*, 2020, *48* (3), 1815–1847.

Liu, Haoyang, Alexander Aue, and Debashis Paul, "On the Marčenko–Pastur law for linear time series," 2015.

Ma, Shujie, Wei Lan, Liangjun Su, and Chih-Ling Tsai, "Testing alphas in conditional time-varying factor models with high-dimensional assets," *Journal of Business & Economic Statistics*, 2020, *38* (1), 214–227.

Marčenko, Vladimir A and Leonid Andreevich Pastur, "Distribution of eigenvalues for some sets of random matrices," *Mathematics of the USSR-Sbornik*, 1967, *1* (4), 457.

Pesaran, M Hashem and Takashi Yamagata, "Testing for Alpha in Linear Factor Pricing Models with a Large Number of Securities*," *Journal of Financial Econometrics*, 02 2023, *22* (2), 407–460.

Srivastava, Muni S, "A test for the mean vector with fewer observations than the dimension under non-normality," *Journal of Multivariate Analysis*, 2009, *100* (3), 518–532.

Yu, Xiufan, Jiawei Yao, and Lingzhou Xue, "Power enhancement for testing multi-factor asset pricing models via Fisher's method," *Journal of Econometrics*, 2024, *239* (2), 105458.

A Proofs of Auxiliary Results

We will need the following known lemma.

Lemma A.1 (Sherman-Morrison formula)

$$\begin{aligned}(A+xx')^{-1} &= A^{-1} - A^{-1}xx'A^{-1}/(1+x'A^{-1}x) \\ (A+xx')^{-1}x &= A^{-1}x/(1+x'A^{-1}x)\end{aligned} \tag{A.1}$$

for any matrix $A \in \mathbb{R}^{P\times P}$ *and any vector* $x \in \mathbb{R}^P$.

Proof of Theorem 1. First, if

$$R_{t+1}^M = \pi_t' R_{t+1}, \tag{A.2}$$

then

$$\mathrm{Var}_t[R_{t+1}^M] = \pi_t'\Sigma_t\pi_t, \tag{A.3}$$

where $\Sigma_t = \mathrm{Cov}_t(R_{t+1}), \mu_t = E_t[R_{t+1}]$, and

$$\pi_t = a_t^{-1}\Sigma_t^{-1}E_t[R_{t+1}] \Leftrightarrow E_t[R_{t+1}] = a_t\Sigma_t\pi_t = a_t\mathrm{Cov}_t(R_{t+1}, R_{t+1}^M) = a_t\mathrm{Var}_t[R_{t+1}^M]\beta_t, \tag{A.4}$$

where

$$a_t\mathrm{Var}_t[R_{t+1}^M] = a_t\pi_t'\Sigma_t\pi_t = a_t^{-1}\mu_t'\Sigma_t^{-1}\mu_t \tag{A.5}$$

while

$$E_t[R^M_{t+1}] \;=\; \pi_t'\mu_t \;=\; a_t^{-1}\mu_t'\Sigma_t^{-1}\mu_t\,. \tag{A.6}$$

Reversing the arguments, we get that the first two items of the theorem are, in fact, equivalent.

Now, suppose R^M_{t+1} is the efficient portfolio. Then,

$$E[R^Z_{t+1}] \;=\; E[Z_tR_{t+1}] \;=\; E[Z_tE_t[R_{t+1}]] \;=\; E[Z_t\beta_tE_t[R^M_t]]\,. \tag{A.7}$$

At the same time, by Lemma A.1, we have that

$$\begin{aligned} E_t[R_{t+1}R_{t+1}']^{-1}\mu_t &= (\Sigma_t+\mu_t\mu_t')^{-1}\mu_t \\ &= \Sigma^{-1}\mu_t/(1+\mu_t'\Sigma_t^{-1}\mu_t) \\ &= \;(1+\theta^2_{M,t})^{-1}\Sigma_t^{-1}\mu_t\,, \end{aligned} \tag{A.8}$$

where $\theta^2_{M,t}=\mu_t'\Sigma_t^{-1}\mu_t$. Hence, if $R^M_{t+1}=R_{t+1}'\pi_t$, the identity

$$E_t[R_{t+1}] \;=\; E_t[R_{t+1}R^M_{t+1}] \tag{A.9}$$

holds if and only if

$$\pi_t \;=\; (1+\theta^2_{M,t})^{-1}\Sigma_t^{-1}\mu_t \tag{A.10}$$

because

$$E_t[R_{t+1}R^M_{t+1}] \;=\; E_t[R_{t+1}R_{t+1}]\pi_t \tag{A.11}$$

Now, standard arguments imply that the conditional identity (A.9) is equivalent to the

unconditional identity

$$E[R^Z_{t+1}] \;=\; E[R^Z_{t+1}R^M_{t+1}] \tag{A.12}$$

holding for any Z. Furthermore,

$$E[R^Z_{t+1}R^M_{t+1}] \;=\; \mathrm{Var}[R^M_{t+1}]\beta^Z \;+\; E[R^Z_{t+1}]E[R^M_{t+1}] \tag{A.13}$$

and hence (A.12) is equivalent to

$$E[R^Z_{t+1}] \;=\; \mathrm{Var}[R^M_{t+1}]\beta^Z \;+\; E[R^Z_{t+1}]E[R^M_{t+1}], \tag{A.14}$$

which is equivalent to

$$E[R^Z_{t+1}] \;=\; \frac{\mathrm{Var}[R^M_{t+1}]}{1-E[R^M_{t+1}]}\beta^Z\,. \tag{A.15}$$

Applying (A.12) to R^M_{t+1}, we get

$$E[(R^M_{t+1})^2] \;=\; E[R^M_{t+1}] \tag{A.16}$$

and, hence, after some algebra,

$$\frac{\mathrm{Var}[R^M_{t+1}]}{1-E[R^M_{t+1}]} \;=\; E[R^M_{t+1}]\,. \tag{A.17}$$

The proof is complete. □

Proof of Theorem 2. Step 1: Setup and Notation. Define the sample averages

$$X_P := \frac{1}{P}\sum_{k=1}^{P} Z_k^2 \quad \text{and} \quad Y_{T-P-1} := \frac{1}{T-P-1}\sum_{k=1}^{T-P-1} \tilde{Z}_k^2.$$

Note that Z_k^2 and $\tilde{Z}_k^2$ are identically distributed χ_1^2 random variables with mean 1 and variance 2.

By the given relation, we have

$$\frac{T(T-P-1)}{P(T-2)}\hat{W} \sim \frac{X_P}{Y_{T-P-1}}.$$

Step 2: Asymptotic Distributions of X_P and Y_{T-P-1}. Since $\{Z_k\}$ are i.i.d. $N(0,1)$, we know Z_k^2 has $\mathbb{E}[Z_k^2] = 1$ and $\text{Var}(Z_k^2) = 2$. By the Central Limit Theorem (CLT), for large P,

$$\sqrt{P}(X_P - 1) = \sqrt{P}\left(\frac{1}{P}\sum_{k=1}^{P} Z_k^2 - 1\right) \xrightarrow{d} N(0,2).$$

Equivalently,

$$X_P = 1 + \sqrt{\frac{2}{P}}N_1,$$

where $N_1 \sim N(0,1)$ asymptotically.

Similarly, since $\{\tilde{Z}_k\}$ are also i.i.d. $N(0,1)$ and independent from $\{Z_k\}$, we apply the CLT to Y_{T-P-1}:

$$\sqrt{T-P-1}(Y_{T-P-1} - 1) = \sqrt{T-P-1}\left(\frac{1}{T-P-1}\sum_{k=1}^{T-P-1} \tilde{Z}_k^2 - 1\right) \xrightarrow{d} N(0,2).$$

Thus,

$$Y_{T-P-1} = 1 + \sqrt{\frac{2}{T-P-1}}N_2,$$

where $N_2 \sim N(0,1)$ and is independent of N_1.

Step 3: Ratio Approximation. Consider the ratio

$$\frac{X_P}{Y_{T-P-1}} = \frac{1 + \sqrt{\frac{2}{P}} N_1}{1 + \sqrt{\frac{2}{T-P-1}} N_2}.$$

For large T and P, both $\sqrt{2/P}$ and $\sqrt{2/(T-P-1)}$ are small. We use a first-order Taylor expansion:

$$\frac{1}{1+x} \to 1 - x \quad \text{for small } x.$$

Set $x = \sqrt{\frac{2}{T-P-1}} N_2$. Then

$$\frac{X_P}{Y_{T-P-1}} \to (1 + \sqrt{\tfrac{2}{P}} N_1)(1 - \sqrt{\tfrac{2}{T-P-1}} N_2).$$

Ignoring terms of order $1/\sqrt{PT}$ and higher, we have

$$\frac{X_P}{Y_{T-P-1}} \to 1 + \sqrt{\frac{2}{P}} N_1 - \sqrt{\frac{2}{T-P-1}} N_2.$$

Since N_1 and N_2 are independent standard normals, the linear combination

$$\sqrt{\frac{2}{P}} N_1 - \sqrt{\frac{2}{T-P-1}} N_2$$

is normal with mean zero and variance

$$2\left(\frac{1}{P} + \frac{1}{T-P-1}\right).$$

Step 4: Including the Scaling Factor. Recall

$$\frac{T(T-P-1)}{P(T-2)} \hat{W} \sim \frac{X_P}{Y_{T-P-1}}.$$

Thus,

$$\hat{W} \sim \frac{P(T-2)}{T(T-P-1)} \frac{X_P}{Y_{T-P-1}}.$$

As $T \to \infty$ with $P/T \to c \in (0,1)$,

$$\frac{P(T-2)}{T(T-P-1)} = \frac{\frac{P}{T}(T-2)}{T-P-1} \to \frac{c}{1-c}.$$

Therefore,

$$\hat{W} \to \frac{c}{1-c}\left(1 + \sqrt{\frac{2}{P}}N_1 - \sqrt{\frac{2}{T-P-1}}N_2\right).$$

Step 5: Computing the Mean and Variance. The mean of $\hat{W}$ is

$$\mathbb{E}[\hat{W}] \to \frac{c}{1-c}.$$

For the variance, we have

$$\text{Var}\left(\sqrt{\frac{2}{P}}N_1 - \sqrt{\frac{2}{T-P-1}}N_2\right) = 2\left(\frac{1}{P} + \frac{1}{T-P-1}\right).$$

Since $P \sim cT$ and $T-P-1 \sim (1-c)T$, this becomes

$$2\left(\frac{1}{cT} + \frac{1}{(1-c)T}\right) = \frac{2}{T}\left(\frac{1}{c} + \frac{1}{1-c}\right) = \frac{2}{Tc(1-c)}.$$

Multiplying by $\left(\frac{c}{1-c}\right)^2$,

$$\lim_{P,T\to\infty, P/T\to c} \text{Var}(\hat{W}) = \left(\frac{c}{1-c}\right)^2 \frac{2}{Tc(1-c)} = \frac{2c}{(1-c)^3 T}.$$

Step 6: Convergence in Distribution. We have shown that as $T \to \infty$,

$$\hat{W} \xrightarrow{d} N\left(\frac{c}{1-c}, \frac{2c}{(1-c)^3 T}\right).$$

Since the variance decreases as T grows, for large T, $\hat{W}$ is well-approximated by this normal distribution.

This completes the proof. □

B The Origins of Non-Sparse Covariance Matrices And Non-Sparse Alpha: Sparsity Is Not Rotation Invariant

Fan et al. (2015) develop powerful techniques for testing asset pricing models when the covariance matrix and/or the α vector of test assets is sparse. Here, we provide a simple argument that if our test assets are portfolios based on a large number of non-linear characteristics as in Didisheim et al. (2024), then sparsity becomes even harder to generate, as is shown by the following result.

Proposition B.1 *Suppose that stock returns are $r_{i,t}$ and test assets are portfolios,*

$$R_{t+1}(\theta) \;=\; \sum_{i=1}^{N} f(X_{i,t};\theta) r_{i,t+1}, \tag{B.1}$$

where $X_{i,t} \in \mathbb{R}^d$ are instruments (stock characteristics) and $f(X_{i,t};\theta_k)$ are bounded and real analytic in θ. Then, for Lebesgue-almost every $\theta, \tilde{\theta}$, $\text{Cov}(R_{t+1}(\theta), R_{t+1}(\tilde{\theta})) \neq 0$.

If $R^M_{t+1} \;=\; \sum_{i=1}^{N} g_i(X_t) R_{t+1}$ for some functions $g_i : \mathbb{R}^{N\times d} \to \mathbb{R}^N$, then

$$\alpha(\theta) \;=\; E[R_{t+1}(\theta)] \;-\; \frac{\text{Cov}(R_{t+1}(\theta), R^M_{t+1})}{\text{Var}[R^M_{t+1}]} E[R^M_{t+1}] \tag{B.2}$$

is either identically zero or is non-zero for Lebesgue-almost every θ.

More generally, in stark contrast to the GRS test, any thresholding-based test is not invariant to linear transformations: If we replace original returns $R_t \in \mathbb{R}^P$ with a linear transformation $AR_t \in \mathbb{R}^K$ for some matrix $A \in \mathbb{R}^{K \times P}$ of portfolio weights, the α of the corresponding portfolio returns AR_t lose the sparsity of the underlying assets' α. The following result holds.

Proposition B.2 (Test Statistic Under Linear Transformations) *Suppose that* $R_t = \alpha + \beta R_t^M + \varepsilon_t$. *Define* $R_t^A = AR_t$.

- *If* $A \in \mathbb{R}^{P \times P}$ *is non-degenerate, then the GRS statistic does not change: the test statistic* (11) *satisfies* $\hat{W}(R) = \hat{W}(R^A)$.
- *If* $A \in \mathbb{R}^{P \times P}$ *is an orthogonal matrix,* $AA' = I$, *then the test statistic* (18) *does not change:* $\hat{W}(R) = \hat{W}(R^A)$.
- *No test statistic involving thresholding can be invariant under any open set of linear transformations.*
- *If* α *is not identically zero, then for almost every orthogonal* $A \in \mathbb{R}^{P \times P}$ *under the Haar measure,* $A\alpha$ *is not sparse: All of its coordinates are non-zero with probability one.*

Proposition B.2 shows that the GRS test has economically intuitive invariance properties: If we replace original assets with portfolios spanning the same set of payoffs, the statistic does not change. Our ridge-penalized statistic has a slightly weaker property of being invariant under rotations:[9] A change of an orthogonal basis of spanning portfolios does not modify the statistic. By contrast, sparsity-based statistics can never have this invariance property, and neither can the sparsity itself: A small, generic rotation of the asset space destroys sparsity.

[9] We conjecture that, for $P > T$, there are no meaningful test statistics that are invariant under linear transformations.

C Random Matrix Theory and Power Under Local Alternatives

A key object of RMT is the eigenvalue distribution $H_A(x)$ of a symmetric matrix $A \in \mathbb{R}^{P\times P}$:

$$H_A(x) \;=\; \frac{1}{P}\sum_{i=1}^{P} \mathbf{1}_{x<\lambda_i(A)}, \tag{C.1}$$

where $\lambda_i(A)$ are the eigenvalues of the matrix A.

Definition 1 *Given a sequence of symmetric matrices $\Sigma_P \in \mathbb{R}^{P\times P}$, $P \to \infty$, we say that their eigenvalue distributions H_{Σ_P} are asymptotically stable if exists a distribution H with compact support in $(0,\infty)$, non-degenerate at zero, such that*

$$\sqrt{P} D_W(H_{\Sigma_P}, H) \to 0, \quad \textit{as } P \to \infty,$$

where $D_W(F_1, F_2)$ denotes the Wasserstein distance between distributions F_1 and F_2, defined as

$$D_W(F_1, F_2) = \sup_f \left\{ \left| \int f\, dF_1 - \int f\, dF_2 \right| : f \textit{ is 1-Lipschitz} \right\}.$$

One trivial example of asymptotic stability is constant: if $\Sigma_P = I_{P\times P}$, then H_{Σ_P} is a δ-function at one. The true Stieltjes transform admits an integral form representation given by

$$m_{\Sigma_P}(-z) \;=\; \int_{\mathbb{R}} \frac{1}{x+z} dH_{\Sigma_P(x)} \tag{C.2}$$

and, hence, under asymptotic stability,

$$\lim_{P\to\infty} m_{\Sigma_P}(-z) \tag{C.3}$$

exists and is given by

$$m(-z) \;=\; \int_{\mathbb{R}} \frac{1}{x+z} dH(x)\,. \tag{C.4}$$

Note that by construction, for $z < 0$, we have that $m(z; c) > 0$.

By Assumption 1 (asymptotic stability of $\Sigma_P,\ P \to \infty$), the limits $P^{-1}\operatorname{tr}(\Sigma_P^k),\ k \geq 0$ also exist. We denote these limits by $m_{\Sigma_P}(-z)$ and $E[\lambda_{\Sigma_P}^k]$, respectively:[10]

$$m_{\Sigma_P}(-z) \;=\; \lim_{P\to\infty} m_{\Sigma_P}(-z)\ ,\ E[\lambda_{\Sigma_P}^k] \;=\; \lim_{P\to\infty} P^{-1}\operatorname{tr}(\Sigma_P^k)\,. \tag{C.5}$$

In the classic case of zero complexity, $\hat{\Sigma}_P$ converges to the true covariance matrix of residuals and, hence, so does the Stieltjes transform:

$$\lim_{T,P\to\infty,\ P/T\to 0} \hat{m}(-z) \;=\; m_{\Sigma_P}(-z)\,. \tag{C.6}$$

A cornerstone result of RMT is that when $P/T \to c > 0$, (C.6) needs to be adjusted: $\hat{m}(-z)$ also converges to a deterministic limit that can be expressed in terms of $m_{\Sigma_P}(-z)$.

Theorem C.1 (Bai and Zhou (2008)) *For each $z > 0$,*

$$\lim_{T,P\to\infty, P/T\to c} \hat{m}(-z) \;=\; m(-z; c) \tag{C.7}$$

exists in probability and $m(-z; c)$ is the unique positive solution to the nonlinear master equation

$$m(-z; c) = \frac{1}{1 \;-\; c \;+\; c\, z\, m(-z; c)}\, m_{\Sigma_P}\left(\frac{-z}{1 \;-\; c \;+\; c\, z\, m(-z; c)},\right)\,. \tag{C.8}$$

As a result, the random quantities $\hat{\xi}, \hat{\xi}', \hat{\Delta}$ converge to their theoretical (constant) counter-

[10] Here, $P^{-1}\operatorname{tr}(\Sigma_P^k) = P^{-1}\sum_{i=1}^{P} \lambda_i(\Sigma_P)^k \to E[\lambda_{\Sigma_P}^k]$.

parts:

$$\begin{aligned}
\hat{\xi}(z;c) &\to \xi(z;c) = (1-c+czm(-z;c))^{-1} - 1 \\
\hat{\xi}'(z;c) &\to \xi'(z;c) = \left(\frac{1}{1-c(1-zm(-z;c))}\right)^2 c(zm'(-z;c) - m(-z;c)) \\
\hat{\Delta}(z;c) &\to \Delta(z;c) = 2(\xi(z;c) + z\xi'(z;c))(1+\xi(z;c))^2
\end{aligned} \tag{C.9}$$

The Master equation (C.8) is a non-linear fixed-point equation. This equation admits a unique positive solution for $z > 0$, that can be found using Newton's method described in Lemma E.3 in the appendix. In some special cases, this equation can be simplified or solved analytically. Here, we consider two such special cases: The case $\Sigma_P = I$ and the case when Σ_P has eigenvalues that are uniformly distributed on $[1, 2]$.

Corollary C.2 *Let $z > 0$. If $\Sigma_P = I_P$, $m(-z;c)$ is given by the Marčenko and Pastur (1967) formula*

$$m(-z;c) = \frac{1-c+z-\sqrt{(1-c+z)^2+4cz}}{-2cz} = \frac{2}{1-c+z+\sqrt{(1-c+z)^2+4cz}}. \tag{C.10}$$

If the eigenvalues of Σ_P are uniformly distributed on $[1, 2]$, then $m(z;c)$ is the unique positive solution to the non-linear equation

$$m(-z;c) = \frac{1}{1-c+czm(-z;c)} \ln\left(\frac{2+\frac{z}{1-c+czm(-z;c)}}{1+\frac{z}{1-c+czm(-z;c)}}\right). \tag{C.11}$$

Many calculations in RMT are based on the following simple concentration lemma for quadratic forms (see, e.g., Kelly et al. (2024a); Didisheim et al. (2024)).

Lemma C.1 (Concentration for Quadratic Forms) *If $\gamma = (\gamma_i)_{i=1}^P$ with γ_i being i.i.d., with $E[\gamma_i] = 0$, $E[\gamma_i^2] = 1$, $E[\gamma_i^4] < \infty$, then*

$$P^{-1}\gamma' A_P \gamma - P^{-1}\operatorname{tr}(A_P) \to 0 \tag{C.12}$$

in probability.

Lemma C.1 implies that expressions involving quadratic forms in high-dimensional vectors tend to converge to non-random limits, and these limits can often we expressed in terms of traces. For example, it explains the identity of the two limits in (36). Below, we will frequently use the approximate equivalence of traces and quadratic forms in high dimensions.

Under Assumption 2, it is possible to show that the random variable $\alpha' \hat{\Sigma}_P(z)^{-1} \alpha$ converges to a non-random limit that can be analytically expressed in terms of Ω_α and Σ_P. To prove this convergence, we need to introduce notation and results from RMT. Let

$$Z_*(z;c) \;=\; \frac{z}{1 \;-\; c \;+\; c\, z\, m(z;c)} \;=\; z\,(1+\xi(z;c))\,. \tag{C.13}$$

Using $Z_*(z;c)$, one can rewrite (C.8) as

$$z\, m(-z;c) \;=\; Z_*(z;c) m_{\Sigma_P}\left(-Z_*(z;c)\right)\,. \tag{C.14}$$

Didisheim et al. (2024) show that Z_* always satisfies $Z_* > z$ and refer to it as the *effective shrinkage.* Formula (C.14) implies that complexity affects the covariance inverse through a form of *implicit regularization,* whereby, for $c > 0$, we have

$$\lim_{P,T\to\infty, P/T\to c} z(zI + \underbrace{\hat{\Sigma}_P}_{random})^{-1} \;=\; \lim_{P\to\infty} Z_*(Z_* I + \underbrace{\Sigma_P}_{deterministic})^{-1}\,. \tag{C.15}$$

That is, inverting a ridge-penalized sample covariance matrix is like inverting the true covariance matrix with a larger ridge penalty. This intuition can be made rigorous in the following sense. Recall that, by Assumption 2,

$$m_{\Sigma_P}(-z;\Omega_\alpha) \;=\; \lim_{P\to\infty} P^{-1} \operatorname{tr}(\Omega_\alpha (zI + \Sigma_P)^{-1})\,. \tag{C.16}$$

Then, the results of Bai and Saranadasa (1996); Li et al. (2020b); Liu et al. (2015) imply that the following is true:

$$\lim_{P\to\infty} P^{-1} z \operatorname{tr}(\Omega_\alpha(zI + \underbrace{\hat{\Sigma}_P}_{random})^{-1}) \;=\; \lim_{P\to\infty} P^{-1} \operatorname{tr}(\Omega_\alpha(Z_* I + \underbrace{\Sigma_P}_{deterministic})^{-1}) \tag{C.17}$$

We formalize this important result in the next theorem (which, in turn, implies Theorem 6).

Theorem C.3 (Bai and Saranadasa (1996); Li et al. (2020b); Liu et al. (2015)) *The following limit exists in probability:*

$$\mathcal{H}(z;c) \;=\; \lim_{T,P\to\infty, P/T\to c} \frac{T^{1/2}\alpha'(zI+\hat{\Sigma}_P)^{-1}\,\alpha}{1+\hat{\theta}_M^2} \;=\; \frac{z^{-1}Z_* m_{\Sigma_P}(-Z_*;\Omega_\alpha)}{1+\hat{\theta}_M^2}\,, \tag{C.18}$$

where m_{Σ_P} is defined in (36).

Formula (C.15) implies that a key determinant of the high-complexity behavior is the effective shrinkage parameter Z_*. As we show in the Appendix G, Lemma G.1, it satisfies

$$Z_*(z;c) \;=\; c\,(P^{-1}\operatorname{tr}(\Sigma_P)) \;+\; O(1) \tag{C.19}$$

as $c \to \infty$. That is, Z_* increases indefinitely with c and, as a result, the regularized inverse gets dominated by Z_*:

$$\begin{aligned}\lim_{P,T\to\infty, P/T\to c} Z_* m_{\Sigma_P}(-Z_*;\Omega_\alpha) \;&=\; \lim_{P,T\to\infty, P/T\to c} Z_* P^{-1}\operatorname{tr}(\Omega_\alpha(Z_* I+\Sigma_P)^{-1}) \;=\; P^{-1}\operatorname{tr}(\Omega_\alpha(I+\Sigma_P/Z_*)^{-1}) \\ =\; \lim_{P,T\to\infty, P/T\to c} P^{-1}\operatorname{tr}(\Omega_\alpha) \;&=\; \lim_{P,T\to\infty, P/T\to c} P^{-1}T^{1/2}E[\|\alpha\|^2]\,.\end{aligned} \tag{C.20}$$

Similarly, Appendix G, Lemma G.2, implies that

$$\Delta(z;c) \;=\; 2cz^{-2}(P^{-1}\operatorname{tr}(\Sigma_P^2)) \;+\; O(1)\,. \tag{C.21}$$

That is, by Theorem 7, the variance of the GRS statistic $\hat{W}(z)$ grows with c. Substituting these expressions into the expressions in Theorem 7, we get Proposition 9.

Proof of Proposition 8. The claim follows from the general results in Li et al. (2020a). □

D Proof of the Main Result (And an Extension to Multiple Factors)

Proof of Propositions 3 and 4. These propositions are classic results in RMT and can be found, for example, in Li et al. (2020a); Kelly et al. (2024a). □

We now state the following important general result from (Li et al., 2020a).

Let $X \in \mathbb{R}^{k \times T}$ be a design matrix of rank k. Let also R

$$R \;=\; BX \;+\; \Sigma_P^{1/2} u, \tag{D.1}$$

where $R = (R_t)_{t=1}^T \in \mathbb{R}^{P \times T}$ are returns and $u = (u_t)$ are residuals satisfying Assumption 1. Define residuals

$$\hat{\alpha} \;=\; R(I - X'(XX')^{-1}X) \;=\; R \;-\; \hat{B}X, \tag{D.2}$$

where

$$\hat{B} \;=\; RX'(XX')^{-1} \tag{D.3}$$

and

$$\hat{\Sigma}_P \;=\; \frac{1}{T}\sum_{t=1}^{T} \hat{\alpha}_t \hat{\alpha}_t' \;=\; T^{-1}\hat{\alpha}\hat{\alpha}' \;=\; T^{-1}R(I - X'(XX')^{-1}X)R', \tag{D.4}$$

and

$$\hat{\Sigma}_P(z) = zI + \hat{\Sigma}_P, \tag{D.5}$$

where we have used that $(I - X'(XX')^{-1}X)$ is a projection,

$$(I - X'(XX')^{-1}X) \;=\; (I - X'(XX')^{-1}X)^2\,. \tag{D.6}$$

We also fix a matrix $C \in k \times q$ defining the set of q constraints: The goal of (Li et al., 2020a) is to jointly test q hypotheses

$$BC \;=\; 0\,. \tag{D.7}$$

Intuitively, we would like to define a test statistic in terms of the "size" of $\hat{B}C$, and testing $BC = 0$ can be done by testing whether $\hat{B}C$ is small. Define

$$\mathcal{T} \;=\; (C'(XX')^{-1}C)^{-1/2}\,. \tag{D.8}$$

This is a simple normalization, controlling for the size of X and C. To this end, (Li et al., 2020a) define the test statistic

$$\begin{aligned}
\hat{M}(z) \;&=\; \frac{1}{T}(\hat{B}C\mathcal{T})'\hat{\Sigma}_P(z)^{-1}(\hat{B}C\mathcal{T}) \\
&=\; \frac{1}{T}\mathcal{T}C'\hat{B}'\hat{\Sigma}_P(z)^{-1}\hat{B}C\mathcal{T} \\
&=\; \frac{1}{T}\mathcal{T}C'(XX')^{-1}XR'\hat{\Sigma}_P(z)^{-1}RX'(XX')^{-1}C\mathcal{T} \;\in\; \mathbb{R}^{q\times q}\,,
\end{aligned} \tag{D.9}$$

which is a $q \times q$ matrix, with q being the number of constraints that are being jointly tested.

An important additional constraint Li et al. (2020a) impose is that $T^{-1}C'(XX')^{-1}C$ be uniformly positive definite: $\liminf_{T\to\infty} \lambda_{\min}(T^{-1}C'(XX')^{-1}C) > 0$.

For $q \geq 1$, denote by $\mathbf{W} = [w_{ij}]_{i,j=1}^{q}$ the *Gaussian Orthogonal Ensemble (GOE)* defined by:

1. $w_{ij} = w_{ji}$;
2. $w_{ii} \sim \mathcal{N}(0,1)$, $w_{ij} \sim \mathcal{N}(0,1/2)$ for $i \neq j$;
3. w_{ij}'s are jointly independent for $1 \leq i \leq j \leq q$.

The following is a consequence of the main result of Li et al. (2020a) (that result is formulated for more general shrinkage functions).

Theorem D.1 *Let W be the Gaussian orthogonal ensemble: We have*

$$\frac{T^{1/2}(\hat{M}(z) - \hat{\xi}(z;c)I_{q\times q})}{\hat{\Delta}(z;c)^{1/2}} \;\rightarrow\; \mathbf{W} \tag{D.10}$$

in the sense of weak convergence.

We now show that Theorem 7 is, in fact, a special case of Theorem D.1. Here, for the sake of completeness, we prove an extension to multiple factors.

Theorem D.2 *Suppose*

$$R \;=\; BR^{M} \;+\; \varepsilon_t, \tag{D.11}$$

with R^M being L-dimensional; otherwise, all other technical conditions of Theorem D.2 hold. Then, $\hat{W}(z)$ has the same asymptotic behavior as in Theorem D.2.

Proof of Theorem D.2. Let

$$C = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{(L+1)\times 1} \tag{D.12}$$

and

$$X \;=\; (\mathbf{1}', R^M) \;\in\; \mathbb{R}^{(L+1)\times T}\,. \tag{D.13}$$

Now, we have

$$\hat{B} = \begin{pmatrix} \bar{\alpha} & \hat{\beta} \end{pmatrix} = RX'(XX')^{-1} \;\in\; \mathbb{R}^{P\times(L+1)}\,, \tag{D.14}$$

where $\hat{\beta}$ are estimated betas with respect to the factors.

Let us define $\bar{R}^M = T^{-1}\sum_{t=1}^{T} R_t^M \in \mathbb{R}^{L\times 1}$. Then,

$$\begin{aligned} XX' &= \begin{pmatrix} \mathbf{1}_T' \\ R^M \end{pmatrix} \begin{pmatrix} \mathbf{1}_T & (R^M)' \end{pmatrix} \\ &= \begin{pmatrix} T & T(\bar{R}^M)' \\ T\bar{R}^M & R^M(R^M)' \end{pmatrix} \end{aligned} \tag{D.15}$$

The formula for the inverse of a block matrix

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

is given by:

$$M^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BS^{-1}CA^{-1} & -A^{-1}BS^{-1} \\ -S^{-1}CA^{-1} & S^{-1} \end{pmatrix}, \tag{D.16}$$

where $S = D - CA^{-1}B$ is the Schur complement of A in M.

Using this formula and defining $T\hat{\Sigma}_M = R^M(R^M)' - T\bar{R}^M(\bar{R}^M)'$, we obtain

$$(XX')^{-1} = \begin{pmatrix} T^{-1}(1 + (\bar{R}^M)'\hat{\Sigma}_M^{-1}\bar{R}^M) & -T^{-1}(\bar{R}^M)'\hat{\Sigma}_M^{-1} \\ -T^{-1}\hat{\Sigma}_M^{-1}\bar{R}^M & T^{-1}\hat{\Sigma}_M^{-1} \end{pmatrix} \tag{D.17}$$

Letting $\bar{R} = T^{-1}\sum_{t=1}^{T} R_t \in \mathbb{R}^{P\times 1}$, we get that (D.14) implies that

$$\begin{aligned} \bar{\alpha} &= \begin{pmatrix} R\mathbf{1}_T & R(R^M)' \end{pmatrix} \begin{pmatrix} T^{-1}(1 + (\bar{R}^M)'\hat{\Sigma}_M^{-1}\bar{R}^M) \\ -T^{-1}\hat{\Sigma}_M^{-1}\bar{R}^M \end{pmatrix} \\ &= \left(\bar{R}(1 + (\bar{R}^M)'\hat{\Sigma}_M^{-1}\bar{R}^M) - T^{-1}R(R^M)'\hat{\Sigma}_M^{-1}\bar{R}^M\right). \end{aligned} \tag{D.18}$$

Similarly,

$$\begin{aligned} \hat{\beta} &= \begin{pmatrix} R\mathbf{1}_T & R(R^M)' \end{pmatrix} \begin{pmatrix} -T^{-1}(\bar{R}^M)'\hat{\Sigma}_M^{-1} \\ T^{-1}\hat{\Sigma}_M^{-1} \end{pmatrix} \\ &= \left(T^{-1}R(R^M)'\hat{\Sigma}_M^{-1} - \bar{R}(\bar{R}^M)'\hat{\Sigma}_M^{-1}\right) \end{aligned} \tag{D.19}$$

Let

$$\hat{\Sigma}_P = T^{-1}\left(RR' - \hat{B}XR'\right) \tag{D.20}$$

By (D.12),

$$
\begin{aligned}
RX'(XX')^{-1}C &= \bar{\alpha} \\
C'(XX')^{-1}C &= T^{-1}(1+(\bar{R}^M)'\hat{\Sigma}_M^{-1}\bar{R}^M) \\
\mathcal{T} &= (C'(XX')^{-1}C)^{-1/2} = (T^{-1}(1+(\bar{R}^M)'\hat{\Sigma}_M^{-1}\bar{R}^M))^{-1/2}
\end{aligned}
\tag{D.21}
$$

Thus, by (D.9),

$$
\begin{aligned}
\hat{M}(z) &= \frac{1}{T}(\hat{B}C\mathcal{T})'\hat{\Sigma}_P(z)^{-1}(\hat{B}C\mathcal{T}) \\
&= \frac{1}{T}\mathcal{T}C'\hat{B}'\hat{\Sigma}_P(z)^{-1}\hat{B}C\mathcal{T} \\
&= \frac{1}{(1+(\bar{R}^M)'\hat{\Sigma}_M^{-1}\bar{R}^M)}\bar{\alpha}'\hat{\Sigma}_P(z)^{-1}\bar{\alpha} \\
&= \hat{W}(z)\,.
\end{aligned}
\tag{D.22}
$$

The proof of Theorem D.2 is complete. □

D.1 Re-expressing The Test Statistic

Let

$$
\tilde{\Sigma}_P = \frac{1}{T}RR' \tag{D.23}
$$

and note that

$$
\hat{\Sigma}_P = \tilde{\Sigma}_P - \frac{1}{T}R(X'(XX')^{-1}X)R'\,. \tag{D.24}
$$

It is convenient to define

$$
\begin{aligned}
U &= X'(XX')^{-1/2} \\
V &= (XX')^{-1/2}C\mathcal{T}\,.
\end{aligned}
\tag{D.25}
$$

A key observation is that $\mathcal{T}$ (the normalization factor) is the symmetric matrix uniquely defined by the condition that

$$V'V \;=\; I \;\Leftrightarrow\; \mathcal{T}(C'(XX')^{-1}C)\mathcal{T} \;=\; I \;\Leftrightarrow\; \mathcal{T}^{-2} = C'(XX')^{-1}C\,. \tag{D.26}$$

Then, we can rewrite

$$\hat{\Sigma}_P \;=\; \tilde{\Sigma}_P \;-\; \frac{1}{T}RUU'R' \tag{D.27}$$

and, by (D.9), we have

$$\begin{aligned}\hat{M}(z) \;&=\; \frac{1}{T}\mathcal{T}C'(XX')^{-1}XR'\hat{\Sigma}_P(z)^{-1}RX'(XX')^{-1}C\mathcal{T}\\ &=\; \frac{1}{T}\mathcal{T}C'(XX')^{-1/2}(XX')^{-1/2}XR'\hat{\Sigma}_P(z)^{-1}RX'(XX')^{-1/2}(XX')^{-1/2}C\mathcal{T}\\ &=\frac{1}{T}V'U'R'\hat{\Sigma}_P(z)^{-1}RUV\,.\end{aligned} \tag{D.28}$$

Dealing with $\hat{\Sigma}_P$ directly is inconvenient because the matrix U inside $\hat{\Sigma}_P$ breaks the independence structure of returns across t. We start with the following auxiliary result.

Lemma D.1 *We have*

$$\frac{1}{T}U'R'\hat{\Sigma}_P(z)^{-1}RU \;=\; \frac{1}{T}U'R'\tilde{\Sigma}_\alpha(z)^{-1}RU(I-\frac{1}{T}U'R'\tilde{\Sigma}_\alpha(z)^{-1}RU)^{-1} \tag{D.29}$$

Proof of Lemma D.1. Let $Y = RU$. The desired identity is equivalent to

$$T^{-1}Y'(zI+T^{-1}RR'-\frac{1}{T}YY')^{-1}Y \;=\; T^{-1}Y'(zI+T^{-1}RR')^{-1}Y(I-T^{-1}Y'(zI+T^{-1}RR')^{-1}Y)^{-1}\,. \tag{D.30}$$

One can see that this is a multi-dimensional extension of the Sherman-Morrison formula and coincides with it when Y is one-dimensional. We will need the following lemma

Lemma D.2

$$(A-B)^{-1} \;=\; A^{-1} \;+\; A^{-1}B(A-B)^{-1} \tag{D.31}$$

Proof of Lemma D.2. multiplying by $A-B$ from the left and by A from the right, we get

$$A \;=\; A-B+B$$

□

Thus,

$$\begin{aligned}
&T^{-1}Y'(zI+T^{-1}RR'-\frac{1}{T}YY')^{-1}Y \\
&= T^{-1}Y'((zI+T^{-1}RR')^{-1}+(zI+T^{-1}RR')^{-1}\frac{1}{T}YY'(zI+T^{-1}RR'-\frac{1}{T}YY')^{-1})Y \\
&= T^{-1}Y'(zI+T^{-1}RR')^{-1}Y+T^{-1}Y'(zI+T^{-1}RR')^{-1}\frac{1}{T}YY'(zI+T^{-1}RR'-\frac{1}{T}YY')^{-1}Y\,.
\end{aligned} \tag{D.32}$$

Let $Q_1 = T^{-1}Y'(zI+T^{-1}RR'-\frac{1}{T}YY')^{-1}Y,\ Q_2 = T^{-1}Y'(zI+T^{-1}RR')^{-1}Y$. Then, we can rewrite this identity as

$$Q_1 \;=\; Q_2+Q_2Q_1 \;\Leftrightarrow\; Q_1 \;=\; Q_2(I-Q_2)^{-1}\,. \tag{D.33}$$

The proof is complete. □

Corollary D.3 *Let*

$$\begin{aligned} \mathcal{R}(z) &= (zI + T^{-1}R'R)^{-1}R'R \in \mathbb{R}^{T\times T}, \\ U &= X'(XX')^{-1/2} \\ V &= (XX')^{-1/2}C\mathcal{T}, \end{aligned} \tag{D.34}$$

and

$$\mathcal{D}(z) = \frac{1}{T}U'\mathcal{R}(z)U\,. \tag{D.35}$$

Then,

$$\hat{M}(z) = V'\mathcal{D}(z)(I - \mathcal{D}(z))^{-1}V\,. \tag{D.36}$$

Proof of Corollary D.3. By (D.28),

$$\hat{M}(z) = \frac{1}{T}V'U'R'\hat{\Sigma}_P(z)^{-1}RUV\,. \tag{D.37}$$

Using Lemma D.1, we thus get

$$\begin{aligned} \hat{M}(z) &= \frac{1}{T}V'U'R'\hat{\Sigma}_P(z)^{-1}RUV \\ &= \frac{1}{T}V'U'R'\tilde{\Sigma}_\alpha(z)^{-1}RU(I - \frac{1}{T}U'R'\tilde{\Sigma}_\alpha(z)^{-1}RU)^{-1}V\,. \end{aligned} \tag{D.38}$$

By the identity

$$R'\tilde{\Sigma}_\alpha(z)^{-1}R = (zI + T^{-1}R'R)^{-1}R'R = \mathcal{R}(z), \tag{D.39}$$

we can rewrite it as

$$\begin{aligned}\hat{M}(z) &= \frac{1}{T}V'U'\mathcal{R}(z)U(I-\frac{1}{T}U'\mathcal{R}(z)U)^{-1}V \\ &= V'\mathcal{D}(z)(I-\mathcal{D}(z))^{-1}V\,.\end{aligned} \tag{D.40}$$

□

Corollary D.4 *Let* $L=1$ *and*

$$X = (\mathbf{1}', R^M) \in \mathbb{R}^{(2)\times T}\,. \tag{D.41}$$

Let also

$$\begin{aligned}\mathcal{R}(z) &= (zI+T^{-1}R'R)^{-1}R'R \in \mathbb{R}^{T\times T}, \\ U &= X'(XX')^{-1/2} \in \mathbb{R}^{T\times 2} \\ V &= (XX')^{-1/2}\begin{pmatrix}1\\0\end{pmatrix}\frac{1}{T^{-1/2}(1+\hat{\theta}_M^2)^{1/2}}\end{aligned} \tag{D.42}$$

and

$$\mathcal{D}(z) = \frac{1}{T}U'\mathcal{R}(z)U\,. \tag{D.43}$$

Then,

$$\hat{M}(z) = V'\mathcal{D}(z)(I-\mathcal{D}(z))^{-1}V\,. \tag{D.44}$$

Proof of Corollary D.4. By (D.28),

$$\hat{M}(z) = \frac{1}{T}V'U'R'\hat{\Sigma}_P(z)^{-1}RUV\,. \tag{D.45}$$

Using Lemma D.1, we thus get

$$\begin{aligned}\hat{M}(z) &= \frac{1}{T}V'U'R'\hat{\Sigma}_P(z)^{-1}RUV \\ &= \frac{1}{T}V'U'R'\tilde{\Sigma}_\alpha(z)^{-1}RU(I-\frac{1}{T}U'R'\tilde{\Sigma}_\alpha(z)^{-1}RU)^{-1}V\,.\end{aligned} \tag{D.46}$$

By the identity

$$R'\tilde{\Sigma}_\alpha(z)^{-1}R \;=\; (zI+T^{-1}R'R)^{-1}R'R \;=\; \mathcal{R}(z), \tag{D.47}$$

we can rewrite it as

$$\begin{aligned}\hat{M}(z) &= \frac{1}{T}V'U'\mathcal{R}(z)U(I-\frac{1}{T}U'\mathcal{R}(z)U)^{-1}V \\ &= V'\mathcal{D}(z)(I-\mathcal{D}(z))^{-1}\,V\,.\end{aligned} \tag{D.48}$$

□

Lemma D.3 *We have*

$$R^{OOS}_{\alpha,t+1}(z) \;=\; q'(zI+T^{-1}R'R)^{-1}U(I-\mathcal{D}(z))^{-1}(XX')^{-1/2}C, \tag{D.49}$$

where

$$q' \;=\; (\hat{\alpha}^{OOS}_{t+1})'R \in\; \mathbb{R}^T\,. \tag{D.50}$$

Proof of Lemma D.3. We have

$$R^{OOS}_{\alpha,t+1}(z) \;=\; (\hat{\alpha}^{OOS}_{t+1})'\hat{\Sigma}_{P,[t-T,t]}(z)^{-1}\bar{\alpha}_{[t-T,t]}\,, \tag{D.51}$$

Let $Y = RU$. Similarly to (D.32), we have

$$\begin{aligned}
&(zI + T^{-1}RR' - \frac{1}{T}YY')^{-1}Y \\
&= ((zI + T^{-1}RR')^{-1} + (zI + T^{-1}RR')^{-1}\frac{1}{T}YY'(zI + T^{-1}RR' - \frac{1}{T}YY')^{-1})Y \\
&= (zI + T^{-1}RR')^{-1}Y + (zI + T^{-1}RR')^{-1}\frac{1}{T}YY'(zI + T^{-1}RR' - \frac{1}{T}YY')^{-1}Y\,,
\end{aligned} \tag{D.52}$$

so that

$$\begin{aligned}
&(I - (zI + T^{-1}RR')^{-1}\frac{1}{T}YY')(zI + T^{-1}RR' - \frac{1}{T}YY')^{-1}Y \\
&= (zI + T^{-1}RR')^{-1}Y,
\end{aligned} \tag{D.53}$$

implying that

$$\begin{aligned}
&(zI + T^{-1}RR' - \frac{1}{T}YY')^{-1}Y \\
&= (I - (zI + T^{-1}RR')^{-1}\frac{1}{T}YY')^{-1}(zI + T^{-1}RR')^{-1}Y.
\end{aligned} \tag{D.54}$$

We will now use the simple identity

$$(I - AB)^{-1}A \;=\; A(I - BA)^{-1} \tag{D.55}$$

with $A = (zI + T^{-1}RR')^{-1}Y,\ B = T^{-1}Y'$ to get

$$\begin{aligned}
&(I - (zI + T^{-1}RR')^{-1}\frac{1}{T}YY')^{-1}(zI + T^{-1}RR')^{-1}Y \\
&= (zI + T^{-1}RR')^{-1}Y(I - Y'(zI + T^{-1}RR')^{-1}\frac{1}{T}Y)^{-1}\,.
\end{aligned} \tag{D.56}$$

We thus have, with $C = (1, 0, \cdots, 0)$, that, with $U = X'(XX')^{-1/2}$,

$$\begin{aligned}
\hat{\Sigma}_{P,[t-T,t]}(z)^{-1}\bar{\alpha}_{[t-T,t]} &= \hat{\Sigma}_P(z)^{-1}RX'(XX')^{-1}C \\
&= (zI + T^{-1}RR' - \frac{1}{T}YY')^{-1}Y(XX')^{-1/2}C \\
&= (zI + T^{-1}RR')^{-1}Y(I - Y'(zI + T^{-1}RR')^{-1}\frac{1}{T}Y)^{-1}(XX')^{-1/2}C \\
&= (zI + T^{-1}RR')^{-1}RU(I - \mathcal{D}(z))^{-1}(XX')^{-1/2}C \\
&= R(zI + T^{-1}R'R)^{-1}U(I - \mathcal{D}(z))^{-1}(XX')^{-1/2}C,
\end{aligned} \tag{D.57}$$

The proof is complete. □

D.2 Size Distortions

Proposition D.5 *Suppose that the distribution of ε_t is symmetric: ε_t and $-\varepsilon_t$ have the same distribution. Then, under null $\alpha = 0$, we have*

$$E[\hat{W}(z)] \geq I\frac{E[\hat{\xi}_t(1+\hat{\xi}_t)^{-1}]}{E[(1+\hat{\xi}_t)^{-1}]}, \tag{D.58}$$

where

$$\hat{\xi}_t = T^{-1}R_t'(zI + T^{-1}\hat{\Sigma}_{T,t})^{-1}R_t\,. \tag{D.59}$$

Proof of Proposition D.5. We have

$$E[\hat{W}(z)] = E[V'\mathcal{D}(z)(I - \mathcal{D}(z))^{-1}V] \geq V'E[\mathcal{D}(z)](I - E[\mathcal{D}(z)])^{-1}V, \tag{D.60}$$

and

$$E[\mathcal{D}(z)] = IE[\hat{\xi}_t(1+\hat{\xi}_t)^{-1}]\,. \tag{D.61}$$

Indeed, the inequality

$$E[V'\mathcal{D}(z)(I-\mathcal{D}(z))^{-1}V] \;\geq\; V'E[\mathcal{D}(z)](I-E[\mathcal{D}(z)])^{-1}V$$

follows directly from the classic matrix Jensen inequality and the matrix convexity of the function $x/(1-x)$ on $[0,1]$. Then, we have

$$E[\mathcal{D}(z)] \;=\; T^{-1}E[U'R'(zI+T^{-1}RR')^{-1}RU]\,. \tag{D.62}$$

To compute this expectation, we note that, for any $a, b \in \mathbb{R}^T$, we will study the following object using the standard Sherman-Morrison transformations:

$$\begin{aligned}
G(z) \;&=\; T^{-1}E[a'R'(zI+T^{-1}RR')^{-1}Rb] \\
&=\; T^{-1}E[\sum_{t_1,t_2} a_{t_1}b_{t_2}R'_{t_1}(zI+T^{-1}RR')^{-1}R_{t_2}] \\
&=\; T^{-1}E[\sum_{t_1\neq t_2} a_{t_1}b_{t_2}R'_{t_1}(zI+T^{-1}\hat{\Sigma}_{T,t_1,t_2})^{-1}R_{t_2}(1+\hat{\xi}_{t_1})^{-1}(1+\hat{\xi}_{t_2})^{-1}] \\
&+\; T^{-1}E[\sum_{t} a_t b_t R'_t(zI+T^{-1}\hat{\Sigma}_{T,t})^{-1}R_t(1+\hat{\xi}_t)^{-1}]
\end{aligned} \tag{D.63}$$

where we have defined

$$\begin{aligned}
\hat{\xi}_{t_1} \;&=\; T^{-1}R'_{t_1}(zI+T^{-1}\hat{\Sigma}_{T,t_1})^{-1}R_{t_1} \\
\hat{\xi}_{t_2} \;&=\; T^{-1}R'_{t_2}(zI+T^{-1}\hat{\Sigma}_{T,t_1,t_2})^{-1}R_{t_2} \\
\hat{\Sigma}_{T,t} \;&=\; T^{-1}(RR'-R_{t_1}R'_{t_1}) \\
\hat{\Sigma}_{T,t} \;&=\; T^{-1}(RR'-R_{t_1}R'_{t_1}-R_{t_2}R'_{t_2})\,.
\end{aligned} \tag{D.64}$$

By the assumed symmetry, we have that the terms with $t_1 \neq t_2$ vanish, and we get

$$\begin{aligned} G(z) &= T^{-1}E[\sum_t a_t b_t R_t'(zI + T^{-1}\hat{\Sigma}_{T,t})^{-1}R_t(1+\hat{\xi}_t)^{-1}] = E[\sum_t a_t b_t \hat{\xi}_t(1+\hat{\xi}_t)^{-1}] \\ &= a'bE[\hat{\xi}_t(1+\hat{\xi}_t)^{-1}], \end{aligned} \tag{D.65}$$

and the claim follows from $U'U = I$. □

E Asymptotic Stieltjes Transform

The following lemma will be useful.

Lemma E.1 *Let $z < 0$ and $c > 0$. Define*

$$\tilde{m}(z;c) = cm(z;c) - (1-c)z^{-1} \tag{E.1}$$

Then we have,

$$\tilde{m}(z;c) > 0 \tag{E.2}$$

Proof of Lemma E.1. We have, for $z < 0$,

$$\tilde{m}(z;c) = \lim_{P\to\infty, T\to\infty, P/T\to c} cP^{-1}\operatorname{tr}((\hat{\Sigma}_P - zI)^{-1}) - (1-c)z^{-1} \tag{E.3}$$

When $c < 1$, we have $-(1-c)z^{-1} > 0$ and hence the result immediately follows. When $c \geq 1$, we have

$$\begin{aligned} \tilde{m}(z;c) &= \lim_{P\to\infty, T\to\infty, P/T\to c} cP^{-1}\operatorname{tr}((\hat{\Sigma}_P - zI)^{-1}) - (1-c)z^{-1} \\ &= \lim_{P\to\infty, T\to\infty, P/T\to c} cP^{-1}\sum_{i=1}^{P}\frac{1}{\lambda_i(\hat{\Sigma}_P) - z} - (1-c)z^{-1} \end{aligned} \tag{E.4}$$

Now, note that when $c \geq 1$, we have $P > T$; hence, $\hat{\Sigma}_P$ has $P - T$ zero eigenvalues. Let us sort the eigenvalues in decreasing order $\lambda_1 \geq ... \geq \lambda_P \geq 0$, where eigenvalue from index $i = T + 1$ to $i = P$ are zero. So,

$$\begin{aligned}
&\lim_{P\to\infty,T\to\infty,P/T\to c} cP^{-1}\sum_{i=1}^{P}\frac{1}{\lambda_i(\hat{\Sigma}_P)-z} - (1-c)z^{-1} \\
&= \lim_{P\to\infty,T\to\infty,P/T\to c} cP^{-1}\sum_{i=1}^{T}\frac{1}{\lambda_i(\hat{\Sigma}_P)-z} + cP^{-1}\sum_{i=T+1}^{P}\frac{1}{\underbrace{\lambda_i(\hat{\Sigma}_P)}_{=0}-z} - (1-c)z^{-1} \\
&= \lim_{P\to\infty,T\to\infty,P/T\to c} cP^{-1}\sum_{i=1}^{T}\frac{1}{\lambda_i(\hat{\Sigma}_P)-z} - cP^{-1}(P-T)\frac{1}{z} - (1-c)z^{-1} \\
&= \lim_{P\to\infty,T\to\infty,P/T\to c} cP^{-1}\sum_{i=1}^{T}\frac{1}{\lambda_i(\hat{\Sigma}_P)-z} - c\frac{1}{z} + \frac{1}{z} - (1-c)z^{-1} \\
&= \lim_{P\to\infty,T\to\infty,P/T\to c} cP^{-1}\sum_{i=1}^{T}\frac{1}{\lambda_i(\hat{\Sigma}_P)-z} > 0\,.
\end{aligned} \tag{E.5}$$

Hence,

$$\tilde{m}(z;c) > 0 \tag{E.6}$$

□

Lemma E.2 *Let $z < 0$ and $c > 0$. Define*

$$\tilde{m}'(z;c) = cm'(z;c) + (1-c)z^{-2} \tag{E.7}$$

Then

$$\tilde{m}'(z;c) > 0 \tag{E.8}$$

Proof of Lemma E.2. By definition of the Stieltjes transform, for $z < 0$, we immediately

have that $m'(z;c) > 0$. So, for $c < 1$, it immediately follows that $\tilde{m}'(z;c) > 0$ as well. When $c \leq 1$, we have, by the Vitali Theorem, that we can differentiate the limit (E.4) and get

$$\tilde{m}'(z;c) = \lim_{P\to\infty, T\to\infty, P/T\to c} cP^{-1} \sum_{i=1}^{P} \frac{1}{(\lambda_i(\hat{\Sigma}_P) - z)^2} + (1-c)z^{-2} \; > \; 0\,. \tag{E.9}$$

Now, note that when $c \geq 1$, we have $P > T$; hence, $\hat{\Sigma}_P$ has $P - T$ zero eigenvalues. Let us sort the eigenvalues in decreasing order $\lambda_1 \geq ... \geq \lambda_P$, where eigenvalue from index $i = T+1$ to $i = P$ are zero. So,

$$\begin{aligned}
&\lim_{P\to\infty, T\to\infty, P/T\to c} cP^{-1} \sum_{i=1}^{P} \frac{1}{(\lambda_i(\hat{\Sigma}_P) - z)^2} + (1-c)z^{-2} \\
&= \lim_{P\to\infty, T\to\infty, P/T\to c} cP^{-1} \sum_{i=1}^{T} \frac{1}{(\lambda_i(\hat{\Sigma}_P) - z)^2} + cP^{-1} \sum_{i=T+1}^{P} \frac{1}{(\underbrace{\lambda_i(\hat{\Sigma}_P)}_{=0} - z)^2} + (1-c)z^{-2} \\
&= \lim_{P\to\infty, T\to\infty, P/T\to c} cP^{-1} \sum_{i=1}^{T} \frac{1}{(\lambda_i(\hat{\Sigma}_P) - z)^2} + cP^{-1}(P-T)\frac{1}{z^2} + (1-c)z^{-2} \\
&= \lim_{P\to\infty, T\to\infty, P/T\to c} cP^{-1} \sum_{i=1}^{T} \frac{1}{(\lambda_i(\hat{\Sigma}_P) - z)^2} > 0\,.
\end{aligned} \tag{E.10}$$

Hence,

$$\tilde{m}'(z;c) > 0 \tag{E.11}$$

□

Theorem E.1 (Implicit Function Theorem) *Let $U \subseteq \mathbb{R}^n \times \mathbb{R}$ be an open neighborhood of a point (g_0, z_0), and let $F : U \to \mathbb{R}$ be a continuously differentiable function, i.e., $F \in C^1(U)$. Suppose that*

$$F(g_0, z_0) = 0$$

and that the partial derivative of F with respect to g at (g_0, z_0) is non-zero:

$$\left.\frac{\partial F}{\partial g}\right|_{(g_0,z_0)} \neq 0.$$

Then, there exists an open neighborhood $V \subseteq \mathbb{R}$ of z_0 and a unique continuously differentiable function $g : V \to \mathbb{R}$ such that

$$g(z_0) = g_0 \quad \textit{and} \quad F(g(z), z) = 0 \quad \textit{for all } z \in V.$$

Moreover, the derivative of g with respect to z is given by

$$g'(z) = -\frac{\partial_z F(g(z), z)}{\partial_g F(g(z), z)} \quad \textit{for all } z \in V.$$

Proof of Theorem E.1. Since F is continuously differentiable in U and

$$\left.\frac{\partial F}{\partial g}\right|_{(g_0,z_0)} \neq 0,$$

by the Implicit Function Theorem, there exists an open neighborhood $V \subseteq \mathbb{R}$ of z_0 and a unique continuously differentiable function $g : V \to \mathbb{R}$ satisfying

$$g(z_0) = g_0$$

and

$$F(g(z), z) = 0 \quad \text{for all } z \in V.$$

To determine the expression for $g'(z)$, differentiate both sides of the equation $F(g(z), z) =$

0 with respect to z:

$$\frac{d}{dz}F(g(z),z) = \frac{\partial F}{\partial g}(g(z),z)\cdot g'(z) + \frac{\partial F}{\partial z}(g(z),z) = 0.$$

Solving for $g'(z)$, we obtain

$$g'(z) = -\frac{\partial_z F(g(z),z)}{\partial_g F(g(z),z)}.$$

This completes the proof. □

Remarks

1. The theorem ensures not only the existence of the function g but also its uniqueness and differentiability within the neighborhood V.

2. The condition

   $$\left.\frac{\partial F}{\partial g}\right|_{(g_0,z_0)} \neq 0$$

   is crucial as it guarantees that the implicit equation $F(g,z) = 0$ can be locally solved for g in terms of z.

3. Higher-order differentiability of F implies higher-order differentiability of the implicit function g.

E.1 Solving the Master Equation

E.1.1 Solving the Master Equation for arbitrary eigenvalue distribution

$m(z;c)$ is the unique positive solution to the nonlinear master equation

$$m(z;c) = \frac{1}{1 \ - \ c \ - \ c\,z\,m(z;c)}\, m\left(\frac{z}{1 \ - \ c \ - \ c\,z\,m(z;c)}\right), \tag{E.12}$$

where

$$m(z) \ = \ \int \frac{dH(x)}{x - z}\,. \tag{E.13}$$

Substituting

$$\begin{aligned} &\tilde{m}(-z;c) = (1-c)z^{-1} \ + \ cm(-z;c) \\ &\Leftrightarrow \ z\tilde{m}(-z;c) \ = \ 1 - c + czm(-z;c) \end{aligned} \tag{E.14}$$

into the Master equation, we get

$$\begin{aligned} &zm(-z;c) = \frac{z}{1 \ - \ c \ + \ c\,z\,m(-z;c)}\, m\left(\frac{z}{1 \ - \ c \ + \ c\,z\,m(-z;c)}\right) \\ &= \ \frac{z}{z\tilde{m}(-z;c)} m(-1/\tilde{m}(-z;c)) \ = \ \frac{1}{\tilde{m}(-z;c)} \int \frac{dH(x)}{x + 1/\tilde{m}(-z;c)}, \end{aligned} \tag{E.15}$$

that is

$$zm(-z;c) \ = \ \int \frac{dH(x)}{\tilde{m}(-z;c)x + 1} \tag{E.16}$$

Rewriting

$$zm(-z;c) \ = \ c^{-1}z\tilde{m}(-z;c) - c^{-1}(1-c) \tag{E.17}$$

and substituting gives

$$c^{-1}z\tilde{m}(-z;c) - c^{-1}(1-c) \;=\; \int \frac{dH(x)}{\tilde{m}(-z;c)x+1}, \tag{E.18}$$

which can be rewritten as

$$c - 1 \;+\; z\tilde{m} - c\int \frac{dH(x)}{(1+\tilde{m}x)} \;=\; 0\,. \tag{E.19}$$

Alternatively, we can also rewrite it as

$$c\int \frac{\tilde{m}xdH(x)}{(1+\tilde{m}x)} - \int \frac{\tilde{m}xdH(x)}{(1+\tilde{m}x)} \;=\; \int \frac{dH(x)}{(1+\tilde{m}x)} - z\tilde{m}, \tag{E.20}$$

which is equivalent to

$$z\tilde{m} + c\int \frac{\tilde{m}xdH(x)}{(1+\tilde{m}x)} = \int \frac{\tilde{m}xdH(x)}{(1+\tilde{m}x)} + \int \frac{dH(x)}{(1+\tilde{m}x)} \;=\; \int \frac{(1+\tilde{m}x)dH(x)}{(1+\tilde{m}x)} \;=\; 1\,. \tag{E.21}$$

We can also rewrite it as an equation for

$$Z_*(z;c) \;=\; 1/\tilde{m}. \tag{E.22}$$

Namely,

$$c^{-1}\tilde{m}(-z;c) - c^{-1}(1-c)z^{-1} \;=\; m \;=\; z^{-1}Z_*\int \frac{dH(x)}{x+Z_*} \tag{E.23}$$

can be rewritten as

$$\begin{aligned} zZ_*(z;c)^{-1} - 1 + c &= Z_* c \int \frac{dH(x)}{x + Z_*} = c \int \frac{Z_* dH(x)}{x + Z_*} = c \int \frac{(-x + x + Z_*) dH(x)}{x + Z_*} \\ &= c - c \int \frac{x dH(x)}{x + Z_*} \\ &\Leftrightarrow zZ_*(z;c)^{-1} + c \int \frac{x dH(x)}{x + Z_*} = 1 \\ &\Leftrightarrow Z_* = z + c \int \frac{x dH(x)}{x + Z_*} . \end{aligned} \tag{E.24}$$

Lemma E.3 *Let* $m(z;c) = c^{-1}\tilde{m}(z;c) + (1-c)z^{-1}c^{-1}, z < 0$ *and*

$$f(\tilde{m}; z, c) = 1 - c + z\tilde{m}(z;c) + c \int \frac{dH(x)}{\tilde{m}x + 1} \tag{E.25}$$

Then, f *has a unique positive root and can be found by Newton's method with* $\tilde{m}_0 = 0$ *and*

$$\tilde{m}_{n+1} = \tilde{m}_n - \frac{f(\tilde{m}_n; z, c)}{f'(\tilde{m}_n; z, c)} . \tag{E.26}$$

Proof of Lemma E.3. We will use $\tilde{m} = \tilde{m}(z;c)$ for notation purposes. We seek to find $\tilde{m}$ such that $f(\tilde{m}; z, c, \Sigma_P) = 0$ as it will give us $m(z;c)$ that solves the Master Equation. By letting $\tilde{m}_0 = 0$ be our initial guess of the root of $f(\tilde{m}; z)$, we can iterate like

$$\tilde{m}_{n+1} = \tilde{m}_n - \frac{f(\tilde{m}_n; z)}{f'(\tilde{m}_n; z)} \tag{E.27}$$

to find the root of the of f. In our case, $f(0) = -1$. If $f(\tilde{m}_n) < 0$, then $\tilde{m}_{n+1} > \tilde{m}_n$. Since f is concave,

$$f(\tilde{m}_{n+1}) < f(\tilde{m}_n) + (\tilde{m}_{n+1} - \tilde{m}_n) f'(\tilde{m}_n) = f(\tilde{m}_n) - \frac{f(\tilde{m}_n; z)}{f'(\tilde{m}_n; z)} f'(\tilde{m}_n) = 0 . \tag{E.28}$$

Thus, we never cross zero and converge monotonically to the unique positive root from below.

□

Lemma E.4 *Let* $m(z;c) = c^{-1}\tilde{m}(z;c) + (1-c)z^{-1}c^{-1}$, $z < 0$, *then*

$$\frac{d\tilde{m}}{dz} = -\frac{\tilde{m}(z;c)}{z - c\int \frac{x\,dH(x)}{(\tilde{m}(z;c)x+1)^2}}. \tag{E.29}$$

Proof of Lemma E.4. **Computation of $\frac{d\tilde{m}}{dz}$ for the Implicit Equation**

Let $\tilde{m}(z;c)$ be the solution to the implicit equation

$$f(\tilde{m};z,c) = 1 - c + z\tilde{m}(z;c) + c\int \frac{dH(x)}{\tilde{m}x+1} = 0.$$

Our goal is to determine the derivative $\frac{d\tilde{m}}{dz}$.

Applying the Implicit Function Theorem

Given that f is continuously differentiable with respect to $\tilde{m}$ and z, and assuming that

$$\frac{\partial f}{\partial \tilde{m}} \neq 0,$$

the Implicit Function Theorem guarantees the existence of a differentiable function $\tilde{m}(z;c)$ in a neighborhood around a point $(\tilde{m}_0, z_0)$ where $f(\tilde{m}_0; z_0, c) = 0$.

Differentiation of the Implicit Equation

Differentiate both sides of the equation $f(\tilde{m};z,c) = 0$ with respect to z:

$$\frac{d}{dz}f(\tilde{m};z,c) = 0.$$

Applying the chain rule, we obtain

$$\frac{\partial f}{\partial z} + \frac{\partial f}{\partial \tilde{m}} \cdot \frac{d\tilde{m}}{dz} = 0.$$

Solving for $\frac{d\tilde{m}}{dz}$ gives

$$\frac{d\tilde{m}}{dz} = -\frac{\frac{\partial f}{\partial z}}{\frac{\partial f}{\partial \tilde{m}}}.$$

Computing the Partial Derivatives

Compute $\frac{\partial f}{\partial z}$:

$$\frac{\partial f}{\partial z} = \tilde{m}(z; c).$$

Compute $\frac{\partial f}{\partial \tilde{m}}$:

$$\frac{\partial f}{\partial \tilde{m}} = z + c \cdot \frac{d}{d\tilde{m}} \left(\int \frac{dH(x)}{\tilde{m}x + 1} \right).$$

Evaluate the derivative inside the integral:

$$\frac{d}{d\tilde{m}} \left(\frac{1}{\tilde{m}x + 1} \right) = -\frac{x}{(\tilde{m}x + 1)^2}.$$

Thus,

$$\frac{\partial f}{\partial \tilde{m}} = z - c \int \frac{x \, dH(x)}{(\tilde{m}x + 1)^2}.$$

Final Expression for $\frac{d\tilde{m}}{dz}$

Substituting the partial derivatives into the expression for $\frac{d\tilde{m}}{dz}$, we obtain

$$\frac{d\tilde{m}}{dz} = -\frac{\tilde{m}(z;c)}{z - c\int \frac{x\, dH(x)}{(\tilde{m}(z;c)x+1)^2}}.$$

This provides the desired derivative of $\tilde{m}$ with respect to z.

Summary

Given the implicit equation

$$1 - c + z\tilde{m}(z;c) + c\int \frac{dH(x)}{\tilde{m}x + 1} = 0,$$

the derivative of $\tilde{m}$ with respect to z is

$$\frac{d\tilde{m}}{dz} = -\frac{\tilde{m}(z;c)}{z - c\int \frac{x\, dH(x)}{(\tilde{m}(z;c)x+1)^2}}.$$

□

E.2 Solving the Master Equation when $\Sigma_P = I$

Lemma E.5 *Let $\Sigma = I_P$ and $z < 0$, we have*

$$m_\Sigma(z) = \int_{\mathbb{R}} \frac{1}{x - z} dH(x) = \frac{1}{1 - z} \tag{E.30}$$

And,

$$m(z;c) = \frac{1 - c - z - \sqrt{(1 - c - z)^2 - 4cz}}{2cz} \tag{E.31}$$

solves the Master Equation. Moreover, we have

$$m'(z;c) = \frac{-\frac{z-c-1}{\sqrt{(1-c-z)^2-4cz}} - 1}{2cz} - \frac{-\sqrt{(1-c-z)^2-4cz} - c - z + 1}{2cz^2} \tag{E.32}$$

Proof of Lemma E.5. The first statement follows from the definition of the Stieltjes transform and the fact that if $\Sigma_P = I_P$, then the eigenvalue distribution of Σ_P has a mass at 1. Then, from the master equation, we have

$$\begin{aligned} m(z;c) &= \frac{1}{1-c-czm(z;c)} m_\Sigma \left(\frac{z}{1-c-czm(z;c)} \right) \\ &= \frac{1}{1-c-czm(z;c)} \frac{1}{1-\frac{z}{1-c-czm(z;c)}} \\ &= \frac{1}{1-c-czm(z;c)-z} \end{aligned} \tag{E.33}$$

Hence,

$$\begin{aligned} & m(z;c) - cm(z;c) - czm(z;c)^2 - zm(z;c) = 1 \\ \implies & czm(z;c)^2 + (z+c-1)m(z;c) + 1 = 0 \end{aligned} \tag{E.34}$$

Solving for $m(z;c)$ gives

$$m_\pm = \frac{1-c-z \pm \sqrt{(1-c-z)^2-4cz}}{2cz} \tag{E.35}$$

Since the Stieltjes transform of an eigenvalue distribution with positive support must be positive and given that $z < 0$ and, hence, $2cz < 0$, we have

$$\sqrt{(1-c-z)-4cz} > 1-c-z \tag{E.36}$$

And therefore, we must keep the negative branch of the solution

$$m(z;c) = \frac{1-c-z-\sqrt{(1-c-z)^2-4cz}}{2cz} \tag{E.37}$$

The result for $m'(z;c)$ follows by differentiating equation (E.37) with respect to z. □

E.3 Solving the Master Equation when $H(\lambda) \sim \mathcal{U}(1,2)$

Lemma E.6 *Let the eigenvalues of Σ_P follow $\mathcal{U}[1,2]$ and $z<0$, then $m(z;c)$ satisfies,*

$$m(z) = \frac{1}{1-c-czm(z;c)} \ln\left(\frac{2-\frac{z}{1-c-czm(z;c)}}{1-\frac{z}{1-c-czm(z;c)}}\right) \tag{E.38}$$

Let $m(z;c) = c^{-1}\tilde{m}(z;c) + (1-c)z^{-1}c^{-1}$, then $\tilde{m}(z;c)$ satisfies

$$c^{-1}\tilde{m}(z;c) + (1-c)z^{-1}c^{-1} + \frac{1}{z\tilde{m}(z;c)} \ln\left(\frac{2\tilde{m}(z;c)+1}{\tilde{m}(z;c)+1}\right) = 0 \tag{E.39}$$

Proof of Lemma E.6. Instead of using the matrix representation Σ_P, we can work directly with $H(\lambda)$, the eigenvalue distribution. We have

$$\begin{aligned} m_\Sigma(z) &= \int_1^2 \frac{1}{x-z} dH(x) \\ &= \int_1^2 \frac{1}{x-z} dx \\ &\underbrace{=}_{u=x-z} \int_{1-z}^{2-z} \frac{1}{u} du \\ &= \ln(2-z) - \ln(1-z) \end{aligned} \tag{E.40}$$

Then, from the master equation, $m(z;c)$ solves

$$\begin{aligned} m(z;c) &= \frac{1}{1-c-czm(z;c)} m_\Sigma\left(\frac{z}{1-c-czm(z;c)}\right) \\ &= \frac{1}{1-c-czm(z;c)} \ln\left(\frac{2-\frac{z}{1-c-czm(z;c)}}{1-\frac{z}{1-c-czm(z;c)}}\right) \end{aligned} \tag{E.41}$$

Let $m(z;c) = c^{-1}\tilde{m}(z;c) + (1-c)z^{-1}c^{-1}$, we have

$$\begin{aligned} c^{-1}\tilde{m}(z;c) + (1-c)z^{-1}c^{-1} &= -\frac{1}{z\tilde{m}(z;c)} \ln\left(\frac{2+\frac{1}{\tilde{m}(z;c)}}{1+\frac{1}{\tilde{m}(z;c)}}\right) \\ &= -\frac{1}{z\tilde{m}(z;c)} \ln\left(\frac{2\tilde{m}(z;c)+1}{\tilde{m}(z;c)+1}\right) \end{aligned} \tag{E.42}$$

Hence,

$$c^{-1}\tilde{m}(z;c) + (1-c)z^{-1}c^{-1} + \frac{1}{z\tilde{m}(z;c)} \ln\left(\frac{2\tilde{m}(z;c)+1}{\tilde{m}(z;c)+1}\right) = 0 \tag{E.43}$$

□

Lemma E.7 *Let*

$$f(\tilde{m};z,c) = c^{-1}\tilde{m}(z;c) + (1-c)z^{-1}c^{-1} + \frac{1}{z\tilde{m}(z;c)} \ln\left(\frac{2\tilde{m}(z;c)+1}{\tilde{m}(z;c)+1}\right) \tag{E.44}$$

where $c > 0, z < 0$. Then, f has a unique positive root that can be found by Newton's method

$$\tilde{m}_{n+1} = \tilde{m}_n - \frac{f(\tilde{m}_n;z,c)}{f'(\tilde{m}_n;z,c)} \tag{E.45}$$

Proof of Lemma E.7. This lemma is a special case of Lemma E.3. We just replaced the sum over eigenvalues of Σ_P by its integral representation since $H(\lambda)$ is available. □

Lemma E.8 *Let the eigenvalues of Σ_P follow $\mathcal{U}[1,2]$ and $z<0$, then $\tilde{m}'(z;c)$ satisfies*

$$\tilde{m}'(z;c) = -\frac{\partial_z f(\tilde{m}(z;c);z,c)}{\partial_{\tilde{m}} f(\tilde{m}(z;c);z,c)} \tag{E.46}$$

Where,

$$\begin{aligned}
\partial_z f(\tilde{m}(z;c);z,c) &= -(1-c)z^{-2}c^{-1} - \frac{1}{z^2\tilde{m}(z;c)}\ln\left(\frac{2\tilde{m}(z;c)+1}{\tilde{m}(z;c)+1}\right) \\
\partial_{\tilde{m}} f(\tilde{m}(z;c);z,c) &= c^{-1} + \frac{1}{z}\frac{\tilde{m}(z;c) - (2\tilde{m}^2(z;c)+3\tilde{m}(z;c)+1)\ln\left(\frac{2\tilde{m}(z;c)+1}{\tilde{m}(z;c)+1}\right)}{\tilde{m}^2(z;c)(2\tilde{m}(z;c)+1)(\tilde{m}(z;c)+1)}
\end{aligned} \tag{E.47}$$

Proof of Lemma E.8. From Lemma E.7, we have

$$f(\tilde{m}(z;c);z,c) = c^{-1}\tilde{m}(z;c) + (1-c)z^{-1}c^{-1} + \frac{1}{z\tilde{m}(z;c)}\ln\left(\frac{2\tilde{m}(z;c)+1}{\tilde{m}(z;c)+1}\right) = 0 \tag{E.48}$$

From the implicit function theorem, we have

$$\tilde{m}'(z;c) = -\frac{\partial_z f(\tilde{m}(z;c);z,c)}{\partial_{\tilde{m}} f(\tilde{m}(z;c);z,c)} \tag{E.49}$$

Where,

$$\begin{aligned}
\partial_z f(\tilde{m}(z;c);z,c) &= -(1-c)z^{-2}c^{-1} - \frac{1}{z^2\tilde{m}(z;c)}\ln\left(\frac{2\tilde{m}(z;c)+1}{\tilde{m}(z;c)+1}\right) \\
\partial_{\tilde{m}} f(\tilde{m}(z;c);z,c) &= c^{-1} + \frac{1}{z}\frac{\tilde{m}(z;c) - (2\tilde{m}^2(z;c)+3\tilde{m}(z;c)+1)\ln\left(\frac{2\tilde{m}(z;c)+1}{\tilde{m}(z;c)+1}\right)}{\tilde{m}^2(z;c)(2\tilde{m}(z;c)+1)(\tilde{m}(z;c)+1)}
\end{aligned} \tag{E.50}$$

□

F Optimal Ridge Penalty

Following Kelly et al. (2024b), we let

$$\begin{aligned}
\hat{\Psi}_T &= \frac{1}{T}\sum_{\tau=1}^{T}\hat{\alpha}_\tau\hat{\alpha}_\tau' \\
\hat{\Psi}_{T,t} &= \hat{\Psi}_T - \frac{1}{T}\hat{\alpha}_t\hat{\alpha}_t' \\
\bar{\alpha}_{T,t} &= \bar{\alpha} - \frac{1}{T}\alpha_t \\
\bar{\Sigma}_{T,t} &= \hat{\Psi}_{T,t} - \bar{\alpha}_{T,t}\bar{\alpha}_{T,t}'\,.
\end{aligned} \tag{F.1}$$

$\bar{\Sigma}_{T,t}$ and $\bar{\alpha}_{T,t}$ are estimated alpha covariance matrix and mean alpha for an investor who does not have access to the observation number t, so that whatever happens at time t is perceived as out-of-sample by this investor (see Kelly et al. (2024b) for details and extensions of this idea).

As a result, we can build an asymptotically unbiased estimator of $\mathcal{H}(z;c)$ using the leave-one-out estimator:

$$\tilde{\mathcal{H}}(z;c) = \frac{1}{T}\sum_{t=1}^{T}\bar{\alpha}_{T,t}(zI+\bar{\Sigma}_{T,t})^{-1}\hat{\alpha}_t \tag{F.2}$$

By the Sherman-Morrison formula,

$$\bar{\alpha}_{T,t}'(zI+\bar{\Sigma}_{T,t})^{-1}\hat{\alpha}_t = \frac{\bar{\alpha}_{T,t}'(zI+\hat{\Psi}_{T,t})^{-1}\hat{\alpha}_t}{1-\bar{\alpha}_{T,t}'(zI+\hat{\Psi}_{T,t})^{-1}\bar{\alpha}_{T,t}} \tag{F.3}$$

For large T, we can approximate

$$\bar{\alpha}_{T,t}'(zI+\hat{\Psi}_{T,t})^{-1}\bar{\alpha}_{T,t} = \bar{\alpha}_T'(zI+\hat{\Psi}_T)^{-1}\bar{\alpha}_T + O(T^{-1})\,, \tag{F.4}$$

whereas

$$\begin{aligned}
\bar{\alpha}'_{T,t}(zI+\hat{\Psi}_{T,t})^{-1}\hat{\alpha}_t &= \bar{\alpha}'_{T,t}(zI+\hat{\Psi}_T)^{-1}\hat{\alpha}_t\frac{1}{1-T^{-1}\hat{\alpha}'_t(zI+\hat{\Psi}_T)^{-1}\hat{\alpha}_t} \\
&= (\bar{\alpha}-T^{-1}\hat{\alpha}_t)(zI+\hat{\Psi}_T)^{-1}\hat{\alpha}_t\frac{1}{1-T^{-1}\hat{\alpha}'_t(zI+\hat{\Psi}_T)^{-1}\hat{\alpha}_t} \\
&= \frac{\bar{W}_t-w_t}{1-w_t},
\end{aligned} \tag{F.5}$$

where

$$\begin{aligned}
w_t &= T^{-1}\hat{\alpha}'_t(zI+\hat{\Psi}_T)^{-1}\hat{\alpha}_t \\
\bar{W}_t &= \bar{\alpha}'(zI+\hat{\Psi}_T)^{-1}\hat{\alpha}_t = \frac{1}{T}\sum_{\tau=1}^{T}\hat{\alpha}'_\tau(zI+\hat{\Psi}_T)^{-1}\hat{\alpha}_t
\end{aligned} \tag{F.6}$$

and we get our estimator

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\bar{\alpha}'_{T,t}(zI+\bar{\Sigma}_{T,t})^{-1}\hat{\alpha}_t \\
&= \frac{1}{T}\sum_{t=1}^{T}\frac{\bar{\alpha}'_{T,t}(zI+\hat{\Psi}_{T,t})^{-1}\hat{\alpha}_t}{1-\bar{\alpha}'_{T,t}(zI+\hat{\Psi}_{T,t})^{-1}\bar{\alpha}_{T,t}} \\
&= \frac{1}{1-\bar{\alpha}'_T(zI+\hat{\Psi}_T)^{-1}\bar{\alpha}_T}\frac{1}{T}\sum_{t=1}^{T}\bar{\alpha}'_{T,t}(zI+\bar{\Sigma}_{T,t})^{-1}\hat{\alpha}_t \\
&= \frac{1}{1-\bar{\alpha}'_T(zI+\hat{\Psi}_T)^{-1}\bar{\alpha}_T}\frac{1}{T}\sum_{t=1}^{T}\frac{\bar{W}_t-w_t}{1-w_t}
\end{aligned} \tag{F.7}$$

Now, let $\hat{\alpha} = (\hat{\alpha}_t)_{t=1}^T \in \mathbb{R}^{P\times T}$. Then, $\hat{\Psi}_T = T^{-1}\hat{\alpha}\hat{\alpha}'$ and, hence,

$$G(z) = T^{-1}\hat{\alpha}'(zI+\hat{\Psi}_T)^{-1}\hat{\alpha} = T^{-1}\hat{\alpha}'(zI+T^{-1}\hat{\alpha}\hat{\alpha}')^{-1}\hat{\alpha} = (zI+T^{-1}\hat{\alpha}'\hat{\alpha})^{-1}T^{-1}\hat{\alpha}'\hat{\alpha} \in \mathbb{R}^{T\times T}. \tag{F.8}$$

Then,

$$w = (w_t)_{t=1}^T \;=\; \mathrm{diag}(G(z)),\; \bar{W} \;=\; (\bar{W}_t)_{t=1}^T = \; G(z)\mathbf{1}\,. \tag{F.9}$$

Similarly,

$$\bar{\alpha}_T'(zI + \hat{\Psi}_T)^{-1}\bar{\alpha}_T \;=\; T^{-2}\mathbf{1}'\hat{\alpha}'(zI + \hat{\Psi}_T)^{-1}\hat{\alpha}\mathbf{1} \;=\; T^{-1}\mathbf{1}'G(z)\mathbf{1} \;=\; T^{-1}\mathbf{1}'\bar{W}\,. \tag{F.10}$$

G Asymptotics for $c \to \infty$

Proposition G.1 *Let $m(z;c) = c^{-1}\tilde{m}(z;c) + (c^{-1}-1)z^{-1}, z < 0$. Let also X be a random variable distributed with $dH(x)$. Then,*

$$\begin{aligned} m(-z;c) \;&=\; (1-\tilde{c})z^{-1} \;+\tilde{c}^2E[X]^{-1} + 0.5\tilde{c}^3(E[X]^{-1}(\frac{E[X^2]}{E[X]^2} - zE[X]^{-1})) + O(\tilde{c}^4) \\ \frac{\partial}{\partial z}(m(-z;c)) \;&=\; -(1-\tilde{c})z^{-2} \;-\tilde{c}^3\tilde{E}[X]^{-2} + O(\tilde{c}^4)\,, \end{aligned} \tag{G.1}$$

where $\tilde{c} = c^{-1}$.

Proof of Proposition G.1. The Master equation for $\tilde{m}(-z;c)$ is

$$z\tilde{m} - 1 + c(1 - \int \frac{dH(x)}{(1+\tilde{m}x)}) \;=\; 0\,. \tag{G.2}$$

Let $\tilde{c} = c^{-1}$. We will use the implicit function theorem to perform Taylor expansion in powers of $\tilde{c}$. Dividing by c, we can rewrite (G.2) as

$$\tilde{c}(z\tilde{m} - 1) + (1 - \int \frac{dH(x)}{(1+\tilde{m}x)}) \;=\; 0, \tag{G.3}$$

implying that $\lim_{\tilde{c}\to 0} \tilde{m} = 0$. Hence,

$$\lim_{c\to\infty} \tilde{m} = 0 \tag{G.4}$$

Moreover, we can write

$$\tilde{m} \;=\; \tilde{c}\tilde{m}_1 \;+\; 0.5\;\tilde{c}^2\,\tilde{m}_2 \;+0.25\tilde{c}^3\tilde{m}_3 +\; O(\tilde{c}^4). \tag{G.5}$$

We will use the fact that

$$\begin{aligned}
&(\tilde{c}\tilde{m}_1 \;+\; 0.5\;\tilde{c}^2\,\tilde{m}_2 \;+0.25\tilde{c}^3\tilde{m}_3 +\; O(\tilde{c}^4))^2 \\
&=\; \tilde{c}^2(\tilde{m}_1 \;+\; 0.5\;\tilde{c}\,\tilde{m}_2 \;+0.25\tilde{c}^2\tilde{m}_3 +\; O(\tilde{c}^3))^2 \\
&=\; \tilde{c}^2(\tilde{m}_1^2 + \tilde{c}\tilde{m}_1\tilde{m}_2 + O(\tilde{c}^2)) \\
&=\; \tilde{c}^2\tilde{m}_1^2 \;+\tilde{c}^3\tilde{m}_1\tilde{m}_2 +\; O(\tilde{c}^4).
\end{aligned} \tag{G.6}$$

And that,

$$\begin{aligned}
&(\tilde{c}\tilde{m}_1 \;+\; 0.5\;\tilde{c}^2\,\tilde{m}_2 \;+0.25\tilde{c}^3\tilde{m}_3 +\; O(\tilde{c}^4))^3 \\
&= \tilde{c}^3(\tilde{m}_1 + 0.5\tilde{c}\tilde{m}_2 + 0.25\tilde{c}^2\tilde{m}_3 + O(\tilde{c}^3))^3 \\
&= \tilde{c}^3(\tilde{m}_1^3 + O(\tilde{c})) \\
&= \tilde{c}^3\tilde{m}_1^3 + O(\tilde{c}^4)
\end{aligned} \tag{G.7}$$

Substituting this into (G.3), we get

$$
\begin{aligned}
0 \;&=\; \tilde{c}(z\tilde{m}-1)+(1-\int \frac{dH(x)}{(1+\tilde{m}x)}) \\
&=\; \tilde{c}(z\tilde{m}-1)+(1-\int(1-\tilde{m}x+\tilde{m}^2x^2-\tilde{m}^3x^3+O(\tilde{c}^4))dH(x)) \\
&=\; \tilde{c}(z(\tilde{c}\tilde{m}_1 \;+\; 0.5\;\tilde{c}^2\,\tilde{m}_2 \;+\; O(\tilde{c}^3))-1) \\
&+\left(1-\int(1-(\tilde{c}\tilde{m}_1 \;+\; 0.5\;\tilde{c}^2\,\tilde{m}_2 \;+0.25\tilde{c}^3\tilde{m}_3+\; O(\tilde{c}^4))x\right. \\
&+(\tilde{c}\tilde{m}_1 \;+\; 0.5\;\tilde{c}^2\,\tilde{m}_2 \;+0.25\tilde{c}^3\tilde{m}_3+\; O(\tilde{c}^4))^2x^2 \\
&\left.-(\tilde{c}\tilde{m}_1 \;+\; 0.5\;\tilde{c}^2\,\tilde{m}_2 \;+0.25\tilde{c}^3\tilde{m}_3+\; O(\tilde{c}^4))^3x^3+O(\tilde{c}^4))dH(x)\right) \\
&=\; \tilde{c}^2z\tilde{m}_1+0.5z\tilde{c}^3\tilde{m}_2+zO(\tilde{c}^4)-\tilde{c} \\
&+\left(\int((\tilde{c}\tilde{m}_1 \;+\; 0.5\;\tilde{c}^2\,\tilde{m}_2 \;+0.25\tilde{c}^3\tilde{m}_3+\; O(\tilde{c}^4))x-(\tilde{c}^2\tilde{m}_1^2 \;+\tilde{c}^3\tilde{m}_1\tilde{m}_2+\; O(\tilde{c}^4))x^2\right. \\
&\left.+(\tilde{c}^3\tilde{m}_1^3+O(\tilde{c}^4))x^3-O(\tilde{c}^4))dH(x)\right) \\
&=\tilde{c}^2z\tilde{m}_1+0.5z\tilde{c}^3\tilde{m}_2+zO(\tilde{c}^4)-\tilde{c} \\
&+(\tilde{c}\tilde{m}_1 \;+\; 0.5\;\tilde{c}^2\,\tilde{m}_2 \;+0.25\tilde{c}^3\tilde{m}_3+\; O(\tilde{c}^4))E[X]-(\tilde{c}^2\tilde{m}_1^2 \;+\tilde{c}^3\tilde{m}_1\tilde{m}_2+\; O(\tilde{c}^4))E[X^2] \\
&+(\tilde{c}^3\tilde{m}_1^3+O(\tilde{c}^4))E[X^3]-O(\tilde{c}^4) \\
&=\; \tilde{c}(\tilde{m}_1E[X]-1) \;+\; \tilde{c}^2(z\tilde{m}_1+0.5\tilde{c}^2\tilde{m}_2E[X]-\tilde{m}_1^2E[X^2]) \\
&+\tilde{c}^3(0.5z\tilde{m}_2+0.25\tilde{m}_3E[X]-\tilde{m}_1\tilde{m}_2E[X^2]+\tilde{m}_1^3E[X^3])+\; O(\tilde{c}^4),
\end{aligned}
\tag{G.8}
$$

This equation must hold for all $\tilde{c}$. Equating the coefficients, we get

$$\begin{aligned}
&\tilde{m}_1 = \frac{1}{E[X]} \\
&\tilde{m}_2 \;=\; 2E[X]^{-1}(\tilde{m}_1^2 E[X^2] - z\tilde{m}_1) \;=\; 2E[X]^{-1}(\frac{E[X^2]}{E[X]^2} - zE[X]^{-1})\,. \\
&\tilde{m}_3 = 4E[X]^{-1}\left(\tilde{m}_1\tilde{m}_2 E[X^2] - 0.5z\tilde{m}_2 - \tilde{m}_1^3 E[X^3]\right) \\
&= 4E[X]^{-1}\left(\tilde{m}_2(\tilde{m}_1 E[X^2] - 0.5z) - \tilde{m}_1^3 E[X^3]\right) \\
&= 4E[X]^{-2}\left(2\left(\frac{E[X^2]}{E[X]^2} - zE[X]^{-1}\right)\left(\frac{E[X^2]}{E[X]} - 0.5z\right) - \frac{E[X^3]}{E[X]^2}\right)
\end{aligned} \tag{G.9}$$

Now,

$$\begin{aligned}
m(-z;c) \;&=\; \tilde{c}\tilde{m}(-z;c) \;+\; (1-\tilde{c})z^{-1} \\
&=\; \tilde{c}(\tilde{c}\tilde{m}_1 + 0.5\tilde{c}^2\tilde{m}_2 + 0.25\tilde{c}^3\tilde{m}_3 + O(\tilde{c}^4)) \;+\; (1-\tilde{c})z^{-1} \\
&=\; (1-\tilde{c})z^{-1} \;+\; \tilde{c}^2\tilde{m}_1 + 0.5\tilde{c}^3\tilde{m}_2 + 0.25\tilde{c}^4\tilde{m}_3 + O(\tilde{c}^5)\,.
\end{aligned} \tag{G.10}$$

Note that, formally, we have computed the Taylor expansion

$$m(z) \;=\; m_0(z) \;+\; m_1(z)\tilde{c} \;+\; m_2(z)\tilde{c}^2 + m_3(z)\tilde{c}^3 \;+\; m_4(z)\tilde{c}^4 + \; O(\tilde{c}^5), \tag{G.11}$$

where

$$m_k(z) \;=\; k!\frac{\partial^k}{\partial\tilde{c}^k}m(z)\,. \tag{G.12}$$

As a result, by analyticity,[11] we also have

$$\frac{\partial}{\partial z}m(z) \;=\; \frac{\partial}{\partial z}m_0(z) \;+\; \frac{\partial}{\partial z}m_1(z)\tilde{c} + \frac{\partial}{\partial z}m_2(z)\tilde{c}^2 + m_3(z)\tilde{c}^3 + \frac{\partial}{\partial z}m_4(z)\tilde{c}^4 + O(\tilde{c}^5), \tag{G.13}$$

[11] Cauchy integral formula implies that when some analytic function $f(z)$ is $O(\tilde{c}^4)$, then so is its derivative.

Thus,

$$
\begin{aligned}
&\frac{\partial}{\partial z}(m(-z;c)) \\
&= \frac{\partial}{\partial z}\left((1-\tilde{c})z^{-1} + \tilde{c}^2\tilde{m}_1 + 0.5\tilde{c}^3\tilde{m}_2 + 0.25\tilde{c}^4\tilde{m}_3 + O(\tilde{c}^5)\right) \\
&= \frac{\partial}{\partial z}\left((1-\tilde{c})z^{-1} + \tilde{c}^2\frac{1}{E[X]} + 0.5\tilde{c}^3\left(2E[X]^{-1}(\frac{E[X^2]}{E[X]^2} - zE[X]^{-1})\right)\right. \\
&\left. + \tilde{c}^4\left(E[X]^{-2}\left(2\left(\frac{E[X^2]}{E[X]^2} - zE[X]^{-1}\right)\left(\frac{E[X^2]}{E[X]} - 0.5z\right) - \frac{E[X^3]}{E[X]^2}\right)\right)\right. \\
&\left. + O(\tilde{c}^5)\right) \\
&= \frac{\partial}{\partial z}\left((1-\tilde{c})z^{-1} + \tilde{c}^2\frac{1}{E[X]} + 0.5\tilde{c}^3\left(2E[X]^{-1}(\frac{E[X^2]}{E[X]^2} - zE[X]^{-1})\right)\right. \\
&\left. + \tilde{c}^4\left(E[X]^{-3}\left(2\left((\frac{E[X^2]}{E[X]})^2 - 1.5\frac{E[X^2]}{E[X]}z + 0.5z^2\right) - \frac{E[X^3]}{E[X]^2}\right)\right)\right. \\
&\left. + O(\tilde{c}^5)\right) \\
&= -(1-\tilde{c})z^{-2} - \tilde{c}^3\tilde{E}[X]^{-2} \\
&- 2\tilde{c}^4E[X]^{-3}\left(1.5\frac{E[X^2]}{E[X]} - z\right) + O(\tilde{c}^5)\,.
\end{aligned}
\tag{G.14}
$$

Note that all terms are negative because $m(-z;c) = \lim_{P,T\to\infty,P/T\to c} P^{-1}\operatorname{tr}((zI + \hat{\Sigma}_P)^{-1})$ is monotone decreasing in $z > 0$. □

Lemma G.1 *Let $m(z;c) = c^{-1}\tilde{m}(z;c) + (c^{-1}-1)z^{-1}$, $z < 0$. Let also be X, a random variable distributed with dH(x). Then,*

$$
\begin{aligned}
\xi(z;c) &= \tilde{c}^{-1}\frac{E[X]}{z} - \left(\frac{E[X^2]}{zE[X]}\right) - \tilde{c}\left(\frac{E[X^2]^2 - zE[X]E[X^2]}{zE[X]^3} - \frac{E[X^3]}{zE[X]^2}\right) + O(\tilde{c}^2) \\
\frac{\partial}{\partial z}\xi(z;c) &= \frac{1}{z^2}\left(-\tilde{c}^{-1}E[X] + \left(\frac{E[X^2]}{E[X]}\right) + \tilde{c}\left(\frac{E[X^2]^2}{E[X]^3} - \frac{E[X^3]}{E[X]^2}\right)\right) + O(\tilde{c}^2)\,.
\end{aligned}
\tag{G.15}
$$

Furthermore,

$$Z_*(z;c) \;=\; \tilde{c}^{-1}E[X] \;+\; O(1)\,. \tag{G.16}$$

Proof of Lemma G.1. Substituting $m(z;c) = c^{-1}\tilde{m}(z;c) + (c^{-1}-1)z^{-1}$ in $\xi(z;c)$ we have

$$1 - c + czm(-z;c) \;=\; 1 - c + z\tilde{m}(-z;c) - (1-c) \;=\; z\tilde{m}(-z;c)\,, \tag{G.17}$$

and by (G.5), we have

$$\begin{aligned}
\xi(z;c) &= \frac{1}{1-c+czm(-z;c)} - 1 \\
&= \frac{1}{z\tilde{m}(-z;c)} - 1 \\
&= \frac{1}{z(\tilde{c}\tilde{m}_1 \;+\; 0.5\;\tilde{c}^2\,\tilde{m}_2 \;+ 0.25\tilde{c}^3\tilde{m}_3 + \;O(\tilde{c}^4))} - 1 \\
&= \tilde{c}^{-1}\frac{1}{z(\tilde{m}_1 \;+\; 0.5\;\tilde{c}\,\tilde{m}_2 \;+ 0.25\tilde{c}^2\tilde{m}_3 + \;O(\tilde{c}^3))} - 1\,.
\end{aligned} \tag{G.18}$$

By Taylor expansion, we get

$$\begin{aligned}
&\tilde{c}^{-1}\frac{1}{z(\tilde{m}_1 \;+\; 0.5\;\tilde{c}\,\tilde{m}_2 \;+ 0.25\tilde{c}^2\tilde{m}_3 + \;O(\tilde{c}^3))} = \tilde{c}^{-1}\frac{1}{z\tilde{m}_1(1 \;+\; 0.5\;\tilde{c}\,\frac{\tilde{m}_2}{\tilde{m}_1} \;+ 0.25\tilde{c}^2\frac{\tilde{m}_3}{\tilde{m}_1} + \;O(\tilde{c}^3))} \\
&= \tilde{c}^{-1}\frac{1}{z\tilde{m}_1}\left(1 - 0.5\tilde{c}\frac{\tilde{m}_2}{\tilde{m}_1} - 0.25\tilde{c}^2\frac{\tilde{m}_3}{\tilde{m}_1} + 0.25\tilde{c}^2\frac{\tilde{m}_2^2}{\tilde{m}_1^2} + O(\tilde{c}^3)\right)
\end{aligned} \tag{G.19}$$

Substituting the expression for $\tilde{m}_1, \tilde{m}_2$ and $\tilde{m}_3$ gives

$$\begin{aligned}
0.5\tilde{c}\frac{\tilde{m}_2}{\tilde{m}_1} &= \tilde{c}\left(\frac{E[X^2]}{E[X]^2} - zE[X]^{-1}\right) = \tilde{c}\left(\frac{E[X^2] - zE[X]}{E[X]^2}\right) \\
0.25\tilde{c}^2\frac{\tilde{m}_3}{\tilde{m}_1} &= \tilde{c}^2E[X]^{-1}\left(2\left(\frac{E[X^2]}{E[X]^2} - zE[X]^{-1}\right)\left(\frac{E[X^2]}{E[X]} - 0.5z\right) - \frac{E[X^3]}{E[X]^2}\right) \\
&= 2\tilde{c}^2E[X]^{-1}\left(\frac{(E[X^2] - zE[X])(E[X^2] - 0.5zE[X])}{E[X]^3} - \frac{E[X^3]}{2E[X]^2}\right) \\
0.25\tilde{c}^2\frac{\tilde{m}_2^2}{\tilde{m}_1^2} &= \tilde{c}^2\left(\frac{E[X^2] - zE[X]}{E[X]^2}\right)^2
\end{aligned} \tag{G.20}$$

Hence,

$$
\begin{aligned}
\xi(z;c) &= \tilde{c}^{-1}\frac{1}{z\tilde{m}_1}\left(1-0.5\tilde{c}\frac{\tilde{m}_2}{\tilde{m}_1}-0.25\tilde{c}^2\frac{\tilde{m}_3}{\tilde{m}_1}+0.25\tilde{c}^2\frac{\tilde{m}_2^2}{\tilde{m}_1^2}+O(\tilde{c}^3)\right)-1\\
&= \tilde{c}^{-1}\frac{E[X]}{z}\left(1-\tilde{c}\left(\frac{E[X^2]-zE[X]}{E[X]^2}\right)\right.\\
&\quad -2\tilde{c}^2E[X]^{-1}\left(\frac{(E[X^2]-zE[X])(E[X^2]-0.5zE[X])}{E[X]^3}-\frac{E[X^3]}{2E[X]^2}\right)\\
&\quad \left.+\tilde{c}^2\left(\frac{E[X^2]-zE[X]}{E[X]^2}\right)^2+O(\tilde{c}^3)\right)-1\\
&= \tilde{c}^{-1}\frac{E[X]}{z}\left(1-\tilde{c}\left(\frac{E[X^2]-zE[X]}{E[X]^2}\right)\right.\\
&\quad -2\tilde{c}^2E[X]^{-3}\left(\frac{E[X^2]^2-1.5zE[X]E[X^2]+0.5z^2E[X]^2}{E[X]}-\frac{E[X^3]}{2}\right)\\
&\quad \left.+\tilde{c}^2\frac{E[X^2]^2-2zE[X]E[X^2]+z^2E[X]^2}{E[X]^4}+O(\tilde{c}^3)\right)-1\\
&= \tilde{c}^{-1}\frac{E[X]}{z}\left(1-\tilde{c}\left(\frac{E[X^2]-zE[X]}{E[X]^2}\right)\right.\\
&\quad -2\tilde{c}^2E[X]^{-3}\\
&\quad \times\left(\frac{E[X^2]^2-1.5zE[X]E[X^2]+0.5z^2E[X]^2}{E[X]}-\frac{E[X^3]}{2}-\frac{0.5E[X^2]^2-zE[X]E[X^2]+0.5z^2E[X]^2}{E[X]}\right)\\
&\quad \left.+O(\tilde{c}^3)\right)-1\\
&= \tilde{c}^{-1}\frac{E[X]}{z}\left(1-\tilde{c}\left(\frac{E[X^2]-zE[X]}{E[X]^2}\right)\right.\\
&\quad -\tilde{c}^2E[X]^{-3}\left(\frac{E[X^2]^2-zE[X]E[X^2]}{E[X]}-E[X^3]\right)\\
&\quad \left.+O(\tilde{c}^3)\right)-1\\
&= \tilde{c}^{-1}\frac{E[X]}{z}-\left(\frac{E[X^2]}{zE[X]}\right)-\tilde{c}\left(\frac{E[X^2]^2-zE[X]E[X^2]}{zE[X]^3}-\frac{E[X^3]}{zE[X]^2}\right)+O(\tilde{c}^2)
\end{aligned}
\tag{G.21}
$$

By analyticity, we can compute $\xi'(z;c)$ by differentiating equation (G.21) with respect to z. We have

$$\begin{aligned}\frac{\partial}{\partial z}\xi(z;c) &= \frac{\partial}{\partial z}\left(\tilde{c}^{-1}\frac{E[X]}{z} - \left(\frac{E[X^2]}{zE[X]}\right) - \tilde{c}\left(\frac{E[X^2]^2 - zE[X]E[X^2]}{zE[X]^3} - \frac{E[X^3]}{zE[X]^2}\right) + O(\tilde{c}^2)\right) \\ &= -\tilde{c}^{-1}\frac{E[X]}{z^2} + \frac{E[X^2]}{z^2E[X]} - \tilde{c}\left(-\frac{E[X^2]^2}{z^2E[X]^3} + \frac{E[X^3]}{z^2E[X]^2}\right) + O(\tilde{c}^2) \\ &= \frac{1}{z^2}\left(-\tilde{c}^{-1}E[X] + \left(\frac{E[X^2]}{E[X]}\right) + \tilde{c}\left(\frac{E[X^2]^2}{E[X]^3} - \frac{E[X^3]}{E[X]^2}\right)\right) + O(\tilde{c}^2)\end{aligned} \tag{G.22}$$

□

Lemma G.2 *We have*

$$\Delta(z;c) = 2\tilde{c}^{-1}\left(\frac{E[X^2]}{z^2} + O(\tilde{c})\right). \tag{G.23}$$

Proof of Lemma G.2. By Lemma G.1, we have

$$\begin{aligned}\xi(z;c) &= \tilde{c}^{-1}\frac{E[X]}{z} - \left(\frac{E[X^2]}{zE[X]}\right) - \tilde{c}\left(\frac{E[X^2]^2 - zE[X]E[X^2]}{zE[X]^3} - \frac{E[X^3]}{zE[X]^2}\right) + O(\tilde{c}^2) \\ \frac{\partial}{\partial z}\xi(z;c) &= \frac{1}{z^2}\left(-\tilde{c}^{-1}E[X] + \left(\frac{E[X^2]}{E[X]}\right) + \tilde{c}\left(\frac{E[X^2]^2}{E[X]^3} - \frac{E[X^3]}{E[X]^2}\right)\right) + O(\tilde{c}^2)\end{aligned} \tag{G.24}$$

Since

$$\Delta(z;c) = 2(\xi(z;c) + z\xi'(z;c))(1+\xi(z;c))^2, \tag{G.25}$$

we get

$$\begin{aligned}
&\xi(z;c) + z\xi'(z;c) = \frac{1}{z}\left(\tilde{c}^{-1}E[X] - \left(\frac{E[X^2]}{E[X]}\right) - \tilde{c}\left(\frac{E[X^2]^2 - zE[X]E[X^2]}{E[X]^3} - \frac{E[X^3]}{E[X]^2}\right)\right) + O(\tilde{c}^2) \\
&+ \frac{1}{z}\left(-\tilde{c}^{-1}E[X] + \left(\frac{E[X^2]}{E[X]}\right) + \tilde{c}\left(\frac{E[X^2]^2}{E[X]^3} - \frac{E[X^3]}{E[X]^2}\right)\right) + O(\tilde{c}^2) \\
&= \tilde{c}\frac{E[X]E[X^2]}{E[X]^3} + O(\tilde{c}^2)\,.
\end{aligned} \tag{G.26}$$

Furthermore,

$$(1+\xi(z;c))^2 = \left(1 + \tilde{c}^{-1}\frac{E[X]}{z} - \left(\frac{E[X^2]}{zE[X]}\right) - \tilde{c}\left(\frac{E[X^2]^2 - zE[X]E[X^2]}{zE[X]^3} - \frac{E[X^3]}{zE[X]^2}\right) + O(\tilde{c}^2)\right)^2 \tag{G.27}$$

Remark that we will get a $O(\tilde{c})$ term by the product of $\tilde{c}^{-1}\frac{E[X]}{z}$ and $O(\tilde{c}^2)$. Therefore, we can collapse all terms of order $O(\tilde{c})$ and above in the computation. We get,

$$\begin{aligned}
&\left(1 + \tilde{c}^{-1}\frac{E[X]}{z} - \left(\frac{E[X^2]}{zE[X]}\right) - \tilde{c}\left(\frac{E[X^2]^2 - zE[X]E[X^2]}{zE[X]^3} - \frac{E[X^3]}{zE[X]^2}\right) + O(\tilde{c}^2)\right)^2 \\
&= 1 + 2\tilde{c}^{-1}\frac{E[X]}{z} - 2\frac{E[X^2]}{zE[X]} + \tilde{c}^{-2}\frac{E[X]^2}{z^2} - 2\tilde{c}^{-1}\frac{E[X^2]}{z^2} - 2\left(\frac{E[X^2]^2 - zE[X]E[X^2]}{z^2E[X]^2} - \frac{E[X^3]}{z^2E[X]}\right) \\
&+ \frac{E[X^2]^2}{z^2E[X]^2} + O(\tilde{c}) \\
&= \tilde{c}^{-2}\left(\frac{E[X]^2}{z^2}\right) + 2\tilde{c}^{-1}\left(\frac{zE[X] - E[X^2]}{z^2}\right) \\
&+ 1 - 2\frac{zE[X]E[X^2]}{z^2E[X]^2} - 2\frac{E[X^2]^2 - zE[X]E[X^2]}{z^2E[X]^2} + 2\frac{E[X^3]}{z^2E[X]} + \frac{E[X^2]^2}{z^2E[X]^2} + O(\tilde{c}) \\
&= \tilde{c}^{-2}\left(\frac{E[X]^2}{z^2}\right) + 2\tilde{c}^{-1}\left(\frac{zE[X] - E[X^2]}{z^2}\right) + \left(1 - \frac{E[X^2]^2}{z^2E[X]^2} + 2\frac{E[X^3]}{z^2E[X]}\right) + O(\tilde{c})
\end{aligned} \tag{G.28}$$

Therefore,

$$\begin{aligned}
\Delta(z;c) &= 2\left(\tilde{c}\frac{E[X]E[X^2]}{E[X]^3}+O(\tilde{c}^2)\right)\\
&\times\left(\tilde{c}^{-2}\left(\frac{E[X]^2}{z^2}\right)+2\tilde{c}^{-1}\left(\frac{zE[X]-E[X^2]}{z^2}\right)+\left(1-\frac{E[X^2]^2}{z^2E[X]^2}+2\frac{E[X^3]}{z^2E[X]}\right)+O(\tilde{c})\right)\\
&= 2\left(\tilde{c}\frac{E[X]E[X^2]}{E[X]^3}+O(\tilde{c}^2)\right)\\
&\times\tilde{c}^{-2}\left(\left(\frac{E[X]^2}{z^2}\right)+2\tilde{c}\left(\frac{zE[X]-E[X^2]}{z^2}\right)+\tilde{c}^2\left(1-\frac{E[X^2]^2}{z^2E[X]^2}+2\frac{E[X^3]}{z^2E[X]}\right)+O(\tilde{c}^3)\right)\\
&= 2\tilde{c}^{-1}\left(\frac{E[X]E[X^2]}{E[X]^3}+O(\tilde{c})\right)\\
&\times\left(\left(\frac{E[X]^2}{z^2}\right)+O(\tilde{c})\right)
\end{aligned} \tag{G.29}$$

□

Lemma G.3 *Under the assumption that $\Omega_\alpha = aI + b\Sigma_P$ and assumption 2, we have*

$$\frac{\mathcal{H}(z;c)}{\Delta(z;c)^{1/2}} = \tilde{c}^{1/2}\frac{1}{1+\hat{\theta}_M^2}\left(\sqrt{\frac{1}{2E[X^2]}}(a+bE[X])+O(\tilde{c})\right) \tag{G.30}$$

Proof of Lemma G.3. Under the assumption that $\Omega_\alpha = aI + b\Sigma_P$ and assumption 2, we have by Lemma 1 that the limit in (C.18) satisfies

$$\mathcal{H}(z;c) = \frac{am(z;c)+b\tilde{c}\xi(z,c)}{1+\hat{\theta}_M^2} \tag{G.31}$$

Therefore, we get

$$
\begin{aligned}
& am(z,c) + b\tilde{c}\xi(z,c) \\
& = a(1-\tilde{c})z^{-1} \ + a\tilde{c}^2 E[X]^{-1} + 0.5a\tilde{c}^3(E[X]^{-1}(\frac{E[X^2]}{E[X]^2} - zE[X]^{-1})) + O(\tilde{c}^4) \\
& + b\frac{E[X]}{z} - \tilde{c}b\left(\frac{E[X^2]}{zE[X]}\right) - \tilde{c}^2 b\left(\frac{E[X^2]^2 - zE[X]E[X^2]}{zE[X]^3} - \frac{E[X^3]}{zE[X]^2}\right) + O(\tilde{c}^3) \\
& = z^{-1}(a + bE[X]) - \tilde{c}\left(az^{-1} + b\left(\frac{E[X^2]}{zE[X]}\right)\right) \\
& + \tilde{c}^2\left(aE[X]^{-1} - b\left(\frac{E[X^2]^2 - zE[X]E[X^2]}{zE[X]^3} - \frac{E[X^3]}{zE[X]^2}\right)\right) + O(\tilde{c}^3)
\end{aligned} \tag{G.32}
$$

Now, by Lemma G.2 we have,

$$
\Delta(z;c) = 2\tilde{c}^{-1}\left(\frac{E[X]E[X^2]}{E[X]^3} + O(\tilde{c})\right) \times \left(\left(\frac{E[X]^2}{z^2}\right) + O(\tilde{c})\right) \tag{G.33}
$$

Therefore,

$$
\sqrt{\Delta(z;c)} = \tilde{c}^{-1/2}\sqrt{2\frac{E[X^2]}{z^2} + O(\tilde{c})} \tag{G.34}
$$

Since the Taylor expansion of $\sqrt{(1+x)}$ around is $x = 0$ is given by

$$
\sqrt{1+bx} = 1 + \frac{1}{2}bx + O(x^2) \tag{G.35}
$$

We get,

$$
\sqrt{2\frac{E[X^2]}{z^2} + O(\tilde{c})} = \sqrt{2\frac{E[X^2]}{z^2}}(1 + O(\tilde{c})) \tag{G.36}
$$

Hence,

$$
\sqrt{\Delta(z;c)} = \tilde{c}^{-1/2}\sqrt{2\frac{E[X^2]}{z^2}} + O(\tilde{c}^{1/2}) \tag{G.37}
$$

Now, we have

$$\begin{aligned}\frac{1}{\tilde{c}^{-1/2}\sqrt{2\frac{E[X^2]}{z^2}}+O(\tilde{c}^{1/2})} &= \frac{1}{\tilde{c}^{-1/2}\sqrt{2\frac{E[X^2]}{z^2}}(1+O(\tilde{c}))} \\ &= \frac{1}{\tilde{c}^{-1/2}\sqrt{2\frac{E[X^2]}{z^2}}}(1-O(\tilde{c})+O(\tilde{c}^2)) \\ &= \tilde{c}^{1/2}\sqrt{\frac{z^2}{2E[X^2]}}(1-O(\tilde{c})+O(\tilde{c}^2)) \\ &= \tilde{c}^{1/2}\sqrt{\frac{z^2}{2E[X^2]}}+O(\tilde{c}^{3/2}))\end{aligned} \tag{G.38}$$

Finally,

$$\begin{aligned}&\frac{\mathcal{H}(z;c)}{\Delta(z;c)^{1/2}} = \tilde{c}^{1/2}\frac{1}{1+\hat{\theta}_M^2}\left(\sqrt{\frac{z^2}{2E[X^2]}}+O(\tilde{c})\right) \\ &\times\left\{z^{-1}(a+bE[X])-\tilde{c}\left(az^{-1}+b\left(\frac{E[X^2]}{zE[X]}\right)\right)\right. \\ &\left.+\tilde{c}^2\left(aE[X]^{-1}-b\left(\frac{E[X^2]^2-zE[X]E[X^2]}{zE[X]^3}-\frac{E[X^3]}{zE[X]^2}\right)\right)+O(\tilde{c}^3)\right\} \\ &= \tilde{c}^{1/2}\frac{1}{1+\hat{\theta}_M^2}\left(\sqrt{\frac{1}{2E[X^2]}}(a+bE[X])+O(\tilde{c})\right)\end{aligned} \tag{G.39}$$

□

H Simple Derivations for the Gaussian Case

In this section, we provide some simple derivations for the underlying statistic for the simpler case when residuals are Gaussian. While proving the Central Limit Theorem is still not trivial, computing the mean and variance of $\hat{W}(z)$ becomes a relatively simple problem.

Proposition H.1 *Let A_P be a uniformly bounded sequence of symmetric matrices and*

$$\bar{Y} \;=\; \bar{\alpha}' A_P \bar{\alpha}\,. \tag{H.1}$$

Suppose also that R_t^M are uniformly bounded for $t \geq 1$ and $\hat{\Sigma}_M^{-1}$ is uniformly bounded. Then,

$$E[\bar{Y}] \;=\; \alpha' A_P \alpha \;+\; \frac{1}{T}\operatorname{tr}(\Sigma_P A_P)(1+\hat{\theta}_M^2) \tag{H.2}$$

and

$$\operatorname{Var}[\bar{Y}] \;=\; \frac{P}{T^2} 2\left(P^{-1}\operatorname{tr}(\Sigma_P A_P \Sigma_P A_P)\right) \;+\; O(T^{-2})\,. \tag{H.3}$$

Proof of Proposition H.1. We have

$$\hat{\alpha}_t \;=\; R_t \;-\; \hat{B} R_t^M, \tag{H.4}$$

where

$$\hat{\Sigma}_M \;=\; \frac{1}{T}\sum_t R_t^M (R_t^M)' - \bar{R}^M (\bar{R}^M)' \;=\; \frac{1}{T}\sum_t R_t^M (R_t^M - \bar{R}^M)' \tag{H.5}$$

implies that

$$\begin{aligned}
\hat{B} \;&=\; T^{-1}\sum_t R_t(R_t^M - \bar{R}^M)'\,\hat{\Sigma}_M^{-1} \\
&=\; T^{-1}\sum_t(\alpha \;+\; BR_t^M \;+\; \varepsilon_t)(R_t^M - \bar{R}^M)'\,\hat{\Sigma}_M^{-1} \\
&=\; B \;+\; T^{-1}\sum_t \underbrace{\varepsilon_t}_{P\times 1}\underbrace{(R_t^M - \bar{R}^M)'\,\hat{\Sigma}_M^{-1}}_{1\times L} \\
&=\; B \;+\; \underbrace{\hat{\varepsilon}}_{P\times L},
\end{aligned} \tag{H.6}$$

where we have used the fact that the true DGP is given by

$$R_t \;=\; \alpha \;+\; BR_t^M \;+\; \varepsilon_t, \tag{H.7}$$

so that

$$\hat{\alpha}_t \;=\; \alpha \;+\; \varepsilon_t \;+\; (B - \hat{B})R_t^M \;=\; \alpha \;+\; \varepsilon_t \;-\hat{\varepsilon}R_t^M \tag{H.8}$$

while

$$\hat{\alpha} - \bar{\alpha} \;=\; \varepsilon_t - \bar{\varepsilon} \;+\; (B - \hat{B})(R_t^M - \bar{R}^M)\,. \tag{H.9}$$

Note that, under Gaussianity, $\hat{\alpha} - \bar{\alpha}$ is independent of $\bar{\alpha}$. Furthermore,

$$\begin{aligned}
\bar{\alpha} \;&=\; \alpha \;+\; \bar{\varepsilon} \;-\; \hat{\varepsilon}\bar{R}^M \\
&=\; \alpha \;+\; T^{-1}\sum_t \varepsilon_t(1 - (R_t^M - \bar{R}^M)'\,\hat{\Sigma}_M^{-1}\bar{R}^M) \\
&=\; \alpha \;+\; T^{-1}\sum_t \varepsilon_t\psi_t \;=\; T^{-1}\sum_t(\alpha \;+\varepsilon_t\psi_t),
\end{aligned} \tag{H.10}$$

where we have defined

$$\psi_t \;=\; (1 - (R_t^M - \bar{R}^M)' \, \hat{\Sigma}_M^{-1} \bar{R}^M) \tag{H.11}$$

and

$$\psi_t^2 \;=\; 1 - 2(R_t^M - \bar{R}^M)' \, \hat{\Sigma}_M^{-1} \bar{R}^M \;+\; ((R_t^M - \bar{R}^M)' \, \hat{\Sigma}_M^{-1} \bar{R}^M)^2 \, . \tag{H.12}$$

Thus, we have

$$E[\bar{\alpha}' A \bar{\alpha}] \;=\; T^{-2} E[\sum_{t_1, t_2} \psi_{t_1} \psi_{t_2} \varepsilon_{t_1}' A \varepsilon_{t_2}] + E[\alpha' A \alpha] \, . \tag{H.13}$$

Without loss of generality, we may assume that $\Sigma_P = I$. Indeed, otherwise, we can write $\varepsilon_t = \Sigma_P^{1/2} z_t$ and redefine $\tilde{A}_P = \Sigma_P^{1/2} A_P \Sigma_P^{1/2}$ to get

$$\bar{\alpha}' A \bar{\alpha} \;=\; \alpha' A \alpha \;+\; T^{-2} \sum_{t_1, t_2} \psi_{t_1} \psi_{t_2} \varepsilon_{t_1}' A \varepsilon_{t_2} \;=\; \alpha' A \alpha \;+\; T^{-2} \sum_{t_1, t_2} \psi_{t_1} \psi_{t_2} z_{t_1}' \tilde{A} z_{t_2} \, . \tag{H.14}$$

Thus, since $E[\varepsilon_t] = 0$ and ε_t are independent across t, we get $E[\varepsilon_t' A \varepsilon_t] = \mathrm{tr}(A)$ and, hence,

$$E[\bar{\alpha}' A \bar{\alpha}] \;=\; \alpha' A \alpha + T^{-2} \sum_t \psi_t^2 \, \mathrm{tr}(A) \, . \tag{H.15}$$

Let

$$\hat{\Sigma}_{M,t} \;=\; R_t^M (R_t^M - \bar{R}^M)', \tag{H.16}$$

so that

$$\hat{\Sigma}_M \;=\; \frac{1}{T} \sum_t \hat{\Sigma}_{M,t} \, . \tag{H.17}$$

Then,

$$\begin{aligned}
\sum_t \psi_t^2 &= \sum_t \left(1 + \left((R_t^M - \bar{R}^M)' \hat{\Sigma}_M^{-1} \bar{R}^M\right)^2\right) \\
&= \sum_t \left(1 + \left((R_t^M - \bar{R}^M)' \hat{\Sigma}_M^{-1} \bar{R}^M (R_t^M - \bar{R}^M)' \hat{\Sigma}_M^{-1} \bar{R}^M\right)\right) \\
&= \sum_t \left(1 + \left((\bar{R}^M)' \hat{\Sigma}_M^{-1} (R_t^M - \bar{R}^M)(R_t^M - \bar{R}^M)' \hat{\Sigma}_M^{-1} \bar{R}^M\right)\right) \\
&= \sum_t \left(1 + \left((\bar{R}^M)' \hat{\Sigma}_M^{-1} \underbrace{R_t^M (R_t^M - \bar{R}^M)'}_{\hat{\Sigma}_{M,t}} \hat{\Sigma}_M^{-1} \bar{R}^M\right)\right) \\
&= T + (\bar{R}^M)' \hat{\Sigma}_M^{-1} (T\hat{\Sigma}_M) \hat{\Sigma}_M^{-1} \bar{R}^M) \\
&= T(1 + (\bar{R}^M)' \hat{\Sigma}_M^{-1} \bar{R}^M) \;=\; T(1 + \hat{\theta}_M^2)\,.
\end{aligned} \tag{H.18}$$

Similarly,

$$\begin{aligned}
&E[(\bar{\alpha}' A \bar{\alpha})^2] = E\left[\left(T^{-2} \sum_{t_1,t_2} \psi_{t_1}\psi_{t_2}\varepsilon_{t_1}' A \varepsilon_{t_2} + \alpha' A \alpha\right)\left(T^{-2} \sum_{t_1,t_2} \psi_{t_1}\psi_{t_2}\varepsilon_{t_1}' A \varepsilon_{t_2} + \alpha' A \alpha\right)\right] \\
&= E\left[T^{-4} \sum_{t_1,t_2,t_3,t_4} \psi_{t_1}\psi_{t_2}\psi_{t_3}\psi_{t_4}\varepsilon_{t_1}' A \varepsilon_{t_2}\varepsilon_{t_3}' A \varepsilon_{t_4}\right] \\
&+ 2E\left[T^{-2} \sum_{t_1,t_2} \psi_{t_1}\psi_{t_2}\varepsilon_{t_1}' A \varepsilon_{t_2} \alpha' A \alpha\right] \\
&+ E[(\alpha' A \alpha)^2] \\
&= Term1 + 2Term2 + Term3
\end{aligned} \tag{H.19}$$

Computation of $Term1$

Since $E[\varepsilon_t] = 0$, $E[\varepsilon_t^2] = 1$ and $E[\varepsilon_t^4] < \infty$, then the summands of

$$T^{-4} \sum_{t_1,t_2,t_3,t_4} E\left[\psi_{t_1}\psi_{t_2}\psi_{t_3}\psi_{t_4}\varepsilon_{t_1}' A \varepsilon_{t_2}\varepsilon_{t_3}' A \varepsilon_{t_4}\right] \tag{H.20}$$

are non-zero only when $(t_1 = t_2 = t_3 = t_4)$, or if there are exactly 2 pairs of identical indexes t. The first case gives

$$T^{-4}\sum_t \psi_t^4 E[\varepsilon_t' A\varepsilon_t \varepsilon_t' A\varepsilon_t] = T^{-4}\sum_t \psi_t^4 \sum_{i,j,k,l} E[\varepsilon_{t,i}A_{ij}\varepsilon_{t,j}\varepsilon_{t,k}A_{kl}\varepsilon_{t,l}] \tag{H.21}$$

This, again, is non-zero only when $i = j = k = l$ or $(i = j \neq k = l), (i = k \neq k = l)$ and $(i = l \neq j = k)$, hence, we get

$$\begin{aligned}
&\sum_{i,j,k,l} E[\varepsilon_{t,i}A_{ij}\varepsilon_{t,j}\varepsilon_{t,k}A_{kl}\varepsilon_{t,l}] \\
&= \sum_i E[\varepsilon_{t,i}^4]A_{ii}^2 + \sum_{i,k,i\neq k} A_{ii}A_{kk} + 2\sum_{i,j,i\neq j} A_{ij}^2 \\
&= \sum_i E[\varepsilon_{t,i}^4]A_{ii}^2 + \left(\sum_i A_{ii}\right)^2 - \sum_i A_{ii}^2 + 2\sum_{i,j} A_{ij}^2 - 2\sum_i A_{ii}^2 \\
&= (E[\varepsilon_{t,i}^4] - 3)\sum_i A_{ii}^2 + \mathrm{tr}(A)^2 + 2\|A\|_2^2
\end{aligned} \tag{H.22}$$

And so,

$$T^{-4}\sum_t \psi_t^4 \sum_{i,j,k,l} E[\varepsilon_{t,i}A_{ij}\varepsilon_{t,j}\varepsilon_{t,k}A_{kl}\varepsilon_{t,l}] = T^{-4}\sum_t \psi_t^4 \left((E[\varepsilon_{t,i}^4] - 3)\sum_i A_{ii}^2 + \mathrm{tr}(A)^2 + 2\|A\|_2^2\right) \tag{H.23}$$

For the second case, we need to distinguish when $(t_1 = t_2 \neq t_3 = t_4)$, $(t_1 = t_3 \neq t_2 = t_4)$ and $(t_1 = t_4 \neq t_2 = t_3)$. Firstly, when $(t_1 = t_2 \neq t_3 = t_4)$, we have

$$\begin{aligned}
T^{-4}\sum_{t_1\neq t_2} \psi_{t_1}^2\psi_{t_2}^2 E[\varepsilon_{t_1}' A\varepsilon_{t_1}\varepsilon_{t_2}' A\varepsilon_{t_2}] &= T^{-4}\sum_{t_1\neq t_2} \psi_{t_1}^2\psi_{t_2}^2 E[\varepsilon_{t_1}' A\varepsilon_{t_1}]E[\varepsilon_{t_2}' A\varepsilon_{t_2}] \\
&= T^{-4}\sum_{t_1\neq t_2} \psi_{t_1}^2\psi_{t_2}^2 \,\mathrm{tr}(A)^2
\end{aligned} \tag{H.24}$$

Then, when $(t_1 = t_3 \neq t_2 = t_4)$, we get

$$T^{-4} \sum_{t_1 \neq t_2} \psi_{t_1}^2 \psi_{t_2}^2 E[\varepsilon'_{t_1} A \varepsilon_{t_2} \varepsilon'_{t_1} A \varepsilon_{t_2}] = T^{-4} \sum_{t_1 \neq t_2} \psi_{t_1}^2 \psi_{t_2}^2 \sum_{i,j,k,l} E[\varepsilon'_{t_1,i} A_{i,j} \varepsilon_{t_2,j} \varepsilon'_{t_1,k} A_{k,l} \varepsilon_{t_2,l}] \tag{H.25}$$

The expectation is non-zero when $(i = k, j = l)$, so we get

$$\sum_{i,j,k,l} E[\varepsilon'_{t_1,i} A_{i,j} \varepsilon_{t_2,j} \varepsilon'_{t_1,k} A_{k,l} \varepsilon_{t_2,l}] = \sum_{i,j} A_{i,j}^2 = \operatorname{tr}(A^2)\,. \tag{H.26}$$

Hence,

$$T^{-4} \sum_{t_1 \neq t_2} \psi_{t_1}^2 \psi_{t_2}^2 E[\varepsilon'_{t_1} A \varepsilon_{t_2} \varepsilon'_{t_1} A \varepsilon_{t_2}] = T^{-4} \sum_{t_1 \neq t_2} \psi_{t_1}^2 \psi_{t_2}^2 \operatorname{tr}(A^2)\,. \tag{H.27}$$

Lastly, when $(t_1 = t_4 \neq t_2 = t_3)$, we obtain,

$$\begin{aligned} T^{-4} \sum_{t_1 \neq t_2} \psi_{t_1}^2 \psi_{t_2}^2 E[\varepsilon'_{t_1} A \varepsilon_{t_2} \varepsilon'_{t_2} A \varepsilon_{t_1}] &\underbrace{=}_{Symmetry\ of\ A} T^{-4} \sum_{t_1 \neq t_2} \psi_{t_1}^2 \psi_{t_2}^2 E[\varepsilon'_{t_1} A \varepsilon_{t_2} \varepsilon'_{t_1} A \varepsilon_{t_2}] \\ &= T^{-4} \sum_{t_1 \neq t_2} \psi_{t_1}^2 \psi_{t_2}^2 \operatorname{tr}(A^2)\,. \end{aligned} \tag{H.28}$$

So, $Term1$ is given by

$$\begin{aligned} &T^{-4} \sum_{t_1,t_2,t_3,t_4} E\left[\psi_{t_1}\psi_{t_2}\psi_{t_3}\psi_{t_4} \varepsilon'_{t_1} A \varepsilon_{t_2} \varepsilon'_{t_3} A \varepsilon_{t_4}\right] = T^{-4} \sum_t \psi_t^4 \left((E[\varepsilon_{t,i}^4] - 3) \sum_i A_{ii}^2 + \operatorname{tr}(A)^2 + 2\|A\|_2^2 \right) \\ &+ T^{-4} \sum_{t_1 \neq t_2} \psi_{t_1}^2 \psi_{t_2}^2 \operatorname{tr}(A)^2 \\ &+ 2T^{-4} \sum_{t_1 \neq t_2} \psi_{t_1}^2 \psi_{t_2}^2 \operatorname{tr}(A^2) \\ &= \; T^{-4} \sum_t \psi_t^4 (E[\varepsilon_{t,i}^4] - 3) \sum_i A_{ii}^2 \\ &+ \; T^{-4} (\sum_t \psi_t^2)^2 (\operatorname{tr}(A)^2 + 2\operatorname{tr}(A^2))\,. \end{aligned} \tag{H.29}$$

Computation of $Term2$

We have

$$\begin{aligned} T^{-2}E[\sum_{t_1,t_2}\psi_{t_1}\psi_{t_2}\varepsilon'_{t_1}A\varepsilon_{t_2}\alpha'A\alpha] &= T^{-2}\sum_{t_1,t_2}\psi_{t_1}\psi_{t_2}E[\varepsilon'_{t_1}A\varepsilon_{t_2}]E[\alpha'A\alpha] \\ &= T^{-2}\sum_{t}\psi_t^2\,\mathrm{tr}(A)\alpha'A\alpha \end{aligned} \tag{H.30}$$

Computation of $Term3$

We have

$$E[(\alpha'A\alpha)^2] = (\alpha'A\alpha)^2 \tag{H.31}$$

Hence, by combining all the terms together, we get

$$\begin{aligned} E[(\bar{\alpha}'A\bar{\alpha})^2] &= Term1 + 2Term2 + Term3 \\ &= T^{-4}\sum_t \psi_t^4(E[\varepsilon_{t,i}^4]-3)\sum_i A_{ii}^2 \\ &+\ T^{-4}(\sum_t\psi_t^2)^2(\mathrm{tr}(A)^2+2\,\mathrm{tr}(A^2)) \\ &+ 2T^{-2}\sum_t\psi_t^2\,\mathrm{tr}(A)\alpha'A\alpha \\ &+ (\alpha'A\alpha)^2 \\ &=\ 2T^{-4}(\sum_t\psi_t^2)^2\,\mathrm{tr}(A^2)\ +\ T^{-4}\sum_t\psi_t^4(E[\varepsilon_{t,i}^4]-3)\sum_i A_{ii}^2 \\ &+\ (E[\bar{\alpha}'A\bar{\alpha}])^2\,, \end{aligned} \tag{H.32}$$

where we have used that, by (H.15),

$$(E[\bar{\alpha}'A\bar{\alpha}])^2\ =\ (\alpha'A\alpha + T^{-2}\sum_t\psi_t^2\,\mathrm{tr}(A))^2\,. \tag{H.33}$$

Hence,

$$\begin{aligned} &Var(\bar{\alpha}'A\bar{\alpha}) = E[(\bar{\alpha}'A\bar{\alpha})^2] - E[\bar{\alpha}'A\bar{\alpha}]^2 \\ &= T^{-4}\sum_t \psi_t^4(E[\varepsilon_{t,i}^4] - 3)\sum_i A_{ii}^2 \\ &+ T^{-4}T^2(1+\hat{\theta}_M^2)^2\,\mathrm{tr}(A^2)\,. \end{aligned} \tag{H.34}$$

Since,

$$\sum_i A_{ii}^2 \leq \sum_{i,j} A_{ij}^2 = \|A\|_2^2 = \mathrm{tr}(A^2) \;\leq\; \|A\|^2 P \;=\; O(P), \tag{H.35}$$

while the assumptions that R_t^M is bounded and $\hat{\Sigma}_M^{-1}$ is bounded imply that ψ_t are uniformly bounded, so that

$$T^{-4}\sum_t \psi_t^4(E[\varepsilon_{t,i}^4] - 3)\sum_i A_{ii}^2 \;=\; T^{-4}O(TP) \;=\; O(T^{-2}) \tag{H.36}$$

□

Proposition H.1 combined with a simple extension of Theorem 2.1 from Srivastava (2009) implies the following result.

Proposition H.2 (Test Statistic With Known Σ_P) *We have that*

$$\widetilde{W}(z) \;=\; \frac{\bar{\alpha}'(zI+\Sigma_P)^{-1}\bar{\alpha}}{1+\hat{\theta}_M^2} \tag{H.37}$$

satisfies

$$\frac{T^{1/2}(\widetilde{W}(z) - \widetilde{\xi}_P(z;c))}{\widetilde{\Delta}_P(z;c)^{1/2}} \;\to\; N(\frac{\widetilde{\mathcal{H}}(z)}{\widetilde{\Delta}(z;c)^{1/2}}, 1) \tag{H.38}$$

is distribution, where

$$\begin{aligned}\widetilde{\xi}(z) &= cP^{-1}\operatorname{tr}((zI+\Sigma_P)^{-1}\Sigma_P)\\ \widetilde{\Delta}(z;c) &= 2c\operatorname{tr}(P^{-1}\operatorname{tr}((zI+\Sigma_P)^{-2}\Sigma_P^2)\\ \widetilde{\mathcal{H}}(z) &= \lim_{P\to\infty}\alpha'(zI+\Sigma_P)^{-1}\alpha,\end{aligned} \tag{H.39}$$

assuming that the last limit exists.

As we can see from Proposition H.2, the behavior of the test statistic differs significantly from that in Theorems 5 and 7. For example, the test statistic is well defined even for $z \to 0$, whereas (by Theorem 2), both the mean and the variance of $\hat{W}$ explode for $z \to 0, c \to 1$.

Proof of Proposition 8 for the Gaussian Case. We have

$$\begin{aligned}E_T[\hat{\alpha}'_{T+1}A\bar{\alpha}] &= E_T[(\alpha+\varepsilon_{T+1}-\hat{\varepsilon}R^M_{T+1})'A\bar{\alpha}]\\ &= (\alpha-\hat{\varepsilon}\mu_T^M)'A\bar{\alpha}\\ &= (\alpha - T^{-1}\sum_t \varepsilon_t(R_t^M-\bar{R}^M)\hat{\Sigma}_M^{-1}\mu_T^M)'AT^{-1}\sum_t(\alpha\ +\varepsilon_t\psi_t),\end{aligned} \tag{H.40}$$

where we have defined

$$\mu_T^M = E_T[R^M_{T+1}]\,. \tag{H.41}$$

Thus,

$$\begin{aligned}&E_T[\hat{\alpha}'_{T+1}A\bar{\alpha}]\\ &= \alpha'A\alpha\ +\ T^{-1}\alpha'A\sum_t\varepsilon_t\psi_t\ -\ T^{-1}\alpha'A\sum_t\varepsilon_t(R_t^M-\bar{R}^M)\hat{\Sigma}_M^{-1}\mu_T^M\\ &+\ T^{-2}\sum_{t_1,t_2}(\varepsilon_{t_1}(R_t^M-\bar{R}^M)\hat{\Sigma}_M^{-1}\mu_T^M)'A\varepsilon_{t_2}\psi_{t_2}\end{aligned} \tag{H.42}$$

Thus,

$$E[E_T[\hat{\alpha}'_{T+1}A\bar{\alpha}]|\alpha, A] \;=\; \alpha' A \alpha \tag{H.43}$$

because

$$\sum_t (R_t^M - \bar{R}^M) \;=\; 0\,. \tag{H.44}$$

The fact that $\mathrm{Var}[E_T[\hat{\alpha}'_{T+1}A\bar{\alpha}]]$ converges to zero follows by the same arguments as above. The proof is complete. □