NBER WORKING PAPER SERIES

A TEST OF THE EFFICIENCY OF A GIVEN PORTFOLIO IN HIGH DIMENSIONS

Mikhail Chernov
Bryan T. Kelly
Semyon Malamud
Johannes Schwab

Working Paper 33565
http://www.nber.org/papers/w33565

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2025, Revised September 2025

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 Magnus Dahlquist, Jianqing Fan, Christian Julliard, Yuan Liao, Andreas Neuhierl, Olivier Scaillet, and Shivam Sharma, and participants at the 2025 ESSFM meeting in Gerzensee, the 2025 SoFiE conference, and the UCLA Mathematical and Statistical Finance seminar 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 or the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w33565

A Test of the Efficiency of a Given Portfolio in High Dimensions
Mikhail Chernov, Bryan T. Kelly, Semyon Malamud, and Johannes Schwab
NBER Working Paper No. 33565
March 2025, Revised September 2025
JEL No. C12, C40, C55, C57, G12

ABSTRACT

We extend the Gibbons–Ross–Shanken test to high-dimensional cases, when the num-ber of test assets far exceeds the sample size and the return covariance matrix is ill-conditioned or singular, as inevitably occurs with large, richly specified test port-folios. In such cases, one must use a regularized (and therefore biased) estimator of the covariance matrix, which distorts the original GRS test statistic. We use Random Matrix Theory to correct for this bias and characterize the asymptotic power of the resulting test. Power increases with the number of test assets and reaches its maximum across a broad range of local alternatives. These findings are supported by extensive simulations. We empirically implement the test on state-of-the-art candidate factor portfolios and test assets to evaluate conditional asset pricing performance.

Mikhail Chernov
University of California, Los Angeles
Anderson School of Management
and CEPR
and also NBER
mikhail.chernov@anderson.ucla.edu

Bryan T. Kelly
Yale University
and NBER
bryan.kelly@yale.edu

Semyon Malamud
Swiss Finance Institute
semyon.malamud@epfl.ch

Johannes Schwab
École Polytechnique Fédérale
de Lausanne (EPFL)
johannes.schwab@epfl.ch

The implementation code is available at
https://github.com/DjoFE2021/CKMSEfficiencyTest

1 Introduction

There is a long-standing tradition in cross-sectional asset pricing of testing tradable factor models by regressing the excess returns of test assets on the factors, assuming constant betas. The seminal Gibbons, Ross and Shanken (1989) (GRS) test evaluates whether the pricing errors (alphas) from these regressions are jointly zero—implying that the factor-spanned portfolio is mean–variance efficient. Moreover, Hansen and Richard (1987) show that if alphas are jointly zero when all feasible portfolios are used as test assets, the factors also price assets conditionally.

However, using an effectively infinite number of test assets poses a number of intertwined challenges for the GRS test. As GRS note, the test's power can degrade sharply when the number of test assets grows relative to the sample size T, prompting a recommendation to limit test assets to no more than one-half of T. This implication creates a tension with respect to testing for conditional pricing per Hansen and Richard (1987).

Recent work addresses this challenge by assuming sparsity in the alphas or residual covariance matrix, with Fan, Liao and Yao (2015) (FLY) 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 open to challenge in our context.

Further, all Wald-type tests that are employed in the context of these linear asset pricing regressions depend on an unbiased estimator of the inverse residual covariance matrix. Asymptotic properties of these tests rely on the crucial assumption that this inverse is uniformly bounded. This assumption is not innocuous: it requires even the smallest principal components to retain non-trivial variance. We show that this condition fails when test assets

include a large set of conditionally feasible portfolios. In such cases, the inverse residual covariance matrix grows unboundedly with the number of test assets. Thus, ill-conditioned covariance matrices are unavoidable when implementing the prescription of Hansen and Richard (1987).

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, and accommodating potentially unbounded inverse of the covariance matrix. Specifically, we propose using a regularized, and thus biased, estimator of the inverse covariance matrix.[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. We apply Random Matrix Theory (RMT) to debias 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 similar to that of GRS. That is, it captures the distance between the Sharpe ratio squared of test assets and the candidate factor combined, 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 already grapples with its test's power, and adding more assets than observations appears to exacerbate the problem. As we demonstrate next, however, the large dimension of test assets actually improves the test's power compared to GRS, despite complicating 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

[1] In fact, when the number of assets exceeds the number of time periods, the bias is present even in the limit of vanishing regularization. Basically, in the over-parameterized 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).

$T^{-1/2}$ (selected based on 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 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 ratio of the number of test assets to sample size, $c = P/T$, 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 full power.

Second, we assess the power of our test via a series of simulation exercises where we consider various sample sizes and alternatives. We explicitly build up portfolios from base assets inside of our simulation thereby leading to non-sparse environment. First, we verify the GRS result that when $c < 1$, the test's power is inversely U-shaped: low with few and with many test assets. Furthermore, the sample size must exceed $T = 300$ to achieve good power properties at intermediate values of c. As we move to our test statistic and consider $c > 1$, the test's power continues to improve as c grows to infinity, consistent with asymptotic power. As we approach sufficiently large sample sizes, with $T > 300$, power approaches one for all but the smallest values of c.

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) (JKP). 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 approximating

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 the BSV and DKKM factors when pricing JKP test assets. However, these factors are rejected in the full sample. BSV and DKKM also fail to price the 25,000 "network" portfolios. We interpret this evidence as indicative of limits-to-learning in high-complexity environments.

Scores of empirical papers are dedicated to identifying and testing efficient portfolios. Our contribution is primarily theoretical and builds on the foundational work of Gibbons, Ross and Shanken (1989) and Hansen and Richard (1987). While our analysis focuses on tradable factors, the test statistic we develop can also be applied to non-tradable factors. Extending our framework to accommodate omitted factors—such as in the three-pass method of Giglio and Xiu (2021)—is an interesting direction for future research.

Like us, Pesaran and Yamagata (2023) aim to extend GRS-type testing to settings with many test assets. They employ a covariance estimator based on the thresholding-based shrinkage technique of Fan, Liao and Mincheva (2011), which assumes sparsity in the covariance matrix. Fan, Liao and Yao (2015) (FLY) propose an influential method for enhancing the power of high-dimensional alpha tests. Specifically, FLY introduce two test statistics for testing zero alphas. The first is a feasible Wald statistic for high dimensions, developed under the assumption of a sparse true covariance matrix. Subsequent work by Feng, Lan, Liu and Ma (2022) and Yu, Yao and Xue (2024) improves the power of this test under sparsity. The second FLY statistic is an alpha-thresholding test that attains maximal power when the alpha vector is sparse, i.e., when only a small number of test assets have non-zero alpha. Kock and Preinerstorfer (2019) investigate the limits of power enhancement achievable through the FLY framework. These threshold-based tests lack rotation invariance and rely on the boundedness of the inverse covariance matrix.

Several papers (e.g., Feng, Lan, Liu and Ma, 2022, Massacci, Sarno, Trapani and Vallarino, 2025) introduce "max-type" tests that look only at the largest, in absolute value, estimated alpha across assets. Because this statistic depends on the selected basis of assets, it is not rotation-invariant. As a result, it has little power against "diffuse-alpha" cases: each asset's alpha is tiny on its own, but a suitable portfolio concentrates these small effects into a meaningful alpha. In exactly these situations, GRS-type tests, which pool information across assets and are rotation-invariant, detect the violation that max-type tests miss.

Lan, Feng and Luo (2018) extend the GRS test to non-sparse covariance matrices by repeatedly projecting test asset returns into low-dimensional subspaces where the GRS test is valid, and then aggregating the test statistics across these projections. This method incurs information loss, particularly regarding the low eigenvalues of the residual covariance matrix, which are rarely sampled. Consequently, the test's asymptotic power becomes independent of the eigenvalue distribution and instead depends inversely on the trace of the covariance matrix. In contrast, our test optimally exploits the joint distribution of eigenvalues and alphas across all principal components. This leads to substantial power gains, especially when low-variance components carry significant alpha, as frequently observed in modern factor zoo frameworks such as Dello-Preite, Uppal, Zaffaroni and Zviadadze (2022), Didisheim, Ke, Kelly and Malamud (2024), Giglio, Xiu and Zhang (2025), and Lettau and Pelger (2020). Notably, the test of Lan, Feng and Luo (2018), like all Wald-type tests to our knowledge, also relies on the boundedness of the inverse covariance matrix.

Our results apply to both sparse and dense (ill-conditioned) covariance matrix environments. This broader applicability is made possible by shrinking the covariance matrix and accounting for the resulting bias via RMT. One implication of our approach is that it achieves maximal power under a wider set of alternatives compared to the FLY feasible Wald statistic. However, our test does lose power under sparse alpha alternatives. That said, this limitation

is not necessarily problematic in practice, since alphas of test portfolios typically do not inherit the sparsity of alphas at the individual asset level.

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.

Recent work by Zhang, Lan, Feng and Feng (2025) extends some of our results to sample-splitting. Assessing the power of these sample-split analogues is a natural next step. In addition, our GRS-like statistic tests a single null of no alpha. Extending it to multiple-hypothesis settings in the spirit of Giglio, Liao and Xiu (2021) remains an open and practically important agenda.

2 The Test Statistic

2.1 Asset-pricing regressions when factors are tradeable

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.

While our headline results feature a single portfolio R_t^M, there is nothing that precludes one from considering multiple portfolios, or tradeable factors. The difference is in the economic interpretation of the exercise. A single factor is literally a candidate for the efficient portfolio. Multiple factors are presumed to span the efficient portfolio, under the null. If betas are constant (time-varying), the spanning is unconditional (conditional). Conditional spanning requires taking an additonal modeling stance on the conditioning information set and how it affects betas. Appendix F derives our results for multiple tradeable factors under the null hypothesis of unconditional spanning, which is the most common application.

2.2 Statistical structure of risk

When the test assets comprise a large number of portfolios, the standard assumptions about the statistical structure of the associated risk need to be revisited. The vast majority of theoretical results rely on boundedness of certain elements of the covariance matrix Σ_P. For example, the approximate $K-$factor structure associated with asymptotic Arbitrage Pricing Theory (APT) posits that, as the number of assets increases, $K \geq 1$ of the largest eigenvalues of the asset covariance matrix grow like P, while the rest $P-K$ remain bounded

(Chamberlain and Rothschild, 1982). The inverse of the covariance matrix of the $P - K$ residuals has to remain bounded as well to preclude no-arbitrage opportunities (unbounded information ratio).

As we show in Appendix B, the behavior of test assets incorporating rich conditional information is more subtle. When test portfolios are generated through many instruments—the set which includes traditional charactetristic-based or PCA-based portfolios—the APT condition does not hold.[2] In fact, all P eigenvalues grow like P. If we redefine test asset returns, normalizing them by $P^{-1/2}$, then Theorem B.1 implies that the eigenvalues of the residual covariance matrix always converge asymptotically to finite limits. The corresponding residual covariance matrix Σ_P has a bounded $\|\Sigma_P\|$ but an unbounded $\|\Sigma_P^{-1}\|$, which violates traditional assumptions used in the context of APT or testing factor-based models via Wald-type statistics. The intuition for this result is straightforward. With infinitely many instruments (non-linear conditioning information) used to generate test assets, some of the resulting dynamic portfolios will always end up having vanishingly small variance.

As a separate challenge, we provide a simple argument in Appendix D that neither the sparsity of α nor the sparsity of Σ, which are assumptions required for the tests introduced in FLY, is 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. A combination of these observations motivates us to seek an alternative testing approach.

Our analysis encompasses two cases:

1. If a researcher considers a classical set of assets where P is finite but large, she needs to decide whether only a few of the largest eigenvalues are exploding. This can be accomplished via the Onatski (2009) test.[3] These extracted top$-K$ factors need to be

[2]For example, we can include standard, linear portfolios such as those in Jensen, Kelly and Pedersen (2023), as well as portfolios based on quadratic interactions of characteristics, cubic interactions, etc., ad infinum.

[3]Note, however, that this test, as most other existing tests, relies on $\|\Sigma_P^{-1}\|$ staying bounded.

used to form R_t^M, and then the researcher may proceed by assuming that the residual covariance matrix Σ_P is bounded. No further normalization is necessary.

2. If the researcher considers a rich test portfolio environment, then, by Theorem B.1, all eigenvalues of the residual covariance matrix grow like P. In this case, test asset returns need to be rescaled by $P^{-1/2}$ (see Equations (B.4) and (B.5) in the Appendix); equivalently we need to be considering Σ_P/P as $P \to \infty$. Thus, in actual empirical work we are dividing test asset returns by $P^{1/2}$ because Theorem B.1 justifies the consideration of these rescaled returns even as $P \to \infty$.

2.3 The Setup

We make the following technical assumptions to establish the inference results and cover both cases discussed here.

Assumption 1 *Returns, perhaps after a rescaling, 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 $P \to \infty$.*
- *$\|\Sigma_P\|$ is uniformly bounded as $P \to \infty$, and the eigenvalue distribution of Σ_P converges to a non-degenerate distribution (see Definition 1 in Appendix E).*
- *We make no assumptions about $\|\Sigma_P^{-1}\|$, which may or may not be bounded as $P \to \infty$.*
- *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 preclude unbounded eigenvalues of the covariance matrix of test assets themselves.

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

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

where

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

We then define the estimated residuals,

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

as well as their sample mean vector

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

and their covariance

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

2.4 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,

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

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. We present novel approach where the covariance matrix regularization shrinks the matrix towards zI (ridge-regularization). This allows us to avoid the bounded inverse requirements and retain rotational invariance, but introduces additional challenges in the form of asymptotic biases in the test statistic. We address these challenges in what follows.

2.5 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:

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

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.[4] 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 E).

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

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

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

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

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

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 $\varepsilon_1, \varepsilon_2 > 0$,[5] 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, $\xi(z; \Sigma_P; c) \in [c\varepsilon_1, c\varepsilon_2]$.

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 $\xi(z; \Sigma_P; c)$ in a way that does not depend on Σ_P.

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

[5] 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).

RMT implies that, although $\hat{\Sigma}_P$ is random, $\hat{m}(-z)$ converges to a deterministic limit. See Appendix E 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

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

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 F 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 E (formula (E.13)), there is an implicit shrinkage parameter $Z_* > z$ such that

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

where

$$Z_* \;=\; Z_*(z;c) \;=\; z(1+\xi(z;c)). \tag{28}$$

Hence, Z_* can be estimated using $\hat{Z}_*(z;c) \;=\; z(1+\hat{\xi}(z;c))$. A similar bias is present in our test statistic:

$$\begin{aligned}\lim_{P,T\to\infty,P/T\to c}\underbrace{\bar{\alpha}'(zI+\hat{\Sigma}_P)^{-1}\bar{\alpha}}_{test\ with\ sample\ \hat{\Sigma}_P} \;&=\; \lim_{P,T\to\infty,P/T\to c}\frac{Z_*}{z}\bar{\alpha}'(Z_*I+\Sigma_P)^{-1}\bar{\alpha}\\ > \lim_{P,T\to\infty,P/T\to c}&\underbrace{\bar{\alpha}'(zI+\Sigma_P)^{-1}\bar{\alpha}}_{test\ with\ true\ \Sigma_P}\,.\end{aligned} \tag{29}$$

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.

2.6 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

$$\hat{\theta}^2_\alpha(z) = \bar{\alpha}' \hat{\Sigma}_P(z)^{-1} \bar{\alpha} \,. \tag{30}$$

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

$$\hat{\theta}^2_*(z) = \hat{\theta}^2_\alpha(z) + \hat{\theta}^2_M \,. \tag{31}$$

Thus, our test statistic can be represented as

$$\hat{W}(z) = \frac{\hat{\theta}^2_\alpha(z)}{1 + \hat{\theta}^2_M} = \frac{\hat{\theta}^2_*(z) - \hat{\theta}^2_M}{1 + \hat{\theta}^2_M} \tag{32}$$

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

$$\begin{aligned} \hat{\alpha}_t &= \alpha + \hat{\varepsilon}_t, \\ \hat{\varepsilon}_t &= (\beta - \hat{\beta}) R_t^M + \varepsilon_t \,, \end{aligned} \tag{33}$$

so that the average residuals are given by

$$\bar{\alpha} = \alpha + \bar{\varepsilon},\ \bar{\varepsilon} = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t, \tag{34}$$

and the test statistic (18) can be decomposed as

$$\begin{aligned} \hat{W}(z) &= \frac{\bar{\alpha}' \hat{\Sigma}_P(z)^{-1} \bar{\alpha}}{1 + \hat{\theta}_M^2} \\ &= (1 + \hat{\theta}_M^2)^{-1} \left(\underbrace{\alpha' \hat{\Sigma}_P(z)^{-1} \alpha}_{true\ alpha} + \underbrace{\bar{\varepsilon}' \hat{\Sigma}_P(z)^{-1} \bar{\varepsilon} + 2\bar{\varepsilon}' \hat{\Sigma}_P(z)^{-1} \alpha}_{noise} \right) . \end{aligned} \tag{35}$$

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 I.1 in Appendix I for a simple proof in the Gaussian case). Further, $(1 + \hat{\theta}_M^2)^{-1} \bar{\varepsilon}' \hat{\Sigma}_P(z)^{-1} \bar{\varepsilon}$ 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

$$\lim_{P,T\to\infty,P/T\to c} \frac{T^{1/2}(\hat{W}(z)-\hat{\xi}(z))}{\hat{\Delta}(z;c)^{1/2}} \sim N\left(\lim_{P,T\to\infty,P/T\to c} \frac{T^{1/2}\,\alpha'\,\hat{\Sigma}_P(z)^{-1}\,\alpha}{\hat{\Delta}(z;c)^{1/2}(1+\hat{\theta}_M^2)},\ 1\right). \tag{36}$$

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 $\alpha'\,\hat{\Sigma}_P(z)^{-1}\,\alpha = O(T^{-1/2})$. We therefore make the following assumption.

Assumption 2 $\alpha = T^{-1/4}\tilde{\alpha}$, *where* $\tilde{\alpha}$ *is uniformly bounded and the limit*

$$m_\Sigma(-z;\alpha) = \lim_{P\to\infty} T^{1/2}\alpha'(zI+\Sigma_P)^{-1}\alpha \tag{37}$$

exists for any $z>0$.

Under Assumption 2, the asymptotic power of the test satisfies

$$\begin{aligned}
&\lim_{P,T\to\infty,P/T\to c} \text{Probability of rejection} \\
&= \lim_{P,T\to\infty,P/T\to c} \text{Prob}\left(\frac{T^{1/2}(\hat{W}(z)-\hat{\xi}(z))}{\hat{\Delta}(z;c)^{1/2}} > \zeta\right) \\
&= \Phi\left(-\zeta + \lim_{P,T\to\infty,P/T\to c} \frac{T^{1/2}\,\alpha'\,\hat{\Sigma}_P(z)^{-1}\,\alpha}{\hat{\Delta}(z;c)^{1/2}(1+\hat{\theta}_M^2)}\right),
\end{aligned} \tag{38}$$

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 (38) 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 E), it is possible to prove the following result.

Theorem 6 *Let Z_* be the effective shrinkage (see (29)). Under Assumption 2, the following limit exists in probability:*

$$\mathcal{H}(z;c) \;=\; \lim_{T,P\to\infty, P/T\to c} \frac{T^{1/2}\alpha'(zI+\hat{\Sigma}_P)^{-1}\,\alpha}{1+\hat{\theta}_M^2} \;=\; \frac{z^{-1}Z_* m_\Sigma(-Z_*;\alpha)}{1+\hat{\theta}_M^2} \tag{39}$$

Theorem 6 implies that the true alpha component $T^{1/2}\alpha'(zI+\hat{\Sigma}_P)^{-1}\,\alpha$ in the test statistic (35) converges to a constant when P, T are large. This constant is independent of the sample and depends only on c, Σ_P, α. 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*

$$\frac{T^{1/2}(\hat{W}(z)-\hat{\xi}(z))}{\hat{\Delta}(z;c)^{1/2}} \;\to\; N\left(\frac{\mathcal{H}(z;c)}{\Delta(z;c)^{1/2}},\; 1\right) \tag{40}$$

in distribution, where $\mathcal{H}(z;c)$ is defined in (39). In particular, the power of the test satisfies

$$\begin{aligned} \textit{Probability of rejection} = \;& \text{Prob}\left(\frac{T^{1/2}(\hat{W}(z)-\hat{\xi}(z))}{\hat{\Delta}(z;c)^{1/2}} > \zeta\right) \\ \to \Gamma_{asym}(z;c) \;&=\; \Phi\left(-\zeta + \frac{\mathcal{H}(z;c)}{\Delta(z;c)^{1/2}}\right), \end{aligned} \tag{41}$$

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 I, Proposition I.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 expected return on the portfolio* $(zI + \hat{\Sigma}_P)^{-1}\bar{\alpha}$ *satisfies*

$$\lim_{T,P\to\infty,P/T\to c} T^{1/2} E_T[\bar{\alpha}'(zI + \hat{\Sigma}_P)^{-1}\hat{\alpha}_{T+1}] \;=\; \mathcal{H}(z;c)\,, \tag{42}$$

where convergence is in probability.

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 (2024c). In the Appendix G, 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 $P^{-1}\operatorname{tr}(A)$. Namely, we let z takes values $z = \tilde{z} \cdot P^{-1}\operatorname{tr}(A),\ \tilde{z} \in grid = [10^{-5}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 20, 10^2]$.

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

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

So, the optimal ridge penalty amounts to scaling $P^{-1}\operatorname{tr}(T^{-1}\hat{\alpha}'\hat{\alpha})$ 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 H, 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 *We have*

$$\frac{\mathcal{H}(z;c)}{\Delta(z;c)^{1/2}} = c^{-1/2}\frac{1}{1+\hat{\theta}_M^2}\left(\kappa(P)\|\alpha\|^2 + O(c^{-1})\right), \tag{47}$$

where

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

while the asymptotic test power from Theorem 7 satisfies

$$\Gamma_{asym}(z;c) \;=\; \Gamma_{approx}(P;T) \;+\; O(c^{-1}), \tag{49}$$

where

$$\Gamma_{approx}(P;T) \;=\; \Phi\left(-\zeta + c^{-1/2}\frac{1}{1+\hat{\theta}_M^2}\kappa(P)\|\alpha\|^2\right) \tag{50}$$

Strikingly, the asymptotic power of the test for high complexity is independent of the ridge penalty and does not vanish in the infinite complexity limit as long as $\kappa(P)\|\alpha\|^2$ grows like $P^{1/2}$ (or faster)—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 and 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.

The growth of our test power with complexity relies heavily on corresponding growth of the size $\|\alpha\|^2$ of α. Such growth only happens when α is dense, so that sufficiently many

returns have non-zero α. By contrast, if α is sparse and only few returns have α, $\|\alpha\|$ may grow very slowly with P. For example, the α-sparsity-based test statistic J_0 in FLY attains full power when $\|\alpha\|^2$ grows as $\log P$.

4 Analysis of power of the portfolio efficiency test

We evaluate the finite-sample power of our proposed test and benchmark it against the GRS test via Monte Carlo simulations calibrated to our managed-portfolio application.

4.1 Data-Generating Process

Let N denote the cross-section of stocks and T the time dimension. For $j = 1, \ldots, N$ and $t = 1, \ldots, T$, we simulate excess returns

$$r_{j,t+1} \;=\; \mu\big(X_{j,t}\big) \;+\; \beta_j \, R^M_{t+1} \;+\; \varepsilon_{j,t+1},$$

where:

- $\beta_j \overset{i.i.d.}{\sim} \mathcal{U}(-3, 3)$ capture heterogeneous exposures to R^M_{t+1}.
- $R^M_{t+1} \sim \mathcal{N}(\mu_M, \sigma^2_M)$ is the monthly excess return with $\mu_M = 0.58\%$ and $\sigma_M = 4.04\%$ selected to match S&P 500 moments over the past 50 years.
- $\varepsilon_{j,t+1} \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$ is idiosyncratic noise.
- $X_{j,t} = (X_{j,t,1}, \ldots, X_{j,t,F})$ are values of F simulated characteristics for stock j, and[6]

$$\mu(x) \;=\; T^{-1/4}\gamma \, \sin\big(x'W\big), \tag{51}$$

[6]Our selection of the functional form $\sin\big(x'W\big)$ for expected returns ensures that these returns are highly non-linear and non-monotonic in stock characteristics, making it hard to learn in finite samples.

where the weight vector W is sampled once, in the beginning of the simulation, according to $W \sim \mathcal{N}\big(0, I_F\big)$.

We use the scale coefficient γ to control the size of α akin to the non-centrality parameter λ in the GRS test ((14)).

In this simulated environment, the return on the maximal residual Sharpe Ratio portfolio is given by

$$R^*_{t+1} \;=\; \sum_{j=1}^{N} \mu\big(X_{j,t}\big)\, r_{j,t+1}. \tag{52}$$

The non-linearity of the unobservable $\mu(x)$ implies that extracting α requires constructing test assets that are managed portfolios with weights given by non-linear functions $f_k(\cdot),\ k \geq 1$, of stock characteristics:

$$R_{k,t+1} = P^{-1/2} \sum_{j=1}^{N} f_k\big(X_{j,t}\big)\, r_{j,t+1}.$$

While we have full freedoom to choose any non-linear functions f_k we like, here we follow Kelly et al. (2024b), and Didisheim et al. (2024) and define then as *Random Fourier Features:*

$$f_k(x) \;=\; \sin\big(x' W_k\big), \quad W_k \overset{i.i.d.}{\sim} \mathcal{N}(0, I_F).$$

As Kelly et al. (2024b), and Kelly and Malamud (2025) explain, there is nothing special about RFFs, we just use them as a simple way of generating a large number of non-linearities.[7]

4.2 Test Statistic and Size Adjustment

To compute expected test power, we repeat the simulation n times and average the results across the simulations. Everywhere in the sequel, we pick $n = 500$. For each simulation

[7] Back and Ober (2025) show that such managed portfolios can be used to discover complex non-linearities in the SDFs originating from economic equilibrium models.

$i = 1, \ldots, n$, we compute our test statistic

$$\hat{G}_i(z) = \frac{\sqrt{T}\big(\hat{W}^{(i)}(z) - \hat{\xi}^{(i)}(z;c)\big)}{\sqrt{\hat{\Delta}^{(i)}(z)}}.$$

Since Theorem 5 is an asymptotic result, the finite sample test-statistic is only approximately $N(0,1)$ and, as such, may exhibit size distortions. To correct for these distortions, we proceed as follows:

1. Simulate n datasets under the null $\mu(\cdot) = 0$, yielding $\{\hat{G}_i^{H_0}(z)\}_{i=1}^n$.

2. Estimate the empirical CDF $F_G^{H_0}(x)$ of the realized test statistics and then set the 95% cutoff

$$\zeta_{0.95} = \inf\big\{x : F_G^{H_0}(x) \geq 0.95\big\}.$$

3. Simulate n datasets under the H_1, assuming $\mu(x) = \gamma \sin(x'W)$ and then use $\{\hat{G}_i^{H_1}(z)\}_{i=1}^n$ to compute the finite sample power of our test, defined as

$$\Gamma_{\text{sim}}(z;c) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\big\{\hat{G}_i^{H_1}(z) > \zeta_{0.95}\big\}. \tag{53}$$

4.3 Asymptotic power of the test

As described by Theorem 6 and equation (38), the asymptotic power of the test is an explicit, monotone transformation of

$$\mathcal{H}(z;c) = \frac{z^{-1} Z_* m_\Sigma(-Z_*;\alpha)}{1 + \hat{\theta}_M^2}\,. \tag{54}$$

As described in Appendix C, we can obtain a closed-form expression for α and Σ_P, which is needed for m_Σ in (37), within our simulation setup.

For each Monte Carlo replication $i = 1, ..., n$, we compute Σ_P and α using equations (C.10) and (C.11), respectively. We then estimate $Z_*(z;c) = z(1+\hat{\xi})$ per (28) using $\hat{\xi}(z;c)$ from (22) and compute the estimated asymptotic power for sample i:[8]

$$\hat{G}_i^{asym} = \Phi\left(-\zeta_{0.95} + \frac{\mathcal{H}_i(z;c)}{\hat{\Delta}_i(z;c)^{1/2}}\right) \tag{55}$$

The estimated asymptotic power of the test is then given by

$$\Gamma_{asym}(z;c) = \frac{1}{n}\sum_{i=1}^{n} \hat{G}_i^{asym} \tag{56}$$

4.4 Design Parameters and Grid

We let $T \in \{60, 240, 360\}$ and make P vary from 1 to $10T$. We set $N = 500$ (approximately S&P 500) and $F = 10$. The parameter γ in (51) controls the amplitude of the α in the stock returns, $\gamma \in \{0.5, 1.0, 2.0\}$. These values correspond to maximal power of GRS when $P = T/2$ for one of the three entertained values of T (in the same order). Table 1 summarizes all the simulation settings.

4.5 Simulation Results

Figure 1 reports the results of our simulations for $T = 60,\ 240,\ 360$. The plots show the simulated test power Γ_{sim} from (53), the estimated asymptotic power Γ_{asym} from (56) 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

[8]Note the asymptotic power varies for each Monte-Carlo simulation because: 1) The Random Features weights W_k and the weights W in $\mu(x)$ change from a sample to another. 2) $\hat{\xi}(z;c)$ used to compute $Z_*(z;c) = z(1+\hat{\xi})$ and $\hat{\Delta}(z;c)$ depend on $\hat{\Sigma}_P$ which also changes for each Monte Carlo run.

Parameter	Description	Value
N	Number of stocks	500
F	Number of characteristics	10
μ_M, σ_M	Market return mean, SD	0.58%, 4.04%
β_i	Portfolio loadings	$\mathcal{U}(-3, 3)$
$\varepsilon_{i,t}$	Resiudals	$\mathcal{N}(0, 1)$
γ	Alpha amplitude	$\{0.5, 1.0, 2.0\}$
P	Number of managed portfolios	$1, 2, \ldots, 10T$
T	Time periods	$\{60, 240, 360\}$
M	Monte Carlo replications	500

Table 1: Simulation parameter values.

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.

Our RMT-based theoretical approximation agrees well with the actual realized test power, overall. In the cases corresponding to small T and/or γ, the Monte Carlo power is higher than asymptotic as c increases. This is intuitive. Our asymptotic theory holds as $T \to \infty$ and naturally breaks down for small T. Similarly, when γ is small, test power becomes much more sensitive to finite-sample noise as α is harder to detect.

We see that the power of the test 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 across all feasible values of c, while the power of our test reaches 20% or 50% for asymptotic or Monte Carlo power, respectively, when $c \approx 10$, i.e., $P = 10T$.

The combination of the sample size T and the magnitude of alphas γ matters for how

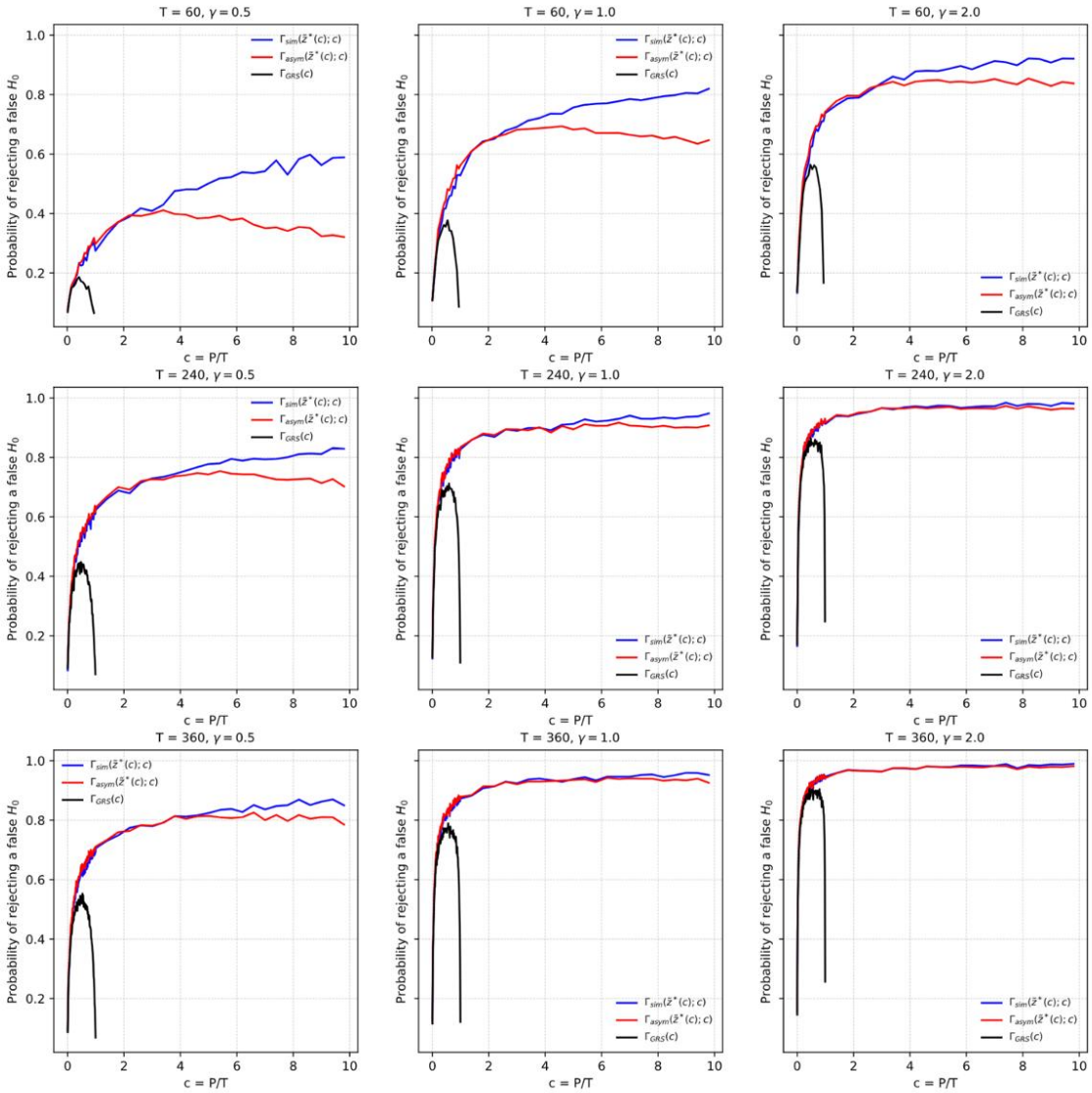

Figure 1: This figure plots the estimated test's power using Monte-Carlo simulation, the asymptotic test's power, and the GRS test's power, as a function of the complexity $c = P/T$ and the optimal ridge penalty, z_*. We pick $T = 60, 240, 360$ (rows), number of simulation paths is $M = 500$ and we let P vary between 10 and $10T$. Different values of γ (columns) capture the variation in alphas.

quickly the test's power reaches the maximum. When T is at about 30 years, this happens very fast, with $c > 1$ delivering high power. With small sample sizes, such as 5 years, the magnitude of alpha is more important with $\gamma = 1.0$ delivering power of 80% at $c = 2$.

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 compiles an extensive collection of stock characteristics from the finance literature.[9] The 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

[9] 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.

both unconditionally and conditionally is given by

$$R^M_{t+1} \;=\; \pi_t' R_{t+1},\; \pi_t \;=\; E_t[R_{t+1}R_{t+1}']^{-1}E_t[R_{t+1}], \tag{57}$$

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

$$R^F_{t+1} \;=\; S_t' R_{t+1} \;\in\; \mathbb{R}^P\,. \tag{58}$$

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

$$\lambda \;=\; E[R^F_{t+1}(R^F_{t+1})']^{-1}E[R^F_{t+1}] \;\in\; \mathbb{R}^P\,. \tag{59}$$

Then,

$$R^M_{t+1} \;=\; \pi_t' R_{t+1} \;=\; \lambda' R^F_{t+1} \;+\; O(1/P). \tag{60}$$

To implement this methodology with real data, we follow Didisheim et al. (2024).[10] 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 (59) via

$$R^M_{t+1}(z_M) \;=\; \hat{\lambda}_t(z_M)' R^F_{t+1}, \tag{61}$$

[10]Note however that our analysis applies directly to other, much more generic Machine Learning settings, such as that from Kelly et al. (2024a).

where

$$\hat{\lambda}_t(z_M) \;=\; \left(z_M I + \frac{1}{T}\sum_{\tau=1}^{T} R^F_{t+1-\tau}(R^F_{t+1-\tau})' \right)^{-1} \frac{1}{T}\sum_{\tau=1}^{T} R^F_{t+1-\tau}\,. \tag{62}$$

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 (58) both as our test assets and as inputs R^F_{t+1} for constructing R^M_{t+1} using (61). 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

$$X_t' R_{t+1} \tag{63}$$

  as *JKP factors.* We also refer to the corresponding portfolio returns R^M constructed in (61) 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).

- $S_t = \mathrm{ReLU}(X_t\gamma) \;\in\; \mathbb{R}^{N_t \times P}$, 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

$$\mathrm{ReLU}(X_t\gamma)' R_{t+1} \tag{64}$$

  as 25,000 RF and to the efficient portfolio constructed from them using (61) as DKKM.[11]

We scale portfolio returns by $P^{-1/2}$ (with $P = 132$ or 25,000) just as in our simulations. But

[11] 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.

because P is fixed in empirical work and the optimal ridge penalty $\tilde{z}_*(c)$ in (46) is defined relative to the variation in α, the test's outcome is entirely invariant to this scaling.

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 F.1, Corollary F.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 2.6). To quantify the economic significance of the alpha, measured by our test statistic, we proceed as follows:

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

$$\hat{\beta}_{t-T,t} \;=\; \frac{1}{T} \sum_{\tau \in [t-T,t]} R_\tau (R^M_\tau - \bar{R}^M_{[t-T,t]}) \tag{65}$$

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

$$\hat{\alpha}^{OOS}_{t+1} \;=\; R_{t+1} \;-\; R^M_{t+1} \hat{\beta}_{t-T,t} \tag{66}$$

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

$$\hat{\alpha}_\tau \;=\; R_\tau \;-\; R^M_\tau \hat{\beta}_{t-T,t},\ \tau \in [t-T, t]\,, \tag{67}$$

  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

$$R^{OOS}_{\alpha,t+1}(z) \;=\; (\hat{\alpha}^{OOS}_{t+1})'\hat{\Sigma}_{P,[t-T,t]}(z)^{-1}\bar{\alpha}_{[t-T,t]}\,, \tag{68}$$

and compute the realized Sharpe ratio of $R^{OOS}_{\alpha,t+1}(z)$ 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 F.3 in the Appendix F.1.

5.3 Time-Series of the Test Statistic

We investigate the ability of our candidate SDFs, FF6, BSV based on (63), and DKKM based on (64) 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 2) and mega stocks (Figure 3).

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 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 (64) 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

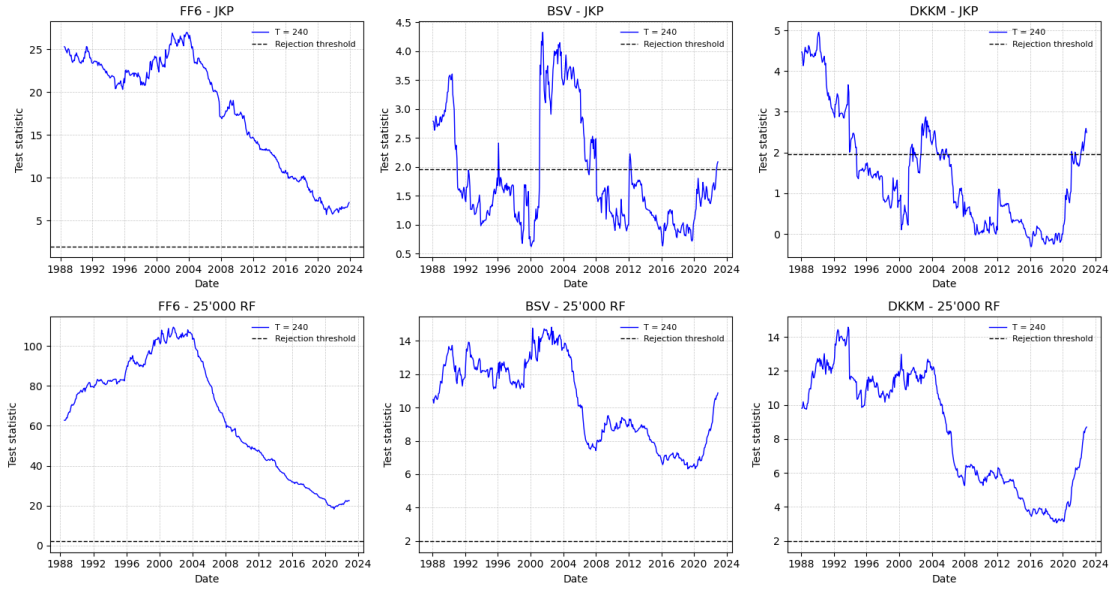

Figure 2: The plot shows the time series dynamics of $\frac{T^{1/2}(\hat{W}(z)-\hat{\xi}(z))}{\hat{\Delta}(z;c)^{1/2}}$ from Theorem 5, computed using $T = 240$. We estimate $\hat{\alpha}$ using (8)–(10). The test assets are non-linear ReLu DKKM factors (64) and linear JKP factors (63) (rows), and we use FF6, BSV, and DKKM (columns) as R_t^M. We consider all 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.

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 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.

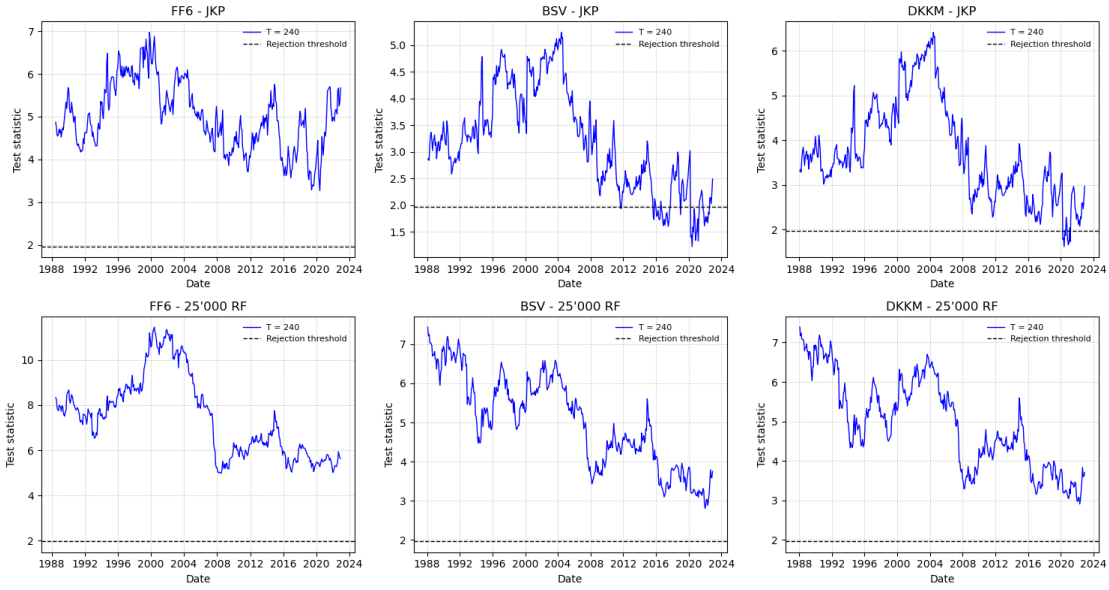

Figure 3: The plot shows the time series dynamics of $\frac{T^{1/2}(\hat{W}(z)-\hat{\xi}(z))}{\hat{\Delta}(z;c)^{1/2}}$ from Theorem 5, computed using $T = 240$. We estimate $\hat{\alpha}$ using (8)–(10). The test assets are non-linear ReLu DKKM factors (64) and linear JKP factors (63) (rows), and we use FF6, BSV, and DKKM (columns) as R_t^M. We consider mega stocks.

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 4 and 5 report the cumulative performance of alpha-portfolios constructed using (68). 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.

Table 3 shows that the realized Sharpe ratios of these alpha portfolios are quite high,

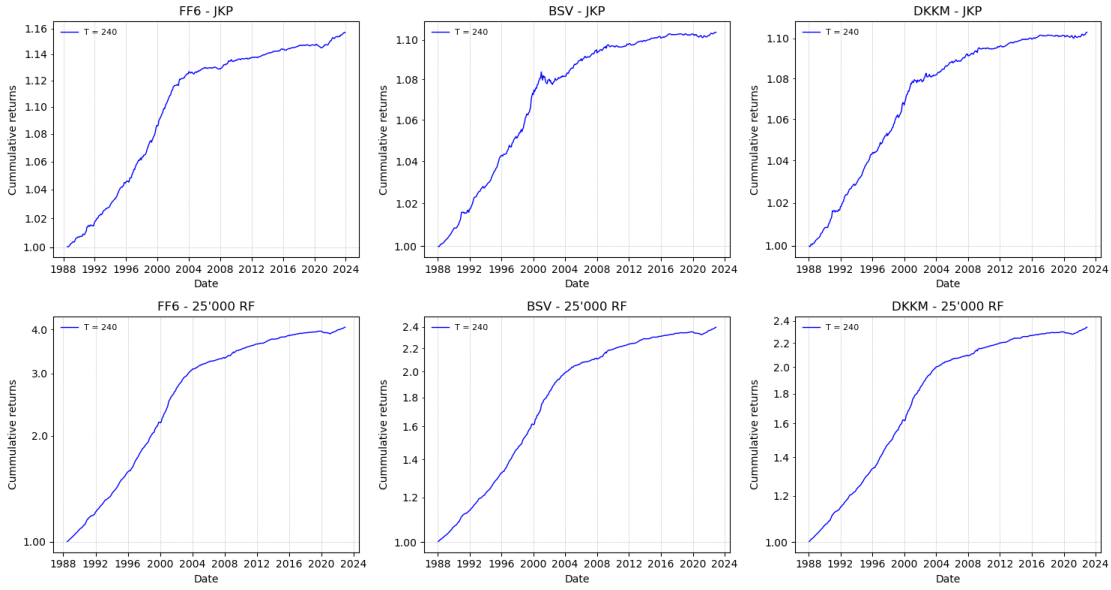

Figure 4: The plot shows the cumulative returns of the α portfolio constructed using (66), computed using $T = 240$. We estimate $\hat{\alpha}$ using (8)–(10). The test assets are non-linear ReLu DKKM factors (64) and linear JKP factors (63) (rows). We use FF6, BSV, and DKKM (columns) as R_t^M. We consider all stocks.

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-sample returns (68) 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.

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 three-fold. First, the original test does not apply to cases where the number of test

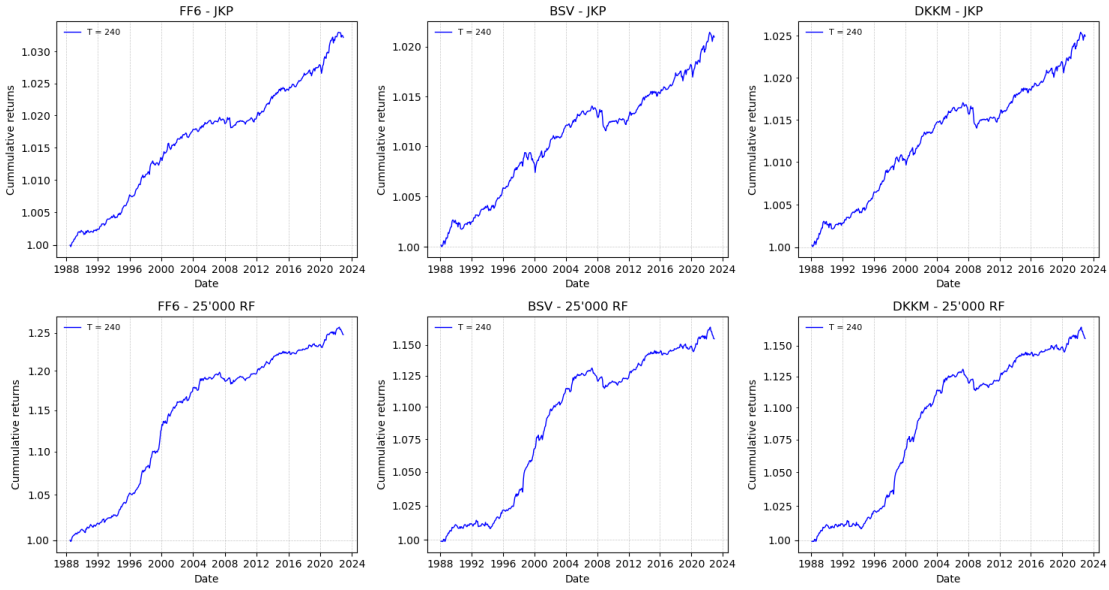

Figure 5: The plot shows the cumulative returns of the α portfolio constructed using (66), computed using $T = 240$. We estimate $\hat{\alpha}$ using (8)–(10). The test assets are non-linear ReLu DKKM factors (64) and linear JKP factors (63) (rows). We use FF6, BSV, and DKKM (columns) as R_t^M. We consider 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.

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

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.

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. Lastly, we demonstrate a third challenge: when one constructs many portfolios, the eigenvalues of the respective covariance matrix are unbounded, or, alternatively, after suitable rescaling of portfolio returns, the eigenvalues of the inverse covariance matrix are unbounded, making almost all existing tests inapplicable.

We manage to address all these issues using two steps. First, we use ridge-regularized covariance matrix, which addresses issues associated with degeneracy or ill-conditioning. Regularization introduces a bias into the GRS statistic, however. Thus, as our second step, we de-bias and rescale the statistic and then use the tools from Random Matrix Theory 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

Back, Kerry and Alexander Ober, "Performance of Factor Models in a Simple Economy," 2025.

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.

Horn, Roger A. and Charles R. Johnson, *Matrix Analysis*, 2nd ed., Cambridge: Cambridge University Press, 2013.

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, Boris Kuznetsov, Semyon Malamud, and Teng Andrea Xu, "Large (and deep) factor models," *arXiv preprint arXiv:2402.06635*, 2024.

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

Kelly, Bryan T and Semyon Malamud, "Understanding The Virtue of Complexity," *Available at SSRN 5346842*, 2025.

___ , ___ **, 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.

Massacci, Daniele, Lucio Sarno, Lorenzo Trapani, and Pierluigi Vallarino, "A general randomized test for Alpha," *arXiv preprint arXiv:2507.17599*, 2025.

Onatski, Alexei, "A formal statistical test for the number of factors in the approximate factor models," *Econometrica*, 2009, *77* (5), 1447–1480.

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.

Zhang, Jun, Wei Lan, Long Feng, and Guanhao Feng, "Testing and Comparing Asset Pricing Factor Models: An Out-of-Sample Perspective," *Available at SSRN 5347838*, 2025.

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:

$$\frac{1}{1+x} = 1 - x + O(x^2) \quad \text{for small } x.$$

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 Breakdown of APT assumptions

In this Appendix, we establish three key properties of managed portfolios (test assets for conditional asset pricing):

- All eigenvalues grow like P for non-linear managed portfolios;
- When multiplied by $1/P$, eigenvalues converge to finite limits.
- These limiting rescaled eigenvalues are never bounded away from zero, implying that $\lim_{P\to\infty} \|\Sigma_P^{-1}\| = \infty$.

Suppose that there are N base securities (stocks) with returns $r_{t+1} = (r_{i,t+1})_{i=1}^N$ and let $X_t \in \mathbb{R}^d$ be an adapted process summarizing conditional information about the first two moments, so that

$$E_t[r_{t+1} r_{t+1}'] = \Sigma(X_t). \tag{B.1}$$

Let also $(\Theta; p(d\theta))$ be a probability space and $f : \mathbb{R}^d \times \Theta \to \mathbb{R}^N$ be a function satisfying the following technical integrability condition: The kernel

$$k(X, \tilde{X}) = \int f(X;\theta) f(\tilde{X};\theta) p(d\theta) \in \mathbb{R}^{N\times N} \tag{B.2}$$

is such that

$$\operatorname{tr} E[k(X,\tilde{X})\Sigma(\tilde{X})k(\tilde{X},X)\Sigma(X)] \;<\; \infty, \tag{B.3}$$

where X and and $\tilde{X}$ are sampled independently from the distribution of X_t. As we show below, this condition also implies that $E[(r'_{t+1}f(X_t;\theta))^2] < \infty$ for $p(d\theta)$-almost every θ, ensuring that the test asset excess returns, defined as

$$\tilde{R}_{k,t+1} \;=\; f_k(X_t;\theta_k)'r_{t+1}\,,\;\; k=1,\cdots,P\,, \tag{B.4}$$

have finite second moments. We also define

$$R_{k,t+1} \;=\; P^{-1/2}\tilde{R}_{k,t+1},\;\; k=1,\cdots,P, \tag{B.5}$$

to be the normalized managed portfolios. Let

$$\begin{aligned}\tilde{R}_{t+1} \;&=\; \tilde{\alpha} \;+\; \tilde{\beta}R^M_{t+1} \;+\; \tilde{\Sigma}_P^{1/2}u_t\\ R_{t+1} \;&=\; \alpha \;+\; \beta R^M_{t+1} \;+\; \Sigma_P^{1/2}u_t\end{aligned} \tag{B.6}$$

be the corresponding $\alpha-\beta-$*residuals* representations, and note that

$$\alpha = P^{-1/2}\tilde{\alpha},\;\; \beta = P^{-1/2}\tilde{\beta},\;\; \Sigma_P = P^{-1}\tilde{\Sigma}_P\,. \tag{B.7}$$

We also denote

$$\Psi_P \;=\; E[R_{t+1}R'_{t+1}] \;=\; P^{-1}\tilde{\Psi} \;=\; P^{-1}E[\tilde{R}_{t+1}\tilde{R}'_{t+1}]\,. \tag{B.8}$$

Theorem B.1 (The Breakdown of APT assumptions) *There exists an infinite sequence*

of positive numbers $\lambda_1 \geq \lambda_2 \geq \cdots$ *such that*

$$\sum_{k=1}^{\infty} \lambda_k^2 \;<\; \infty, \;\; \lim_{k\to\infty} \lambda_k = 0\,. \tag{B.9}$$

The eigenvalues $\lambda_1(\Psi_P) \geq \cdots \geq \lambda_P(\Psi_P)$ *of the second moment matrix* (B.8) *converge to* λ_k *in the sense that*

$$\lim_{P\to\infty} \lambda_k(\Psi_P) \;=\; \lambda_k\,. \tag{B.10}$$

By the Cauchy Interlacing Theorem (Horn and Johnson (2013), Theorem 4.3.17, p. 236), the eigenvalues $\lambda_k(\Sigma_P)$ *of the residual covariance matrix* (B.6) *satisfy*

$$\max(0, \lambda_{k-2}(\Psi_P)) \;\leq\; \lambda_k(\Sigma_P) \;\leq\; \lambda_{k+2}(\Sigma_P)\,. \tag{B.11}$$

In particular, for any $\varepsilon > 0$, *there exist* $K > 0, P^* > 0$ *such that*

$$\lambda_k(\Psi_P) \;<\; \varepsilon \tag{B.12}$$

for all $k > K,\; P > P^*$ *and, hence,*

$$\lim_{P\to\infty} \|\Sigma_P^{-1}\| \;=\; \infty\,. \tag{B.13}$$

Proof of Theorem B.1. Let $Y_{t+1} = (r_{t+1}, X_t)$, and let

$$g(Y;\theta) \;=\; f_k(X;\theta)'r\,. \tag{B.14}$$

Define the positive-definite kernel (Kelly et al. (2024a) call it the "portfolio kernel")

$$K(Y,\tilde{Y}) \;=\; \int g(Y;\theta)g(\tilde{Y};\theta)p(d\theta) \;=\; r'k(X,\tilde{X})\tilde{r}, \tag{B.15}$$

where

$$k(X,\tilde{X}) \;=\; \int f(X;\theta)X(\tilde{X};\theta)'p(d\theta) \;\in\; \mathbb{R}^{N\times N} \tag{B.16}$$

is a positive-definite matrix-valued kernel. Let T_K be the corresponding operator acting on L_2 :

$$T_K f(Y) \;=\; E[K(Y,Y_{t+1})f(Y_{t+1})]\,. \tag{B.17}$$

We have its Monte-Carlo approximation

$$K_P(Y,\tilde{Y}) \;=\; \frac{1}{P}\sum_{k=1}^{P} g(Y;\theta_k)g(\tilde{Y};\theta_k)\,. \tag{B.18}$$

By direct calculation, K_P is a finite-rank kernel and the non-zero eigenvalues of the corresponding kernel operator T_{K_P} on $L_2(\Omega)$ coincide with those of the matrix Σ_P. To prove continuity of eigenvalues and apply the Mercer theorem, it suffices to prove that K is a Hilbert-Schmidt operator, or, equivalently, that the kernel K is square integrable. Then, the operator sequence $T_{K_P}, P \geq 0$, weakly converges to T_K and the convergence of eigenvalues follows.

Let $Y, \tilde{Y}$ be two independent random vectors sampled from the distribution of (r, X).

$$\begin{aligned} & E[(K(Y,\tilde{Y}))^2] \\ = \; & E[r'k(X,\tilde{X})\tilde{r}\tilde{r}'k(\tilde{X},X)r] \\ = \; & E[r'k(X,\tilde{X})\Sigma(\tilde{X})k(\tilde{X},X)r] \\ = \; & \operatorname{tr} E[k(X,\tilde{X})\Sigma(\tilde{X})k(\tilde{X},X)rr'] \\ = \; & \operatorname{tr} E[k(X,\tilde{X})\Sigma(\tilde{X})k(\tilde{X},X)\Sigma(X)] \end{aligned} \tag{B.19}$$

To complete the proof, it remains to establish the result for residuals. The last claim follows because the covariance matrix of residuals is a finite rank perturbation has eigenvalues that interlace original eigenvalues. □

B.1 Managed Portfolios and Eigenvalues of Kernels

Suppose that stock returns satisfy $r_{i,t+1} = \mu(X_{i,t}) + \beta_i R^M_{t+1} + \varepsilon_{i,t+1}$, $\varepsilon_{t+1} \sim N(0,1)$. Let $k(x,\tilde{x})$ be a positive definite kernel admitting a random feature representation (B.2). We can then build managed portfolio returns

$$R_{k,t+1}(1) \;=\; \sum_i f_k(X_{i,t})\, R_{i,t+1}\,. \tag{B.20}$$

We then have, under standard independence assumptions, that

$$E[R_{k,t+1}] \;=\; \underbrace{\sum_i E[f_k(X_{i,t})]\beta_i}_{\beta_k}\, E[R^M_{t+1}] \;+\; \underbrace{\sum_i E[f_k(X_{i,t})\mu(X_{i,t})]}_{\alpha_k}\,. \tag{B.21}$$

The results from above as well as those of Kelly et al. (2024a) imply the following result.

Proposition B.2 *Define the portfolio kernel*

$$K((X,r),(\tilde{X},\tilde{r})) \;=\; r' k(X,\tilde{X})\tilde{r}\,. \tag{B.22}$$

Then, in the limit as $P \to \infty$, *the eigenvalues of the matrix* $P^{-1}(E[R_{k_1,t+1}R_{k_2,t+1}])_{k_1,k_2=1}^{P}$ *converge to those of the kernel operator* T_K. *In particular, returns need to be rescaled by* $P^{-1/2}$.

C Details and Derivations for the Simulation

In this section, we derive explicit expressions for α and Σ in the framework of the simulation described in Section 4.1. We use these expressions in the main text to compute asymptotic test power.

Recall the setup in Section 4.1. Let N denote the cross-section of stocks and T the time dimension. For $j = 1,\ldots,N$ and $t = 1,\ldots,T$, we simulate excess returns

$$r_{j,t+1} \;=\; \mu\big(X_{j,t}\big) \;+\; \beta_j\, R^M_{t+1} \;+\; \varepsilon_{j,t+1},$$

where

$$\mu(x) \;=\; T^{-1/4}\gamma\,\sin\big(x'W\big), \tag{C.1}$$

where the weight vector W is sampled once, in the beginning of the simulation, according to $W \sim \mathcal{N}\big(0, I_F\big)$. Test assets are managed portfolios with weights given by non-linear functions $f_k(\cdot),\ k \geq 1$, of stock characteristics:

$$R_{k,t+1} = P^{-1/2}\sum_{j=1}^{N} f_k\big(X_{j,t}\big)\, r_{j,t+1}.$$

with

$$f_k(x) \;=\; \sin\big(x'W_k\big), \quad W_k \overset{i.i.d.}{\sim} \mathcal{N}(0, I_F).$$

Theorem C.1 *Under the data-generating process described in Section 4.1, we have*

$$\begin{aligned} \alpha &= \Big(T^{-1/4}P^{-1/2}\gamma N e^{-\frac{1}{2}(\|W_k\|^2+\|W\|^2)}\,\sinh(W_k'W)\Big)_{k=1}^{P} \\ \Sigma_P &= E[R_{t+1}R_{t+1}'] - \alpha\alpha' \end{aligned} \tag{C.2}$$

Where

$$E[R_{k,t+1}R_{k_1,t+1}] = \begin{cases} \frac{N}{2P}\big(1-e^{-2\|W_k\|^2}\big) + \gamma^2 T^{-1/2}P^{-1}\big(N(N-1)A(W_k,W) + NB(W_k,W)\big), & k = k_1, \\ NP^{-1}e^{-\frac{1}{2}(\|W_k\|^2+\|W_{k_1}\|^2)}\sinh(W_k'W_{k_1}) & \\ \quad +\gamma^2 T^{-1/2}P^{-1}\big(N(N-1)C(W_k,W_{k_1},W) + ND(W_k,W_{k_1},W)\big), & k \neq k_1, \end{cases} \tag{C.3}$$

With,

$$\begin{aligned} A(w_1,w_2) &= e^{-(\|w_1\|^2+\|w_2\|^2)}\sinh(w_1'w_2)^2, \\ B(w_1,w_2) &= \tfrac{1}{4}\left(1 - e^{-2\|w_1\|^2} - e^{-2\|w_2\|^2} + e^{-2(\|w_1\|^2+\|w_2\|^2)}\cosh(4w_1'w_2)\right), \\ C(w_1,w_2,w_3) &= e^{-\frac{1}{2}(\|w_1\|^2+\|w_2\|^2+2\|w_3\|^2)}\sinh(w_1'w_3)\sinh(w_2'w_3), \\ D(w_1,w_2,w_3) &= \tfrac{1}{2}e^{-\frac{1}{2}(\|w_1\|^2+\|w_2\|^2)}\sinh(w_1'w_2) \\ &\quad - \tfrac{1}{4}e^{-\frac{1}{2}(\|w_1\|^2+\|w_2\|^2+4\|w_3\|^2)}\Big(e^{w_1'w_2}\cosh(2w_3'(w_1-w_2)) - e^{-w_1'w_2}\cosh(2w_3'(w_1+w_2))\Big). \end{aligned} \tag{C.4}$$

Proof of Theorem C.1. We start from the definition

$$\begin{aligned} E[R_{k,t+1}] &= P^{-1/2} E[f_k' r_{t+1}] \\ &= P^{-1/2} \underbrace{E[f_k' \mu]}_{\alpha_k} + P^{-1/2} E[f_k']\, \beta \mu_M, \end{aligned} \tag{C.5}$$

where the second equality follows from the decomposition $r_{t+1} = \mu + \beta\mu_M + \varepsilon_{t+1}$ and the fact that $E[f_k'\varepsilon_{t+1}] = 0$. To compute $E[f_k']$, we use the following observation.

Lemma C.1 *Let $h(x)$ be an odd function, i.e., $h(-x) = -h(x)$, and let Z be a random variable symmetric around 0, i.e., its density f satisfies $f(z) = f(-z)$. Then*

$$E[h(Z)] = 0.$$

Proof of Lemma C.1. By symmetry of Z and oddness of h,

$$\begin{aligned} E[h(Z)] &= \int_{-\infty}^{\infty} h(z) f(z)\, dz \\ &= \int_{-\infty}^{0} h(z) f(z)\, dz + \int_{0}^{\infty} h(z) f(z)\, dz \\ &= \int_{0}^{\infty} h(-z) f(-z)\, dz + \int_{0}^{\infty} h(z) f(z)\, dz \\ &= \int_{0}^{\infty} \big[-h(z) \big] f(z)\, dz + \int_{0}^{\infty} h(z) f(z)\, dz \\ &= 0. \end{aligned} \tag{C.6}$$

□

Since each component of f_k is of the form $\sin(X_{i,t}' W_k)$, which is odd in $X_{i,t}$, and $X_{i,t}$ is symmetric around zero, Lemma C.1 implies $E[f_k'] = 0$. Therefore the term $E[f_k']\,\beta\mu_M$ vanishes, and we only need to compute $E[f_k'\mu]$.

By definition of f_k and μ,

$$\begin{aligned} T^{1/4}E[f_k'\mu] &= E\left[\begin{pmatrix}\sin(X_{1,t}'W_k) & \cdots & \sin(X_{N,t}'W_k)\end{pmatrix}\begin{pmatrix}\gamma\sin(X_{1,t}'W)\\ \vdots \\ \gamma\sin(X_{N,t}'W)\end{pmatrix}\right] \\ &= \gamma\, E\left[\sum_{i=1}^{N}\sin(X_{i,t}'W_k)\,\sin(X_{i,t}'W)\right] \\ &= \gamma N\, E\big[\sin(X_{i,t}'W_k)\,\sin(X_{i,t}'W)\big]\,, \end{aligned} \tag{C.7}$$

where the last equality uses the fact that the summands are i.i.d. across i. To compute the expectation in the last display, we will need

Lemma C.2 *Let $X \sim N(0, I_d)$ and let $a, b \in \mathbb{R}^d$. Then*

$$E[\sin(a'X)\sin(b'X)] = e^{-\frac{1}{2}(\|a\|^2+\|b\|^2)}\,\sinh(a'b).$$

In particular, if $a = b$, then,

$$E\big[\sin(a'X)^2\big] = \frac{1}{2}\left(1 - e^{-2\|a\|^2}\right)$$

Proof of Lemma C.2. Use the trigonometric identity $\sin u \sin v = \frac{1}{2}\{\cos(u-v)-\cos(u+v)\}$ to write

$$E[\sin(a'X)\sin(b'X)] = \tfrac{1}{2}\Big(E[\cos\big((a-b)'X\big)] - E[\cos\big((a+b)'X\big)]\Big).$$

For any $c \in \mathbb{R}^d$, the characteristic function of $X \sim N(0, I_d)$ gives

$$E\left[e^{ic'X}\right] = e^{-\frac{1}{2}\|c\|^2} \quad \Longrightarrow \quad E[\cos(c'X)] = \Re\, E\left[e^{ic'X}\right] = e^{-\frac{1}{2}\|c\|^2}. \tag{C.8}$$

Hence

$$E[\sin(a'X)\sin(b'X)] = \tfrac{1}{2}\left(e^{-\frac{1}{2}\|a-b\|^2} - e^{-\frac{1}{2}\|a+b\|^2}\right).$$

Setting $a = b$ in the equation above proves the second part of the claim. For the first part, let us expand the norms: $\|a-b\|^2 = \|a\|^2 + \|b\|^2 - 2a'b$ and $\|a+b\|^2 = \|a\|^2 + \|b\|^2 + 2a'b$. Factor out $e^{-\frac{1}{2}(\|a\|^2+\|b\|^2)}$ to obtain

$$E[\sin(a'X)\sin(b'X)] = \tfrac{1}{2}\, e^{-\frac{1}{2}(\|a\|^2+\|b\|^2)}\left(e^{a'b} - e^{-a'b}\right) = e^{-\frac{1}{2}(\|a\|^2+\|b\|^2)}\,\sinh(a'b),$$

as claimed. □

Using Lemma C.2 with $a = W_k$, $b = W$, and $X = X_{i,t} \sim N(0, I_d)$ yields

$$T^{1/4}E[f_k'\mu] = \gamma N\, e^{-\frac{1}{2}(\|W_k\|^2+\|W\|^2)}\,\sinh(W_k'W). \tag{C.9}$$

Substituting into the expression for $E[R_{k,t+1}]$ yields

$$E[R_{k,t+1}] = P^{-1/2}T^{-1/4}\gamma N e^{-\frac{1}{2}(\|W_k\|^2+\|W\|^2)}\,\sinh(W_k'W) = \alpha_k, \tag{C.10}$$

as claimed. By the definition of Σ_P,

$$(\Sigma_P)_{k,k_1} = E[R_{k,t+1}R_{k_1,t+1}] - E[R_{k,t+1}]\,E[R_{k_1,t+1}]. \tag{C.11}$$

Since $E[R_{k,t+1}]$ was computed in (C.10), it remains to evaluate $E[R_{k,t+1}R_{k_1,t+1}]$. From the representation $R_{k,t+1} = P^{-1/2}f_k'(\mu + \varepsilon_{t+1})$ with $E[f_k'\varepsilon_{t+1}] = 0$, we have

$$\begin{aligned} E[R_{k,t+1}R_{k_1,t+1}] &= P^{-1}E\left[f_k'(\mu+\varepsilon_{t+1})(\mu'+\varepsilon_{t+1}')f_{k_1}\right] \\ &= P^{-1}E[f_k'f_{k_1}] + P^{-1}E[(f_k'\mu)(f_{k_1}'\mu)]. \end{aligned} \tag{C.12}$$

We first compute $E[f_k' f_{k_1}]$, distinguishing two cases.

Case 1: $k = k_1$**.** We have

$$E[f_k' f_k] = E\left[\sum_{i=1}^{N} \sin^2(X_{i,t}' W_k)\right] = N\, E\big[\sin^2(X_{i,t}' W_k)\big]. \tag{C.13}$$

Since $X_{i,t}' W_k \sim N(0, \|W_k\|^2)$ and W_k is fixed after sampling, Lemma C.2 yields

$$E\big[\sin^2(X_{i,t}' W_k)\big] = \tfrac{1}{2}\big(1 - e^{-2\|W_k\|^2}\big), \tag{C.14}$$

so that

$$E[f_k' f_k] = \tfrac{N}{2}\big(1 - e^{-2\|W_k\|^2}\big). \tag{C.15}$$

Case 2: $k \neq k_1$**.** Here,

$$E[f_k' f_{k_1}] = E\left[\sum_{i=1}^{N} \sin(X_{i,t}' W_k)\, \sin(X_{i,t}' W_{k_1})\right] = N\, E\big[\sin(X_{i,t}' W_k)\, \sin(X_{i,t}' W_{k_1})\big]. \tag{C.16}$$

Applying Lemma C.2 gives

$$E[f_k' f_{k_1}] = N\, e^{-\frac{1}{2}(\|W_k\|^2 + \|W_{k_1}\|^2)}\, \sinh(W_k' W_{k_1}). \tag{C.17}$$

This completes the computation of $E[f_k' f_{k_1}]$ in both cases. We now turn to the calculation of $E[(f_k'\mu)(f_k\mu)]$. We will distinguish 2 cases: $k = k_1$ or $k \neq k_1$.

Case 1: $k = k_1$**.** By definition of f_k and μ,

$$\begin{aligned} T^{1/2}E[(f_k'\mu)^2] &= \gamma^2 E\left[\left(\sum_{i=1}^{N} \sin(X_{i,t}'W_k)\ \sin(X_{i,t}'W)\right)^2\right] \\ &= \gamma^2 E\left[\sum_{i,j=1}^{N} \sin(X_{i,t}'W_k)\sin(X_{i,t}'W)\ \sin(X_{j,t}'W_k)\sin(X_{j,t}'W)\right] \\ &= \gamma^2 N(N-1)\cdot \text{Term1} \ +\ \gamma^2 N \cdot \text{Term2}. \end{aligned} \tag{C.18}$$

Term1. For $i \neq j$, independence of $(X_{i,t})$ across i implies

$$\text{Term1} = \left(E[\sin(X_{i,t}'W_k)\sin(X_{i,t}'W)]\right)^2.$$

By Lemma C.2 (with $a = W_k$, $b = W$),

$$\text{Term1} = e^{-(\|W_k\|^2+\|W\|^2)}\ \sinh(W_k'W)^2 := A(W_k, W)$$

Term2. For $i = j$, we have $E[\sin^2(X_{i,t}'W_k)\sin^2(X_{i,t}'W)]$, which is not covered by Lemma C.2. Using $\sin^2 u = \frac{1}{2}(1-\cos 2u)$ and $\cos u \cos v = \frac{1}{2}\{\cos(u-v)+\cos(u+v)\}$ together with

Gaussian moment equation (C.8), we obtain

$$\begin{aligned}
\text{Term2} &= E[\sin^2(X'_{i,t}W_k)\sin^2(X'_{i,t}W)] \\
&= E[\sin^2(U)\sin^2(V)] \\
&= \frac{1}{4}E[(1-\cos(2U))(1-\cos(2V))] \\
&= \frac{1}{4}E[1-\cos(2U)-\cos(2V)+\cos(2U)\cos(2V)] \\
&= \frac{1}{4}E[1-\cos(2U)-\cos(2V)+\frac{1}{2}(\cos(2(U-V))+\cos(2(U+V)))] \\
&= \frac{1}{4}\left(1-e^{-2\|W_k\|^2}-e^{-2\|W\|^2}+\frac{1}{2}(e^{-2\|W_k-W\|^2}+e^{-2\|W_k+W\|^2})\right) \\
&= \frac{1}{4}\left(1-e^{-2\|W_k\|^2}-e^{-2\|W\|^2}+\frac{1}{2}e^{-2(\|W_k\|^2+\|W\|^2)}(e^{4W'_kW}+e^{-4W'_kW})\right) \\
&= \frac{1}{4}\left(1-e^{-2\|W_k\|^2}-e^{-2\|W\|^2}+e^{-2(\|W_k\|^2+\|W\|^2)}\cosh(4W'_kW)\right) := B(W_k,W)
\end{aligned} \tag{C.19}$$

Therefore, in the case $k = k_1$,

$$T^{1/2}E[(f'_k\mu)^2] = \gamma^2 N(N-1)A(W_k,W) \;+\; \gamma^2 N\, B(W_k,W), \tag{C.20}$$

Case 2: $k \neq k_1$**.** Similarly,

$$T^{1/2}E[(f'_k\mu)(f'_{k_1}\mu)] = \gamma^2 N(N-1)\cdot\text{Term1} \;+\; \gamma^2 N\cdot\text{Term2}. \tag{C.21}$$

Term1. For $i \neq j$,

$$\text{Term1} = E[\sin(X'_{i,t}W_k)\sin(X'_{i,t}W)]\; E[\sin(X'_{j,t}W_{k_1})\sin(X'_{j,t}W)].$$

By Lemma C.2,

$$\text{Term1} = e^{-\frac{1}{2}(\|W_k\|^2+\|W\|^2)}\sinh(W'_kW)\cdot e^{-\frac{1}{2}(\|W_{k_1}\|^2+\|W\|^2)}\sinh(W'_{k_1}W) := C(W_k,W_{k_1},W)$$

Term2. For $i = j$, we have $E[\sin(X'_{i,t}W_k)\sin^2(X'_{i,t}W)\sin(X'_{i,t}W_{k_1})]$. Writing $\sin^2 v = \frac{1}{2}(1 - \cos 2v)$ and applying Lemma C.2 for the $\sin(\cdot)\sin(\cdot)$ part, we get

$$\begin{aligned}
&\text{Term2} = E[\sin(X'_{i,t}W_k)\sin(X'_{i,t}W)^2\sin(X'_{i,t}W_{k_1})] \\
&= E[\sin(U)\sin(V)^2\sin(Z)] \\
&= E[\sin(U)\frac{1-\cos(2V)}{2}\sin(Z)] \\
&= \frac{1}{2}\left(E[\sin(U)\sin(Z)] - E[\sin(U)\sin(Z)\cos(2V)]\right) \\
&= \frac{1}{2}\left(e^{-\frac{1}{2}(\|W_k\|^2+\|W_{k_1}\|^2)}\sinh(W'_kW_{k_1}) - E[\sin(U)\sin(Z)\cos(2V)]\right)
\end{aligned} \tag{C.22}$$

Then, using that $\sin(u)\sin(v) = \frac{1}{2}\{\cos(u-v)-\cos(u+v)\}$ and $\cos(u)\sin(v) = \frac{1}{2}(\cos(u+v)+\cos(u-v))$ together with equation (C.8), we get

$$\begin{aligned}
&E[\sin(U)\sin(Z)\cos(2V)] \\
&= \frac{1}{2}\left(E[\cos(U-Z)\cos(2V)] - E[\cos(U+Z)\cos(2V)]\right) \\
&= \frac{1}{4}\left(E[\cos(U-Z+2V)] + E[\cos(U-Z-2V)] - E[\cos(U+Z+2V)] - E[\cos(U+Z-2V)\right) \\
&= \frac{1}{4}\left(e^{-\frac{1}{2}\|W_k-W_{k_1}+2W\|^2} + e^{-\frac{1}{2}\|W_k-W_{k_1}-2W\|^2} - e^{-\frac{1}{2}\|W_k+W_{k_1}+2W\|^2} - e^{-\frac{1}{2}\|W_k+W_{k_1}-2W\|^2}\right) \\
&= \frac{1}{4}\Big(e^{-\frac{1}{2}(\|W_k\|^2+\|W_{k_1}\|^2+4\|W\|^2-2W'_kW_{k_1}+4W'_kW-4W'_{k_1}W)} + e^{-\frac{1}{2}(\|W_k\|^2+\|W_{k_1}\|^2+4\|W\|^2-2W'_kW_{k_1}-4W'_kW+4W'_{k_1}W)} \\
&- e^{-\frac{1}{2}(\|W_k\|^2+\|W_{k_1}\|^2+4\|W\|^2+2W'_kW_{k_1}+4W'_kW+4W'_{k_1}W)} - e^{-\frac{1}{2}(\|W_k\|^2+\|W_{k_1}\|^2+4\|W\|^2+2W'_kW_{k_1}-4W'_kW-4W'_{k_1}W)}\Big) \\
&= \frac{1}{2}\Big(e^{-\frac{1}{2}(\|W_k\|^2+\|W_{k_1}\|^2+4\|W\|^2-2W'_kW_{k_1})}\cosh(2W'(W_k-W_{k_1})) \\
&- e^{-\frac{1}{2}(\|W_k\|^2+\|W_{k_1}\|^2+4\|W\|^2+2W'_kW_{k_1})}\cosh(2W'(W_k+W_{k_1}))\Big) \\
&= \frac{1}{2}e^{-\frac{1}{2}(\|W_k\|^2+\|W_{k_1}\|^2+4\|W\|^2)}\left(e^{W'_kW_{k_1}}\cosh(2W'(W_k-W_{k_1})) - e^{-W'_kW_{k_1}}\cosh(2W'(W_k+W_{k_1}))\right)
\end{aligned} \tag{C.23}$$

So finally,

$$\begin{aligned}
\text{Term2} &= \frac{1}{2}e^{-\frac{1}{2}(\|W_k\|^2+\|W_{k_1}\|^2)}\sinh(W_k'W_{k_1}) \\
&- \frac{1}{4}e^{-\frac{1}{2}(\|W_k\|^2+\|W_{k_1}\|^2+4\|W\|^2)}\Big(e^{W_k'W_{k_1}}\cosh(2W'(W_k-W_{k_1})) - e^{-W_k'W_{k_1}}\cosh(2W'(W_k+W_{k_1}))\Big) \\
&:= D(W_k, W_{k_1}, W)
\end{aligned} \tag{C.24}$$

Thus, in the case $k \neq k_1$,

$$T^{1/2}E[(f_k'\mu)(f_{k_1}'\mu)] = \gamma^2 N(N-1)C(W_k, W_{k_1}, W) \;+\; \gamma^2 N\, D(W_k, W_{k_1}, W), \tag{C.25}$$

Summary. Gathering the results for both $E[f_k'f_{k_1}]$ and $E[(f_k'\mu)(f_{k_1}'\mu)]$, we have

$$E[R_{k,t+1}R_{k_1,t+1}] = \begin{cases} \frac{N}{2P}\big(1-e^{-2\|W_k\|^2}\big) + \gamma^2 T^{-1/2}P^{-1}\big(N(N-1)A(W_k,W) + NB(W_k,W)\big), & k = k_1, \\ NP^{-1}e^{-\frac{1}{2}(\|W_k\|^2+\|W_{k_1}\|^2)}\sinh(W_k'W_{k_1}) & \\ \quad +\gamma^2 T^{-1/2}P^{-1}\big(N(N-1)C(W_k,W_{k_1},W) + ND(W_k,W_{k_1},W)\big), & k \neq k_1, \end{cases} \tag{C.26}$$

with

$$\begin{aligned}
A(w_1,w_2) &= e^{-(\|w_1\|^2+\|w_2\|^2)}\sinh(w_1'w_2)^2, \\
B(w_1,w_2) &= \tfrac{1}{4}\left(1 - e^{-2\|w_1\|^2} - e^{-2\|w_2\|^2} + e^{-2(\|w_1\|^2+\|w_2\|^2)}\cosh(4w_1'w_2)\right), \\
C(w_1,w_2,w_3) &= e^{-\frac{1}{2}(\|w_1\|^2+\|w_2\|^2+2\|w_3\|^2)}\sinh(w_1'w_3)\sinh(w_2'w_3), \\
D(w_1,w_2,w_3) &= \tfrac{1}{2}e^{-\frac{1}{2}(\|w_1\|^2+\|w_2\|^2)}\sinh(w_1'w_2) \\
&\quad - \tfrac{1}{4}e^{-\frac{1}{2}(\|w_1\|^2+\|w_2\|^2+4\|w_3\|^2)}\Big(e^{w_1'w_2}\cosh(2w_3'(w_1-w_2)) - e^{-w_1'w_2}\cosh(2w_3'(w_1+w_2))\Big).
\end{aligned} \tag{C.27}$$

which proves the claim. □

C.1 Computing Eigenvalues

We have (in the notation of Section B) that

$$K(Y,\tilde{Y}) \;=\; r'k(X,\tilde{X})\tilde{r} \tag{C.28}$$

and the corresponding operator

$$T_K f(Y) \;=\; E[r'k(X,\tilde{X})\tilde{r}f(\tilde{X},\tilde{r})]\,. \tag{C.29}$$

If $f(X,r) \;=\; g(X)'r$, we get

$$\begin{aligned} T_K f(Y) \;&=\; E[r'k(X,\tilde{X})\tilde{r}\tilde{r}'g(\tilde{X})] \\ &=\; r'E[k(X,\tilde{X})(\mu(\tilde{X})\mu(\tilde{X})'+I)g(\tilde{X})] \\ &=\; r'E[k(X,\tilde{X})\mu(\tilde{X})\mu(\tilde{X})'g(\tilde{X})] \\ &+\; r'E[k(X,\tilde{X})g(\tilde{X})]\,. \end{aligned} \tag{C.30}$$

Thus, eigenfunctions are $r'g$ where g satisfies

$$\begin{aligned} \lambda g(x) \;&=\; \sum_i E[k(x,\tilde{x}_i)\mu(\tilde{x}_i)\sum_j \mu(\tilde{x}_j)'g(\tilde{x}_j)] \\ &+\; \sum_i E[k(x,\tilde{x}_i)g(\tilde{x}_i)]\,. \end{aligned} \tag{C.31}$$

We will assume for simplicity that x_i are independent across stocks. Then, we can rewrite it as

$$\begin{aligned}\lambda g(x) &= E[k(x,\tilde{x})\mu(\tilde{x})^2 g(\tilde{x})] \\ &+ E[k(x,\tilde{x})\mu(\tilde{x})](N-1)E[\mu(\tilde{x})g(\tilde{x})] + NE[k(x,\tilde{x})g(\tilde{x})] \\ &= E[k(x,\tilde{x})(N+\mu(\tilde{x})^2)g(\tilde{x})] \\ &+ E[k(x,\tilde{x})\mu(\tilde{x})](N-1)E[\mu(\tilde{x})g(\tilde{x})]\end{aligned} \tag{C.32}$$

Define

$$q(x) = E[k(x,\tilde{x})\mu(\tilde{x})]\,. \tag{C.33}$$

Then,

$$\begin{aligned}\lambda g(x) &= E[k(x,\tilde{x})(N+\mu(\tilde{x})^2)g(\tilde{x})] \\ &+ q(x)(N-1)E[\mu(\tilde{x})g(\tilde{x})]\end{aligned} \tag{C.34}$$

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

returns

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

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* $\theta, \tilde{\theta}$, $\mathrm{Cov}(R_{t+1}(\theta), R_{t+1}(\tilde{\theta})) \neq 0$.

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*

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

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 D.2 (Test Statistic Under Linear Transformations) *Suppose that* $R_t = \alpha + \beta R^M_t + \varepsilon_t$*. Define* $R^A_t = AR_t$*.*

- *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 D.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:[12] 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.

E Random Matrix Theory and Power Under Local Alternatives

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

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

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*

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

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

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

as

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

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

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

and, hence, under asymptotic stability,

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

exists and is given by

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

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

By Assumption 1 (asymptotic stability of Σ_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:[13]

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

In the classic case of zero complexity, $\hat{\Sigma}_P$ converges to the true covariance matrix of residuals

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

and, hence, so does the Stieltjes transform:

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

A cornerstone result of RMT is that when $P/T \to c > 0$, (E.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 E.1 (Bai and Zhou (2008)) *For each $z > 0$,*

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

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

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

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

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

The Master equation (E.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.

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

Lemma E.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*

$$P^{-1}\gamma' A_P \gamma \ - \ P^{-1}\operatorname{tr}(A_P) \ \to \ 0 \tag{E.10}$$

in probability.

Lemma E.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 (37). 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. To prove this convergence, we need to introduce notation and results from RMT. Let

$$Z_*(z;c) \ = \ \frac{z}{1 \ - \ c \ + \ c\, z\, m(z;c)} \ = \ z\,(1 + \xi(z;c))\,. \tag{E.11}$$

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

$$z\, m(-z;c) \ = \ Z_*(z;c) m_{\Sigma_P}\left(-Z_*(z;c)\right)\,. \tag{E.12}$$

Didisheim et al. (2024) show that Z_* always satisfies $Z_* > z$ and refer to it as the *effective shrinkage.* Formula (E.12) implies that complexity affects the covariance inverse through a form of *implicit regularization,* whereby, for $c > 0$, we have

$$\lim_{P,T\to\infty, P/T\to c} z(zI + \underbrace{\hat{\Sigma}_P}_{random})^{-1} \ = \ \lim_{P\to\infty} Z_*(Z_* I + \underbrace{\Sigma_P}_{deterministic})^{-1}\,. \tag{E.13}$$

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,

$$m_{\Sigma_P}(-z;\alpha) \;=\; \lim_{P\to\infty} \alpha'(zI+\Sigma_P)^{-1}\alpha \tag{E.14}$$

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

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

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

Theorem E.2 (Bai and Saranadasa (1996); Li et al. (2020b); Liu et al. (2015)) *The following limit exists in probability:*

$$\mathcal{H}(z;c) \;=\; \lim_{T,P\to\infty,P/T\to c} \frac{T^{1/2}\alpha'(zI+\hat{\Sigma}_P)^{-1}\,\alpha}{1+\hat{\theta}_M^2} \;=\; \frac{T^{1/2}z^{-1}Z_* m_{\Sigma_P}(-Z_*;\alpha)}{1+\hat{\theta}_M^2}\,, \tag{E.16}$$

where m_{Σ_P} is defined in (37).

Formula (E.13) implies that a key determinant of the high-complexity behavior is the effective shrinkage parameter Z_*. As we show in the Appendix H, Lemma H.1, it satisfies

$$Z_*(z;c) \;=\; c\,(P^{-1}\,\mathrm{tr}(\Sigma_P)) \;+\; O(1) \tag{E.17}$$

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

gets dominated by Z_*:

$$\begin{aligned}
\lim_{P,T\to\infty,P/T\to c} Z_* m_{\Sigma_P}(-Z_*;\alpha) &= \lim_{P,T\to\infty,P/T\to c} Z_*\alpha'(Z_*I+\Sigma_P)^{-1})\alpha \\
&= \alpha'(I+\Sigma_P/Z_*)^{-1})\alpha \\
&= \lim_{P,T\to\infty,P/T\to c} \|\alpha\|^2 .
\end{aligned} \tag{E.18}$$

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

$$\Delta(z;c) \;=\; 2cz^{-2}(P^{-1}\operatorname{tr}(\Sigma_P^2)) \;+\; O(1)\,. \tag{E.19}$$

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). □

F 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. (2024b). □

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

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

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[14]

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

where

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

and

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

and

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

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

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

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

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

Intuitively, we would like to define a test statistic in terms of the "size" of $\hat{B}C$, and testing

[14]Note that, in the main text, we define residuals $\hat{\alpha}$ using only $X = R^M$ as the design matrix. However, we that define the residual covariance matrix $\hat{\Sigma}_P$ using the demeaned residuals, $\hat{\Sigma}_P = \hat{\alpha}\hat{\alpha}' - \bar{\alpha}\bar{\alpha} = (\hat{\alpha} - \bar{\alpha})(\hat{\alpha} - \bar{\alpha})'$. By direct calculation, $\hat{\alpha} - \bar{\alpha}$ are the residuals if we use $X = [1, R^M]$ as the design matrix. Below, we use this notation.

$BC = 0$ can be done by testing whether $\hat{B}C$ is small. Define

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

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

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

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

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

in the sense of weak convergence.

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

Suppose

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

with R^M being L-dimensional. Let also

$$\hat{\beta} \;=\; \widehat{\mathrm{Cov}}(R, R^M)\hat{\Sigma}_M^{-1} \;\in\; \mathbb{R}^{P\times L}, \tag{F.12}$$

where

$$\begin{aligned}
\bar{R} \;&=\; \frac{1}{T}\sum_{t=1}^{T} R_t \\
\bar{R}^M \;&=\; \frac{1}{T}\sum_{t=1}^{T} R_t^M \in \mathbb{R}^L \\
\widehat{\mathrm{Cov}}(R, R^M) \;&=\; \frac{1}{T}\sum_{t=1}^{T} R_t(R_t^M)' \;-\; \bar{R}\,(\bar{R}^M)' \in \mathbb{R}^{P\times L} \\
\hat{\Sigma}_M \;&=\; \widehat{\mathrm{Var}}[R^M] \;=\; \frac{1}{T}\sum_{t=1}^{T} R_t^M(R_t^M)' \;-\; \bar{R}^M(\bar{R}^M)' \in \mathbb{R}^{L\times L}.
\end{aligned} \tag{F.13}$$

We then define the estimated residuals,

$$\hat{\alpha}_t \;=\; R_t \;-\; \hat{\beta}R_t^M \tag{F.14}$$

as well as their sample mean vector

$$\bar{\alpha} \;=\; \frac{1}{T}\sum_{t=1}^{T} \hat{\alpha}_t \tag{F.15}$$

and their covariance

$$\hat{\Sigma}_P \;=\; \frac{1}{T}\sum_{t=1}^{T}\hat{\alpha}_t\hat{\alpha}_t' \;-\; \bar{\alpha}\bar{\alpha}'\,. \tag{F.16}$$

We also define

$$\hat{W}(z) \;=\; \frac{1}{(1+(\bar{R}^M)'\hat{\Sigma}_M^{-1}\bar{R}^M)}\bar{\alpha}'\hat{\Sigma}_P(z)^{-1}\bar{\alpha}\,. \tag{F.17}$$

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

$$\frac{T^{1/2}(\hat{W}(z)-\hat{\xi}(z))}{\hat{\Delta}(z;c)^{1/2}} \;\to\; N\left(\frac{\mathcal{H}(z;c)}{\Delta(z;c)^{1/2}},\;1\right) \tag{F.18}$$

in distribution, where $\mathcal{H}(z;c)$ is defined in (39). In particular, the power of the test satisfies

$$\begin{aligned} \textit{Probability of rejection} = \;& \text{Prob}\left(\frac{T^{1/2}(\hat{W}(z)-\hat{\xi}(z))}{\hat{\Delta}(z;c)^{1/2}} > \zeta\right) \\ \to \Gamma_{asym}(z;c) \;&=\; \Phi\left(-\zeta+\frac{\mathcal{H}(z;c)}{\Delta(z;c)^{1/2}}\right)\,, \end{aligned} \tag{F.19}$$

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

Proof of Theorem F.2. Let

$$C = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{(L+1)\times 1} \tag{F.20}$$

and

$$X \;=\; (\mathbf{1}', R^M) \;\in\; \mathbb{R}^{(L+1)\times T}\,. \tag{F.21}$$

Now, we have

$$\hat{B} = \begin{pmatrix} \bar{\alpha} & \hat{\beta} \end{pmatrix} = RX'(XX')^{-1} \;\in\; \mathbb{R}^{P\times(L+1)}\,, \tag{F.22}$$

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,

$$\begin{aligned} XX' &= \begin{pmatrix} \mathbf{1}_T' \\ R^M \end{pmatrix} \begin{pmatrix} \mathbf{1}_T & (R^M)' \end{pmatrix} \\ &= \begin{pmatrix} T & T(\bar{R}^M)' \\ T\bar{R}^M & R^M(R^M)' \end{pmatrix} \end{aligned} \tag{F.23}$$

The formula for the inverse of a block matrix

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

is given by:

$$M^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BS^{-1}CA^{-1} & -A^{-1}BS^{-1} \\ -S^{-1}CA^{-1} & S^{-1} \end{pmatrix}, \tag{F.24}$$

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

$$(XX')^{-1} = \begin{pmatrix} T^{-1}(1 + (\bar{R}^M)'\hat{\Sigma}_M^{-1}\bar{R}^M) & -T^{-1}(\bar{R}^M)'\hat{\Sigma}_M^{-1} \\ -T^{-1}\hat{\Sigma}_M^{-1}\bar{R}^M & T^{-1}\hat{\Sigma}_M^{-1} \end{pmatrix} \tag{F.25}$$

Letting $\bar{R} = T^{-1}\sum_{t=1}^{T} R_t \in \mathbb{R}^{P\times 1}$, we get that (F.22) implies that

$$\begin{aligned} \bar{\alpha} &= \begin{pmatrix} R\mathbf{1}_T & R(R^M)' \end{pmatrix} \begin{pmatrix} T^{-1}(1 + (\bar{R}^M)'\hat{\Sigma}_M^{-1}\bar{R}^M) \\ -T^{-1}\hat{\Sigma}_M^{-1}\bar{R}^M \end{pmatrix} \\ &= \left(\bar{R}(1 + (\bar{R}^M)'\hat{\Sigma}_M^{-1}\bar{R}^M) - T^{-1}R(R^M)'\hat{\Sigma}_M^{-1}\bar{R}^M\right). \end{aligned} \tag{F.26}$$

Similarly,

$$\begin{aligned} \hat{\beta} &= \begin{pmatrix} R\mathbf{1}_T & R(R^M)' \end{pmatrix} \begin{pmatrix} -T^{-1}(\bar{R}^M)'\hat{\Sigma}_M^{-1} \\ T^{-1}\hat{\Sigma}_M^{-1} \end{pmatrix} \\ &= \left(T^{-1}R(R^M)'\hat{\Sigma}_M^{-1} - \bar{R}(\bar{R}^M)'\hat{\Sigma}_M^{-1}\right) \end{aligned} \tag{F.27}$$

Let

$$\hat{\Sigma}_P = T^{-1}\left(RR' - \hat{B}XR'\right) \tag{F.28}$$

By (F.20),

$$\begin{aligned} RX'(XX')^{-1}C &= \bar{\alpha} \\ C'(XX')^{-1}C &= T^{-1}(1 + (\bar{R}^M)'\hat{\Sigma}_M^{-1}\bar{R}^M) \\ \mathcal{T} &= (C'(XX')^{-1}C)^{-1/2} = (T^{-1}(1 + (\bar{R}^M)'\hat{\Sigma}_M^{-1}\bar{R}^M))^{-1/2} \end{aligned} \tag{F.29}$$

Thus, by (F.9),

$$\begin{aligned}\hat{M}(z) &= \frac{1}{T}(\hat{B}C\mathcal{T})'\hat{\Sigma}_P(z)^{-1}(\hat{B}C\mathcal{T}) \\ &= \frac{1}{T}\mathcal{T}C'\hat{B}'\hat{\Sigma}_P(z)^{-1}\hat{B}C\mathcal{T} \\ &= \frac{1}{(1+(\bar{R}^M)'\hat{\Sigma}_M^{-1}\bar{R}^M)}\bar{\alpha}'\hat{\Sigma}_P(z)^{-1}\bar{\alpha} \\ &= \hat{W}(z)\,.\end{aligned} \tag{F.30}$$

The proof of Theorem F.2 is complete. □

F.1 Re-expressing The Test Statistic

Let

$$\tilde{\Sigma}_P = \frac{1}{T}RR' \tag{F.31}$$

and note that

$$\hat{\Sigma}_P = \tilde{\Sigma}_P - \frac{1}{T}R(X'(XX')^{-1}X)R'\,. \tag{F.32}$$

It is convenient to define

$$\begin{aligned}U &= X'(XX')^{-1/2} \\ V &= (XX')^{-1/2}C\mathcal{T}\,.\end{aligned} \tag{F.33}$$

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

$$V'V = I \Leftrightarrow \mathcal{T}(C'(XX')^{-1}C)\mathcal{T} = I \Leftrightarrow \mathcal{T}^{-2} = C'(XX')^{-1}C\,. \tag{F.34}$$

Then, we can rewrite

$$\hat{\Sigma}_P \;=\; \tilde{\Sigma}_P \;-\; \frac{1}{T}RUU'R' \tag{F.35}$$

and, by (F.9), we have

$$\begin{aligned}\hat{M}(z) \;&=\; \frac{1}{T}\mathcal{T}C'(XX')^{-1}XR'\hat{\Sigma}_P(z)^{-1}RX'(XX')^{-1}C\mathcal{T}\\ &=\; \frac{1}{T}\mathcal{T}C'(XX')^{-1/2}(XX')^{-1/2}XR'\hat{\Sigma}_P(z)^{-1}RX'(XX')^{-1/2}(XX')^{-1/2}C\mathcal{T}\\ &=\frac{1}{T}V'U'R'\hat{\Sigma}_P(z)^{-1}RUV\,.\end{aligned} \tag{F.36}$$

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 F.1 *We have*

$$\frac{1}{T}U'R'\hat{\Sigma}_P(z)^{-1}RU \;=\; \frac{1}{T}U'R'\tilde{\Sigma}_\alpha(z)^{-1}RU(I-\frac{1}{T}U'R'\tilde{\Sigma}_\alpha(z)^{-1}RU)^{-1} \tag{F.37}$$

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

$$T^{-1}Y'(zI+T^{-1}RR'-\frac{1}{T}YY')^{-1}Y \;=\; T^{-1}Y'(zI+T^{-1}RR')^{-1}Y(I-T^{-1}Y'(zI+T^{-1}RR')^{-1}Y)^{-1}\,. \tag{F.38}$$

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 F.2

$$(A-B)^{-1} \;=\; A^{-1} \;+\; A^{-1}B(A-B)^{-1} \tag{F.39}$$

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

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

□

Thus,

$$\begin{aligned}
&T^{-1}Y'(zI+T^{-1}RR'-\frac{1}{T}YY')^{-1}Y \\
&= T^{-1}Y'((zI+T^{-1}RR')^{-1}+(zI+T^{-1}RR')^{-1}\frac{1}{T}YY'(zI+T^{-1}RR'-\frac{1}{T}YY')^{-1})Y \\
&= T^{-1}Y'(zI+T^{-1}RR')^{-1}Y+T^{-1}Y'(zI+T^{-1}RR')^{-1}\frac{1}{T}YY'(zI+T^{-1}RR'-\frac{1}{T}YY')^{-1}Y\,.
\end{aligned} \tag{F.40}$$

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

$$Q_1 \;=\; Q_2+Q_2Q_1 \;\Leftrightarrow\; Q_1 \;=\; Q_2(I-Q_2)^{-1}\,. \tag{F.41}$$

The proof is complete. □

Corollary F.3 *Let*

$$\begin{aligned}
\mathcal{R}(z) \;&=\; (zI+T^{-1}R'R)^{-1}R'R \;\in\; \mathbb{R}^{T\times T}, \\
U \;&=\; X'(XX')^{-1/2} \\
V \;&=\; (XX')^{-1/2}C\mathcal{T},
\end{aligned} \tag{F.42}$$

and

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

Then,

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

Proof of Corollary F.3. By (F.36),

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

Using Lemma F.1, we thus get

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

By the identity

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

we can rewrite it as

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

□

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

$$X \;=\; (\mathbf{1}', R^M) \;\in\; \mathbb{R}^{(2)\times T}\,. \tag{F.49}$$

Let also

$$\begin{aligned} \mathcal{R}(z) &= (zI + T^{-1}R'R)^{-1}R'R \in \mathbb{R}^{T\times T}, \\ U &= X'(XX')^{-1/2} \in \mathbb{R}^{T\times 2} \\ V &= (XX')^{-1/2}\begin{pmatrix}1\\0\end{pmatrix}\frac{1}{T^{-1/2}(1+\hat{\theta}_M^2)^{1/2}} \end{aligned} \tag{F.50}$$

and

$$\mathcal{D}(z) = \frac{1}{T}U'\mathcal{R}(z)U\,. \tag{F.51}$$

Then,

$$\hat{M}(z) = V'\mathcal{D}(z)(I-\mathcal{D}(z))^{-1}V\,. \tag{F.52}$$

Proof of Corollary F.4. By (F.36),

$$\hat{M}(z) = \frac{1}{T}V'U'R'\hat{\Sigma}_P(z)^{-1}RUV\,. \tag{F.53}$$

Using Lemma F.1, we thus get

$$\begin{aligned} \hat{M}(z) &= \frac{1}{T}V'U'R'\hat{\Sigma}_P(z)^{-1}RUV \\ &= \frac{1}{T}V'U'R'\tilde{\Sigma}_\alpha(z)^{-1}RU(I-\frac{1}{T}U'R'\tilde{\Sigma}_\alpha(z)^{-1}RU)^{-1}V\,. \end{aligned} \tag{F.54}$$

By the identity

$$R'\tilde{\Sigma}_\alpha(z)^{-1}R = (zI+T^{-1}R'R)^{-1}R'R = \mathcal{R}(z), \tag{F.55}$$

we can rewrite it as

$$\begin{aligned}\hat{M}(z) &= \frac{1}{T}V'U'\mathcal{R}(z)U(I-\frac{1}{T}U'\mathcal{R}(z)U)^{-1}V \\ &= V'\mathcal{D}(z)(I-\mathcal{D}(z))^{-1}V\,.\end{aligned} \tag{F.56}$$

□

Lemma F.3 *We have*

$$R^{OOS}_{\alpha,t+1}(z) = q'(zI+T^{-1}R'R)^{-1}U(I-\mathcal{D}(z))^{-1}(XX')^{-1/2}C, \tag{F.57}$$

where

$$q' = (\hat{\alpha}^{OOS}_{t+1})'R \in \mathbb{R}^T\,. \tag{F.58}$$

Proof of Lemma F.3. We have

$$R^{OOS}_{\alpha,t+1}(z) = (\hat{\alpha}^{OOS}_{t+1})'\hat{\Sigma}_{P,[t-T,t]}(z)^{-1}\bar{\alpha}_{[t-T,t]}\,, \tag{F.59}$$

Let $Y = RU$. Similarly to (F.40), we have

$$\begin{aligned}&(zI+T^{-1}RR'-\frac{1}{T}YY')^{-1}Y \\ &= ((zI+T^{-1}RR')^{-1}+(zI+T^{-1}RR')^{-1}\frac{1}{T}YY'(zI+T^{-1}RR'-\frac{1}{T}YY')^{-1})Y \\ &= (zI+T^{-1}RR')^{-1}Y+(zI+T^{-1}RR')^{-1}\frac{1}{T}YY'(zI+T^{-1}RR'-\frac{1}{T}YY')^{-1}Y\,,\end{aligned} \tag{F.60}$$

so that

$$\begin{aligned}&(I-(zI+T^{-1}RR')^{-1}\frac{1}{T}YY')(zI+T^{-1}RR'-\frac{1}{T}YY')^{-1}Y \\ &= (zI+T^{-1}RR')^{-1}Y,\end{aligned} \tag{F.61}$$

implying that

$$\begin{aligned}
&(zI + T^{-1}RR' - \frac{1}{T}YY')^{-1}Y \\
&= (I - (zI + T^{-1}RR')^{-1}\frac{1}{T}YY')^{-1}(zI + T^{-1}RR')^{-1}Y.
\end{aligned} \tag{F.62}$$

We will now use the simple identity

$$(I - AB)^{-1}A = A(I - BA)^{-1} \tag{F.63}$$

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

$$\begin{aligned}
&(I - (zI + T^{-1}RR')^{-1}\frac{1}{T}YY')^{-1}(zI + T^{-1}RR')^{-1}Y \\
&= (zI + T^{-1}RR')^{-1}Y(I - Y'(zI + T^{-1}RR')^{-1}\frac{1}{T}Y)^{-1}.
\end{aligned} \tag{F.64}$$

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

$$\begin{aligned}
\hat{\Sigma}_{P,[t-T,t]}(z)^{-1}\bar{\alpha}_{[t-T,t]} &= \hat{\Sigma}_P(z)^{-1}RX'(XX')^{-1}C \\
&= (zI + T^{-1}RR' - \frac{1}{T}YY')^{-1}Y(XX')^{-1/2}C \\
&= (zI + T^{-1}RR')^{-1}Y(I - Y'(zI + T^{-1}RR')^{-1}\frac{1}{T}Y)^{-1}(XX')^{-1/2}C \\
&= (zI + T^{-1}RR')^{-1}RU(I - \mathcal{D}(z))^{-1}(XX')^{-1/2}C \\
&= R(zI + T^{-1}R'R)^{-1}U(I - \mathcal{D}(z))^{-1}(XX')^{-1/2}C,
\end{aligned} \tag{F.65}$$

The proof is complete. □

F.2 Size Distortions

Proposition F.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*

$$E[\hat{W}(z)] \;\geq\; I\frac{E[\hat{\xi}_t(1+\hat{\xi}_t)^{-1}]}{E[(1+\hat{\xi}_t)^{-1}]}, \tag{F.66}$$

where

$$\hat{\xi}_t \;=\; T^{-1}R_t'(zI + T^{-1}\hat{\Sigma}_{T,t})^{-1}R_t\,. \tag{F.67}$$

Proof of Proposition F.5. We have by (F.56) that

$$E[\hat{W}(z)] \;=\; E[V'\mathcal{D}(z)(I-\mathcal{D}(z))^{-1}V] \;\geq\; V'E[\mathcal{D}(z)](I-E[\mathcal{D}(z)])^{-1}V, \tag{F.68}$$

and

$$E[\mathcal{D}(z)] \;=\; IE[\hat{\xi}_t(1+\hat{\xi}_t)^{-1}]\,. \tag{F.69}$$

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

$$E[\mathcal{D}(z)] \;=\; T^{-1}E[U'R'(zI + T^{-1}RR')^{-1}RU]\,. \tag{F.70}$$

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:

$$\begin{aligned} G(z) &= T^{-1}E[a'R'(zI+T^{-1}RR')^{-1}Rb] \\ &= T^{-1}E[\sum_{t_1,t_2} a_{t_1}b_{t_2}R'_{t_1}(zI+T^{-1}RR')^{-1}R_{t_2}] \\ &= T^{-1}E[\sum_{t_1\neq t_2} a_{t_1}b_{t_2}R'_{t_1}(zI+T^{-1}\hat{\Sigma}_{T,t_1,t_2})^{-1}R_{t_2}(1+\hat{\xi}_{t_1})^{-1}(1+\hat{\xi}_{t_2})^{-1}] \\ &+ T^{-1}E[\sum_{t} a_t b_t R'_t(zI+T^{-1}\hat{\Sigma}_{T,t})^{-1}R_t(1+\hat{\xi}_t)^{-1}] \end{aligned} \tag{F.71}$$

where we have defined

$$\begin{aligned} \hat{\xi}_{t_1} &= T^{-1}R'_{t_1}(zI+T^{-1}\hat{\Sigma}_{T,t_1})^{-1}R_{t_1} \\ \hat{\xi}_{t_2} &= T^{-1}R'_{t_2}(zI+T^{-1}\hat{\Sigma}_{T,t_1,t_2})^{-1}R_{t_2} \\ \hat{\Sigma}_{T,t_1} &= T^{-1}(RR'-R_{t_1}R'_{t_1}) \\ \hat{\Sigma}_{T,t_1,t_2} &= T^{-1}(RR'-R_{t_1}R'_{t_1}-R_{t_2}R'_{t_2})\,. \end{aligned} \tag{F.72}$$

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

$$\begin{aligned} G(z) &= T^{-1}E[\sum_t a_t b_t R'_t(zI+T^{-1}\hat{\Sigma}_{T,t})^{-1}R_t(1+\hat{\xi}_t)^{-1}] \quad = E[\sum_t a_t b_t \hat{\xi}_t(1+\hat{\xi}_t)^{-1}] \\ &= a'bE[\hat{\xi}_t(1+\hat{\xi}_t)^{-1}]\,, \end{aligned} \tag{F.73}$$

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

G Optimal Ridge Penalty

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

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

$\bar{\Sigma}_{T,t}$ 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. (2024c) 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:

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

By the Sherman-Morrison formula,

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

For large T, we can approximate

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

whereas

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

where

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

and we get our estimator

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

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,

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

Then,

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

Similarly,

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

H Asymptotics for $c \to \infty$

Proposition H.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,*

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

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

Proof of Proposition H.1. Substituting $m(z;c) = c^{-1}\tilde{m}(z;c) + (c^{-1}-1)z^{-1}$ into (E.8), we get that the Master equation for $\tilde{m}(-z;c)$ is

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

where dH is the asymptotic distribution of eigenvalues of Σ_P. 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 (H.2) as

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

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

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

Moreover, we can write

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

We will use the fact that

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

And that,

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

Substituting this into (H.3), we get

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

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

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

Now,

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

Note that, formally, we have computed the Taylor expansion

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

where

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

As a result, by analyticity,[15] we also have

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

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

Thus,

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

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 H.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,*

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

Furthermore,

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

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

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

and by (H.5), we have

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

By Taylor expansion, we get

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

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

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

Hence,

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

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

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

□

Lemma H.2 *We have*

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

Proof of Lemma H.2. By Lemma *H.1*, we have

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

Since

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

we get

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

Furthermore,

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

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,

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

Therefore,

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

□

Lemma H.3 *Under Assumption 2, we have*

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

Proof of Lemma H.3. Recall that $\tilde{c}=c^{-1}$ and that, by (E.18),

$$\mathcal{H}(z;c) \;=\; z^{-1}\|\alpha\|^2 \;+\; O(\tilde{c}). \tag{H.31}$$

Now, by Lemma H.2 we have,

$$\Delta(z;c)=2\tilde{c}^{-1}\left(\frac{E[X]E[X^2]}{E[X]^3}+O(\tilde{c})\right)\times\left(\left(\frac{E[X]^2}{z^2}\right)+O(\tilde{c})\right) \tag{H.32}$$

Therefore,

$$\sqrt{\Delta(z;c)} = \tilde{c}^{-1/2}\sqrt{2\frac{E[X^2]}{z^2} + O(\tilde{c})} \tag{H.33}$$

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

$$\sqrt{1+bx} = 1 + \frac{1}{2}bx + O(x^2) \tag{H.34}$$

We get,

$$\sqrt{2\frac{E[X^2]}{z^2} + O(\tilde{c})} = \sqrt{2\frac{E[X^2]}{z^2}}(1 + O(\tilde{c})) \tag{H.35}$$

Hence,

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

Now, we have

$$\begin{aligned}
\frac{1}{\tilde{c}^{-1/2}\sqrt{2\frac{E[X^2]}{z^2}} + O(\tilde{c}^{1/2})} &= \frac{1}{\tilde{c}^{-1/2}\sqrt{2\frac{E[X^2]}{z^2}}(1 + O(\tilde{c}))} \\
&= \frac{1}{\tilde{c}^{-1/2}\sqrt{2\frac{E[X^2]}{z^2}}}(1 - O(\tilde{c}) + O(\tilde{c}^2)) \\
&= \tilde{c}^{1/2}\sqrt{\frac{z^2}{2E[X^2]}}(1 - O(\tilde{c}) + O(\tilde{c}^2)) \\
&= \tilde{c}^{1/2}\sqrt{\frac{z^2}{2E[X^2]}} + O(\tilde{c}^{3/2}))
\end{aligned} \tag{H.37}$$

Finally,

$$\begin{aligned}
\frac{\mathcal{H}(z;c)}{\Delta(z;c)^{1/2}} &= \tilde{c}^{1/2}\frac{1}{1+\hat{\theta}_M^2}\left(\sqrt{\frac{z^2}{2E[X^2]}}+O(\tilde{c})\right) \\
&\times\left\{z^{-1}\|\alpha\|^2+O(\tilde{c})\right\} \\
&= \tilde{c}^{1/2}\frac{1}{1+\hat{\theta}_M^2}\left(\sqrt{\frac{1}{2E[X^2]}}\|\alpha\|^2+O(\tilde{c})\right)
\end{aligned} \tag{H.38}$$

□

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

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

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

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

and

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

Proof of Proposition I.1. We have

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

where

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

implies that

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

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

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

so that

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

while

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

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

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

where we have defined

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

and

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

Thus, we have

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

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

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

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,

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

Let

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

so that

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

Then,

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

Similarly,

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

Computation of $Term1$

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

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

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

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

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

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

And so,

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

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

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

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

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

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

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

Hence,

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

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

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

So, $Term1$ is given by

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

Computation of $Term2$

We have

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

Computation of $Term3$

We have

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

Hence, by combining all the terms together, we get

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

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

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

Hence,

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

Since,

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

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

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

□

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

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

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

satisfies

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

is distribution, where

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

assuming that the last limit exists.

As we can see from Proposition I.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

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

where we have defined

$$\mu^M_T \ = \ E_T[R^M_{T+1}] \, . \tag{I.41}$$

Thus,

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

Thus,

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

because

$$\sum_t (R^M_t - \bar{R}^M) \ = \ 0 \, . \tag{I.44}$$

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. □