Complexity in Factor Pricing Models

Antoine Didisheim, Shikun Ke, Bryan Kelly, and Semyon Malamud*

March 13, 2023

Abstract

We theoretically characterize the behavior of machine learning asset pricing models. We prove that expected out-of-sample model performance—in terms of SDF Sharpe ratio and average pricing errors—is improving in model parameterization (or "complexity"). Our results predict that the best asset pricing models (in terms of expected out-of-sample performance) have an extremely large number of factors (more than the number of training observations or base assets). Our empirical findings verify the theoretically predicted "virtue of complexity" in the cross-section of stock returns and find that the best model combines tens of thousands of factors. We also derive the *feasible Hansen-Jagannathan (HJ) bound*: The maximal Sharpe ratio achievable by a feasible portfolio strategy. The infeasible HJ bound massively overstates the achievable maximal Sharpe ratio due to a *complexity wedge* that we characterize.

Keywords: Portfolio choice, asset pricing tests, optimization, expected returns, predictability

JEL: C3, C58, C61, G11, G12, G14

*Antoine Didisheim is at the University of Lausanne. Shikun Ke is at Yale School of Management. Bryan Kelly is at Yale School of Management, AQR Capital Management, and NBER; www.bryankellyacademic.org. Semyon Malamud is at Swiss Finance Institute, EPFL, and CEPR, and is a consultant to AQR. We are grateful for helpful comments from seminar participants at the Wharton School of Management, Temple University, and the University of California San Diego. We are incredibly grateful to Mohammad Pourmohammadi for his numerous constructive comments and suggestions. Semyon Malamud gratefully acknowledges the financial support of the Swiss Finance Institute and the Swiss National Science Foundation, Grant 100018_192692. AQR Capital Management is a global investment management firm that may or may not apply similar investment techniques or methods of analysis as described herein. The views expressed here are those of the authors and not necessarily those of AQR.

1 Introduction

The finance literature has recently seen rapid advances in return prediction and SDF estimation using highly parameterized machine learning (ML) models (see the survey of (Giglio et al., 2022)). The notable empirical gains of financial ML clash with traditional principles of statistical modeling in finance that espouse a philosophy of parsimony.[1] Until recently, a clear theoretical justification for employing heavy model parameterizations has been lacking. (Kelly et al., 2021) (KMZ henceforth) makes a first step in building the theoretical case for high-dimensional models in financial applications. They prove that under fairly general conditions, the performance of time series forecasting models—both in terms of forecast accuracy and market timing strategy returns—is increasing in model complexity (i.e., in the number of model parameters).

This paper builds upon KMZ in two critical ways. First, we move from a single asset time series setting to a panel setting with an arbitrary number of risky assets. Second, we reorient the statistical objective from time series forecasting to stochastic discount factor (SDF) optimization. These innovations provide a statistical theory of machine learning asset pricing models. Like KMZ, we study a class of high-dimensional ridge estimators that provide an analytical link with the random matrix theory necessary to characterize properties of the SDF estimator when the number of model parameters becomes large. We explicitly derive an SDF's expected out-of-sample Sharpe ratio and pricing errors (i.e., its ability to explain cross-sectional differences in average returns) as a function of its complexity.

The Virtue of Complexity in Asset Pricing Models

Our theoretical development arrives at surprising conclusions about asset pricing model complexity. The central result is that expected out-of-sample SDF performance, both in

[1] "It is important, in practice, that we employ the smallest possible number of parameters for adequate representations" (Box and Jenkins, Time Series Analysis: Forecasting and Control)

terms of Sharpe ratio and pricing errors, is strictly improving in SDF complexity when appropriate shrinkage is employed.

To build intuition for this result, imagine a researcher studying N risky assets (with excess returns R_{t+1}) in a training sample of T observations. She posits an SDF taking the form

$$M^{\star}_{t+1} = 1 - w^{\star}(X_t)' R_{t+1}, \tag{1}$$

noting that this representation is without loss of generality (Hansen and Richard, 1987).[2] The researcher has access to conditioning variables X_t that span the time t information set, but does not know the functional form $w^{\star}$ that relates conditioning variables to SDF weights.

To model the SDF, the researcher opts for a "universal approximator" of $w^{\star}$, such as a wide two-layer neural network, knowing that this provides an arbitrarily close approximation of $w^{\star}$ when sufficiently parameterized (assuming suitable regularity on $w^{\star}$). This approach eschews the alternative of fixing a parametric model, which may be parsimonious but likely introduces specification errors. The approximating model for an individual asset weight $w^{\star}_{i,t}$ is $w_{i,t} = \lambda' S_{i,t}$. Specifically, $S_{i,t} = f(X_{i,t}) = (f_k(X_{i,t}))_{k=1}^{P}$ is a vector of P generated regressors that result from propagating the raw conditioners $X_{i,t}$ through the neural network, while λ is the vector of coefficients that aggregate generated regressors into final SDF weights (see Figure 1). At last, the approximating SDF model may be written

$$M_{t+1} \;=\; 1 \;-\; \underbrace{\lambda'}_{1\times P}\,\underbrace{S_t'}_{P\times N}\,\underbrace{R_{t+1}}_{N\times 1} = 1 - \lambda' F_{t+1}\,, \tag{2}$$

where $w_t \;=\; S_t\lambda$ and S_t is a $N \times P$ matrix that stacks together generated features for all assets. The second equality in (2) highlights that the neural network approximating model

[2] An SDF can be equivalently represented as its projection on the base assets, with the resulting portfolio lying on the mean-variance efficient frontier.

is a high-dimensional factor pricing model. The product $F_{t+1} = S_t' R_{t+1}$ is a vector of P factor portfolio returns, one for each nonlinear "characteristic" in S. In turn, λ interprets the vector of risk prices corresponding to the nonlinear factors.

At this point, the researcher must make a decision. She has already opted out of using a specific parametric model. But now she must decide how large to make the approximating model and faces a cost-benefit tradeoff. With a simple approximating model ($P << T$), the model will generally suffer specification bias, limiting its ability to represent the true SDF. But with P/T close to zero, the variance of the parameter estimates will be controlled. Additionally, the law of large numbers will apply, so the in-sample performance of the SDF will be indicative of the expected out-of-sample performance (assuming data stationarity).

On the other hand, the researcher can use a complex approximating model with $P/T >> 0$. The added flexibility of the complex model allows it to approximate the true SDF better. The cost, of course, is that a large number of parameters will result in estimates with high variance. And the rich parameterization will overfit the training data, and thus in-sample performance will exceed the model's expected out-of-sample performance (by a potentially large margin). Keep in mind that for any chosen degree of model complexity, the researcher has the option of shrinking parameter estimates to manage their variability.

Faced with this dilemma—enjoy the low variance of a parsimonious model, or enjoy the accurate approximation of a complex model—what course should the researcher take? The answer we show is that the model with the highest possible complexity maximizes the expected out-of-sample performance of their SDF. Using an ultra-high dimensional factor model for the SDF in (2) together with ridge shrinkage, the researcher achieves a higher expected out-of-sample Sharpe ratio and lower out-of-sample pricing errors than is possible with fewer parameters.

Comparing a parsimonious parameterization of equation (2) (using, say, P_1 parameters) with a complex specification (with $P > P_1$ parameters) sheds light on why complex models are beneficial in general, and also why they tend to dominate small parametric specifications. When employing a complex model, the researcher decides on model shrinkage *after* seeing the training data. The complex model casts a wide net in model specification to detect which of the many (P) generated features are most effective. Then, through ridge shrinkage, the researcher restricts the effective parameterization by proportionally scaling down coefficients on all features. This accompaniment of ridge shrinkage allows the researcher to control the variance of the complex model.

Absent any prior knowledge of the functional form of $w^{\star}$, a researcher's specification choice for a parsimonious model is analogous to drawing a small random subset P_1 from the larger set of P generated features and discarding the rest. Here, the researcher controls parameter variance by imposing parsimony, which can be considered another form of shrinkage. In essence, a parsimonious specification shrinks the model *before* seeing the data (by forcing $P - P_1$ of the coefficients to precisely zero). It is certainly possible that a researcher can get lucky and select a high-performance parsimonious specification that beats the complex model. But you cannot be lucky on average. On average, the complex model is more informed and, thus, the better bet. It achieves its variance reduction more judiciously because it gathers information from the training data before deciding on (shrunken) parameter estimates.

Our main theoretical contribution—proving the virtue of complexity in asset pricing models—has important research implications. Unless the researcher knows the correct functional form a priori (which requires heroic assumptions, particularly in complicated systems like financial markets), complex models provide a more reliable out-of-sample understanding of the cross-section of returns. In the lingua franca of asset pricing, forty years of research have produced a “factor zoo” of a few hundred characteristic-based factors. Our theory shows that expanding this small zoo into a teeming Noah's ark of factors is optimal by transforming raw asset characteristics into a rich variety of nonlinear signals (buttressed by appropriate shrinkage). Doing so improves the out-of-sample Sharpe ratio of the SDF and reduces out-of-sample pricing errors.

From a technical standpoint, we have overcome a number of new theoretical hurdles relative to KMZ. The panel aspect of the problem means that the behavior of high-dimensional models has some fundamental differences versus the time series problem of KMZ. In time series regression, the random matrix behavior of the time series covariance of signals dictates the behavior of complex models and associated trading portfolios. In the panel problem, behavior is determined by time series properties and equally by the covariance of signals *across assets*. Our analysis is significantly more involved than that in KMZ. It has required the development of novel mathematical techniques to tackle the ultra-high-dimensional model where the number of stocks, periods, and characteristics per stock is comparably large. In this case, describing the joint behavior of estimated factor risk premia and factor covariances presents a significant challenge that we overcome. First, we show that the managed portfolio approach is indeed efficient in recovering the conditionally efficient portfolio *without the need of estimating the conditional covariance matrix* of stock returns. We prove that the out-of-sample performance of our managed-portfolio-based SDF only depends on two objects: The eigenvalue distribution of the signal covariance matrix and the distribution of Sharpe ratios of factor principal components. Perhaps surprisingly, neither the true conditional covariance of stock returns nor the structure of the latent factors driving those returns impacts the SDF performance in the high-dimensional regime. Second, we formalize the known intuition in machine learning that *over-parametrized models have an implicit regularization effect*: In the interpolation regime, having more degrees of freedom allows the machine learning model to choose better, more regular (e.g., smaller norms, less subject to outliers) interpolators. See, e.g., (Belkin, 2021). This paper provides an exact mathematical formalization for this intuition for ridge-penalized portfolios and SDFs in the high-complexity regime. This implicit regularization is responsible for the virtue of complexity: As we increase complexity, the high-dimensional ridge regularizes more, improving the out-of-sample performance.

Empirical Findings

We design data experiments that mirror our theoretical environment in order to evaluate the role of complexity in the performance of empirical asset pricing models. We study the sample of monthly US stocks and a fixed set of 110 stock-level predictors from (Jensen et al., Forthcoming), which correspond to the raw conditioning variables X_t in (1). To bring our theory to the data, we need to study models ranging from parsimonious to highly complex *while holding the information set fixed*. For this, we adapt the machine learning method of random features regression (as used in KMZ) to the SDF estimation problem. This converts the fixed set of 110 raw stock characteristics into any desired number P of "random features." The random features are an augmented set of stock-level characteristics that make flexible use of the information in the raw data by including an arbitrarily rich set of nonlinear transformations of the raw variables. Random features are equivalent to the features engineered in the hidden layer of a wide two-layer neural network.[3] A convenient fact of using random features to vary the complexity of the empirical model is that conditioning information from *all* of the raw features X_t is distributed impartially to *each* of the random features S_t. As a result, each random feature has an ex-ante identical expected contribution to the overall conditioning information in the complex model. An implication is that the order of the random features is irrelevant—the first random feature is on the exact same footing as the last in terms of its predictive potential—so as we vary model size P from one hundred to a hundred thousand, it doesn't matter how we work our way through the list of features. In short, the key point of our random features SDF formulation is that we can

[3]In the first layer of the network, fixed weights (randomly drawn, as opposed to estimated) aggregate the raw inputs X_t which are then fed through a nonlinear activation function to produce the "random features" S_t. In the second layer, the random features are combined with estimated weights to optimize the SDF performance objective (with ridge shrinkage).

evaluate the empirical effect of complexity by simply varying the number of random features in the model.

We summarize this through the main empirical results. First, we document an empirical virtue of complexity in pricing the cross-section of returns. We find that realized out-of-sample performance of the empirical SDF is generally increasing in model complexity. Increasing the number of model parameters consistently raises the out-of-sample SDF Sharpe ratio and reduces its out-of-sample pricing errors in a manner that closely tracks our theoretical predictions. Our empirical "VOC (virtue of complexity) curves," which plot model performance as a function of model complexity, data support the intuition outlined above that the empirical gains from incorporating nonlinearities are large and that improvements in approximation accuracy from larger P dominate the statistical costs of estimating more parameters. Furthermore, our high-complexity model outperforms standard benchmark models (like the Fama-French-Carhart six-factor model) by a large margin.

The virtue of complexity in our empirical asset pricing models appears highly robust. It is not driven by any particular subset of the stock universe. We find nearly identical VOC curves when SDFs are estimated from subsets of the broader sample (for example, among stocks broken into mega, large, small, and micro capitalization groups). Furthermore, contrary to existing critiques of machine learning models arguing that they produce infeasible trading strategies, our results are robust to excluding fast signals: Even when we remove the 20% of fastest moving characteristics (including short-term reversal and idiosyncratic volatility), the performance of the high-complexity model is barely affected.

Next, recent work by (Kozak et al., 2020) suggests that a successful SDF does not require many factors per se because the asset pricing properties of those factors are adequately summarized by a small number of their principal components (PCs).[4] Their "sparse PC-based SDF" cleverly avoids model complexity through a dimension reduction of the factors.

[4]Relatedly, papers such as (Kelly et al., 2020; Lettau and Pelger, 2020; Gu et al., 2020a) demonstrate the success of dimension reduction methods when estimating asset pricing models with a large number of candidate factors.

This begs the question: Can the complex models we study be similarly reduced to achieve similar performance with potentially many fewer parameters? We show that this is not possible. For every model size P, we consider replacing the P generated factors with a smaller number K of their principal components. We show that dimension reduction significantly impaired model performance across all choices of P and K. In other words, attempting to reduce model complexity inevitably sacrifices model performance. Two important properties of high-dimensional models drive this effect. First, the eigenvalue distribution of the factor covariance matrix is so dense that even very large eigenvalues get absorbed by the bulk of the spectrum, making it impossible to estimate the corresponding PCs efficiently. Second, perhaps surprisingly, contrary to the conventional wisdom, even low-variance PCs have a significant Sharpe ratio and, hence, dropping them leads to a drop in performance. While this seems counter-intuitive from the point of view of arbitrage pricing theory, these high Sharpe ratios are infeasible to achieve because low-variance PCs are impossible to estimate. Thus, we should include all of them in the portfolio.

Literature

Our paper is related to several strands of literature about the growing "factor zoo" (see (Cochrane, 2011), (Harvey et al., 2016), (McLean and Pontiff, 2016), (Hou et al., 2020), (Feng et al., 2020), (Jensen et al., Forthcoming)) and modern statistical and machine learning methods for analyzing it. See (Giglio et al., 2022) for a recent overview.

Many papers in this literature focus on predicting asset returns using complex, non-linear models; see (Moritz and Zimmermann, 2016), (Chinco et al., 2019), (Han et al., 2019), (Gu et al., 2020b), (Kozak et al., 2020), (Freyberger et al., 2020), (Avramov et al., 2021), (Guijarro-Ordonez et al., 2021), (Leippold et al., 2022), (Didisheim et al., 2022), and (Kelly et al., 2022). This approach is agnostic about the link between expected returns and the return (conditional) covariance structure, which is necessary for constructing the stochastic discount factor.

Another stream of literature focuses on using machine learning methods to directly construct the SDF from characteristics-based factors, focusing on the explicit link between the pricing kernel and the conditionally efficient portfolio. See, for example, (Chen et al., 2019; Bryzgalova et al., 2020; Liu et al., 2020). Our paper provides a theoretical foundation for this approach. The idea of using principal components of characteristics-based factors to shrink the cross-section of returns is exploited in (Kelly et al., 2020), (Kozak et al., 2018), (Kozak et al., 2020), (Lettau and Pelger, 2020), and (Giglio and Xiu, 2021), who argue that retaining only a few top principal components is sufficient to explain the cross-section of returns. See also (Gagliardini et al., 2016). As we explain above, our empirical results suggest that PC-sparse SDFs are inefficient and cannot capture the nature of non-linearities in the true SDF.

(Kelly et al., 2020) introduce an econometric framework where stock characteristics are explicitly linked to risk because betas concerning latent factors are (linear) functions of characteristics. A series of recent papers extend the analysis of (Kelly et al., 2020) to the case of a non-linear dependence of betas on characteristics. See, e.g., (Chen et al., 2021), (Fan et al., 2022), and (Ma, 2021).[5] All these papers provide evidence that introducing nonlinearities into the latent factor betas improves pricing efficiency. Motivated by the dangers of overfitting, and in stark contrast to our paper, all these papers operate in a low-complexity regime where the number of parameters is small relative to the panel size. Under such low complexity conditions, these papers prove that the true conditional pricing kernel can be efficiently estimated. In this paper, we show that, despite the true SDF being impossible to estimate, high-complexity models do a great job of extracting non-linearities due to the *virtue of complexity.*

Our results are consistent with the recent findings of (Lettau and Pelger, 2020) and

[5] See also (Gagliardini and Ma, 2019) and (Gagliardini et al., 2020) for an overview.

(Bryzgalova et al., 2023)) who argue that many asset pricing factors are *weak* (see also (Giglio et al., 2021)); that is, factor risk premia are too small to be efficiently estimated even when the number of assets in the cross-section is large. Our paper provides empirical evidence for the pervasive nature of the weak factor hypothesis. While conventional wisdom suggests that a few strong factors dominate the cross-section, our findings offer an alternative picture. We argue that the conditional expected returns might be determined by tens of thousands of weak factors. The virtue of complexity is the most direct illustration of this empirical fact: Every weak factor adds a little bit to the out-of-sample performance, but their joint effect is very large.

As in (Kozak et al., 2020), we construct the feasible proxy for the SDF from the maximal Sharpe ratio portfolio of factors. The problem of finding the highest Sharpe ratio combination of characteristics-based factors is equivalent to the problem of finding the optimal parametric portfolio policy in the language of (Brandt et al., 2009). This point of view is exploited in (DeMiguel et al., 2020) and (Jensen et al., 2022) (taking transaction costs into account), in (Simon et al., 2022) (with parametric portfolio weights based on deep learning), in (Chen et al., 2019) using adversarial training, and in (Cong et al., 2021) using reinforcement learning. Our results provide a theoretical basis for the empirical analysis performed in these papers.

Our paper also belongs to the emergent literature about the *limits to learning:* The fact that, in high-dimensional settings, asset pricing models cannot be efficiently estimated, and there exists a *wedge* between the feasible and infeasible model performances. See, (Da et al., 2022) and (Didisheim et al., 2022). In this paper, we explicitly compute the *complexity wedge for the SDF,* offering a framework for a deeper theoretical understanding of ultra-high-dimensional models for conditional SDFs.

2 Complex Pricing Kernel

In this section, we lay down our assumptions and demonstrate how a high-dimensional factor model is equivalent to a two-layer neural network model for the SDF. We start with the following assumption about the relationship between stock returns and characteristics.[6]

Assumption 1 (Complex Pricing Kernel) *There are N assets with excess returns $R_{t+1} = (R_{i,t+1})_{i=1}^N$. Each asset i has a vector of characteristics $X_{i,t} \in \mathbb{R}^d$, and there exists a non-linear function $w : \mathbb{R}^d \to \mathbb{R}$ such that*

$$M_{t,t+1} \equiv 1 - \sum_{i=1}^{N} w^*(X_{i,t}) R_{i,t+1} , \tag{3}$$

is a tradable pricing kernel, so that the excess returns satisfy[7]

$$E_t[R_{i,t+1} M_{t,t+1}] = 0, \; i = 1, \cdots, N . \tag{4}$$

The random matrices $X_t \in \mathbb{R}^{N\times d}$ and random vectors $R_{t+1} \in \mathbb{R}^N$ are independent and identically distributed over time.

The nonlinearity of w^* makes the dependence of the SDF on characteristics *complex* because it drastically increases the potential dimensionality of the function spaces needed to model $w^*(X)$ in (3). We model $w^*(X)$ as belonging to a parametric family of non-linear functions (such as, e.g., neural networks of a given depth): $w^*(X) = w^*(X;\theta)$, $\theta \in \mathbb{R}^P$, and then relate *model complexity* to the number P of parameters needed to characterize the nonlinearity. The larger P is, the more complex the model. When $w^*(X;\theta)$ has enough

[6] In Appendix B, we describe a class of data-generating processes consistent with the pricing kernel (3).

[7] The assumption that the weight of stock i only depends on the characteristics of stock i can be easily relaxed. For example, we may assume to include macroeconomic variables into $X_{i,t}$ for each stock i. In the Appendix B, we provide a setting where the true pricing kernel has approximately the form (3), up to several terms involving symmetric functions of $X_{i,t}$ across stocks. As we show in our proofs, these terms are asymptotically negligible.

expressive power, the family $w^*(X;\theta)$ becomes a universal approximator, allowing us to generate any form of non-linearity. In this paper, we focus on a particular parametric class of w^*: Namely, we assume that w^* can be represented as a combination of *features*,

$$w^*(X_{i,t}) \;=\; \sum_{\ell=1}^{P} \lambda_\ell \, S_{i,\ell,t},\;\; k=1,\cdots,q\,, \tag{5}$$

where

$$S_{i,\ell,t} = f_\ell(X_{i,t}),\;\; \ell = 1,\cdots,P \tag{6}$$

are given by non-linear transformations $f_\ell(\cdot)$ of the original covariates X_t. It is known that the specification (5) has a very large expressive power: With a properly chosen *basis* of nonlinear functions f_ℓ, any sufficiently regular function w can be approximated by a linear expression (5). For example, $f_\ell(\cdot)$ could be chosen as a spline basis, as in (Chen et al., 2021), or deep neural networks (as in (Fan et al., 2022)). In our empirical analysis, we choose $f_\ell(\cdot)$ to be random features (see (Kelly et al., 2021) and Section 6 for details), in which case (5) is equivalent to approximating $w^*(X)$ with a two-layer neural network (see Figure 1). Independent of the choice of the $f_\ell(\cdot)$, we need a large number P of nonlinear characteristics S_ℓ in (6) to be able to approximate a generic non-linear function w^*.

Let

$$F_{k,t+1} \;=\; \sum_{i=1}^{N} S_{i,k,t} R_{i,t+1} \tag{7}$$

be the *managed portfolios.* Henceforth, we refer to $F_{k,t+1},\;\; k=1,\cdots,P$, as *factors.* In the

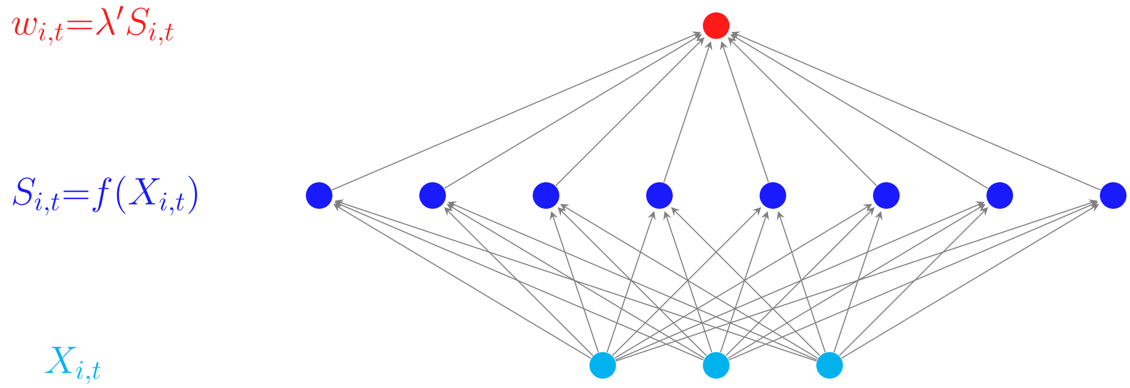


Figure 1: This diagram illustrates how (5) is equivalent to a two-layer neural network with one input layer, one hidden layer, and a single-neuron output layer.

matrix notation,

$$F_{t+1} \;=\; S_t' R_{t+1} \;\in \mathbb{R}^P \,. \tag{8}$$

Substituting (5) into (3), we arrive at the *factor representation for the SDF*:

$$M_{t+1} \;=\; 1 \;-\; \lambda' S_t' R_{t+1} \;=\; 1 \;-\; F_{t+1}'\lambda \,, \tag{9}$$

and the pricing equation (4) implies

$$\begin{aligned} 0 \; &\underbrace{=}_{(4)} S_t' E_t[R_{t+1} M_{t+1}] \; \underbrace{=}_{(9)} \; E_t[S_t' R_{t+1}(1 \;-\; \lambda' F_{t+1})] \\ &\underbrace{=}_{(8)} \; E_t[F_{t+1}(1 - F_{t+1}'\lambda)] \;=\; E_t[F_{t+1}] \;-\; E_t[F_{t+1}F_{t+1}']\,\lambda \,, \end{aligned} \tag{10}$$

and hence

$$E[F_{t+1}] \;=\; E[F_{t+1}F_{t+1}']\,\lambda \,, \tag{11}$$

implying that

$$\lambda \;=\; E[FF']^{-1}E[F]\,. \tag{12}$$

This calculation is based on a key observation: The factor structure of the SDF reduces the problem of computing the conditional SDF to an unconditional problem. Equivalently, managed portfolios efficiently incorporate all conditional information.[8] We will use

$$\Psi \;=\; E[FF'] \;-\; E[F]E[F]' \tag{13}$$

to denote the variance-covariance matrix of factors. This matrix's eigenvalue decomposition captures the factors' risk structure and will play a key role in our analysis. Finally, we will use

$$R^{infeas}_{T+1} \;=\lambda' F_{t+1}\,. \tag{14}$$

to denote the *infeasible* efficient portfolio of an investor with access to an infinite amount of data.

3 Feasible Factor Portfolios, Ridge, and Random Matrix Theory

The principal object of our studies is the finite sample counterpart of the efficient portfolio (12), defined as the *in-sample* solution to a penalized version of the (Britten-Jones, 1999) regression

$$\hat{\lambda}_{INS}(z) \;=\; \arg\min_{\hat{\lambda}}\{\sum_{t=1}^{T}(1-\hat{\lambda}'F_t)^2 \;+\; z\,\|\hat{\lambda}\|^2\}, \tag{15}$$

[8]See, Appendix B for technical details showing that this fact indeed holds for a large class of data-generating processes for stock returns and characteristics.

where z is the *ridge penalty,* used as a regularization to prevent in-sample over-fitting. The subscript INS emphasizes the in-sample nature of $\hat{\lambda}_{INS}$. The solution to (15) is given by

$$\hat{\lambda}_{INS}(z) \;=\; (zI + B_T)^{-1}\bar{F}_T\,, \tag{16}$$

where

$$\bar{F}_T \;=\; \frac{1}{T}\sum_{t=1}^{T} F_t \;\in\; \mathbb{R}^P\,, \tag{17}$$

and

$$B_T \;=\; \frac{1}{T}\sum_{t=1}^{T} F_t F_t' \;\in\; \mathbb{R}^{P\times P} \tag{18}$$

is the sample second moment matrix of factors. We also define the finite sample (feasible) counterpart of the efficient portfolio return (14):

$$R^F_{T+1}(z) \;=\; \hat{\lambda}_{INS}(z)' F_{T+1}\,. \tag{19}$$

Intuitively, we expect that, as T increases, finite sample estimates converge to their population values, and in-sample quantities converge to out-of-sample quantities. This assumption has governed most of the existing asset pricing literature, and our paper could have been considerably shorter under it. Alas, contrary to conventional wisdom, when $P/T \not\to 0$, we have $B_T \not\approx E[FF']$, $\bar{F}_T \not\approx E[F]$, and

$$(zI + B_T)^{-1}\bar{F}_T \;\not\approx\; (zI + E[FF'])^{-1}E[F]\,. \tag{20}$$

The counter-intuitive reality of high-dimensional portfolios renders much of the standard statistical arsenal obsolete in the high-complexity regime. Fortunately, another branch of mathematics allows us to study the properties of equation (19): Random Matrix Theory (RMT). As the name suggests, this branch of mathematics discusses the theoretical properties of large random matrices, such as the $P \times P$ matrix B_T. While some of RMT's predictions are complex, its key insight is remarkably simple: most of the theoretical properties of the quantities, such as (19), can be expressed in quantities known as the *Stieltjes transforms* that we now introduce.[9]

We consider a sequence of models indexed by $P \to \infty$ Each model is characterized by a covariance matrix $\Psi = \Psi_P$ in (13) and a vector of risk premia $\lambda = \lambda_P$. The only assumption we make is that both Ψ_P and λ_P are uniformly bounded as $P \to \infty$. We then introduce the Stieltjes transforms for Ψ and B_T,

$$\begin{aligned} m_\Psi(-z) &= \lim_{P\to\infty} \frac{1}{P} \operatorname{tr}\left((\Psi + zI)^{-1}\right) \\ m(-z;c) &= \lim_{P\to\infty} \frac{1}{P} \operatorname{tr}\left((B_T + zI)^{-1}\right), \end{aligned} \tag{21}$$

provided the limits exist.[10] Since $(\Psi + zI)^{-1}$ and $(B_T + zI)^{-1}$ are the regularized inverse covariance matrices appearing in the two portfolios (20), the two Stieltjes transforms (21) capture the total amount of risk reduction achieved by inverting these matrices for a given level of ridge penalty. We will also need the quantity

$$A(z) = \lim_{P\to\infty} E[F]'(zI + \Psi)^{-1}E[F], \tag{22}$$

describing the dependence of the expected returns of the infeasible portfolio, $F_{t+1}(zI + E[FF'])^{-1}E[F]$, on the shrinkage parameter z.

[9]See KMZ for applications of the Stieltjes transform to high-dimensional regression problems.

[10]As KMZ show, both limits exist when the eigenvalue distribution of Ψ weakly converges to a limit distribution as $P \to \infty$. Furthermore, $m(z;c)$ indeed only depends on z and c.

3.1 Implicit Regularization and the Expected Return of the Efficient Portfolio

We will need an additional technical condition to ensure we can apply RMT to factors.

Assumption 2 *We have that*

$$\frac{1}{P}\big(F_t' A_P F_t \;-\; \mathrm{tr}(\Psi A_P)\big) \;\to\; 0 \tag{23}$$

in L_2 for any uniformly bounded sequence of matrices A_P that are independent of F_t. In particular, the random variables $\frac{1}{P}F_t' A_P F_t$ converge to a non-random limit in probability.

Assumption 2 plays the role of the law of large numbers for the *cross-section of factors*. While in standard applications of the law of large numbers, we estimate quantities by averaging over multiple observations (time), the high complexity limit as $P \to \infty$ allows us to compute non-random averages in the cross-section of factors even though the realization of the factor vector F_t at each period t is random. Indeed, using the identity $F' A_P F_t = \mathrm{tr}(A_P F_t F_t')$, we can argue that, when P is large enough, $P^{-1} F_t F_t' \approx P^{-1}\Psi$ and, hence,

$$P^{-1}\,\mathrm{tr}(A_P F_t F_t') \approx \; P^{-1}\,\mathrm{tr}(A_P \Psi). \tag{24}$$

Assumption 2 formalizes this intuition.[11] We now introduce a key object that will be crucial for understanding the out-of-sample properties of factor portfolios.

So far, we have defined two important portfolios: The infeasible portfolio, only accessible when the true moments of the cross-section of factors are known (as given in equation (14)), and the feasible (penalized) portfolio, which can be estimated in finite samples (as given in equation (19)). To understand the impact of complexity on the feasible portfolio

[11] Establishing (23) for managed portfolios (8) is highly non-trivial. We prove it in the Appendix, Lemma 12. Its proof is extremely complex and has required developing novel techniques for dealing with a joint limit of large N, large T, and large P.

performance, we need to introduce a third one: *The penalized infeasible portfolio*, given by

$$R^{infeas}_{T+1}(z) = \lambda(z)' F_{t+1}, \qquad \text{where } \lambda(z) = (E[FF'] + zI)^{-1} E[F]. \tag{25}$$

Note that the infeasible portfolio (25) is a special case of the penalized infeasible portfolio, with $R^{infeas}_{T+1} = R^{infeas}_{T+1}(0)$. Thus, this portfolio can be thought of as an intermediary between the infeasible portfolio and its feasible counterpart. Intuitively, the penalized infeasible portfolio always underperforms the true infeasible portfolio, as penalization is unnecessary when the true moments are known. See Lemma 2 the appendix.

As we show below, an intricate link exists between the expected return of the feasible portfolio and the expected return of the penalized infeasible portfolio. For any degree c of the complexity of the factor model and any penalization of the feasible portfolio z, there exists a $Z^*(z;c) > z$, such that $E[R^F_{T+1}(z)] = E[R^{infeas}_{T+1}(Z^*(z;c))]$. Remarkably, we can characterize $Z^*(z;c)$ in close form. The following is true.

Theorem 1 *In the limit as $P,\ T \to \infty,\ P/T \to c$, we have*

$$E[R^F_{T+1}(z)] \;\to\; \mathcal{R}_1(z;c) \;=\; \mathcal{R}^{infeas}_1(Z^*(z;c))\,, \tag{26}$$

where $\mathcal{R}^{infeas}_1(z) = E[R^{infeas}_{T+1}(z)] = \mathcal{R}_1(z;0) = \frac{A(z)}{1+A(z)}$ is the expected return on the infeasible portfolio.[12] *The function $Z^*(z;c)$ is the effective shrinkage given by*

$$Z^*(z;c) \;=\; z\,(1+\xi(z;c)) \;\in (z,\ z+c)\,, \tag{27}$$

with

$$\xi(z;c) \;=\; \lim_{P,\ T\to\infty,\ P/T\to c} \frac{1}{T}\,\mathrm{tr}((zI + B_T)^{-1}\Psi)\,. \tag{28}$$

[12]This return corresponds the case when the number of observations T is large relative to P, so that $c = P/T \to 0$.

Furthermore, $Z^*(z;c)$ is monotone increasing in z and c. In the ridgeless limit as $z \to 0$, we have

$$Z^*(z;c) \;\to\; \begin{cases} 0, & c < 1 \\ 1/\tilde{m}(c), & c > 1 \end{cases} \tag{29}$$

where $\tilde{m}(c) > 0$ is the unique positive solution to

$$c - 1 \;=\; \frac{\int \frac{dH(x)}{\tilde{m}(1+\tilde{m}\,x)}}{\int \frac{x dH(x)}{1+\tilde{m}\,x}}, \tag{30}$$

and

To understand the intuition behind the formula (26), consider the effect of increasing the estimation window from $T-1$ to T by adding another observation F_T to our estimation of the in-sample moments in (16). By the Sherman-Morrison formula,[13]

$$(zI + B_T)^{-1} F_T \underbrace{\approx}_{(159)} \frac{1}{1+\xi(z;c)} (zI + B_{T-1})^{-1} F_T\,. \tag{31}$$

When complexity $c = P/T$ is close to zero, so is $\xi(z;c)$. However, with high complexity, the factor $\frac{1}{1+\xi(z;c)}$ acts as an effective shrinkage, dampening the effect of each new observation F_T on the estimation of (16). Theorem 1 formalizes the idea of implicit regularization: In the high complexity regime, $E[R^F_{T+1}(z)]$ behaves like $E[R^{infeas}_{T+1}(z)]$, but with z replaced by $Z_*(z;c)$. Since $Z_*(z;c) \geq z$, high dimensional models shrink (regularize) eigenvalues more. Thus, contrary to conventional wisdom,

$$\begin{aligned} E[R^F_{T+1}(z)] \;&=\; E[F'_{T+1}(zI + B_T)^{-1}\bar{F}_T] \\ \approx\; & E[F]'(\mathbf{Z}^*(\mathbf{z};\mathbf{c})I + E[FF'])^{-1}E[F] \;<\; E[F]'(\mathbf{z}I + E[FF'])^{-1}E[F]\,. \end{aligned} \tag{32}$$

[13] $(zI + B_T)^{-1}F_T \;=\; \frac{1}{1+\frac{1}{T}F'_T(zI+B_{T-1})^{-1}F_T}(zI + B_{T-1})^{-1}F_T$, see (124) in the Appendix.

Strikingly, in the complex regime when $c > 1$, (29) implies that $Z^*(z; c)$ is uniformly bounded away from zero. Thus, even in the *ridgeless limit* when $z \to 0$, the estimated efficient portfolio (19) performs an *implicit regularization* of the highly degenerate $P \times P$ covariance matrix B_T that has rank at most $T < P$, leading to a significant drop in out-of-sample expected returns: When with $z = 0$, expected returns behave like those for the infeasible portfolio with shrinkage $1/\tilde{m}(c)$.

3.2 The Risk of High-Complexity Efficient Portfolios

We will now discuss the second moment of the feasible portfolio, $R^F_{T+1}(z)$. To provide intuition, we first consider the corner case where $E[F] = 0$ and $E[F^2_{k,t+1}] \neq 0$ for all $k = 1, \cdots, P$. In this extreme scenario, every factor has an expected return of zero and a non-zero variance. Consequently, the mean-variance efficient strategy involves buying no assets, i.e., $\lambda = 0$. In the low complexity regime where $P/T \approx 0$, the feasible portfolio converges to this solution: $\hat{\lambda}(z) \approx 0$ for all z. However, in the high complexity case, where an agent has a finite number of observations T and $P/T > 0$, the total estimation error aggregated across all factors can be large. Despite approximately estimating $E[F_{k,t+1}]$ with an error of the order $1/T^{1/2}$ for each fixed k, the total error for estimating $E[F] \in \mathbb{R}^P$ can be significant. In other words, when $P/T > 0$, the feasible portfolio will have a non-zero variance even when the data has zero predictability. The following is true.

Proposition 2 (Estimation Risk) *Suppose that $E[F] = 0$. Then,*

$$\lim_{P\to\infty,\ P/T\to c} E[R^F_{t+1}(z)] \;=\; 0\,, \tag{33}$$

whereas

$$G(z;c) \;=\; \lim_{P\to\infty,\ P/T\to c} E[(R^F_{t+1}(z))^2] \;=\; (z\xi(z;c))' \;>\; 0\,. \tag{34}$$

$G(z;c)$ is monotone decreasing in z and increasing in c, and satisfies $G(z;c) \leq \; c\,z^{-2}$.

To enhance intuition, we consider a simple portfolio strategy that invests proportionally to the historical mean returns, with portfolio weights' vector given by $\bar{F}_T'$ (see (17)):

$$R_{t+1}^M \;=\; \bar{F}_T' \, F_{T+1} \,. \tag{35}$$

Then, under the assumption that $E[F]=0$,

$$E[R_{t+1}^M] \;=\; E[\bar{F}_T' \, F_{T+1}] \;=\; E[\bar{F}_T] \, E[F_{T+1}] \;=\; 0 \,. \tag{36}$$

Yet,

$$\begin{aligned} E[(R_{t+1}^M)^2] \;&=\; E[(\bar{F}_T' \, F_{T+1})^2] \;=\; \operatorname{tr} E[\bar{F}_T \bar{F}_T' F_{T+1} F_{T+1}'] \\ =\; \operatorname{tr} E[\bar{F}_T \bar{F}_T' \Psi] \;&=\; \frac{1}{T^2} \sum_t \operatorname{tr} E[F_t F_t' \Psi] \;=\; \frac{1}{T} \operatorname{tr}(\Psi^2) \geq 0 \end{aligned} \tag{37}$$

As we explain in Proposition 2, numerous small estimation errors accumulate, creating significant risk for the portfolio. The case where $\Psi = I$ and $\frac{1}{T}\operatorname{tr}(\Psi^2) = P/T \to c$ highlights how these errors accumulate across P and increase with complexity.

Using the insight gained from the simpler case of $E[F]=0$, we can now describe the risk of high-complexity portfolios for the general case where $E[F] \neq 0$.

Theorem 3 *We have*

$$E[(R_{T+1}^F(z))^2] \;\to\; \underbrace{\mathcal{R}_2^{infeas}(Z^*(z;c))}_{implicit\ regularization} \;+\; \underbrace{G(z;c)(1 - 2\mathcal{R}_1^{infeas}(Z^*(z;c)) + \mathcal{R}_2^{infeas}(Z^*(z;c)))}_{estimation\ risk}, \tag{38}$$

where

$$\mathcal{R}_2^{infeas}(z) = \mathcal{R}_2(z;0) \;=\; \frac{d}{dz}\left(\frac{zA(z)}{1+A(z)}\right) \tag{39}$$

is the second moment of the return on the infeasible portfolio, $F'_{T+1}(\mathbf{z}I + E[FF'])^{-1}E[F]$, *estimated using* $T = \infty$.

Theorem 3 shows that the variance of the feasible portfolio can be characterized in two terms. The first term, $\mathcal{R}_2^{infeas}(Z^*(z;c))$, is the second moment of the infeasible portfolio with implicit regularization provided by $Z^*(z;c)$, similarly to Theorem 26. Through this regularization, complexity reduces risk and may improve the risk-return tradeoff. At the same time, the risk of the feasible portfolio is impacted by the estimation risk, $G(z;c)$, characterized in Proposition 2. Estimation risk is bounded from above by complexity: By Proposition 2, *it only depends on the eigenvalue distribution of* Ψ, stays positive even when $E[F] = 0$, and satisfies $G(z;c) \le cz^{-2}$. This surprising tradeoff between implicit regularization and estimation risk is the quintessence of complexity and its impact on out-of-sample portfolio performance.

3.3 The Complexity Wedge and the Feasible Hansen-Jagannathan Bound

Low complexity ($c \approx 0$) describes settings with many more observations than parameters to estimate. This is the purview of traditional econometrics. In these conditions, the law of large numbers kicks in, and appropriate estimators tend to recover the true model parameters.[14] Importantly, when $c \approx 0$, a model's in-sample performance indicates its expected out-of-sample performance.

In high complexity settings ($c >> 0$), parameterization is large relative to the number of observations. This is the machine learning case. Here, the law of large numbers breaks down.

[14] If the model is correctly specified. If the model is misspecified, estimators recover the nearest "pseudo-true" parameters.

"Best" estimators fail to converge on the actual parameters because there is not enough data to go around. As a result, complexity creates two wedges critical to understanding model performance, and our first theoretical contribution is to characterize these wedges explicitly.

The first wedge is "overfit." When the number of parameters is large relative to the number of observations, in-sample model performance is exaggerated. Overfit is defined as the difference between the in-sample performance (in terms of SDF Sharpe ratio or pricing errors) of the trained model versus the performance of the true but infeasible model.

$$\text{Overfit} = \text{In-sample Performance} - \text{True Performance}\,.$$

Note that overfit is more about a dearth of training observations than excessive parameters—it exists even for correctly specified models with no "excess" parameters.[15]

Intuitively, heavy parameterization inflates in-sample performance. But by how much? Surprisingly, we can explicitly quantify the extent of overfit using only our knowledge of the training environment. By taking into account the details of the model (specification and the number of parameters) and the training data conditions (the amount of training data and its covariance structure), our theory infers the expected overfit using random matrix limit theory *by studying the in-sample performance as a function of the ridge penalty.* A fascinating implication is that we can, in turn, infer the amount of predictability in the true underlying model and, along with it, the expected performance of a trading strategy based on the true model. This answers economic questions like "if an investor had full information about the prediction environment, including the true model structure and parameter values, how much predictability would she see in financial markets, and how profitable would this information be to trade on?"

Second, high complexity creates a wedge between the true model and the expected out-

[15] For example, when $c \geq 1$, one can achieve perfect in-sample predictions because there are enough parameters to fit each training observation exactly.

of-sample performance of the trained model. The failure of the law of large numbers when $c > 0$ means that, even in the limit, there remains a gap between the trained model and the truth. This limits how much the trained model can learn, even in large samples and with a correctly specified training model. Because the training model needs to hone in on the truth fully, it falls short of the true model's performance. We call this wedge "limits to learning." It describes how far the trained model is from the best possible model regarding expected performance.

$$\text{Limits to Learning} = \text{True Predictability} - \text{Out-of-sample Performance}$$

Like overfit, this wedge is due to the statistical cost of estimating many parameters when training data is relatively scarce. Our theory quantifies learning limits based on the training model's properties and training data as with overfit. The most important implication of this result is that we can characterize the expected out-of-sample performance of the training model. Another interesting fact is that even if there are arbitrage (or simply very high Sharpe ratio) opportunities implied by the true model, limits to learning make these inaccessible to real-world investors. In a realistic empirical setting, we demonstrate that attainable Sharpe ratios are attenuated by order of magnitude (or more) relative to the true data-generating process due to the costs of estimating complex statistical relationships. Finally, we define the overall complexity wedge as the difference in in-sample model performance versus expected out-of-sample performance:

$$\begin{aligned}\text{Complexity Wedge} &= \text{In-sample Performance} - \text{Out-of-sample Performance}\\ &= \text{Overfit} + \text{Limits to Learning}\end{aligned}$$

which is determined by the amount of overfitting plus the limits to learning.

To summarize, when $c = 0$, the law of large numbers eliminates these wedges, and in-sample, out-of-sample, and true predictability are equalized. But when $c > 0$, complexity causes the trained model's in-sample predictability to exceed that of the true model, which in turn exceeds the trained model's expected out-of-sample predictability. The sum of overfit and limits to learning equals the difference between in-sample and out-of-sample performance.

Complexity wedges can be partially mitigated by ridge shrinkage. Positive shrinkage can reduce the extent of overfit and improve the limits to learning, but as long as complexity is greater than zero, limits to learning are always positive.[16] Much of the work in machine learning goes into identifying the extent of shrinkage that maximizes a model's expected out-of-sample performance, typically using data-driven methods like cross-validation. Another contribution of our theoretical approach is that it identifies the optimal degree of shrinkage for a given set of training conditions to maximize expected out-of-sample model performance. Theorem 3 implies the following expression for the limits to learning.

Corollary 4 (Limits To Learning) *Let* $SR(z;c) = \lim E[R^F_{T+1}(z)]/(Var[(R^F_{T+1}(z))^2])^{1/2}$ *be the asymptotic unconditional Sharpe ratio of the feasible efficient portfolio. Then,*

$$\frac{1}{SR^2(z;c)} = (1+G(z;c))\frac{1}{SR^2_{infeas}(Z_*(z;c))} + \frac{G(z;c)}{A(Z_*(z;c))^2} \tag{40}$$

where we have defined

$$SR^2_{infeas}(0) = MaxSR^2 = E[F]'\mathrm{Var}[F]^{-1}E[F] \tag{41}$$

to be the maximal achievable (infeasible) unconditional squared Sharpe ratio.

As explained above, the infeasible Sharpe ratio, $SR_{infeas}(z)$, is monotone, decreasing in

[16] When the shrinkage parameter is very large, overfit may become negative, but the out-of-sample performance is always strictly smaller that the performance of the "true" model.

z. Corollary 4 shows how both the implicit regularization and the estimation risk create a wedge between feasible and infeasible performance.

We now discuss the implications of the complexity wedge for asset pricing. Suppose for simplicity that an economic agent has a flat Gaussian prior $\mu_0$about the distribution of factor returns, $\mu_0(F) \sim N(0, zI)$. In this case, the standard Bayes rule for Gaussian distributions implies that the posterior density $\mu_T(F)$ is also Gaussian, $\mu_T(F) \sim N(\bar{F}_T, zI + B_T)$ and therefore the feasible efficient portfolio maximizes quadratic utility,

$$\hat{\lambda}_{INS}(z) \;=\; \arg\max_{\pi} E[\pi' F_{T+1} - 0.5(\pi' F_{T+1})^2 | \mathcal{F}_T], \tag{42}$$

where $\mathcal{F}_T$ is the agent's information set, described by past factor returns $\{F_\tau, \tau \leq T\}$. Consider now an arbitrary, $\mathcal{F}_T$-measurable portfolio strategy π_T. Then, the law of iterated expectations implies the inequality for the unconditional Sharpe ratio $SR(X) = E[X]/(E[X^2])^{1/2}$:

$$SR(\pi_T' F_{T+1}) \;\leq\; SR(\hat{\lambda}_{INS}(z)' F_{T+1}) \tag{43}$$

for any feasible portfolio strategy π_T. We now discuss the implications of our findings to the classic Hansen and Jagannathan (1991) bounds. The asset pricing equation $E_t[R_{t+1} M_{t,t+1}] = 0$ implies the classic bound

$$\frac{STD(M_{t,t+1})}{E[M_{t,t+1}]} \;=\; \max_{\pi_T} SR(\pi_T' F_{T+1}) \tag{44}$$

where the maximum is over *all (both feasible and infeasible)* portfolio strategies π_T. However, in reality, investors only have access to feasible portfolio strategies. In this case, (44) severely overstates the maximal attainable Sharpe ratio. The following is true.

Proposition 5 (Feasible Hansen-Jagannathan Bounds) *Suppose that the econometri-*

cian has a Gaussian prior $\mu_0 \sim N(0, zI)$ *for the distribution of* F_t*. Then, any feasible strategy* π_T *satisfies*

$$\frac{1}{SR(\pi_T' F_{T+1})} \geq \frac{1}{SR^2(z;c)} = \frac{E[M]}{STD(M)} + ComplexityWedge(z;c) \geq \frac{E[M]}{STD(M)}, \quad (45)$$

where the complexity wedge is strictly positive, and the last inequality is always strict for $c > 0$*.*

It is instructive to relate our findings to the statistical limit of arbitrage introduced in Rui Da and Xiu (2022). Contrary to our paper, Rui Da and Xiu (2022) assume that the true pricing kernel, M, is observable and investigate the possibility of discovering arbitrage opportunities by detecting mispriced stocks. By contrast, our focus here is on the pricing kernel's complexity and an economic agent's inability to learn a highly complex SDF in finite samples. Developing econometric techniques for estimating the complexity wedge and its link to the true nature of the data-generating process is an important direction for future research.

4 Mis-Specified Models and the Virtue of Complexity

So far, we have implicitly assumed that formula (9) is a correctly specified model for the SDF. Equivalently, stock returns have an exact, P-dimensional factor structure given by factors F_t with unknown risk premia that we attempt to learn. We now explore a more realistic setting where only a fraction $q = \frac{P_1}{P}$ of factors is observable, with some $P_1 < P$. The subset of factors, $F_{t+1}(q) = (F_{i,t+1})_{i=1}^{P_1}$, has a covariance matrix $\Psi(q) \in \mathbb{R}^{P_1 \times P_1}$. We then define

$$\hat{\lambda}_{INS}(z;q) = (zI + B_T(q))^{-1} \bar{F}_T(q), \quad (46)$$

where

$$\bar{F}_T(q) \;=\; \frac{1}{T}\sum_{t=1}^{T} F_t(q) \;\in\; \mathbb{R}^{P_1}\,, \tag{47}$$

and

$$B_T(q) \;=\; \frac{1}{T}\sum_{t=1}^{T} F_t(q)F_t(q)' \;\in\; \mathbb{R}^{P_1\times P_1}\,. \tag{48}$$

In this case, all of the above expressions in Theorems 1 and 3 hold true, with the key functions $A(z;q)$ and $\xi(z;c;q)$ depending explicitly on q through $E[F(q)]$ and the eigenvalue distribution of $\Psi(q)$. It is straightforward to show that $A(z;q)$ is always monotone increasing in q: For the infeasible portfolio, having more factors is always beneficial and always improves the infeasible maximal Sharpe ratio. In the high-complexity case, this improvement only holds when the marginal benefit of an additional factor is large enough to compensate for the higher estimation error. The following result is an analog of the *virtue of complexity* (VoC) principle of KMZ for factor portfolios.

Theorem 6 (The Virtue of Complexity) *Suppose that* $\frac{\partial}{\partial q}SR_{infeas}(z;q)$ *is sufficiently large relative to* $\frac{\partial}{\partial z}SR_{infeas}(z;q)$*. Then, the OOS Sharpe ratio is monotone increasing in* q*.*

Recall that P_1 represents both the number of factors and the number of parameters in our model. According to Theorem 6, as $q = P_1/P$ and model complexity P_1/T increase, so does the model performance. This highly counterintuitive result suggests that the zoo of factors should be expanded rather than controlled. Equivalently, rather than restricting the weights $w^*(X)$ in (3), we should *expand their parameterization, saturating it with conditioning information.* Our extensive numerical simulations and empirical results suggest that the conditions of Theorem 6 are almost always satisfied in real data.

5 Pricing Errors

We now analyze asset pricing errors using the (Hansen and Jagannathan, 1997) (HJ) distance to complete our discussion of stochastic discount factors' performance. In the high complexity regime, exact details of computing the distance are important, including how we define the HJ distance weighting matrix. Consistent with the fact that we test the efficiency of a portfolio based on its out-of-sample performance, pricing errors also need to be computed out-of-sample (OOS) using the out-of-sample factor moments. The distinction between in- and out-of-sample performance is an essential ingredient in the analysis of any high-complexity model. See, e.g., (Martin and Nagel, 2021) for a related discussion.

Let $E_{OOS}[\cdot]$ denote out-of-sample averages:

$$E_{OOS}[X] \;=\; \frac{1}{T_{OOS}} \sum_{t\in(T,T+T_{OOS}]} X_t\,. \tag{49}$$

We will also need

$$\bar{F}_{OOS} \;=\; E_{OOS}[F] \;\in\; \mathbb{R}^P,\; B_{OOS} \;=\; E_{OOS}[FF'] \;\in\; \mathbb{R}^{P\times P}\,. \tag{50}$$

Then, as in the previous section, we consider an expanding set of factor models indexed by $q\in(0,1)$, and define

$$M_t(z) \;=\; 1 \;-\; R_t^F(z;q),\;\; with\;\; R_t^F(z;q) \;=\; \hat{\lambda}_{INS}(z;q)'F_t(q) \tag{51}$$

and $\hat{\lambda}_{INS}(z;q)\in\mathbb{R}^{P_1}$ from (46), with $P_1<P$. We evaluate the ability of this P_1-factor SDF to price all P factors by computing the OOS pricing error vector:

$$\mathcal{E}_{OOS}(z;q) \;=\; \frac{1}{T_{OOS}} \sum_{t\in(T,T+T_{OOS}]} F_t M_t(z;q) \;\in\; \mathbb{R}^P\,. \tag{52}$$

The HJ distance is then given by

$$\mathcal{D}^{HJ}_{OOS}(z;q) \;=\; \mathcal{E}_{OOS}(z;q)'\, B^{+}_{OOS}\, \mathcal{E}_{OOS}(z;q)\,, \tag{53}$$

where B^{+}_{OOS} is the Moore-Penrose quasi-inverse of the highly degenerate (of rank$\leq T_{OOS}$) matrix B_{OOS}. The following result follows by direct calculation.

Proposition 7 *We have*

$$\mathcal{D}^{HJ}_{OOS}(z;cq;q) \;-\; \bar{F}'_{OOS}B^{-1}_{OOS}\bar{F}_{OOS} \;=\; -2E_{OOS}[R^F(z;q)] \;+\; E_{OOS}[(R^F(z;q))^2] \tag{54}$$

When $P > T_{OOS}$ and both are sufficiently large, we have

$$\bar{F}'_{OOS}B^{-1}_{OOS}\bar{F}_{OOS} \;\approx\; 1 \tag{55}$$

and hence

$$\mathcal{D}^{HJ}_{OOS}(z;q) \;\approx\; E_{OOS}[(1 - M_t(z;q))^2]\,. \tag{56}$$

In particular, pricing errors are independent of the set of test factors.

Proposition 7 shows how the HJ distance is directly linked to the OOS performance of the efficient portfolio. In particular, Theorems 1 and 3 allow us to derive explicit asymptotic expressions for this distance and obtain an analog of the VoC result from Theorem 6.

Theorem 8 (The Virtue of Complexity for Pricing Errors) *The asymptotic expected pricing error is given by*

$$\mathcal{D}^{HJ}_{OOS}(z;cq;q) \;\approx\; (1 + G(z;cq))\mathcal{D}^{HJ}_{OOS}(Z^*(z;cq);0;q)\,. \tag{57}$$

Suppose that $\frac{\partial}{\partial q}\mathcal{D}^{HJ}_{OOS}(z;0;q)$ is sufficiently large relative to $\frac{\partial}{\partial z}\mathcal{D}^{HJ}_{OOS}(z;0;q)$. Then, $\mathcal{D}^{HJ}_{OOS}(z;cq;q)$ is monotone increasing in q.

Theorem 8 shows a surprising identity for the expected pricing errors: The high complexity error is proportional to the infeasible error, subject to the implicit regularization (i.e., z is replaced by $Z^*(z;cq)$), and the proportionality factor equals one plus the estimation risk. Similarly to Theorem 6, this defines a tradeoff between the ability of the model to approximate the truth and the estimation risk due to complexity. When the gain from higher q in $\mathcal{D}^{HJ}_{OOS}(z;0;q)$ is large enough, the former effect dominates, and we obtain the VoC.

6 Empirics

Our monthly frequency dataset comes from (Jensen et al., Forthcoming) and contains 153 characteristics and realized returns for US publicly traded stocks from 1963-01-31 to 2019-12-31.

Many of the 153 characteristics from (Jensen et al., Forthcoming) have significant fractions of missing values, especially in the early parts of the sample. For this reason, we first pre-select 130 characteristics with the smallest percentage of missing values. This ensures that the characteristics composition is more homogeneous over time. Among those 130 characteristics, we select and exclude 20 with the highest turnover.[17] We do this on purpose to address the existing critiques based on the apparent tendency of machine learning models to generate extremely fast-varying and hence hard-to-trade signals (see, e.g., Chinco et al. (2019), and Jensen et al. (2022) for a potential remedy based on machine learning methods that properly account for trading costs). This leaves us with $d = 110$ characteristics. Then, for each date, we only keep stocks for which less than 30% characteristic values are missing, ensuring that, on each date, each stock has at least 77 characteristics as an input to our

[17] See Appendix **??** the definition of turnover and the list of these excluded characteristics.

machine learning models. In the sequel, we use N_t to denote the number of such "eligible" stocks available at time t.

Every date, for each characteristic $k = 1, \cdots, d$, there are $n_t(k)$ stocks with non-missing values cross-sectionally rank these characteristics (not including the missing values), replacing them with their rank between 0 and $n_t(k)$. We then divide this rank by $n_t(k)$ and subtract 0.5, to get a normalized rank in $[-0.5, 0.5]$. We then fill in missing values of the k-th characteristic of the remaining $N_t - n_t(k)$ stocks with zeros. This way, we obtain a panel of characteristics $X_t = (X_{i,k,t})_{i,k} \in \mathbb{R}^{N_t \times d},\ d = 110$, taking values in $[-0.5, 0.5]$ and we assume a pricing kernel of the form (3).

We now define the $P >> d$ non-linear features (6) that serve as an input to our pricing kernel expansion (5). Our goal is to capture features *with varying degrees of non-linearity.* This is important: Linear features (i.e., those given by linear combinations of X_t) contain information about future expected returns, as is shown by (Jensen et al., Forthcoming). Following KMZ, we control the degree of non-linearity by introducing a grid of G scale parameters, $\gamma_g,\ i = 1, \cdots, G$. In our analysis, we use the $[0.5, 0.6, 0.7, 0.8, 0.9, 1.0]$ grid. For each scale parameter γ_g, we draw a random weight matrix

$$W_g \sim \mathbb{N}(0, \gamma_g) \in \mathbb{R}^{d, \frac{P}{2G}}, \tag{58}$$

Next, we define,

$$\hat{S}_t(\gamma_g) = \mathit{concatenate}\left(\cos(X_t W_g), \sin(X_t W_g)\right) \in \mathbb{R}^{N_t \times \frac{P}{G}}. \tag{59}$$

We then concatenate all these feature groups to produce

$$\hat{S}_t = \mathit{concatenate}\left(\hat{S}_t(\gamma_1), \cdots, \hat{S}_t(\gamma_G)\right) \in \mathbb{R}^{N_t \times P}. \tag{60}$$

We *randomly permute* the order of random features so that features with different activation functions (*cos* or *sin*) and different degrees γ of non-linearity appear uniformly spread across the feature universe. This is important for our analysis of the *virtue of complexity* where we expand the set of random features from $P_1 = 1$ to $P_1 = P$.

We now perform the same ranking procedure as above in the random features $\hat{S}_t$. Now, there are always precisely N_t values for each random feature k, and we rank them, normalize them by N_t, and then subtract 0.5 to obtain our final random features

$$S_t \;=\; N_t^{-1} rank(\hat{S}_t) \;-\; 0.5 \;\in \mathbb{R}^{N_t \times P}\,. \tag{61}$$

Finally, we define the random factors,

$$F_{t+1} = \frac{1}{N_t^{1/2}} R'_{t+1} S_t \in R^P, \tag{62}$$

where $R_{t+1} \in R^{N_t}$ is a vector stock returns. The normalization by $N_t^{1/2}$ ensures that the random vector F has a well-defined, bounded covariance matrix (see the Appendix for details).

We define a sequence of P_1, gradually increasing from T to P and define $q = P_1/P \in [0,1]$ and $F_{t+1}(q)$ to be the first P_1 factors out of P. We pick a rolling window of $T = 360$ months and define our rolling estimators for the empirical mean and covariance matrix of the managed portfolios:

$$B_t(q) \;=\; \frac{1}{T}\sum_{\tau=t-T}^{t} F_\tau(q)F_\tau(q)' \;=\; \frac{1}{T}F_{[t-T:t]}(q)'F_{[t-T:t]}(q) \;\in\; P_1 \times P_1,\; F(q) \in \mathbb{R}^{T\times P_1} \tag{63}$$

and

$$\bar{F}_t(q) \;=\; N_t^{-1/2} \sum_{\tau=t-T}^{t} F_\tau(q) N_\tau^{1/2}, \tag{64}$$

and then

$$\hat{\lambda}_t(z;q) \;=\; (zI + B_t(q))^{-1}\bar{F}_t(q) \tag{65}$$

and the corresponding efficient portfolio returns,

$$R^F_{t+1}(q) \;=\; \hat{\lambda}_t(z;q)' F_{t+1}(q)\,. \tag{66}$$

While the matrix $B_t(q)$ is easy to define, computing $(zI + B_t(q))^{-1}\bar{F}_t(q)$ is far from trivial due to the gigantic dimension of the matrix $B_t(q)$. We use the following lemma to circumvent this problem.

Lemma 1 *Let $F \in \mathbb{R}^{T\times P_1}$ and consider the matrix*

$$\tilde{B} \;=\; \frac{1}{T} F F' \;\in\; \mathbb{R}^{T\times T}\,. \tag{67}$$

Let $\tilde{B} = UDU'$ be the eigenvalue decomposition of $\tilde{B}$. Let also $B = \frac{1}{T}F'F \in \mathbb{R}^{P\times P}$. Define

$$\tilde{U} \;=\; F'UD^{-1/2} \;\in\; \mathbb{R}^{P\times T}\,. \tag{68}$$

Then,

$$(zI + B)^{-1}Y \;=\; \tilde{U}(D + zI)^{-1}(\tilde{U}'Y) \;+\; z^{-1}(I - \tilde{U}\tilde{U}')Y \tag{69}$$

for any vector $Y \in \mathbb{R}^P$. This can be computed efficiently by first computing $\hat{Y} = \tilde{U}'Y \in \mathbb{R}^T$ using only $P \times T$ operations; then computing $\tilde{Y} = (\mathrm{diag}(D + z))^{-1} \circ \hat{Y}$ using only T

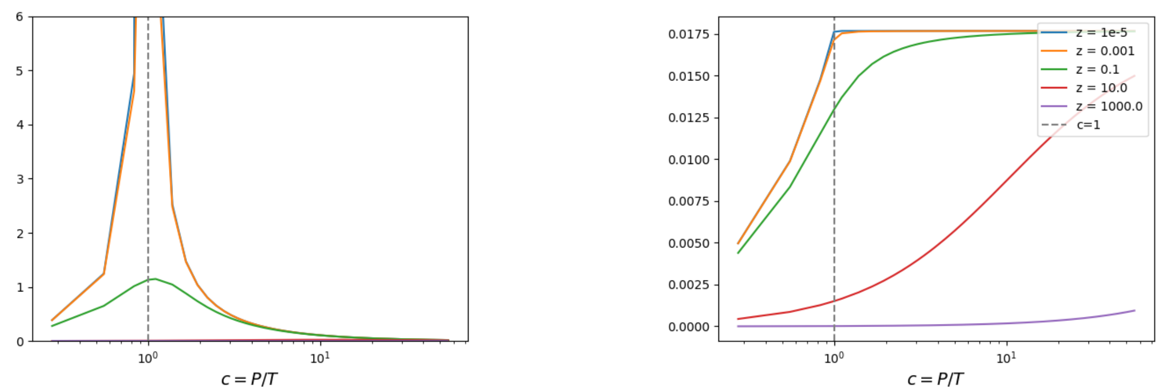
operations;[18] *and then*

$$(zI+B)^{-1}Y \;=\; \underbrace{\tilde{U}(\tilde{Y}-z^{-1}\hat{Y})}_{P\times T\ operations} + z^{-1}Y \tag{70}$$

[18] $A \circ B$ is the elementwise (Hadamard) product of two matrices A, B.

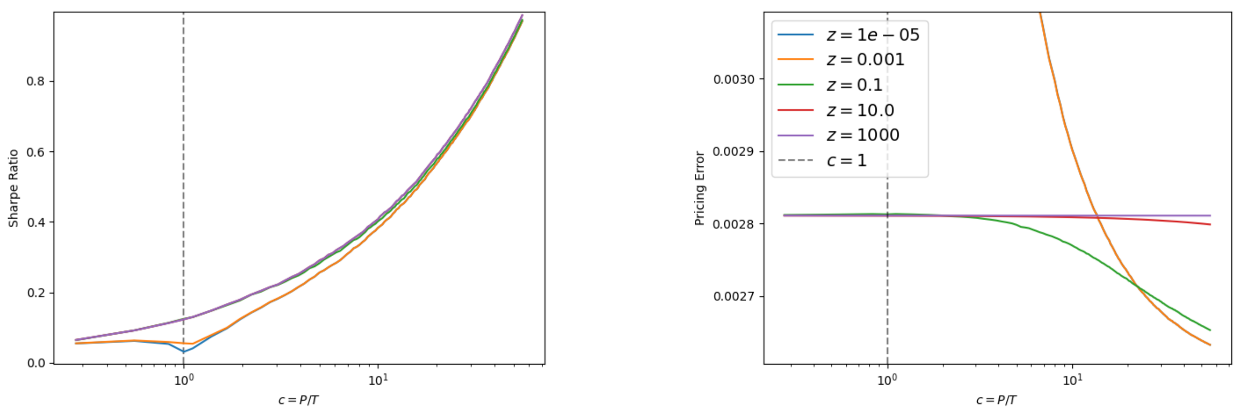
6.1 Simulation

To provide evidence in support of our theory, we start with simulations, generating factor returns satisfying the data-generating process assumptions. Figure 2, Panel (a), shows the behavior of out-of-sample realized standard deviation of the efficient portfolio. Beyond the interpolation boundary (when $P > T$ and $c > 1$), we observe the key phenomenon responsible for the virtue of complexity principle discovered in KMZ: With the growing complexity, implicit regularization of high-dimensional models leads to a reduction in out-of-sample risk. At the same time, as the model becomes a closer approximation to the true complex model, Figure 2, Panel (b) shows how the out-of-sample expected return is monotonically increasing in c. Not surprisingly, these patterns are mapped into a monotonic improvement in the Sharpe Ratio and Pricing Errors past the interpolation boundary, as can be seen from Figure 3.



(a) Volatility **(b)** Expected Return

Figure 2: Simulation results: Volatility and Expected Returns with $\Psi = I$ and $\lambda \sim N(0, I)$



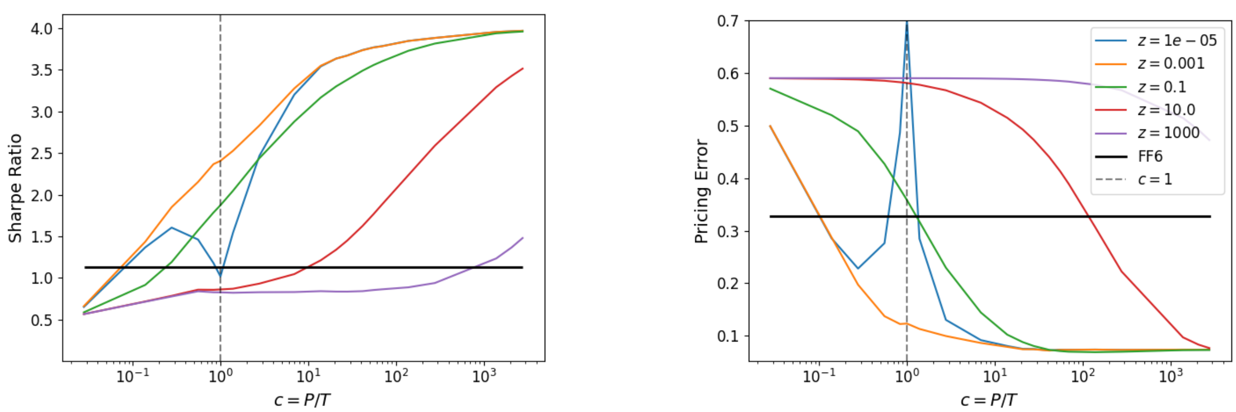
(a) Sharpe Ratio

(b) Pricing Error

Figure 3: Simulation results: SDF Performance with $\Psi = I$ and $\lambda \sim N(0, I)$

6.2 Full Sample

We now repeat the experiment from the previous section with real data, using signals and factors constructed in (61)-(62) and the efficient portfolios from (66). In this experiment, we use *all* stocks in our sample. For all degrees of complexity below the maximum ($q < 1$), we conducted the experiment 20 times by randomly selecting P_1 features out of a maximum of $1e6$. The results presented below and in the rest of this section represent the average performance across those 20 experiments. Figure 4 reports the realized OOS Sharpe ratios and pricing errors. The remarkable monotonic patterns observed with real data offer compelling empirical evidence for the complexity principle: increasing the number of factors significantly enhances the out-of-sample performance of factor models. The extremely high Sharpe ratio (above 4) achieved by the highest-complexity models reflects significant frictions (such as illiquidity and short-sale constraints) associated with trading small and micro-stocks. To understand the role of these frictions, we analyze the virtue of complexity separately for each size group in the next section.

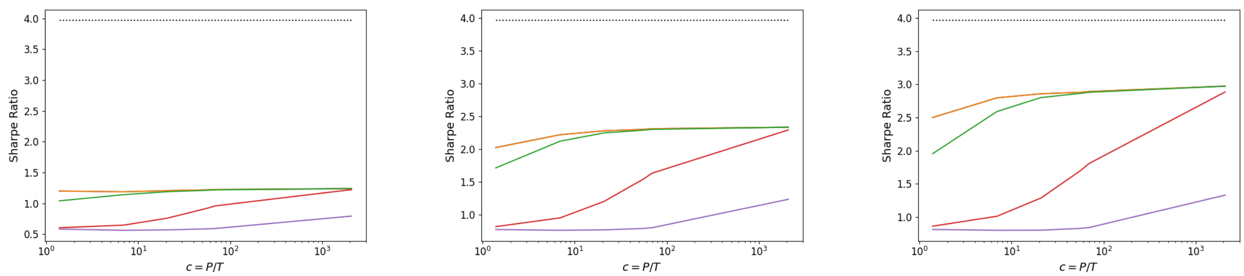


(a) Sharpe Ratios

(b) Pricing Errors

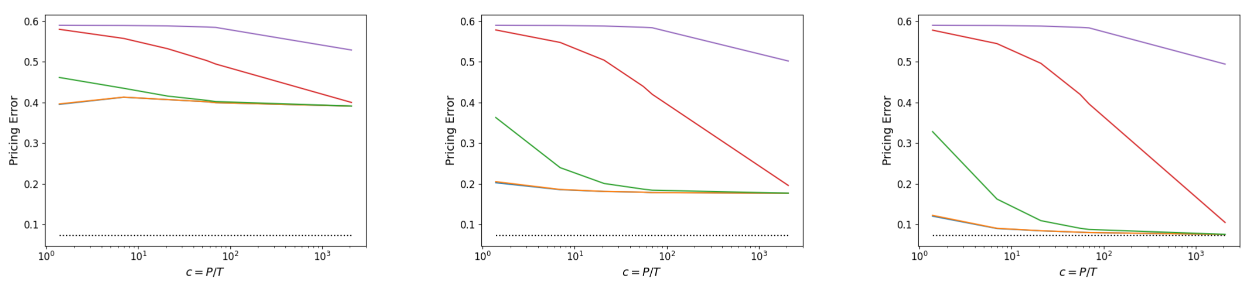
Figure 4: Sharpe ratios (a) and pricing error (b) of (66) computed on our full sample.

One may ask whether other forms or sparse representations for the pricing kernel exist. One natural form of sparsity could be potentially achieved by using only a few most important principal components of factors, as in (Kelly et al., 2020), (Kozak et al., 2018), (Kozak et al., 2020), (Lettau and Pelger, 2020), and (Giglio and Xiu, 2021), who argue that retaining only a few top principal components is sufficient to explain the cross-section of returns. See also (Gagliardini et al., 2016). We investigate this approach empirically and report the corresponding results in Figures 5 and 6. As one can see, even with the top 25 principal components, the performance gap relative to the full high-complexity model is very large, with the Sharpe ratio dropping from 4 to 3.



(a) $K = 5$ **(b)** $K = 10$ **(c)** $K = 25$

Figure 5: Shrinking with PCA: Sharpe Ratios of top-PCs-based version of (66). K indicates the number of top PCs used. PCs are computed in the same rolling window as (66)



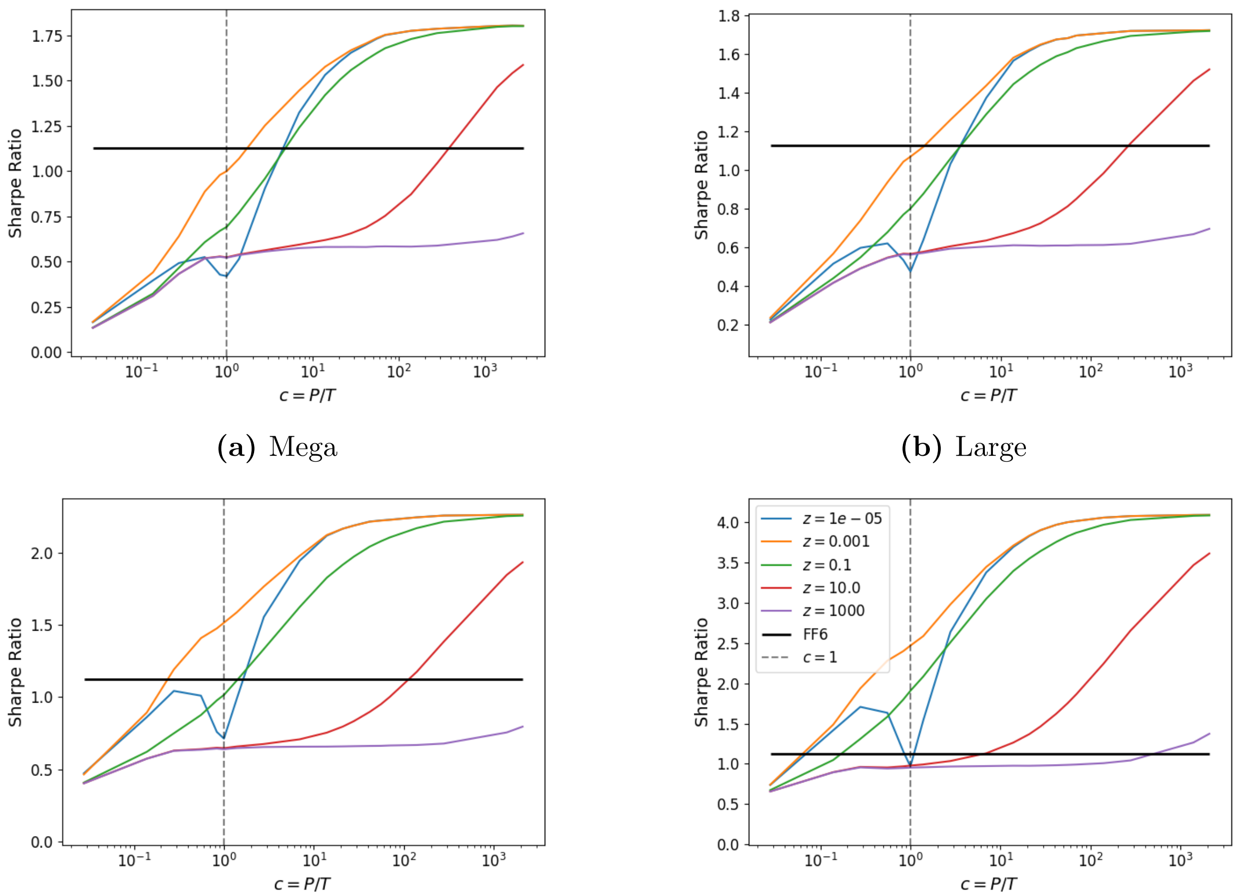
(a) $K = 5$ **(b)** $K = 10$ **(c)** $K = 25$

Figure 6: Shrinking with PCA: Pricing Error of top-PCs-based version of (66). K indicates the number of top PCs used. PCs are computed in the same rolling window as (66)

6.3 Empirical Analysis by Size Group

In the previous section, we performed our analysis on the full cross-section of stocks. We now perform our experiments separately for four groups of stocks, selecting according to their market capitalization (size): mega (largest 20% of stocks based on NYSE breakpoints at each time period), large (between 80% and 50% percentile of NYSE breakpoints), small (between 50% and 20% percentile of NYSE breakpoints), and micro (between 20% and 1% percentile).

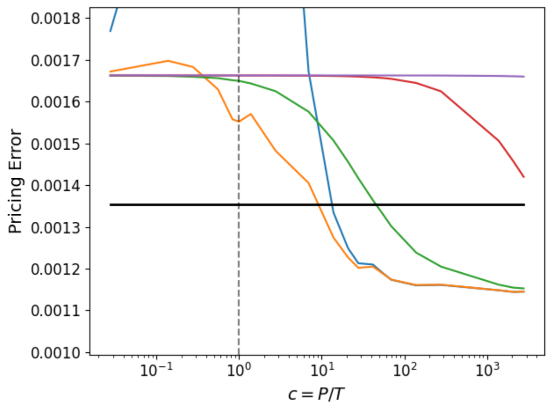
We construct our random feature and optimal portfolio for various P_1 using the methods outlined in the previous section and calculate the Sharpe Ratio per complexity level and size group (see Figure 7), as well as the pricing error (see Figure 8). Both figures demonstrate that the VoC holds for all subsamples in terms of both pricing error and Sharpe Ratios. Not surprisingly, The Sharpe ratio achieves its highest values for the micro group of stocks. The more realistic Sharpe ratio of 1.75 achieved by the highest-complexity model trading only mega stocks (about 300-500 largest stocks in the US economy) is broadly consistent with the net Sharpe ratio of roughly 1.4 reported in Jensen et al. (2022) after accounting for transaction costs.



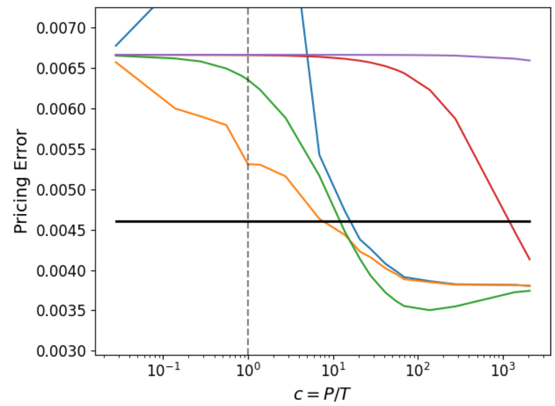
(a) Mega **(b)** Large

(c) Small **(d)** Micro

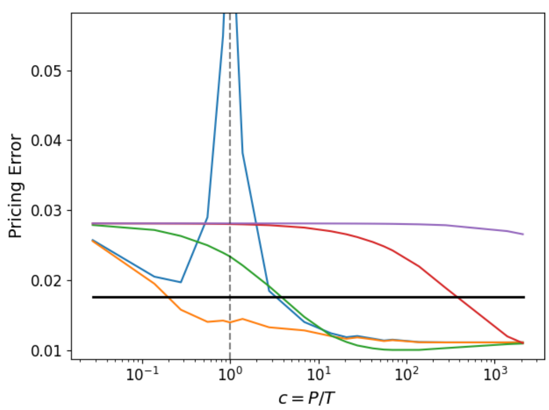
Figure 7: Sharpe ratios of (66) for different Market Capitalization Subsamples



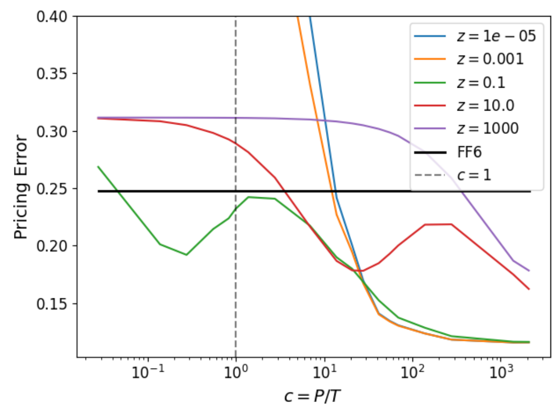
(a) Mega



(b) Large



(c) Small



(d) Micro

Figure 8: Pricing Errors of (66) for different Market Capitalization Subsamples

6.4 Complexity vs Simpler Models

So far, we have only discussed the absolute performance of the high-complexity efficient portfolios. But how does it relate to standard benchmarks such as Fama-French factors and simpler efficient portfolios based on "linear" characteristics from (Jensen et al., Forthcoming) that serve as an input to our high-complexity features? To answer this question, we first define

$$F_t^{linear} \;=\; X_t' R_{t+1}\,, \tag{71}$$

where $X_t \in \mathbb{R}^{N_t \times 110}$ is the matrix of rank-standardized characteristics taking values in $[-0.5, 0.5]$ (see Section 6 for details). X_t are built so that F_t^{linear} is expected to have positive mean returns. Next, we build two simple benchmarks: The $1/N$ portfolio (see DeMiguel et al. (2009)), defined as the equal-weighted portfolio of F_t^{linear}

$$R_t^{EW} \;=\; \frac{1}{N_{t-1}} \sum_{k=1}^{110} F_{t,k}^{linear}\,, \tag{72}$$

and the efficient portfolio of linear factors built using the same methodology as (66):

$$R_{t+1}^{linear}(z) \;=\; \hat{\lambda}_t^{linear}(z;q)' F_{t+1}^{linear}(q) \tag{73}$$

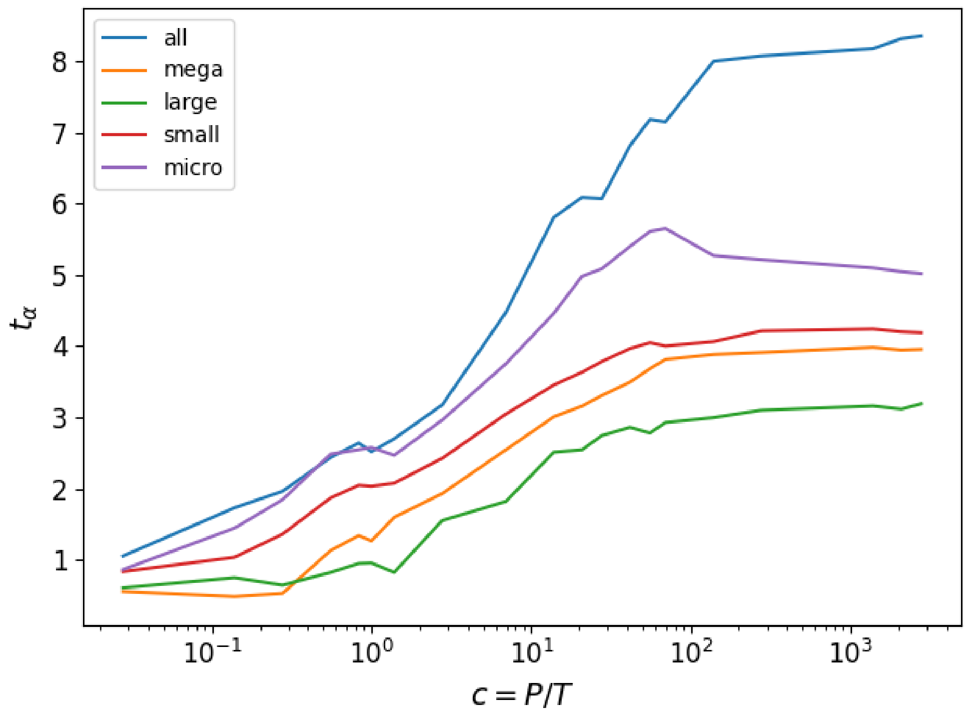
where

$$\hat{\lambda}_t^{linear}(z;q) \;=\; (zI + B_t^{linear}(q))^{-1} \bar{F}_t^{linear}(q)\,,$$

is constructed using the respective moments of F_t^{linear}. Additionally, we use six standard Fama-French and momentum factors as benchmarks: CMA_t, HML_t, MKT_t, MOM_t, RMW_t, and SMB_t. For both the benchmarks, $R_t^{linear}(z)$ and the dependent variable $R_t^F(z;q)$, we select the penalty parameters that maximize the Sharpe Ratio over the entire sample. We denote these optimal penalty parameters by z_*^{linear} and $z_*^{complex}$, respectively. We then run the multi-variate regression

$$
\begin{aligned}
R_t^F(z_*^{complex}; q) \ &= \ \alpha \ + \ \beta^{EW} R_t^{EW} + \beta^{linear} R_t^{linear}(z_*^{linear}) \\
&+ MKT_t + SMB_t + HML_t + CMA_t + RMW_t + MOM_t + \varepsilon_t
\end{aligned}
\tag{74}
$$

The results of these regressions are summarized in Figure 9a and Table 1. The figure displays the heteroskedasticity-adjusted t-statistics of the regressions for various degrees of complexity of our complex pricing kernel for the full sample (all) and the sample defined by market capitalization (see section 6.3). The table presents the regression's α and β, along with their standard deviations for the maximum complexity across all stock groups. It is evident that high-complexity efficient portfolios significantly outperform the extremely demanding benchmark in (74), with both economic and statistical significance. We interpret these findings as evidence that the true pricing kernel is highly non-linear in characteristics, and our high-complexity model can capture some of these non-linearities.



(a) Heteroskedasticity-adjusted (with five lags) t-statistics of α from the regression (74) for the full stock (all) universe as well as for each size subgroup.

	All	Mega	Large	Small	Micro
α	0.1805***	0.077***	0.0649***	0.104***	0.1588***
	(0.0218)	(0.0191)	(0.0204)	(0.0248)	(0.0316)
R_t^{linear}	0.6362***	0.7844***	0.7184***	0.5021***	0.6413***
	(0.046)	(0.0476)	(0.0411)	(0.0307)	(0.0626)
R_t^{EW}	-0.0012	-0.0023	-0.0002	0.006	-0.0087
	(0.0053)	(0.0071)	(0.0132)	(0.0069)	(0.0131)
CMA_t	0.9779	-0.2212	-1.5575	0.9845	1.1355
	(0.8746)	(1.1865)	(1.3362)	(1.6593)	(1.0378)
HMK_t	0.0957	1.8198**	2.362**	1.1499	-0.2733
	(0.7252)	(0.7632)	(0.9345)	(1.0122)	(0.6836)
MKT_t	1.7219***	0.4767	1.7409***	1.7501***	0.4887
	(0.3484)	(0.5621)	(0.6091)	(0.6476)	(0.3876)
MOM_t	1.2361***	-0.1158	0.5778	3.2479***	0.4871
	(0.425)	(0.5837)	(0.648)	(0.9557)	(0.44)
RMW_t	4.0977***	2.9615***	5.901***	5.2083***	2.9756***
	(0.816)	(1.0195)	(1.3512)	(1.3278)	(1.0467)
SMB_t	0.4532	-0.496	-0.1987	-1.7623*	-0.6249
	(0.5297)	(0.7108)	(0.891)	(0.9314)	(0.5023)
Observations	322	322	322	322	322
R^2	0.7924	0.644	0.7453	0.6655	0.7838

Table 1: Heteroskedasticity-adjusted (with five lags) t-statistics of α and β coefficients for the regression (74) for the full stock (all) universe as well as for each size subgroup, with $q = 1$—that is $c = \frac{P}{T} = \frac{1e6}{360}$. Note: *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Standard errors are reported in parentheses.

7 WIP Turnover

We define the turnover of characteristic k as the average absolute change of the variable across firm and time:

$$Turnover_k = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=2}^{T}|X_{i,k,t} - X_{i,k,t-1}|\right) \tag{75}$$

Table 2 shows the turnover value for each of the 130 characteristics.

	Variable	Turnover	Variable	Turnover	Variable	Turnover	Variable	Turnover
0	ret_1_0	0.351	nfna_gr1a	0.055	aliq_mat	0.034	eqpo_me	0.022
1	seas_1_1an	0.337	dsale_drec	0.054	ebitda_mev	0.034	capx_gr3	0.022
2	rskew_21d	0.322	niq_be	0.052	netis_at	0.034	lti_gr1a	0.022
3	coskew_21d	0.322	saleq_gr1	0.052	sale_gr1	0.034	eqnetis_at	0.022
4	iskew_ff3_21d	0.321	dbnetis_at	0.051	chcsho_12m	0.034	qmj_safety	0.022
5	iskew_capm_21d	0.320	betadown_252d	0.051	dgp_dsale	0.033	div12m_me	0.021
6	iskew_hxz4_21d	0.308	mispricing_perf	0.051	eqnpo_12m	0.032	sale_me	0.021
7	rmax5_rvol_21d	0.301	taccruals_at	0.048	sale_emp_gr1	0.032	at_me	0.021
8	beta_dimson_21d	0.283	cowc_gr1a	0.048	ocf_at	0.030	beta_60m	0.02
9	ret_3_1	0.241	taccruals_ni	0.047	lnoa_gr1a	0.030	netdebt_me	0.02
10	rmax1_21d	0.230	fnl_gr1a	0.046	fcf_me	0.030	o_score	0.02
11	rmax5_21d	0.211	f_score	0.046	sti_gr1a	0.028	sale_gr3	0.019
12	bidaskhl_21d	0.201	col_gr1a	0.045	be_me	0.028	intrinsic_value	0.019
13	ivol_ff3_21d	0.196	niq_at	0.045	ppeinv_gr1a	0.028	ope_bel1	0.019
14	ivol_capm_21d	0.192	market_equity	0.043	aliq_at	0.028	ebit_bev	0.019
15	ivol_hxz4_21d	0.192	tax_gr1a	0.041	zero_trades_126d	0.028	op_at	0.019
16	rvol_21d	0.180	coa_gr1a	0.040	capx_gr2	0.027	op_atl1	0.019
17	resff3_6_1	0.151	prc	0.040	turnover_126d	0.027	ope_be	0.018
18	prc_highprc_252d	0.151	at_gr1	0.040	noa_at	0.027	debt_me	0.018
19	ret_6_1	0.150	ocf_me	0.039	ivol_capm_252d	0.026	at_be	0.018
20	ret_12_7	0.145	noa_gr1a	0.039	dsale_dsga	0.026	zero_trades_252d	0.018
21	resff3_12_1	0.120	inv_gr1	0.039	pi_nix	0.026	sale_bev	0.015
22	ret_9_1	0.117	nncoa_gr1a	0.039	debt_gr3	0.026	tangibility	0.015
23	zero_trades_21d	0.102	mispricing_mgmt	0.039	cop_atl1	0.026	z_score	0.015
24	ret_12_1	0.097	ncoa_gr1a	0.039	cop_at	0.026	gp_at	0.015
25	seas_1_1na	0.083	capx_gr1	0.038	eqnpo_me	0.025	gp_atl1	0.014
26	niq_su	0.081	ncol_gr1a	0.038	bev_mev	0.025	ebit_sale	0.014
27	niq_at_chg1	0.075	oaccruals_at	0.037	cash_at	0.025	kz_index	0.013
28	niq_be_chg1	0.073	inv_gr1a	0.036	qmj_prof	0.024	at_turnover	0.013
29	turnover_var_126d	0.069	oaccruals_ni	0.036	dolvol_126d	0.024	opex_at	0.013
30	dolvol_var_126d	0.065	ni_me	0.036	ami_126d	0.024	age	0.006
31	saleq_su	0.062	eq_dur	0.035	emp_gr1	0.024		
32	ocf_at_chg1	0.058	be_gr1a	0.035	ni_be	0.023		

Table 2: The table above show the turnover defined in Equation (75) for each of our 130 characteristics.

References

Avramov, Doron, Si Cheng, and Lior Metzker, "Machine learning versus economic restrictions: Evidence from stock return predictability," *Available at SSRN 3450322*, 2021.

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

Bartlett, Maurice S, "An inverse matrix adjustment arising in discriminant analysis," *The Annals of Mathematical Statistics*, 1951, *22* (1), 107–111.

Belkin, Mikhail, "Fit without fear: remarkable mathematical phenomena of deep learning through the prism of interpolation," *Acta Numerica*, 2021, *30*, 203–248.

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.

Britten-Jones, Mark, "The sampling error in estimates of mean-variance efficient portfolio weights," *The Journal of Finance*, 1999, *54* (2), 655–671.

Bryzgalova, Svetlana, Jiantao Huang, and Christian Julliard, "Bayesian solutions for the factor zoo: We just ran two quadrillion models," *The Journal of Finance*, 2023, *78* (1), 487–557.

—, **Markus Pelger, and Jason Zhu**, "Forest through the trees: Building cross-sections of stock returns," *Available at SSRN 3493458*, 2020.

Chen, Luyang, Markus Pelger, and Jason Zhu, "Deep learning in asset pricing," *arXiv preprint arXiv:1904.00745*, 2019.

Chen, Qihui, Nikolai Roussanov, and Xiaoliang Wang, "Semiparametric Conditional Factor Models: Estimation and Inference," *arXiv preprint arXiv:2112.07121*, 2021.

Chinco, Alex, Adam D Clark-Joseph, and Mao Ye, "Sparse signals in the cross-section of returns," *The Journal of Finance*, 2019, *74* (1), 449–492.

Cochrane, John H, "Presidential address: Discount rates," *The Journal of finance*, 2011, *66* (4), 1047–1108.

Cong, Lin William, Ke Tang, Jingyuan Wang, and Yang Zhang, "AlphaPortfolio: Direct construction through deep reinforcement learning and interpretable AI," *SSRN Electronic Journal. https://doi. org/10.2139/ssrn*, 2021, *3554486.*

Da, Rui, Stefan Nagel, and Dacheng Xiu, "The Statistical Limit of Arbitrage," Technical Report, Technical Report, Chicago Booth 2022.

Da, Stefan Nagel Rui and Dacheng Xiu, "The Statistical Limit of Arbitrage," *Working paper*, 2022.

DeMiguel, Victor, Alberto Martin-Utrera, Francisco J Nogales, and Raman Uppal, "A transaction-cost perspective on the multitude of firm characteristics," *The Review of Financial Studies*, 2020, *33* (5), 2180–2222.

___, Lorenzo Garlappi, and Raman Uppal, "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?," *The review of Financial studies*, 2009, *22* (5), 1915–1953.

Didisheim, Antoine, Shikun Ke, Bryan Kelly, and Semyon Malamud, "Complexity for the Cross-Section," *Swiss Finance Institute Research Paper*, 2022, (22-57).

Fan, Jianqing, Zheng Tracy Ke, Yuan Liao, and Andreas Neuhierl, "Structural Deep Learning in Conditional Asset Pricing," *Available at SSRN 4117882*, 2022.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu, "Taming the factor zoo: A test of new factors," *The Journal of Finance*, 2020, *75* (3), 1327–1370.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber, "Dissecting characteristics nonparametrically," *The Review of Financial Studies*, 2020, *33* (5), 2326–2377.

Gagliardini, Patrick and Hao Ma, "Extracting statistical factors when betas are time-varying," *Swiss Finance Institute Research Paper*, 2019, (19-65).

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

——, ——, and ——, "Estimation of large dimensional conditional factor models in finance," *Handbook of econometrics*, 2020, *7*, 219–282.

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

——, **Bryan Kelly, and Dacheng Xiu**, "Factor models, machine learning, and asset pricing," *Annual Review of Financial Economics*, 2022, *14*.

——, **Dacheng Xiu, and Dake Zhang**, "Test assets and weak factors," Technical Report, National Bureau of Economic Research 2021.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, "Autoencoder Asset Pricing Models," *Journal of Econometrics*, 2020.

——, ——, **and** ——, "Empirical asset pricing via machine learning," *The Review of Financial Studies*, 2020, *33* (5), 2223–2273.

Guijarro-Ordonez, Jorge, Markus Pelger, and Greg Zanotti, "Deep learning statistical arbitrage," *arXiv preprint arXiv:2106.04028*, 2021.

Han, Yufeng, Ai He, David Rapach, and Guofu Zhou, "Expected stock returns and firm characteristics: E-LASSO, assessment, and implications," *SSRN*, 2019.

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

—— **and** ——, "Assessing specification errors in stochastic discount factor models," *The Journal of Finance*, 1997, *52* (2), 557–590.

—— **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.

Harvey, Campbell R, Yan Liu, and Heqing Zhu, "... and the cross-section of expected returns," *The Review of Financial Studies*, 2016, *29* (1), 5–68.

Hou, Kewei, Chen Xue, and Lu Zhang, "Replicating anomalies," *The Review of Financial Studies*, 2020, *33* (5), 2019–2133.

Jensen, Theis Ingerslev, Bryan T Kelly, and Lasse Heje Pedersen, "Is there a replication crisis in finance?," Technical Report, Journal of Finance Forthcoming.

___ , ___ , Semyon Malamud, and Lasse Heje Pedersen, "Machine Learning and the Implementable Efficient Frontier," *Available at SSRN 4187217*, 2022.

Kelly, Bryan, Semyon Malamud, and Kangying Zhou, "The Virtue of Complexity in Return Prediction," *Swiss Finance Institute Research Paper*, 2021, (21-90).

___ , ___ , and ___ , "The Virtue of Complexity Everywhere," *Swiss Finance Institute Research Paper*, 2022, (22-57).

___ , Seth Pruitt, and Yinan Su, "Characteristics are Covariances: A Unified Model of Risk and Return," *Journal of Financial Economics*, 2020.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, "Interpreting Factor Models," *The Journal of Finance*, 2018, *73* (3), 1183–1223.

___ , ___ , and ___ , "Shrinking the cross-section," *Journal of Financial Economics*, 2020, *135* (2), 271–292.

Leippold, Markus, Qian Wang, and Wenyu Zhou, "Machine learning in the Chinese stock market," *Journal of Financial Economics*, 2022, *145* (2), 64–82.

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.

Liu, Yang, Guofu Zhou, and Yingzi Zhu, "Maximizing the Sharpe ratio: A genetic programming approach," *Available at SSRN 3726609*, 2020.

Ma, Hao, "Conditional Latent Factor Models Via Econometrics-Based Neural Networks," *Available at SSRN*, 2021.

Martin, Ian WR and Stefan Nagel, "Market efficiency in the age of big data," *Journal of Financial Economics*, 2021.

McLean, R David and Jeffrey Pontiff, "Does academic research destroy stock return predictability?," *The Journal of Finance*, 2016, *71* (1), 5–32.

Moritz, Benjamin and Tom Zimmermann, "Tree-based conditional portfolio sorts: The relation between past and future stock returns," *Available at SSRN 2740751*, 2016.

Silverstein, Jack W and ZD Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate analysis*, 1995, *54* (2), 175–192.

Simon, Frederik, Sebastian Weibels, and Tom Zimmermann, "Deep Parametric Portfolio Policies," *Available at SSRN 4150292*, 2022.

A Properties of the Infeasible Portfolio

By a direct calculation,[19]

$$\lambda = \frac{1}{1+MaxSR^2}\,\mathrm{Var}[F]^{-1}E[F]\,, \tag{76}$$

where we have defined

$$MaxSR^2 \;=\; E[F]'\mathrm{Var}[F]^{-1}E[F] \tag{77}$$

to be the maximal achievable unconditional squared Sharpe ratio. Most existing papers perform their analysis assuming that the population moments of the factors are directly observable and, hence, so is the vector of factor risk premia, λ. The corresponding portfolio satisfies

$$E[\lambda' F_{t+1}] \;=\; E[(\lambda' F_{t+1})^2] \;=\; E[F]'E[FF']^{-1}E[F] \;=\; \frac{MaxSR^2}{1+MaxSR^2}\,. \tag{78}$$

It will be instructive for our subsequent analysis to decompose the maximal Sharpe ratio into the contributions coming from the factor principal components. Given the eigenvalue decomposition $\mathrm{Var}[F] \;=\; U\,\mathrm{diag}(\mu)U'$, we can define PC_i to be the i-th column of $U'F$. In the sequel, we will use

$$\theta \;=\; U'E[F] \tag{79}$$

[19] See the Sherman-Morrison formula (124).

to denote the vector of mean returns of the PCs. Then, we can rewrite the maximal Sharpe ratio (77) as

$$MaxSR^2 \;=\; \sum_i \frac{\theta_i^2}{\mu_i} \;=\; \sum_i (SR(PC_i))^2\,. \tag{80}$$

We will now use this representation to understand the effect of ridge shrinkage on the performance of the *infeasible* efficient portfolio,

$$R_{t+1}^{infeas}(z) \;=\; E[F]'(zI + \mathrm{Var}[F])^{-1}F_{t+1}\,. \tag{81}$$

We call this portfolio *infeasible* because, in the big data regime, when $P > T$, neither $E[F] \in \mathbb{R}^P$ nor $E[FF'] \in \mathbb{R}^{P\times P}$ can be efficiently estimated from only T observations. By construction, $R_{t+1}^{infeas}(0) = \lambda' F_{t+1}$ achieves the $MaxSR$, and

$$\mathcal{R}_1^{infeas}(z) \;=\; E[R^{infeas}(z)] \;=\; E[F]'(zI + E[FF'])^{-1}E[F] \;=\; \frac{A(z)}{1+A(z)}\,, \tag{82}$$

where we have defined

$$A(z) \;=\; E[F]'(zI + \mathrm{Var}[F])^{-1}E[F] \;=\; \sum_i (SR(PC_i))^2 \frac{\mu_i}{\mu_i + z} \tag{83}$$

and

$$A'(z) \;=\; -\sum_i \theta_i^2 \frac{1}{(\mu_i+z)^2}\,. \tag{84}$$

The function $A(z)$ will be important in understanding ridge-regularization in the high-complexity case. It turns out that the risk of the efficient portfolio can be expressed in terms of the derivative of $A(z)$: Defining

$$(zA(z))' \;=\; \sum_i (SR(PC_i))^2 \left(\frac{\mu_i}{\mu_i+z}\right)^2 , \tag{85}$$

a somewhat tedious calculation implies that

$$\mathrm{Var}[R^{infeas}(z)] = \frac{(zA(z))'}{(1+A(z))^2} . \tag{86}$$

and

$$\begin{aligned}
\mathcal{R}_2^{infeas}(z) \;&=\; E[(R^{infeas}(z))^2] \\
&= \frac{1}{(1+A(z))^2} E[(E[F]'(zI+\Psi)^{-1}F_t)^2] \;=\; \frac{1}{(1+A(z))^2} E[E[F]'(zI+\Psi)^{-1}F_tF_t'(zI+\Psi)^{-1}E[F]] \\
&= \frac{1}{(1+A(z))^2} E[E[F]'(zI+\Psi)^{-1}F_tF_t'(zI+\Psi)^{-1}E[F]] \\
&= E[F]'(zI+\Psi)^{-1}\Psi(zI+\Psi)^{-1}E[F] \;+\; \mathcal{R}_1(z)^2 \\
&= \frac{1}{(1+A(z))^2} \sum_i \theta_i^2 (z+\mu_i)^{-2}\mu_i \;+\; \left(\frac{A(z)}{1+A(z)}\right)^2 \\
&= \frac{A(z)+zA'(z)+A^2(z)}{(1+A(z))^2} \\
&= \frac{(A(z)+zA'(z))(1+A(z))-zA(z)A'(z)}{(1+A(z))^2} \\
&= \frac{d}{dz}\left(\frac{zA(z)}{1+A(z)}\right) .
\end{aligned} \tag{87}$$

Since the weights $\frac{\mu_i}{\mu_i+z}$ are monotone increasing in μ_i, we see that all that the ridge shrinkage does it re-weights principal components, giving a larger weight to higher-variance PCs. The following is a simple but important observation, implying that ridge shrinkage is always detrimental to performance.

Lemma 2 *The Sharpe ratio $SR^{infeasible}(z) \; = \; SR(R^{infeasible}(z))$ is monotone decreasing in z.*

We now use properties of $\mathcal{R}_1, \mathcal{R}_2$ to prove Corollary 4.

Proof of Corollary 4. Define

$$\Phi(x,y) \; = \; \frac{\mathcal{R}_2(x)}{\mathcal{R}_1(x)^2} \; + \; y\, \frac{1-2\mathcal{R}_1(x)+\mathcal{R}_2(x)}{\mathcal{R}_1(x)^2} \tag{88}$$

where

$$\mathcal{R}_1(x) \; = \; \frac{A(x)}{1+A(x)}, \; A(x) \; = \; \sum_i \theta_i^2(\mu_i+x)^{-1} \tag{89}$$

and

$$\mathcal{R}_2(x) \; = \; \frac{A(x)+xA'(x)+A^2(x)}{(1+A(x))^2} \tag{90}$$

so that

$$\begin{aligned} \frac{1-2\mathcal{R}_1(x)+\mathcal{R}_2(x)}{\mathcal{R}_1(x)^2} \; &= \; \frac{1-2\frac{A(x)}{1+A(x)}+\frac{A(x)+xA'(x)+A^2(x)}{(1+A(x))^2}}{A(x)^2}(1+A(x))^2 \\ = \; & \frac{1+2A(x)+A(x)^2-2A(x)(1+A(x))+A(x)+xA'(x)+A^2(x)}{A(x)^2} \\ = \; & \frac{1+A(x)+xA'(x)}{A(x)^2} \end{aligned} \tag{91}$$

so that

$$\begin{aligned} \Phi(x,y) \; &= \; 1+\frac{A(x)+xA'(x)}{A(x)^2} \; + \; y\frac{1+A(x)+xA'(x)}{A(x)^2} \\ = \; 1 \; &+ \; \frac{y}{A(x)^2} \; + \; (1+y)\frac{A(x)+xA'(x)}{A(x)^2} \end{aligned} \tag{92}$$

Then,

$$\Phi_x(x,y) \;=\; 1 \;-\; \frac{2yA'(x)}{A(x)^3} \;+\; (1+y)\frac{(2A'(x)+xA''(x))A(x)^2 - 2A(x)A'(x)(A(x)+xA'(x))}{A(x)^4}$$
$$=\; 1 \;-\; \frac{2yA'(x)A(x)}{A(x)^4} \;+\; (1+y)\frac{xA''(x)A(x)^2 - 2xA(x)(A'(x))^2}{A(x)^4} \tag{93}$$

□

B Data Generating Process Consistent with the Factor Structure

Definition 1 (Strongly uncorrelated variables) *We say that $f_i,\ i=1,\cdots,K$ are strongly uncorrelated if $E[f_{i_1}f_{i_2}]=0$ for all $i_1\neq i_2$, $E[f_{i_1}f_{i_2}f_{i_3}]=0$ for any i_1,i_2,i_3 and $E[f_{i_1}f_{i_2}f_{i_3}f_{i_4}]=0$ unless the set $\{i_1,i_2,i_3,i_4\}$ contains exactly two different elements.*

Lemma 3 *Suppose that $X=(X_i)_{i=1}^P$ with X_i being strongly uncorrelated according to Definition 1. Suppose also that $E[X_i^2]=1, E[X_i^4]\leq k$, and let A_P be random matrices independent of X and such that $\|A_P\|_2=o(1)$. Let also*

$$Y_t \;=\; X_t'A_PX_t\,. \tag{94}$$

Then,

$$(1)\ Y_t=\operatorname{tr}(A_PX_tX_t') \tag{95}$$

$$(2)\ \lim_{P\to\infty} E[(Y_t-\operatorname{tr}(A_P))^2|A_P]=0 \tag{96}$$

In particular, If $A_P=B_P/P$ where $\|B_P\|\leq K$, we have $\|A_P\|_2^2\leq P\|B_P\|^2/P^2\leq K$, and *hence*

$$\lim_{P\to\infty} E[(X_t' B_P X_t - \mathrm{tr}(B_P))^2 | B_P]/P^2 \;=\; 0\,. \tag{97}$$

Proof of Lemma 3.

(1):

$$X_t' A X_t \in R \Rightarrow X_t' A X_t = \mathrm{tr}(X_t' A X_t)$$

$$\mathrm{tr}(AB) = \mathrm{tr}(BA) \Rightarrow \mathrm{tr}(X_t' A X_t) = \mathrm{tr}(A X_t X_t')$$

(2): Define $Y_t = X_t A_P X_t$. We have

$$E[Y_t] \;=\; E[\mathrm{tr}(A_P(X_t X_t'))|A_P] = \mathrm{tr}(A_P E[X_t X_t']) = \mathrm{tr}(A_P)\,,$$

and hence

$$E[(Y_t - \mathrm{tr}(A_P))^2 | A_P] \;=\; \mathrm{Var}[Y_t|A_P] \;=\; E[Y_t^2|A_P] - E[Y_t|A_P]^2 \tag{98}$$

and hence it suffices to prove that

$$E[Y_t^2|A_P] \;-(\mathrm{tr}(A_P))^2 \;\to\; 0 \tag{99}$$

For simplicity, we assume from now on that A_P is deterministic, and write $A_P = (A_{i,j})_{i,j=1}^P$. Then,

$$Y_t \;=\; \sum_{i,j} X_i X_j A_{i,j} \tag{100}$$

and therefore

$$Y_t^2 \;=\; \sum_{i_1,j_1,i_2,j_2} X_{i_1}X_{j_1}A_{i_1,j_1}A_{i_2,j_2}X_{i_2}X_{j_2} \tag{101}$$

Now we attempt to compute the expectation:

$$\begin{aligned} E[Y_t^2] \;&=\; \sum_{i_1,j_1,i_2,j_2} A_{i_1,j_1}A_{i_2,j_2}E[X_{i_1}X_{j_1}X_{i_2}X_{j_2}] \\ &= (\sum_i A_{i,i}^2 E[X_i^4] + \sum_{i,j}(A_{i,j}^2 + A_{i,i}A_{j,j})E[X_i^2X_j^2] \\ &= (\sum_i kA_{i,i}^2 + \sum_{i,j} A_{i,j}^2 + 2A_{i,i}A_{j,j}) \\ &= ((k-1)\sum_i A_{i,i}^2 + \sum_{i,j} A_{i,j}^2 + \mathrm{tr}(A)^2) \end{aligned} \tag{102}$$

We have

$$\sum_i A_{i,i}^2 \;\leq\; \sum_{i,j} A_{i,j}^2 \;=\; \|A\|_2^2\,, \tag{103}$$

and therefore

$$|E[Y_t^2] - \mathrm{tr}(A)^2| \;\leq\; k\|A_2\|_2^2\,, \tag{104}$$

and the proof is complete. □

Assumption 3 *There exist independent random matrices $X_t \in \mathbb{R}^{N\times P}$ with six finite first moments, and two symmetric, nonnegative-definite matrices $\Sigma \in \mathbb{R}^{N\times N}$ and $\Psi \in \mathbb{R}^{P\times P}$, such that*

$$S_t = \frac{1}{N^{1/2}}\Sigma^{1/2}X_t\Psi^{1/2}\,. \tag{105}$$

Furthermore, $E[X_{i,k,t}] = 0$*, and they are strongly uncorrelated. Finally, we assume that the sixth moments are uniformly bounded:* $\max_{i,k} E[X_{i,k,t}^6] \le K$ *for some* $K > 0$*.*

Assumption 3 implies that Ψ and Σ are identifiable only up to a multiplicative constant. Indeed, multiplying Σ by a constant and dividing Ψ by the same constant does not change S_t. Up to this constant, Ψ and Σ can be identified using the identities

$$E[S'S] \;=\; \mathrm{tr}(\Sigma/N)\Psi \;\in\; \mathbb{R}^{P\times P} \quad \text{and} \quad E[SS'] \;=\; \mathrm{tr}(\Psi)\Sigma/N \in\; \mathbb{R}^{N\times N}. \tag{106}$$

While Ψ captures the covariance structure of characteristics across characteristics, Σ captures the covariance structure of signals across assets. The latter defines the cross-sectional diversification capacity of the characteristics-based portfolios. For example, suppose that $\mathrm{rank}\Sigma = 1$, so that $\Sigma^{1/2} = \pi\pi'$ for some $\pi \in \mathbb{R}^N$. Then, $S_t = N^{-1/2}\pi\pi' X_t \Psi^{1/2}$ and therefore all factors are given by

$$F_{t+1} \;=\; \Psi^{1/2} X_t' \pi(\pi' R_{t+1})\,, \tag{107}$$

implying that all factor returns are proportional to returns on a single portfolio, $\pi' R_{t+1}$. Thus, there are no diversification benefits from constructing a portfolio of factors. The same happens when Σ has only a few large eigenvalues. Our next technical assumption ensures that this pathological situation cannot occur.

Assumption 4 (Diversification) *We have* $\mathrm{tr}(\Sigma/N) \to 1$[20] *and* $\mathrm{tr}(\Sigma^2/N^2) \to 0$.

This assumption implies that the signals for asset returns $R_{i,t+1}$ are sufficiently diversified because a few top principal components do not dominate the factor portfolio returns. As an illustration, consider the case when $\mathrm{rank}(\Sigma) = 1$. In this case, $\sigma_* = 1$ means that $\mathrm{tr}(\Sigma) = N$ and $\mathrm{tr}(\Sigma^2) = N^2$. Let π be the corresponding eigenvector. Then, $\pi' S_t$ is the only linear

[20] This normalization is without loss of generality.

combination of signals with non-zero variance, and hence $F_{t+1} = S_t' R_{t+1} = (S_t'\pi)\pi' R_{t+1}$. That is, all factor returns are proportional to the returns on just one portfolio, $\pi' R_{t+1}$. In this case of an extreme concentration of predictive power of the signals, there are no diversification benefits from a large cross-section: In fact, there is effectively only one asset, with return $\pi' R_{t+1}$, and our results do not apply. Empirically, we find strong support for this assumption, finding that the *Herfindahl index*, $\mathrm{tr}(\Sigma^2)/(\mathrm{tr}(\Sigma)^2)$ is around $1/N$ is all samples we consider.

In order to proceed further, we make assumptions about the conditional covariance matrix of returns.

Assumption 5 *We assume that* $R_{t+1} = S_t \tilde{F}_{t+1} + \varepsilon_{t+1}$ *where* $E[\varepsilon_{t+1}] = 0$ *and* $E[\varepsilon_{t+1}\varepsilon_{t+1}'] = \Sigma_\varepsilon$ *is uniformly bounded. The latent factors satisfy* $E[\tilde{F}_{t+1}] = \lambda$ *and* $\Sigma_F = E[\tilde{F}_{t+1}\tilde{F}_{t+1}']$ *satisfies* $\mathrm{tr}(\Sigma_F) = O(1)$. *We also use* $\Sigma_F^* = E[\tilde{F}_{t+1}\tilde{F}_{t+1}'] - \lambda\lambda'$ *to denote the covariance matrix of latent factors.*

We now define

$$F_{t+1} \;=\; N^{1/2} S_t' R_{t+1} \tag{108}$$

In this section, we investigate the feasible counter-part of the efficient factor portfolio, with both $E[F]$ and $E[FF']$ estimated in finite samples: Namely, we define

$$\hat{\lambda}(z) \;=\; (zI + B_T)^{-1} \frac{1}{NT} \sum_{t=1}^{T} F_t \tag{109}$$

where

$$B_T \;=\; \frac{1}{NT} \sum_{t=1}^{T} F_t F_t'\,. \tag{110}$$

The ridge regularization zI is necessary to take care of the fact that the matrix B_T degenerates when $T < P$. When $z = 0$, portfolio (109) is the natural finite sample counterpart of the infeasible efficient portfolio. The corresponding realized (out-of-sample) returns are then given by

$$R^F_{T+1}(z) \;=\; \hat{\lambda}(z)' F_{t+1} \;=\; (N^{1/2} S_t \hat{\lambda}(z))' R_{t+1}\,. \tag{111}$$

Our goal in this paper is to understand the performance of this portfolio in the limit as $T, P \to \infty$. Standard arguments based on the law of large numbers imply that

$$\lim_{T\to\infty,\ P/T\to 0} \hat{\lambda}(z) \;=\; (zI + E[F_t F_t'])^{-1} E[F_t]\,. \tag{112}$$

In particular, when $z = 0$, we have $\hat{\lambda}(z) \to \pi_F$ and Proposition 10 implies that $\hat{\lambda}(0)$ achieves the maximal possible conditional Sharpe ratio, coinciding with that of the conditionally efficient portfolio. The condition $P/T \to 0$ is key to this result. It corresponds to a limit of *zero complexity.* By contrast, in this paper we are interested in the high complexity limit, corresponding to $P/T \to c > 0$.

The first step in our analysis is to understand the asymptotic behavior of the empirical factor covariance matrix, B_T, defined in (110). As we show below, a key role in our results is played by the eigenvalue distribution of B_T. We start with the following technical lemma.

Lemma 4 *Suppose that $E[(UX_{i,k})^4] = \xi_k$ is independent of i, where U is the eigenmatrix of Σ. We have*

$$\begin{aligned} E[B_T] \;=\; \frac{1}{N} E[F_t F_t'] \;=\; \frac{1}{N^2}\Bigg(&((\operatorname{tr}\Sigma)^2 + \operatorname{tr}(\Sigma^2))\Psi N^{-1}\Sigma_F \Psi \\ +\; \operatorname{tr}(\Sigma^2)\Psi^{1/2}\operatorname{diag}(\xi - 2)\operatorname{diag}(\Psi^{1/2} N^{-1}\Sigma_F \Psi^{1/2})\Psi^{1/2} &+ \Psi\Big(N \operatorname{tr}(\Sigma\Sigma_\varepsilon) + \operatorname{tr}(\Psi N^{-1}\Sigma_F)\operatorname{tr}(\Sigma^2)\Big)\Bigg) \end{aligned} \tag{113}$$

While one might hope that the eigenvalue distribution of B_T coincides with that of $\frac{1}{N}E[F_tF_t']$ in the $T \to \infty$ limit, this is only true in the zero complexity limit when $P/T \to 0$. Once $P/T \to c > 0$, the eigenvalue distribution of B_T and $\frac{1}{N}E[F_tF_t']$ diverge. The following is true.

Theorem 9 *The eigenvalue distribution of $\frac{1}{N}E[F_tF_t']$ converges to that of $\Psi\sigma_*$ where $\sigma_* = \lim N^{-1}\operatorname{tr}(\Sigma\Sigma_\varepsilon)$ in the limit as $N, P, T \to \infty,\ P/T \to c$, so that*

$$\frac{1}{P}\operatorname{tr}\left((zI + \frac{1}{N}E[F_tF_t'])^{-1}\right) \ \to\ \sigma_*^{-1}m_\Psi(-z/\sigma_*) \ =\ \frac{1}{P}\operatorname{tr}((zI+\sigma_*\Psi)^{-1})\,. \tag{114}$$

whereas

$$\frac{1}{P}\operatorname{tr}((zI+B_T)^{-1}) \ \to\ m(-z;c)\,, \tag{115}$$

where, for each $z < 0$, we have that $m(z;c)$ is the unique positive solution to the non-linear master equation

$$m(z;c) = \frac{1}{1\ -\ c\ -\ c\,z\,m(z;c)}\,m_{\sigma_*\Psi}\left(\frac{z}{1\ -\ c\ -\ c\,z\,m(z;c)}\right)\,. \tag{116}$$

Perhaps surprisingly, the $((\operatorname{tr}\Sigma)^2 + \operatorname{tr}(\Sigma^2))\Psi N^{-1}\Sigma_F\Psi$ term from (113) is "lost" because it has rank one and therefore does not affect the eigenvalue distribution. See, for example, Lemma 2.4 in (Silverstein and Bai, 1995). As we show in Lemma 10 in the Appendix, the kurtosis term also has no impact on the asymptotic eigenvalue distribution. This theorem's proof is non-trivial and based on techniques from the random matrix theory from (Bai and Zhou, 2008). Applying standard results from random matrix theory to F_t is not straightforward because of the complex cross-dependence in higher moments of F_t introduced by the signals. Namely, even if R_{t+1} are conditionally independent, $S_t'R_{t+1}$ have very strong cross-dependencies.

Proof of Proposition 7. We have

$$
\begin{aligned}
PricingError(z;cq;q) &= E[F'(1-\lambda(z;q)'F(q))]\,E[FF']^{-1}E[(1-\lambda(z;q)'F)F] \\
&= (E[F]-E[FF(q)']\lambda(z;q))'E[FF']^{-1}(E[F]-E[FF(q)']\lambda(z;q)) \\
&= E[F]'E[FF']^{-1}E[F]-2\,\underbrace{E[R^F(z;q)F']E[FF']^{-1}E[F]}_{directional} \\
&+ \underbrace{E[R^F(z;q)F']E[FF']^{-1}E[R^F(z;q)F]}_{risk} \\
&= E[F]'E[FF']^{-1}E[F]-2E[R^F(z;q)] \;+\; E[(R^F(z;q))^2]
\end{aligned}
\tag{117}
$$

We have

$$
E\left[\hat{\lambda}(z;q)'\left(\frac{1}{\hat{T}}\sum_\tau (F_\tau(q))F_\tau'\right)((0+)I+\hat{B}_{\hat{T}})^{-1}\left(\frac{1}{\hat{T}}\sum_\tau F_\tau\right)\right] \tag{118}
$$

Now, all matrices here have a block structure:

$$
\left(\frac{1}{\hat{T}}\sum_\tau (F_\tau(q))F_\tau'\right) \;=\; [\hat{B}_{\hat{T}}(q)+(0+)I\,,\,\hat{\Psi}_{1,2}] \tag{119}
$$

where $\hat{\Psi}_{1,2}\in\mathbb{R}^{P_1\times(P-P_1)}$ and, assuming for simplicity that

$$
\left(\frac{1}{\hat{T}}\sum_\tau (F_\tau(q))F_\tau'\right)((0+)I+\hat{B}_{\hat{T}})^{-1} \;=\; [I_{P_1\times P_1},0_{P_1\times(P-P_1)}] \tag{120}
$$

by the definition of the inverse matrix. Namely,

$$
(A,B)\begin{pmatrix} A & B\\ C & D\end{pmatrix}^{-1} \;=\; (I,0) \tag{121}
$$

Thus,

$$E[R^F(z;q)F']E[FF']^{-1} \;=\; \hat{\lambda}(z;q)'(I,0) \tag{122}$$

and hence

$$\begin{aligned}
&E[R^F(z;q)F']E[FF']^{-1}E[R^F(z;q)F] \\
&= \; E[R^F(z;q)F']E[FF']^{-1}E[FF']E[FF']^{-1}E[R^F(z;q)F] \\
&= \; \hat{\lambda}(z;q)'E[F(q)F(q)']\hat{\lambda}(z;q)\,.
\end{aligned} \tag{123}$$

□

C Proofs for the Infinite Sample

We will frequently be using the Sherman-Morrison formula

$$(A+xx')^{-1} \;=\; A^{-1} - A^{-1}xx'A^{-1}/(1+x'Ax) \tag{124}$$

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

Lemma 5 *We have*

$$(A+B)^{-1} \;=\; B^{-1} - (A+B)^{-1}AB^{-1}\,, \tag{125}$$

and

$$(A+B)^{-1}AB^{-1} \;\le\; A \tag{126}$$

in the sense of positive semi-definite order.

Proof of Lemma 5. We have

$$(A+B)^{-1}AB^{-1} \;=\; B^{-1/2}(\hat{A}+I)^{-1}\hat{A}B^{-1/2} \;\leq\; B^{-1/2}\hat{A}B^{-1/2} = B^{-1}AB^{-1} \tag{127}$$

$\square$

Proposition 10 *Let*

$$R'_{t+1}\pi_t^{MV} \;=\; F'_{t+1}(S_t\Sigma_F^*S_t' \;+\; \Sigma_\varepsilon)^{-1}S_t\lambda \tag{128}$$

be the conditionally efficient portfolio of stocks, and

$$F'_{t+1}\bar{\lambda} \;=\; F'_{t+1}(\Psi+\Psi\Sigma_F^*\Psi)^{-1}\Psi\lambda \tag{129}$$

be the unconditionally efficient portfolio of factors. Suppose also that

$$\lambda' A\lambda \;\to\; 0 \tag{130}$$

for any A with uniformly bounded $\|A\|_1$-norm. For example, this is then case when $\lambda \sim N(0,\Sigma_\lambda/P)$ for some uniformly bounded Σ_λ. Then, in the limit as $P\to\infty$, the two portfolios' squared Sharpe ratios both converge to

$$\frac{\lambda'\Psi\lambda}{1+\lambda'\Psi\lambda}$$

Proof of Proposition 10. We have

$$((\Sigma_F)^{-1}+S_t'S_t)^{-1} \;\leq\; ((\Sigma_F)^{-1})^{-1}$$

Hence, defining

$$Q_t \;=\; (S_t\Sigma_F^* S_t' \;+\; \Sigma_\varepsilon)^{-1} \;=\; \Sigma_\varepsilon^{-1} \;-\; (S_t\Sigma_F^* S_t' \;+\; \Sigma_\varepsilon)^{-1} S_t\Sigma_F^* S_t'\Sigma_\varepsilon^{-1}\,, \tag{131}$$

we get

$$\begin{aligned}
&E[R_{t+1}'\pi_t^{MV}] \\
&=\; E[(S_t\widetilde{F}_{t+1} \;+\; \varepsilon_{t+1})'(S_t(\Sigma_F)S_t' \;+\; \Sigma_\varepsilon)^{-1}S_t\lambda] \\
&=\; E[\lambda' S_t'(S_t(\Sigma_F)S_t' \;+\; \Sigma_\varepsilon)^{-1}S_t\lambda] \\
&=\; E[\lambda' S_t'(S_t(\lambda\lambda' + \Sigma_F^*)S_t' \;+\; \Sigma_\varepsilon)^{-1}S_t\lambda] \\
&=\; E[\lambda' S_t'((S_t\lambda)(S_t\lambda)' + (S_t\Sigma_F^* S_t' \;+\; \Sigma_\varepsilon))^{-1}S_t\lambda] \\
&\underbrace{=}_{(124)}\; E[\lambda' S_t'(Q_t - Q_tS_t\lambda\lambda' S_t'Q_t(1+\lambda' S_t'Q_tS_t\lambda)^{-1})S_t\lambda] \\
&=\; E[Z_t - Z_t^2(1+Z_t)^{-1}] \;=\; E[Z_t/(1+Z_t)]\,,
\end{aligned} \tag{132}$$

where

$$Z_t \;=\; \lambda' S_t' Q_t S_t \lambda \;=\; \lambda' S_t'\Sigma_\varepsilon^{-1}S_t\lambda - q\,, \tag{133}$$

where, by Lemma 5,

$$(S_t\Sigma_F^* S_t' \;+\; \Sigma_\varepsilon)^{-1}S_t\Sigma_F^* S_t'\Sigma_\varepsilon^{-1} \;\le\; \Sigma_\varepsilon^{-1}S_t\Sigma_F^* S_t'\Sigma_\varepsilon^{-1} \tag{134}$$

and hence

$$q \;=\; \lambda' S_t(S_t\Sigma_F^* S_t' \;+\; \Sigma_\varepsilon)^{-1}S_t\Sigma_F^* S_t'\Sigma_\varepsilon^{-1}S_t\lambda \;\le\; \lambda' S_t\Sigma_\varepsilon^{-1}S_t\Sigma_F^* S_t'\Sigma_\varepsilon^{-1}S_t\lambda\,. \tag{135}$$

We have that, by Corollary 11,

$$\begin{aligned}
E[\lambda' S_t' \Sigma_\varepsilon^{-1} S_t A S_t' \Sigma_\varepsilon^{-1} S_t \lambda] \;&=\; \frac{1}{N^2}\lambda' \Bigg(((\operatorname{tr}\hat{\Sigma})^2 + \operatorname{tr}(\hat{\Sigma}^2))\Psi A \Psi + \operatorname{tr}(\hat{\Sigma}^2)\operatorname{tr}(\Psi A)\Psi \\
&+\; \operatorname{tr}(\hat{\Sigma}^2)\,(\kappa-2)\Psi^{1/2}\operatorname{diag}(\Psi^{1/2} A \Psi^{1/2})\Psi^{1/2} \Bigg)\lambda \\
&\approx \lambda' \Psi A \Psi \lambda \\
&=\; \lambda' U D U' A U D U' \lambda \;\to\; 0
\end{aligned} \tag{136}$$

by (130), with

$$A \;=\; \Sigma_F^* \tag{137}$$

and

$$\hat{\Sigma} \;=\; \Sigma^{1/2}\Sigma_\varepsilon^{-1}\Sigma^{1/2} \tag{138}$$

Thus, by assumption, $E[q_t] \to 0$ and hence $q_t \to 0$ in probability. As a result, $Z_t - \lambda' S_t' S_t \lambda \to 0$ is probability, and hence

$$\frac{Z_t}{1+Z_t} \;\to\; \frac{\lambda'\Psi\lambda}{1+\lambda'\Psi\lambda} \tag{139}$$

in probability, and the dominated convergence theorem implies that the same holds in expectation. Similarly,

$$\begin{aligned}
&E[(\pi_t^{MV})' R_{t+1} R_{t+1}' \pi_t^{MV}] \\
&=\; E[\lambda' S_t' (S_t(\Sigma_F)S_t' \;+\; \Sigma_\varepsilon)^{-1}(S_t(\Sigma_F)S_t' \;+\; \Sigma_\varepsilon)(S_t(\Sigma_F)S_t' \;+\; \Sigma_\varepsilon)^{-1} S_t \lambda] \\
&=\; E[R_{t+1}' \pi_t^{MV}]
\end{aligned} \tag{140}$$

Now, for the factor portfolios, we have

$$E[F_t] \;=\; N^{1/2}E[S_t'R_{t+1}] \;=\; N^{1/2}E[S_t'(S_t\widetilde{F}_{t+1}+\varepsilon_{t+1})] \;=\; \frac{1}{N^{1/2}}E[\Psi^{1/2}X_t'\Sigma X_t\Psi^{1/2}\lambda] \;=\; N^{1/2}\Psi\lambda \tag{141}$$

and, again by Corollary 11, we have

$$\begin{aligned}
&\frac{1}{N}E[F_tF_t'|\lambda] \;=\; E[S_t'(S_t\widetilde{F}_{t+1} \;+\; \varepsilon_{t+1})(S_t\widetilde{F}_{t+1} \;+\; \varepsilon_{t+1})'S_t|\lambda] \;=\; E[S_t'(S_t(\Sigma_F)S_t' \;+\; \Sigma_\varepsilon)S_t] \\
&\approx\; \sigma_*\Psi \;+\Psi(\Sigma_F)\Psi\,.
\end{aligned} \tag{142}$$

For the brevity, we assume that $\sigma_* = 1$. Then, defining

$$Q \;=\; (\Psi \;+\; \Psi\Sigma_F^*\Psi)^{-1}\,, \tag{143}$$

we get that the efficient portfolio of factors is given by

$$\begin{aligned}
&\pi_F = (\Psi \;+\; \Psi(\Sigma_F)\Psi)^{-1}\Psi\lambda \\
&\{(124)\} \;=\; (Q - Q\Psi\lambda\lambda'\Psi Q(1+\lambda'\Psi Q\Psi\lambda)^{-1})\Psi\lambda \\
&=\; \frac{1}{1+Z}Q\Psi\lambda\,,
\end{aligned} \tag{144}$$

where

$$Z \;=\; \lambda'\Psi Q\Psi\lambda\,. \tag{145}$$

and we get, by the same argument as above, that $Z \to \lambda'\Psi\lambda$ because Σ_*^F has a small trace.

We then have $E[F_{t+1}] = N^{1/2}\Psi\lambda$ and, hence,

$$N^{-1/2}E[\pi_F' F_{t+1}] \;=\; E[\lambda' \frac{1}{1+Z}\Psi Q \Psi \lambda] \approx \; \frac{Z}{1+Z}\,, \tag{146}$$

while

$$N^{-1}E[\pi_F' F_{t+1} F_{t+1}' \pi_F] \;=\; E[\pi_F'(\Psi \;+\Psi(\Sigma_F)\Psi)\pi_F] = \; N^{-1/2}E[\pi_F' F_{t+1}]\,, \tag{147}$$

and the proof is complete. $\square$

Everywhere in the sequel, we abuse the notation and use the equivalent formulation where $S_t = \Sigma^{1/2}X_t\Psi^{1/2}$, while $\widetilde{F}$ is rescaled by $N^{-1/2}$, so that Σ_λ and Σ_F^* are both multiplied by $1/N$. For simplicity, we also assume that $\lambda = \lambda_P \sim N(0, \Sigma_\lambda/P)$ for some uniformly bounded sequence of matrices $\Sigma_\lambda = \Sigma_\lambda(P)$. In this case,

$$\lambda' A \lambda \;\approx P^{-1}\,\mathrm{tr}(A\Sigma_\lambda) \tag{148}$$

in probability (and in L_2). All our results hold under the more general condition (130), and all expressions can be rewritten without Σ_λ using (148). Defining $\beta_{t+1} = N^{-1/2}\widetilde{F}_{t+1}$, we can reformulate our key assumptions as

Assumption 6 *We have*

$$R_{t+1} \;=\; S_t\beta_{t+1} \;+\; \varepsilon_{t+1} \tag{149}$$

where $S_t = \Sigma^{1/2}X_t\Psi^{1/2}$, *and* $E[\beta_{t+1}] = N^{-1/2}\lambda$ *where* $E[\lambda\lambda'] = P^{-1}\Sigma_\lambda$; *and* $E[(\beta_{t+1} - \lambda)(\beta_{t+1} - \lambda)'] = N^{-1}\Sigma_F^*$. *We will also use the notation* $b_{*,1} \;=\; \mathrm{tr}(\Sigma_F^*) + P^{-1}\,\mathrm{tr}(\Sigma_\lambda)$, *and* $b_* = b_{*,1}/N$.

Lemma 6 *We have*

$$N(\beta'_{t+1}A_P\beta_{t+1} \; - \; \operatorname{tr}((\Sigma_F^* A_P) + P^{-1}\operatorname{tr}(A_P\Sigma_\lambda))) \;\to\; 0 \tag{150}$$

is L_2 and hence in probability, for any sequence of bounded matrices A_P.

Proof of Lemma 6. The proof follows directly from Lamme 3. □

We will need the following lemma, whose proof follows by direct calculation.

Lemma 7 *Suppose that $X_t \in \mathbb{R}^{N\times P}$ is a matrix with i.i.d. elements satisfying $E[X_{i,k}X_{j,l}] = \delta_{(i,k),(j,l)}$. Then,*

$$E[X_t'\Sigma X_t] \;=\; \operatorname{tr}(\Sigma)\, I_{P\times P}\,.$$

We can now prove

Lemma 8 (Expected Factor Moments) *We have*

$$E[S_t'\Sigma_\varepsilon S_t] \;=\; \operatorname{tr}(\Sigma\Sigma_\varepsilon)\Psi$$

and

$$\begin{aligned} E[F_{t+1}F_{t+1}'] \;&=\; ((\operatorname{tr}\Sigma)^2 + \operatorname{tr}(\Sigma^2))\Psi N^{-1}\Sigma_F\Psi \\ &+\; \operatorname{tr}(\Sigma^2)\Psi^{1/2}\operatorname{diag}(\kappa-2)\operatorname{diag}(\Psi^{1/2}N^{-1}\Sigma_F\Psi^{1/2})\Psi^{1/2} + \Psi\Big(\operatorname{tr}(\Sigma\Sigma_\varepsilon) + \operatorname{tr}(\Psi N^{-1}\Sigma_F)\operatorname{tr}(\Sigma^2)\Big) \end{aligned} \tag{151}$$

Thus,

$$\|N^{-1}E[F_{t+1}F_{t+1}'] - (\Psi\Sigma_F\Psi + \sigma_*\Psi)\| \;\to\; 0 \tag{152}$$

because

$$\mathrm{tr}(\Psi N^{-1}\Sigma_F) = O(1/N)$$

and

$$\|\,\mathrm{diag}(\Psi^{1/2}N^{-1}\Sigma_F\Psi^{1/2})\| \;=\; O(1/N)$$

Proof of Lemma 8. We have

$$E[F_{t+1}F_{t+1}'] \;=\; E[S_t'(S_t\widetilde{F}+\varepsilon)(S_t\widetilde{F}+\varepsilon)'S_t] \;=\; E[S_t'S_t\Sigma_F S_t'S_t] \;+\; E[S_t'\Sigma_\varepsilon S_t]\,,$$

and

$$E[S_t'\Sigma_\varepsilon S_t] \;=\; E[\Psi^{1/2}X_t'\Sigma^{1/2}\Sigma_\varepsilon\Sigma^{1/2}X_t\Psi^{1/2}] \;=\; \Psi^{1/2}E[X_t'\Sigma^{1/2}\Sigma_\varepsilon\Sigma^{1/2}X_t]\Psi^{1/2} \;=\; \Psi\,\mathrm{tr}(\Sigma\Sigma_\varepsilon)\,,$$

Defining $\tilde{\beta} = \Psi^{1/2}\beta_{t+1}$, we get

$$\begin{aligned} E[S_t'S_t\tilde{\beta}\tilde{\beta}'S_t'S_t] \;&=\; E[\Psi^{1/2}X_t'\Sigma X_t\Psi^{1/2}\tilde{\beta}\tilde{\beta}'\Psi^{1/2}X_t'\Sigma X_t\Psi^{1/2}] \;=E[\Psi^{1/2}X_t'\Sigma X_t\tilde{\beta}\tilde{\beta}'X_t'\Sigma X_t\Psi^{1/2}] \\ \;&=\; \Psi^{1/2}E[\tilde{X}_t'D\tilde{X}_t\tilde{\beta}\tilde{\beta}'\tilde{X}_t'D\tilde{X}_t]\Psi^{1/2}\,, \end{aligned} \tag{153}$$

where we have defined $\Sigma = U'DU$ and D is diagonal and U is orthogonal and $\tilde{X} = UX$ are still have the same moments as X by the assumptions made.

Now,

$$E[\tilde{X}_t'D\tilde{X}_t\tilde{\beta}\tilde{\beta}'\tilde{X}_t'D\tilde{X}_t]_{k_1,k_2} \;=\; E[\sum_{i_1,i_2,l_1,l_2} D_{i_1}D_{i_2}\,\tilde{X}_{i_1,k_1}\tilde{X}_{i_1,l_1}\tilde{\beta}_{l_1}\tilde{\beta}_{l_2}\tilde{X}_{i_2,l_2}\tilde{X}_{i_2,k_2}]\,.$$

First we study the terms with $i_1 \neq i_2$:

$$\sum_{i_1 \neq i_2} D_{i_1} D_{i_2} \, E[\sum_{l_1,l_2} \tilde{X}_{i_1,k_1} \tilde{X}_{i_1,l_1} \tilde{\beta}_{l_1} \tilde{\beta}_{l_2} \tilde{X}_{i_2,l_2} \tilde{X}_{i_2,k_2}] \; = \; \sum_{i_1 \neq i_2} D_{i_1} D_{i_2} \tilde{\beta}_{k_1} \tilde{\beta}_{k_2} = ((\mathrm{tr}\,\Sigma)^2 - \mathrm{tr}(\Sigma^2)) \tilde{\beta}_{k_1} \tilde{\beta}_{k_2}$$

At the same time,

$$\sum_{i_1 = i_2} D_{i_1}^2 \; E[\sum_{l_1,l_2} \tilde{X}_{i_1,k_1} \tilde{X}_{i_1,l_1} \tilde{\beta}_{l_1} \tilde{\beta}_{l_2} \tilde{X}_{i_2,l_2} \tilde{X}_{i_2,k_2}]$$

depends on whether $k_1 = k_2$. If $k_1 = k_2$, then we have

$$\sum_{i_1 = i_2} D_{i_1}^2 \; E[\sum_{l_1,l_2} \tilde{X}_{i_1,k_1}^2 \tilde{X}_{i_1,l_1} \tilde{\beta}_{l_1} \tilde{\beta}_{l_2} \tilde{X}_{i_1,l_2}] \; = \; \mathrm{tr}(\Sigma^2)(\kappa \tilde{\beta}_{k_1}^2 + \|\tilde{\beta}\|^2)$$

and if $k_1 \neq k_2$ then we need that ℓ_1, ℓ_2 coincide with k_1, k_2, so that

$$\sum_{i_1 = i_2} D_{i_1} D_{i_2} E[\sum_{l_1,l_2} \tilde{X}_{i_1,k_1} \tilde{X}_{i_1,l_1} \tilde{\beta}_{l_1} \tilde{\beta}_{l_2} \tilde{X}_{i_2,l_2} \tilde{X}_{i_2,k_2}] \; = \; 2 \sum_{i_1 = i_2} D_{i_1}^2 E[\tilde{X}_{i_1,k_1}^2 \tilde{X}_{i_1,k_2}^2] \tilde{\beta}_{k_1} \tilde{\beta}_{k_2} \; = \; 2\,\mathrm{tr}(\Sigma^2) \tilde{\beta}_{k_1} \tilde{\beta}_{k_2}$$

Thus,

$$\begin{aligned} & E[\tilde{X}_t' D \tilde{X}_t \tilde{\beta} \tilde{\beta}' \tilde{X}_t' D \tilde{X}_t]_{k_1,k_2} \\ &= \; ((\mathrm{tr}\,\Sigma)^2 - \mathrm{tr}(\Sigma^2)) \tilde{\beta}_{k_1} \tilde{\beta}_{k_2} \; + 2\,\mathrm{tr}(\Sigma^2) \tilde{\beta}_{k_1} \tilde{\beta}_{k_2} (1 - \delta_{k_1,k_2}) + \; \mathrm{tr}(\Sigma^2)(\kappa \tilde{\beta}_{k_1}^2 + \|\tilde{\beta}\|^2) \delta_{k_1,k_2} \\ &= \; ((\mathrm{tr}\,\Sigma)^2 + \mathrm{tr}(\Sigma^2)) \tilde{\beta}_{k_1} \tilde{\beta}_{k_2} + \mathrm{tr}(\Sigma^2)((\kappa - 2) \tilde{\beta}_{k_1}^2 + \|\tilde{\beta}\|^2) \delta_{k_1,k_2} \end{aligned} \tag{154}$$

Thus, by formula (153), we get

$$E[S_t' S_t \lambda \lambda' S_t' S_t] \; = \; ((\mathrm{tr}\,\Sigma)^2 + \mathrm{tr}(\Sigma^2)) \Psi N^{-1} \Sigma_F \Psi + \mathrm{tr}(\Sigma^2)((\kappa - 2) \Psi^{1/2} \, \mathrm{diag}(\tilde{\beta}_{k_1}^2) \Psi^{1/2} + \|\tilde{\beta}\|^2 \Psi) \tag{155}$$

and the claim follows because $\|\tilde{\beta}\|^2 = \lambda' \Psi \lambda$.

Corollary 11 *We have*

$$E[S_t'S_tAS_t'S_t] \;=\; ((\operatorname{tr}\Sigma)^2 + \operatorname{tr}(\Sigma^2))\Psi A\Psi + \operatorname{tr}(\Sigma^2)\operatorname{tr}(\Psi A)\Psi$$
$$+\;\operatorname{tr}(\Sigma^2)\,\Psi^{1/2}\operatorname{diag}(\kappa-2)\operatorname{diag}(\Psi^{1/2}A\Psi^{1/2})\Psi^{1/2} \tag{156}$$

where $\operatorname{diag}(\Psi^{1/2}A\Psi^{1/2})$ *is the diagonal matrix with diagonal coinciding with that of* $\operatorname{diag}(\Psi^{1/2}A\Psi^{1/2})$.

Proof. Writing

$$A \;=\; \sum_i \lambda_i\beta_i\beta_i'$$

we can apply the calculations for rank-one A. □

The proof of Lemma 8 is complete.

□

We will also need the following lemma from KMZ.

Lemma 9 *Define* $\xi(z;c)$ *through*

$$\frac{c^{-1}\xi(z;c)}{1+\xi(z;c)} \;=\; 1\;-\;m(-z;c)z\,. \tag{157}$$

Then,

$$\frac{1}{T}\operatorname{tr}((zI+B_T)^{-1}\Psi)\;\to\;\xi(z;c) \tag{158}$$

almost surely and

$$\frac{1}{T}F_{T+1}'(zI+B_T)^{-1}F_{T+1}\;\to\;\xi(z;c) \tag{159}$$

in probability. Furthermore, $\xi(z;c)<c/z$.

Define the effective shrinkage

$$Z^*(z;c) \;=\; z\,(1+\xi(z;c)) \;\in (z,\; z+c) \tag{160}$$

Then, $Z^*(z;c)$ *is monotone increasing in* z *and* c. *In the ridgeless limit as* $z \to 0$, *we have*

$$Z^*(z;c) \;\to\; \begin{cases} 0, & c < 1 \\ 1/\tilde{m}(c), & c > 1 \end{cases} \tag{161}$$

where $\tilde{m}(c) > 0$ *is the unique positive solution to*

$$c - 1 \;=\; \frac{\int \frac{dH(x)}{\tilde{m}(1+\tilde{m}\,x)}}{\int \frac{x dH(x)}{1+\tilde{m}\,x}} \tag{162}$$

D The Proof that Managed Portfolio Returns Satisfy Assumption 2 of RMT

Lemma 10 *Let* X_P *be a sequence of positive semi-definite matrices with* $\mathrm{tr}(X_P) \le K$. *Then,*

$$\lim_{M\to\infty} (\frac{1}{P}\,\mathrm{tr}(zI + A_P + X_P)^{-1} - \frac{1}{P}\,\mathrm{tr}(zI + A_P)^{-1}) \;=\; 0$$

for any positive semi-definite matrices A_P.

Proof. We have

$$\frac{1}{P}\,\mathrm{tr}(zI + A_P + X_P)^{-1} - \frac{1}{P}\,\mathrm{tr}(zI + A_P)^{-1} \;=\; \frac{1}{P}\,\mathrm{tr}((zI + A_P + X_P)^{-1} - (zI + A_P)^{-1})$$

and the claim follows because

$$\frac{1}{P}\,\mathrm{tr}((zI + A_P + X_P)^{-1} - (zI + A_P)^{-1}) \;=\; -\frac{1}{P}\,\mathrm{tr}((zI + A_P + X_P)^{-1} X_P (zI + A_P)^{-1})$$

and

$$
\begin{aligned}
\operatorname{tr}((zI + A_P + X_P)^{-1} X_P (zI + A_P)^{-1}) \;&=\; \operatorname{tr}(X_P (zI + A_P)^{-1} (zI + A_P + X_P)^{-1}) \\
\le\; \operatorname{tr}(X_P) \|(zI + A_P)^{-1} (zI + A_P + X_P)^{-1}\| \;&\le\; K z^{-2}
\end{aligned} \tag{163}
$$

Thus, the difference is bounded in absolute value by Kz^{-2}/M. □

We will need the following auxiliary lemma.

Lemma 11 *Let ε be a random vector with independent $N(0,1)$ coordinates. We have*

$$E[\varepsilon Z' \varepsilon] \;=\; Z$$

and

$$E[\varepsilon' Z \varepsilon'] \;=\; Z'$$

for any vector Z. Furthermore,

$$E[\varepsilon' A \varepsilon] \;=\; \operatorname{tr}(A)$$

for any matrix A. Furthermore,

$$E[\varepsilon_t' B \varepsilon_t \varepsilon_t' B \varepsilon_t] \;=\; (\kappa_\varepsilon - 1)\, 0.5 (\operatorname{tr}(BB) + \operatorname{tr}(B'B)) \;+\; \operatorname{tr}(B)^2 \tag{164}$$

and

$$E[\varepsilon_t \varepsilon_t' B \varepsilon_t \varepsilon_t'] \;=\; (\kappa_\varepsilon - 1) 0.5 (B + B') \;+\; \operatorname{tr}(B)$$

where $\kappa_\varepsilon = E[\tilde{\varepsilon}^4]$.

Proof. We have

$$E[\varepsilon Z'\varepsilon]_{i,j} \;=\; E[\varepsilon_i \sum_j Z_j \varepsilon_j] \;=\; \sum_j \Sigma_{\varepsilon,i,j} Z_j$$

and the first claim follows. The second claim follows because

$$E[\varepsilon' Z \varepsilon'] \;=\; E\varepsilon Z'\varepsilon]' \,.$$

For the third claim, we have

$$E[\varepsilon' A \varepsilon] \;=\; \operatorname{tr} E[\varepsilon' A \varepsilon] = \operatorname{tr} E[A\varepsilon\varepsilon'] = \operatorname{tr}(A) \tag{165}$$

For the last claim: first, we do a transformation $\varepsilon_t = \tilde{\varepsilon}_t$ and then we make the observation that, for any matrix B,

$$\varepsilon' B \varepsilon \;=\; 0.5\varepsilon'(B + B')\varepsilon.$$

Since $0.5(B + B')$ is symmetric, we can diagonalize it: $\tilde{B} = (0.5(B + B'))$. Then,

$$E[\varepsilon_t' B \varepsilon_t \varepsilon_t' B \varepsilon_t] \;=\; E[(\sum_i \varepsilon_{i,t}^2 \lambda_i(0.5(B + B')))^2] \;=\; (\kappa_\varepsilon - 1)\, \operatorname{tr}(\tilde{B}^2) \;+\; \operatorname{tr}(\tilde{B})^2 \,, \tag{166}$$

and we have

$$\operatorname{tr}(\tilde{B}^2) \;=\; \operatorname{tr}((0.5(B + B'))(0.5(B + B'))) \;=\; 0.25(\operatorname{tr}(BB) + 2(\operatorname{tr} B'B) + \operatorname{tr}(B'B'))$$

and

$$\operatorname{tr}(B'B') \;=\; \operatorname{tr}(B'B') \;=\; \operatorname{tr}(BB) \,.$$

Let $\varepsilon = \tilde{\varepsilon}$ and $\tilde{B} = U\Lambda U'$ and $\hat{\varepsilon} = U'\tilde{\varepsilon}$

$$\begin{aligned} & E[\varepsilon_t \varepsilon_t' B \varepsilon_t \varepsilon_t'] \\ = \; & E[\tilde{\varepsilon}\tilde{\varepsilon}' \tilde{B} \tilde{\varepsilon}\tilde{\varepsilon}'] \\ = \; & U E[\hat{\varepsilon}\hat{\varepsilon}' \Lambda \hat{\varepsilon}\hat{\varepsilon}'] U' \\ = \; & U E[\hat{\varepsilon} \sum_i \hat{\varepsilon}_{i_1}^2 \lambda_{i_1}(\tilde{B}) \hat{\varepsilon}'] U' \\ = \; & (\kappa_\varepsilon - 1)\tilde{B} \; + \; \mathrm{tr}(\tilde{B}) \\ = \; & (\kappa_\varepsilon - 1) 0.5 (B + B') \; + \; \mathrm{tr}(B) \end{aligned} \tag{167}$$

□

Lemma 12 (Managed Portfolios Satisfy Assumption 2) *Let A_P be a sequence of symmetric $P \times P$ matrices such that $\|A_P\| \le K$ and A_P are independent of F_t. Then, $\frac{1}{N} E[F_t F_t']$ is uniformly bounded and*

$$\mathrm{Var}[\frac{1}{TN} F_t' A_P F_t] \; \to \; 0 \,, \tag{168}$$

so that

$$\frac{1}{TN} \left(F_t' A_P F_t \; - \; \mathrm{tr}\left(A_P \, \sigma_* \Psi \right) \right) \; \to \; 0$$

in probability. That is, averaging across P factors leads to constant risk, no matter which matrix A we use to measure it.

An important observation is that, by Lemma 8,

$$\frac{1}{TN} \mathrm{tr}(A_P E[F_t F_t']) \; \approx \; \frac{1}{TN} \mathrm{tr}(A_P (\Psi \Sigma_F \Psi + \sigma_* \Psi)) \,. \tag{169}$$

However, since Σ_F has a uniformly bounded trace norm, we have

$$\frac{1}{TN}\,\mathrm{tr}(A_P(\Psi\Sigma_F\Psi+\sigma_*\Psi)) \;\approx\; \frac{1}{TN}\,\mathrm{tr}(A_P(\sigma_*\Psi)) \tag{170}$$

Similarly, the following is a direct consequence of Lemma 8.

Lemma 13 *Let A_P, B_P be sequences of symmetric $P\times P$ matrices such that $\|A_P\|,\ \|B_P\|\le K$, and $A_P,\ B_P$ are independent of F_t. Then,*

$$\frac{1}{N}\,(\lambda' E[A_P F_t F_t' B_P]\lambda \;-\; \lambda' A_P(\Psi\Sigma_F\Psi+\sigma_*\Psi)B_P\lambda) \;\to\; 0\,.$$

If λ satisfies the technical condition (130), then

$$\frac{1}{N}\,(\lambda' E[A_P F_t F_t' B_P]\lambda \;-\; \lambda' A_P(\Psi\lambda\lambda'\Psi+\sigma_*\Psi)B_P\lambda) \;\to\; 0$$

because $\mathrm{tr}(\Sigma_F^*)$ is bounded.

Note that $\mathrm{tr}\,(A_P\,F_tF_t') \;=\; F_t'A_PF_t$.

Proof of Lemma 12. For simplicity, we will assume that A_P is deterministic.[21] We can also assume that A_P is symmetric because $F_t'A_PF_t = F_t 0.5(A_P+A_P')F_t$. We need to prove that

$$\frac{1}{(TN)^2}E[F_t'A_PF_tF_t'A_PF_t] \;-\; \left(\frac{1}{NT}E[F_t'A_PF_t]\right)^2 \;\to\; 0$$

For simplicity, we will assume that $\Sigma_\varepsilon = I$. We have by Lemma 8 that

$$\begin{aligned}
E[F_tF_t'] \;&=\; ((\mathrm{tr}\,\Sigma)^2+\mathrm{tr}(\Sigma^2))\Psi N^{-1}\Sigma_F\Psi \\
&+\mathrm{tr}(\Sigma^2)\Psi^{1/2}\,\mathrm{diag}(\kappa-2)\,\mathrm{diag}(\Psi^{1/2}N^{-1}\Sigma_F\Psi^{1/2})\Psi^{1/2} \;+\Psi\Big(\,\mathrm{tr}(\Sigma)+\mathrm{tr}(\Psi N^{-1}\Sigma_F)\,\mathrm{tr}(\Sigma^2)\Big)
\end{aligned}$$

[21]Otherwise, we replace all expectations below by expectations conditional on A_P.

$$(171)$$

and, with Σ_F having uniformly bounded traces and Assumption 4, we get

$$\begin{aligned}
\frac{1}{NT}E[F_t'A_PF_t] \;&=\; \frac{1}{NT}\operatorname{tr}E[A_PF_tF_t'] \\
&\approx\; \frac{1}{N^2T}\operatorname{tr}\Bigg(A_P\Bigg((\operatorname{tr}\Sigma)^2\Psi\Sigma_F\Psi + \operatorname{tr}(\Sigma^2)\Psi^{1/2}\operatorname{diag}(\kappa-2)\operatorname{diag}(\Psi^{1/2}\Sigma_F\Psi^{1/2})\Psi^{1/2} \\
&\quad + \Psi\Big(N\operatorname{tr}(\Sigma) + \operatorname{tr}(\Psi\Sigma_F)\operatorname{tr}(\Sigma^2)\Big)\Bigg)\Bigg) \\
&\approx\; T^{-1}\operatorname{tr}(A_P\Psi)
\end{aligned} \tag{172}$$

since

$$\frac{1}{TP}\operatorname{tr}(\Psi A_P\Psi\Sigma_F) \;=\; O(1/T),$$

and, similarly, the kurtosis term does not matter because it has a uniformly bounded trace.

Now, we have

$$\begin{aligned}
F_tF_t' \;&=\; S_{t-1}'(S_{t-1}\beta\beta'S_{t-1}' + \varepsilon_t\beta'S_{t-1}' + S_{t-1}\beta\varepsilon_t' + \varepsilon_t\varepsilon_t')S_{t-1} \\
&=\; Z_t\beta\beta'Z_t + S_{t-1}'\varepsilon_t\beta'Z_t + Z_t\beta\varepsilon_t'S_{t-1} + S_{t-1}'\varepsilon_t\varepsilon_t'S_{t-1}\,.
\end{aligned} \tag{173}$$

with $Z_t = S_{t-1}'S_{t-1}$. Then, using the fact that ε and all third moments of ε have zero expectations as well as Lemma 11, we have

$$
\begin{aligned}
&\frac{1}{N^2T^2}E[F_t'AF_tF_t'AF_t] \;=\; \frac{1}{N^2T^2}\,\mathrm{tr}\,E[F_tF_t'AF_tF_t'A] \\
&=\; \frac{1}{N^2T^2}\,\mathrm{tr}\,E[(Z_t\beta\beta'Z_t + S_{t-1}'\varepsilon_t\beta'Z_t + Z_t\beta\varepsilon_t'S_{t-1} + S_{t-1}'\varepsilon_t\varepsilon_t'S_{t-1})A \\
&(Z_t\beta\beta'Z_t + S_{t-1}'\varepsilon_t\beta'Z_t + Z_t\beta\varepsilon_t'S_{t-1} + S_{t-1}'\varepsilon_t\varepsilon_t'S_{t-1})A] \\
&=\; \frac{1}{N^2T^2}\,\mathrm{tr}\,E[Z_t\beta\beta'Z_tAZ_t\beta\beta'Z_tA] \\
&+\; \frac{1}{N^2T^2}2\,\mathrm{tr}\,E[Z_t\beta\beta'Z_tAS_{t-1}'\varepsilon_t\varepsilon_t'S_{t-1}A] \\
&+ \frac{1}{N^2T^2}2\,\mathrm{tr}\,E[S_{t-1}'\varepsilon_t\beta'Z_tAS_{t-1}'\varepsilon_t\beta'Z_tA] \\
&+ \frac{1}{N^2T^2}2\,\mathrm{tr}\,E[S_{t-1}'\varepsilon_t\beta'Z_tAZ_t\beta\varepsilon_t'S_{t-1}A] \\
&+ \frac{1}{N^2T^2}\,\mathrm{tr}\,E[S_{t-1}'\varepsilon_t\varepsilon_t'S_{t-1}AS_{t-1}'\varepsilon_t\varepsilon_t'S_{t-1}A] \\
&=\; \frac{1}{N^2T^2}\,\mathrm{tr}\,E[Z_t\beta\beta'Z_tAZ_t\beta\beta'Z_tA] \\
&+\; \frac{1}{N^2T^2}2\,\mathrm{tr}\,E[Z_t\beta\beta'Z_tAZ_tA] \\
&+ \frac{1}{N^2T^2}2\,\mathrm{tr}\,E[Z_tAZ_t\beta\beta'Z_tA] \\
&+ \frac{1}{N^2T^2}2\,\mathrm{tr}\,E[(\beta'Z_tAZ_t\beta)Z_tA] \\
&+ \frac{1}{N^2T^2}((\kappa_\varepsilon - 1)\,\mathrm{tr}\,E[Z_tAZ_tA] \;+\; \mathrm{tr}\,E[\mathrm{tr}(Z_tA)Z_tA]) \\
&=\; \frac{1}{N^2T^2}\,\mathrm{tr}\,E[Z_t\beta\beta'Z_tAZ_t\beta\beta'Z_tA] \\
&+\; \frac{1}{N^2T^2}4\,\mathrm{tr}\,E[Z_t\beta\beta'Z_tAZ_tA] \\
&+ \frac{1}{N^2T^2}2\,\mathrm{tr}\,E[(\beta'Z_tAZ_t\beta)Z_tA] \\
&+ \frac{1}{N^2T^2}((\kappa_\varepsilon - 1)\,\mathrm{tr}\,E[Z_tAZ_tA] \;+\; \mathrm{tr}\,E[\mathrm{tr}(Z_tA)Z_tA]) \\
&=\; Term1 \;+\; Term2 \;+\; Term3 \;+\; Term4 \;+\; Term5\,,
\end{aligned}
\tag{174}
$$

where in the last term we have used Lemma 11 to show that

$$
\begin{aligned}
&\operatorname{tr} E[S_{t-1}'\varepsilon_t\varepsilon_t' S_{t-1} A S_{t-1}'\varepsilon_t\varepsilon_t' S_{t-1} A] \\
&= \operatorname{tr} E\left[S_{t-1}'\left((\kappa_\varepsilon - 1)(S_{t-1} A S_{t-1}') \;+\; \operatorname{tr}((S_{t-1} A S_{t-1}'))\right) S_{t-1} A\right] \\
&= (\kappa_\varepsilon - 1)\operatorname{tr} E[Z_t A Z_t A] \;+\; \operatorname{tr} E[\operatorname{tr}(Z_t A) Z_t A]\,.
\end{aligned} \tag{175}
$$

In our proofs, we will be using Newton's identities.

Lemma 14 (Newton's identities) *Let A be a matrix with eigenvalues λ_i. Then,*

$$
\begin{aligned}
\sum_{i_1, i_2, i_1 \neq i_2} \lambda_{i_1}\lambda_{i_2} &= (\operatorname{tr} A)^2 - \operatorname{tr}(A^2) \\
\sum_{i_1, i_2, i_3 \ \text{all different}} \lambda_{i_1}\lambda_{i_2}\lambda_{i_3} &= (\operatorname{tr} A)^3 - 3\operatorname{tr}(A)\operatorname{tr}(A^2) \;+\; 2\operatorname{tr}(A^3) \\
\sum_{i_1, i_2, i_3, i_4 \ \text{all different}} \lambda_{i_1}\lambda_{i_2}\lambda_{i_3}\lambda_{i_4} & \\
= (\operatorname{tr} A)^4 - 6(\operatorname{tr}(A))^2\operatorname{tr}(A^2) \;+\; 3(\operatorname{tr}(A^2))^2 &+ 8(\operatorname{tr} A)(\operatorname{tr}(A^3)) - \; 6\operatorname{tr}(A^4)\,.
\end{aligned} \tag{176}
$$

We also note that Assumption 4 implies

$$
\operatorname{tr}(\Sigma^3) \;\le\; \operatorname{tr}(\Sigma^2)\operatorname{tr}(\Sigma) \;=\; o(N^3),\;\; \operatorname{tr}(\Sigma^4) \le (\operatorname{tr}(\Sigma^2))^2 \;=\; o(N^4) \tag{177}
$$

D.1 $Term1$ in (174)

We start with the first term. We have

$$
\frac{1}{T^2}\operatorname{tr} E[Z_t\beta\beta' Z_t A Z_t\beta\beta' Z_t A] \;=\; \frac{1}{T^2}E[(\beta' Z_t A Z_t \beta)^2]\,. \tag{178}
$$

Writing

$$
Z_t = S_{t-1}'S_{t-1} \;=\; \Psi^{1/2}X_{t-1}'\Sigma X_{t-1}\Psi^{1/2}
$$

and defining

$$\tilde{\beta} = \Psi^{1/2}\beta\,,$$

and

$$\tilde{A} \;=\; \Psi^{1/2} A \Psi^{1/2}\,,$$

and then using rotational invariance of all moments up to eight, we may assume that $\tilde{A}$ is diagonal and Σ is diagonal and $\tilde{\beta} = e_1\|\tilde{\beta}\| = (1, 0, \cdots, 0)\|\tilde{\beta}\|$. Note that

$$\|\tilde{\beta}\|^2 \;=\; \beta'\Psi\beta \;\sim\; b_*\frac{1}{P}\,\mathrm{tr}(\Psi)\,.$$

Then, setting $\lambda_k = \lambda_k(\tilde{A})$ we get

$$\begin{aligned}
&\frac{1}{N^2T^2}\,\mathrm{tr}\,E[Z_t\beta\beta' Z_t A Z_t \beta\beta' Z_t A] \;=\; \frac{1}{T^2}E[(\beta' Z_t A Z_t \beta)^2]\\
&=\; \frac{1}{N^2T^2}\|\tilde{\beta}\|^4 E[\Bigg(\sum_{i_1,j_1,k_1} X_{i_1,1}\lambda_{i_1}(\Sigma)X_{i_1,k_1}\lambda_{k_1}X_{j_1,k_1}\lambda_{j_1}(\Sigma)X_{j_1,1}\Bigg)^2\,]\\
&=\; \frac{1}{N^2T^2}\|\tilde{\beta}\|^4 E[\Bigg(\sum_{i_1,j_1,k_1} X_{i_1,1}\lambda_{i_1}(\Sigma)X_{i_1,k_1}\lambda_{k_1}X_{j_1,k_1}\lambda_{j_1}(\Sigma)X_{j_1,1}\Bigg)^2\,]\\
&=\; \frac{1}{N^2T^2}\|\tilde{\beta}\|^4 E[\sum_{i_2,j_2,k_2}\sum_{i_1,j_1,k_1} X_{i_1,1}\lambda_{i_1}(\Sigma)X_{i_1,k_1}\lambda_{k_1}X_{j_1,k_1}\lambda_{j_1}(\Sigma)X_{j_1,1}X_{i_2,1}\lambda_{i_2}(\Sigma)X_{i_2,k_2}\lambda_{k_2}X_{j_2,k_2}\lambda_{j_2}(\Sigma)X_{j_2,1}]
\end{aligned} \tag{179}$$

- First, consider the terms with $k_1 = k_2$ in (179):

$$\frac{1}{N^2T^2}\|\tilde{\beta}\|^4 E[\sum_{i_2,j_2}\sum_{i_1,j_1,k_1} X_{i_1,1}\lambda_{i_1}(\Sigma)X_{i_1,k_1}\lambda_{k_1}X_{j_1,k_1}\lambda_{j_1}(\Sigma)X_{j_1,1}X_{i_2,1}\lambda_{i_2}(\Sigma)X_{i_2,k_1}\lambda_{k_1}X_{j_2,k_1}\lambda_{j_2}(\Sigma)X_{j_2,1}] \tag{180}$$

Using Newton's identities, we get that the contribution of terms with $k_1 = 1$ is given by

$$
\begin{aligned}
&\|\tilde{\beta}\|^4 \frac{1}{N^2T^2} E[\sum_{i_2,j_2}\sum_{i_1,j_1} X_{i_1,1}^2 \lambda_{i_1}(\Sigma)\lambda_1^2 X_{j_1,1}^2 \lambda_{j_1}(\Sigma) X_{i_2,1}^2 \lambda_{i_2}(\Sigma) X_{j_2,1}^2 \lambda_{j_2}(\Sigma)] \\
=\ & \|\tilde{\beta}\|^4 \frac{1}{N^2T^2}\lambda_1^2 \Bigg(E[\sum_{i_2,j_2,i_1,j_1 \ all\ different} X_{i_1,1}^2 \lambda_{i_1}(\Sigma) X_{j_1,1}^2 \lambda_{j_1}(\Sigma) X_{i_2,1}^2 \lambda_{i_2}(\Sigma) X_{j_2,1}^2 \lambda_{j_2}(\Sigma)] \\
&+ E[\sum_{i_2,j_2,i_1,j_1 \ only\ two\ are\ equal} X_{i_1,1}^2 \lambda_{i_1}(\Sigma) X_{j_1,1}^2 \lambda_{j_1}(\Sigma) X_{i_2,1}^2 \lambda_{i_2}(\Sigma) X_{j_2,1}^2 \lambda_{j_2}(\Sigma)] \\
&+ E[\sum_{i_2,j_2,i_1,j_1 \ only\ three\ are\ equal} X_{i_1,1}^2 \lambda_{i_1}(\Sigma) X_{j_1,1}^2 \lambda_{j_1}(\Sigma) X_{i_2,1}^2 \lambda_{i_2}(\Sigma) X_{j_2,1}^2 \lambda_{j_2}(\Sigma)] \\
&+ E[\sum_{i_2,j_2,i_1,j_1 \ all\ four\ are\ equal} X_{i_1,1}^2 \lambda_{i_1}(\Sigma) X_{j_1,1}^2 \lambda_{j_1}(\Sigma) X_{i_2,1}^2 \lambda_{i_2}(\Sigma) X_{j_2,1}^2 \lambda_{j_2}(\Sigma)] \Bigg) \\
=\ & \|\tilde{\beta}\|^4 \lambda_1^2 \frac{1}{N^2T^2} \Bigg((\operatorname{tr}\Sigma)^4 - 6(\operatorname{tr}\Sigma)^2(\operatorname{tr}(\Sigma^2)) \ + \ 8(\operatorname{tr}\Sigma)(\operatorname{tr}(\Sigma^3)) + 3(\operatorname{tr}(\Sigma^2))^2 - 6\operatorname{tr}(\Sigma^4) \\
&+ \binom{4}{2} E[X^4] \sum_j \lambda_j(\Sigma)^2 \sum_{i_1,j_1 \neq j, i_1 \neq j_1} \lambda_{i_1}(\Sigma)\lambda_{j_1}(\Sigma) \\
&+ 4E[X^6] \sum_j \lambda_j(\Sigma)^3 \sum_{i_1 \neq j} \lambda_{i_1}(\Sigma) \\
&+ E[X^8]\operatorname{tr}(\Sigma^4) \Bigg) \\
=\ & \|\tilde{\beta}\|^4 \lambda_1^2 \frac{1}{N^2T^2} \Bigg((\operatorname{tr}\Sigma)^4 - 6(\operatorname{tr}\Sigma)^2(\operatorname{tr}(\Sigma^2)) \ + \ 8(\operatorname{tr}\Sigma)(\operatorname{tr}(\Sigma^3)) + 3(\operatorname{tr}(\Sigma^2))^2 - 6\operatorname{tr}(\Sigma^4) \\
&+ \binom{4}{2} E[X^4] \sum_j \lambda_j(\Sigma)^2 ((\operatorname{tr}(\Sigma) - \lambda_j)^2 - (\operatorname{tr}(\Sigma^2) - \lambda_j^2) \\
&+ 4E[X^6](\operatorname{tr}(\Sigma)\operatorname{tr}(\Sigma^3) - \operatorname{tr}(\Sigma^4)) + E[X^8]\operatorname{tr}(\Sigma^4) \Bigg) \\
=\ & O\Bigg((\operatorname{tr}\Sigma)^4 (\tilde{\beta}'\tilde{A}\tilde{\beta})^2/(N^2T^2) \Bigg) \ = \ O(1/T^2)
\end{aligned}
\tag{181}
$$

Here, we have used the fact that

$$(\operatorname{tr}\Sigma)^4(\tilde{\beta}'\tilde{A}\tilde{\beta})^2 \;=\; O(N^2)$$

because $(\operatorname{tr}\Sigma)^2 b_*/N$ converges to a finite limit. The rest terms with $k_1 = k_2 \neq 1$ must have i_1, i_2, j_1, j_2 have at least two identical pairs. The first contribution would be

$$\begin{aligned} &\|\tilde{\beta}\|^4 E[\sum_{i_1=i_2\neq j_1=j_2;k_1} X^2_{i_1,1}\lambda^2_{i_1}(\Sigma)X^2_{i_1,k_1}\lambda^2_{k_1}X^2_{j_1,k_1}\lambda^2_{j_1}(\Sigma)X^2_{j_1,1}] \\ &\sim\; \|\tilde{\beta}\|^4 \operatorname{tr}(\tilde{A}^2)\left((\operatorname{tr}(\Sigma^2))^2 \;-\operatorname{tr}(\Sigma^4)\right) \;\sim\; \|\tilde{\beta}\|^4 \operatorname{tr}(\tilde{A}^2)\,(\operatorname{tr}(\Sigma^2))^2\,, \end{aligned} \tag{182}$$

there will be *three* contributions like this, corresponding to the three cases: $i_1 = i_2$, $i_1 = j_1$, and $i_1 = j_2$.

In the case when more than two out of i_1, i_2, j_1, j_2 are identical, they would all have to be identical. This contribution would be negligible because it would give

$$\|\tilde{\beta}\|^4 E[X^4]\operatorname{tr}(\tilde{A}^2)\,(\operatorname{tr}(\Sigma^4)) \;=\; O(PN^2)$$

which is negligible.

- We can now focus on the case $k_1 \neq k_2$ in (179). First, consider the terms with $k_1 = 1$. By symmetry, terms with $k_2 = 1$ give the same contribution. Since $k_2 \neq 1$ and

$\|\tilde{\beta}\|^2\lambda_1 = \tilde{\beta}'\tilde{A}\tilde{\beta}$, Newton's identities imply that

$$
\begin{aligned}
&\lambda_1 \frac{1}{N^2T^2}\|\tilde{\beta}\|^4 E[\sum_{i_2,j_2,k_2\neq 1}\sum_{i_1,j_1} X_{i_1,1}^2\lambda_{i_1}(\Sigma)X_{j_1,1}^2\lambda_{j_1}(\Sigma)X_{i_2,1}\lambda_{i_2}(\Sigma)X_{i_2,k_2}\lambda_{k_2}X_{j_2,k_2}\lambda_{j_2}(\Sigma)X_{j_2,1}] \\
&\sim \lambda_1 \frac{1}{N^2T^2}\|\tilde{\beta}\|^4 E[\sum_{i_2,k_2}\sum_{i_1,j_1} X_{i_1,1}^2 X_{j_1,1}^2\lambda_{i_1}(\Sigma)\lambda_{j_1}(\Sigma)X_{i_2,1}^2\lambda_{i_2}(\Sigma)^2X_{i_2,k_2}^2\lambda_{k_2}] \\
&\sim (\tilde{\beta}'\tilde{A}\tilde{\beta})\,\|\tilde{\beta}\|^2\,\frac{1}{N^2T^2}\,\mathrm{tr}(\tilde{A})\left(E[\sum_{i_2}\sum_{i_1,j_1} X_{i_1,1}^2 X_{j_1,1}^2\lambda_{i_1}(\Sigma)\lambda_{j_1}(\Sigma)X_{i_2,1}^2\lambda_{i_2}(\Sigma)^2]\right) \\
&=\ (\tilde{\beta}'\tilde{A}\tilde{\beta})\,\|\tilde{\beta}\|^2\,\frac{1}{N^2T^2}\,\mathrm{tr}(\tilde{A})\Bigg(\sum_{i_2,i_1,j_1\ all\ different}\lambda_{i_1}(\Sigma)\lambda_{j_1}(\Sigma)\lambda_{i_2}(\Sigma)^2 \\
&+\ \sum_{i_1=j_1\neq i_2} E[X^4]\lambda_{i_1}(\Sigma)^2\lambda_{i_2}(\Sigma)^2 \\
&+\ 2\sum_{i_1\neq j_1=i_2} E[X^4]\lambda_{i_1}(\Sigma)\lambda_{i_2}(\Sigma)^3 \\
&+\ E[X^6]\,\mathrm{tr}(\Sigma^4)\Bigg) \\
&=\ (\tilde{\beta}'\tilde{A}\tilde{\beta})\,\|\tilde{\beta}\|^2\,\frac{1}{N^2T^2}\,\mathrm{tr}(\tilde{A})\Bigg(\sum_{i_2}\lambda_{i_2}(\Sigma)^2((\mathrm{tr}(\Sigma)-\lambda_{i_2})^2-(\mathrm{tr}(\Sigma^2)-\lambda_{i_2}^2)) \\
&+\ E[X^4]((\mathrm{tr}(\Sigma^2))^2-\mathrm{tr}(\Sigma^4)) \\
&+\ 2E[X^4]\sum_{i_2}\lambda_{i_2}(\Sigma)^3(\mathrm{tr}(\Sigma)-\lambda_{i_2}) \\
&+\ E[X^6]\,\mathrm{tr}(\Sigma^4)\Bigg) \\
&=\ (\tilde{\beta}'\tilde{A}\tilde{\beta})\,\|\tilde{\beta}\|^2\,\frac{1}{N^2T^2}\,\mathrm{tr}(\tilde{A})\Bigg((\mathrm{tr}(\Sigma)^2)\,\mathrm{tr}(\Sigma^2)-2(\mathrm{tr}\,\Sigma)(\mathrm{tr}(\Sigma^3))+2\,\mathrm{tr}(\Sigma^4)-(\mathrm{tr}(\Sigma^2))^2 \\
&+\ E[X^4]((\mathrm{tr}(\Sigma^2))^2-\mathrm{tr}(\Sigma^4)) \\
&+\ 2E[X^4]((\mathrm{tr}\,\Sigma)(\mathrm{tr}(\Sigma^3))-\mathrm{tr}(\Sigma^4))+\ E[X^6]\,\mathrm{tr}(\Sigma^4)\Bigg)
\end{aligned}
\tag{183}
$$

because the rest terms are zero. And this term gets multiplied by 2 when we add the contribution of the $k_2 = 1$ case. As above, all these terms are

$$O(\|\lambda\|^4 (\mathrm{tr}(\Sigma))^4 \, \mathrm{tr}(\tilde{A})/(N^2 T^2)) \;=\; O(P/T^2)$$

and hence are negligible.

- Now, in the case when $k_1 \neq k_2$ and both are different from 1 in (179), we immediately get that (i_1, i_2, j_1, j_2) must either be all identical, or come in two identical pairs. The first case gives a contribution of

$$\|\tilde{\beta}\|^4 E[\sum_{i,k_1 \notin \{k_2,1\}} X_{i,1}^4 X_{i,k_1}^2 X_{i,k_2}^2 \lambda_i(\Sigma)^4 \lambda_{k_1} \lambda_{k_2}] \;\sim\; \|\tilde{\beta}\|^4 E[X^4] \, (\mathrm{tr}(\tilde{A})^2 - \mathrm{tr}(\tilde{A}^2)) \; \mathrm{tr}(\Sigma^4) \;=\; o(P^2 N^2)\,.$$

  The second one ought to have $i_1 = j_1, i_2 = j_2$ because $k_1 \neq k_2$ and both are not equal to 1, giving

$$\begin{aligned} & \|\tilde{\beta}\|^4 E[\sum_{i_2,k_2} \sum_{i_1,k_1} X_{i_1,1}^2 X_{i_1,k_1}^2 \lambda_{k_1} \lambda_{i_1}^2(\Sigma) \lambda_{i_2}^2(\Sigma) X_{i_2,1}^2 X_{i_2,k_2}^2 \lambda_{k_2}] \\ & \sim \; \|\tilde{\beta}\|^4 ((\mathrm{tr}\, \tilde{A})^2 - \mathrm{tr}(\tilde{A}^2)) \left(E[\sum_{i_2} \sum_{i_1} X_{i_1,1}^2 \lambda_{i_1}^2(\Sigma) \lambda_{i_2}^2(\Sigma) X_{i_2,1}^2] \right) \\ & = \; \|\tilde{\beta}\|^4 ((\mathrm{tr}\, \tilde{A})^2 - \mathrm{tr}(\tilde{A}^2))((\mathrm{tr}(\Sigma^2))^2 - \mathrm{tr}(\Sigma^4)) \\ & \sim \; \|\tilde{\beta}\|^4 ((\mathrm{tr}\, \tilde{A})^2 - \mathrm{tr}(\tilde{A}^2))(\mathrm{tr}(\Sigma^2))^2 \end{aligned} \tag{184}$$

Summarizing, the dominant terms are (182) (multiplied by 3) and (184), so that

$$
\begin{aligned}
& Term1 \\
\sim\ & 3\|\tilde\beta\|^4\,\mathrm{tr}(\tilde A^2)\,(\mathrm{tr}(\Sigma^2))^2\frac{1}{N^2T^2}\ +\ \|\tilde\beta\|^4 E[X^4]\,\mathrm{tr}(\tilde A^2)\,(\mathrm{tr}(\Sigma^4))\frac{1}{N^2T^2} \\
+\ & 2(\tilde\beta'\tilde A\tilde\beta)\,\|\tilde\beta\|^2\,\frac{1}{N^2T^2}\,\mathrm{tr}(\tilde A)\Bigg(\mathrm{tr}(\Sigma^2)(\mathrm{tr}(\Sigma))^2-2(\mathrm{tr}\,\Sigma)(\mathrm{tr}(\Sigma^3))+2\,\mathrm{tr}(\Sigma^4)-(\mathrm{tr}(\Sigma^2))^2 \\
+\ & E[X^4]((\mathrm{tr}(\Sigma^2))^2-\mathrm{tr}(\Sigma^4)) \\
+\ & 2E[X^4]((\mathrm{tr}\,\Sigma)(\mathrm{tr}(\Sigma^3))-\mathrm{tr}(\Sigma^4))+\ E[X^6]\,\mathrm{tr}(\Sigma^4)\Bigg)\frac{1}{N^2T^2} \\
+\ & \|\tilde\beta\|^4 E[X^4]\,(\mathrm{tr}(\tilde A)^2-\mathrm{tr}(\tilde A^2))\ \mathrm{tr}(\Sigma^4)\frac{1}{N^2T^2} \\
+\ & \|\tilde\beta\|^4((\mathrm{tr}\,\tilde A)^2-\mathrm{tr}(\tilde A^2))(\mathrm{tr}(\Sigma^2))^2\frac{1}{N^2T^2} \\
\sim\ & \|\tilde\beta\|^4((\mathrm{tr}\,\tilde A)^2+2\,\mathrm{tr}(\tilde A^2))(\mathrm{tr}(\Sigma^2))^2/(N^2T^2)\ \sim\ \|\tilde\beta\|^4(\mathrm{tr}\,\tilde A)^2(\mathrm{tr}(\Sigma^2))^2/(N^2T^2)
\end{aligned}
\tag{185}
$$

because $\mathrm{tr}(\tilde A^2)=O(P)$.

D.2 $Term2$ in (174)

We now proceed with the second term (note that *it comes with a factor of four*). We have

$$
E[\lambda' Z_tAZ_tAZ_t\lambda]\ =\ \|\tilde\beta\|^2E[\sum X_{i_1,1}\lambda_{i_1}(\Sigma)X_{i_1,k_1}\lambda_{k_1}X_{i_2,k_1}\lambda_{i_2}(\Sigma)X_{i_2,k_2}\lambda_{k_2}X_{i_3,k_2}\lambda_{i_3}(\Sigma)X_{i_3,1}]\,. \tag{186}
$$

- Suppose first that $k_1=k_2\neq 1$ in (186). The respective contribution is

$$
\|\tilde\beta\|^2E[\sum X_{i_1,1}\lambda_{i_1}(\Sigma)X_{i_1,k_1}\lambda_{k_1}X_{i_2,k_1}^2\lambda_{i_2}(\Sigma)\lambda_{k_1}X_{i_3,k_1}\lambda_{i_3}(\Sigma)X_{i_3,1}]\,, \tag{187}
$$

and hence $i_1 = i_3$ for non-zero terms, so that this contribution becomes

$$\begin{aligned}
&\|\tilde{\beta}\|^2 E[\sum X_{i_1,1}^2 \lambda_{i_1}(\Sigma)^2 X_{i_1,k_1}^2 \lambda_{k_1}^2 X_{i_2,k_1}^2 \lambda_{i_2}(\Sigma)] \\
&= \ \|\tilde{\beta}\|^2 \left(\sum_{i_1 \neq i_2, k_1 \neq 1} \lambda_{i_1}(\Sigma)^2 \lambda_{k_1}^2 \lambda_{i_2}(\Sigma) \ + \ E[X^4] \sum_{i_1, k_1 \neq 1} \lambda_{i_1}(\Sigma)^3 \lambda_{k_1}^2 \right) \\
&\sim \|\tilde{\beta}\|^2 \operatorname{tr}(\tilde{A}^2)((E[X^4]-1)\operatorname{tr}(\Sigma^3) + \operatorname{tr}(\Sigma)\operatorname{tr}(\Sigma^2)) \ = \ O(P(b_*(\operatorname{tr}\Sigma)^2)\operatorname{tr}\Sigma) \ = \ O(PN^2)
\end{aligned} \tag{188}$$

- The terms with $k_1 = k_2 = 1$ in (186) give

$$\begin{aligned}
&\lambda_1^2 \|\tilde{\beta}\|^2 E[\sum X_{i_1,1}^2 \lambda_{i_1}(\Sigma) X_{i_2,1}^2 \lambda_{i_2}(\Sigma) X_{i_3,1}^2 \lambda_{i_3}(\Sigma)] \\
&\sim \ \lambda_1^2 \|\tilde{\beta}\|^2 \Bigg(\sum_{i_1,i_2,i_3 \ pairwise \ different} \lambda_{i_1}(\Sigma)\lambda_{i_2}(\Sigma)\lambda_{i_3}(\Sigma) \\
&+ \ 3 \sum_{i_1,i_2 \ different} E[X^4]\lambda_{i_1}^2(\Sigma)\lambda_{i_2}(\Sigma) \ + \ E[X^6]\operatorname{tr}(\Sigma^3) \Bigg) \\
&= (\tilde{\beta}'\tilde{A}\tilde{\beta})^2 \|\tilde{\beta}\|^2 \Bigg((\operatorname{tr}\Sigma)^3 - 3(\operatorname{tr}\Sigma)\operatorname{tr}(\Sigma^2) \ + \ 2\operatorname{tr}(\Sigma^3) \\
&+ \ 3E[X^4]((\operatorname{tr}\Sigma)\operatorname{tr}(\Sigma^2) - \operatorname{tr}(\Sigma^3)) + E[X^6]\operatorname{tr}(\Sigma^3) \Bigg) \ = \ O(b_*(\operatorname{tr}\Sigma)^2 \operatorname{tr}\Sigma) \ = \ O(N^2)
\end{aligned} \tag{189}$$

by Newton's identities, where $3\sum_{i_1,i_2 \ different}$ appears because there are three possibilities for a coincidence of of pair among i_1, i_2, i_3, and where we have used that $\|\tilde{\beta}\|^2\lambda_1 = \tilde{\beta}'\tilde{A}\tilde{\beta}$.

- For the terms with $k_1 \neq k_2$ and none of them equal to 1 in in (186), we must have

$i_1 = i_2 = i_3$ for them to be non-zero, giving

$$\begin{aligned}
&\|\tilde\beta\|^2 E[\sum X_{i_1,1}^2 \lambda_{i_1}(\Sigma)^3 X_{i_1,k_1}^2 \lambda_{k_1} X_{i_1,k_2}^2 \lambda_{k_2}] \ \sim\ \|\tilde\beta\|^2((\operatorname{tr}(\tilde A))^2 - \operatorname{tr}(\tilde A^2)) \operatorname{tr}(\Sigma^3) \\
&=\ o(P^2N^2)
\end{aligned} \tag{190}$$

since $((\operatorname{tr}(\tilde A))^2 - \operatorname{tr}(\tilde A^2)) = O(P^2)$.

- If $k_1 \neq k_2 = 1$ in (186), then we get the contribution

$$\begin{aligned}
&\|\tilde\beta\|^2 E[\sum X_{i_1,1}\lambda_{i_1}(\Sigma) X_{i_1,k_1}\lambda_{k_1} X_{i_2,k_1}\lambda_{i_2}(\Sigma) X_{i_2,1}\lambda_1 \lambda_{i_3}(\Sigma) X_{i_3,1}^2] \\
&= \tilde\beta' \tilde A \tilde\beta E[\sum X_{i_1,1}\lambda_{i_1}(\Sigma) X_{i_1,k_1}\lambda_{k_1} X_{i_2,k_1}\lambda_{i_2}(\Sigma) X_{i_2,1} \lambda_{i_3}(\Sigma) X_{i_3,1}^2] \\
&=\ \{\textit{only terms with } i_1 = i_2 \textit{ survive}\} \\
&=\ \tilde\beta' \tilde A \tilde\beta E[\sum X_{i_1,1}^2 \lambda_{i_1}^2(\Sigma) X_{i_1,k_1}^2 \lambda_{k_1} \lambda_{i_3}(\Sigma) X_{i_3,1}^2] \\
&\sim\ \tilde\beta' \tilde A \tilde\beta\, (\operatorname{tr} \tilde A) \left(\operatorname{tr}(\Sigma)(\operatorname{tr}(\Sigma^2)) \ +\ (E[X^4]-1)\operatorname{tr}(\Sigma^3) \right) \ =\ O(Pb_*(\operatorname{tr}(\Sigma))^3) \ =\ O(PN^2)
\end{aligned} \tag{191}$$

and there is an identical contribution with $k_1 = 1 \neq k_2$.

Thus,

$$\begin{aligned}
&\frac{1}{4} Term2 \ \sim\ \|\tilde\beta\|^2 \operatorname{tr}(\tilde A^2)((E[X^4]-1)\operatorname{tr}(\Sigma^3) + \operatorname{tr}(\Sigma)\operatorname{tr}(\Sigma^2)) \\
&+ \|\tilde\beta\|^2((\operatorname{tr}(\tilde A))^2 - \operatorname{tr}(\tilde A^2)) \operatorname{tr}(\Sigma^3) \\
&+\ 2\tilde\beta' \tilde A \tilde\beta\, (\operatorname{tr} \tilde A) \left(\operatorname{tr}(\Sigma)(\operatorname{tr}(\Sigma^2)) \ +\ (E[X^4]-1)\operatorname{tr}(\Sigma^3) \right) \\
&\sim o(T^2N^2)\,.
\end{aligned} \tag{192}$$

D.3 $Term3$ in (174)

We now proceed with the third term. We have

$$
\begin{aligned}
& 2\frac{1}{N^2T^2}E[\mathrm{tr}(AZ_t)\,\lambda' Z_t A Z_t \lambda] \\
&= \; 2\|\tilde\beta\|^2\frac{1}{N^2T^2}E[\sum_k \lambda_k(\tilde A)\sum_i \lambda_i(\Sigma)X_{i,k}^2 \sum_{i_1,k_1,i_2} X_{i_1,1}\lambda_{i_1}(\Sigma)X_{i_1,k_1}\lambda_{k_1}(\tilde A)X_{i_2,k_1}\lambda_{i_2}(\Sigma)X_{i_2,1}]
\end{aligned}
\tag{193}
$$

- First consider the terms with $k_1 = 1$ in (193). This gives

$$
\begin{aligned}
& 2\|\tilde\beta\|^2\frac{1}{N^2T^2}E[\sum_k \lambda_k(\tilde A)\sum_i \lambda_i(\Sigma)X_{i,k}^2 \sum_{i_1,i_2} X_{i_1,1}^2\lambda_{i_1}(\Sigma)\lambda_1(\tilde A)\lambda_{i_2}(\Sigma)X_{i_2,1}^2] \\
&\sim \; 2\frac{1}{N^2T^2}(\tilde\beta'\tilde A\tilde\beta)\,(\mathrm{tr}\,\tilde A)\,(\mathrm{tr}\,\Sigma)\,E[\sum_{i_1,i_2} X_{i_1,1}^2\lambda_{i_1}(\Sigma)\lambda_{i_2}(\Sigma)X_{i_2,1}^2] \\
&= \; 2\frac{1}{N^2T^2}(\tilde\beta'\tilde A\tilde\beta)\,(\mathrm{tr}\,\tilde A)\,(\mathrm{tr}\,\Sigma)((\mathrm{tr}(\Sigma))^2 + (E[X^4]-1)\,\mathrm{tr}(\Sigma^2)) \\
&= \; O(Pb_*(\mathrm{tr}\,\Sigma)^3) \; = \; O(PN^2)
\end{aligned}
\tag{194}
$$

  where in the transition from the first to the second line we have used that λ_1 is a negligible fraction of $\mathrm{tr}\,\tilde A$.

- If $k_1 \neq 1$ in in (193), the only non-zero terms are with $i_1 = i_2$ and they give

$$\begin{aligned}
&2\|\tilde{\beta}\|^2 \frac{1}{N^2T^2} E[\sum_k \lambda_k(\tilde{A}) \sum_i \lambda_i(\Sigma) X_{i,k}^2 \sum_{i_1,k_1\neq 1} X_{i_1,1}^2 \lambda_{i_1}^2(\Sigma) X_{i_1,k_1}^2 \lambda_{k_1}(\tilde{A})] \\
&\sim\; 2\|\tilde{\beta}\|^2 \frac{1}{N^2T^2} E[\sum_{k\neq 1} \lambda_k(\tilde{A}) \sum_i \lambda_i(\Sigma) X_{i,k}^2 \sum_{i_1,k_1\neq 1} X_{i_1,1}^2 \lambda_{i_1}^2(\Sigma) X_{i_1,k_1}^2 \lambda_{k_1}(\tilde{A})] \\
&=\; 2\|\tilde{\beta}\|^2 \frac{1}{N^2T^2}\, E[\sum_{k\neq 1} \lambda_k(\tilde{A}) \sum_i \lambda_i(\Sigma) X_{i,k}^2 \sum_{i_1,k_1\neq 1} \lambda_{i_1}^2(\Sigma) X_{i_1,k_1}^2 \lambda_{k_1}(\tilde{A})] \\
&=\; 2\|\tilde{\beta}\|^2 \frac{1}{N^2T^2} \Bigg(E[\sum_{k\neq 1} \lambda_k^2(\tilde{A}) \sum_i \lambda_i(\Sigma) X_{i,k}^2 \sum_{i_1} \lambda_{i_1}^2(\Sigma) X_{i_1,k}^2] \\
&+\; E[\sum_{k\neq 1} \lambda_k(\tilde{A}) \sum_{i,k_1\neq 1,k} \lambda_i(\Sigma) X_{i,k}^2 \sum_{i_1} \lambda_{i_1}^2(\Sigma) X_{i_1,k_1}^2 \lambda_{k_1}(\tilde{A})] \Bigg) \\
&=\; 2\|\tilde{\beta}\|^2 \frac{1}{N^2T^2} \Bigg(E[X^4]\,\mathrm{tr}(\tilde{A}^2)\,\mathrm{tr}(\Sigma^3) + \sum_{k\neq 1} \lambda_k^2(\tilde{A}) \sum_i \lambda_i(\Sigma) \sum_{i_1\neq i} \lambda_{i_1}^2(\Sigma) \\
&+\; \sum_{k\neq 1} \lambda_k(\tilde{A}) \sum_{i,k_1\neq 1,k} \lambda_i(\Sigma) \sum_{i_1} \lambda_{i_1}^2(\Sigma) \lambda_{k_1}(\tilde{A}) \Bigg) \\
&\sim\; 2\|\tilde{\beta}\|^2 \frac{1}{N^2T^2} \Bigg(\mathrm{tr}(\tilde{A}^2)\Big((E[X^4]-1)\,\mathrm{tr}(\Sigma^3) + \mathrm{tr}(\Sigma)\,\mathrm{tr}(\Sigma^2)\Big) \\
&+\; ((\mathrm{tr}\,\tilde{A})^2 - \mathrm{tr}(\tilde{A}^2))\,\mathrm{tr}(\Sigma)\,\mathrm{tr}(\Sigma^2) \Bigg) \;\sim\; 2\|\tilde{\beta}\|^2 \frac{1}{N^2T^2} (\mathrm{tr}\,\tilde{A})^2\,\mathrm{tr}(\Sigma)\,\mathrm{tr}(\Sigma^2)\,.
\end{aligned} \tag{195}$$

Thus,

$$\begin{aligned}
&Term3 \sim\; 2\frac{1}{N^2T^2}(\tilde{\beta}'\tilde{A}\tilde{\beta})\,(\mathrm{tr}\,\tilde{A})\,(\mathrm{tr}\,\Sigma)((\mathrm{tr}(\Sigma))^2 \\
&+ (E[X^4]-1)\,\mathrm{tr}(\Sigma^2)) + 2\|\tilde{\beta}\|^2 \frac{1}{N^2T^2}(\mathrm{tr}\,\tilde{A})^2\,\mathrm{tr}(\Sigma)\,\mathrm{tr}(\Sigma^2) \\
&\sim\; 2\frac{1}{N^2T^2}(\tilde{\beta}'\tilde{A}\tilde{\beta})\,(\mathrm{tr}\,\tilde{A})\,(\mathrm{tr}\,\Sigma)^3 + 2\|\tilde{\beta}\|^2 \frac{1}{N^2T^2}(\mathrm{tr}\,\tilde{A})^2\,\mathrm{tr}(\Sigma)\,\mathrm{tr}(\Sigma^2) \\
&\sim\; 2\|\tilde{\beta}\|^2 \frac{1}{N^2T^2}(\mathrm{tr}\,\tilde{A})^2\,\mathrm{tr}(\Sigma)\,\mathrm{tr}(\Sigma^2)
\end{aligned} \tag{196}$$

D.4 $Term4$ and $Term5$ in (174)

We have

$$
\begin{aligned}
&E[(E[\varepsilon^4]-1)\operatorname{tr}(AZ_tAZ_t) \;+\; (\operatorname{tr}(AZ_t))^2] \\
&= \;(E[\varepsilon^4]-1)E[\sum \lambda_k(\tilde{A})X_{i,k}\lambda_i(\Sigma)X_{i,k_1}\lambda_{k_1}(\tilde{A})X_{i_1,k_1}\lambda_{i_1}(\Sigma)X_{i_1,k}] \\
&+ E[(\sum_k \lambda_k(\tilde{A})\sum_i \lambda_i(\Sigma)X_{i,k}^2)^2]
\end{aligned} \tag{197}
$$

We have

$$
\begin{aligned}
&E[(\sum_k \lambda_k(\tilde{A})\sum_i \lambda_i(\Sigma)X_{i,k}^2)^2] \\
&= \;E[\sum_{k,k_1,i,i_1} \lambda_k(\tilde{A})\lambda_{k_1}(\tilde{A})\lambda_{i_1}(\Sigma)X_{i_1,k_1}^2\lambda_{i_2}(\Sigma)X_{i_2,k_2}^2] \\
&= \;E[\sum_k \lambda_k^2(\tilde{A})\sum_{i_1,i_2}\lambda_{i_1}\lambda_{i_2}X_{i_1,k}^2X_{i_2,k}^2] \;+\; \sum_{k_1\neq k_2}\lambda_{k_1}(\tilde{A})\lambda_{k_2}(\tilde{A})(\operatorname{tr}(\Sigma))^2 \\
&\sim \;\operatorname{tr}(\tilde{A}^2)((E[X^4]-1)\operatorname{tr}(\Sigma^2)+(\operatorname{tr}\Sigma)^2) \;+\; ((\operatorname{tr}(\tilde{A}))^2-\operatorname{tr}(\tilde{A}^2))(\operatorname{tr}\Sigma)^2
\end{aligned} \tag{198}
$$

Similarly,

$$
\begin{aligned}
&(E[\varepsilon^4]-1)E[\sum \lambda_k(\tilde{A})X_{i,k}\lambda_i(\Sigma)X_{i,k_1}\lambda_{k_1}(\tilde{A})X_{i_1,k_1}\lambda_{i_1}(\Sigma)X_{i_1,k}] \\
&= \;(E[\varepsilon^4]-1)E[\sum_{k_1=k}\lambda_k(\tilde{A})^2X_{i,k}^2\lambda_i(\Sigma)\lambda_{i_1}(\Sigma)X_{i_1,k}^2] \\
&+ \;(E[\varepsilon^4]-1)E[\sum_{k\neq k_1}\sum_i \lambda_k(\tilde{A})X_{i,k}^2\lambda_i^2(\Sigma)X_{i,k_1}^2\lambda_{k_1}(\tilde{A})] \\
&\sim \;(E[\varepsilon^4]-1)\operatorname{tr}(\tilde{A}^2)((E[X^4]-1)\operatorname{tr}(\Sigma^2)+(\operatorname{tr}\Sigma)^2) \;+\; (E[\varepsilon^4]-1)((\operatorname{tr}(\tilde{A}))^2-\operatorname{tr}(\tilde{A}^2))\operatorname{tr}(\Sigma^2)
\end{aligned} \tag{199}
$$

Thus,

$$
\begin{aligned}
&Term4 + Term5 \;\sim\; \operatorname{tr}(\tilde{A}^2)((E[X^4]-1)\operatorname{tr}(\Sigma^2) + (\operatorname{tr}\Sigma)^2) \;+\; ((\operatorname{tr}(\tilde{A}))^2 - \operatorname{tr}(\tilde{A}^2))(\operatorname{tr}\Sigma)^2\\
&+\; (E[\varepsilon^4]-1)\operatorname{tr}(\tilde{A}^2)((E[X^4]-1)\operatorname{tr}(\Sigma^2) + (\operatorname{tr}\Sigma)^2) \;+\; (E[\varepsilon^4]-1)((\operatorname{tr}(\tilde{A}))^2 - \operatorname{tr}(\tilde{A}^2))\operatorname{tr}(\Sigma^2)\\
&\sim\; (\operatorname{tr}(\tilde{A}^2)(\operatorname{tr}\Sigma)^2 \;+\; ((\operatorname{tr}(\tilde{A}))^2 - \operatorname{tr}(\tilde{A}^2))(\operatorname{tr}\Sigma)^2)\frac{1}{N^2T^2}\\
&+ (E[\varepsilon^4]-1)\Big(\operatorname{tr}(\tilde{A}^2)(\operatorname{tr}\Sigma)^2 + ((\operatorname{tr}(\tilde{A}))^2 - \operatorname{tr}(\tilde{A}^2))\operatorname{tr}(\Sigma^2)\Big)\frac{1}{N^2T^2}\\
&=\; (\operatorname{tr}(\tilde{A}))^2(\operatorname{tr}\Sigma)^2\frac{1}{N^2T^2}\\
&+ (E[\varepsilon^4]-1)\Big(\operatorname{tr}(\tilde{A}^2)(\operatorname{tr}\Sigma)^2 + ((\operatorname{tr}(\tilde{A}))^2 - \operatorname{tr}(\tilde{A}^2))\operatorname{tr}(\Sigma^2)\Big)\frac{1}{N^2T^2}\\
&\sim (\operatorname{tr}(\tilde{A}))^2(\operatorname{tr}\Sigma)^2/(N^2T^2)
\end{aligned}
\tag{200}
$$

because $\operatorname{tr}(\Sigma^2)/(\operatorname{tr}(\Sigma))^2 \to 0$.

D.5 Equating the terms

By (172),

$$
\begin{aligned}
&(\frac{1}{NT}\operatorname{tr} E[A_P F_t F_t'])^2 \;\sim\; \frac{1}{T^2N^2}\operatorname{tr}(\tilde{A})^2(\operatorname{tr}\Sigma + \|\tilde{\beta}\|^2\ \operatorname{tr}(\Sigma^2))^2\\
&=\; \frac{1}{T^2N^2}\operatorname{tr}(\tilde{A})^2\Big((\operatorname{tr}\Sigma)^2 + 2\|\tilde{\beta}\|^2(\operatorname{tr}\Sigma)\ \operatorname{tr}(\Sigma^2) + \|\tilde{\beta}\|^4\,(\operatorname{tr}(\Sigma^2))^2\Big)
\end{aligned}
\tag{201}
$$

and the claim follows from (185), (192), (196), and (200).

The proof of Lemma 12 is complete. □

E Proof of Theorem 9

Proof of Theorem 9. The first claim follows because, by Lemma 10, the other contributions do not impact eigenvalue distribution.

To prove the claim about the eigenvalue distribution of B_T, we use a remarkable Theorem of (Bai and Zhou, 2008). According to (Bai and Zhou, 2008), defining $Z_t = N^{-1/2}F_t = S_t'R_{t+1}$, we need to verify the following technical conditions:

(1) $E[Z_tZ_t'] = A_P$ for some matrix A_P

(2) $E[(Z_t'BZ_t - \mathrm{tr}(A_PB_P))^2] \;=\; o(T^2)$ for any bounded matrix sequence $B_P,\ P > 0$.

(3) The norm of A_P is uniformly bounded, and its eigenvalue distribution converges as $P \to \infty$.

The only non-trivial claim here is item (3), which in turn follows from Lemma 12. The proof of Theorem 9 is complete. □

F Technical Lemmas for Computing Higher Moments

The following lemma is a direct consequence of (174) and the polarization identity

$$ab \;=\; 0.25((a+b)^2 - (a-b)^2)\,.$$

Lemma 15 *Let $Z_t = S_{t-1}'S_{t-1}$. Recall also that, according to Assumption 6,*

$$R_{t+1} \;=\; S_t\beta \;+\; \varepsilon_{t+1} \tag{202}$$

where, for brevity, we omit the time index for $\beta = \beta_{t+1}$. Thus,

$$F_t \;=\; Z_t\beta \;+\; S_{t-1}'\varepsilon_t\,. \tag{203}$$

For any two matrices A, B with A being symmetric, we have

$$
\begin{aligned}
&\frac{1}{N^2T}E[F_t'AF_tF_t'BF_t]\\
&= \frac{1}{N^2T}\operatorname{tr}E[Z_t\beta\beta'Z_tAZ_t\beta\beta'Z_tB]\\
&+ \frac{1}{N^2T}2\operatorname{tr}(E[\beta'Z_tAZ_tBZ_t\beta]+E[\beta'Z_tBZ_tAZ_t\beta])\\
&+\frac{1}{N^2T}\operatorname{tr}(E[(\beta'Z_tAZ_t\beta)Z_tB]+E[(\beta'Z_tBZ_t\beta)Z_tA])\\
&+\frac{1}{N^2T}((\kappa_\varepsilon-1)\operatorname{tr}E[Z_tAZ_tB]+ E[\operatorname{tr}(Z_tA)\operatorname{tr}(Z_tB)])\\
&= Term1 + Term2 + Term3 + Term4 + Term5\,.
\end{aligned}
\tag{204}
$$

Proof. When A, B are symmetric, (174) implies

$$
\begin{aligned}
&\frac{1}{N^2T}E[F_t'AF_tF_t'BF_t]\\
&= \frac{1}{N^2T}\operatorname{tr}E[Z_t\beta\beta'Z_tAZ_t\beta\beta'Z_tB]\\
&+ \frac{1}{N^2T}2\operatorname{tr}(E[Z_t\beta\beta'Z_tAZ_tB]+E[Z_t\beta\beta'Z_tBZ_tA])\\
&+\frac{1}{N^2T}\operatorname{tr}(E[(\beta'Z_tAZ_t\beta)Z_tB]+E[(\beta'Z_tBZ_t\beta)Z_tA])\\
&+\frac{1}{N^2T}((\kappa_\varepsilon-1)\operatorname{tr}E[Z_tAZ_tB] + E[\operatorname{tr}(Z_tA)\operatorname{tr}(Z_tB)])
\end{aligned}
\tag{205}
$$

The general case follows because

$$
\begin{aligned}
&\frac{1}{N^2T}E[F_t'AF_tF_t'BF_t] \;=\; \frac{1}{N^2T}E[F_t'0.5(A+A')F_tF_t'0.5(B+B')F_t] \\
&=\; \frac{1}{N^2T}\operatorname{tr}E[Z_t\beta\beta'Z_t0.5(A+A')Z_t\beta\beta'Z_t0.5(B+B')] \\
&+\; \frac{1}{N^2T}2\operatorname{tr}(E[Z_t\beta\beta'Z_t0.5(A+A')Z_t0.5(B+B')]+E[Z_t\beta\beta'Z_t0.5(B+B')Z_t0.5(A+A')]) \\
&+\frac{1}{N^2T}\operatorname{tr}(E[(\beta'Z_t0.5(A+A')Z_t\beta)Z_t0.5(B+B')]+E[(\beta'Z_t0.5(B+B')Z_t\beta)Z_t0.5(A+A')]) \\
&+\frac{1}{N^2T}((\kappa_\varepsilon-1)\operatorname{tr}E[Z_t0.5(A+A')Z_t0.5(B+B')] \;+\; E[\operatorname{tr}(Z_t0.5(A+A'))\operatorname{tr}(Z_t0.5(B+B'))]) \\
&=\; \frac{1}{N^2T}\operatorname{tr}E[Z_t\beta\beta'Z_tAZ_t\beta\beta'Z_tB] \\
&+\; \frac{1}{N^2T}\operatorname{tr}(E[\beta'Z_tAZ_tBZ_t\beta]+E[\beta'Z_tBZ_tAZ_t\beta]+E[\beta'Z_tA'Z_tBZ_t\beta]+E[\beta'Z_tAZ_tB'Z_t\beta]) \\
&+\frac{1}{N^2T}\operatorname{tr}(E[(\beta'Z_tAZ_t\beta)Z_tB]+E[(\beta'Z_tBZ_t\beta)Z_tA]) \\
&+\frac{1}{N^2T}((\kappa_\varepsilon-1)0.5\operatorname{tr}(E[Z_tAZ_tB]+E[Z_tA'Z_tB])+\; E[\operatorname{tr}(Z_tA)\operatorname{tr}(Z_tB)])
\end{aligned}
\tag{206}
$$

□

Lemma 16 *For any two matrices A, B, we have*

$$
\begin{aligned}
&\frac{1}{N^2T}\operatorname{tr} E[Z_t\beta\beta' Z_t A Z_t\beta\beta' Z_t B] \\
&\sim \ \Big((\tilde\beta'\tilde A\tilde\beta)\operatorname{tr}(\tilde B) \ + \ (\tilde\beta'\tilde B\tilde\beta)\operatorname{tr}(\tilde A)\Big)\,\|\tilde\beta\|^2 \ \operatorname{tr}(\Sigma^2)(\operatorname{tr}(\Sigma))^2\frac{1}{N^2T} \\
&+ \ \|\tilde\beta\|^4((\operatorname{tr}\tilde A)(\operatorname{tr}\tilde B) + 2\operatorname{tr}(\tilde A\tilde B))(\operatorname{tr}(\Sigma^2))^2\frac{1}{N^2T} \\
&+ \|\tilde\beta\|^4 E[X^4] \ \operatorname{tr}(\tilde A)\operatorname{tr}(\tilde B)\operatorname{tr}(\Sigma^4)\frac{1}{N^2T} \\
&\frac{1}{N^2T}2\operatorname{tr}(E[Z_t\beta\beta' Z_t A Z_t B] + E[Z_t\beta\beta' Z_t B Z_t A]) \\
&\sim \ \frac{1}{N^2T}4\|\tilde\beta\|^2\operatorname{tr}(\tilde A\tilde B)\operatorname{tr}(\Sigma)\operatorname{tr}(\Sigma^2) \\
&+ \frac{1}{N^2T}4\|\tilde\beta\|^2(\operatorname{tr}(\tilde A)\operatorname{tr}(\tilde B) - \operatorname{tr}(\tilde A\tilde B))\operatorname{tr}(\Sigma^3) \\
&+ \ \frac{1}{N^2T}4\Big(\tilde\beta'\tilde A\tilde\beta\,(\operatorname{tr}\tilde B) + \tilde\beta'\tilde B\tilde\beta\,(\operatorname{tr}\tilde A)\Big)\operatorname{tr}(\Sigma)(\operatorname{tr}(\Sigma^2)) \\
&\frac{1}{N^2T}\operatorname{tr}(E[(\beta' Z_t A Z_t\beta)Z_t B] + E[(\beta' Z_t B Z_t\beta)Z_t A]) \\
&\sim \ \frac{1}{N^2T}\Big(\tilde\beta'\tilde A\tilde\beta\,(\operatorname{tr}\tilde B) + \tilde\beta'\tilde B\tilde\beta\,(\operatorname{tr}\tilde A)\Big)(\operatorname{tr}\Sigma)^3 + 2\|\tilde\beta\|^2\frac{1}{N^2T}(\operatorname{tr}\tilde A)(\operatorname{tr}\tilde B)\operatorname{tr}(\Sigma)\operatorname{tr}(\Sigma^2) \\
&\frac{1}{N^2T}((\kappa_\varepsilon - 1)\operatorname{tr} E[Z_t A Z_t B] \ + \ E[\operatorname{tr}(Z_t A)\operatorname{tr}(Z_t B)]) \\
&\sim \ \Big((\operatorname{tr}\tilde A)(\operatorname{tr}\tilde B) \ + \ (E[\varepsilon^4]-1)\operatorname{tr}(\tilde A\tilde B)\Big)(\operatorname{tr}\Sigma)^2\frac{1}{N^2T}
\end{aligned}
\tag{207}
$$

with $\tilde A = \Psi^{1/2}A\Psi^{1/2}$ and $\tilde B = \Psi^{1/2}B\Psi^{1/2}$.

Proof of Lemma 16. Using (185), (192), (196), and (200) , we get the following result:

$$
\begin{aligned}
&\frac{1}{N^2T}\operatorname{tr} E[Z_t\beta\beta' Z_t A Z_t \beta\beta' Z_t B] \;\sim\; 3\|\tilde\beta\|^4 \operatorname{tr}(\tilde A\tilde B)\,(\operatorname{tr}(\Sigma^2))^2\frac{1}{N^2T} \;+\; \|\tilde\beta\|^4 E[X^4]\operatorname{tr}(\tilde A\tilde B)\,(\operatorname{tr}(\Sigma^4))\frac{1}{N^2T}\\
&+\;\Big((\tilde\beta'\tilde A\tilde\beta)\operatorname{tr}(\tilde B)\;+\;(\tilde\beta'\tilde B\tilde\beta)\operatorname{tr}(\tilde A)\Big)\|\tilde\beta\|^2\Bigg(\operatorname{tr}(\Sigma^2)(\operatorname{tr}(\Sigma))^2 - 2(\operatorname{tr}\Sigma)(\operatorname{tr}(\Sigma^3)) + 2\operatorname{tr}(\Sigma^4) - (\operatorname{tr}(\Sigma^2))^2\\
&+\; E[X^4]((\operatorname{tr}(\Sigma^2))^2 - \operatorname{tr}(\Sigma^4))\\
&+\; 2E[X^4]((\operatorname{tr}\Sigma)(\operatorname{tr}(\Sigma^3)) - \operatorname{tr}(\Sigma^4)) + \; E[X^6]\operatorname{tr}(\Sigma^4)\Bigg)\frac{1}{N^2T}\\
&+\; \|\tilde\beta\|^4 E[X^4]\,(\operatorname{tr}(\tilde A)\operatorname{tr}(\tilde B) - \operatorname{tr}(\tilde A\tilde B))\;\operatorname{tr}(\Sigma^4)\frac{1}{N^2T}\\
&+\; \|\tilde\beta\|^4((\operatorname{tr}\tilde A)\operatorname{tr}(\tilde B) - \operatorname{tr}(\tilde A\tilde B))(\operatorname{tr}(\Sigma^2))^2\frac{1}{N^2T}\\
&\frac{1}{N^2T}2\operatorname{tr}(E[Z_t\beta\beta' Z_t A Z_t B] + E[Z_t\beta\beta' Z_t B Z_t A])\\
&\sim\; \frac{1}{N^2T}4\|\tilde\beta\|^2\operatorname{tr}(\tilde A\tilde B)((E[X^4]-1)\operatorname{tr}(\Sigma^3) + \operatorname{tr}(\Sigma)\operatorname{tr}(\Sigma^2))\\
&+\frac{1}{N^2T}4\|\tilde\beta\|^2(\operatorname{tr}(\tilde A)\operatorname{tr}(\tilde B) - \operatorname{tr}(\tilde A\tilde B))\operatorname{tr}(\Sigma^3)\\
&+\; \frac{1}{N^2T}4\Big(\tilde\beta'\tilde A\tilde\beta\,(\operatorname{tr}\tilde B) + \tilde\beta'\tilde B\tilde\beta\,(\operatorname{tr}\tilde A)\Big)\Bigg(\operatorname{tr}(\Sigma)(\operatorname{tr}(\Sigma^2)) \;+\; (E[X^4]-1)\operatorname{tr}(\Sigma^3)\Bigg)\\
&\frac{1}{N^2T}\operatorname{tr}(E[(\beta' Z_t A Z_t\beta)Z_t B] + E[(\beta' Z_t B Z_t \beta) Z_t A])\\
&\sim\; \frac{1}{N^2T}\Big(\tilde\beta'\tilde A\tilde\beta\,(\operatorname{tr}\tilde B) + \tilde\beta'\tilde B\tilde\beta\,(\operatorname{tr}\tilde A)\Big)(\operatorname{tr}\Sigma)^3 + 2\|\tilde\beta\|^2\frac{1}{N^2T^2}(\operatorname{tr}\tilde A)(\operatorname{tr}\tilde B)\operatorname{tr}(\Sigma)\operatorname{tr}(\Sigma^2)\\
&\frac{1}{N^2T}((\kappa_\varepsilon - 1)\operatorname{tr} E[Z_t A Z_t B] \;+\; E[\operatorname{tr}(Z_t A)\operatorname{tr}(Z_t B)])\\
&\sim\; \Big((\operatorname{tr}\tilde A)(\operatorname{tr}\tilde B) \;+\; (E[\varepsilon^4]-1)\operatorname{tr}(\tilde A\tilde B)\Big)(\operatorname{tr}\Sigma)^2\frac{1}{N^2T}\\
&\;+ (E[\varepsilon^4]-1)\Big((\operatorname{tr}\tilde A)(\operatorname{tr}\tilde B) \;-\; \operatorname{tr}(\tilde A\tilde B)\Big)\operatorname{tr}(\Sigma^2)\frac{1}{N^2T^2}
\end{aligned}
\tag{208}
$$

where we have used that

$$\begin{aligned}
&\Big(\operatorname{tr}(\Sigma^2)(\operatorname{tr}(\Sigma))^2 - 2(\operatorname{tr}\Sigma)(\operatorname{tr}(\Sigma^3)) + 2\operatorname{tr}(\Sigma^4) - (\operatorname{tr}(\Sigma^2))^2 \\
&+\ E[X^4]((\operatorname{tr}(\Sigma^2))^2 - \operatorname{tr}(\Sigma^4)) \\
&+\ 2E[X^4]((\operatorname{tr}\Sigma)(\operatorname{tr}(\Sigma^3)) - \operatorname{tr}(\Sigma^4)) +\ E[X^6]\operatorname{tr}(\Sigma^4)\Big) \ \sim\ \operatorname{tr}(\Sigma^2)(\operatorname{tr}(\Sigma))^2
\end{aligned} \tag{209}$$

□

Lemma 17 *Define $\psi_{*,1}$ through the equation*

$$b_*\psi_{*,1} \ = \ N^{-1}\operatorname{tr}((\Sigma_F^*\Psi) + \lambda'\Psi\lambda))\,. \tag{210}$$

Then, we have

$$\frac{1}{TN^2}\operatorname{tr} E[\beta\beta' F_{t_1}F_{t_1}' F_{t_1}F_{t_1}' Q] \ \sim\ \frac{1}{TN^2}\operatorname{tr}(\Psi)\,(\operatorname{tr}(\Sigma))^2(b_*\operatorname{tr}\Sigma\psi_{*,1} + 1)E[\beta'\Psi Q\beta]$$

for any uniformly bounded Q that is independent of F.

Proof of Lemma 17. We have

$$\frac{1}{TN^2}\operatorname{tr} E[\beta\beta' F_{t_1}F_{t_1}' F_{t_1}F_{t_1}' Q] \ = \ \frac{1}{TN^2}\operatorname{tr} E[F_{t_1}'F_{t_1}F_{t_1}'Q\beta\beta' F_{t_1}] \tag{211}$$

and hence we are in a position to apply Lemmas 15 and 16 with the two matrices given by $A = I$ and $B = \Psi^{1/2}Q\beta\beta'\Psi^{1/2}$ so that $\tilde{A} = \Psi$ and $\tilde{B} = \Psi^{1/2}Q\beta\beta'\Psi^{1/2}$. Thus, (211) is the sum of the following terms:

$$
\begin{aligned}
&\frac{1}{N^2T}\operatorname{tr} E[Z_t\beta\beta' Z_t A Z_t \beta\beta' Z_t B] \\
&\sim \Big((\tilde\beta'\Psi\tilde\beta)\operatorname{tr}(\Psi^{1/2}Q\beta\beta'\Psi^{1/2}) \;+\; (\tilde\beta'\Psi^{1/2}Q\beta\beta'\Psi^{1/2}\tilde\beta)\operatorname{tr}(\Psi)\Big)\,\|\tilde\beta\|^2\;\operatorname{tr}(\Sigma^2)(\operatorname{tr}(\Sigma))^2\frac{1}{N^2T} \\
&+\;\|\tilde\beta\|^4((\operatorname{tr}\Psi)(\operatorname{tr}\Psi^{1/2}Q\beta\beta'\Psi^{1/2}) + 2\operatorname{tr}(\Psi\Psi^{1/2}Q\beta\beta'\Psi^{1/2}))(\operatorname{tr}(\Sigma^2))^2\frac{1}{N^2T} \\
&\frac{1}{N^2T}2\operatorname{tr}(E[Z_t\beta\beta' Z_t A Z_t B] + E[Z_t\beta\beta' Z_t B Z_t A]) \\
&\sim\;\frac{1}{N^2T}4\|\tilde\beta\|^2\operatorname{tr}(\Psi\Psi^{1/2}Q\beta\beta'\Psi^{1/2})\operatorname{tr}(\Sigma)\operatorname{tr}(\Sigma^2) \\
&+\frac{1}{N^2T}4\|\tilde\beta\|^2(\operatorname{tr}(\Psi)\operatorname{tr}(\Psi^{1/2}Q\beta\beta'\Psi^{1/2}) - \operatorname{tr}(\Psi\Psi^{1/2}Q\beta\beta'\Psi^{1/2}))\operatorname{tr}(\Sigma^3) \\
&+\;\frac{1}{N^2T}4\Big(\tilde\beta'\Psi\tilde\beta\,(\operatorname{tr}\Psi^{1/2}Q\beta\beta'\Psi^{1/2}) + \tilde\beta'\Psi^{1/2}Q\beta\beta'\Psi^{1/2}\tilde\beta\,(\operatorname{tr}\Psi)\Big)\operatorname{tr}(\Sigma)(\operatorname{tr}(\Sigma^2)) \\
&\frac{1}{N^2T}\operatorname{tr}(E[(\beta' Z_t A Z_t\beta)Z_t B] + E[(\beta' Z_t B Z_t \beta) Z_t A]) \\
&\sim\;\frac{1}{N^2T}\Big(\tilde\beta'\Psi\tilde\beta\,(\operatorname{tr}\Psi^{1/2}Q\beta\beta'\Psi^{1/2}) + \tilde\beta'\Psi^{1/2}Q\beta\beta'\Psi^{1/2}\tilde\beta\,(\operatorname{tr}\Psi)\Big)(\operatorname{tr}\Sigma)^3 \\
&+2\|\tilde\beta\|^2\frac{1}{N^2T}(\operatorname{tr}\Psi)(\operatorname{tr}\Psi^{1/2}Q\beta\beta'\Psi^{1/2})\operatorname{tr}(\Sigma)\operatorname{tr}(\Sigma^2) \\
&\frac{1}{N^2T}((E[\varepsilon^4]-1)\operatorname{tr}E[Z_tAZ_tB]\;\;+\;\;E[\operatorname{tr}(Z_tA)\operatorname{tr}(Z_tB)]) \\
&\sim\;\Big((\operatorname{tr}\Psi)(\operatorname{tr}\Psi^{1/2}Q\beta\beta'\Psi^{1/2})\;\;+\;\;(E[\varepsilon^4]-1)\operatorname{tr}(\Psi\Psi^{1/2}Q\beta\beta'\Psi^{1/2})\Big)(\operatorname{tr}\Sigma)^2\frac{1}{N^2T}
\end{aligned}
\tag{212}
$$

Now, $\operatorname{tr}(\beta\beta' D)$ is uniformly bounded almost surely for any bounded D. In addition, Assumption 4 implies that $\operatorname{tr}(\Sigma^2) = o(\operatorname{tr}(\Sigma)^2)$ and $\operatorname{tr}(\Sigma^3) = o(\operatorname{tr}(\Sigma)\ \operatorname{tr}(\Sigma^2))$. As a result, many terms become negligible and we get

$$
\begin{aligned}
&\frac{1}{N^2T}\operatorname{tr} E[Z_t\beta\beta' Z_t A Z_t \beta\beta' Z_t B] \\
&\sim\ (\tilde\beta'\Psi^{1/2}Q\beta\beta'\Psi^{1/2}\tilde\beta)\operatorname{tr}(\Psi)\,\|\tilde\beta\|^2\ \operatorname{tr}(\Sigma^2)(\operatorname{tr}(\Sigma))^2\frac{1}{N^2T} \\
&\frac{1}{N^2T}2\operatorname{tr}(E[Z_t\beta\beta' Z_t A Z_t B] + E[Z_t\beta\beta' Z_t B Z_t A]) \\
&\sim\ \frac{1}{N^2T}4\tilde\beta'\Psi^{1/2}Q\beta\beta'\Psi^{1/2}\tilde\beta\,(\operatorname{tr}\Psi)\operatorname{tr}(\Sigma)(\operatorname{tr}(\Sigma^2)) \\
&\frac{1}{N^2T}\operatorname{tr}(E[(\beta' Z_t A Z_t\beta) Z_t B] + E[(\beta' Z_t B Z_t \beta) Z_t A]) \\
&\sim\ \frac{1}{N^2T}\tilde\beta'\Psi^{1/2}Q\beta\beta'\Psi^{1/2}\tilde\beta\,(\operatorname{tr}\Psi)(\operatorname{tr}\Sigma)^3 \\
&\frac{1}{N^2T}((\kappa_\varepsilon-1)\operatorname{tr}E[Z_tAZ_tB]\ +\ E[\operatorname{tr}(Z_tA)\operatorname{tr}(Z_tB)]) \\
&\sim\ (\operatorname{tr}\Psi)(\operatorname{tr}\Psi^{1/2}Q\beta\beta'\Psi^{1/2})(\operatorname{tr}\Sigma)^2\frac{1}{N^2T}
\end{aligned}
\tag{213}
$$

Recall that $b_* = \operatorname{tr} E[\beta\beta'] = \operatorname{tr}((\Sigma_F^*\Psi)+\lambda'\Psi\lambda))$. The first term is of the order $b_*^3 M \operatorname{tr}(\Sigma)\operatorname{tr}(\Sigma^2)$. The second term is of the order $b_*^2 M\operatorname{tr}(\Sigma)\operatorname{tr}(\Sigma^2)$. The third term is of the order of $b_*^2M(\operatorname{tr}\Sigma)^3$ and hence it dominates the second term as well as the first term because $\operatorname{tr}(\Sigma^2) = o((\operatorname{tr}(\Sigma))^2)$. Thus, we are left with

$$
\begin{aligned}
&\frac{1}{N^2T}\tilde\beta'\Psi^{1/2}Q\beta\beta'\Psi^{1/2}\tilde\beta\,(\operatorname{tr}\Psi)(\operatorname{tr}\Sigma)^3 + (\operatorname{tr}\Psi)(\operatorname{tr}\Psi^{1/2}Q\beta\beta'\Psi^{1/2})(\operatorname{tr}\Sigma)^2\frac{1}{N^2T} \\
&\sim \frac{1}{TN^2}b_*\psi_{*,1}\operatorname{tr}(\Psi)\,(\operatorname{tr}(\Sigma))^3E[\beta'\Psi Q\beta]\ +\ (\operatorname{tr}\Psi)E[\beta'\Psi Q\beta](\operatorname{tr}\Sigma)^2\frac{1}{N^2T}
\end{aligned}
\tag{214}
$$

where we have used that, by Lemma 6, $\beta'\Psi^{1/2}\tilde\beta \approx N^{-1}\operatorname{tr}((\Sigma_F^*\Psi)+\lambda'\Psi\lambda))$ The proof of Lemma 17 is complete. ☐

G The Martingale Lemma and $\xi(z;c)$

We start with the following Lemma from KMZ.

Lemma 18 *We have*

$$P^{-1}\operatorname{tr}(A_1(zI+B_T)^{-1}A_2) \;-\; P^{-1}\operatorname{tr}E[A_1(zI+B_T)^{-1}A_2] \;\to\; 0$$

almost surely for any bounded $A_1,\ A_2$ that are independent of F_t.

The next lemma is a non-trivial modification of Lemma 18.

Lemma 19 *Let $\hat{F}_t = F_t -$*

Proof of Lemma 19. The proof follows by the same arguments as in (Bai and Zhou, 2008). Our first key observation is that $\|\lambda\|^2$ is almost surely bounded as $P\to\infty$.

Let $B_{T,t}=\frac{1}{T}\sum_{\tau\neq t}F_\tau F_\tau'$. By the Sherman-Morrison formula (see (Bartlett, 1951)),

$$(zI+B_T)^{-1} = (zI+B_{T,t})^{-1} - \frac{1}{NT}(zI+B_{T,t})^{-1}F_tF_t'(zI+B_{T,t})^{-1}\frac{1}{1+(NT)^{-1}F_t'(zI+B_{T,t})^{-1}F_t} \tag{215}$$

Let E_τ denote the conditional expectation given $F_{\tau+1},\cdots,F_T$. Let also

$$q_T(z)=\lambda' A_1(zI+B_T)^{-1}A_2\lambda$$

With this notation, since $B_{T,t}$ is independent of F_t, we have

$$(E_{t-1}-E_t)[\lambda' A_1(zI+B_{T,t})^{-1}A_2\lambda|\lambda]=0\,,$$

(below we omit the conditioning on λ for the sake of simplicity) and therefore

$$
\begin{aligned}
E[q_T(z)] - q_T(z) &= E_0[q_T(z)] - E_T[q_T(z)] \\
&= \sum_{t=1}^{T}(E_{t-1}[q_T(z)] - E_t[q_T(z)]) \\
&= \sum_{t=1}^{T}(E_{t-1} - E_t)[q_T(z)] \\
&= \sum_{t=1}^{T}(E_{t-1} - E_t)[q_T(z) - \lambda' A_1 (zI + B_{T,t})^{-1} A_2 \lambda] \\
&= \sum_{t=1}^{T}(E_{t-1} - E_t)[\lambda' A_1 (zI + B_T)^{-1} A_2 \lambda - \lambda' A_1 (zI + B_{T,t})^{-1} A_2 \lambda] \\
&= -\sum_{\tau=1}^{T}(E_{t-1} - E_t)[\gamma_t]\,,
\end{aligned}
\tag{216}
$$

where we have used (215) and defined

$$
\gamma_t = \lambda' A_1 \left(\frac{1}{NT}(zI + B_{T,t})^{-1} F_t (I + \frac{1}{NT} F_t' (zI + B_{T,t})^{-1} F_t)^{-1} F_t' (zI + B_{T,t})^{-1} A_2 \lambda \right). \tag{217}
$$

Now, we substitute

$$
F_t = \Psi \lambda + \hat{F}_t \tag{218}
$$

where $\hat{F}_t$ satisfies the RMT conditions.

Almost sure convergence follows with $q > 2$ from the following lemma.

Lemma 20 *Suppose that*

$$
E[|X_T|^q] \leq T^{-\alpha}
$$

for some $\alpha > 1$ and some $q > 0$. Then, $X_T \to 0$ almost surely.

Proof. It is known that if

$$\sum_{T=1}^{\infty} Prob(|X_T| > \varepsilon) \; < \; \infty$$

for any $\varepsilon > 0$, then $X_T \to 0$ almost surely. In our case, the Chebyshev inequality implies that

$$Prob(|X_T| > \varepsilon) \; \leq \; \varepsilon^{-q} E[|X_T|^q] \; \leq \; T^{-\alpha}$$

and convergence follows because $\alpha > 1$. □

The proof of Lemma 19 is complete □

Lemma 21 *Let*

$$\frac{1}{T}\operatorname{tr}((zI + B_T)^{-1}\Psi\sigma_*) \; \to \; \xi(z;c) \tag{219}$$

almost surely and

$$\frac{1}{NT} F_t'(zI + B_{T,t})^{-1} F_t \; \to \; \xi(z;c)\,, \tag{220}$$

in probability, where

$$\frac{c^{-1}\xi(z;c)}{1+\xi(z;c)} \; = \; 1 \; - \; m(-z;c)z \tag{221}$$

Proof. First, Lemma 12 implies that

$$\frac{1}{NT} F_t'(zI + B_{T,t})^{-1} F_t \; - \; \frac{1}{T}\operatorname{tr}((zI + B_{T,t})^{-1}\frac{1}{N}E[F_tF_t']) \; \to \; 0\,.$$

in probability. Next Lemma 18 applied to our setting implies that for any bounded matrix

Q_T independent of $B_{T,t}$ we have

$$\frac{1}{T}\operatorname{tr}((zI+B_{T,t})^{-1}Q_T) \; - \; \frac{1}{T}E[\operatorname{tr}((zI+B_{T,t})^{-1}Q_T)] \; \to \; 0$$

almost surely. At the same time, by Lemma 8,

$$\begin{aligned}\frac{1}{N}E[F_tF_t'] \; &= \; ((\operatorname{tr}\Sigma/N)^2N+\operatorname{tr}(\Sigma^2/N))\Psi N^{-1}\Sigma_F\Psi \\ + \; \operatorname{tr}(\Sigma^2/N)(\kappa-2)\Psi^{1/2}&\operatorname{diag}(\Psi^{1/2}N^{-1}\Sigma_F\Psi^{1/2})\Psi^{1/2}+\Psi\Big(\operatorname{tr}(\Sigma\Sigma_\varepsilon/N)+\operatorname{tr}(\Psi N^{-1}\Sigma_F)\operatorname{tr}(\Sigma^2/N)\Big)\end{aligned} \tag{222}$$

We have

$$\frac{1}{T}\operatorname{tr}((zI+B_{T,t})^{-1}(\operatorname{tr}\Sigma/N)^2\Psi\Sigma_F^*\Psi) \; = \; O(1/T) \tag{223}$$

The same argument applies to the second term because the trace of

$$\operatorname{tr}(\Sigma^2/N)(\kappa-2)\Psi^{1/2}\operatorname{diag}(\Psi^{1/2}N^{-1}\Sigma_F\Psi^{1/2})\Psi^{1/2}$$

is also uniformly bounded. Thus, we get

$$\begin{aligned}&\frac{1}{NT}F_t'(zI+B_{T,t})^{-1}F_t \; \sim \; \frac{1}{T}\operatorname{tr}((zI+B_{T,t})^{-1}\frac{1}{N}E[F_tF_t']) \\ &\sim \; T^{-1}\operatorname{tr}[(zI+B_{T,t})^{-1}\Psi\sigma_*] \; \to \; \xi(z;c)\,.\end{aligned} \tag{224}$$

Now, we have

$$\begin{aligned}1 \; &= \; P^{-1}\operatorname{tr}E[(zI+B_T)^{-1}(zI+B_T)] \\ &= \; zm(-z;c) \; + \; \frac{1}{P}\operatorname{tr}\frac{1}{T}\sum_t\frac{1}{N}E[(zI+B_T)^{-1}F_tF_t'] \\ &= \; zm(-z;c)+\frac{1}{P}\operatorname{tr}\frac{1}{N}E[(zI+B_T)^{-1}F_tF_t']\end{aligned} \tag{225}$$

where we have used symmetry across t in the last step. Using the Sherman-Morrison formula, we get

$$\frac{1}{NT}\,\mathrm{tr}\,E[(zI+B_T)^{-1}F_t'F_t] \;=\; E[\frac{\frac{1}{NT}F_t'(zI+B_{T,t})^{-1}F_t}{1+\frac{1}{NT}F_t'(zI+B_{T,t})^{-1}F_t}]\,,$$

where

$$B_{T,t} \;=\; \frac{1}{NT}\sum_{\tau\neq t}F_\tau F_\tau'.$$

Furthermore, since all functions involved are uniformly bounded, a standard argument implies that we can replace

$$\frac{1}{NT}F_t'(zI+B_{T,t})^{-1}F_t$$

with

$$\xi(z;c)$$

by (224).[22] □

H Expected Return on the Feasible Portfolio

Proposition 12 *We have*

$$E[R^F_{t+1}(z)] \;=\; \frac{\Gamma_{1,1}(z)}{1+\xi(z;c)}\,, \tag{226}$$

[22]Indeed, $E[\frac{Y_T}{1+Y_T}-\frac{Z_T}{1+Z_T}] = \frac{Y_T-Z_T}{(1+Y_T)(1+Z_T)}$ for any random variables $Y_T,\ Z_T$. If $Y_T, Z_T \geq 0$ then $\frac{|Y_T-Z_T|}{(1+Y_T)(1+Z_T)} \leq 1$ and hence convergence $Y_T-Z_T\to 0$ in probability implies convergence of expectations.

where

$$\Gamma_{1,1}(z) \;=\; \lim_{T,P\to\infty} \lambda' E[\Psi(zI+B_T)^{-1}\Psi]\lambda\,. \tag{227}$$

Proof of Proposition 12. We start by computing

$$E[F_{t+1}] \;=\; E[S_t' R_{t+1}] = E[S_t'(S_t\widetilde{F}_{t+1}+\varepsilon_{t+1})] \;=\; N^{-1/2}\,\mathrm{tr}(\Sigma)\Psi\lambda \tag{228}$$

and therefore, by (111), we have

$$\begin{aligned} &E[R^F_{t+1}(z)] \;=\; E[\hat\beta(z)' F_{t+1}] \\ &=\; \mathrm{tr}(\Sigma) E[\frac{1}{NT}\sum_t F_t'(zI+B_T)^{-1}]\Psi\lambda \;\sim\; E[\frac{1}{T}\sum_t F_t'(zI+B_T)^{-1}]\Psi\lambda N^{-1/2}\,, \end{aligned} \tag{229}$$

where we have used the normalization $N^{-1}\,\mathrm{tr}\,\Sigma = 1$. Now, by the interchangeability of F_t across t and the Sherman-Morrison formula, we have

$$\begin{aligned} &N^{-1/2}E[\frac{1}{T}\sum_t F_t'(zI+B_T)^{-1}]\Psi\lambda \\ &=\; N^{-1/2}E[F_t'(zI+B_T)^{-1}\Psi]\lambda \;=\; N^{-1/2}E[F_t'(zI+B_{T,t})^{-1}\frac{1}{1+(NT)^{-1}F_t'(zI+B_{T,t})^{-1}F_t}\Psi]\lambda\,, \end{aligned} \tag{230}$$

where

$$B_{T,t} \;=\; \frac{1}{NT}\sum_{\tau\neq t} F_\tau F_\tau'.$$

By Lemma 21,

$$(NT)^{-1}F_t'(zI+B_{T,t})^{-1}F_t \;\to\; \xi(z;c)$$

is probability and therefore

$$N^{-1/2}E[F_t'(zI+B_{T,t})^{-1}\frac{1}{1+(NT)^{-1}F_t'(zI+B_{T,t})^{-1}F_t}\Psi]\lambda \;\sim\; N^{-1/2}\frac{E[F_t'(zI+B_{T,t})^{-1}\Psi\lambda]}{1+\xi(z;c)}\,, \tag{231}$$

whereas $E[F_t'] = \mathrm{tr}(\Sigma\Sigma_\varepsilon)\Psi\lambda N^{-1/2}$ implies

$$N^{-1/2}E[F_t'(zI+B_{T,t})^{-1}\Psi\lambda] \;=\; N^{-1}\,\mathrm{tr}(\Sigma)\lambda' E[\Psi(zI+B_{T,t})^{-1}\Psi\lambda] \;\sim\; \Gamma_{1,1}(z)\,. \tag{232}$$

The proof of Proposition 12 is complete.

□

I Computing the Quasi-Moments

Lemma 22 *Let*

$$\psi_{*,k} \;=\; \lim P^{-1}\,\mathrm{tr}(\Psi^k\Sigma_\lambda) \tag{233}$$

and

$$\Gamma_{k,l,T}(z) \;\equiv\; \lambda' E[\Psi^k(zI+B_T)^{-1}\Psi^\ell]\lambda\,. \tag{234}$$

We have

$$\psi_{*,k+\ell} \sim \; z\,\Gamma_{k,\ell,T}(z) \;+\; \Bigg(\psi_{*,k+1}\Gamma_{1,\ell,T}(z) \;+\; \sigma_*\Gamma_{k+1,\ell,T}\Bigg)(1+\xi(z;c))^{-1} \tag{235}$$

Proof of Lemma 22. Using the Sherman-Morrison formula and Lemma 21, we get

$$F_t'(zI+B_T)^{-1} \;=\; F_t'(zI+B_{T,t})^{-1}(1+(NT)^{-1}F_t'(zI+B_{T,t})^{-1}F_t)^{-1} \sim\; F_t'(zI+B_{T,t})^{-1}(1+\xi(z;c))^{-1}$$

We also have

$$
\begin{aligned}
\frac{1}{N}E[F_tF_t'] \;&=\; ((\operatorname{tr}\Sigma/N)^2+\operatorname{tr}(\Sigma^2/N^2))\Psi\Sigma_F\Psi \\
+\;& \operatorname{tr}(\Sigma^2/N^2)(\kappa-2)\Psi^{1/2}\operatorname{diag}(\Psi^{1/2}\Sigma_F\Psi^{1/2})\Psi^{1/2}+\Psi\Big(\operatorname{tr}(\Sigma\Sigma_\varepsilon/N)+\operatorname{tr}(\Psi\Sigma_F N^{-1})\operatorname{tr}(\Sigma^2/N)\Big) \\
=\;& \widehat{\Sigma}_F \;+\Psi\Sigma_F\Psi \;+\; \sigma_*\Psi\,,
\end{aligned}
\tag{236}
$$

where $\|\widehat{\Sigma}_F\|=o(1)$, and

$$
\Sigma_F \;=\; \lambda\lambda' \;+\; \Sigma_F^*\,. \tag{237}
$$

We will need the following important observation:

Lemma 23 *For any sequence*

$$
\lambda' A_P Q_P \lambda \;\to\; 0 \tag{238}
$$

in probability, for any uniformly bounded Q_P (even if they correlate with λ) and any A_P with a uniformly bounded trace norm, such that A_P is independent of λ.

Proof of Lemma 23. We have

$$
\begin{aligned}
\lambda' A_P Q_P\lambda \;&=\; \operatorname{tr}(\lambda\lambda' A_P Q_P) \\
\le\;& \|\lambda\lambda' A_P Q_P\|_1 \;\le\; \|Q_P\|_\infty \|\lambda\lambda' A_P\|_1 \\
=\;& \|Q_P\|_\infty \operatorname{tr}((\lambda\lambda' A_P A_P'\lambda\lambda')^{1/2}) \;=\; \|Q_P\|_\infty(\lambda' A_P A_P'\lambda)^{1/2}\operatorname{tr}((\lambda\lambda')^{1/2}) \;=\; (\lambda' A_P A_P'\lambda)^{1/2}\|\lambda\| \\
=\;& (\operatorname{tr}(A_P A_P'\lambda\lambda'))^{1/2}\|\lambda\| \;\to\; (P^{-1}\operatorname{tr}(\Sigma_\lambda))^{1/2}(P^{-1}\operatorname{tr}(A_P A_P'\Sigma_\lambda))^{1/2} \\
\le\;& (P^{-1}\operatorname{tr}(\Sigma_\lambda))^{1/2}\|\Sigma_\lambda\|^{1/2}(P^{-1}\operatorname{tr}(A_P A_P'))^{1/2} \;\to\; 0
\end{aligned}
\tag{239}
$$

The proof of Lemma 23 is complete. $\square$

Thus, for any A_P with bounded trace norm, we get

$$
\begin{aligned}
&\psi_{*,k+\ell} \;=\; P^{-1}\operatorname{tr}(\Psi^{k+\ell}\Sigma_\lambda) \;\approx\; \lambda'\Psi^{k+\ell}\lambda \;=\; \lambda' E[\Psi^k (zI+B_T)(zI+B_T)^{-1}\Psi^\ell]\lambda \\
&=\; z\Gamma_{k,\ell,T}(z) \;+\; \lambda' E[\Psi^k B_T (zI+B_T)^{-1}\Psi^\ell]\lambda \\
&\underbrace{=}_{symmetry\ over\ t} \quad z\Gamma_{k,\ell,T}(z) \;+\; \frac{1}{N}\lambda' E[\Psi^k F_t F_t'(zI+B_T)^{-1}\Psi^\ell]\lambda \\
&\underbrace{=}_{(124)} \; z\Gamma_{k,\ell,T}(z) \;+\; \frac{1}{N}\lambda' E[\Psi^k F_t F_t'(zI+B_{T,t})^{-1}(1+(NT)^{-1}F_t'(zI+B_{T,t})^{-1}F_t)^{-1}\Psi^\ell]\lambda \\
&\underbrace{\sim}_{Lemma\ 21} \; z\Gamma_{k,\ell,T}(z) \;+\; \frac{1}{N}\lambda' E[\Psi^k F_t F_t'(zI+B_{T,t})^{-1}\Psi^\ell]\lambda(1+\xi(z;c))^{-1} \\
&\underbrace{\sim}_{(236)} \; z\Gamma_{k,\ell,T}(z) \;+\; \lambda' E[\Psi^k (\widehat{\Sigma}_F \;+\Psi\Sigma_F\Psi \;+\; \sigma_*\Psi)(zI+B_{T,t})^{-1}\Psi^\ell]\lambda(1+\xi(z;c))^{-1} \qquad (240) \\
&\sim\; z\Gamma_{k,\ell,T}(z) \;+\; \lambda' E[\Psi^k (\Psi(\Sigma_F+\lambda\lambda')\Psi \;+\; \sigma_*\Psi)(zI+B_T)^{-1}\Psi^\ell]\lambda(1+\xi(z;c))^{-1} \\
&\underbrace{\sim}_{(238)} \; z\Gamma_{k,\ell,T}(z) \;+\; \lambda' E[\Psi^k (\Psi\lambda\lambda'\Psi \;+\; \sigma_*\Psi)(zI+B_T)^{-1}\Psi^\ell]\lambda(1+\xi(z;c))^{-1} \\
&=\; z\Gamma_{k,\ell,T}(z) \;+\; \lambda'\Psi^{k+1}\lambda E[\lambda'\Psi(zI+B_T)^{-1}\Psi^\ell]\lambda(1+\xi(z;c))^{-1} \\
&+\; \lambda'\Psi^{k+1}\sigma_*(zI+B_T)^{-1}\Psi^\ell\lambda(1+\xi(z;c))^{-1} \\
&\sim\; z\,\Gamma_{k,\ell,T}(z) \;+\; \left(\psi_{*,k+1}\Gamma_{1,\ell,T}(z) \;+\; \sigma_*\Gamma_{k+1,\ell,T}\right)(1+\xi(z;c))^{-1}
\end{aligned}
$$

$\square$

Lemma 24 *Let*

$$
\delta(z) \;=\; -\sigma_* z^{-1}(1+\xi(z;c))^{-1}\,. \qquad (241)
$$

Then,

$$
\Gamma_{1,l}(z) \;=\; \frac{z^{-1}P^{-1}\operatorname{tr}(\Psi^{1+\ell}(I-\Psi\delta(z))^{-1}\Sigma_\lambda)}{1-\delta(z)P^{-1}\operatorname{tr}(\Psi^2(I-\Psi\delta(z))^{-1}\Sigma_\lambda)} \qquad (242)
$$

and

$$\Gamma_{k,\ell} \;=\; z^{-1}P^{-1}\,\mathrm{tr}(\Psi^{k+\ell}(I-\Psi\delta(z))^{-1}\Sigma_\lambda)-z^{-1}P^{-1}\,\mathrm{tr}(\Psi^{k+1}(I-\Psi\delta(z))^{-1}\Sigma_\lambda)\Gamma_{1,\ell}(1+\xi(z;c))^{-1} \tag{243}$$

Proof. We have

$$\Gamma_{k,\ell} \;=\; a_{k+1} + \delta\,\Gamma_{k+1,\ell} \tag{244}$$

where

$$a_{k+1,\ell} \;=\; z^{-1}(\psi_{*,k+\ell} - \psi_{*,k+1}\Gamma_{1,\ell}(1+\xi(z;c))^{-1}),\;\; \delta(z) \;=\; -\sigma_* z^{-1}(1+\xi(z;c))^{-1}\,. \tag{245}$$

Let us pick $z > \max(1, \|\Psi\|)$ sufficiently large, so that $\sigma_* z^{-1}(1+\xi(z;c))^{-1} < 1$ and[23]

$$|\delta^k \Gamma_{k,\ell}(z)| \;\le\; z^{-k+1}\|\lambda\|^2\|\Psi\|^{k+\ell} \;\to_{k\to\infty}\; 0\,. \tag{246}$$

Then, since iterating forward, we get

$$\Gamma_{k,\ell} \;=\; \sum_{\tau=0}^{\infty} a_{k+\tau+1,\ell}\delta^\tau\,. \tag{247}$$

Now,

$$a_{k+\tau+1,\ell} \;=\; z^{-1}(\psi_{*,k+\tau+\ell} - \psi_{*,k+\tau+1}\Gamma_{1,\ell}(1+\xi(z;c))^{-1}),\;\; \delta(z) \;=\; -\sigma_* z^{-1}(1+\xi(z;c))^{-1}\,. \tag{248}$$

[23]This uniform exponential decay also implies that the infinite sum of the limits equals the limit of the infinite sum, as we pass to the $P \to \infty$ limit.

$$
\begin{aligned}
\Gamma_{1,\ell} &= \sum_{\tau=0}^{\infty} a_{\tau+2,\ell}\delta^{\tau} \\
&= \sum_{\tau=0}^{\infty} z^{-1}(\psi_{*,1+\tau+\ell} - \psi_{*,1+\tau+1}\Gamma_{1,\ell}(1+\xi(z;c))^{-1})\delta^{\tau} \\
&= \sum_{\tau=0}^{\infty}(z^{-1}(P^{-1}\operatorname{tr}(\Psi^{\tau+\ell+1}\Sigma_{\lambda}) - P^{-1}\operatorname{tr}(\Psi^{\tau+2}\Sigma_{\lambda})\Gamma_{1,\ell}(1+\xi(z;c))^{-1}))\delta^{\tau} \\
&= z^{-1}P^{-1}\operatorname{tr}(\Psi^{1+\ell}(I-\Psi\delta(z))^{-1}\Sigma_{\lambda}) - z^{-1}P^{-1}\operatorname{tr}(\Psi^{2}(I-\Psi\delta(z))^{-1}\Sigma_{\lambda})\Gamma_{1,\ell}(1+\xi(z;c))^{-1}\,,
\end{aligned} \tag{249}
$$

implying that

$$
\Gamma_{1,l} = \frac{z^{-1}P^{-1}\operatorname{tr}(\Psi^{1+\ell}(I-\Psi\delta(z))^{-1}\Sigma_{\lambda})}{1-\delta(z)P^{-1}\operatorname{tr}(\Psi^{2}(I-\Psi\delta(z))^{-1}\Sigma_{\lambda})} \tag{250}
$$

Then, the same argument implies

$$
\Gamma_{k,\ell} = z^{-1}P^{-1}\operatorname{tr}(\Psi^{k+\ell}(I-\Psi\delta(z))^{-1}\Sigma_{\lambda}) - z^{-1}P^{-1}\operatorname{tr}(\Psi^{k+1}(I-\Psi\delta(z))^{-1}\Sigma_{\lambda})\Gamma_{1,\ell}(1+\xi(z;c))^{-1} \tag{251}
$$

$\square$

Proof of Lemma 36. We are using the representation of returns where the missing factors are absorbed into ε. As we show in (**??**), it does not affect σ_*, and the cross-terms are negligible by the $\operatorname{tr}(\Psi_{1,2}\Psi_{2,1}) = o(P)$ condition, and therefore all calculations stay the same:

$$
\begin{aligned}
&\psi_{*,k+\ell}(q) \;=\; P^{-1}\,\mathrm{tr}(\Psi_{1,1}^{k+\ell}\Sigma_\lambda^{(1)})) \;\approx\; (\lambda^{(1)})'\Psi_{1,1}^{k+\ell}(\lambda^{(1)}) \;=\; (\lambda^{(1)})'E[\Psi_{1,1}^{k}(zI+B_T^{(1)})(zI+B_T^{(1)})^{-1}\Psi_{1,1}^{\ell}](\lambda^{(1)})\\
&=\; z\Gamma_{k,\ell,T}(z) + \; (\lambda^{(1)})'E[\Psi_{1,1}^{k}B_T^{(1)}(zI+B_T^{(1)})^{-1}\Psi_{1,1}^{\ell}](\lambda^{(1)})\\
&\underbrace{=}_{symmetry\ over\ t} \quad z\Gamma_{k,\ell,T}(z)\\
&+\; \frac{1}{N}(\lambda^{(1)})'E[\Psi_{1,1}^{k}F_tF_t'(zI+B_T^{(1)})^{-1}\Psi_{1,1}^{\ell}](\lambda^{(1)})\\
&\underbrace{=}_{(124)} z\Gamma_{k,\ell,T}(z)\\
&+\; \frac{1}{N}(\lambda^{(1)})'E[\Psi_{1,1}^{k}F_tF_t'(zI+B_{T,t})^{-1}(1+(NT)^{-1}F_t'(zI+B_{T,t})^{-1}F_t)^{-1}\Psi_{1,1}^{\ell}](\lambda^{(1)})\\
&\underbrace{\sim}_{Lemma\ 21} \quad z\Gamma_{k,\ell,T}(z) \;+\; \frac{1}{N}(\lambda^{(1)})'E[\Psi_{1,1}^{k}F_tF_t'(zI+B_{T,t})^{-1}\Psi_{1,1}^{\ell}](\lambda^{(1)})(1+\xi(z;cq))^{-1}\\
&\underbrace{\sim}_{(236)} z\Gamma_{k,\ell,T}(z) \;+\; (\lambda^{(1)})'E[\Psi_{1,1}^{k}(\widehat{\Sigma}_F^{(1)} \;+\Psi_{1,1}\Sigma_F^{(1)}\Psi_{1,1} \;+\; \sigma_*\Psi_{1,1})(zI+B_{T,t})^{-1}\Psi_{1,1}^{\ell}](\lambda^{(1)})(1+\xi(z;cq))^{-1}\\
&\sim\; z\Gamma_{k,\ell,T}(z) \;+\; (\lambda^{(1)})'E[\Psi_{1,1}^{k}(\Psi_{1,1}(\Sigma_F^{(1)}+(\lambda^{(1)})(\lambda^{(1)})')\Psi_{1,1} \;+\; \sigma_*\Psi_{1,1})(zI+B_T)^{-1}\Psi_{1,1}^{\ell}](\lambda^{(1)})(1+\xi(z;cq))^{-1}\\
&\underbrace{\sim}_{(238)} z\Gamma_{k,\ell,T}(z) \;+\; (\lambda^{(1)})'\Psi_{1,1}^{k+1}(\lambda^{(1)})E[(\lambda^{(1)})'\Psi_{1,1}(zI+B_T)^{-1}\Psi_{1,1}^{\ell}](\lambda^{(1)})(1+\xi(z;cq))^{-1}\\
&+(\lambda^{(1)})'E[\Psi_{1,1}^{k}\sigma_*\Psi_{1,1}(zI+B_T)^{-1}\Psi_{1,1}^{\ell}](\lambda^{(1)})(1+\xi(z;cq))^{-1}\\
&\sim\; z\,\Gamma_{k,\ell,T}(z) \;+\; \Bigg(\psi_{*,k+1}(q)\Gamma_{1,\ell,T}(z) \;+\; \sigma_*\Gamma_{k+1,\ell,T}\Bigg)(1+\xi(z;cq))^{-1}
\end{aligned}
\tag{252}
$$

and the claim follows by the same argument as in the correctly specified case. $\square$

J Proof of Theorem 3: Second Moment of the Feasible Efficient Portfolio

Let

$$\overline{F_t} \;=\; \sum_t F_t\,.$$

Without loss of generality, we assume that $\kappa = 2$ because all kurtosis terms vanish asymptotically due to their vanishing trace norm. Using Assumption ?? and Lemma 8, we get[24]

$$\begin{aligned}
&E[(R^F_{t+1}(z))^2] \;=\; E[\frac{1}{NT}\overline{F_t}'(zI+B_T)^{-1}F_{t+1}F_{t+1}'(zI+B_T)^{-1}\frac{1}{NT}\overline{F_t}]\\
&=\; E[\frac{1}{NT}\overline{F_t}'(zI+B_T)^{-1}E_{t-}[F_{t+1}F_{t+1}'](zI+B_T)^{-1}\frac{1}{NT}\overline{F_t}]\\
&\underbrace{=}_{Lemma\ 8}\; E[\frac{1}{NT}\overline{F_t}'(zI+B_T)^{-1}\Bigg(((\operatorname{tr}\Sigma)^2+\operatorname{tr}(\Sigma^2))\Psi N^{-1}\Sigma_F\Psi+\Psi\Big(\operatorname{tr}(\Sigma\Sigma_\varepsilon)+\operatorname{tr}(\Psi N^{-1}\Sigma_F)\operatorname{tr}(\Sigma^2)\Big)\Bigg)\\
&(zI+B_T)^{-1}\frac{1}{NT}\overline{F_t}]\\
&\approx\; E[\frac{1}{NT}\overline{F_t}'(zI+B_T)^{-1}\Bigg((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)\Bigg)\\
&(zI+B_T)^{-1}\frac{1}{NT}\overline{F_t}]\\
&=\; \frac{1}{N^2T^2}\sum_{t_1,t_2}E[F_{t_1}(zI+B_T)^{-1}\Bigg((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)(zI+B_T)^{-1}F_{t_2}]\\
&\sim\; Term1 \;+\; Term2
\end{aligned} \tag{253}$$

[24] E_{t-} denotes the expectation averaging over realizations of S_t and R_{t+1}.

with

$$Term1 \;=\; \frac{1}{N^2T}E[F_{t_1}'(zI+B_T)^{-1}\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)(zI+B_T)^{-1}F_{t_1}] \tag{254}$$

and

$$Term2 \;=\; \frac{1}{N^2}\frac{T(T-1)}{T^2}E[F_{t_1}'(zI+B_T)^{-1}\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)(zI+B_T)^{-1}F_{t_2}] \tag{255}$$

for any $t_1 \neq t_2$.

J.1 $Term1$ in (254)

We first deal with the first term. Using the Sherman-Morrison formula and Lemma 21, and Lemma 8, we get

$$\begin{aligned}
Term1 \;&=\; \frac{1}{N^2T}\operatorname{tr}E[\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)(zI+B_T)^{-1}F_{t_1}F_{t_1}'(zI+B_T)^{-1}]\\
&\sim\; \frac{1}{N^2T}\operatorname{tr}E[\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)(zI+B_{T,t_1})^{-1}F_{t_1}F_{t_1}'(zI+B_{T,t_1})^{-1}](1+\xi(z;c))^{-2}\\
&\sim \frac{1}{N^2T}\operatorname{tr}E[\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)(zI+B_{T,t})^{-1}\\
&\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)(zI+B_{T,t_1})^{-1}](1+\xi(z;c))^{-2}
\end{aligned} \tag{256}$$

We can now split this expression into several terms. We have

$$\begin{aligned}
&\frac{1}{N^2T}\operatorname{tr}E[(\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi(zI+B_{T,t})^{-1}(\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi(zI+B_{T,t})^{-1}](1+\xi(z;c))^{-2}\\
&=\; \frac{1}{T}\operatorname{tr}E[\Psi\Sigma_F\Psi(zI+B_{T,t})^{-1}\Psi\Sigma_F\Psi(zI+B_{T,t})^{-1}](1+\xi(z;c))^{-2} \;\to 0
\end{aligned} \tag{257}$$

because

$$\operatorname{tr}(\Sigma_F) \;=\; \operatorname{tr}(\Sigma_F^*) \;+\; P^{-1}\|\lambda\|^2 \;=\; o(P) \;+\; O(1) = o(T)\,,$$

and all other matrices involved are uniformly bounded. The second term is

$$\frac{1}{N^2T}\operatorname{tr} E[(\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi(zI+B_{T,t})^{-1}\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Psi(zI+B_{T,t})^{-1}]/(1+\xi(z;c))^2 \;=\; O(T^{-1}) \tag{258}$$

by the same argument. Finally, the last term is

$$\frac{1}{N^2}(\operatorname{tr}(\Sigma\Sigma_\varepsilon))^2\frac{1}{T}\operatorname{tr} E[\Psi(zI+B_{T,t})^{-1}\Psi(zI+B_{T,t})^{-1}]/(1+\xi(z;c))^2 \tag{259}$$

and it needs to be evaluated directly.

Lemma 25 *We have*

$$\begin{aligned}
&\frac{1}{PN^2}\operatorname{tr} E[F_{t_1}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}F_{t_2}F_{t_2}'(zI+B_{T,t_1,t_2})^{-1}]\\
&\sim\; \sigma_*^2\frac{1}{P}\operatorname{tr} E[\Psi(zI+B_T)^{-1}\Psi(zI+B_T)^{-1}]\\
&\to\; \Gamma_3(z) \;=\; \left(1-(-z^2m'(-z;c)+2zm(-z;c)+c^{-1}\left(\frac{\xi(z;c)}{1+\xi(z;c)}\right)^2)\right)(1+\xi(z;c))^4
\end{aligned} \tag{260}$$

Proof. We have by the Sherman-Morrison formula that

$$\begin{aligned}
&\frac{1}{P}\frac{1}{N^2T}\operatorname{tr} E[F_{t_1}F_{t_1}'(zI+B_T)^{-1}F_{t_1}F_{t_1}'(zI+B_T)^{-1}]\\
&\sim \frac{1}{c}\frac{1}{N^2T^2}E[F_{t_1}'(zI+B_T)^{-1}F_{t_1}F_{t_1}'(zI+B_T)^{-1}F_{t_1}]\\
&=\; c^{-1}E\left[\left(\frac{\frac{1}{NT}F_{t_1}'(zI+B_{T,t_1})^{-1}F_{t_1}}{1+\frac{1}{NT}F_{t_1}'(zI+B_{T,t_1})^{-1}F_{t_1}}\right)^2\right]\\
&\sim\; c^{-1}\left(\frac{\xi(z;c)}{1+\xi(z;c)}\right)^2
\end{aligned} \tag{261}$$

by Lemma 21. Now,

$$m'(-z;c) \; = \; \lim P^{-1} \operatorname{tr} E[(zI + B_T)^{-2}] \tag{262}$$

and hence

$$\begin{aligned}
1 \; &= \; \frac{1}{P} \operatorname{tr} E[(zI + B_T)(zI + B_T)^{-1}(zI + B_T)(zI + B_T)^{-1}] \\
&= \; \frac{1}{P} z^2 \operatorname{tr} E[(zI + B_T)^{-2}] \; + \; 2z \, \frac{1}{P} \operatorname{tr} E[(zI + B_T)^{-2} B_T] \\
&+ \; \frac{1}{P} \operatorname{tr} E[B_T(zI + B_T)^{-1} B_T (zI + B_T)^{-1}] \\
&\sim \; z^2 m'(-z;c) \; + \; 2z \, \frac{1}{P} \operatorname{tr} E[(zI + B_T)^{-2}(B_T + zI - zI)] \\
&+ \; \frac{1}{P} \frac{1}{N^2 T^2} \sum_{t_1, t_2} \operatorname{tr} E[F_{t_1} F_{t_1}' (zI + B_T)^{-1} F_{t_2} F_{t_2}' (zI + B_T)^{-1}] \\
&= \; -z^2 m'(-z;c) + 2z m(-z;c) + \frac{1}{P} \frac{1}{N^2 T} \operatorname{tr} E[F_{t_1} F_{t_1}' (zI + B_T)^{-1} F_{t_1} F_{t_1}' (zI + B_T)^{-1}] \\
&+ \; \frac{1}{P} \frac{1}{N^2} \frac{T(T-1)}{T^2} \operatorname{tr} E[F_{t_1} F_{t_1}' (zI + B_T)^{-1} F_{t_2} F_{t_2}' (zI + B_T)^{-1}] \\
&\sim \; -z^2 m'(-z;c) + 2z m(-z;c) + c^{-1} \left(\frac{\xi(z;c)}{1 + \xi(z;c)} \right)^2 \qquad (263) \\
&+ \; \frac{1}{P} \frac{1}{N^2} \operatorname{tr} E[F_{t_1} F_{t_1}' (zI + B_T)^{-1} F_{t_2} F_{t_2}' (zI + B_T)^{-1}] \\
&\sim \; -z^2 m'(-z;c) + 2z m(-z;c) + c^{-1} \left(\frac{\xi(z;c)}{1 + \xi(z;c)} \right)^2 \\
&+ \; \frac{1}{P} \frac{1}{N^2} \operatorname{tr} E[F_{t_1} F_{t_1}' (zI + B_{T,t_1})^{-1} F_{t_2} F_{t_2}' (zI + B_{T,t_2})^{-1}] / (1 + \xi(z;c))^2 \\
&\sim \; -z^2 m'(-z;c) + 2z m(-z;c) + c^{-1} \left(\frac{\xi(z;c)}{1 + \xi(z;c)} \right)^2 \\
&+ \; \frac{1}{P} \frac{1}{N^2} E[F_{t_1}' (zI + B_{T,t_1,t_2})^{-1} F_{t_2} F_{t_2}' (zI + B_{T,t_1,t_2})^{-1} F_{t_1}] / (1 + \xi(z;c))^4 \\
&= \; -z^2 m'(-z;c) + 2z m(-z;c) + c^{-1} \left(\frac{\xi(z;c)}{1 + \xi(z;c)} \right)^2 \\
&+ \; \frac{1}{P} \frac{1}{N^2} \operatorname{tr} E[F_{t_1} F_{t_1}' (zI + B_{T,t_1,t_2})^{-1} F_{t_2} F_{t_2}' (zI + B_{T,t_1,t_2})^{-1}] / (1 + \xi(z;c))^4
\end{aligned}$$

where we have defined

$$B_{T,t_1,t_2} \;=\; \frac{1}{NT}\sum_{\tau\notin\{t_1,t_2\}} F_\tau F_\tau' \,. \tag{264}$$

We also used that

$$F_{t_1}'(zI+B_T)^{-1} \;\sim\; F_{t_1}'(zI+B_{T,t_1})^{-1}/(1+\xi(z;c))$$

by Lemma 21 and the Sherman-Morrison formula.

Now,

$$\begin{aligned}
&\frac{1}{P}\frac{1}{N^2}\operatorname{tr} E[F_{t_1}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}F_{t_2}F_{t_2}'(zI+B_{T,t_1,t_2})^{-1}]\\
&=\;\frac{1}{P}\frac{1}{N^2}\operatorname{tr} E\Bigg[\Bigg(((\operatorname{tr}\Sigma)^2+\operatorname{tr}(\Sigma^2))\Psi N^{-1}\Sigma_F\Psi\\
&+\Psi\Bigg(\operatorname{tr}(\Sigma\Sigma_\varepsilon)+\operatorname{tr}(N^{-1}\Sigma_F\Psi)\operatorname{tr}(\Sigma^2)\Bigg)\Bigg)(zI+B_{T,t_1,t_2})^{-1}\Bigg(((\operatorname{tr}\Sigma)^2+\operatorname{tr}(\Sigma^2))\Psi N^{-1}\Sigma_F\Psi\\
&+\Psi\Bigg(\operatorname{tr}(\Sigma\Sigma_\varepsilon)+\operatorname{tr}(N^{-1}\Sigma_F\Psi)\operatorname{tr}(\Sigma^2)\Bigg)\Bigg)(zI+B_{T,t_1,t_2})^{-1}\Bigg]
\end{aligned} \tag{265}$$

which coincides with the expression in (256). By the derivations in formulas (257) and (258), we get

$$\begin{aligned}
&\frac{1}{PN^2}\operatorname{tr} E[F_{t_1}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}F_{t_2}F_{t_2}'(zI+B_{T,t_1,t_2})^{-1}]\\
&\sim\sigma_*^2\frac{1}{P}\operatorname{tr} E[\Psi(zI+B_T)^{-1}\Psi(zI+B_T)^{-1}]\,,
\end{aligned} \tag{266}$$

and hence

$$\begin{aligned} 1 &= -z^2 m'(-z;c) + 2zm(-z;c) + c^{-1}\left(\frac{\xi(z;c)}{1+\xi(z;c)}\right)^2 \\ &+ \sigma_*^2 \frac{1}{P} \operatorname{tr} E[\Psi(zI+B_T)^{-1}\Psi(zI+B_T)^{-1}]/(1+\xi(z;c))^4 \end{aligned} \tag{267}$$

Finally,

$$\frac{\xi(z;c)}{1+\xi(z;c)} = c(1 - zm(-z;c)) \tag{268}$$

The proof of Lemma 25 is complete. □

We conclude that the first term from (253) characterized in (256) satisfies

$$\begin{aligned} Term1 &= \frac{1}{N^2 T} E[F_{t_1}'(zI+B_T)^{-1}\Big((\operatorname{tr}\Sigma)^2 \Psi N^{-1}\Sigma_F \Psi + \Psi \operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)(zI+B_T)^{-1}F_{t_1}] \\ &\sim (1+\xi(z;c))^{-2} c\Gamma_3(z) \end{aligned} \tag{269}$$

because $1/T \sim c/P$.

J.2 $Term2$ in (255)

We now proceed with the second term (255). By the Sherman-Morrison formula and Lemma 21,

$$
\begin{aligned}
&\frac{1}{N^2}E[F_{t_1}'(zI+B_T)^{-1}\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)(zI+B_T)^{-1}F_{t_2}]\\
&\sim\ \frac{1}{N^2}E[F_{t_1}'(zI+B_{T,t_1})^{-1}\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)(zI+B_{T,t_2})^{-1}F_{t_2}]/(1+\xi(z;c))^2\\
&\sim\ \frac{1}{N^2}E[F_{t_1}'\Big((zI+B_{T,t_1,t_2})^{-1}-\frac{\frac{1}{NT}(zI+B_{T,t_1,t_2})^{-1}F_{t_2}F_{t_2}'(zI+B_{T,t_1,t_2})^{-1}}{1+\frac{1}{NT}F_{t_2}'(zI+B_{T,t_1,t_2})^{-1}F_{t_2}}\Big)\\
&\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)\Big((zI+B_{T,t_1,t_2})^{-1}\\
&-\frac{\frac{1}{NT}(zI+B_{T,t_1,t_2})^{-1}F_{t_1}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}}{1+\frac{1}{NT}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}F_{t_1}}\Big)F_{t_2}]/(1+\xi(z;c))^2\\
&=\ Term1\ +\ Term2\ +\ Term3
\end{aligned}
\tag{270}
$$

where

$$
\begin{aligned}
Term1\ &=\ \frac{1}{N^2}E[F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}\\
&\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)(zI+B_{T,t_1,t_2})^{-1}F_{t_2}]/(1+\xi(z;c))^2\\
Term2\ &=\ -\frac{1}{N^2}2E[F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}\\
&\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)\\
&\times\frac{\frac{1}{NT}(zI+B_{T,t_1,t_2})^{-1}F_{t_1}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}}{1+\frac{1}{NT}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}F_{t_1}}F_{t_2}]/(1+\xi(z;c))^2\\
Term3\ &=\ \frac{1}{N^2}E[F_{t_1}'\frac{\frac{1}{NT}(zI+B_{T,t_1,t_2})^{-1}F_{t_2}F_{t_2}'(zI+B_{T,t_1,t_2})^{-1}}{1+\frac{1}{NT}F_{t_2}'(zI+B_{T,t_1,t_2})^{-1}F_{t_2}}\\
&\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)\frac{\frac{1}{NT}(zI+B_{T,t_1,t_2})^{-1}F_{t_1}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}}{1+\frac{1}{NT}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}F_{t_1}}F_{t_2}]/(1+\xi(z;c))^2
\end{aligned}
\tag{271}
$$

We now analyze each term separately.

J.3 $Term1$ in (271)

We will need the following lemma.

Lemma 26 *We have*

$$F(A) \ = \ \lambda' E[(zI + B_T)^{-1} A (zI + B_T)^{-1}]\lambda \ \to \ 0 \tag{272}$$

for any A with uniformly bounded trace norm, with A independent of λ.

Proof of Lemma 26. We know from Lemma 23 that $\lambda' E[A(zI + B_T)^{-1}]\lambda \to 0$. Furthermore,

$$
\begin{aligned}
&\lambda' E[A(zI+B_T)^{-1}]\lambda \;=\; \lambda' E[(zI+B_T)^{-1}(zI+B_T)A(zI+B_T)^{-1}]\lambda \\
&\underbrace{=}_{symmetry}\; z\lambda' E[(zI+B_T)^{-1}A(zI+B_T)^{-1}]\lambda \;+\; \frac{1}{NT}\lambda' E[(zI+B_T)^{-1}F_tF_t'A(zI+B_T)^{-1}\lambda] \\
&=\; z\lambda' E[(zI+B_T)^{-1}A(zI+B_T)^{-1}]\lambda \\
&+\; N^{-1}E\Big[\Big((zI+B_{T,t})^{-1} - \frac{\frac{1}{NT}(zI+B_{T,t})^{-1}F_tF_t'(zI+B_{T,t})^{-1}}{1+\frac{1}{NT}F_t'(zI+B_{T,t})^{-1}F_t}\Big)F_tF_t'A(zI+B_T)^{-1}\lambda\Big] \\
&\approx\; z\lambda' E[(zI+B_T)^{-1}A(zI+B_T)^{-1}]\lambda \;+\; (1+\xi(z;c))^{-1}N^{-1}\lambda' E[(zI+B_{T,t})^{-1}F_tF_t'A \\
&\times \Big((zI+B_{T,t})^{-1} - \frac{\frac{1}{NT}(zI+B_{T,t})^{-1}F_tF_t'(zI+B_{T,t})^{-1}}{1+\xi(z;c)}\Big)\lambda] \\
&=\; z\lambda' E[(zI+B_T)^{-1}A(zI+B_T)^{-1}]\lambda \\
&+\; (1+\xi(z;c))^{-1}N^{-1}\lambda' E[(zI+B_{T,t})^{-1}\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi + \Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)A(zI+B_{T,t})^{-1}]\lambda \\
&-(1+\xi(z;c))^{-2}N^{-1}\lambda' E[(zI+B_{T,t})^{-1}F_tF_t'A\frac{1}{NT}(zI+B_{T,t})^{-1}F_tF_t'(zI+B_{T,t})^{-1}]\lambda \\
&\approx\; z\lambda' E[(zI+B_T)^{-1}A(zI+B_T)^{-1}]\lambda \\
&+\; (1+\xi(z;c))^{-1}N^{-1}\lambda' E[(zI+B_{T,t})^{-1}\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi + \Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)A(zI+B_{T,t})^{-1}]\lambda \\
&-Q(z)(1+\xi(z;c))^{-2}N^{-1}\lambda' E[(zI+B_{T,t})^{-1}F_tF_t'(zI+B_{T,t})^{-1}]\lambda
\end{aligned}
\tag{273}
$$

where

$$
Q(z) \;=\; F_t'A\frac{1}{NT}(zI+B_{T,t})^{-1}F_t \;\to\; T^{-1}\operatorname{tr}E[\Psi A(zI+B_{T,t})^{-1}] \;\to 0 \tag{274}
$$

because $\|A\|_1 = o(P)$ by assumption, and

$$
\begin{aligned}
&\lambda' E[(zI+B_{T,t})^{-1}F_tF_t'(zI+B_{T,t})^{-1}]\lambda \\
&=\; N^{-1}\lambda' E[(zI+B_{T,t})^{-1}\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi + \Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)(zI+B_{T,t})^{-1}]\lambda \;=\; O(1)\,.
\end{aligned}
\tag{275}
$$

Thus, we get

$$o(1) \;\approx\; zF(A) \;+\; (1+\xi(z;c))^{-1}\, F((\Psi\Sigma_F\Psi+\Psi)A) \tag{276}$$

where $o(1)$ is uniform, and the same iterative argument as in the proof of Lemma 24 give a power series representation for $F((\Psi\Sigma_F\Psi+\Psi)^k A)$ for all k, and the same uniform boundedness argument implies that $F(A)=0$. The proof of Lemma 26 is complete. $\square$

Now, $E[F_t] = N^{-1/2}\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Psi\lambda$ and therefore

$$\begin{aligned}
&(1+\xi(z;c))^2 Term1 \;=\; \frac{1}{N^2}E[F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}\\
&\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi + \Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)(zI+B_{T,t_1,t_2})^{-1}F_{t_2}]\\
&\sim\; \frac{1}{N^3}(\operatorname{tr}(\Sigma))^2\lambda'\Psi\, E[(zI+B_{T,t_1,t_2})^{-1}\\
&\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}(\Sigma_F^*+\lambda\lambda')\Psi + \Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)(zI+B_{T,t_1,t_2})^{-1}]\Psi\lambda\\
&=\; \frac{1}{N^4}(\operatorname{tr}(\Sigma))^2\lambda'\Psi\, E[(zI+B_{T,t_1,t_2})^{-1}(\operatorname{tr}\Sigma)^2\Psi\Sigma_F^*\Psi(zI+B_{T,t_1,t_2})^{-1}]\Psi\lambda\\
&+\frac{1}{N^4}(\operatorname{tr}(\Sigma))^2\lambda'\Psi\, E[(zI+B_{T,t_1,t_2})^{-1}(\operatorname{tr}\Sigma)^2\Psi\lambda\lambda'\Psi(zI+B_{T,t_1,t_2})^{-1}]\Psi\lambda\\
&+\frac{1}{N^3}(\operatorname{tr}(\Sigma))^2\lambda'\Psi\, E[(zI+B_{T,t_1,t_2})^{-1}(\operatorname{tr}\Sigma\Sigma_\varepsilon)\Psi(zI+B_{T,t_1,t_2})^{-1}]\Psi\lambda\\
&\sim\; \Gamma_{1,1}(z)^2 \;+\; \Gamma_{4,T}(z)\,,
\end{aligned} \tag{277}$$

where Γ_4 is defined in the following lemma.

Lemma 27 *We have*

$$\begin{aligned}
&\sigma_*\lambda'\Psi E[(zI+B_{T,t_1,t_2})^{-1}\Psi(zI+B_{T,t_1,t_2})^{-1}]\Psi\lambda \;=\; \Gamma_{4,T}(z)\\
&\to\; \Gamma_4(z) \;=\; \frac{\Gamma_{1,1}(z)+z\Gamma_{1,1}'(z)-(\Gamma_{1,1}(z))^2(1+\xi(z;c))^{-2}}{(1+\xi(z;c))^{-2}}
\end{aligned} \tag{278}$$

Proof. We have by the symmetry across t and the Sherman-Morrison formula and Lemma

21 that

$$
\begin{aligned}
\Gamma_{1,1}(z) \;&\sim\; \lambda' E[\Psi(zI+B_T)^{-1}\Psi]\lambda \;=\; \lambda' E[\Psi(zI+B_T)^{-1}(zI+B_T)(zI+B_T)^{-1}\Psi]\lambda \\
&=\; z\,\lambda' E[\Psi(zI+B_T)^{-1}(zI+B_T)^{-1}\Psi]\lambda \;+\; \lambda' E[\Psi(zI+B_T)^{-1}B_T(zI+B_T)^{-1}\Psi]\lambda \\
&=\; -z\,\Gamma'_{1,1,T}(z) \;+\; \lambda' E[\Psi(zI+B_T)^{-1}\frac{1}{NT}\sum_t F_tF_t'(zI+B_T)^{-1}\Psi]\lambda \\
&=\; -z\,\Gamma'_{1,1,T}(z) \;+\; \frac{1}{N}\lambda' E[\Psi(zI+B_T)^{-1}F_tF_t'(zI+B_T)^{-1}\Psi]\lambda \\
&\sim\; -z\,\Gamma'_{1,1,T}(z) \;+\; \frac{1}{N}\lambda' E[\Psi(zI+B_{T,t})^{-1}F_tF_t'(zI+B_{T,t})^{-1}\Psi]\lambda(1+\xi(z;c))^{-2} \\
&=\; -z\,\Gamma'_{1,1,T}(z) \\
&+\; \frac{1}{N}\lambda' E[\Psi(zI+B_{T,t})^{-1}\Bigg(((\mathrm{tr}\,\Sigma)^2+\mathrm{tr}(\Sigma^2))\Psi N^{-1}\Sigma_F\Psi \\
&+\; \Psi\Big(\,\mathrm{tr}(\Sigma\Sigma_\varepsilon)+\mathrm{tr}(N^{-1}\Sigma_F\Psi)\,\mathrm{tr}(\Sigma^2)\Big)\Bigg)(zI+B_{T,t})^{-1}\Psi]\lambda(1+\xi(z;c))^{-2} \\
&\sim\; -z\,\Gamma'_{1,1,T}(z) \;+\; (\Gamma_{1,1}(z))^2(1+\xi(z;c))^{-2} \\
&+\; \Gamma_{4,T}(z)(1+\xi(z;c))^{-2}
\end{aligned}
\tag{279}
$$

The claim follows now because $\Gamma'_{1,1,T}(z)\to\Gamma'_{1,1}(z)$ by standard properties of analytic functions. The proof of Lemma 27 is complete. $\square$

J.4 $Term2$ in (271)

The next term in (271) is (note the factor of 2 as it appears two times):

$$\begin{aligned}
Term2 \;&=\; -\frac{1}{N^2}2E[F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}\\
&\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)\\
&\times\frac{\frac{1}{NT}(zI+B_{T,t_1,t_2})^{-1}F_{t_1}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}}{1+\frac{1}{NT}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}F_{t_1}}F_{t_2}]/(1+\xi(z;c))^2\\
&=\; -\frac{1}{N^2}2E[F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}\\
&\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)\\
&\times\frac{\frac{1}{NT}(zI+B_{T,t_1,t_2})^{-1}F_{t_1}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}}{1+\frac{1}{NT}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}F_{t_1}}\Psi\lambda N^{-1/2}]\operatorname{tr}(\Sigma)/(1+\xi(z;c))^2\\
&\sim\; -2\frac{1}{N}E[F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}\\
&\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)\\
&\times\frac{\frac{1}{NT}(zI+B_{T,t_1,t_2})^{-1}F_{t_1}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}}{1+\frac{1}{NT}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}F_{t_1}}\Psi\lambda N^{-1/2}]/(1+\xi(z;c))^2\\
&= -2(1+\xi(z;c))^{-2}E[X_TY_T],
\end{aligned} \tag{280}$$

where we have used that

$$E[F_{t_2}] \;=\; \Psi\lambda N^{-1/2}\,, \tag{281}$$

and where

$$\begin{aligned}
Y_T \;&=\; N^{-1/2}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}\lambda\\
X_T \;&=\; N^{-1}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}\\
&\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)\\
&\times\frac{\frac{1}{NT}(zI+B_{T,t_1,t_2})^{-1}F_{t_1}}{1+\frac{1}{NT}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}F_{t_1}}
\end{aligned} \tag{282}$$

We will need the following technical lemma whose proof follows directly from the Cauchy-Schwarz inequality.

Lemma 28 *If $X_T \to X$ in probability and is uniformly bounded and $E[Y_T^2]$ is uniformly bounded. Then,*

$$E[(X_T - X)Y_T] \ \to \ 0$$

Then, we will need

Lemma 29 *We have*

$$E[(Y_T)^2]$$

is uniformly bounded in L_2, whereas

$$E[Y_T] = E[\frac{1}{N^{1/2}} F_{t_1}'(zI + B_{T,t_1,t_2})^{-1}\Psi\lambda] \ \to \ \Gamma_{1,1}(z)\,. \tag{283}$$

Proof. Recall that

$$\lambda'\Psi^k(zI + B_T)^{-1}\Psi^\ell\lambda \ \to \ \Gamma_{k,l}(z) \tag{284}$$

by Lemma 19 and 24.

We have

$$
\begin{aligned}
&\frac{1}{N}E[\big(F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}\Psi\lambda\big)^2]\\
&= \frac{1}{N}E[F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}\Psi\lambda\lambda'(zI+B_{T,t_1,t_2})^{-1}F_{t_1}]\\
&= \frac{1}{N}\operatorname{tr}E[(zI+B_{T,t_1,t_2})^{-1}\Psi\lambda\lambda'(zI+B_{T,t_1,t_2})^{-1}F_{t_1}F_{t_1}']\\
&\sim \frac{1}{N}\operatorname{tr}E[(zI+B_{T,t_1,t_2})^{-1}\Psi\lambda\lambda'(zI+B_{T,t_1,t_2})^{-1}\\
&\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)]\\
&= E[\lambda'(zI+B_{T,t_1,t_2})^{-1}\Psi(\Sigma_F^*+\lambda\lambda')\Psi(zI+B_{T,t_1,t_2})^{-1}\Psi\lambda]\\
&+ E[\lambda'(zI+B_{T,t_1,t_2})^{-1}\Psi(zI+B_{T,t_1,t_2})^{-1}\Psi\lambda]\\
&\sim \Gamma_1(z)\Gamma_{1,1}(z) + \Gamma_3(z)
\end{aligned} \tag{285}
$$

by Lemma 25 (and Lemma 26 makes sure that the Σ_F^* contribution is zero).

The proof of Lemma 29 is complete. □

Recall that

$$
Y_T = \frac{1}{N^{1/2}}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}\Psi\lambda
$$

and

$$
\begin{aligned}
X_T &= N^{-1}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}\\
&\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)\\
&\times\frac{\frac{1}{NT}(zI+B_{T,t_1,t_2})^{-1}F_{t_1}}{1+\frac{1}{NT}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}F_{t_1}}
\end{aligned} \tag{286}
$$

Now, we know from the proof of Lemma 12 that

$$
\frac{1}{NT}F_t'AF_t-\frac{1}{NT}\operatorname{tr}(AE[F_tF_t']) \to 0
$$

in L_2 and

$$
\begin{aligned}
& N^{-1}F'_{t_1}(zI+B_{T,t_1,t_2})^{-1} \\
& \Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)\frac{1}{NT}(zI+B_{T,t_1,t_2})^{-1}F_{t_1} \\
\sim\ & \frac{1}{T}\operatorname{tr}E[(zI+B_{T,t_1,t_2})^{-1}\Big(\Psi(\Sigma_F^*+\lambda\lambda')\Psi+\sigma_*\Psi\Big) \\
& \times(zI+B_{T,t_1,t_2})^{-1}\Big(\Psi(\Sigma_F^*+\lambda\lambda')\Psi+\sigma_*\Psi\Big)] \\
& \underbrace{\sim}_{(238)\ \textit{and Lemma}\ 26} \\
& \frac{1}{T}\operatorname{tr}E[(zI+B_{T,t_1,t_2})^{-1}\Big(\Psi\lambda\lambda'\Psi+\sigma_*\Psi\Big) \\
& \times(zI+B_{T,t_1,t_2})^{-1}\Big(\Psi\lambda\lambda'\Psi+\sigma_*\Psi\Big)] \\
\sim\ & \frac{1}{T}\operatorname{tr}E[(zI+B_{T,t_1,t_2})^{-1}\Psi\lambda\lambda'\Psi(zI+B_{T,t_1,t_2})^{-1}\Psi\lambda\lambda'\Psi] \\
+\ & 2\frac{1}{T}\operatorname{tr}E[(zI+B_{T,t_1,t_2})^{-1}\Psi\lambda\lambda'\Psi(zI+B_{T,t_1,t_2})^{-1}\Psi\sigma_*] \\
+\ & \sigma_*^2\frac{1}{T}\operatorname{tr}E[(zI+B_{T,t_1,t_2})^{-1}\Psi(zI+B_{T,t_1,t_2})^{-1}\Psi] \\
\sim\ & c\Gamma_3(z)
\end{aligned}
\tag{287}
$$

by Lemma (25) because the λ-terms are $O(T^{-1})$. Furthermore, X_T is uniformly bounded by the Cauchy-Schwarz inequality. Thus,

$$
X_T \ \to\ \frac{c\Gamma_3(z)}{1+\xi(z;c)}
$$

and

$$
E[Y_T]\to\Gamma_{1,1}(z)
$$

by Lemma 29, and Lemma 28 and formula (280) imply that

$$Term2 \ \sim -2\frac{c\Gamma_3(z)\Gamma_{1,1}(z)}{(1+\xi(z;c))^3}\,. \tag{288}$$

J.5 $Term3$ in (271)

Finally, we now deal with $Term3$ in (271).

Lemma 30 *$Term3$ in (271) converges to zero.*

Proof of Lemma 30. We have

$$\begin{aligned} Term3 \ &= \ \frac{1}{N^2}E[F_{t_1}'\frac{\frac{1}{NT}(zI+B_{T,t_1,t_2})^{-1}F_{t_2}F_{t_2}'(zI+B_{T,t_1,t_2})^{-1}}{1+\frac{1}{NT}F_{t_2}'(zI+B_{T,t_1,t_2})^{-1}F_{t_2}} \\ &\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)\frac{\frac{1}{NT}(zI+B_{T,t_1,t_2})^{-1}F_{t_1}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}}{1+\frac{1}{NT}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}F_{t_1}}F_{t_2}]/(1+\xi(z;c))^2 \\ &= \ E[X_TY_T]\,/(1+\xi(z;c))^2\,, \end{aligned} \tag{289}$$

where we have defined

$$X_T \ = \ \frac{\big(\frac{1}{NT}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}F_{t_2}\big)^2}{(1+\frac{1}{NT}F_{t_1}'(zI+B_{T,t_1,t_2})^{-1}F_{t_1})(1+\frac{1}{NT}F_{t_2}'(zI+B_{T,t_1,t_2})^{-1}F_{t_2})}$$

and

$$Y_T \ = \ \frac{1}{N}F_{t_2}'(zI+B_{T,t_1,t_2})^{-1}\Bigg(\Psi\Sigma_F\Psi \ + \ \sigma_*\Psi\Bigg)(zI+B_{T,t_1,t_2})^{-1}F_{t_1}\,.$$

The first observation is that X_T is uniformly bounded by the Cauchy-Schwarz inequality and has a $O(1/T)$ L_2-norm by Lemma 31. Since the first component of Y_T,

$$\frac{1}{N}F_{t_2}'(zI+B_{T,t_1,t_2})^{-1}\Psi\Sigma_F\Psi(zI+B_{T,t_1,t_2})^{-1}F_{t_1}\,.$$

has a $o(T)$ L_2-norm, we get that this part is negligible by Lemma 28.

Lemma 31 *We have that*

$$E[(\frac{1}{N}F'_{t_1}AF_{t_2})^2] \;=\; O(\|A\|_1\,\|A\|_\infty)\,.$$

for any A. Thus,

$$\left(\frac{1}{NT}F'_{t_1}(zI+B_{T,t_1,t_2})^{-1}F_{t_2}\right)^2$$

converges to zero in L_1, while

$$\frac{1}{N}F'_{t_2}(zI+B_{T,t_1,t_2})^{-1}\Psi\Sigma_F\Psi(zI+B_{T,t_1,t_2})^{-1}F_{t_1}$$

has a uniformly bounded L_2-norm because $\mathrm{tr}(\Sigma_F)=o(T)$.

Proof. We have

$$\begin{aligned}
&E[(N^{-1}F'_{t_1}AF_{t_2})^2] \;=\; N^{-2}E[F'_{t_1}AF_{t_2}F'_{t_2}AF_{t_1}]\\
&=\; N^{-2}\,\mathrm{tr}\,E[AF_{t_2}F'_{t_2}AF_{t_1}F'_{t_1}]\\
&\sim\; \mathrm{tr}\,E[A\left(\Psi\Sigma_F\Psi \;+\; \sigma_*\Psi\right)\\
&\times A\left(\Psi\Sigma_F\Psi \;+\; \sigma_*\Psi\right)\Big)]
\end{aligned} \tag{290}$$

The proof of Lemma 31 is complete. □

Lemma 32 *We have*

$$E[(N^{-1}F'_{t_1}AF_{t_2})^4] \;=\; O(P^2)$$

for any uniformly bounded A.

Indeed, Lemma 32 implies that

$$E[X_T^2] \;\le\; T^{-4}E[(N^{-1}F'_{t_1}(zI+B_{T,t_1,t_2})^{-1}F_{t_2})^4] = O(P^2/T^4)$$

while Lemma 31 implies that

$$E[Y_T^2] \;=\; O(P)\,.$$

Thus,

$$|E[X_TY_T]|^2 \;\le\; E[X_T^2]E[Y_T^2] \;=\; O(P^2/T^4)\,O(P) \;\to\; 0$$

and the claim follows.

Proof of Lemma 32. Without loss of generality, we may assume that A is symmetric. Recall that

$$R_t \;=\; S_{t-1}\beta_t \;+\; \varepsilon_t, \tag{291}$$

and

$$F_t \;=\; S'_{t-1}R_t \;=\; S'_{t-1}S_{t-1}\beta_t + S'_{t-1}\varepsilon_t \;=\; Z_t\beta \;+\; S'_{t-1}\varepsilon_t \tag{292}$$

and therefore

$$F_tF'_t \;=\; Z_t\beta\beta' Z_t + S'_{t-1}\varepsilon_t\beta' Z_t + Z_t\beta\varepsilon'_t S_{t-1} + S'_{t-1}\varepsilon_t\varepsilon'_t S_{t-1}\,. \tag{293}$$

and formula (174) applied to $t = t_1$ implies

$$
\begin{aligned}
E[(F_{t_1}' A F_{t_2})^4] \;&=\; E[F_{t_1}' A F_{t_2} F_{t_2}' A F_{t_1} F_{t_1}' A F_{t_2} F_{t_2}' A F_{t_1}] \\
=&\; \operatorname{tr} E[F_{t_1} F_{t_1}' A F_{t_2} F_{t_2}' A F_{t_1} F_{t_1}' A F_{t_2} F_{t_2}' A] \\
=&\; \operatorname{tr} E[Z_{t_1} \beta\beta' Z_{t_1} A F_{t_2} F_{t_2}' A Z_{t_1} \beta\beta' Z_{t_1} A F_{t_2} F_{t_2}' A] \\
+&\; \operatorname{tr} E[Z_{t_1} \beta\beta' Z_{t_1} A F_{t_2} F_{t_2}' A Z_{t_1} A F_{t_2} F_{t_2}' A] \\
+&\, 2 \operatorname{tr} E[(\beta' Z_{t_1} A F_{t_2} F_{t_2}' A Z_{t_1} \beta) Z_{t_1} A F_{t_2} F_{t_2}' A] \\
+&\, ((\kappa_\varepsilon - 1) \operatorname{tr} E[Z_{t_1} A F_{t_2} F_{t_2}' A Z_{t_1} A F_{t_2} F_{t_2}' A] \\
+&\; E[\operatorname{tr}(Z_{t_1} A F_{t_2} F_{t_2}' A)^2]
\end{aligned}
\tag{294}
$$

We then again apply (174) to $t = t_2$. It is then straightforward to show that the leading contribution will be

$$
\begin{aligned}
E[\operatorname{tr}(Z_{t_1} A Z_{t_2} A)^2] \;&=\; E[\left(\sum X_{i_1,k_1,t_1} \lambda_{i_1}(\Sigma) X_{i_1,k_2,t_1} \lambda_{k_2}(\tilde{A}) X_{i_2,k_2,t_2} \lambda_{i_2}(\Sigma) X_{i_2,k_1,t_2} \lambda_{k_1}(\tilde{A})\right)^2] \\
=&\; E[\sum X_{i_1,k_1,t_1} \lambda_{i_1}(\Sigma) X_{i_1,k_2,t_1} \lambda_{k_2}(\tilde{A}) X_{i_2,k_2,t_2} \lambda_{i_2}(\Sigma) X_{i_2,k_1,t_2} \lambda_{k_1}(\tilde{A}) \\
\times&\, X_{\tilde{i}_1,\tilde{k}_1,t_1} \lambda_{\tilde{i}_1}(\Sigma) X_{\tilde{i}_1,\tilde{k}_2,t_1} \lambda_{\tilde{k}_2}(\tilde{A}) X_{\tilde{i}_2,\tilde{k}_2,t_2} \lambda_{i_2}(\Sigma) X_{\tilde{i}_2,\tilde{k}_1,t_2} \lambda_{\tilde{k}_1}(\tilde{A})]
\end{aligned}
\tag{295}
$$

Non-zero terms must have that $(i_1, k_1), (i_1, k_2), (\tilde{i}_1, \tilde{k}_1), (\tilde{i}_2, \tilde{k}_2)$ is coming in at least two identical pairs. For example, $k_1 = k_2,\ \tilde{k}_1 = \tilde{k}_2$ will give $\operatorname{tr}(\Sigma)^4 (\operatorname{tr}(\tilde{A}^2))^2$. All other terms will be even smaller because more indices should be equal. For example, if $k_1 = \tilde{k}_1$ we ought to have $i_1 = \tilde{i}_1$. The proof of Lemma 32 is complete. □

Thus, (289) converges to zero.

The proof of Lemma 30 is complete. □

Summarizing, we get from (280) and (277), (288), that

$$Term2 \;=\; (1+\xi(z;c))^{-2}(\Gamma_{1,1}(z)^2+\Gamma_4(z)) \;-2\frac{c\Gamma_3(z)\Gamma_{1,1}(z)}{(1+\xi(z;c))^3} \tag{296}$$

and (253) implies

$$\begin{aligned} &E[(R^F_{t+1}(z))^2] \underbrace{\sim}_{(253)} Term1 \;+\; Term2 \\ &\underbrace{\sim}(269)\; (1+\xi(z;c))^{-2}c\Gamma_3(z) \;+\; Term2 \\ &\underbrace{\sim}_{(296)}(1+\xi(z;c))^{-2}c\Gamma_3(z) \;+\; (1+\xi(z;c))^{-2}(\Gamma_{1,1}(z)^2+\Gamma_4(z)) \;-2\frac{c\Gamma_3(z)\Gamma_{1,1}(z)}{(1+\xi(z;c))^3} \end{aligned} \tag{297}$$

and the final expression follows from Lemma 27:

$$\Gamma_{1,1}(z)^2+\Gamma_4(z) \;=\; \Gamma_{1,1}(z)^2+\frac{\Gamma_{1,1}(z)+z\Gamma'_{1,1}(z)-(\Gamma_{1,1}(z))^2(1+\xi(z;c))^{-2}}{(1+\xi(z;c))^{-2}} \tag{298}$$

K Proof of Theorem ??

The same argument as in (253) implies that

$$\begin{aligned} &E[R^F_{t+1}(z_1)R^F_{t+1}(z_2)] \\ &\sim\; Term1 \;+\; Term2 \end{aligned} \tag{299}$$

with

$$Term1 \;=\; \frac{1}{N^2T}E[F'_{t_1}(z_1I+B_T)^{-1}\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi+\Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)(z_2I+B_T)^{-1}F_{t_1}] \tag{300}$$

and

$$
Term2 \; = \; \frac{1}{N^2}\frac{T(T-1)}{T^2}E[F_{t_1}'(z_1 I + B_T)^{-1}\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi + \Psi \operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)(z_2 I + B_T)^{-1}F_{t_2}] \tag{301}
$$

for any $t_1 \neq t_2$.

The same argument as above implies that

$$
Term1 \; \sim \; (1+\xi(z_1))^{-1}(1+\xi(z_2))^{-1}c\Gamma_3(z_1,z_2) \tag{302}
$$

where

Lemma 33 *We have*

$$
\begin{aligned}
&\frac{1}{PN^2}\operatorname{tr} E[F_{t_1}F_{t_1}'(z_1 I + B_{T,t_1,t_2})^{-1}F_{t_2}F_{t_2}'(z_2 I + B_{T,t_1,t_2})^{-1}] \\
&\sim \; \frac{1}{P}\operatorname{tr} E[\Psi(z_1 I + B_T)^{-1}\Psi(z_2 I + B_T)^{-1}] \\
&\to \; \Gamma_3(z_1,z_2) \; = \; \Big(1 - \Big(\frac{z_2^2 m(-z_2;c) - z_1^2 m(-z_1;c)}{z_2 - z_1} + c^{-1}\frac{\xi(z_1)}{1+\xi(z_1)}\frac{\xi(z_1)}{1+\xi(z_1)}\Big)\Big)((1+\xi(z_1))(1+\xi(z_2)))^2\,.
\end{aligned} \tag{303}
$$

Proof. We have by the Sherman-Morrison formula that

$$
\begin{aligned}
&\frac{1}{P}\frac{1}{N^2T}\operatorname{tr} E[F_{t_1}F_{t_1}'(z_1 I + B_T)^{-1}F_{t_1}F_{t_1}'(z_2 I + B_T)^{-1}] \\
&\sim \frac{1}{c}\frac{1}{N^2T^2}E[F_{t_1}'(z_1 I + B_T)^{-1}F_{t_1}F_{t_1}'(z_2 I + B_T)^{-1}F_{t_1}] \\
&= \; c^{-1}E\left[\frac{\frac{1}{NT}F_{t_1}'(z_1 I + B_{T,t_1})^{-1}F_{t_1}}{1+\frac{1}{NT}F_{t_1}'(z_1 I + B_{T,t_1})^{-1}F_{t_1}}\frac{\frac{1}{NT}F_{t_1}'(z_2 I + B_{T,t_1})^{-1}F_{t_1}}{1+\frac{1}{NT}F_{t_1}'(z_2 I + B_{T,t_1})^{-1}F_{t_1}}\right] \\
&\sim \; c^{-1}\frac{\xi(z_1)}{1+\xi(z_1)}\frac{\xi(z_2)}{1+\xi(z_2)}
\end{aligned} \tag{304}
$$

by Lemma 21. Now,

$$
\begin{aligned}
1 \;&=\; \frac{1}{P}\operatorname{tr} E[(z_1 I + B_T)(z_1 I + B_T)^{-1}(z_2 I + B_T)(z_2 I + B_T)^{-1}] \\
&=\; f(z_1, z_2) \\
&+\; \frac{1}{P}\operatorname{tr} E[B_T(z_1 I + B_T)^{-1} B_T (z_2 I + B_T)^{-1}] \\
&\sim\; f(z_1, z_2) \;+\; \frac{1}{P}\frac{1}{N^2 T^2}\sum_{t_1, t_2}\operatorname{tr} E[F_{t_1}F_{t_1}'(z_1 I + B_T)^{-1} F_{t_2}F_{t_2}'(z_2 I + B_T)^{-1}] \\
&=\; f(z_1, z_2) + \frac{1}{P}\frac{1}{N^2 T}\operatorname{tr} E[F_{t_1}F_{t_1}'(z_1 I + B_T)^{-1} F_{t_1}F_{t_1}'(z_2 I + B_T)^{-1}] \\
&+\; \frac{1}{P}\frac{1}{N^2}\frac{T(T-1)}{T^2}\operatorname{tr} E[F_{t_1}F_{t_1}'(z_1 I + B_T)^{-1} F_{t_2}F_{t_2}'(z_2 I + B_T)^{-1}] \\
&\sim\; f(z_1, z_2) + c^{-1}\frac{\xi(z_1)}{1+\xi(z_1)}\frac{\xi(z_1)}{1+\xi(z_1)} \\
&+\; \frac{1}{P}\frac{1}{N^2}\operatorname{tr} E[F_{t_1}F_{t_1}'(z_1 I + B_T)^{-1} F_{t_2}F_{t_2}'(z_2 I + B_T)^{-1}] \\
&\sim\; f(z_1, z_2) + c^{-1}\frac{\xi(z_1)}{1+\xi(z_1)}\frac{\xi(z_1)}{1+\xi(z_1)} \\
&+\; \frac{1}{P}\frac{1}{N^2}\operatorname{tr} E[F_{t_1}F_{t_1}'(z_1 I + B_{T,t_1})^{-1} F_{t_2}F_{t_2}'(z_2 I + B_{T,t_2})^{-1}]/((1+\xi(z_1))(1+\xi(z_2))) \\
&\sim\; f(z_1, z_2) + c^{-1}\frac{\xi(z_1)}{1+\xi(z_1)}\frac{\xi(z_1)}{1+\xi(z_1)} \\
&+\; \frac{1}{P}\frac{1}{N^2} E[F_{t_1}'(z_2 I + B_{T,t_1,t_2})^{-1} F_{t_2}F_{t_2}'(z_1 I + B_{T,t_1,t_2})^{-1} F_{t_1}]/((1+\xi(z_1))(1+\xi(z_2)))^2 \\
&=\; f(z_1, z_2) \;+\; c^{-1}\frac{\xi(z_1)}{1+\xi(z_1)}\frac{\xi(z_1)}{1+\xi(z_1)} \\
&+\; \frac{1}{P}\frac{1}{N^2}\operatorname{tr} E[F_{t_1}F_{t_1}'(z_1 I + B_{T,t_1,t_2})^{-1} F_{t_2}F_{t_2}'(z_2 I + B_{T,t_1,t_2})^{-1}]/((1+\xi(z_1))(1+\xi(z_2)))^2
\end{aligned}
\tag{305}
$$

where we have defined

$$
B_{T,t_1,t_2} \;=\; \frac{1}{NT}\sum_{\tau \notin \{t_1, t_2\}} F_\tau F_\tau' \,. \tag{306}
$$

and

$$f(z_1,z_2) \;=\; \frac{1}{P} z_1 z_2 \operatorname{tr} E[(z_1 I{+}B_T)^{-1}(z_2 I{+}B_T)^{-1}] + (z_1{+}z_2)\,\frac{1}{P}\operatorname{tr} E[(z_1 I{+}B_T)^{-1}(z_2 I{+}B_T)^{-1}B_T]\,. \tag{307}$$

We also used that

$$F_{t_1}'(zI + B_T)^{-1} \;\sim\; F_{t_1}'(zI + B_{T,t_1})^{-1}/(1+\xi(z;c))$$

by Lemma 21 and the Sherman-Morrison formula.

Now,

$$\begin{aligned}
&\frac{1}{P}\frac{1}{N^2}\operatorname{tr} E[F_{t_1}F_{t_1}'(z_1 I + B_{T,t_1,t_2})^{-1}F_{t_2}F_{t_2}'(z_2 I + B_{T,t_1,t_2})^{-1}]\\
&= \;\frac{1}{P}\frac{1}{N^2}\operatorname{tr} E\Big[\Bigg(((\operatorname{tr}\Sigma)^2 + \operatorname{tr}(\Sigma^2))\Psi N^{-1}\Sigma_F\Psi\\
&+\Psi\Big(\operatorname{tr}(\Sigma) + \operatorname{tr}(N^{-1}\Sigma_F\Psi)\operatorname{tr}(\Sigma^2)\Big)\Bigg)(z_1 I + B_{T,t_1,t_2})^{-1}\Bigg(((\operatorname{tr}\Sigma)^2 + \operatorname{tr}(\Sigma^2))\Psi N^{-1}\Sigma_F\Psi\\
&+\Psi\Big(\operatorname{tr}(\Sigma) + \operatorname{tr}(N^{-1}\Sigma_F\Psi)\operatorname{tr}(\Sigma^2)\Big)\Bigg)(z_2 I + B_{T,t_1,t_2})^{-1}\Big]
\end{aligned} \tag{308}$$

which coincides with the expression in (256). By the derivations in formulas (257) and (258), we get

$$\begin{aligned}
&\frac{1}{PN^2}\operatorname{tr} E[F_{t_1}F_{t_1}'(z_1 I + B_{T,t_1,t_2})^{-1}F_{t_2}F_{t_2}'(z_2 I + B_{T,t_1,t_2})^{-1}]\\
&\sim \frac{1}{P}\operatorname{tr} E[\Psi(z_1 I + B_T)^{-1}\Psi(z_2 I + B_T)^{-1}]\,,
\end{aligned} \tag{309}$$

and hence

$$
\begin{aligned}
1 \ &\sim \ f(z_1, z_2) + c^{-1}\frac{\xi(z_1)}{1+\xi(z_1)}\frac{\xi(z_1)}{1+\xi(z_1)} \\
&+ \ \frac{1}{P}\operatorname{tr} E[\Psi(zI+B_T)^{-1}\Psi(zI+B_T)^{-1}]/((1+\xi(z_1))(1+\xi(z_2)))^2\,,
\end{aligned}
\tag{310}
$$

Finally,

$$
\begin{aligned}
f(z_1, z_2) \ &= \ \frac{1}{P}z_1z_2\operatorname{tr} E[(z_1I+B_T)^{-1}(z_2I+B_T)^{-1}] \\
&+ \ (z_1+z_2)\frac{1}{P}\operatorname{tr} E[(z_1I+B_T)^{-1}(z_2I+B_T)^{-1}(B_T+z_2I-z_2I)] \\
&= \ P^{-1}(z_1z_2-(z_1+z_2)z_2)(z_2-z_1)^{-1}(m(-z_1;c)-m(-z_2;c)) \ + \ (z_1+z_2)m(-z_1;c) \\
&= \ \frac{z_2^2m(-z_2;c)-z_1^2m(-z_1;c)}{z_2-z_1}\,.
\end{aligned}
\tag{311}
$$

□

Thus,

$$
Term1 \ \sim \ \frac{\Gamma_3(z_1,z_2)}{(1+\xi(z_1))(1+\xi(z_2))}\,. \tag{312}
$$

L $Term2$ in (299)

We now proceed with $Term2$ in (299):

$$
\begin{aligned}
&((1+\xi(z_1))(1+\xi(z_2))) \ \times \ Term2 \ from \ (299) \\
&\sim \ \frac{1}{N^2}E[F_{t_1}'\Big((z_1I+B_{T,t_1,t_2})^{-1} - \frac{\frac{1}{NT}(z_1I+B_{T,t_1,t_2})^{-1}F_{t_2}F_{t_2}'(z_1I+B_{T,t_1,t_2})^{-1}}{1+\frac{1}{NT}F_{t_2}'(z_1I+B_{T,t_1,t_2})^{-1}F_{t_2}}\Big) \\
&\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi + \Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)\Big((z_2I+B_{T,t_1,t_2})^{-1} \\
&- \frac{\frac{1}{NT}(z_2I+B_{T,t_1,t_2})^{-1}F_{t_1}F_{t_1}'(z_2I+B_{T,t_1,t_2})^{-1}}{1+\frac{1}{NT}F_{t_1}'(z_2I+B_{T,t_1,t_2})^{-1}F_{t_1}}\Big)F_{t_2}] \\
&= \ Term1 \ + \ Term2 \ + \ Term3
\end{aligned}
\tag{313}
$$

where

$$
\begin{aligned}
Term1 \;&=\; \frac{1}{N^2}E[F_{t_1}'(z_1 I + B_{T,t_1,t_2})^{-1} \\
&\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi + \Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)(z_2 I + B_{T,t_1,t_2})^{-1}F_{t_2}] \\
Term2 \;&=\; \tau(z_1,z_2) \;+\; \tau(z_2,z_1) \\
\tau(z_1,z_2) \;&=\; -\frac{1}{N^2}E[F_{t_1}'(z_1 I + B_{T,t_1,t_2})^{-1} \\
&\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi + \Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big) \\
&\times \frac{\frac{1}{NT}(z_2 I + B_{T,t_1,t_2})^{-1}F_{t_1}F_{t_1}'(z_2 I + B_{T,t_1,t_2})^{-1}}{1+\frac{1}{NT}F_{t_1}'(z_2 I + B_{T,t_1,t_2})^{-1}F_{t_1}}F_{t_2}] \\
Term3 \;&=\; \frac{1}{N^2}E[F_{t_1}'\frac{\frac{1}{NT}(z_1 I + B_{T,t_1,t_2})^{-1}F_{t_2}F_{t_2}'(z_1 I + B_{T,t_1,t_2})^{-1}}{1+\frac{1}{NT}F_{t_2}'(z_1 I + B_{T,t_1,t_2})^{-1}F_{t_2}} \\
&\Big((\operatorname{tr}\Sigma)^2\Psi N^{-1}\Sigma_F\Psi + \Psi\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big)\frac{\frac{1}{NT}(z_2 I + B_{T,t_1,t_2})^{-1}F_{t_1}F_{t_1}'(z_2 I + B_{T,t_1,t_2})^{-1}}{1+\frac{1}{NT}F_{t_1}'(z_2 I + B_{T,t_1,t_2})^{-1}F_{t_1}}F_{t_2}]
\end{aligned}
\tag{314}
$$

The same argument as above implies that $Term3$ is asymptotically negligible.

We now analyze each term separately.

L.1 $Term1$ in (314)

First, $E[F_t] = N^{-1/2}\operatorname{tr}(\Sigma\Sigma_\varepsilon)\Psi\lambda$ and therefore

$$
\begin{aligned}
&Term1 \\
&=\; \frac{1}{N^4}(\operatorname{tr}(\Sigma))^2\lambda'\Psi\, E[(z_1 I + B_{T,t_1,t_2})^{-1}(\operatorname{tr}\Sigma)^2\Psi\Sigma_F^*\Psi(z_2 I + B_{T,t_1,t_2})^{-1}]\Psi\lambda \\
&+ \frac{1}{N^4}(\operatorname{tr}(\Sigma))^2\lambda'\Psi\, E[(z_1 I + B_{T,t_1,t_2})^{-1}(\operatorname{tr}\Sigma)^2\Psi\lambda\lambda'\Psi(z_2 I + B_{T,t_1,t_2})^{-1}]\Psi\lambda \\
&+ \frac{1}{N^3}(\operatorname{tr}(\Sigma))^2\lambda'\Psi\, E[(z_1 I + B_{T,t_1,t_2})^{-1}(\operatorname{tr}\Sigma\Sigma_\varepsilon)\Psi(z_2 I + B_{T,t_1,t_2})^{-1}]\Psi\lambda \\
&\sim\; \Gamma_{1,1}(z_1)\Gamma_{1,1}(z_2) \;+\; \Gamma_{4,T}(z_1,z_2)\,,
\end{aligned}
\tag{315}
$$

where Γ_4 is defined in the following lemma. Here, we have used Lemma 26.

Lemma 34 *We have*

$$\lambda'\Psi E[(z_1 I + B_{T,t_1,t_2})^{-1}\Psi(z_2 I + B_{T,t_1,t_2})^{-1}]\Psi\lambda \;=\; \Gamma_{4,T}(z_1,z_2)$$
$$\to\; \Gamma_4(z_1,z_2) \;=\; \frac{\frac{z_2\Gamma_{1,1,T}(z_2)-z_1\Gamma_{1,1,T}(z_1)}{z_2-z_1} - \frac{\Gamma_{1,1}(z_1)\Gamma_{1,1}(z_2)}{(1+\xi(z_1))(1+\xi(z_2))}}{(1+\xi(z_1))^{-1}(1+\xi(z_2))^{-1}} \tag{316}$$

Proof. We have by the symmetry across t and the Sherman-Morrison formula and Lemma 21 that

$$\begin{aligned}
\Gamma_{1,1}(z_1) \;&\sim\; \lambda' E[\Psi(z_1 I + B_T)^{-1}\Psi]\lambda \;=\; \lambda' E[\Psi(z_1 I + B_T)^{-1}(z_2 I + B_T)(z_2 I + B_T)^{-1}\Psi]\lambda \\
&=\; z_2\,\lambda' E[\Psi(z_1 I + B_T)^{-1}(z_2 I + B_T)^{-1}\Psi]\lambda \;+\; \lambda' E[\Psi(z_1 I + B_T)^{-1}B_T(z_2 I + B_T)^{-1}\Psi]\lambda \\
&=\; -z_2\,\frac{\Gamma_{1,1,T}(z_2)-\Gamma_{1,1,T}(z_1)}{z_2-z_1} \;+\; \lambda' E[\Psi(z_1 I + B_T)^{-1}\frac{1}{NT}\sum_t F_t F_t'(z_2 I + B_T)^{-1}\Psi]\lambda \\
&=\; -z_2\,\frac{\Gamma_{1,1,T}(z_2)-\Gamma_{1,1,T}(z_1)}{z_2-z_1} \;+\; \frac{1}{N}\lambda' E[\Psi(z_1 I + B_T)^{-1}F_t F_t'(z_2 I + B_T)^{-1}\Psi]\lambda \\
&\sim\; -z_2\,\frac{\Gamma_{1,1,T}(z_2)-\Gamma_{1,1,T}(z_1)}{z_2-z_1} \;+\; \frac{1}{N}\lambda' E[\Psi(z_1 I + B_{T,t})^{-1}F_t F_t'(z_2 I + B_{T,t})^{-1}\Psi]\lambda(1+\xi(z_1))^{-1}(1+\xi(z_2))^{-1} \\
&=\; -z_2\,\frac{\Gamma_{1,1,T}(z_2)-\Gamma_{1,1,T}(z_1)}{z_2-z_1} \\
&+\; \frac{1}{N}\lambda' E[\Psi(z_1 I + B_{T,t})^{-1}\Bigg(((\operatorname{tr}\Sigma)^2 + \operatorname{tr}(\Sigma^2))\Psi N^{-1}\Sigma_F\Psi \\
&+\; \Psi\Big(\operatorname{tr}(\Sigma) + \operatorname{tr}(N^{-1}\Sigma_F\Psi)\operatorname{tr}(\Sigma^2)\Big)\Bigg)(z_2 I + B_{T,t})^{-1}\Psi]\lambda(1+\xi(z_1))^{-1}(1+\xi(z_2))^{-1} \\
&\sim\; -z_2\,\frac{\Gamma_{1,1,T}(z_2)-\Gamma_{1,1,T}(z_1)}{z_2-z_1} \;+\; \Gamma_{1,1}(z_1)\Gamma_{1,1}(z_2)(1+\xi(z_1))^{-1}(1+\xi(z_2))^{-1} \\
&+\; \Gamma_{4,T}(z_1,z_2)(1+\xi(z_1))^{-1}(1+\xi(z_2))^{-1}
\end{aligned} \tag{317}$$

The claim follows now because $\Gamma'_{1,1,T}(z) \to \Gamma'_{1,1}(z)$ by standard properties of analytic functions. The proof of Lemma 34 is complete. □

L.2 $Term2$ in (314)

The next term in (314) is

$$\begin{aligned}
\tau(z_1, z_2) \;=\; & -\frac{1}{N^2} E[F_{t_1}'(z_1 I + B_{T,t_1,t_2})^{-1} \\
& \Big((\operatorname{tr}\Sigma)^2 \Psi N^{-1} \Sigma_F \Psi + \Psi \operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big) \\
& \times \frac{\frac{1}{NT}(z_2 I + B_{T,t_1,t_2})^{-1} F_{t_1} F_{t_1}'(z_2 I + B_{T,t_1,t_2})^{-1}}{1 + \frac{1}{NT} F_{t_1}'(z_2 I + B_{T,t_1,t_2})^{-1} F_{t_1}} F_{t_2}]/((1+\xi(z_1))(1+\xi(z_2))) \\
& = \; -\frac{1}{N^2} E[F_{t_1}'(z_1 I + B_{T,t_1,t_2})^{-1} \\
& \Big((\operatorname{tr}\Sigma)^2 \Psi N^{-1} \Sigma_F \Psi + \Psi \operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big) \\
& \times \frac{\frac{1}{NT}(z_2 I + B_{T,t_1,t_2})^{-1} F_{t_1} F_{t_1}'(z_2 I + B_{T,t_1,t_2})^{-1}}{1 + \frac{1}{NT} F_{t_1}'(z_2 I + B_{T,t_1,t_2})^{-1} F_{t_1}} \Psi \lambda N^{-1/2}] \operatorname{tr}(\Sigma)//((1+\xi(z_1))(1+\xi(z_2))) \\
& \sim \; -\frac{1}{N} E[F_{t_1}'(z_1 I + B_{T,t_1,t_2})^{-1} \\
& \Big((\operatorname{tr}\Sigma)^2 \Psi N^{-1} \Sigma_F \Psi + \Psi \operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big) \\
& \times \frac{\frac{1}{NT}(z_2 I + B_{T,t_1,t_2})^{-1} F_{t_1} F_{t_1}'(z_2 I + B_{T,t_1,t_2})^{-1}}{1 + \frac{1}{NT} F_{t_1}'(z_2 I + B_{T,t_1,t_2})^{-1} F_{t_1}} \Psi \lambda N^{-1/2}]//((1+\xi(z_1))(1+\xi(z_2))) \\
& = -((1+\xi(z_1))(1+\xi(z_2)))^{-1} E[X_T Y_T],
\end{aligned} \tag{318}$$

where we have used that

$$E[F_{t_2}] \;=\; \Psi \lambda N^{-1/2}\,, \tag{319}$$

and where

$$
\begin{aligned}
Y_T &= N^{-1/2} F_{t_1}'(z_1 I + B_{T,t_1,t_2})^{-1}\lambda \\
X_T &= N^{-1} F_{t_1}'(z_2 I + B_{T,t_1,t_2})^{-1} \\
&\Big((\operatorname{tr}\Sigma)^2 \Psi N^{-1}\Sigma_F \Psi + \Psi \operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big) \\
&\times \frac{\frac{1}{NT}(z_2 I + B_{T,t_1,t_2})^{-1} F_{t_1}}{1 + \frac{1}{NT} F_{t_1}'(z_2 I + B_{T,t_1,t_2})^{-1} F_{t_1}}
\end{aligned}
\tag{320}
$$

Recall that

$$
Y_T = \frac{1}{N^{1/2}} F_{t_1}'(z_1 I + B_{T,t_1,t_2})^{-1}\Psi\lambda
$$

and

$$
\begin{aligned}
X_T &= N^{-1} F_{t_1}'(z_2 I + B_{T,t_1,t_2})^{-1} \\
&\Big((\operatorname{tr}\Sigma)^2 \Psi N^{-1}\Sigma_F \Psi + \Psi \operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big) \\
&\times \frac{\frac{1}{NT}(z_2 I + B_{T,t_1,t_2})^{-1} F_{t_1}}{1 + \frac{1}{NT} F_{t_1}'(z_2 I + B_{T,t_1,t_2})^{-1} F_{t_1}}
\end{aligned}
\tag{321}
$$

Now, we know from the proof of Lemma 12 that

$$
\frac{1}{NT} F_t' A F_t - \frac{1}{NT} \operatorname{tr}(A E[F_t F_t']) \to 0
$$

in L_2 and

$$
\begin{aligned}
&N^{-1} F_{t_1}'(zI + B_{T,t_1,t_2})^{-1} \\
&\Big((\operatorname{tr}\Sigma)^2 \Psi N^{-1}\Sigma_F \Psi + \Psi \operatorname{tr}(\Sigma\Sigma_\varepsilon)\Big) \frac{1}{NT}(zI + B_{T,t_1,t_2})^{-1} F_{t_1} \\
&\sim c\Gamma_3(z)
\end{aligned}
\tag{322}
$$

by (287).

Furthermore, X_T is uniformly bounded by the Cauchy-Schwarz inequality. Thus,

$$X_T \;\to\; \frac{c\Gamma_3(z_2)}{1+\xi(z_2)}$$

and

$$E[Y_T] \to \Gamma_{1,1}(z_1)$$

by Lemma 29, and Lemma 28 implies

$$Term2 \;\sim -c\frac{\Gamma_3(z_2)\Gamma_{1,1}(z_1)(1+\xi(z_2))^{-1}+\Gamma_3(z_1)\Gamma_{1,1}(z_2)(1+\xi(z_1))^{-1}}{(1+\xi(z_1))(1+\xi(z_2))}\,. \tag{323}$$

M Proofs for the Mis-Specified Model

Lemma 35 *Define*

$$m(-z;cq) \;=\; \lim_{P_1\to\infty,\ P_1/P\to q} P_1^{-1}\,\mathrm{tr}((zI+B_T^{(1)})^{-1}) \tag{324}$$

and let $\xi(z;cq)$ be uniquely defined through

$$\frac{(cq)^{-1}\xi(z;cq)}{1+\xi(z;cq)} \;=\; 1 \;-\; m(-z;cq)z\,. \tag{325}$$

Then,

$$\frac{1}{T}\,\mathrm{tr}((zI+B_T^{(1)})^{-1}\sigma_*\Psi_{1,1}) \;\to\; \xi(z;cq) \tag{326}$$

almost surely and

$$\frac{1}{NT}(F_{T+1}^{(1)})'(zI+B_T^{(1)})^{-1}F_{T+1}^{(1)} \;\to\; \xi(z;cq) \tag{327}$$

in probability.

Lemma 36 *Let*

$$\Gamma_{1,1}(z;q) \;=\; \lim(\lambda^{(1)})'\Psi_{1,1}(zI + B_T^{(1)})^{-1}\Psi_{1,1}(\lambda^{(1)})\,. \tag{328}$$

Then, this limit exists almost surely and is non-random. Let

$$\delta(z;q) \;=\; -\sigma_* z^{-1}(1+\xi(z;cq))^{-1}\,. \tag{329}$$

Then,

$$\Gamma_{1,1}(z;q) \;=\; q\frac{z^{-1}P_1^{-1}\operatorname{tr}(\Psi_{1,1}^2(I-\Psi_{1,1}\delta(z;q))^{-1}\Sigma_\lambda^{(1)})}{1-\delta(z;q)qP_1^{-1}\operatorname{tr}(\Psi_{1,1}^2(I-\Psi_{1,1}\delta(z;q))^{-1}\Sigma_\lambda^{(1)})}\,. \tag{330}$$