Surprises in Modern Portfolio Theory

Oliver Hellum, Theis Ingerslev Jensen, Bryan Kelly, and Semyon Malamud*

First version: September 10, 2025

This version: March 16, 2026

Abstract

The literature has long wrestled with the practical usefulness of Modern Portfolio Theory (MPT), and extensive evidence shows its performance decreases rapidly with the number of assets (N). We present several new and counterintuitive facts about MPT. Most importantly, the performance of MPT in fact *increases* with N once the number of assets exceeds the number of training observations (T). This finding holds in a variety of settings: in conjunction with popular portfolio regularization methods, in a variety of asset universes, and when T is large or small.

*Hellum is at Copenhagen Business School. Jensen is at Yale School of Management; www.tijensen.com. Kelly is at Yale School of Management, AQR Capital Management, and NBER; www.bryankellyacademic.org. Malamud is at Swiss Finance Institute, EPFL, and CEPR, and is a consultant at AQR. We are grateful for helpful comments from seminar participants at Copenhagen Business School. Hellum and Jensen gratefully acknowledge support from the Center for Big Data in Finance (grant no. DNRF167).

1 Introduction

More than 70 years ago, Markowitz (1952) proposed his solution to selecting an optimal portfolio of risky assets. The environment is simple: There are N risky assets with mean returns $\mu = E[R_t]$ and covariance matrix $\Sigma = Cov(R_t)$, and these are known by the investor. The investor's objective is to select portfolio weights w to earn the highest possible return per unit of risk. The formalization of this problem is

$$\max_w w'\mu - \frac{1}{2}w'\Sigma w,$$

and the "Markowitz" solution for the mean-variance efficient portfolio is

$$w^{\text{eff}*} = \Sigma^{-1}\mu.$$

Likewise, Markowitz showed that an investor wishing to minimize the riskiness of their portfolio (but who must allocate a pre-specified proportion of their wealth to risky assets, e.g., fixing $w'\iota = 1$) achieves minimum variance with the portfolio

$$w^{\text{mvp}*} = \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota}.$$

These insights are the basis of what has come to be known as "modern portfolio theory," or MPT.

The literature has long cast doubt on the practical usefulness of MPT.[1] The crux of the problem is that means and covariances of returns are, of course, *unknown* to the investor and must be estimated. This is only a minor issue when N is just a few assets and T is large. In this case, "plug-in" portfolios that substitute sample moments for the unknown true moments converge to the true optimal portfolios thanks to the law of large numbers,

$$\hat{w}^{\text{eff}} = \hat{\Sigma}^{-1}\hat{\mu} \underset{T\to\infty}{\to} w^{\text{eff}*} \quad \text{and} \quad \hat{w}^{\text{mvp}} = \frac{\hat{\Sigma}^{-1}\iota}{\iota'\hat{\Sigma}^{-1}\iota} \underset{T\to\infty}{\to} w^{\text{mvp}*}. \tag{1}$$

The critical condition for convergence in (1) is that the statistical "complexity" of the estimation problem, $c = N/T$, approaches zero. When N is large relative to T (i.e., $c > 0$), the law of large numbers breaks down, and MPT portfolios can remain disastrously far from

[1] For early examples, see, among others, Cohen and Pogue (1967), Elton and Gruber (1973), and Michaud (1989).

the true optimum. Even a small amount of complexity is problematic. Britten-Jones (1999) shows that when allocating across a mere $N = 11$ country-level equity indices in a sample of $T = 120$ months (so $c = 0.09$), the estimated optimal percentage of wealth to invest in the US equity index has a standard error of 47%!

Our empirical analysis shows that as the number of assets approaches T from below ($c \to 1$), the performance of MPT portfolios collapses, consistent with prior literature highlighting large-N deficiencies of MPT. We then show the counter-intuitive result that, by increasing the number of assets beyond T, MPT performance recovers. It continues to improve as the set of stocks N increases. In other words, in portfolio selection problems with high statistical complexity ($c > 1$), more assets are a blessing rather than a curse.

We organize our findings into a list of new MPT facts in data sets where $N > T$. First, we find that the out-of-sample Sharpe ratio of the Markowitz portfolio increases with N when $N > T$. Second, we find that the out-of-sample variance of the minimum variance portfolio (MVP) decreases with N. Third, we find that in most experiments, the MVP achieves a higher out-of-sample Sharpe ratio than the Markowitz portfolio when N is large. Fourth, we show that the Markowitz portfolio can eventually dominate the MVP when T is sufficiently large (while still requiring $N > T$).

A common approach to improve the empirical performance of MPT is to first regularize portfolio behavior using various forms of Bayesian shrinkage and portfolio constraints.[2] Our fifth fact is that Markowitz and minimum variance portfolios continue to improving with N even when shrinkage and portfolio constraints are also applied.

In an influential paper, DeMiguel et al. (2009a) show that a naively diversified "$1/N$" portfolio achieves consistently higher out-of-sample Sharpe than a wide variety of practical MPT refinements, including shrinkage methods, portfolio constraints, and portfolio combinations. Indeed, our analysis shows that when $c \approx 1$, the $1/N$ portfolio dominates the Markowitz and MVP both in terms of out-of-sample Sharpe ratio and variance. Our sixth fact, however, is that Markowitz and minimum-variance portfolios both dominate the $1/N$ Sharpe ratio when N is sufficiently large, and that their performance relative to $1/N$ increases with N.

These empirical facts are robust. They hold in different asset universes: While we focus our analysis on a large set of anomaly factors from Jensen et al. (2023), the same patterns are

[2]Examples of shrinkage-based portfolio optimization include Jobson and Korkie (1980); Michaud (1989); Pástor (2000); Pástor and Stambaugh (2000); and Ledoit and Wolf (n.d., 2004). Examples of portfolio optimization with constraints (such as excluding short sales) include Chopra et al. (1993); Jagannathan and Ma (2003); Jorion (2003); and DeMiguel et al. (2009b).

found when building portfolios from individual stocks. The same patterns hold for various training sample sizes, even with as little as one year of monthly observations. We see that, holding complexity fixed, the benefits of complexity are larger when T is larger. And they apply at different data frequencies (daily and monthly).

Why does the MVP often achieve a higher Sharpe ratio than the Markowitz portfolio when N is large? The Markowitz portfolio suffers from estimation noise in both sample means and sample covariances. When $N > T$, however, the optimization problem becomes high-dimensional, and—as shown in Didisheim et al. (2024)—the Markowitz solution implicitly applies a form of ridge shrinkage to the covariance matrix. This unintended regularization partially mitigates covariance-estimation noise and can improve out-of-sample risk estimates. What it does not do is regularize the mean. In high dimensions, sample means are extremely noisy, and the Markowitz portfolio loads heavily on this noise. The MVP sidesteps this noise by replacing the sample mean vector with a constant vector, effectively imposing a dogmatic shrinkage of expected returns to a fixed value. When mean estimates are very noisy—as is typical in the $N \gg T$ regime—this mean-shrinkage benefit outweighs the loss of information, leading the MVP to outperform Markowitz in the Sharpe ratio. When T is sufficiently large, the Markowitz portfolio can overtake the MVP: once means are estimated with enough precision, the dogmatic shrinkage in MVP becomes unnecessary and even detrimental.

The fact that MPT performance improves with N reveals something fundamental about the nature of asset markets. Implicit shrinkage alone is not sufficient to produce gains from complexity. It also requires that the benefits of adding each additional asset overwhelm the statistical costs of having to estimate more parameters.

1.1 Literature Review

A large literature explores the challenges of MPT and creative methods for regularizing the sample covariance matrix in order to stabilize the out-of-sample performance of MPT. The prevailing view of practical is aptly summarized by Michaud and Michaud (2007):

> *While theoretically important for modern finance, [mean-variance] optimization's sensitivity to uncertainty in risk-return estimates typically results in an unstable asset management framework, ambiguous portfolio optimality, and poor out-of-sample performance...dominated by equal weighting and [has] essentially no practical investment value.*

Ledoit and Wolf (2003) reach a similarly severe conclusion: "The central message of this paper is that nobody should be using the sample covariance matrix for the purpose of

portfolio optimization." The literature traces the failure of MPT to settings in which N is large relative to T. Brandt (2010) emphasizes that the quality of MPT portfolios "deteriorates drastically with the number of assets held in the portfolio. Intuitively, this is because, as the number of assets increases, the number of unique elements of the return covariance matrix increases at a quadratic rate." DeMiguel et al. (2009a) argue that MPT can only be successful under the conditions that

> "(i) the estimation window is long; (ii) the ex ante (true) Sharpe ratio of the mean-variance efficient portfolio is substantially higher than that of the $1/N$ portfolio; and (iii) the number of assets is small. The first two conditions are intuitive. The reason for the last condition is that a smaller number of assets implies fewer parameters to be estimated and, therefore, less room for estimation error. Moreover, other things being equal, a smaller number of assets makes naive diversification less effective relative to optimal diversification."

The received wisdom for the failure of MPT—that it emerges when there are too many assets N relative to the number of time series observations T—is based on empirical analyses in which N approaches T *from below*. For example, DeMiguel et al. (2009a) use a simulation to demonstrate sample sizes in which MPT is likely to fail, but their analysis never considers values of c above about 0.4. Another such example is Kan and Zhou (2007), who show that MPT performance is decreasing in N, and again the largest c they consider is around 0.4. This focus on $c \leq 1$ is found throughout the MPT literature. While correct in its conclusions about the failure of MPT when $c \leq 1$, the prior literature illustrates only one component of a bigger picture. The central contribution of our paper is to demonstrate the surprisingly strong performance of MPT in complex samples, when N is much larger than T ($c \gg 1$).

Much of the MPT literature shows that Bayesian (or other shrinkage) techniques and portfolio constraints are critical for improving the empirical performance of MPT. Leading examples of include Jobson and Korkie (1980), Jorion (1986), Pástor and Stambaugh (2000), Jagannathan and Ma (2003), Garlappi et al. (2007), and Ledoit and Wolf (2017), among many others. A key contribution of our paper is demonstrating the great benefit of *implicit regularization* that arises in complex portfolio environments and how this resurrects the practical viability of MPT. We also show that further explicit regularization (of the kinds studied in the literature referenced above) is complementary to complexity's implicit regularization and can further boost Markowitz portfolio performance.

Our paper also extends the literature on the virtue of complex statistical models for empirical asset pricing, including Kelly et al. (2024a) and Didisheim et al. (2024), among

others. Those papers focus on the benefits of complexity arising from machine learning designs for predictive regressions and asset pricing models. In contrast, our paper is not about prediction or conditional models. Instead we demonstrate the benefits of statistical complexity in the vanilla *unconditional* portfolio choice problem, absent any role of machine learning. In doing so, we leverage and extend our understanding of implicit shrinkage, limits to learning, and other complex phenomena documented in the prior literature.

2 New Facts in Modern Portfolio Theory

2.1 Data and Methods

In this section, we study how out-of-sample MPT performance depends on the number of assets under consideration, N. Our main analysis uses monthly returns for 153 US factor portfolios from Jensen et al. (2023) (JKP henceforth) from 1972 to 2023. The start date in 1972 ensures that we have non-missing returns for all factors. We use a rolling training window of $T = 12$ training observations to estimate the sample mean vector and variance-covariance matrix. We compute optimal portfolios using $N = 1, ..., 153$, thus c ranges from approximately zero to 13. To ensure that our results are not driven by the ordering of the JKP factors, for each $N < 153$ we recompute optimal portfolios using 100 randomly permuted factor orderings. Our figures and tables report the average portfolio performance across all permutations.

In Sections 2.8 and 2.9, we extend our analysis to other asset cross-sections and longer training windows. For individual stocks, we use monthly equity returns from CRSP and sample from the largest 400 stocks 100 times each month. Similar to the other figures, we report the average portfolio performance across all samples. In Section 2.11, we perform our analysis using daily return data for the JKP factors.

2.2 MPT When $c \leq 1$

We first analyze the performance of MPT when the number of assets N is less than or equal to the number of observations T, reported in Panel A of Figure 1. The left plot shows the out-of-sample Sharpe ratio of the Markowitz portfolio, and the right plot shows the out-of-sample volatility of the MVP. When c is close to zero, the performance of both portfolios improves from increasing the number of assets in the cross-section. Increasing the number of assets allows the Markowitz portfolio to allocate more weight to assets with higher average returns, and it allows both Markowitz and MVP to reduce risk by finding more diversification

opportunities. These gains max out when c is around 0.5. Beyond this, there are not enough observations to precisely estimate the increasing number of parameters. As a result, portfolio estimates are dominated by noise. When $c = 1$, the Sharpe ratio of the Markowitz portfolio falls close to zero because it fails to identify the relatively high return assets and because noisy weight estimates drive up the portfolio's variance. The MVP is so severely corrupted by noise that its volatility exceeds the volatility of individual assets (the left-most point in the figure corresponds to a single-asset portfolio).

In the absence of estimation noise, we would expect Markowitz and MVP performance to improve (weakly) with N as it has more diversification opportunities at its disposal. Thus, when viewing Figure 1, it is useful to keep in mind that the cratering performance of MPT is *net* of the diversification benefits from larger N.

These findings are neither new nor surprising. Instead, Panel A of Figure 1 establishes a baseline understanding of the challenges faced by MPT when using even moderate numbers of assets. The failure of MPT when N approaches T has motivated the decades-long search for portfolio optimization refinements (shrinkage, constraints, and so forth) as articulated in our literature review.

2.3 MPT When $c > 1$

Next, we analyze out-of-sample MPT performance when the number of assets is allowed to exceed the number of observations (i.e., $c \geq 1$). In this case, the inverse of the sample covariance $\hat{\Sigma}$ in equation (1) is undefined, thus we instead compute the inverse covariance using the Moore-Penrose pseudo-inverse, $\hat{\Sigma}^{+}$.

Panel B of Figure 1 extends the portfolio analysis to allow the number of assets to exceed the number of training observations ($c > 1$). In this case, the Sharpe ratio of the Markowitz portfolio is increasing in N, and does not flatten out until the ratio of assets to data points reaches approximately $c = 4$. Likewise, the volatility of the MVP is decreasing in N, flattening around $c = 2$. Evidently, the failures of MPT emphasized in prior literature are not about large N, but are instead about $N \approx T$. In fact, a higher N is a virtue when it comes to MPT, as long as N is larger than the number of training observations.

2.4 The Role of Shrinkage

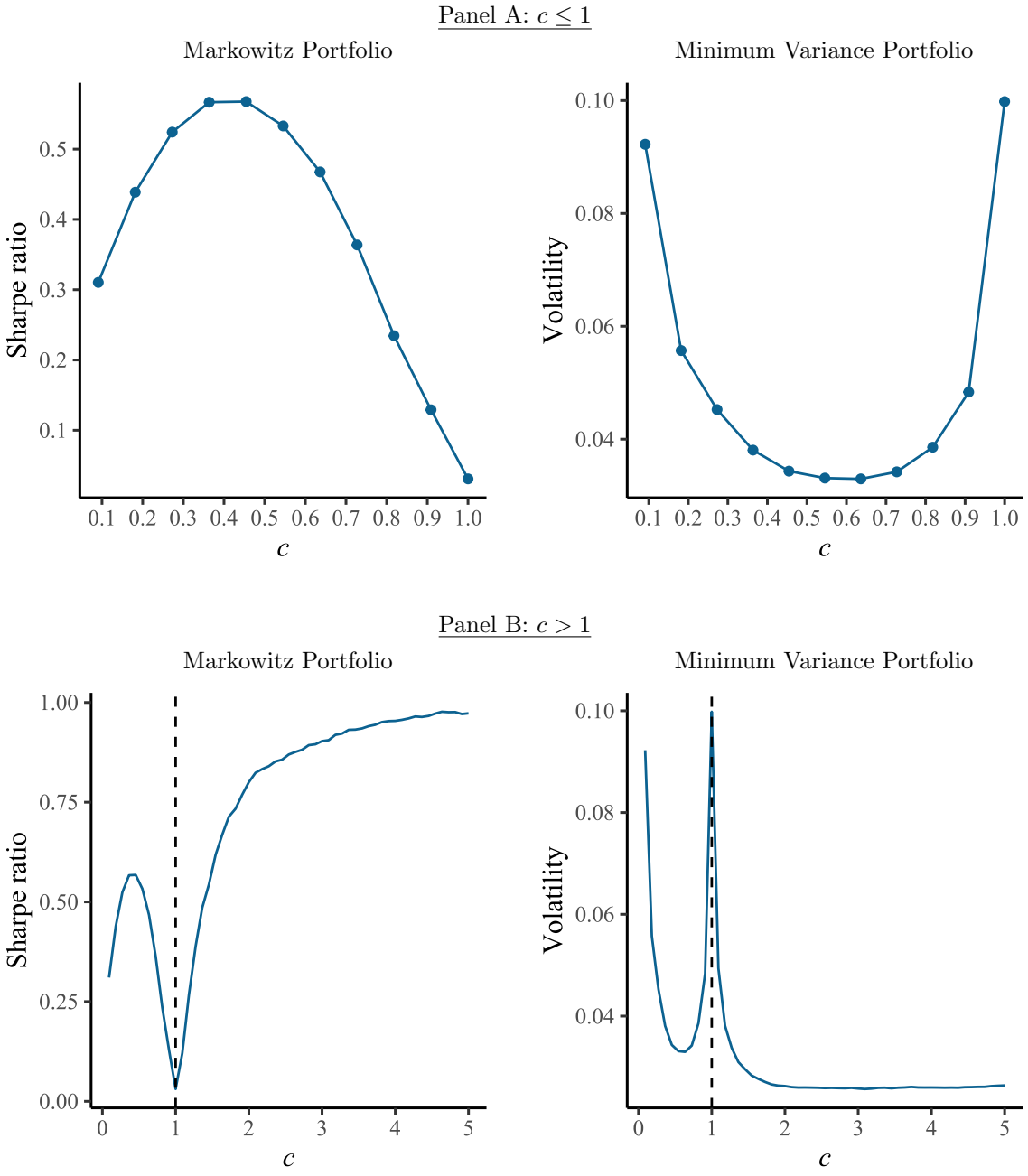

Figure 1: Out-of-sample MPT Performance.

The figure shows the performance of the Markowitz and minimum variance portfolio based on a varying number of equity factors. Specifically, we randomly select N factors and, each month, create the portfolios based on their sample mean and variance-covariance matrix estimated over the past $T = 12$ months. For every N, we repeat the random selection process 100 times and average the performance statistics across permutations. The x-axis shows complexity defined as $c = N/(T-1)$. The y-axis for figures on the left show the annualized out-of-sample Sharpe ratio of the Markowitz portfolio. The y-axis on the right shows the annualized out-of-sample volatility of the minimum variance portfolio. Performance is calculated on data from 1972 to 2023.

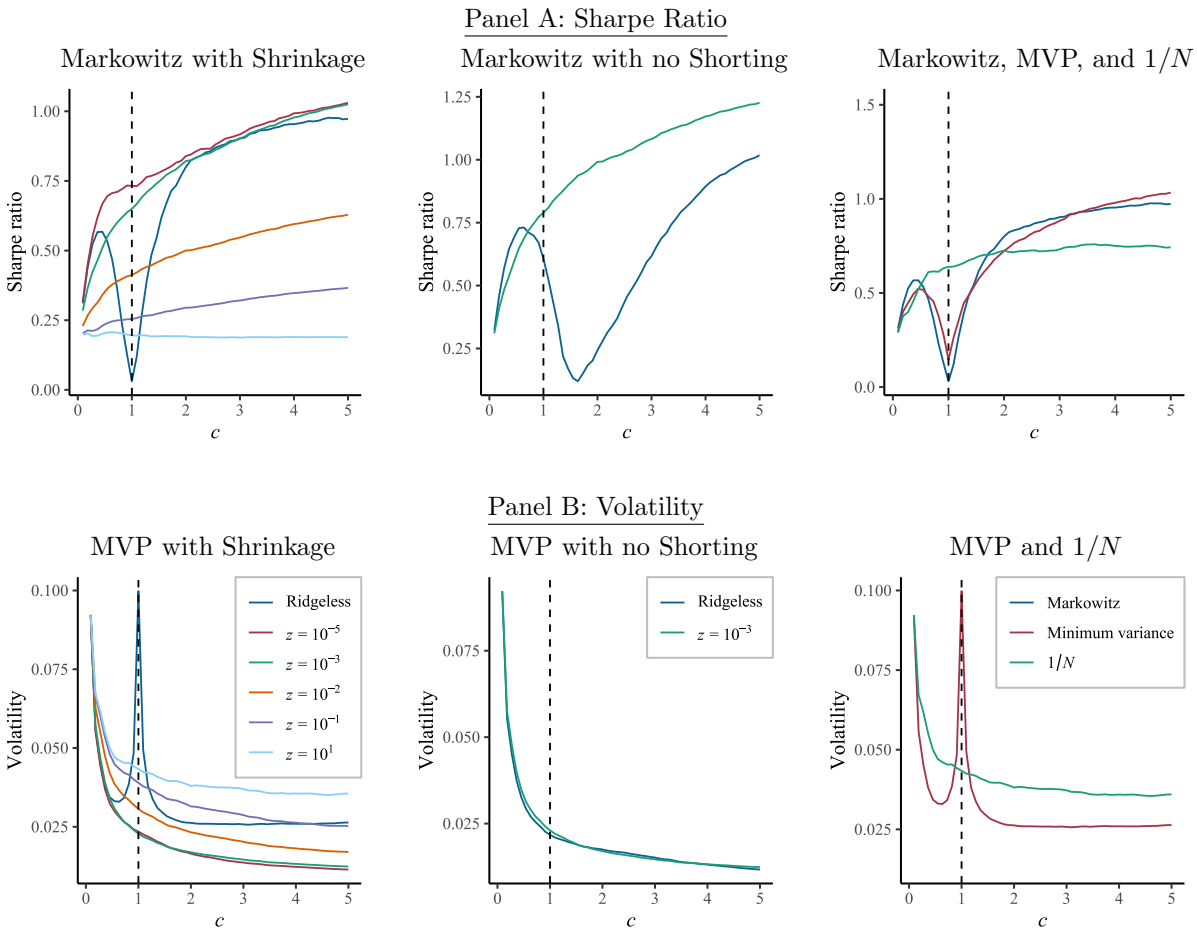

Figure 2: MPT Performance with Shrinkage.

Panel A shows as a function of complexity, the effects on Sharpe ratio of shrinkage and a no-shorting constraint on the Markowitz portfolio. Lastly, it compares the Sharpe ratio of Markowitz, MVP, and $1/N$. Panel B shows the effects on the volatility of the MVP and compares the volatility of MVP and $1/N$. For ridgeless shrinkage we invert the covariance matrix using the pseudo-inverse whenever complexity exceeds one. In all other cases we add an explicit ridge penalty of z.

A large body of literature proposes regularization methods to improve the empirical performance of MPT portfolios. One major strand of this literature advocates covariance shrinkage to improve out-of-sample portfolio performance. Thus, we investigate how covariance shrinkage alters the behavior of MPT in high complexity settings and affects our conclusions from Section 2.3. We focus our analysis on ridge shrinkage, which is among the oldest, most well-known, and most commonly employed covariance regularization methods.[3] As a counterpart to the unregularized portfolios in (1), ridge regularized MPT portfolios with shrinkage parameter $z > 0$ are given by

$$\hat{w}^{\text{eff}}(z) = (zI + \hat{\Sigma})^{-1}\hat{\mu} \quad \text{and} \quad \hat{w}^{\text{mvp}}(z) = (zI + \hat{\Sigma})^{-1}\iota. \tag{2}$$

Interestingly, ridge regularization has a close connection with the high complexity MPT portfolios in Section 2.3 because the pseudo-inverse has an equivalent interpretation as infinitesimal ridge regularization. In particular, we can rewrite the sample MPT portfolios as

$$\hat{w}^{\text{eff}} = \hat{\Sigma}^{+}\hat{\mu} = \lim_{z\to 0+}(zI + \hat{\Sigma})^{-1}\hat{\mu} \quad \text{and} \quad \hat{w}^{\text{mvp}} = \hat{\Sigma}^{+}\iota = \lim_{z\to 0+}(zI + \hat{\Sigma})^{-1}\iota. \tag{3}$$

When $c \leq 1$, the limit is exactly zero shrinkage and corresponds to the sample portfolio using the standard matrix inverse. When $c > 1$, the standard inverse does not exist, so the limit in (3) is discontinuous. In this case, the pseudo-inverse can be defined as a limit as ridge shrinkage approaches zero—i.e., it adds just enough identity matrix to the covariance so that their sum is invertible. Following Kelly et al. (2024a), we refer to MPT portfolios that use infinitesimal shrinkage as "ridgeless" portfolios, and we denote these by $\hat{w}^{\text{eff}}(0^+)$ and $\hat{w}^{\text{mvp}}(0^+)$. Thus, the ridge portfolios in (2) are simply an extension of the analysis in Figure **??** to incorporate non-trivial positive shrinkage.

Panel A of Figure 2 illustrates how the performance of ridge-regularized MPT portfolios changes with the number of assets. There are three key findings from this figure. First, when $c > 1$, the performance of MPT portfolios is increasing with N for *any* amount of shrinkage (including the ridgeless case with shrinkage approaching zero). Second, the primary benefit of non-trivial ridge shrinkage is that it robustifies portfolio performance in the neighborhood of $c = 1$. Third, non-trivial ridge shrinkage further amplifies the benefits of very large N. While we find that the out-of-sample performance of the ridgeless Markowitz portfolio

[3]There is a wide gamut of covariance shrinkage methods proposed in the literature, and analyzing how each of these behaves for different degrees of statistical complexity is an interesting question but beyond our scope.

flattens around $c = 4$ and the performance of the ridgeless MVP flattens around $c = 2$, heavier shrinkage (e.g., $z = 10^{-3}$) extends the large N performance benefits even further and pushes the Sharpe ratio above 1.0 (and higher than the ridgeless case). Complexity gains from non-trivial shrinkage are even larger for MVP portfolios. For example, a ridge parameter of $z = 10^{-3}$ reduces MVP volatility by about half at $c = 5$, and the volatility curve continues to improve beyond $c = 5$.

2.5 Minimum Variance Can Maximize Sharpe Ratio

Thus far, we have evaluated the Markowitz portfolio in terms of its out-of-sample Sharpe ratio and the MVP in terms of its out-of-sample volatility. We show that not only is the volatility of the MVP improving with N, but so is its Sharpe ratio (see Figure 6 of the Internet Appendix). So much, in fact, that the MVP Sharpe ratio dominates that of the Markowitz portfolio when c is large (left panel). Adding non-trivial ridge shrinkage (center and right panels) leads to even starker gains in MVP Sharpe ratio, both in absolute terms and relative to Markowitz.

Since Markowitz and MVP share the same inverse covariance, the difference in their Sharpe ratios traces to their treatment of mean returns. Markowitz uses sample means, while the MVP can be interpreted as a Markowitz portfolio that constrains all assets to have the same mean. The MVP can thus be viewed as a dogmatic form of mean shrinkage. In the training samples reported in Figure 6, the noisiness of sample means stifles the Markowitz performance so much that the extreme mean shrinkage of MVP is preferred. With larger training samples, the performance of Markowitz and MVP may reverse, as we show next. And, while outside the scope of this paper, our findings suggest that intermediate mean shrinkage between that of Markowitz (zero mean shrinkage) and MVP (100% shrinkage to a constant mean) can further improve portfolio performance (see e.g. Pedersen et al., 2021).

2.6 The Role of Constraints

Another literature proposes portfolio constraints as a means of improving the empirical performance of MPT. The leading example is Jagannathan and Ma (2003) who show that imposing a non-negativity constraint on the portfolio optimization problem improves its out-of-sample performance.[4] To understand how complexity interacts with portfolio

[4]For individual assets such as stocks, this is equivalent to a no-shorting constraint. But when the base assets are themselves long-short factor portfolios, the non-negativity constraint applies at the factor level, which still allows stock-level shorting.

constraints, we modify our high-dimensional MPT problems, restricting weights to be non-negative. Panel B of Figure 2 shows that even in the presence of a non-negativity constraint, MPT performance improves when the number of assets becomes large. This is true in the "ridgeless" and the non-trivial ridge cases. With $z = 10^{-3}$, the constrained complex Markowitz Sharpe ratio rises by nearly 25% compared to the unconstrained complex model in Figure 2.

Interestingly, the collapse of the ridgeless Markowitz Sharpe ratio no longer occurs exactly at $c = 1$ but closer to $c = 1.5$. In the unconstrained model, the Markowitz portfolio achieves an infinite in-sample Sharpe ratio at $c = 1$; in other words, it "interpolates" the training returns when $N \geq T$ (see, e.g. Kelly and Malamud, 2025). But with a non-negativity constraint, the Markowitz solution (typically) requires a larger number of assets to interpolate, which is why the performance dip occurs at a higher value of c.

2.7 Resilience of $1/N$

DeMiguel et al. (2009a) masterfully demonstrate failures of applied MPT by showing that a naively diversified portfolio that equally weights assets, the "$1/N$" portfolio, dominates a wide variety of optimal portfolios, including those employing various forms of regularization. In the previous section, we interpreted the MVP as a Markowitz portfolio with dogmatic mean shrinkage. Likewise, the $1/N$ portfolio can be interpreted as a Markowitz portfolio with the same dogmatic mean shrinkage as MVP and with dogmatic variance shrinkage to the identity matrix. We now examine the role of this shrinkage design for large-N portfolios.

Panel C of Figure 2 compares the performance of the $1/N$ with MPT portfolios in terms of out-of-sample Sharpe ratio (top panels) and volatility (bottom panels). Indeed, the $1/N$ portfolio is extremely resilient as the number of assets grows, achieving an attractive Sharpe ratio and low volatility. When compared to ridgeless MPT (left panels), $1/N$ far exceeds the performance of the Markowitz portfolio and MVP around $c = 1$. Evidently, it is the breakdown of MPT when $N \approx T$ that drives the outperformance of $1/N$. When $c \ll 1$ or $c \gg 1$, MPT portfolios are superior to $1/N$ in terms of both Sharpe ratio and volatility. For non-trivial positive shrinkage of $z = 10^{-3}$, MPT outperforms $1/N$ by an substantially larger margin and is even superior to $1/N$ in the critical region $c \approx 1$ (shown in Internet Appendix Figure 7 due to space constraints).

2.8 The Role of T

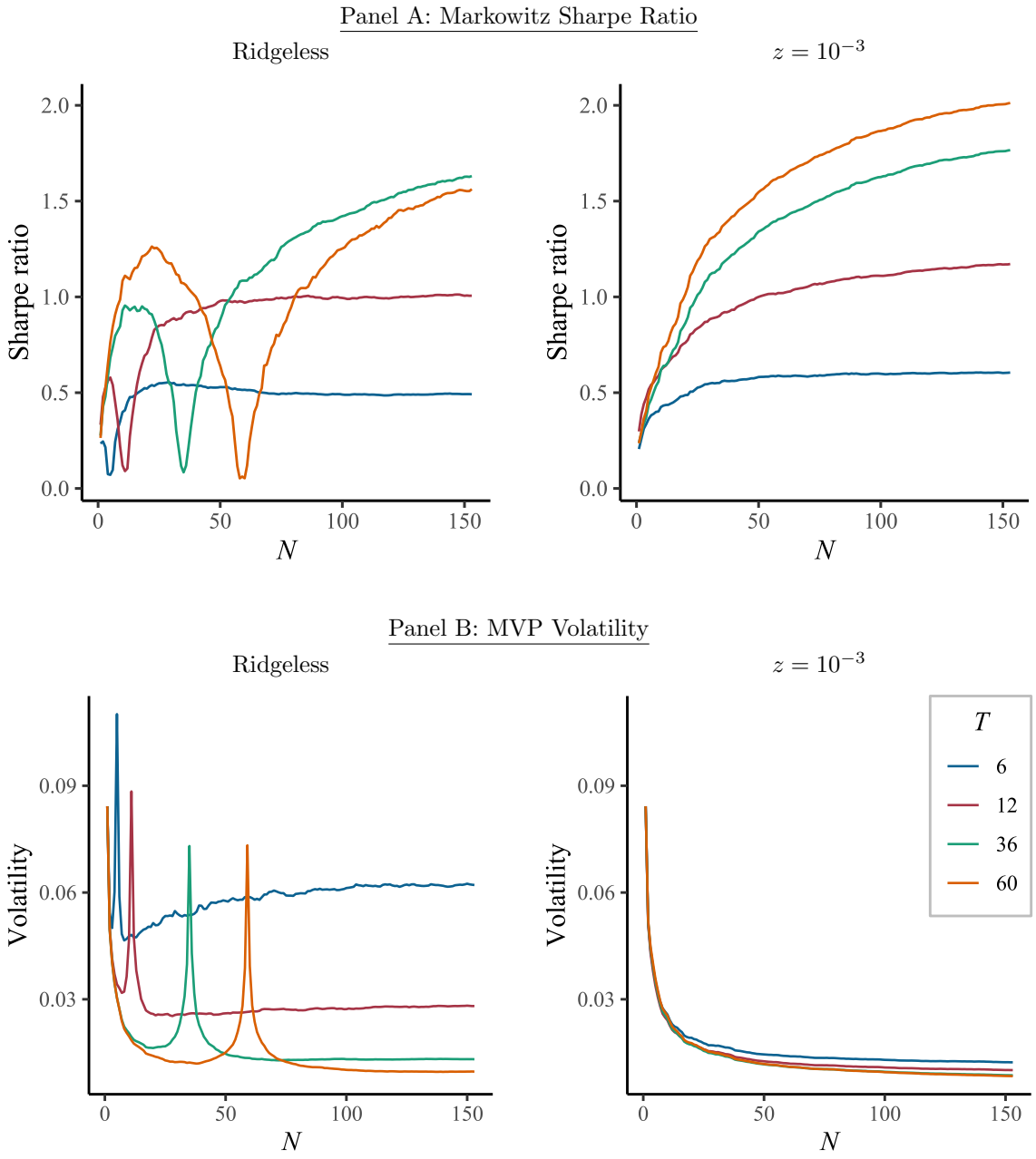

Figure 3: MPT Effects of Sample Size (T).

Panels A and B show the out-of-sample Sharpe ratio of the Markowitz portfolio and the volatility of minimum variance portfolio, respectively, as we vary the number of time periods T in the rolling training window. Each point is averaged across 25 random permutations.

In our analysis so far, we have fixed the sample size to a mere $T = 12$ monthly training observations. We now explore how our previous conclusions are affected by varying the

number of training observations. We repeat the preceding analyses that evaluate portfolio performance as we gradually increase the number of assets in the cross-section.

Panel A of Figure 3 plots out-of-sample Sharpe ratios for MPT portfolios. Because our analysis now varies both N and T, we use N to delineate the x-axis and use different lines for each choice of $T = 6$, 12, 36, and 120 months. We report Sharpe ratios for both Markowitz and minimum variance portfolios (left and right columns, respectively) and for the ridgeless case (top row) and non-trivial ridge ($z = 10^{-3}$, middle row).

Panel A illustrates four main findings. First, there are portfolio benefits to increasing cross-section size N for any sample size T. Second, the scale of the Sharpe ratio curves is increasing in T. In other words, larger sample sizes magnify the benefits of large N MPT portfolios. While this is reminiscent of traditional asymptotic theory (fixing N and increasing T), larger T cannot avoid MPT collapse in the critical region $c \approx 1$. Third, with sufficiently large T, the performance of the high complexity Markowitz portfolio eventually exceeds that of the MVP, which occurs in Figure 3 for $T \geq 36$. Fourth, with non-trivial ridge shrinkage (bottom row of Figure 3), MPT performance strictly improves in both N and T. Finally, Panel B reports the effect of sample size T on volatility of the MVP. These results are consistent with the conclusions of Panel A: Out-of-sample MVP volatility improves as T increases.

2.9 MPT and Individual Stocks

Our analysis to this point has focused on 153 factor portfolios as base assets. In this section, we examine the behavior of large-N MPT using individual stocks. Figure 8 in the Internet Appendix shows how the MPT portfolios perform as a function of complexity. As shown in DeMiguel et al. (2009a), the $1/N$ portfolio performs well out-of-sample, achieving a higher Sharpe ratio than both the Markowitz portfolio and MVP when $T = 12$. However, when the estimation window is large enough—e.g. at $T = 60$—the MVP outperforms $1/N$ in the high-complexity setting $N \gg T$. The Markowitz portfolio still struggles even with this longer estimation window, in contrast to the results that used factor portfolios. This is consistent with individual stock returns being less stationary than factor returns—the historic average factor returns are informative about the expected factor returns, while the average stock returns are bad estimates of expected stock returns going forward. The MVP sidesteps this issue by dogmatically shrinking the means to a constant.

2.10 Large N Risk Modeling

Our analysis so far has focused on the performance of optimized portfolios. The strong performance of MPT in high complexity settings traces in large part to the stable behavior of the sample covariance. In this section, we analyze another application of the high complexity covariance matrix—its ability to create hedging portfolios. In other words, the sample covariance plays a broader role as a risk model. We now ask whether high complexity is as beneficial for out-of-sample risk modeling as it is for portfolio performance.

We are interested in hedging specific risks that are not spanned by the JKP factors. In particular, we attempt to hedge changes in the 10y-1y term spread, crude oil returns, and VIX returns using the FRED-MD data. The hedging portfolio is estimated by the coefficients of the regression

$$y_t = w' F_t + \varepsilon_t, \tag{4}$$

where y_t is the risk of interest and F_t is the vector of factor returns. In vector form

$$y = Fw + \varepsilon. \tag{5}$$

The hedging portfolio is then

$$\hat{w}^{\text{hedge}} = (F'F)^{+} y. \tag{6}$$

We estimate the hedging portfolio using a 12-month rolling window.

Figure 4 shows the realized correlations between the hedging portfolios and their targets as a function of their statistical complexity, $c = N/T$. We again see the familiar virtue of complexity curve—performance is hump-shaped in the $c < 1$ region, collapsing at $c = 1$, and increasing in the $c > 1$ region.

These empirical results can be understood in light of Kelly et al. (2024b). The more included assets, the better they can capture complex time-series patterns of the risk of interest. This ability is more important than the higher statistical costs associated with more assets, leading to more assets increasing the out-of-sample correlation.

For hedging portfolio problems with high statistical complexity, more assets are once again a blessing rather than a curse.

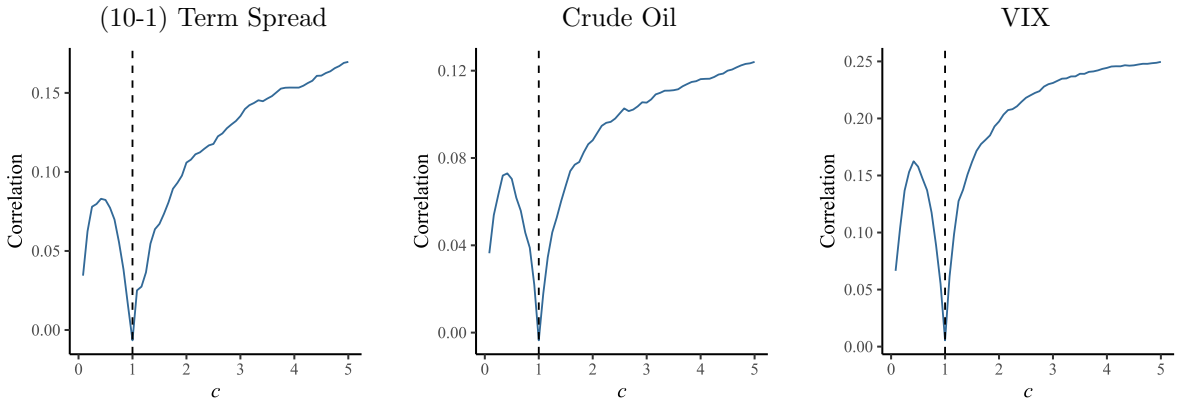

Figure 4: Correlations of Hedging Portfolios.

The figure shows correlations of different macro variables and their hedging portfolios. Specifically, we randomly select N JKP factors and, each month, create the portfolio with the highest correlation to the macro variable in question over the past $T = 12$ months. For every N, we repeat the random selection process 100 times, and the performance statistics are averaged across these permutations. The x-coordinate shows the complexity, defined as $c = N/T$. The correlations are calculated using data from 1977 to 2023.

2.11 Data Frequency and Non-stationarity

So far, we use monthly observations to form monthly portfolios. Our results and discussion highlight that the gains from complexity comes through the covariance matrix. Since covariances can be estimated with higher frequency data, we now explore the effects of using daily data to construct MPT portfolios. With the exception of the data frequency, our analysis is otherwise unchanged (i.e., we continue to rebalance portfolios monthly, maintain the same sample horizons and assets, and so on).

In Figure 5, we plot the performance of Markowitz, MVP, and $1/N$ for $N = 50, 100, 153$, while varying T (now measured in days). Consistent with the findings in Section 2.8, we see that the performance of Markowitz increases with T. However, the performance of MVP flattens out quickly with T—in the low complexity setting—and never recovers to its level under $T < N$. This is consistent with the covariance matrix being dynamic, making higher T less useful for estimation. This becomes more problematic the higher the N. On the other hand, if the mean returns are not as dynamic, they can be better estimated with a longer time-series, driving the improvements in performance for the Markowitz portfolio.

A large literature models dynamic "spot" covariance using high frequency data (e.g. Barndorff-Nielsen and Shephard, 2004; Bollerslev et al., 2018; Andersen et al., 2006; Hautsch et al., 2012). Estimators typically take some form of a sample covariance estimated in a short window from even shorter-horizon returns (e.g., daily "realized" covariances estimated from

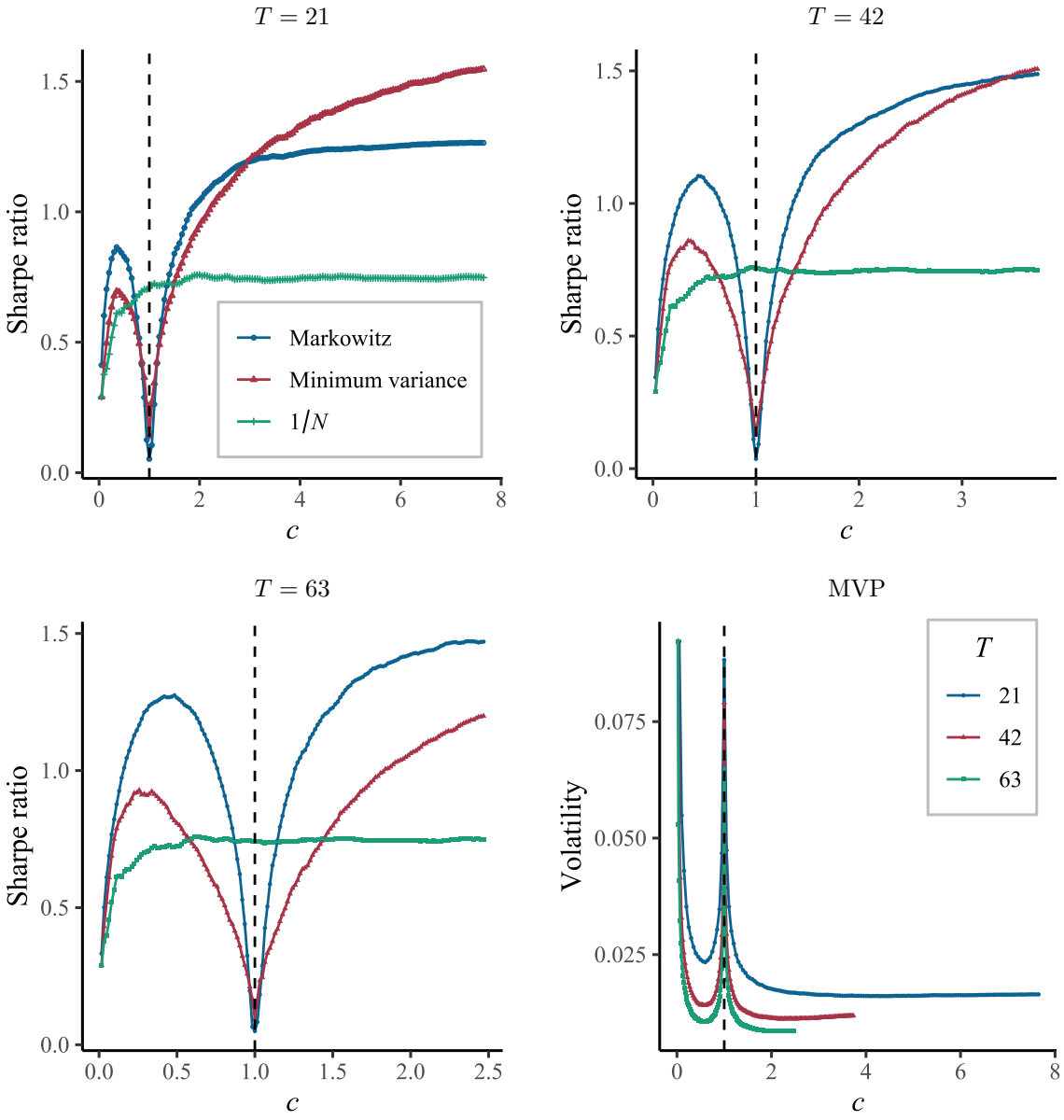

Figure 5: MPT and Daily Data.

This figure shows the performance of Markowitz, MVP, and $1/N$, where portfolio weights are estimated using daily returns. We randomly select N JKP factors and, each month, create the portfolios based on their sample mean and covariance matrix estimated using daily data for the past T days. The portfolios are rebalanced monthly. For every N, we repeat the random selection 100 times, and the performance statistics are averaged across these permutations. The y-coordinate shows the annualized Sharpe ratio, except for the last graph, which plots the volatility of the MVP across complexity for various T. The performance statistics are based on out-of-sample data from 1977 to 2023.

5-minute returns within the day). While such high frequency applications of complex MPT are beyond the scope of this paper, the benefits of complexity can be especially valuable in

such short estimation windows and suggest a valuable direction for future work on modeling high-dimensional conditional covariances (see e.g. Bauwens et al., 2006; Engle et al., 2019).

3 Theory

We use $X \approx Y$ to denote that $X - Y \to 0$ in probability.

Assumption 1 *Suppose that* $R_t = \mu + \Sigma^{1/2}\varepsilon_t$, *where* $\varepsilon_t = (\varepsilon_{i,t})_{i=1}^N$ *are i.i.d. with* $E[\varepsilon_{i,t}] = 0$, $E[\varepsilon_{i,t}^2] = 1$, $E[\varepsilon_{i,t}^{4+\varepsilon}] < \infty$ *for some* $\varepsilon > 0$. *We assume that* $\Sigma = \Sigma(N)$ *is uniformly bounded, and that its eigenvalue distribution weakly converges to a non-degenerate distribution* $H(dx)$ *on* $\mathbb{R}_+$ *as* $N \to \infty$. *Let let* $\mu(N_1), \Sigma(N_1)$ *be the mean and covariance of the first* N_1 *stocks. We assume that, in the limit as* $N, N_1 \to \infty, N_1/N \to q$, *the eigenvalue distribution of* $\Sigma(N_1)$ *converges to the same limit as that of* Σ. *Furthermore, given the eigenvalue decomposition* $\Sigma(N_1) = U(N_1)D(N_1)U(N_1)'$, *we have that the distribution of the coordinates of the vector* $(U(N_1)'\mu(N_1))^2/\|\mu(N_1)\|^2$ *weakly converges to the same limit that is independent of* q *ans is absolutely continuous with respect to* $H(dx)$, *while* $\lim \frac{\|\mu(N_1)\|^2}{\|\mu\|^2} = q$, *for any* $q > 0$.

Theorem 1 (Ridgeless Limit) *Suppose we sample uniformly without repetion* N_1 *stocks from* $[1, N]$. *In the limit as* $N_1, T \to \infty$, $N_1/T \to c > 0$ *and* $z \to 0+$, *we have:*

- *for* $c < 1$,

$$\frac{E[R'_{T+1}\hat{w}^{eff}]^2}{\mathrm{Var}[R'_{T+1}\hat{w}^{eff}]} \approx c(1-c)\,\mu'\Sigma^{-1}\mu\frac{1}{\frac{N}{T}+\frac{N^2}{T^2}} \tag{7}$$

  Consistent with Figure 1, maximum is achieved for $c = 0.5$. *Similarly,*

$$N_1\mathrm{Var}[R'_{T+1}\hat{w}^{mvp}] \approx \frac{1}{1-c}\frac{1}{\iota'\Sigma(N_1)^{-1}\iota/N_1}\,. \tag{8}$$

- *For large* c,

$$\frac{E[R'_{T+1}\hat{w}^{eff}]^2}{\mathrm{Var}[R'_{T+1}\hat{w}^{eff}]} \approx \frac{\|\mu\|^4}{T^{-1}\,\mathrm{tr}(\Sigma^2) \;+\; \mu'\Sigma\mu} \;+\; O(c^{-2}) \tag{9}$$

Thus, the Sharpe Ratio for large c *dominates that for the best classical one* $(c = 0.5)$ *as long*

as N is large enough. Similarly,

$$N\text{Var}[R'_{T+1}\hat{w}^{mvp}] \;\approx\; \frac{\iota'\Sigma\iota}{N} \;+\; O(c^{-1}) \tag{10}$$

for large c.

3.1 Economic Interpretation

Theorem 1 highlights a sharp phase transition in portfolio performance as model complexity increases. In particular, the Sharpe ratio exhibits a double-descent pattern as the ratio $c = N_1/T$ grows. In the classical regime ($N_1 < T$), portfolio performance follows the familiar bias–variance tradeoff. As the number of assets increases, diversification initially improves performance, but estimation error eventually dominates. Consequently, the Sharpe ratio is hump-shaped and, remarkably, symmetric around $c = 0.5$: it increases for $N_1 < 0.5T$ and declines to zero as N_1 approaches T. Even for $N = T$, the loss from estimation error is substantial. The highest achievable squared Sharpe ratio in the classical regime is roughly eight times smaller than the infeasible population benchmark $\mu'\Sigma^{-1}\mu$.

Beyond the interpolation threshold ($c > 1$), the behavior changes fundamentally. Once the model becomes sufficiently over-parameterized, portfolio performance improves again. This phenomenon corresponds to the *virtue of complexity* documented in Kelly et al. (2024b); Didisheim et al. (2024): the Sharpe ratio in the high-complexity regime exceeds the best level attainable in the classical regime.

The mechanism behind this improvement is implicit regularization. When $c > 1$, the optimization problem effectively imposes a strong penalty on the covariance matrix (Didisheim et al. (2024); Chernov et al. (2025)). As a result, estimation risk grows only linearly in N, whereas in the classical regime it grows quadratically. This implicit penalization offsets the otherwise increasing estimation variance and eliminates the classical bias–variance tradeoff.

A similar pattern arises for the minimum-variance portfolio. In the classical regime, diversification $\iota'\Sigma(N_1)^{-1}\iota$ initially offsets the overfitting factor $1/(1-c)$, but overfitting eventually dominates, causing the variance to explode as $c \uparrow 1$. In contrast, in the high-complexity regime, implicit regularization prevents this explosion. Instead, for typical covariance configurations, the variance collapses toward zero as $N \to \infty$. Indeed, since

$\|\iota\|^2 = N$,

$$\frac{\iota'\Sigma\iota}{N^2} \leq \frac{\|\Sigma\|}{N} \to 0.$$

Taken together, these results suggest that the virtue of complexity is a structural feature of high-dimensional portfolio optimization rather than an artefact of a particular dataset. In sufficiently high dimensions, increasing the number of assets can improve portfolio performance not despite, but precisely because of the implicit regularization induced by over-parameterization.

4 Conclusion

We present an empirical resurrection of Modern Portfolio Theory (MPT) when the number of assets is large. We show in a variety of settings that the performance of MPT increases with the number of assets once enough assets are included. We analyze the role of complexity, shrinkage, constraints, and the length of the estimation window. We show that performance gains come through implicit bias in the covariance matrix and not mean estimates. These results are only consistent with the cross-section of returns being governed by a high-dimensional factor structure.

References

Andersen, Torben G, Tim Bollerslev, Francis X Diebold, and Ginger Wu, "Realized Beta: Persistence and Predictability," *Advances in Econometrics*, 2006, pp. 1–39.

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

Barndorff-Nielsen, Ole E and Neil Shephard, "Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics," *Econometrica*, 2004, *72* (3), 885–925.

Bauwens, Luc, Sébastien Laurent, and Jeroen VK Rombouts, "Multivariate GARCH models: a survey," *Journal of applied econometrics*, 2006, *21* (1), 79–109.

Bollerslev, Tim, Andrew J Patton, and Rogier Quaedvlieg, "Modeling and forecasting (un) reliable realized covariances for more reliable financial decisions," *Journal of Econometrics*, 2018, *207* (1), 71–91.

Brandt, Michael W, "Portfolio choice problems," in "Handbook of financial econometrics: Tools and techniques," Elsevier, 2010, pp. 269–336.

Britten-Jones, Mark, "The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights," *Journal of Finance*, 1999, *54* (2), 655–670.

Chen, Zhimin, Bryan Kelly, and Semyon Malamud, "Limits To (Machine) Learning," *arXiv preprint arXiv:2512.12735*, 2025.

Chernov, Mikhail, Bryan T Kelly, Semyon Malamud, and Johannes Schwab, "A test of the efficiency of a given portfolio in high dimensions," Technical Report, National Bureau of Economic Research 2025.

Chopra, Vijay K, Chris R Hensel, and Andrew L Turner, "Massaging mean-variance inputs: returns from alternative global investment strategies in the 1980s," *Management Science*, 1993, *39* (7), 845–855.

Cohen, Kalman J and Jerry A Pogue, "An empirical evaluation of alternative portfolio-selection models," *The Journal of Business*, 1967, *40* (2), 166–193.

DeMiguel, Victor, 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.

___ , ___ , **Francisco J Nogales, and Raman Uppal**, "A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms," *Management science*, 2009, *55* (5), 798–812.

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

Elton, Edwin J and Martin J Gruber, "Estimating the dependence structure of share prices–implications for portfolio selection," *The Journal of Finance*, 1973, *28* (5), 1203–1232.

Engle, Robert F, Olivier Ledoit, and Michael Wolf, "Large dynamic covariance matrices," *Journal of Business & Economic Statistics*, 2019, *37* (2), 363–375.

Garlappi, Lorenzo, Raman Uppal, and Tan Wang, "Portfolio selection with parameter and model uncertainty: A multi-prior approach," *The Review of Financial Studies*, 2007, *20* (1), 41–81.

Hautsch, Nikolaus, Lada M Kyj, and Roel CA Oomen, "A blocking and regularization approach to high-dimensional realized covariance estimation," *Journal of Applied Econometrics*, 2012, *27* (4), 625–645.

Jagannathan, Ravi and Tongshu Ma, "Risk reduction in large portfolios: Why imposing the wrong constraints helps," *The journal of finance*, 2003, *58* (4), 1651–1683.

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

Jobson, J David and Bob Korkie, "Estimation for Markowitz efficient portfolios," *Journal of the American Statistical Association*, 1980, *75* (371), 544–554.

Jorion, Philippe, "Bayes-Stein estimation for portfolio analysis," *Journal of Financial and Quantitative analysis*, 1986, *21* (3), 279–292.

___ , "Portfolio optimization with tracking-error constraints," *Financial Analysts Journal*, 2003, *59* (5), 70–82.

Kan, Raymond and Guofu Zhou, "Optimal portfolio choice with parameter uncertainty," *Journal of Financial and Quantitative Analysis*, 2007, *42* (3), 621–656.

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

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

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

Knowles, Antti and Jun Yin, "Anisotropic local laws for random matrices," *Probability Theory and Related Fields*, 2017, *169* (1), 257–352.

Ledoit, O and M Wolf, "Honey, I shrunk the sample covariance matrix, 2004," *J. Portf. Management*, *31*, 110.

Ledoit, Olivier and Michael Wolf, "Honey, I shrunk the sample covariance matrix," *UPF economics and business working paper*, 2003, (691).

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

___ **and** ___ , "Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks," *The Review of Financial Studies*, 2017, *30* (12), 4349–4388.

Markowitz, Harry, "Portfolio Selection," *The Journal of Finance*, 1952, *7* (1), 77–91.

Michaud, Richard O, "The Markowitz optimization enigma: Is 'optimized'optimal?," *Financial analysts journal*, 1989, *45* (1), 31–42.

___ **and Robert Michaud**, "Estimation error and portfolio optimization: a resampling solution," *Available at SSRN 2658657*, 2007.

Pástor, L'uboš, "Portfolio selection and asset pricing models," *The Journal of Finance*, 2000, *55* (1), 179–223.

___ **and Robert F Stambaugh**, "Comparing asset pricing models: an investment perspective," *Journal of Financial Economics*, 2000, *56* (3), 335–381.

Pedersen, Lasse Heje, Abhilash Babu, and Ari Levine, "Enhanced portfolio optimization," *Financial Analysts Journal*, 2021, *77* (2), 124–151.

Internet Appendix

A Additional Results

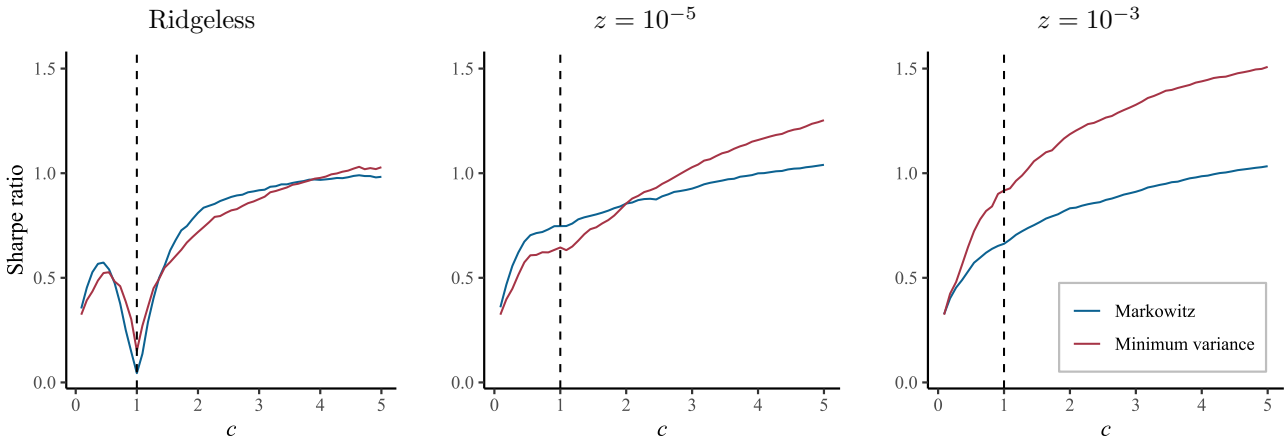

Figure 6: Sharpe Ratio Comparison of Markowitz and Minimum Variance Portfolios.

The figure shows the Sharpe ratio of Markowitz and minimum variance as a function of shrinkage and complexity.

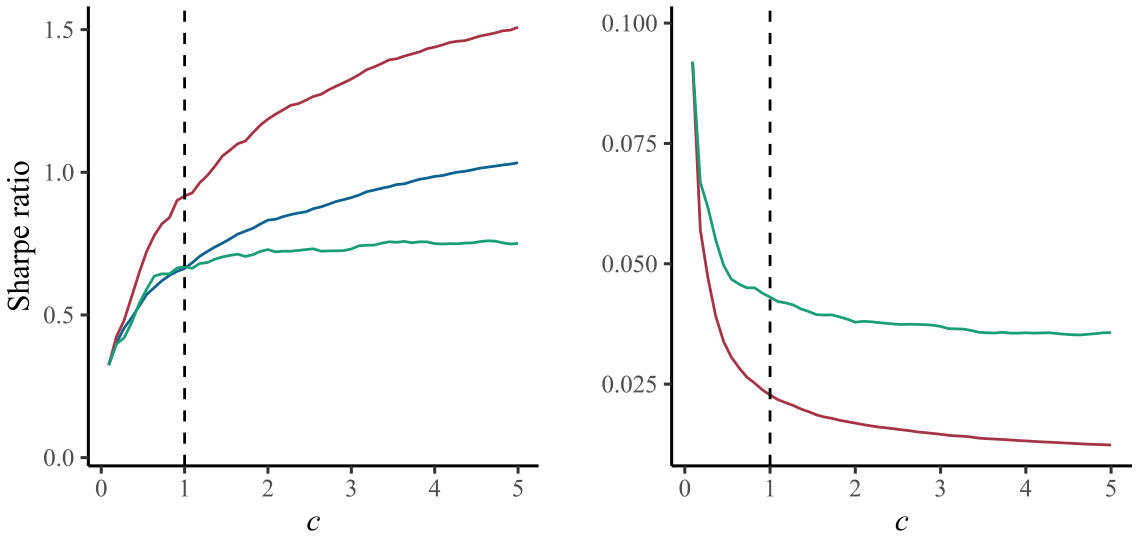

Figure 7: Complex MPT and $1/N$ With Ridge Shrinkage.

The figure shows the performance of Markowitz, MVP, and $1/N$ for MPT with ridge shrinkage of $z = 10^{-3}$.

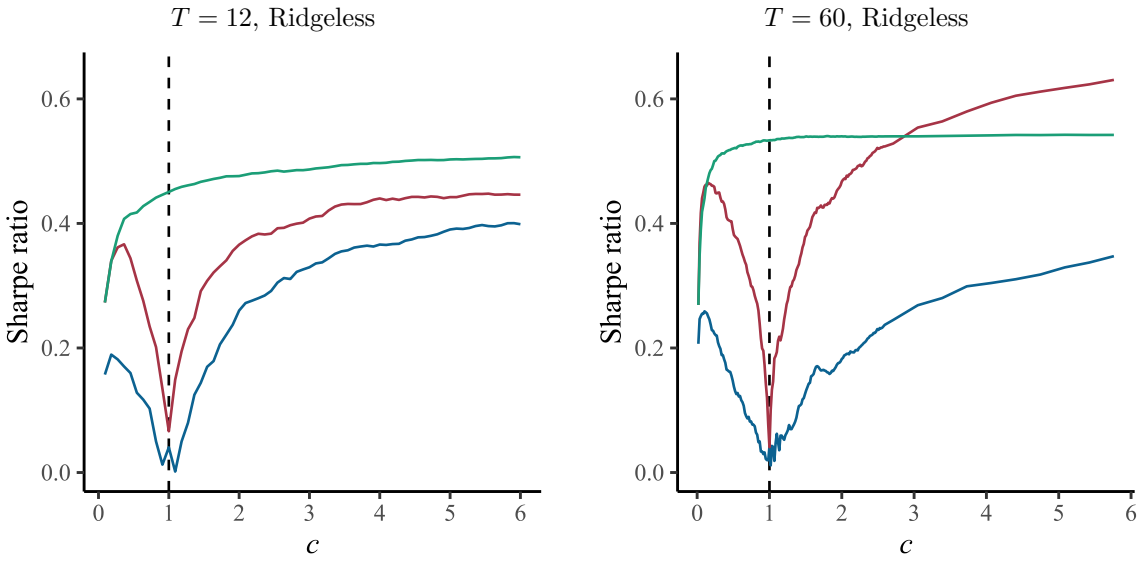

Figure 8: MPT and Individual Stocks.

The figure shows the performance of Markowitz and MVP based on a varying number of individual stocks. Specifically, we randomly select N stocks and, each month, create the portfolios based on their sample mean and variance-covariance matrix estimated over the past $T = 12$ or $T = 60$ months. For every N, we repeat the random selection process 100 times, and the performance statistics are averaged across these permutations. The x-coordinate shows the complexity, defined as $c = N/(T-1)$. The y-coordinate in the left panel shows the annualized Sharpe ratio of the Markowitz portfolio, and the y-coordinate in the right panel shows the annualized volatility of the minimum variance portfolio. The performance statistics are based on data from 1977 to 2023.

B Proofs

Didisheim et al. (2024) derive the asymptotic theory in a much more complex setting where the traded assets are themselves managed portfolios originating from a large panel of signals. Their results trivially apply in our setting under Assumption 1. We state these results here for the readers' convenience.

Let

$$\begin{aligned} \mathcal{E}_*(z) &= \lim_{N\to\infty} \mu'(zI + \Sigma^{-1}\mu \\ \mathcal{V}_*(z) &= \lim_{N\to\infty} \mu'\Sigma(zI + \Sigma)^{-2}\mu \\ m(-z) &= \lim_{N\to\infty} N^{-1}\operatorname{tr}((zI + \Sigma)^{-1}) . \end{aligned} \tag{11}$$

Under Assumption 1, we have

$$\begin{aligned}
\lim_{N_1,N\to\infty,N1/N\to q} \mu(N_1)'(zI+\Sigma(N_1))^{-1}\mu(N_1) = q\,\mathcal{E}_*(z) &= \mathcal{E}(z;q) \\
\lim_{N_1,N\to\infty,N1/N\to q} \mu(N_1)\Sigma(N_1)'(zI+\Sigma(N_1))^{-2}\mu(N_1) = q\,\mathcal{V}_*(z) &= \mathcal{V}(z;q) \\
m(-z) &= \lim_{N_1\to\infty} N_1^{-1}\operatorname{tr}((zI+\Sigma(N_1))^{-1})\,.
\end{aligned} \tag{12}$$

The following results follow from the results in Didisheim et al. (2024). We use $\hat{mu}(N_1)$ and $\hat{\Sigma}(N_1)$ to denote the sample mean and covariance matrix of the N_1 assets.

Lemma 1 (Bai and Zhou (2008)) *In the limit as $N_1, T \to \infty$, $N_1/T \to c$, we have*

$$\lim N_1^{-1}\operatorname{tr}((zI+\hat{\Sigma}(N_1))^{-1}) = m(-z;c) \tag{13}$$

in probability, where $m(-z;c)$ is the unique positive solution to

$$m(-z;c) = \frac{1}{1-c+czm(-z;c)} m\left(\frac{-z}{1-c+czm(-z;c)}\right) \tag{14}$$

Let

$$\xi(z;c) = \frac{c(1 - m(-z;c)z)}{1-c(1 - m(-z;c)z)} \tag{15}$$

and

$$Z^*(z;c) = z(1+\xi(z;c))\,. \tag{16}$$

We will also need

Theorem 2 (Knowles and Yin (2017)) *For any sequence of bounded vectors β, we have*

$$\beta'(zI+\hat{\Sigma})^{-1}\beta - (1+\xi(z;c))\beta'(Z_*(z;c)I+\Sigma)^{-1}\beta \to 0 \tag{17}$$

and

$$\beta'(zI+\hat{\Sigma})^{-1}\Sigma(zI+\hat{\Sigma})^{-1}\beta - (1+\xi(z;c))^2(1+G(z;c)\beta'(Z_*(z;c)I+\Sigma)^{-2}\Sigma\beta \to 0 \tag{18}$$

in probability. Here,

$$G(z;c) \;=\; \xi(z;c) \;+\; z\,\xi'(z;c) \tag{19}$$

is the limits-to-learning gap of Chen et al. (2025).

Importantly, Didisheim et al. (2024) construct the Stochastic Discount Factor (SDF) using a non-demeaned covariance matrix (i.e., the second moment matrix), and the corresponding efficient portfolios are proportional due to the Sherman-Morrison formula. Let

$$\tilde{w} \;=\; (zI + \hat{E}[RR'])^{-1}\hat{\mu} \tag{20}$$

be the portfolio from Didisheim et al. (2024). Then,

$$\tilde{w} \;=\; \frac{\hat{w}^{\text{eff}}}{1 + \hat{\mu}'(zI + \hat{\Sigma})^{-1}\hat{\mu}}\,. \tag{21}$$

Furthermore, the results of Didisheim et al. (2024) imply that

$$\hat{\mu}'(zI + \hat{\Sigma})^{-1}\hat{\mu} \;\approx\; \mu'(zI + \hat{\Sigma})^{-1}\mu \;+\; \xi(z;c) \tag{22}$$

whereas

$$\mu'(zI + \hat{\Sigma})^{-1}\mu \;\approx\; (1 + \xi(z;c))\mu'(Z_*(z;c)I + \Sigma)^{-1}\mu \;\approx\; (1 + \xi(z;c))\mathcal{E}_*(Z_*(z;c))\,, \tag{23}$$

Similarly,

$$E[\hat{w}'R_{T+1}] \;=\; E[\hat{w}'\mu] \;\approx\; \mu'(zI + \hat{\Sigma})^{-1}\mu \tag{24}$$

so that

$$E[\tilde{w}'R_{T+1}] \;\approx\; \frac{E[\hat{w}'R_{T+1}]}{1 + (1 + \xi(z;c))\mathcal{E}_*(Z_*(z;c)) + \xi(z;c)} \;\approx\; \frac{1}{1 + \xi(z;c)}\frac{\mathcal{E}_*(Z_*(z;c))}{1 + \mathcal{E}_*(Z_*(z;c))} \tag{25}$$

Similarly,

$$E[(\hat{w}'R_{T+1})^2] \;=\; E[\hat{w}'\Sigma\hat{w}] \tag{26}$$

and

$$E[\mu'(zI+\hat{\Sigma})^{-1}\Sigma(zI+\hat{\Sigma})^{-1}\mu] \approx \ (1+\xi(z;c))^2(1+G(z;c))\,\mathcal{V}(Z^*(z;c)), \tag{27}$$

and, analogously,

$$E[\iota'(zI+\hat{\Sigma})^{-1}\Sigma(zI+\hat{\Sigma})^{-1}\iota] \approx \ (1+\xi(z;c))^2(1+G(z;c))\,\lim \iota'(zI+\Sigma)^{-2}\Sigma\iota\,. \tag{28}$$

Furthermore, by (21),

$$E[(\tilde{w}'R_{T+1})^2] \ \approx \ \frac{E[(\hat{w}'R_{T+1})^2]}{(1+(1+\xi(z;c))\mathcal{E}_*(Z_*(z;c))+\xi(z;c))^2}\,. \tag{29}$$

The following is then a direct consequence of the main theoretical result in Didisheim et al. (2024).

Theorem 3 (Didisheim et al. (2024)) *We have*

$$\lim \frac{\text{Var}[\hat{w}'R_{T+1}]}{E[\hat{w}'R_{T+1}]^2} \ = \ (1+G(z;c))\frac{1}{\mathcal{S}^2(Z^*;q)} + \ G(z;c)\left(\frac{1}{\mathcal{E}_*(Z^*;q)}\right)^2,$$
$$\textit{where } \mathcal{S}^2(z;q) \ = \ \mathcal{E}^2(z;q)/\mathcal{V}(z;q)$$

Our next goal is to compute the ridgeless limit of these expressions

Proof of Theorem 1. Substituting $m(z;c) = c^{-1}\tilde{m}(z;c) + (c^{-1}-1)z^{-1}$ into (14), we get that the Master equation for $\tilde{m}(-z;c)$ is

$$z\tilde{m} - 1 + c(1 - \int \frac{dH(x)}{(1+\tilde{m}x)}) \ = \ 0\,, \tag{30}$$

where dH is the asymptotic distribution of eigenvalues of Σ_P. Differentiating this identity, we get that, for $c<1$, $\tilde{m} \sim z^{-1}(1-c)$, and $Z_* = 1/\tilde{m} \to 0$ goes to zero in the ridgeless limit, while $G \to c/(1-c)$ for $c<1$. As a result,

$$\lim \frac{\text{Var}[\hat{w}'R_{T+1}]}{E[\hat{w}'R_{T+1}]^2} \ = \ \frac{1}{1-c}\frac{1}{\mathcal{S}^2(0;q)} + \ \frac{c}{1-c}\left(\frac{1}{\mathcal{E}(0;q)}\right)^2 \tag{31}$$

for $c<1$. Recall that $q = N_1/N$. Then, $c = N_1/T = qN/T$ so that $q = c(T/N)$. Substituting

this expression, we get

$$\begin{aligned}&\frac{1}{1-c}\frac{1}{\mathcal{S}_*(0)^2 c(T/N)} + \frac{c}{1-c}\left(\frac{1}{(\mathcal{S}_*(0)c(T/N))^2}\right)^2 \\ &= \frac{1}{c(1-c)}\frac{1}{\mathcal{S}_*(0)^2}\left(\frac{1}{(T/N)}+\frac{1}{(T/N)^2}\right)\end{aligned} \tag{32}$$

Since ι is not bounded, we use $\tilde{\iota} = \iota/N^{1/2}$ instead, so that

$$(\iota'(zI+\hat{\Sigma}(N_1))^{-1}\iota)^{-1}\iota'(zI+\hat{\Sigma}(N_1))^{-1}R_{T+1} \;=\; N_1^{-1/2}(\tilde{\iota}'(zI+\hat{\Sigma}(N_1))^{-1}\tilde{\iota})^{-1}\tilde{\iota}'(zI+\hat{\Sigma}(N_1))^{-1}R_{T+1}$$

Then, Theorem 2 implies that

$$\begin{aligned}&\mathrm{Var}[N_1^{1/2}(\iota'(zI+\hat{\Sigma}(N_1))^{-1}\iota)^{-1}\iota'(zI+\hat{\Sigma})^{-1}R_{T+1}] \\ &\approx N_1((Z_*/z)\iota'(Z_*I+\Sigma(N_1))^{-1}\iota)^{-2}\iota'(zI+\hat{\Sigma}(N_1))^{-1}\Sigma(N_1)(zI+\hat{\Sigma}(N_1))^{-1}\iota \\ &\approx N_1(\iota'(Z_*I+\Sigma(N_1))^{-1}\iota)^{-2}(1+G)\iota'(Z_*I+\Sigma(N_1))^{-2}\Sigma(N_1)\iota\end{aligned} \tag{33}$$

In the ridgeless limit, for $c<1$, we have

$$\frac{1}{1-c}\frac{1}{\iota'\Sigma(N_1)^{-1}\iota}\,. \tag{34}$$

B.1 Large$-c$ asymptotics

Our goal is to derive the asymptotic behavior of

$$\begin{aligned}&\lim\frac{\mathrm{Var}[\hat{w}'R_{T+1}]}{E[\hat{w}'R_{T+1}]^2} \;=\; (1+G(N_1/T))\frac{1}{(N_1/N)\mathcal{S}^2(0)} + G(N_1/T)\left(\frac{1}{(N_1/N)\mathcal{E}(0)}\right)^2 \text{ and} \\ &N_1(\iota'(Z_*(N_1/T)I+\Sigma(N_1))^{-1}\iota)^{-2}(1+G(N_1/T))\iota'(Z_*(N_1/T)I+\Sigma(N_1))^{-2}\Sigma(N_1)\iota\,.\end{aligned} \tag{35}$$

for $N_1 = N$. This gives

$$\begin{aligned}&\lim\frac{\mathrm{Var}[\hat{w}'R_{T+1}]}{E[\hat{w}'R_{T+1}]^2} \;=\; (1+G(N/T))\frac{1}{\mathcal{S}^2(0)} + G(N/T)\left(\frac{1}{\mathcal{E}(0)}\right)^2 \text{ and} \\ &N(\iota'(Z_*(N/T)I+\Sigma)^{-1}\iota)^{-2}(1+G(N/T))\iota'(Z_*(N/T)I+\Sigma)^{-2}\Sigma\iota\,.\end{aligned} \tag{36}$$

To study the large $-c$ asymptotics, we will need some auxiliary results.

Lemma 2 (Large-c asymptotics of $\tilde{m}(c)$) *Let H be a probability measure supported on*

$[0,\infty)$, *and define for* $c>1$ *the unique* $\tilde{m}(c)>0$ *satisfying*

$$c\int \frac{dH(x)}{1+\tilde{m}(c)\,x} = c-1. \tag{37}$$

Assume the moments

$$\mu_k := \int x^k\, dH(x)$$

exist and are finite.

1. *If* $\mu_1 \in (0,\infty)$, *then as* $c\to\infty$,

$$\tilde{m}(c) \sim \frac{1}{\mu_1\, c}.$$

2. *If moreover* $\mu_2<\infty$, *then as* $c\to\infty$,

$$\tilde{m}(c) = \frac{1}{\mu_1\, c} + \frac{\mu_2}{\mu_1^3\, c^2} + O(c^{-3}),$$

and in particular

$$\tilde{m}(c) = \frac{1}{\mu_1\, c} + \frac{\mu_2}{\mu_1^3\, c^2} + o(c^{-2}).$$

Proof. Define

$$I(m) := \int \frac{dH(x)}{1+mx}, \qquad m\ge 0.$$

Dividing (37) by c gives

$$I(\tilde{m}(c)) = 1-\frac{1}{c}. \tag{38}$$

Since $x\ge 0$, the map $m\mapsto (1+mx)^{-1}$ is decreasing for each x, hence I is decreasing and continuous on $[0,\infty)$ with $I(0)=1$ and $\lim_{m\to\infty} I(m) = H(\{0\}) \in [0,1]$. Therefore, for every $c>1$ there is a unique $\tilde{m}(c)>0$ solving (38). Moreover, (38) implies $I(\tilde{m}(c))\uparrow 1$ as $c\to\infty$, and by continuity and strict monotonicity of I near 0 we conclude $\tilde{m}(c)\downarrow 0$.

Next, for $m \geq 0$ and $x \geq 0$,

$$\frac{1}{1+mx} = 1 - mx + \frac{m^2x^2}{1+mx}.$$

Integrating with respect to H yields

$$I(m) = 1 - \mu_1 m + m^2 R(m), \qquad R(m) := \int \frac{x^2}{1+mx}\, dH(x). \tag{39}$$

If $\mu_2 < \infty$, then $0 \leq R(m) \leq \mu_2$ and $R(m) \to \mu_2$ as $m \downarrow 0$ by dominated convergence.

Step 1 (leading order). From (38) and (39) with $m = \tilde{m}(c)$ we get

$$\mu_1 \tilde{m}(c) - \tilde{m}(c)^2 R(\tilde{m}(c)) = \frac{1}{c}.$$

Because $R(\tilde{m}(c)) \geq 0$, this implies $\mu_1 \tilde{m}(c) \geq c^{-1}$, i.e. $\tilde{m}(c) \geq (\mu_1 c)^{-1}$. If $\mu_2 < \infty$, then $R(\tilde{m}(c)) \leq \mu_2$ and hence

$$\mu_1 \tilde{m}(c) - \mu_2 \tilde{m}(c)^2 \leq \frac{1}{c}.$$

Since $\tilde{m}(c) \to 0$, the quadratic term is $o(\tilde{m}(c))$, so dividing by $\tilde{m}(c)$ gives $\mu_1 = (c\tilde{m}(c))^{-1} + o(1)$, equivalently $c\,\tilde{m}(c) \to \mu_1^{-1}$, proving $\tilde{m}(c) \sim (\mu_1 c)^{-1}$. (The same conclusion follows under the weaker condition $\mu_1 < \infty$ by standard differentiability at zero arguments for I, but we keep the moment expansion route.)

Step 2 (second order). Assume $\mu_2 < \infty$. Expand $I(m)$ to second order using (39) and the limit $R(m) \to \mu_2$:

$$I(m) = 1 - \mu_1 m + \mu_2 m^2 + o(m^2) \qquad (m \downarrow 0).$$

Insert $m = \tilde{m}(c)$ and use (38):

$$1 - \mu_1 \tilde{m}(c) + \mu_2 \tilde{m}(c)^2 + o(\tilde{m}(c)^2) = 1 - \frac{1}{c},$$

so

$$\mu_1 \tilde{m}(c) - \mu_2 \tilde{m}(c)^2 = \frac{1}{c} + o(c^{-2}), \tag{40}$$

where we used $\tilde{m}(c) \asymp c^{-1}$ from Step 1. Now set $\tilde{m}(c) = \frac{a}{c} + \frac{b}{c^2} + o(c^{-2})$ and substitute into

(40):

$$\mu_1\Big(\frac{a}{c}+\frac{b}{c^2}\Big)-\mu_2\Big(\frac{a^2}{c^2}\Big)=\frac{1}{c}+o(c^{-2}).$$

Matching coefficients yields $a=\mu_1^{-1}$ and $b=\mu_2\mu_1^{-3}$, which gives

$$\tilde{m}(c)=\frac{1}{\mu_1\,c}+\frac{\mu_2}{\mu_1^3\,c^2}+o(c^{-2}).$$

If additionally $\mu_3<\infty$, one can continue the expansion to obtain an $O(c^{-3})$ remainder (by a third-order Taylor/moment expansion of I at 0), which we record in the statement. □

Lemma 3 (Large-c asymptotics of $G(c)$) *Let H be a probability measure supported on $[0,\infty)$, and let $\tilde{m}(c)>0$ be the unique solution of*

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

Define

$$G(c)\;:=\;-1+\frac{1}{c\displaystyle\int\frac{\tilde{m}(c)\,x}{\big(1+\tilde{m}(c)\,x\big)^2}\,dH(x)}.$$

Let $\mu_k:=\int x^k\,dH(x)$.

1. *If $\mu_1\in(0,\infty)$ and $\mu_2<\infty$, then as $c\to\infty$,*

$$G(c)=\frac{\mu_2}{\mu_1^2}\,\frac{1}{c}+o\Big(\frac{1}{c}\Big).$$

2. *If moreover $\mu_3<\infty$, then as $c\to\infty$,*

$$G(c)=\frac{\mu_2}{\mu_1^2}\,\frac{1}{c}+\Big(\frac{5\mu_2^2}{\mu_1^4}-\frac{3\mu_3}{\mu_1^3}\Big)\,\frac{1}{c^2}+o\Big(\frac{1}{c^2}\Big).$$

Proof. Set, for $m\ge 0$,

$$J(m):=\int\frac{mx}{(1+mx)^2}\,dH(x).$$

Then by definition,

$$G(c) = -1 + \frac{1}{c\, J(\tilde{m}(c))}.$$

Step 1: small-m expansion of $J(m)$. For $u \geq 0$,

$$\frac{u}{(1+u)^2} = u - 2u^2 + \frac{3u^3}{(1+u)^2},$$

hence with $u = mx$,

$$\frac{mx}{(1+mx)^2} = mx - 2m^2x^2 + 3m^3\frac{x^3}{(1+mx)^2}.$$

Integrating yields

$$J(m) = \mu_1 m - 2\mu_2 m^2 + 3m^3 R_3(m), \qquad R_3(m) := \int \frac{x^3}{(1+mx)^2}\, dH(x). \tag{41}$$

If $\mu_3 < \infty$, then $0 \leq R_3(m) \leq \mu_3$ and $R_3(m) \to \mu_3$ as $m \downarrow 0$ by dominated convergence, so

$$J(m) = \mu_1 m - 2\mu_2 m^2 + 3\mu_3 m^3 + o(m^3) \qquad (m \downarrow 0). \tag{42}$$

If only $\mu_2 < \infty$, then the decomposition

$$\frac{mx}{(1+mx)^2} = mx - 2m^2\frac{x^2}{1+mx}$$

implies

$$J(m) = \mu_1 m - 2\mu_2 m^2 + o(m^2) \qquad (m \downarrow 0), \tag{43}$$

since $x^2/(1+mx) \leq x^2$ and $\int x^2\, dH < \infty$.

Step 2: plug in $\tilde{m}(c)$. From the previous lemma (already established),

$$\tilde{m}(c) = \frac{1}{\mu_1 c} + \frac{\mu_2}{\mu_1^3 c^2} + o(c^{-2}). \tag{44}$$

(i) First-order expansion. Using (43) and (44),

$$c\,J(\tilde{m}(c)) = c\Big(\mu_1\tilde{m}(c) - 2\mu_2\tilde{m}(c)^2 + o(\tilde{m}(c)^2)\Big) = 1 - \frac{\mu_2}{\mu_1^2}\frac{1}{c} + o\left(\frac{1}{c}\right).$$

Therefore,

$$\frac{1}{c\,J(\tilde{m}(c))} = \frac{1}{1 - \frac{\mu_2}{\mu_1^2}\frac{1}{c} + o(c^{-1})} = 1 + \frac{\mu_2}{\mu_1^2}\frac{1}{c} + o\left(\frac{1}{c}\right),$$

and hence

$$G(c) = -1 + \frac{1}{c\,J(\tilde{m}(c))} = \frac{\mu_2}{\mu_1^2}\frac{1}{c} + o\left(\frac{1}{c}\right).$$

(ii) Second-order expansion. Assume $\mu_3 < \infty$. Combine (42) with (44):

$$c\,J(\tilde{m}(c)) = c\Big(\mu_1\tilde{m}(c) - 2\mu_2\tilde{m}(c)^2 + 3\mu_3\tilde{m}(c)^3 + o(\tilde{m}(c)^3)\Big) = 1 - \frac{a}{c} + \frac{b}{c^2} + o(c^{-2}),$$

where

$$a := \frac{\mu_2}{\mu_1^2}, \qquad b := \frac{3\mu_3}{\mu_1^3} - \frac{4\mu_2^2}{\mu_1^4}.$$

Now expand $(1 - \frac{a}{c} + \frac{b}{c^2})^{-1}$:

$$\frac{1}{c\,J(\tilde{m}(c))} = \frac{1}{1 - \frac{a}{c} + \frac{b}{c^2} + o(c^{-2})} = 1 + \frac{a}{c} + \frac{a^2 - b}{c^2} + o(c^{-2}).$$

Thus

$$G(c) = -1 + \frac{1}{c\,J(\tilde{m}(c))} = \frac{a}{c} + \frac{a^2 - b}{c^2} + o(c^{-2}) = \frac{\mu_2}{\mu_1^2}\frac{1}{c} + \left(\frac{5\mu_2^2}{\mu_1^4} - \frac{3\mu_3}{\mu_1^3}\right)\frac{1}{c^2} + o(c^{-2}).$$

□

Let $v(x)$ be the asymptotic density of the distribution of $(U'\mu)^2$ with respect to $dH(x)$.

Then,

$$\begin{aligned} \mathcal{E}_*(c) &= \int \frac{v(x)}{\tilde{m}(c)^{-1}+x}\, dH(x) \\ \mathcal{V}_*(c) &= \int \frac{v(x)\, x}{(\tilde{m}(c)^{-1}+x)^2}\, dH(x) \end{aligned} \tag{45}$$

Lemma 4 (Asymptotics of S) *Let H be a probability measure supported on $[0,\infty)$ with moments*

$$\mu_k := \int x^k\, dH(x), \qquad \mu_1 \in (0,\infty),\ \mu_2 < \infty.$$

Let $\tilde{m}(c) > 0$ be the unique solution of

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

Let $v : [0,\infty) \to \mathbb{R}$ be such that the weighted moments

$$\nu_k := \int |v(x)|\, x^k\, dH(x)$$

are finite for $k = 0, 1$.

Define

$$\mathcal{E}_*(c) := \int \frac{v(x)}{\tilde{m}(c)^{-1}+x}\, dH(x).$$

Then, as $c \to \infty$,

$$\mathcal{E}_*(c) = \frac{\nu_0}{\mu_1}\,\frac{1}{c} + \left(\frac{\mu_2\,\nu_0}{\mu_1^3} - \frac{\nu_1}{\mu_1^2}\right)\frac{1}{c^2} + o\!\left(\frac{1}{c^2}\right),$$

where $\nu_0 = \int v(x)\, dH(x)$ and $\nu_1 = \int v(x)\, x\, dH(x)$ (these are finite under the stated assumptions).

Proof. Write $m := \tilde{m}(c)$. Since

$$\frac{1}{m^{-1}+x} = \frac{m}{1+mx},$$

we have

$$\mathcal{E}_*(c) = m \int \frac{v(x)}{1+mx}\, dH(x).$$

For $m \downarrow 0$ and $x \geq 0$,

$$\frac{1}{1+mx} = 1 - mx + \frac{m^2x^2}{1+mx},$$

hence

$$\int \frac{v(x)}{1+mx}\, dH(x) = \nu_0 - m\nu_1 + m^2 R_v(m), \qquad R_v(m) := \int v(x)\, \frac{x^2}{1+mx}\, dH(x).$$

Under $\int |v(x)|x^2\, dH(x) < \infty$, we have $R_v(m) = \nu_2 + o(1)$ as $m \downarrow 0$ by dominated convergence, and in any case (with only $\nu_0, \nu_1 < \infty$) the first two terms are valid with a remainder $o(m)$.

Thus,

$$\mathcal{E}_*(c) = m\nu_0 - m^2\nu_1 + o(m^2).$$

From the previous lemma,

$$m = \tilde{m}(c) = \frac{1}{\mu_1 c} + \frac{\mu_2}{\mu_1^3 c^2} + o(c^{-2}), \qquad m^2 = \frac{1}{\mu_1^2 c^2} + o(c^{-2}).$$

Plugging into $\mathcal{E}_*(c) = m\nu_0 - m^2\nu_1 + o(m^2)$ gives

$$\mathcal{E}_*(c) = \left(\frac{1}{\mu_1 c} + \frac{\mu_2}{\mu_1^3 c^2} + o(c^{-2})\right)\nu_0 - \left(\frac{1}{\mu_1^2 c^2} + o(c^{-2})\right)\nu_1 + o(c^{-2}),$$

which simplifies to the claimed expansion. □

Lemma 5 (Asymptotics of $\mathcal{V}_*(c)$) *Let H be a probability measure supported on $[0, \infty)$ with moments*

$$\mu_k := \int x^k\, dH(x), \qquad \mu_1 \in (0, \infty),\ \mu_2 < \infty.$$

Let $\tilde{m}(c) > 0$ be the unique solution of

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

Let $v : [0, \infty) \to \mathbb{R}$ *satisfy*

$$\int |v(x)|\, x\, dH(x) < \infty, \qquad \int |v(x)|\, x^2\, dH(x) < \infty.$$

Define

$$\mathcal{V}_*(c) := \int \frac{v(x)\, x}{\left(\tilde{m}(c)^{-1} + x\right)^2}\, dH(x).$$

Then, as $c \to \infty$*,*

$$\mathcal{V}_*(c) = \frac{\nu_1}{\mu_1^2}\, \frac{1}{c^2} + \left(\frac{2\mu_2\, \nu_1}{\mu_1^4} - \frac{2\nu_2}{\mu_1^3} \right) \frac{1}{c^3} + o\left(\frac{1}{c^3} \right),$$

where

$$\nu_1 := \int v(x)\, x\, dH(x), \qquad \nu_2 := \int v(x)\, x^2\, dH(x).$$

Proof. Write $m := \tilde{m}(c)$. Using

$$\frac{1}{(m^{-1} + x)^2} = \frac{m^2}{(1 + mx)^2},$$

we get

$$\mathcal{V}_*(c) = m^2 \int \frac{v(x)\, x}{(1 + mx)^2}\, dH(x).$$

For $u \geq 0$,

$$\frac{1}{(1 + u)^2} = 1 - 2u + \frac{3u^2}{(1 + u)^2},$$

hence with $u = mx$,

$$\frac{x}{(1 + mx)^2} = x - 2mx^2 + 3m^2 \frac{x^3}{(1 + mx)^2}.$$

Multiplying by $v(x)$ and integrating gives

$$\int \frac{v(x)\,x}{(1+mx)^2}\,dH(x) = \nu_1 - 2m\nu_2 + 3m^2 R_v(m), \qquad R_v(m) := \int v(x)\,\frac{x^3}{(1+mx)^2}\,dH(x).$$

Under $\int |v(x)|x^2\,dH(x) < \infty$ we at least have

$$\int \frac{v(x)\,x}{(1+mx)^2}\,dH(x) = \nu_1 - 2m\nu_2 + o(m), \qquad (m \downarrow 0),$$

and if additionally $\int |v(x)|x^3\,dH(x) < \infty$ then the remainder is $O(m^2)$.

Therefore,

$$\mathcal{V}_*(c) = m^2\nu_1 - 2m^3\nu_2 + o(m^3).$$

From the asymptotics of $\tilde{m}(c)$,

$$m = \frac{1}{\mu_1 c} + \frac{\mu_2}{\mu_1^3 c^2} + o(c^{-2}), \qquad m^2 = \frac{1}{\mu_1^2 c^2} + \frac{2\mu_2}{\mu_1^4 c^3} + o(c^{-3}), \qquad m^3 = \frac{1}{\mu_1^3 c^3} + o(c^{-3}).$$

Plugging into $\mathcal{V}_*(c) = m^2\nu_1 - 2m^3\nu_2 + o(m^3)$ yields

$$\mathcal{V}_*(c) = \left(\frac{1}{\mu_1^2 c^2} + \frac{2\mu_2}{\mu_1^4 c^3} + o(c^{-3})\right)\nu_1 - 2\left(\frac{1}{\mu_1^3 c^3} + o(c^{-3})\right)\nu_2 + o(c^{-3}),$$

which simplifies to the stated expansion. □

Lemma 6 (Asymptotics of $\mathcal{V}_*(c)$) *Let H be a probability measure supported on $[0,\infty)$ with moments*

$$\mu_k := \int x^k\,dH(x), \qquad \mu_1 \in (0,\infty),\ \mu_2 < \infty.$$

Let $\tilde{m}(c) > 0$ be the unique solution of

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

Let $v : [0,\infty) \to \mathbb{R}$ satisfy

$$\int |v(x)|\,x\,dH(x) < \infty, \qquad \int |v(x)|\,x^2\,dH(x) < \infty.$$

Define

$$\mathcal{V}_*(c) := \int \frac{v(x)\,x}{\left(\tilde{m}(c)^{-1} + x\right)^2}\, dH(x).$$

Then, as $c \to \infty$*,*

$$\mathcal{V}_*(c) = \frac{\nu_1}{\mu_1^2}\frac{1}{c^2} + \left(\frac{2\mu_2\,\nu_1}{\mu_1^4} - \frac{2\nu_2}{\mu_1^3}\right)\frac{1}{c^3} + o\left(\frac{1}{c^3}\right),$$

where

$$\nu_1 := \int v(x)\,x\,dH(x), \qquad \nu_2 := \int v(x)\,x^2\,dH(x).$$

Proof. Write $m := \tilde{m}(c)$. Using

$$\frac{1}{(m^{-1}+x)^2} = \frac{m^2}{(1+mx)^2},$$

we get

$$\mathcal{V}_*(c) = m^2 \int \frac{v(x)\,x}{(1+mx)^2}\, dH(x).$$

For $u \geq 0$,

$$\frac{1}{(1+u)^2} = 1 - 2u + \frac{3u^2}{(1+u)^2},$$

hence with $u = mx$,

$$\frac{x}{(1+mx)^2} = x - 2mx^2 + 3m^2\frac{x^3}{(1+mx)^2}.$$

Multiplying by $v(x)$ and integrating gives

$$\int \frac{v(x)\,x}{(1+mx)^2}\, dH(x) = \nu_1 - 2m\nu_2 + 3m^2 R_v(m), \qquad R_v(m) := \int v(x)\,\frac{x^3}{(1+mx)^2}\, dH(x).$$

Under $\int |v(x)|x^2\,dH(x) < \infty$ we at least have

$$\int \frac{v(x)\,x}{(1+mx)^2}\,dH(x) = \nu_1 - 2m\nu_2 + o(m), \qquad (m \downarrow 0),$$

and if additionally $\int |v(x)|x^3\,dH(x) < \infty$ then the remainder is $O(m^2)$.

Therefore,

$$\mathcal{V}_*(c) = m^2\nu_1 - 2m^3\nu_2 + o(m^3).$$

From the asymptotics of $\tilde{m}(c)$,

$$m = \frac{1}{\mu_1 c} + \frac{\mu_2}{\mu_1^3 c^2} + o(c^{-2}), \qquad m^2 = \frac{1}{\mu_1^2 c^2} + \frac{2\mu_2}{\mu_1^4 c^3} + o(c^{-3}), \qquad m^3 = \frac{1}{\mu_1^3 c^3} + o(c^{-3}).$$

Plugging into $\mathcal{V}_*(c) = m^2\nu_1 - 2m^3\nu_2 + o(m^3)$ yields

$$\mathcal{V}_*(c) = \left(\frac{1}{\mu_1^2 c^2} + \frac{2\mu_2}{\mu_1^4 c^3} + o(c^{-3})\right)\nu_1 - 2\left(\frac{1}{\mu_1^3 c^3} + o(c^{-3})\right)\nu_2 + o(c^{-3}),$$

which simplifies to the stated expansion. □

Lemma 7 (Asymptotics of the combined ratio) *Let H be a probability measure supported on $[0,\infty)$ and let $\tilde{m}(c) > 0$ solve*

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

Define the moments $\mu_k := \int x^k\,dH(x)$ and the v–weighted moments

$$\nu_k := \int v(x)\,x^k\,dH(x).$$

Assume $\mu_1 \in (0,\infty)$, $\mu_2 < \infty$, and $\nu_0 \neq 0$, $\nu_1 < \infty$, $\nu_2 < \infty$. Let

$$\mathcal{E}_*(c) := \int \frac{v(x)}{\tilde{m}(c)^{-1} + x}\,dH(x), \qquad \mathcal{V}_*(c) := \int \frac{v(x)\,x}{(\tilde{m}(c)^{-1} + x)^2}\,dH(x),$$

and

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

Consider

$$R(c) := \frac{(1+G(c))\, \mathcal{V}_*(c) + G(c)}{(\mathcal{E}_*(c)\, c)^2}.$$

Then, as $c \to \infty$*,*

$$R(c) = \frac{\mu_2 + \nu_1}{\nu_0^2}\, \frac{1}{c} + o\left(\frac{1}{c}\right). \tag{46}$$

If in addition $\mu_3 < \infty$*, then the second-order expansion holds:*

$$R(c) = \frac{\mu_2 + \nu_1}{\nu_0^2}\, \frac{1}{c} + \left[\frac{3\mu_2^2 + \mu_2\nu_1}{\mu_1^2\, \nu_0^2} - \frac{2\nu_2 + 3\mu_3}{\mu_1\, \nu_0^2} + \frac{2(\mu_2+\nu_1)\nu_1}{\mu_1\, \nu_0^3}\right] \frac{1}{c^2} + o\left(\frac{1}{c^2}\right).$$

Proof. From the previous lemmas, as $c \to \infty$,

$$\mathcal{E}_*(c) = \frac{\nu_0}{\mu_1}\, \frac{1}{c} + \left(\frac{\mu_2\nu_0}{\mu_1^3} - \frac{\nu_1}{\mu_1^2}\right) \frac{1}{c^2} + o(c^{-2}), \tag{47}$$

$$\mathcal{V}_*(c) = \frac{\nu_1}{\mu_1^2}\, \frac{1}{c^2} + \left(\frac{2\mu_2\nu_1}{\mu_1^4} - \frac{2\nu_2}{\mu_1^3}\right) \frac{1}{c^3} + o(c^{-3}), \tag{48}$$

$$G(c) = \frac{\mu_2}{\mu_1^2}\, \frac{1}{c} + o(c^{-1}), \tag{49}$$

and if $\mu_3 < \infty$,

$$G(c) = \frac{\mu_2}{\mu_1^2}\, \frac{1}{c} + \left(\frac{5\mu_2^2}{\mu_1^4} - \frac{3\mu_3}{\mu_1^3}\right) \frac{1}{c^2} + o(c^{-2}). \tag{50}$$

Introduce shorthand

$$A_0 := \frac{\nu_0}{\mu_1}, \qquad A_1 := \frac{\mu_2\nu_0}{\mu_1^3} - \frac{\nu_1}{\mu_1^2}, \qquad t_1 := \frac{\nu_1}{\mu_1^2}, \qquad g_1 := \frac{\mu_2}{\mu_1^2}.$$

Then (47) implies

$$\mathcal{E}_*(c)\,c = A_0 + \frac{A_1}{c} + o(c^{-1}), \qquad (\mathcal{E}_*(c)\,c)^2 = A_0^2 + \frac{2A_0A_1}{c} + o(c^{-1}).$$

Also, from (48) and (49),

$$\mathcal{V}_*(c) = \frac{t_1}{c^2} + o(c^{-3}), \qquad G(c) = \frac{g_1}{c} + o(c^{-1}), \qquad 1 + G(c) = 1 + o(1).$$

Hence the numerator satisfies

$$(1 + G(c))\,\mathcal{V}_*(c)\,c + G(c) = \frac{t_1 + g_1}{c} + o(c^{-1}).$$

Therefore,

$$R(c) = \frac{\frac{t_1+g_1}{c} + o(c^{-1})}{A_0^2 + o(1)} = \frac{t_1 + g_1}{A_0^2}\,\frac{1}{c} + o(c^{-1}).$$

Substituting back $t_1 = \nu_1/\mu_1^2$, $g_1 = \mu_2/\mu_1^2$, and $A_0 = \nu_0/\mu_1$ gives the leading term

$$\frac{t_1 + g_1}{A_0^2} = \frac{(\nu_1 + \mu_2)/\mu_1^2}{\nu_0^2/\mu_1^2} = \frac{\mu_2 + \nu_1}{\nu_0^2},$$

which proves the first claim.

For the second-order term (assuming $\mu_3 < \infty$), keep one more term in each expansion:

$$\mathcal{E}_*(c)\,c = A_0 + \frac{A_1}{c} + o(c^{-1}),$$

$$\mathcal{V}_*(c)\,c = \frac{t_1}{c} + \frac{t_2}{c^2} + o(c^{-2}), \quad t_2 := \frac{2\mu_2\nu_1}{\mu_1^4} - \frac{2\nu_2}{\mu_1^3},$$

and from (50),

$$G(c) = \frac{g_1}{c} + \frac{g_2}{c^2} + o(c^{-2}), \quad g_2 := \frac{5\mu_2^2}{\mu_1^4} - \frac{3\mu_3}{\mu_1^3}.$$

Then

$$(1 + G(c))\,\mathcal{V}_*(c)\,c = \left(1 + \frac{g_1}{c} + o(c^{-1})\right)\left(\frac{t_1}{c} + \frac{t_2}{c^2} + o(c^{-2})\right) = \frac{t_1}{c} + \frac{t_2 + g_1t_1}{c^2} + o(c^{-2}),$$

so the numerator equals

$$\frac{t_1 + g_1}{c} + \frac{t_2 + g_1 t_1 + g_2}{c^2} + o(c^{-2}).$$

Moreover,

$$\frac{1}{(\mathcal{E}_*(c)\, c)^2} = \frac{1}{A_0^2}\left(1 - \frac{2A_1}{A_0}\frac{1}{c} + o(c^{-1})\right).$$

Multiplying yields

$$R(c) = \frac{t_1 + g_1}{A_0^2}\frac{1}{c} + \left[\frac{t_2 + g_1 t_1 + g_2}{A_0^2} - \frac{t_1 + g_1}{A_0^2} \cdot \frac{2A_1}{A_0}\right]\frac{1}{c^2} + o(c^{-2}),$$

and substituting $A_0, A_1, t_1, t_2, g_1, g_2$ gives exactly the stated c^{-2} coefficient. □

Finalizing the large$-c$ derivations For large c, we know that

$$Z_* \;\sim\; cP^{-1}\,\mathrm{tr}(\Sigma) \tag{51}$$

and $G(c) = O(c^{-1})$, so that

$$\begin{aligned}&(\iota'(Z_* I + \Sigma)^{-1}\iota)^{-2}(1+G)\iota'(Z_* I + \Sigma)^{-2}\Sigma\iota \\ &\approx\; \frac{\iota'\Sigma\iota}{\|\iota\|^4}\end{aligned} \tag{52}$$

and we have proved (10).

The proof of (9) follows directly from (46).

□