Portfolio Choice with Model Misspecification: A Foundation for Alpha and Beta Portfolios*

Raman Uppal Paolo Zaffaroni

March 15, 2016

Abstract

Our objective is to formalize the effect of model misspecification on mean-variance portfolios and to show how asset-pricing theory and asymptotic analysis can be used to provide powerful solutions to mitigate it. In particular we show how to design mean-variance portfolios that perform well out of sample in the presence of model misspecification. Our key insight is that, instead of treating misspecification directly in the mean-variance portfolio, it is better to first decompose the portfolio into two components, and to then treat misspecification in the two components separately using different methods. The starting point of our analysis is the Arbitrage Pricing Theory (APT). We first extend the APT to show that it can capture not just small pricing errors that are independent of factors, but also large pricing errors from mismeasured or missing factors. Then, we decompose the mean-variance portfolio into two components that correspond to the two components of returns in the APT: an “alpha” portfolio that depends only on pricing errors and a “beta” portfolio that depends on factor risk premia. For the alpha portfolio, we treat misspecification by imposing the APT restriction on alphas, which serves both as an identification condition and a shrinkage constraint, achieving substantial improvement in the precision of the estimated pricing errors; for the beta portfolio, we treat misspecification using asymptotic analysis: as the number of assets increases, the weights of the alpha portfolio dominate those of the beta portfolio, providing an expression for mean-variance portfolio weights that is immune to beta misspecification. Finally, we demonstrate that our approach leads to significant improvement in out-of-sample performance.

JEL classification: G11, G12, C58, C53.

Keywords: Active and passive portfolios, pricing errors, factor models, factor investing, mean-variance portfolio, estimation error, robust estimation.

*Uppal is affiliated with CEPR and EDHEC Business School; Email: Raman.Uppal@edhec.edu. Zaffaroni is affiliated with Imperial College Business School; Email: P.Zaffaroni@imperial.ac.uk. We gratefully acknowledge comments from Dante Amengual, Harjoat Bhamra, Victor DeMiguel, Rene Garcia, Christian Gouriéroux, Ralph Koijen, Xiaoji Lin, Lionel Martellini, Vasant Naik, Lubos Pástor, Hashem Pesaran, Cesare Robotti, Jay Shanken, Eric Shiribini, Claudio Tebaldi, Nikolaos Tissaromatis, Allan Timmermann, and Motohiro Yogo, and participants at presentations at: Imperial College Business School, Judge Business School, Norges Bank Investment Management, the Financial Econometrics Conference at Toulouse School of Economics, the NBER Summer Institute workshop on Forecasting & Empirical Methods, the UBC Summer Finance Conference, Advances in the Analysis of Hedge Fund Strategies conference at Imperial College London, and 8th Annual Hedge Fund Conference
## Contents

1 Introduction 3  
2 Related Literature 6  
3 Notation, Factor Model for Returns, and APT 10  
  3.1 Notation 10  
  3.2 Linear factor model for asset returns with misspecification 11  
  3.3 Arbitrage Pricing Theory (APT) 12  
  3.4 Extending the APT 13  
  3.5 Different forms of model misspecification 17  
4 Mitigating Model Misspecification in Mean-Variance Portfolios 19  
  4.1 Decomposing the mean-variance portfolio 20  
  4.2 Mitigating misspecification in beta component of returns 23  
  4.3 Mitigating misspecification in the alpha component of returns 28  
5 Simulation Results 33  
  5.1 Model misspecification in the alpha component of returns 33  
  5.2 Model misspecification in the beta component of returns 37  
6 Conclusion 38  
A Proofs for Theorems 40  
B Appendix with Additional Results 49  
  B.1 Mitigating Model Misspecification in Global-Minimum-Variance Portfolios 49  
  B.2 Mitigating Model Misspecification in Markowitz Frontier Portfolios 53  
  B.3 Extension of Roll (1980) 62  
  B.4 Some Useful Results on the APT 63  
  B.5 Examples of Estimation of Models with Misspecification 64  
References 73
1 Introduction

In this paper, our objective is to formalize the effect of model misspecification on mean-variance portfolios and to show how asset-pricing theory and asymptotic analysis can be used to provide powerful solutions to mitigate it. In particular we show how to design mean-variance portfolios that perform well out of sample in the presence of model misspecification. In our context, one form of model misspecification is represented by the \textit{pricing error}, often referred to as alpha.\footnote{As in Hansen and Jagannathan (1997), this alpha could arise if the asset-pricing model is correct (that is, we have the correct set of factors) but the factors are measured with error (Roll (1977), Green (1986a)) or if the factors are measured without error but the model is incorrect in the sense that some factors are missing (MacKinlay and Pástor, 2000) or are latent (Connor and Korajczyk, 1986; Lehmann and Modest, 1988). These alphas could arise also if the views of the investor disagree with the predictions of the asset-pricing model (Black and Litterman, 1990, 1992).} Another form of misspecification is associated with the beta component of returns; this arises even when all relevant factors are observable but one specifies the incorrect risk premia or covariances for the factors. Our key insight is that, instead of treating misspecification in the mean-variance portfolio, it is better to first decompose the mean-variance portfolio into two parts that correspond to the alpha and beta components of returns, and then to treat misspecification in the two parts using different methods.

We study portfolio choice assuming that asset returns satisfy the Arbitrage Pricing Theory (APT). The APT is particularly well-suited for our purpose because it allows for the possibility of model misspecification, and hence, mispricing, while still imposing no arbitrage. Moreover, the APT is a very general asset-pricing model that can accommodate a variety of observed factors. The factors could be statistical, for example, based on a principal-component decomposition of returns; macroeconomic, for example, shocks to inflation, interest rates, and exchange rates; or, characteristic-based, for instance, industry, country, size, value, return momentum, and liquidity.\footnote{For further details of the variety of applications of the APT, see Connor, Goldberg, and Korajczyk (2010, Ch. 4–6). For the importance of factor investing, see the excellent discussion in Ang (2014).} The factor model for returns also allows for correlated residual errors across assets.

Our work makes three contributions. Our first contribution is to extend the notion of the pricing errors in the APT and to show that the APT applies much more broadly than is typically assumed. Traditionally, the APT is interpreted as applying to pricing errors that are small and unrelated to factors, implying that the covariance matrix of residuals has bounded eigenvalues. Green and Hollifield (1992) extend this to allow for slowly-increasing eigenvalues while requiring that portfolios be well diversified. We show that the APT can capture not just small but also large pricing errors, such as those arising from latent pervasive factors. This implies that not all eigenvalues of the residual-covariance matrix
are bounded, and hence, even well-spread portfolios would not be well diversified. Thus, the factor model for returns that we consider is general enough to allow for missing or mismeasured factors, in addition to pure pricing errors that are unrelated to factors. In fact, our work shows that the APT is much more than just a statistical model of returns, and our results allow us to illustrate the deep economic content of the APT.

Our second contribution is to demonstrate how this extended version of the APT can be used to capture, and thus to mitigate, model misspecification in the class of mean-variance portfolios. We treat model misspecification in three steps. In the first step, we show that under the APT the optimal mean-variance portfolio can be decomposed into an “alpha” portfolio, which is a strategy that depends only on pricing errors with zero exposure to common risk, and a “beta” portfolio, which is a strategy that depends only on factor risk premia and their loadings. We characterize several properties of these two portfolios. For instance, even though the alpha and beta portfolios themselves are not on the Markowitz efficient frontier, we show that they satisfy an optimality condition: each is the minimum-variance portfolio that is orthogonal to the other, extending the result in Roll (1980) about orthogonal portfolios to the case where a risk-free asset is available. We relate our results to MacKinlay (1995) and MacKinlay and Pástor (2000) who also study orthogonal portfolios. The portfolio decomposition also allows us to show that the APT no-arbitrage restriction is equivalent to imposing a bound on the square of the Sharpe ratio of the alpha portfolio, which turns out to be identical also to the expression in Gibbons, Ross, and Shanken (1989) for the difference in the squared slope of the efficient frontier and the squared Sharpe ratio of the benchmark portfolio. The bound on the square of the Sharpe ratio of the alpha portfolio can be seen as providing a theoretical rationalization for the no-good-deal bound in Cochrane and Saa-Requejo (2001). In our discussion of the related literature, we also explain how imposing the APT restriction is analogous to the approach adopted in Garlappi, Uppal, and Wang (2007), where one accounts for parameter uncertainty in portfolio choice using the minmax approach originally proposed in Gilboa and Schmeidler (1989).

In the second step, we treat misspecification in the beta portfolio using asymptotic analysis as the number of assets increases. In the environment with an asymptotically large number of assets, we show that the weights of the alpha portfolio typically dominate

---

3These insights are important because the world’s largest hedge funds, such as Bridgewater Associates, offer alpha and beta portfolios. Similarly, sovereign-wealth funds, such as Norges Bank, separate the management of their alpha and beta funds. In fact, today most asset managers offer “portable alpha” products and a large proportion of institutional investors have invested in these products. The returns on these alpha portfolios are often referred to as “absolute returns” because they are supposed to remain positive under all market conditions.

4The asymptotic analysis as the number of assets increases is not just an abstract mathematical tool, but also corresponds to practice: hedge funds and sovereign-wealth funds hold a large number of assets in their portfolios; for instance, the portfolio of Norges Bank has over 9,000 assets.
the weights of the beta portfolio. That is, as the number of assets increases, the weight of a typical asset in the alpha portfolio diminishes slower than the corresponding weight in the beta portfolio.\footnote{This result complements that of Green and Hollifield (1992) and Jagannathan and Ma (2003) who study the magnitude of efficient portfolio weights when alpha is zero.} These asymptotic results allow us to construct portfolio weights that are functionally independent of the conditional distribution of the observed factors. This asymptotic analysis of portfolio weights also shows the dominant role played by the pricing errors as the number of assets in a portfolio gets large, which implies that it is critical to estimate the pricing errors precisely.

In the third step, we treat misspecification in the alpha portfolio. We do this by imposing the APT no-arbitrage restriction on the pricing errors, which reduces the error in estimating the parameters of the factor model generating returns. In our analysis, we consider both the case of pricing errors that are unrelated to factors and pricing errors that arise from latent factors. For the latter case, the APT no-arbitrage restriction plays a second, even more fundamental, role: in the absence of this condition, the model is not (econometrically) identified, and hence, cannot be estimated. This part of our work extends the rich insights in MacKinlay and Pástor (2000), who study estimation of models with missing factors.

Our third contribution is to demonstrate how these results about the decomposition of portfolio weights, the asymptotic analysis of these weights, together with the restriction arising from the extended APT, can and should be used to improve the estimation of the return-generating model and the portfolio weights in the presence of model misspecification. Our simulation results show that it is possible to take advantage of our theoretical insights to achieve significant improvement in the out-of-sample performance of mean-variance portfolios: the Sharpe ratios of our portfolios are 50\% to 100\% higher than the Sharpe ratio of the equal-weighted portfolio.

The rest of the paper is organized as follows. In Section 2, we discuss the literature related to our work. Our contributions are presented in the next three sections. In Section 3, we specify the linear factor model for asset returns, summarize the results in the existing literature for the APT, and then extend these results to the case of unbounded eigenvalues for the covariance matrix of residuals. In Section 4, we describe the three steps that allow one to mitigate model misspecification in mean-variance portfolios. In Section 5, we demonstrate how these results can be applied to improve the estimation of portfolio weights that achieve superior out-of-sample performance. We conclude in Section 6. Proofs and technical details for all our results in the main text are collected in Appendix A. Additional results, including the mitigation of model misspecification for the global-minimum-variance portfolio and the frontier portfolios on the Markowitz efficient frontier, are collected in Appendix B.
2 Related Literature

Each of our three contributions is related to a distinct stream of the literature, which we discuss below.

Ross (1976, 1977) develops the APT by showing that if asset returns have a strict factor structure, then the mean returns are approximately linear functions of factor loadings. Huberman (1982) formalizes the argument proposed for linear factor pricing in Ross. He also shows that the arbitrage portfolios used to prove the APT need not be well diversified (that is, these portfolios could have idiosyncratic risk); instead, they need to satisfy only two conditions: they are zero-cost portfolios and their weights are orthogonal to factor loadings (which implies that they have zero factor risk). Chamberlain (1983), Chamberlain and Rothschild (1983), and Ingersoll (1984) extend the results of Ross (1976, 1977) and Huberman (1982) to a setting where returns need to satisfy only an approximate factor structure; that is, the idiosyncratic components of returns are allowed to be correlated across assets. Chamberlain also shows that arbitrage pricing is exact if and only if there is a risky, well-diversified portfolio on the mean-variance frontier. Moreover, if all mean-variance efficient portfolios are well-diversified, then the well-diversified portfolios provide mutual-fund separation. Chamberlain and Rothschild (1983) also show that if the covariance matrix of the asset returns has only $K$ unbounded eigenvalues, then there is an approximate factor structure and it is unique.\footnote{See Reisman (1988, 1992a,b) and Shanken (1992) for further extensions of the APT.} Just as in Chamberlain (1983), Chamberlain and Rothschild (1983), and Ingersoll (1984), we do not restrict the covariance matrix of the residuals to be diagonal; that is, we allow for correlated error terms. However, in contrast to Chamberlain (1983) and Chamberlain and Rothschild (1983), we study also the case where the maximum eigenvalue of the covariance matrix of the residuals are not restricted to be bounded as the number of assets increases, which is the case when the pricing errors are large, for instance, when they are related to some latent factors.

The reason that Chamberlain (1983) and Chamberlain and Rothschild (1983) do not consider the case where the maximum eigenvalue of the covariance matrix are unbounded when the number of assets is large is that they view all factors as latent (in fact, they advocate the use of principal components to estimate the overall factor structure); consequently, the covariance matrix of the residual has necessarily bounded maximum eigenvalue. In contrast, the usual practice is to consider that returns are driven by a set of observed factors, which are assumed to be the only source of commonality. In this case, if there are missing factors, then the mean of the missing factors will show up in the pricing errors.
and the covariance of the missing factors will contribute to the residual covariance matrix. Therefore, the maximum eigenvalue of the residual covariance matrix could be unbounded as the number of assets increases.

The second part of our analysis studies the implications for portfolio selection of mispricing, as modeled by the APT. There is a large literature that studies how one should form portfolios in the presence of mispricing, which could arise because the index portfolio is inefficient (Dybvig and Ross (1985a), Green (1986a)) or because of superior information, analyst recommendations, or managerial skill (see, for example, Dybvig and Ross (1985b), Grinblatt and Titman (1989), and Kosowski, Timmermann, Wermers, and White (2006)).

The seminal paper in this literature is Treynor and Black (1973), that takes as its starting point the single-factor model with a diagonal residual covariance matrix in Sharpe (1963), and asks whether or not it is desirable to form an “active” portfolio by going long underpriced assets and shorting overpriced assets so that market risk is fully hedged; or, should one invest in a diversified “passive” portfolio completely so that it is exposed only to market risk. Most importantly, they highlight the vital role in active asset management of pricing errors, which they call “independent returns.”

Treynor and Black (1973), and the papers that build on it, study the case of the mean-variance portfolios with a target mean for the case of a finite number of assets that may be overpriced or underpriced, with these pricing errors uncorrelated to the single factor. But, these papers do not restrict the pricing errors in any way so it is not clear how exactly arbitrage opportunities are ruled out. They also consider the case where the number of risky assets is fixed, so one cannot analyze the situation where the number of risky assets is increasing asymptotically. Moreover, the pricing errors they consider are unrelated to latent factors. Our work extends their analysis along these dimensions. We provide a firm theoretical foundation for the important issue of active versus passive portfolio management highlighted in Treynor and Black (1973) by placing this analysis in the context of a no-arbitrage asset pricing model such as the APT. We allow for multiple factors, do not restrict the covariance matrix for the residual errors to be diagonal, which allows us to capture a much wider class of misspecification, and we study the case of both pricing errors that are independent of factors and those that arise from latent factors. Finally, we show that the

---

7 For work showing how beliefs about an asset-pricing model can influence portfolio choice, see Pástor (2000) and Pástor and Stambaugh (2000).

8 See, for example, Kane, Kim, and White (2003) that uses shrinkage estimators to improve the econometric methods used in constructing portfolios, He (2007) that extends the Treynor and Black model to a Bayesian setting, and Brown and Tiu (2010) that uses the Treynor and Black model to assess ex post whether the proper active-passive allocation strategy was adopted by the portfolio manager.

9 For surveys of this literature, see Kothari and Shanken (2002), Aragon and Ferson (2006), and Ferson (2013).
no-arbitrage restriction imposed by APT plays a central role in the estimation of optimal active and passive portfolios.

There is also a literature that studies whether weights in mean-variance portfolios can be extreme. In a setting without zero pricing errors, Green and Hollifield (1992) show that the presence of a single pervasive factor in asset returns would lead to extreme positive and negative weights in mean-variance efficient portfolios.\textsuperscript{10} We complement this result by showing that, in the presence of pricing error, the long-short alpha portfolio weights dominate the efficient portfolio weights as the number of assets increases. Pesaran and Zaffaroni (2009) show, under some technical conditions, that the limiting properties of the mean-variance portfolio and a factor-neutral portfolio can be similar.

The third part of our analysis examines the implications of the no-arbitrage constraint imposed by the APT and the insights from the asymptotic analysis of portfolio weights for the empirical estimation of mean-variance portfolio weights in order to improve their out-of-sample performance. It is well known that mean-variance efficient portfolios that are based on sample estimates of first and second moments perform poorly out of sample; see, for example, Jobson and Korkie (1980), Frost and Savarino (1986), Michaud (1989), Black and Litterman (1990), and DeMiguel, Garlappi, and Uppal (2009).

Of the many approaches considered to improve the out-of-sample performance of mean-variance portfolios, one is to impose portfolio constraints. For example, Frost and Savarino (1988) find that imposing shortsale constraints can lead to significant improvement in performance. In an insightful paper, Jagannathan and Ma (2003) explain in the context of the global-minimum-variance portfolio that the reason for the improved performance is that imposing shortsale constraints is equivalent to shrinking the covariance matrix, and that this constraint can help even when returns are driven by a dominant factor. DeMiguel, Garlappi, Nogales, and Uppal (2009) show that further gains are possible by imposing a more general form of the shortsale constraint, a norm constraint on the portfolio weights; they also show that these constraints can be interpreted as leading to portfolio weights with Bayesian shrinkage, just as Tibshirani (1996) does for the lasso and ridge-regression techniques. Olivares-Nadal and DeMiguel (2015) show that the portfolio optimization problem with a constraint that is motivated by transaction costs can be interpreted in three ways: as a robust portfolio optimization problem, a robust regression problem, and a Bayesian portfolio problem.

\textsuperscript{10}Green (1986b) provides necessary and sufficient conditions for portfolios on the minimum-variance frontier to have positive weights in all assets.
MacKinlay and Pástor (2000) recognize that a missing factor implies the presence of the pricing error in the residual covariance matrix and that taking this into account in the estimation improves portfolio selection. They also find that, even if the true covariance matrix of returns is not the identify matrix, using it as the covariance matrix, which is another kind of restriction, improves portfolio performance. Pettenuzzo, Timmermann, and Valkanov (2014) demonstrate that economic constraints such as restricting the equity risk premium to be positive and bounding the Sharpe ratio improve the estimation of time-series forecasts of the equity risk premium. In our work, the constraint to be imposed when estimating asset returns follows directly from the APT.

As highlighted by the literature described above, an investor choosing the optimal portfolio weights faces both model and parameter uncertainty.\(^\text{11}\) Two features of the APT allow us to deal with both sources of uncertainty. The first feature is the presence of the pricing error, \(\alpha\). Allowing for the presence of \(\alpha\) resolves model uncertainty because it tells us how the incorrectly-specified mean and the variance of returns are to be adjusted in the presence of model uncertainty.

The second feature of the APT is the restriction on the magnitude of the pricing error: \(\alpha'\Sigma^{-1}\alpha \leq \delta < \infty\), where \(\alpha\) is the vector of pricing errors, \(\Sigma^{-1}\) is the inverse of the covariance matrix of residuals and \(\delta\) is an arbitrary positive constant. By imposing the APT restriction when estimating the parameters, one ensures that the “approximating model” lies within the set of no-arbitrage models.\(^\text{12}\) This restriction limits the magnitude of estimates for \(\alpha\), and hence, allows one to deal also with parameter uncertainty associated with \(\alpha\).\(^\text{13}\) Interestingly, imposing the APT restriction is analogous to the approach adopted in Garlappi, Uppal, and Wang (2007) where one accounts for parameter uncertainty using the minmax approach originally proposed in Gilboa and Schmeidler (1989). In the minmax approach, portfolio weights are obtained by first minimizing the mean-variance objective function, equivalent to minimizing the Sharpe ratio, over the set of expected returns subject to the constraint that these expected returns are not too distant from the estimated mean returns, and then maximizing over the portfolio weights—see, for example, Garlappi, Uppal, and Wang (2007, their equation (14)). Analogously, our approach can be interpreted as first

\(^{11}\)Model uncertainty arises when the investor is not confident about the true data-generating process. Parameter uncertainty, on the other hand, refers to not knowing the true parameter values of the data-generating process. Thus, parameter uncertainty arises even in the absence of model uncertainty: that is, even if one knows the true data-generating process, one may not know the parameter values for this process. Conditional on knowing the true model, one can always resolve parameter uncertainty with an infinite number of observations.

\(^{12}\)Adopting the terminology of Hansen and Sargent (2008), this means that any “approximating model” that is estimated with the data must lie within the set bounded by the APT restriction.

\(^{13}\)The uncertainty arising from having to estimate the parameters associated with the beta component of returns, referred to as beta misspecification, is dealt with using asymptotic analysis as described later in the text.
estimating the parameters subject to the constraint that the maximum-likelihood estimates satisfy the APT restriction, which is equivalent to restricting the Sharpe ratio of the alpha portfolio. We then plug in the estimated parameters into the mean-variance objective function and choose portfolio weights to maximize the Sharpe ratio. Thus, the key difference between our approach and that of the minmax approach is that we constrain the alpha component of expected returns, while the minmax approach constrains the total expected returns. Garlappi, Uppal, and Wang (2007) and DeMiguel, Garlappi, Nogales, and Uppal (2009) provide also a Bayesian interpretation of the portfolio weights that account for parameter uncertainty. Specifically, all these approaches—Bayesian, minmax, and imposing the APT restriction—that address the problem of parameter uncertainty lead to shrinkage-type estimators; see, for example, Bawa, Brown, and Klein (1979), Jorion (1986), Pástor (2000), Pástor and Stambaugh (2000), and Garlappi, Uppal, and Wang (2007).

3 Notation, Factor Model for Returns, and APT

In this section, we explain our notation and the linear factor model we use for returns. Our analysis is founded on precisely the same assumptions as the ones underlying the APT. In particular, we use the same model for returns as the APT, we define alpha as in the APT, and we use the APT no-arbitrage restriction in our empirical analysis. After stating the assumptions and main result of the APT, we show how the APT can be extended to incorporate the case with large pricing errors, that is, where the covariance matrix of returns has unbounded eigenvalues implying that even a well-spread portfolio will not be diversified, in addition to the standard case where the eigenvalues are bounded. We conclude this section by describing how the APT model can capture misspecification arising from the alpha and beta components of returns.

3.1 Notation

Let \( r_f \) denote the return on the risk-free asset and let the \( N \)-dimensional vector \( r_t = (r_{1t}, r_{2t}, \ldots, r_{Nt})' \) denote the vector of returns on risky assets.

Given an arbitrary portfolio strategy \( a \) with weights \( w^a = (w_1^a, w_2^a, \ldots, w_N^a)' \) of \( N \) risky assets, and using \( e \) to denote an \( N \)-dimensional vector of ones, we define the associated portfolio return as

\[
r_t^a = r_f'w^a + r_f(1 - e'w^a),
\]
with finite conditional mean, standard deviation, and Sharpe ratio indicated by

\[ \mu^a = E(r^a_t) = E(r_t)'w^a + r_f(1 - c'w^a), \]

\[ \sigma^a = \sqrt{\text{var}(r^a_t)}, \]

\[ SR^a = \frac{\mu^a - r_f}{\sigma^a}. \]

### 3.2 Linear factor model for asset returns with misspecification

We start our analysis with the assumption of a linear factor structure for returns. Just as in Chamberlain (1983), Chamberlain and Rothschild (1983), and Ingersoll (1984), we do not restrict the covariance matrix of the residuals, \( \Sigma \), to be diagonal; that is, we allow for correlated error terms. Furthermore, in contrast to Chamberlain (1983) and Chamberlain and Rothschild (1983), we study also the case where not all eigenvalues of \( \Sigma \) are restricted to be bounded when \( N \) is large.

**Assumption 3.1 (Linear factor model).** We assume the \( N \)-dimensional vector \( r_t \) of asset returns can be characterized by the following linear factor model

\[ r_t = \mu + B z_t + \varepsilon_t, \]  

where \( z_t = (z_{1t}, z_{2t}, ..., z_{Kt})' \) is the \( K \times 1 \) vector of common unobserved factors, assumed to be with zero (conditional) mean, without loss of generality, as explained below; \( B = (\beta_1, \beta_2, \ldots, \beta_N)' \) is an \( N \times K \) full-rank matrix of factor loadings with \( i \)th row \( \beta_i' \); \( \varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})' \) is an \( N \times 1 \) vector of residuals, and the \( N \times 1 \) vector \( \mu \) represents the mean of the vector of returns, \( r_t \).\(^{14}\) At any time \( t \),

\[ z_t \sim (0, \Omega) \quad \text{and} \quad \varepsilon_t \sim (0, \Sigma), \]

where the \( K \times K \) covariance matrix \( \Omega \) and \( N \times N \) covariance matrix \( \Sigma \) are positive definite. Moreover, \( \varepsilon_t \) and \( z_t \) are uncorrelated; that is, \( E(\varepsilon_t z_t') = 0 \).

Assumption 3.1 implies that the conditional variance-covariance matrix for asset returns is:

\[ E[(r_t - \mu)(r_t - \mu)'] = V = B \Omega B' + \Sigma. \]

Notice that in the original formulation of the APT in (1), no factor is assumed to be observed; instead, the error term \( r_t - \mu = B z_t + \varepsilon_t \) has a latent factor structure through the

\(^{14}\)Throughout the paper, it will be assumed that \( K < N \).
common innovation $z_t$. In contrast, nowadays factor models are typically used assuming the existence of a given number of priced observed factors. This means that rather than the model in (1), one considers

$$r_t - r_f e = B(f_t - r_f e_K) + \varepsilon_t$$  \hspace{1cm} (2)

for $K$-dimensional factors $f_t$, all of which are observed. Notice that (2) and (1) are equivalent if $f_t - r_f e_K = \lambda + z_t$ and $\mu - r_f e = B\lambda$.

3.3 Arbitrage Pricing Theory (APT)

In the definition below, as well as throughout the paper, we use $\delta$ to denote an arbitrary positive scalar.

Definition 3.1 (Asymptotic arbitrage). A sequence of portfolios is said to generate an asymptotic arbitrage opportunity if along some subsequence $N'$:

$$\text{var}(r'w^a) \to 0 \quad \text{as} \quad N' \to \infty \quad \text{and} \quad (\mu - r_f e)'w^a \geq \delta > 0 \quad \text{for all} \quad N'.$$

Assumption 3.2 (No asymptotic arbitrage). There are no asymptotic arbitrage opportunities.

We now state the APT result. This result is derived in Huberman (1982) and Ingersoll (1984), so we do not include the proof.

Theorem 3.1 (APT (Huberman, 1982; Ingersoll, 1984)). Assumptions 3.1 and 3.2 imply that for all $N$ there exists some positive number $\delta$ such that the weighted sum of the squared pricing errors is uniformly bounded:

$$\hat{\alpha}'\Sigma^{-1}\hat{\alpha} \leq \delta < \infty,$$

where the vector of pricing errors is

$$\hat{\alpha} = (\mu - r_f e) - B\hat{\lambda},$$  \hspace{1cm} (4)

and the vector of risk premia is

$$\hat{\lambda} = (B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}(\mu - r_f e).$$  \hspace{1cm} (5)

Remark 3.1.1. Observe that $\hat{\lambda}$ is the projection coefficient when projecting $(\mu - r_f e)$ on $B$ and $\hat{\alpha}$ is the projection residual that satisfies

$$B'\Sigma^{-1}\hat{\alpha} = 0.$$  \hspace{1cm} (6)

Note that the ^ symbol in $\hat{\lambda}$ and $\hat{\alpha}$ is used to indicate that these are obtained from a projection; the ^ symbol does not mean that these quantities are estimates.
Remark 3.1.2. We interpret the projection residual $\hat{\alpha}$ as the “pricing error.” For instance, Roll (1980) points out that such a pricing error could arise in the Black (1972) linear model of returns if the index with respect to which risk (beta) is measured is not mean-variance efficient. This could happen if the theory was correct but the true value-weighted market portfolio was measured with error (Roll (1977), Green (1986a)). But, the pricing error could arise also if the index was measured without error but the theory was incorrect. These pricing errors would also be present if the asset-pricing model had missing factors (MacKinlay and Pástor, 2000) or if some of the factors were latent (Connor and Korajczyk, 1986; Lehmann and Modest, 1988).

3.4 Extending the APT

In the sections above, we have described results from the existing literature. In this section, we describe our first contribution, which is to extend the APT to the case where the residual covariance matrix does not necessarily have bounded eigenvalues; that is, even a well-spread portfolio may not be well-diversified. This occurs precisely because of model misspecification.

Just like in Chamberlain and Rothschild (1983) and Ingersoll (1984), we study a market with an infinite number of assets.\(^\text{15}\) That is, instead of considering a sequence of distinct economies, we consider a fixed infinite economy in which we study a sequence of nested subsets of assets. That is, for the $N$th step, one new asset is added to the first $N-1$ assets whose parameters do not change with the addition of the new assets. These unchanging parameters can be interpreted as the parameters one would get in the limit as the number of assets becomes asymptotically large.

We adopt the following notation. Consider a symmetric $M \times M$ matrix $A$. Let $g_{iM}(A)$ denote the $i$-th eigenvalue of $A$ in decreasing order for $1 \leq i \leq M$. Thus, the maximum eigenvalue is $g_{1M}(A)$ and the minimum eigenvalue is $g_{MM}(A)$.

We start with the definition of a regular factor economy.

Definition 3.2 (Regular factor economy (Ingersoll, 1984, p. 1028 and footnote 10)). A factor economy is regular if the maximum eigenvalue $g_{1K}(B_N^T \Sigma_N^{-1} B_N)^{-1} \rightarrow 0$ as $N \rightarrow \infty$. If the limit is positive instead of zero, then the factor representation is irregular. Equivalently, the minimum eigenvalue $g_{KK}(B_N^T \Sigma_N^{-1} B_N)$ is diverging as $N \rightarrow \infty$.

\(^{15}\)For this reason, we index the parameters that depend on the number of assets by the subscript $N$, except for random variables, such as $\varepsilon_t$ and $r_t$, by $N$.  

13
The above definition implies that in a regular economy the risk arising from factors cannot be diversified away; see Connor, Goldberg, and Korajczyk (2010) for additional details.

Next, we introduce the limit of \( \hat{\lambda} \) as \( N \to \infty \), which we label \( \lambda \), and the associated vector of pricing errors, \( \alpha_N = (\mu_N - r_f e) - B_N \lambda \).\(^{16}\) Under Assumptions 3.1 and 3.2 and the assumption of a regular factor economy, Ingersoll (1984, Theorem 3 and fttn. 10) shows that \( \lambda \) is unique and prices assets with bounded squared error:

\[
\alpha_N' \Sigma_N^{-1} \alpha_N \leq \delta < \infty,
\]

where \( \delta \) represents some positive arbitrary scalar. Observe that the APT condition in (3) is expressed in terms of \( \hat{\alpha} \), while the condition in (7) is expressed in terms of \( \alpha_N \); in Appendix B.4, we show the equivalence between these two conditions.

Observe that, as discussed in Ingersoll (1984, footnote 10), the weighting of the squared pricing error uses \( \Sigma_N^{-1} \). The intuition for why the squared-error bound in (7) is weighted by (the inverse of) the matrix of residual variances, \( \Sigma_N \), is that an asset’s usefulness in arbitrages is limited by its residual variance. As Ingersoll (1984) explains: “Roughly speaking, the smaller an asset’s residual variance, the more extreme its weighting can be in a portfolio, while maintaining diversification, and the more extreme is its weighting, the bigger is its effect on the portfolio’s expected return. Thus, an asset with a small residual variance can have a major impact on a diversified portfolio’s expected return, and to prevent arbitrage its own expected return must be more nearly in line with the prediction of the APT. Assets with larger residual variances can only have small effects on a diversified portfolio’s expected return, so their own expected returns need not be as closely in line with the prediction.”

The condition in (7), which is a consequence of asymptotic no arbitrage, links the pricing error \( \alpha_N \) to the idiosyncratic covariance matrix, \( \Sigma_N \). There are two possible, complementary, cases for \( \Sigma_N \) as \( N \to \infty \) that we will examine. In the first case, for large \( N \) all the eigenvalues of \( \Sigma_N \) are bounded, and in the second case, at least one of the eigenvalues is unbounded. The existing APT literature has focused on studying the case in which all the eigenvalues of \( \Sigma_N \) are bounded, which—for completeness—is restated in the theorem below, with the proof for this provided in Huberman (1982) and Ingersoll (1984).

**Theorem 3.2** (Constraint imposed on \( \alpha_N \) by no arbitrage for case with bounded eigenvalues; Huberman (1982); Ingersoll (1984)). *If \( \Sigma_N \) has bounded eigenvalues for large \( N \), then the condition in (7) requires the elements of the pricing-error vector \( \alpha_N \) to become small*

\(^{16}\)Note that \( \alpha_N \) is an \( N \)-dimensional vector that is a component of the infinite-dimensional vector \( \alpha \); as \( N \) increases, the number of elements in \( \alpha_N \) increases, but the elements themselves do not change.
for large \(N\) in the following sense:

\[
\alpha' N \alpha N \leq g_{1N}(\Sigma_N)(\alpha' N \Sigma^{-1}_N \alpha N) \leq \delta < \infty,
\]

but without \(\alpha N\) being tied down to \(\Sigma N\).

The assumption that the covariance matrix for the residuals of returns, \(\Sigma N\), has bounded eigenvalues for large \(N\) implies that the return of any well-diversified portfolio (sometimes referred to as a “well-spread” portfolio) has zero residual risk in the limit. The above theorem then says that the squared sum of the pricing errors is bounded, which is equivalent to saying that the squared pricing errors, \(\alpha N\), get smaller as the number of assets increases.

We now show that the APT model in the existing literature can be extended to the case where, as \(N\) increases, some of the eigenvalues of \(\Sigma N\) are not bounded. This typically occurs when there are missing or mismeasured factors; see Section 3.5.

**Theorem 3.3** (Constraint imposed on \(\alpha N\) by no arbitrage for case with unbounded eigenvalues). Suppose that the vector of asset returns, \(r_t\), satisfies Assumptions 3.1 and 3.2. Suppose that for some finite \(1 \leq p < N\) the following three conditions hold: (i) \(\sup N g_p N(\Sigma N) = \infty\); (ii) \(\sup N g_{p+1 N}(\Sigma N) \leq \delta < \infty\); and, (iii) \(\inf N g_{NN}(\Sigma N) \geq \delta > 0\). Then, the APT condition in equation (7) is satisfied by the pricing error \(\alpha N\), represented as

\[
\alpha N = A_N m + a_N,
\]

where \(A_N\) is a \(N \times p\) matrix whose \(j\)th column equals \(g_{jN}^{\frac{1}{2}} v_{jN}(\Sigma N)\), where \(1 \leq j \leq p\), \(v_{jN}(\Sigma N)\) is the eigenvector of \(\Sigma N\) associated with the eigenvalue \(g_{jN}(\Sigma N)\), \(m\) is some \(p \times 1\) vector, and \(a_N\) is some non-zero \(N \times 1\) vector that satisfies

\[
a' N \Sigma^{-1}_N a N \leq \delta < \infty.
\]

One can interpret the two components of \(\alpha N\) in (8) in a variety of ways. The first term, \(A_N m\), could be associated with \(p\) latent or missing pervasive factors, where \(A_N\) are the factor loadings and \(m\) are the risk premia for the missing factors.\(^{17}\) The second term, \(a_N\), is the idiosyncratic part of the pricing error \(\alpha N\); for instance, \(a_N\) could be interpreted as representing managerial skills or views of analysts. Under Assumptions 3.1 and 3.2, the expected excess return can be written as:

\[
E(r_t - r e) = \mu N - r e = \alpha N + B_N \lambda = (A_N m + a_N) + B_N \lambda.
\]

\(^{17}\)Pervasiveness of the latent missing factors with loadings \(A_N\) follows from the fact that \(g_{jN}\) for \(1 \leq j \leq p\) are diverging for large \(N\).
Observe that, compared to equation (2) where we specify the model with respect to observed factors, equation (9) contains the extra term $\alpha_N = (A_N m + a_N)$, which represents the effect of misspecification of the priced observed factors.

More importantly, Theorem 3.3 shows that the common perception that the pricing errors in $\alpha_N$ need to be small in the APT is not accurate. In particular, if the maximum eigenvalue of the residual covariance matrix is not bounded, then the pricing errors can also be large without violating the no-arbitrage condition given in (7). What Theorem 3.3 states is that if the maximum eigenvalue of $\Sigma_N$ is asymptotically unbounded, then the contribution of the pricing error to the portfolio return could be large, but for this to satisfy the no-arbitrage condition, any portfolio earning this high return would not be well diversified and would be bearing idiosyncratic risk. To see this, observe that by Chamberlain and Rothschild (1983, Theorem 4), under the assumptions made for deriving Theorem 3.3, the covariance matrix of residuals, $\Sigma_N$, has the following approximate $p$-factor structure:

$$\Sigma_N = A_N A_N' + C_N,$$

where $C_N$ is a $N \times N$ positive semi-definite matrix with bounded eigenvalues. Therefore, the residual variance of the return on any portfolio weights $w_N$ is given by

$$w_N' \Sigma_N w_N = w_N' A_N A_N' w_N + w_N' C_N w_N.$$

Whereas the second term, $w_N' C_N w_N$, goes to zero for well-spread portfolios (that is, for portfolios with $w_N' w_N \to 0$), there is no guarantee that the same occurs for the first term, $w_N' A_N A_N' w_N$.

We wish to extend the definition of a regular economy in two different dimensions. The earlier definition of regularity was applied to $B_N$ and $\Sigma_N$. We extend the definition to any arbitrary matrix of dimension $N \times K$, such as $D_N$, and an arbitrary positive-definite $N \times N$ matrix, $C_N$. Second, we impose that the eigenvalues are diverging at precisely the same rate.

**Definition 3.3 (C_N-regularity).** A matrix $D_N$ is C_N-regular if there exists an increasing function of $N$, $f(N)$, such that for any $1 \leq j \leq K$, the eigenvalues $g_{jK}(\frac{1}{f(N)}D_N' C_N^{-1} D_N) \to \delta_j > 0$, where $\delta_j$ is some finite positive constant.

Ingersoll (1984, p. 1026) defines the setting of Theorem 3.2 as one with bounded residual variation; we label the setting of Theorem 3.3 as one with unbounded residual variation.

---

18 Under the assumptions of Chamberlain and Rothschild (1983, Theorem 4), $C_N$ will be non-singular for a sufficiently large $N$, which implies that $\Sigma_N$ is also nonsingular for a sufficiently large $N$.

19 A special case of the definition below is the notion of pervasive factors defined in Connor and Korajczyk (1986, Assumption 6) and Connor, Goldberg, and Korajczyk (2010, p. 85). Both papers use $f(N) = N$, but Connor and Korajczyk require all the eigenvalues to diverge at least at that rate, whereas Connor, Goldberg, and Korajczyk is similar to our definition in the sense that it requires all eigenvalues to diverge at precisely the same rate.
3.5 Different forms of model misspecification

In this section, we describe three forms of model misspecification that are captured by the framework we have described above. The first is related to the “beta” component of returns. The other two are related to the “alpha” component of returns and can be interpreted as special cases of bounded and unbounded residual variation. Throughout this section, assume that the true model satisfies

\[ r_t - r_f e = \alpha^* + B^* (f^*_{t} - r_f e_{K^*}) + \varepsilon_t \]  

(11)

for \( K^* \)-dimensional factors, of which \( 0 \leq K \leq K^* \) are observed. In the above equation, \( f^*_t \) are the set of all common factors affecting returns, \( B^* \) are the true loadings on these factors, and \( \alpha^* \) are the true firm-specific components of expected returns that satisfy the APT condition. For expositional ease, we will assume that \( \alpha^* \) is zero in this section.

**Case 1: Incorrect means or covariances for the factors**

We start by considering the situation in which all of the \( K^* \) factors are observed, implying that \( K^* = K \) and that there is no error in measuring these factors because our objective is to study misspecification in the “beta” component of returns; that is, misspecification in the assumed distribution of the factors, namely \( \lambda \) or \( \Omega \). In particular, suppose that we specify the mean and covariance matrix of the observed factors incorrectly. This implies that the portfolio weights will also be incorrectly specified, even in population, despite the fact that the model in (11) can be correctly estimated using OLS to obtain estimates of \( B^* \) and \( \Sigma \), where \( \Sigma \) is the covariance matrix of the residuals.

**Case 2: Missing factors**

Suppose now that of the \( K^* \) factors, only \( K \) are observed and \( p = K^* - K \) are missing and suppose for simplicity that the observed and missing factors are uncorrelated. Then, the model in (11) can be re-written as follows:

\[ r_t - r_f e = A_{\text{miss}} (f_{\text{miss},t} - r_f e_p) + B (f_t - r_f e_K) + \varepsilon_t, \]

where

\[ f^*_t = \begin{pmatrix} f_{\text{miss},t} \\ f_t \end{pmatrix} \hspace{1cm} \text{and} \hspace{1cm} B^* = \begin{pmatrix} A_{\text{miss}} \\ B \end{pmatrix}, \]

and \( f_{\text{miss},t} \) are the missing factors. Given \( f_{\text{miss},t} - r_f e_p = \lambda_{\text{miss}} + z_{\text{miss},t} \), it follows that

\[ r_t - r_f e = A_{\text{miss}} \lambda_{\text{miss}} + B (f_t - r_f e_K) + (\varepsilon_t + A_{\text{miss}} z_{\text{miss},t}), \]

(12)
where \(E[z_{\text{miss},t}] = 0\) and \(E[z_{\text{miss},t} z'_{\text{miss},t}] = \Omega_{\text{miss}}\). There are two possible cases for the set of missing factors, \(f_{\text{miss},t}\). Some of them, for instance \(f_{\text{miss1},t}\) could possibly be pervasive while others, \(f_{\text{miss2},t}\), are non-pervasive, where in turn we can split the columns of \(A_{\text{miss}} = (A_{\text{miss1}}; A_{\text{miss2}})\) and analogously for \(\lambda_{\text{miss}}\) and \(z_{\text{miss},t}\). Therefore, the model in (12) coincides with the model in (9)–(10), once we set

\[
A = A_{\text{miss1}}, \quad m = \lambda_{\text{miss1}}, \quad a = A_{\text{miss2}} \lambda_{\text{miss2}}, \quad \lambda = E[(f_t - r_f e_K)], \quad \\
\Omega = E[z_{\text{miss1},t} z'_{\text{miss1},t}], \quad C = E[\varepsilon_t \varepsilon'_t] + A_{\text{miss2}} (E[z_{\text{miss2},t} z'_{\text{miss2},t}]) A_{\text{miss2}},
\]

where, for expositional convenience, we have assumed that the factors \(f_{\text{miss1},t}\) and \(f_{\text{miss2},t}\) are uncorrelated. Note that the non-pervasive “missing factors” referred to in the above example include also firm-specific characteristics, such as managerial skill. Usually such characteristics are modeled without imposing a factor structure.\(^{21}\)

**Case 3: Mismeasured factors**

Finally, consider the case where all \(K^*\) factors are measured with error. In particular, the observed factors satisfy \(f_t = f'_t + \eta_t\), where the measurement error \(\eta_t\) has mean \(E[\eta_t] = \mu_\eta\) and covariance matrix \(E[(\eta_t - \mu_\eta)(\eta_t - \mu_\eta)'] = \Sigma_\eta\).\(^{22}\) Then, the model in (11) can be re-written as follows:

\[
r_t - r_f e = -B \mu_\eta + B (f_t - r_f e_K) + (\varepsilon_t - B (\eta_t - \mu_\eta)). \tag{13}
\]

Therefore, the model in (13) coincides with the model in (9)–(10), once we set

\[
A = -B, \quad m = \mu_\eta, \quad a = 0, \quad \lambda = E[(f_t - r_f e_K)], \quad \\
\Omega = \Sigma_\eta, \quad C = E[\varepsilon_t \varepsilon'_t].
\]

The econometric problem in estimating (13) is akin to an errors-in-variables problem: the residual is correlated with the observed factors through the measurement error, \(\eta_t\). Observe also that there is a constraint between \(\alpha\) and \(B\) (in fact, \(B\) appears in three terms of the model in (13)), which can be used to test for the presence of mismeasurement and can also be exploited in the estimation.

\(^{20}\)Using our definition of regularity, one obtains, as \(N \to \infty\) that \(g_{112} (A_{\text{miss1}} (E[\varepsilon_t \varepsilon'_t])^{-1} A_{\text{miss1}}) \to \infty\) and \(g_{122} (A_{\text{miss2}} (E[\varepsilon_t \varepsilon'_t])^{-1} A_{\text{miss2}}) \leq \delta < \infty\), where \(E[\varepsilon_t \varepsilon'_t]\) has bounded maximum eigenvalue and \(p_1 + p_2 = p\).

\(^{21}\)Here, non-pervasiveness allows us to retain the factor structure even for such characteristics. In particular, note that \(\Omega_{\text{miss2}}\) could be equal to 0, implying that \(A_{\text{miss2}}\) and \(\lambda_{\text{miss2}}\) cannot be identified separately.

\(^{22}\)Of course, it is possible that some of the factors are measured without any error; in that case the means and variances of the \(\eta_t\) associated with these factors are zero.
4 Mitigating Model Misspecification in Mean-Variance Portfolios

We study the weights and returns for the family of mean-variance portfolios. In particular, the portfolios we study are: (1) the mean-variance efficient portfolio when a risk-free asset is available, $w_{mv}$; (2) the global minimum-variance portfolio when a risk-free asset is not available, $w_{gmv}$; and (3) the mean-variance efficient portfolios in the absence of a risk-free asset, $w_{fp}$, which are the Markowitz frontier portfolios that have the smallest variance for a given target mean. These three portfolio are displayed in Figure 1 and the asymptotic behavior of these portfolios as the number of available assets increases is displayed in Figure 2.23

For each of these three portfolios, we treat model misspecification in three steps. First, in Section 4.1, we show how the weights and returns of these portfolios can be decomposed into two components: a component that depends on the risk premia (beta) and another component that depends on the pricing error (alpha). Second, in Section 4.2, we show how misspecification in the beta component of returns can be mitigated by studying the case where the number of assets increases asymptotically, in which case the portfolio weights become functionally independent of the distribution of the observed factors, $\lambda$ and $\Omega$. Third, in Section 4.3, we show how misspecification in the alpha component of returns can be mitigated by imposing the APT restriction on the estimation and prove that this restriction coincides with the identification condition of the model for asset returns; we also show that this restriction leads to an improvement in the estimation of the model parameters because it leads to shrinkage of some of the model parameters.

Below, we provide the details of these three steps for the mean-variance efficient portfolio in the presence of a risk-free asset. To conserve space, the analysis of the global-minimum-variance portfolio (the results for which are stronger than those for the mean-variance portfolio because the global-minimum-variance portfolio depends only on covariances of returns and not on expected returns) and the frontier portfolios (the results for which are a combination of those for the mean-variance and global-minimum-variance portfolios) is presented in Appendices B.1 and B.2.

23 Figure 2 shows that as the number of assets increases, the volatility and expected return of the global minimum-variance portfolio decline, and it converges to the risk-free asset. Similarly, the frontier portfolios converge to the capital market lines. This linearity for large $N$ allows the use of Hilbert-space mathematics in Chamberlain (1983).
4.1 Decomposing the mean-variance portfolio

The mean-variance efficient portfolio in the presence of a risk-free asset is defined by the solution to the following optimization problem:

\[ w_{mv} = \arg\max_w \left( (w'\mu + (1 - w'e)r_f) - \frac{\gamma}{2}w'Vw \right), \]

where \( 0 < \gamma < \infty \) is the coefficient of risk aversion; \( w_{mv} = (w_{mv1}, \ldots, w_{mvN})' \) is the vector of portfolio weights in the \( N \) risky assets; and, the investment in the risk-free asset is given by \( 1 - e'w_{mv} \). Alternatively, if one wished to formulate the above problem in terms of a constraint that required the portfolio to achieve a target mean of \( \mu^* \), one needs to set

\[ \gamma = \frac{(\mu - r_f e)'V^{-1}(\mu - r_f e)}{\mu^* - r_f}. \]

The solution to the above optimization problem is:

\[ w_{mv} = \frac{1}{\gamma}V^{-1}(\mu - r_f e). \]

By standard arguments, its associated portfolio return has conditional mean, standard deviation, and Sharpe ratio given by the following three expressions:

\[ \mu_{mv} - r_f = \gamma^{-1}(\mu - r_f e)'V^{-1}(\mu - r_f e), \quad (14) \]
\[ \sigma_{mv} = \gamma^{-1}(\mu - r_f e)'V^{-1}(\mu - r_f e)^{1/2}, \quad (15) \]
\[ SR_{mv} = \left( (\mu - r_f e)'V^{-1}(\mu - r_f e) \right)^{1/2}. \quad (16) \]

The following theorem, which is valid for any finite \( N \), establishes the relations that exist across the mean-variance portfolio, \( w_{mv} \), and two portfolios, \( w^\alpha \) and \( w^\beta \), of which \( w^\alpha \) depends on only the pricing error and \( w^\beta \) depends only on factor risk premia. The mean-variance portfolio and its decomposition into the “alpha” and “beta” portfolios, is displayed in Figure 4.

**Theorem 4.1** (Decomposing weights of mean-variance portfolio). *Suppose that the vector of asset returns, \( \mathbf{r}_t \), satisfies Assumptions 3.1 and 3.2. Then for any finite \( N > K \) and

\[ w^{lan} = w_{mv} e'w_{mv} = V^{-1}(\mu - r_f e) = \frac{e'V^{-1}(\mu - r_f e)}{e'V^{-1} e} \]

A special case of the mean-variance portfolio is the tangency portfolio, which has zero wealth invested in the risk-free asset.
\( \mu^*>r_f \), the mean-variance portfolio weights satisfies the following decomposition:

\[
\mathbf{w}^{\text{mv}} = \delta^\alpha \mathbf{w}^\alpha + \delta^\beta \mathbf{w}^\beta,
\]

where

\[
\mathbf{w}^\alpha = \frac{1}{\gamma^\alpha} \Sigma^{-1} \hat{\alpha}, \quad (17)
\]
\[
\mathbf{w}^\beta = \frac{1}{\gamma^\beta} \mathbf{V}^{-1} \hat{B} \lambda, \quad (18)
\]

with \( \gamma^\alpha = \frac{\hat{\alpha}' \Sigma^{-1} \hat{\alpha}}{\mu^* - r_f e}, \gamma^\beta = \gamma - \gamma^\alpha = \frac{\hat{\lambda}' \mathbf{V}^{-1} \hat{B} \lambda}{\mu^* - r_f e}, \delta^\alpha = \frac{\gamma^\alpha}{\gamma} = \frac{\hat{\alpha}' \Sigma^{-1} \hat{\alpha}}{(\mu - r_f e) \mathbf{V}^{-1} (\mu - r_f e)}, \) and \( \delta^\beta = \frac{\gamma^\beta}{\gamma} = 1 - \delta^\alpha. \) Furthermore, the portfolios \( \mathbf{w}^\alpha \) and \( \mathbf{w}^\beta \) satisfy the orthogonality condition,

\[
(\mathbf{w}^\alpha)' \mathbf{V} \mathbf{w}^\beta = (\mathbf{w}^\alpha)' \Sigma \mathbf{w}^\beta = 0, \quad (19)
\]

and \( \mathbf{w}^\beta \) is the minimum-variance portfolio that is orthogonal to \( \mathbf{w}^\alpha \), and \( \mathbf{w}^\alpha \) is the minimum-variance portfolio orthogonal to \( \mathbf{w}^\beta \).

**Remark 4.1.1.** One can interpret \( \gamma^\alpha \) as the ratio of the share of the contribution of \( \mathbf{w}^\alpha \) to the expected return on the Markovitz portfolio with unit risk aversion, \( (\mu - r_f e) \mathbf{V}^{-1} (\mu - r_f e) \), over the target excess mean return, \( \mu^* - r_f \).

**Remark 4.1.2.** Note that the portfolio \( \mathbf{w}^\alpha \), defined in (17), depends only on the pricing error but not on the risk premia or the factor-covariance matrix, which is why we label this portfolio as the “alpha” portfolio. Analogously, the portfolio strategy \( \mathbf{w}^\beta \), defined in (18), depends on factor exposures and their risk premia, but not on the pricing errors.

**Remark 4.1.3.** The orthogonality condition in (19) above says that the two portfolios \( \mathbf{w}^\alpha \) and \( \mathbf{w}^\beta \) are uncorrelated, both conditional on the factors and also unconditionally. For practical portfolio construction, what this implies is that if the alpha portfolio is constructed using the method described above, then it will not have any exposure to factor risk.

**Remark 4.1.4.** The last part of the theorem states that, in addition to \( \mathbf{w}^\alpha \) and \( \mathbf{w}^\beta \) being orthogonal to each other, if one searched for the minimum-variance portfolio that is orthogonal to \( \mathbf{w}^\alpha \), the resulting portfolio would be \( \mathbf{w}^\beta \), and vice versa. That is, even though the \( \mathbf{w}^\alpha \) and \( \mathbf{w}^\beta \) portfolios are obtained simply by relying on the APT decomposition of the total mean return into the component that depends on the pricing error and the component that depends on the factor exposure, these portfolios can be characterized also as being the result of an optimization, which is described in Appendix B.3 that extends Roll (1980) to the case where, in addition to investing in risky assets, one can invest also in a risk-free asset. Note that the role of the \( \delta^\alpha \) and \( \delta^\beta \) coefficients is to ensure that the \( \mathbf{w}^\alpha \) and \( \mathbf{w}^\beta \)
portfolios have the same target mean as \( w^{mv} \). Observe that three-fund separation holds; that is, the Capital Market Line is spanned by a linear combination of the risk free asset and any inefficient \( w^\alpha \) and \( w^\beta \) portfolios.

We now characterize the returns of the two components of the mean-variance portfolio.

**Theorem 4.2** (Decomposing returns of mean-variance portfolio). Suppose that the vector of asset returns, \( r_t \), satisfies Assumptions 3.1 and 3.2 and \( \alpha_N \neq 0 \). Then for any finite \( N > K \), and assuming \( \mu^* > r_f \), the returns on the portfolios \( w^\alpha \) and \( w^\beta \) have mean, volatility (standard deviation), and Sharpe ratio that have the same functional form as the corresponding expressions for \( w^{mv} \) given in equations (14), (15), and (16). Moreover, these three quantities are functions of the same quadratic form, and satisfy, respectively:

\[
\begin{align*}
\mu^\alpha - r_f &= \frac{1}{\gamma^\alpha} (\hat{\alpha}' \Sigma^{-1} \hat{\alpha})^\frac{1}{2} = \mu^* - r_f; \\
\mu^\beta - r_f &= \frac{1}{\gamma^\beta} (\hat{\lambda}' \hat{B}' \hat{V}^{-1} \hat{B} \hat{\lambda})^\frac{1}{2} = \mu^* - r_f; \\
\sigma^\alpha &= \frac{1}{\gamma^\alpha} (\hat{\alpha}' \Sigma^{-1} \hat{\alpha})^\frac{1}{2}; \\
\sigma^\beta &= \frac{1}{\gamma^\beta} (\hat{\lambda}' \hat{B}' \hat{V}^{-1} \hat{B} \hat{\lambda})^\frac{1}{2}; \\
SR^\alpha &= (\hat{\alpha}' \Sigma^{-1} \hat{\alpha})^\frac{1}{2}; \\
SR^\beta &= (\hat{\lambda}' \hat{B}' \hat{V}^{-1} \hat{B} \hat{\lambda})^\frac{1}{2}.
\end{align*}
\]

Moreover, the Sharpe ratios satisfy \( 0 \leq SR^\alpha < \infty \) and \( 0 \leq SR^\beta < \infty \) and

\[
(SR^{mv})^2 = (SR^\alpha)^2 + (SR^\beta)^2.
\]

**Remark 4.2.1.** Observe that the excess return on the portfolio \( w^\alpha \) is independent of the distribution of the factors. The portfolio \( w^\alpha \) can be interpreted as one that is “factor neutral;” that is, it is the portfolio whose return is not exposed to the risks arising from the common factors, \((r_t - r_f e)'w^\alpha = (\hat{\alpha} + \varepsilon_t)' \Sigma^{-1} \hat{\alpha}\), which follows from (6).

**Remark 4.2.2.** The expected excess return for the \( w^\alpha \) and \( w^\beta \) portfolios equal the target mean in excess of the risk-free rate, \( \mu^* - r_f \), by construction. However, their Sharpe ratios are positive because they are given by quadratic forms.

**Remark 4.2.3.** An important insight is that the quantity on the left-hand side of the APT restriction in (3) is exactly the expression one would get from squaring the Sharpe ratio for the \( w^\alpha \) portfolio, \( SR^\alpha \), in (22). Thus, the APT restriction in (3), which is typically interpreted as a bound on the pricing errors, can instead be interpreted as a bound on the Sharpe ratio of the \( w^\alpha \) portfolio. The decomposition of the square of the Sharpe ratio of the mean-variance portfolio in (23) is obtained also in Treynor and Black (1973) for the case of the single index model with a diagonal covariance matrix for the residuals. Gibbons, Ross, and Shanken (1989, p. 1150) also recognize that \( \hat{\alpha}' \Sigma^{-1} \hat{\alpha} = (SR^{mv})^2 - (SR^\beta)^2 \), but do not interpret the left-hand side as the square of the Sharpe ratio of a portfolio, in particular, the alpha portfolio.
4.2 Mitigating misspecification in beta component of returns

To treat misspecification arising from the beta component of returns, we study the mean-variance portfolio weights for the case where the number of assets is asymptotically large; Figure 2 shows how the mean-variance portfolio behaves as one increases the number of assets.

**Theorem 4.3** (Weights of alpha, beta, and mean-variance portfolios for large $N$). Suppose that the vector of asset returns, $r_i$, satisfies Assumptions 3.1 and 3.2 and $\alpha_N \neq 0$. Suppose also that $A_N$ and $B_N$ are $C_N$-regular with the same scaling factor $f(N)$ and $A_N$ and $B_N$ are not asymptotically collinear. As $N \to \infty$:

(i) Then $0 \leq \delta^\alpha \leq 1$, $0 \leq \delta^\beta \leq 1$, and

\[
\frac{w^\beta_i}{w^\alpha_i} \to 0. \tag{24}
\]

(ii) The sum of the squared components of the mean-variance portfolio vectors $w^\alpha_N'w^\alpha_N$ is always bounded, whereas $w^\beta_N'w^\beta_N$ converges to zero.

(iii) The sum of the components of the mean-variance portfolio vectors $|e_i'w^\alpha_N|$ can diverge to infinity, whereas $|e_i'w^\beta_N|$ is always bounded.

(iv) The vector of weights for the mean-variance portfolio are asymptotically equivalent element-by-element to the weights of $\delta^\alpha_N w^\alpha_N$:

\[
w_{\text{mv}}^N = (1 - \delta^\beta_N)w^\alpha_N + \delta^\beta_N w^\beta_N \sim (1 - \delta^\beta_N)w^\alpha_N = \delta^\alpha_N w^\alpha_N. \tag{25}
\]

**Remark 4.3.1.** For the bounded-residual-variation case, the result in (24) follows from the fact that, under Assumptions 3.1 and 3.2, the absolute value of the components of the mean-variance portfolio vectors decrease at most at the rate:

\[
|w^\alpha_i| = O \left( |e_i'\Sigma^{-1}_N\alpha_N| + \frac{\|e_i'\Sigma^{-1}_N B_N\|}{f^\frac{1}{2}(N)} \right); \tag{26}
\]

\[
|w^\beta_i| = O \left( \frac{\|e_i'\Sigma^{-1}_N B_N\|}{f(N)} \right), \tag{27}
\]

where $e_i$ is an $N$-dimensional vector in which the $i$th element is one and the rest of the elements are zero. From equations (26) and (27), we see that $w^\alpha_i$ can dominate $w^\beta_i$ as the number of assets increases. In particular, $w^\alpha_i$ dominates $w^\beta_i$ when the pricing-error term,
$|\mathbf{e}'(\Sigma_N^{-1}\alpha_N)|$, goes to zero slowly as the number of assets increases. Recall that the APT bounds the pricing error from above; that is, $\alpha_N'\Sigma_N^{-1}\alpha_N \leq \delta < \infty$. But, the APT is silent about whether $\alpha_N'\Sigma_N^{-1}\alpha_N$ is bounded away from zero. When this expression is bounded away from zero, then one can show that the ratio $w^\beta_i/w^\alpha_i$ always decreases at a rate that is equal or faster than $1/f_{1/2}(N)$.

**Remark 4.3.2.** For the unbounded-residual-variation case (that is, where $\alpha_N = A_Nm + a_N$ and $\Sigma_N = A_NA_N' + C_N$), the result in (24) follows from the fact that, under Assumptions 3.1 and 3.2, the absolute value of the components of the mean-variance portfolio vectors decrease at most at the rate:

$$|w^\alpha_i| = O \left( |\mathbf{e}'C_N^{-1}a_N| + \|a_N\| \frac{||e'C_N^{-1}A_N|| + ||e'C_N^{-1}B_N||}{f_{1/2}(N)} + ||m|| \frac{||e'C_N^{-1}A_N||}{f(N)} \right); \quad (28)$$

$$|w^\beta_i| = O \left( \frac{||e'C_N^{-1}A_N|| + ||e'C_N^{-1}B_N||}{f(N)} \right). \quad (29)$$

Recall that $m$ can be interpreted as the risk-premia on the unobserved factors with loadings $A_N$ whereas the vector $a_N$ represents the pure pricing error that is not associated with a factor structure. The $a_N$ component dominates the behavior of the portfolio weights $w^\alpha_i$ in (28), whereas the risk premia $m$ component declines to zero faster. In general, the portfolio weight $w^\beta_i$ in (29) declines at the same, fast, rate as the risk-premia $m$ component of $w^\alpha_i$. When $a_N$ is non-zero then, as before, the $w^\alpha_N$ portfolio dominates the $w^\beta_N$ portfolio across all three norms considered in the theorem above.

**Remark 4.3.3.** Note that the $w^\alpha_N$ portfolio dominates the $w^\beta_N$ portfolio across all three norms considered in the theorem above. The notion of diversification used in part (ii) of the theorem is the sum of the squares, which is the same notion adopted in Chamberlain (1983). Because $w^\beta_N'w^\beta_N \geq \sup_i |w^\beta_i|$, it follows that the $w^\beta_N$ portfolio is diversified according to the sup norm criterion, which is the norm used in Green and Hollifield (1992). In contrast, the $w^\alpha_N$ portfolio is not necessarily diversified according to the squared norm; that is, this portfolio may not be fully diversified.\(^{25}\) Part (iii) of the theorem studies how the total investment in risky assets is allocated between the $w^\alpha_N$ and $w^\beta_N$ portfolios. The result in the theorem shows that $\mathbf{e}'w^\alpha$ could be greater than 1 and it could be growing without bound as $N$ increases, implying that it may be optimal to lever up unboundedly the investment.

\(^{25}\)For example, there could be a finite number of assets with a sufficiently large alpha, in which case the weights of these assets will not go to zero. Alternatively, even if none of the assets have a particularly large alpha, the weights of the $w^\alpha$ portfolio can go to zero at a sufficiently slow rate, as slow as $1/\sqrt{N}$, as shown in the corollary below.
in the \( w_N^\alpha \) portfolio.\(^{26}\) On the other hand, the investment in the \( w_N^\beta \) portfolio is bounded, and hence, is associated with a finite amount of leverage always.

**Remark 4.3.4.** Equation (25) shows that the \( w^\alpha \) dominates \( w^\beta \) for large \( N \). Observe that \( w^\alpha \) is functionally independent of the factor risk premia, \( \lambda \), and the factor covariance matrix, \( \Omega \), making it robust to model misspecification by construction. In contrast, portfolio \( w^\beta \) depends on both \( \lambda \) and \( \Omega \). Observe that \( \delta_N^\alpha \) is the ratio of \( \hat{\alpha} \Sigma^{-1} \hat{\alpha} \) to the variance of the return on the market portfolio, and hence, can be estimated from the data without using the beta portfolio.

In the corollary below, we look at a special case to get a better understanding of the result in (24); we consider only this part of the theorem, because the other parts of the theorem are unchanged under the special case.

**Corollary 4.3.1**  (Weights of alpha and beta portfolios for large \( N \) with \( f(N) = N \)). Suppose that the assumptions of Theorem 4.3 are satisfied and that the rows of \( A_N, B_N \) and \( C_N^{-1} \) are uniformly bounded. Then, as \( N \to \infty \), for the case of bounded-residual variation the absolute value of the components of the mean-variance portfolio vectors, \( w_N^\alpha \) and \( w_N^\beta \), decrease at most at the rate:

\[
|w_i^\alpha| = O \left( \left| \mathbf{e}' \Sigma_N^{-1} \alpha_N \right| + \frac{1}{N^2} \right) \quad \text{and} \quad |w_i^\beta| = O \left( \frac{1}{N} \right),
\]

and for the case of unbounded-residual variation they decrease at most at the rate:

\[
|w_i^\alpha| = O \left( \left| \mathbf{e}' C_N^{-1} \mathbf{a}_N \right| + \frac{\| \mathbf{a}_N \|}{N^{1.2}} + \frac{\| \mathbf{m} \|}{N} \right) \quad \text{and} \quad |w_i^\beta| = O \left( \frac{1}{N} \right).
\]

**Remark 4.3.5.** Examining the expression above, we see that \( w_i^\alpha \) will always go to zero at a slower rate than \( w_i^\beta \). The expression for \( |w_i^\beta| \) represents the rate at which the mean-variance portfolio weights decay to zero as we increase the number of assets under the traditional setting of exact pricing where the pricing error \( \alpha_N \) is zero. The expression for \( |w_i^\alpha| \) shows how the pricing error impacts portfolio weights for large \( N \). The asymptotic no-arbitrage condition of the APT in (7) allows us to say only that \( \mathbf{e}' \Sigma_N^{-1} \alpha_N \) decays to zero, without restricting the rate. On the other hand, mathematically speaking, \( \mathbf{e}' \Sigma_N^{-1} \alpha_N \) cannot decay slower than \( N^{-(1/2-\epsilon)} \), for any small \( \epsilon > 0 \).

To understand the intuition for the results about the dominance of the \( w_N^\alpha \) portfolio weights, recall that the objective of mean-variance portfolio optimization is to maximize

\(^{26}\)To see that \( \mathbf{e}' \mathbf{w}^\alpha \) can diverge, consider the following example in which the \( \alpha_N = \mathbf{e}_N / \sqrt{N} \), \( \Sigma_N = \sigma^2 I_N \), and there is a single factor with \( \beta_i \) with iid distribution with mean 1 and variance \( \sigma_\beta^2 > 0 \). Then, \( \mathbf{e}' \mathbf{w}^\alpha \sim \sqrt{N} \sigma_\beta^2 / (1 + \sigma_\beta^2) \) which goes to infinity with \( N \).
the Sharpe ratio, which entails increasing the mean of the portfolio return and/or reducing the variance of the portfolio return. Let us start by looking at the $w_N^\alpha$ portfolio. In order to minimize the portfolio variance, this portfolio will reduce the variance arising from factor exposure by being orthogonal to $B_N$. But, this is accomplished irrespective of the rate at which the weights decrease because of orthogonality of the $w_N^\alpha$ portfolio weights to $B_N$.

Because the weights can decrease at any rate, ideally one would like the weights to decrease at a rate slower than $1/N$ in order to maximize the tradeoff between the mean and variance of the portfolio. But, the APT no-arbitrage condition does not allow the rate at which the weights decrease to be slower than $N^{-\frac{1}{2}}$. However, there is no guarantee that idiosyncratic exposure is diversified away completely when the weights reduce at this rate; for this to happen, the rate of change in portfolio weights would have to be slightly faster than $N^{-\frac{1}{2}}$.

Let us now look at the $w_N^\beta$ portfolio. In order to minimize portfolio variance, this portfolio needs to reduce the variance arising from idiosyncratic exposure, $\Sigma_N$. Because this exposure is idiosyncratic, to diversify this exposure requires that the portfolio weights decrease at a sufficiently fast rate, faster than $N^{-\frac{1}{2}}$. On the other hand, if the weights decrease at any rate slower than $1/N$, then the systematic exposure explodes, both in the mean and the variance of the portfolio. If the weights decrease faster than $1/N$, then the portfolio mean declines to the risk-free rate. So, the rate of $1/N$ strikes the correct balance between optimizing the risk and return of the portfolio.\footnote{We can see the above argument also in the expression of the portfolio weight: $w_N^\beta = V_N^{-1}(B_N\hat{\lambda}_N) = (B_N\Sigma_N^\prime B_N + \Sigma_N)^{-1}B_N\Sigma_N^\prime \hat{\lambda}_N$. Because $B_N$ appears twice in denominator of $w_N^\beta$, it causes its faster decay to zero. On the other hand, $w_N^\alpha = \Sigma_N^{-1}\hat{\alpha}_N$. However, only the $A_N^m$ part of $\alpha_N$ can appear in the denominator of $w_N^\alpha$, implying that $w_N^\alpha$ decays to zero slowly whenever $a_N$ is not zero.}

We now study the properties of the returns of the mean-variance portfolio for the case in which the number of assets is large.

**Theorem 4.4** (Weight and Sharpe ratio of mean-variance portfolio for large $N$). Suppose that the vector of asset returns, $r_t$, satisfies Assumptions 3.1 and 3.2 and $\alpha_N \neq 0$. Suppose further that the investor holds a benchmark portfolio $w_N^{bench}$ with target mean $\mu^*$ satisfying the following properties:

$$
(w_N^{bench})'\alpha_N \to 0, \quad B_N'w_N^{bench} \to c^{bench}, \quad (w_N^{bench})'\Sigma_Nw_N^{bench} \to 0, \quad (30)
$$

where $c^{bench}$ is a $K \times 1$ vector of constants, different from the zero vector, satisfying $\lambda'c^{bench} \neq 0$.

(i) If $c^{bench}$ is perfectly proportional to the vector $\Omega^{-1}\lambda$, then:

$$
(w_{mv}^N)'(1 - \delta_{N}^{bench})w_N^\alpha + \delta_{N}^{bench}w_N^{bench} \sim (1 - \delta_{N}^{bench})w_N^\alpha,
$$

$$
(SR_{mv}^N)^2 \sim (SR_N^\alpha)^2 + (SR_N^{bench})^2.
$$
(ii) If $K = 1$, then $c^\text{bench}$ is always perfectly proportional to the vector $\Omega^{-1}\lambda$.

(iii) If $c^\text{bench}$ is not proportional to the vector $\Omega^{-1}\lambda$, then $w^\text{mvN}$ is not asymptotically equivalent to $(1 - \delta^\text{bench}_N)w^\alpha_N + \delta^\text{bench}_N w^\text{bench}_N$ and

$$(SR^\text{mvN}_N)^2 > (SR^\alpha_N)^2 + (SR^\text{bench}_N)^2.$$ 

Remark 4.4.1. The first assumption in (30) implies that the benchmark portfolio is asymptotically orthogonal to the alphas. The second assumption rules out that the benchmark portfolio return is equal to the risk-free return in the limit. The third assumption requires that the benchmark portfolio be well diversified. Note that for the unbounded-variation case, the first assumption is satisfied whenever $(w^\text{bench}_N)'a_N \to 0$ and $(w^\text{bench}_N)'A_N \to 0$, where the latter condition ensures that $w^\text{bench}_N$ diversifies away the contribution of the latent factors, $A_N$, to $\Sigma_N$.

Remark 4.4.2. To mitigate the effects of model misspecification in the $w^\beta$ portfolio within a large $N$ environment one can rely on the insights of Treynor and Black (1973) and DeMiguel, Garlappi, and Uppal (2009). The results of Treynor and Black (1973) can be interpreted as saying that $w^\beta$ can be approximated by a portfolio that is similar to the market portfolio, $w^\text{mkt}$. Alternatively, the empirical findings of DeMiguel, Garlappi, and Uppal (2009) suggest that, in the absence of alpha, one hold an equally weighted portfolio. In the second part of the theorem above, we formalize these two proposals by reporting the results for any arbitrary benchmark portfolio. We show the condition under which a benchmark portfolio, combined with the alpha portfolio, will coincide asymptotically with the optimal mean-variance portfolio. This condition is always satisfied with $K = 1$.

Remark 4.4.3. Observe that the two portfolios, $w^\alpha_N$ and $w^\beta_N$, generate portfolio returns with bounded Sharpe ratios. These Sharpe ratios are, in general, of the same order of magnitude even for large $N$. This is in striking contrast to the portfolio weights themselves, where the components of $w^\alpha_N$ asymptotically dominate those of $w^\beta_N$, in general. For instance, focusing on the case of bounded residual variation, the $w^\beta_N$ portfolio decays to zero at a fast rate equal to $1/N$ (see Corollary 4.3.1), whereas the $w^\alpha_N$ portfolio decays to zero at a rate equal to $1/N^{1/2}$ at most. The reason for this contrast is that, for finite $N$, the portfolio return associated with $w^\alpha_N$ is

$$r^\alpha_t - r_f = (r_t - r_f)e^\alpha \sim (\hat{\alpha}'_N + \varepsilon'_t)w^\alpha_N = (\alpha'_N + \varepsilon'_t)w^\alpha_N,$$

and whose mean and variance are given in equations (20) and (21), which are bounded by the no-arbitrage condition asymptotically. Note that the $B_N$ is absent in the expression for the portfolio return $r^\alpha_t$ because of orthogonality of the $\alpha_N$ and $B_N$; if they were not
orthogonal, then their product would have generated an explosive term in the expression for the return on portfolio $w^\alpha_N$, in which case $r^\alpha_t$ would dominate the return on the mean-variance portfolio. Note that the contribution of the idiosyncratic component to the return on the $w^\alpha_N$ portfolio, $\varepsilon'_t w^\alpha_N$, will not necessarily be diversified away by arbitrage. The reason for this is that the mean of the excess return on the $w^\alpha_N$ portfolio, which from Theorem 4.2 is $\alpha'_N w^\alpha_N = \alpha'_N \tilde{V}_N \alpha_N$, cannot be zero unless the variance is also zero because the variance of $r^\alpha_t w^\alpha_N$ is also $\alpha'_N \tilde{V}_N \alpha_N$.

In contrast, the return on the portfolio $w^\beta_N$ is $r^\beta_t - r_f = (\hat{\lambda}'_N + z'_t) B'_N w^\beta_N + \varepsilon'_t w^\beta_N$, whose mean and variance are given in equations (20) and (21), which are bounded by no arbitrage asymptotically. Again by orthogonality, $\hat{\alpha}_N$ is absent in the last expression. Moreover, the contribution of $\varepsilon_t$ in the expression for $r^\beta_t$ will vanish for well-spread (diversified) portfolios, in contrast to its contribution to $r^\alpha_t$, where it does not vanish in general. Observe that, because $z_t$ is random, the return $r^\beta_t$ is random even if the variance of the idiosyncratic component of the return, $\varepsilon'_t w^\beta_N$, is diversified away, and hence, the mean of the excess return, $r^\beta_t - r_f$, is not constrained to go to zero with diversification. For well-spread portfolios, under the conditions of Theorem 3.2 the variance of the term $\varepsilon'_t w^\beta_N$ is of order $\mathcal{O}(g^{-1}_{KK}(B'_N \Sigma^{-1}_N B_N))$, which decays to zero if $B_N$ is $\Sigma_N$-regular.

4.3 Mitigating misspecification in the alpha component of returns

The class of misspecified models considered in our theoretical analysis are framed by the APT, and in particular, by the constraint imposed on $\alpha$ by no-arbitrage, which we reproduce below:

$$\alpha' \Sigma^{-1} \alpha \leq \delta < \infty. \tag{31}$$

At the same time, the vector of pricing errors $\alpha$ plays a dominant role in the choice of optimal portfolio weights and their returns. This suggests that it is vital to estimate $\alpha$ precisely. The APT constraint provides exactly the condition that must be imposed in the estimation of the factor model generating returns.\(^{29}\)

The way in which we impose the APT constraint in (31) depends on whether we are in the case of bounded or unbounded residual variation. We propose a multi-step procedure to

---

\(^{28}\)Observe that if we had expressed the return $r^\beta_t$ in terms of $\alpha_N$ instead of $\hat{\alpha}_N$, then $\alpha_N$ would be eliminated by no arbitrage only asymptotically.

\(^{29}\)In principle, the constraint in (31) binds when $N \to \infty$ because the theory does not specify a particular value for $\delta$ (though, from the result in (23) we know that $\delta$ is less than the square of the Sharpe ratio of the market portfolio, if the market is an efficient portfolio). Following the common practice of using asymptotic standard errors even though the sample size is finite, we impose the APT restriction when estimating the model, which of course is always for a finite number of assets.
determine in which case we are. In the first step, one estimates the parameters of the factor model conditional on the factor realizations without imposing the APT restriction. Some of the observed factors could be mismeasured, requiring one to use an estimation procedure that accounts for the possibility of errors in variables. Having obtained consistent estimates of the parameters, the next step is to analyze the possibility of pervasive missing factors by studying the eigenvalues associated with the estimated $\Sigma$. This part uses conventional principal-component analysis of $\Sigma$ and it allows one to determine the number of it latent pervasive factors, $p$; see, for example, Anderson (1984). In the next two subsections, we provide the details of how to ultimately estimate the model when $p = 0$ (bounded-variation case) and when $p > 0$ (unbounded-variation case), after imposing the APT constraint.

### 4.3.1 Estimation for the case of bounded residual variation ($p = 0$)

In the bounded-residual-variation case, the true unconditional means and covariances of returns satisfy

$$E(r_t - r_f) = \mu_0 - r_f = \alpha_0 + B_0\lambda_0, \quad \text{var}(r_t) = \mathbf{V}_0 = B_0\Omega_0B_0' + \Sigma_0,$$

where the subscript “0” indicates the true value of a parameter, $\lambda_0 = E(f_t) - r_f$ is the vector of risk premia and $\Omega_0 = \text{var}(f_t)$, assuming stationarity of the $K$ factors, $f_t$, which are assumed to be traded without loss of generality. In order to identify $\lambda_0$ and $\Omega_0$, one needs to consider also the information stemming from the sample observations for $f_t$. Although our argument applies to virtually any estimation procedure, we will illustrate it with respect to the (pseudo) Gaussian maximum-likelihood (ML) estimator. This is a very natural estimator for our model when the first two moments of asset returns are specified correctly, although distributional assumptions (such as normality) are not required.\(^{30}\)

The (pseudo) ML estimator, based on the unconditional joint distribution of $(r_t - r_f e_{N+K})$ and assuming i.i.d. residuals for simplicity, is:\(^{31}\)

$$L(\theta) = -\frac{1}{2} \log(\det(\Sigma)) - \frac{1}{2T} \sum_{t=1}^{T} \left( r_t - r_f e - \alpha - B(f_t - r_f e_K) \right)' \Sigma^{-1} \left( r_t - r_f e - \alpha - B(f_t - r_f e_K) \right)$$

\(^{30}\)In turn, this requires one to have identified whether one faces the case of bounded or unbounded residual variation and, in the latter case, the exact number of missing factors.

\(^{31}\)Notice that we have expressed the joint distribution as the product of a conditional distribution and a marginal distribution. Relaxing the i.i.d. assumption requires specification of either time-varying conditional means or conditional variances or both.
\[-\frac{1}{2} \log(\det(\Omega)) - \frac{1}{2T} \sum_{t=1}^{T} \left( \hat{f}_t - r^eK - \lambda \right)' \Omega^{-1} \left( \hat{f}_t - r^eK - \lambda \right),\]

where \( \theta = (\alpha', \text{vec}(B)'', \text{vech}(\Sigma)'', \lambda', \text{vech}(\Omega)'') \). Therefore, the ML estimator for \( \alpha_0, B_0, \Sigma_0 \) coincide with the OLS estimator, conditional on the realization of the factors. On the other hand, the ML estimators for \( \lambda_0 \) and \( \Omega_0 \) are the sample mean and sample covariance of \( \hat{f}_t \).

However, because the APT restriction is not guaranteed to hold, one should consider the maximum-likelihood constrained (MLC) estimator:

\[
\hat{\theta}_{MLC} = \arg\max_{\theta} L(\theta) \text{ such that } \alpha' \Sigma^{-1} \alpha \leq \delta. \tag{32}
\]

Because the parameter \( \alpha_0 \) is constrained only by the APT restriction, imposing this constraint may lead \( \hat{\theta}_{MLC} \) to be a more precise estimator of the true parameter values compared to the unconstrained estimator, \( \hat{\theta}_{ML} \). The theory does not tell us what \( \delta \) should be in (32). As discussed earlier, we know that the upper bound of \( \delta \) must be less than the square of the Sharpe ratio of the mean-variance efficient portfolio; in our empirical application, we will choose \( \delta \) using cross-validation techniques.

To impose the APT restriction one can consider a penalized log-likelihood function as follows.

**Theorem 4.5** (Parameter identification by imposing asset-pricing restriction: Bounded-variation case). Given any \( \kappa > 0 \),

\[
\hat{\theta}_{MLC} = \arg\max_{\theta} L(\theta) - \kappa(\alpha' \Sigma^{-1} \alpha - \delta).
\]

If \( \left( \sum_{t=1}^{T} \hat{f}_t \hat{f}_t' \right) \) is nonsingular, then \( \theta_{MLC} = (\tilde{\alpha}'_{MLC}, \text{vec}(\tilde{\mathbf{B}}_{MLC})', \text{vech}(\tilde{\Sigma}_{MLC})', \tilde{\lambda}'_{MLC}, \text{vech}(\tilde{\Omega}_{MLC})')' \) exists, where:

\[
\tilde{\alpha}_{MLC} = \frac{1}{1 + \kappa} \left[ \bar{\mathbf{r}} - r^e \mathbf{e} - \tilde{\mathbf{B}}_{MLC} \hat{\mathbf{f}} \right],
\]

\[
\tilde{\mathbf{B}}_{MLC} = \left( \sum_{t=1}^{T} \hat{f}_t \hat{f}_t' \right)^{-1} \left( \sum_{t=1}^{T} \hat{f}_t \hat{f}_t' \right),
\]

\[
\tilde{\Sigma}_{MLC} = \frac{1}{T} \sum_{t=1}^{T} (\hat{\mathbf{r}}_t - \tilde{\mathbf{B}}_{MLC} \hat{\mathbf{f}}_t)(\hat{\mathbf{r}}_t - \tilde{\mathbf{B}}_{MLC} \hat{\mathbf{f}}_t)',
\]

where \( \hat{\mathbf{f}}_t = \mathbf{f}_t - r^e \mathbf{e}_K - \frac{1}{1 + \kappa}(\bar{\mathbf{r}} - r^e \mathbf{e}_K), \hat{\mathbf{r}}_t = \mathbf{r}_t - r^e \mathbf{e} - \frac{1}{1 + \kappa}(\bar{\mathbf{r}} - r^e \mathbf{e}) \), and the MLC estimators \( \tilde{\lambda}'_{MLC} \) and \( \text{vech}(\tilde{\Omega}_{MLC})' \) coincide with the sample mean and covariance of the factors \( \mathbf{f}_t \).

---

Note that \( \det(\cdot) \) denotes the determinant, \( \text{vec}(\cdot) \) denotes the operator that stacks the columns of a matrix into a single column vector, and \( \text{vech}(\cdot) \) denotes the operator that stacks the unique elements of the columns of a symmetric matrix into a single column vector.

---

32 Note that \( \det(\cdot) \) denotes the determinant, \( \text{vec}(\cdot) \) denotes the operator that stacks the columns of a matrix into a single column vector, and \( \text{vech}(\cdot) \) denotes the operator that stacks the unique elements of the columns of a symmetric matrix into a single column vector.
Remark 4.5.1. The constrained estimator $\hat{\alpha}_{MLC}$ turns out to be precisely the ridge estimator for $\alpha_0$, because $\kappa > 0$. The estimators of $\hat{B}_{MLC}$ and $\hat{\Sigma}_{MLC}$ are also functions of $\kappa$ because of the APT constraint, in contrast to $\hat{\lambda}_{MLC}$ and $\hat{\Omega}_{MLC}$, which are simply the sample mean and sample covariance of the $f_t$ because the APT constraint does not affect the distribution of the factors $f_t$.

Remark 4.5.2. Alternatively, one can solve for the Karush-Kuhn-Tucker multiplier $\hat{\kappa}_{MLC}$, which is given in the appendix. An iterative procedure is required to solve for $\hat{\kappa}_{MLC}$ and $\hat{B}_{MLC}$ jointly; then the estimators $\hat{\alpha}_{MLC}$ and $\hat{\Sigma}_{MLC}$ follow. We follow an alternative estimation approach consists of choosing $\kappa$ by cross-validation. The formulae for the estimators of the other parameters remain the same. The cross-validation approach is closer to a LASSO-type formulation where one considers the penalized log-likelihood $L(\theta) - \kappa(\alpha'\Sigma^{-1}\alpha)$, which turns out to be computationally simpler than the Karush-Kuhn-Tucker estimator.

4.3.2 Estimation for the case of unbounded residual variation ($p > 0$)

Suppose that $a$, the pricing error unrelated to the $p$ missing factors, is zero in (8). Then, $\alpha = Am$ and $\Sigma = AA' + C$, where $m$ is the risk premia corresponding to the missing factors, and $C$ is an $N \times N$ positive-definite matrix with bounded eigenvalues that represents the covariance matrix of the pure idiosyncratic component of the error returns. Observe that $\alpha$ is a component of the expected return, $\mu$; likewise, $\Sigma$ is a component of the return-covariance matrix, $V$. Hence, $A$ appears in both the mean and covariance matrix of returns. MacKinlay and Pástor (2000) use this insight to improve the precision of the estimated $A$ parameters when estimating the model for asset returns by maximum likelihood (ML), which, in turn, improves substantially the estimated portfolios’ performance.\footnote{Note that because we are interpreting the missing factors as unobserved, without loss of generality one can assume that $AA'$ represents the contribution of the missing factors to the residual variance $\Sigma$ because the missing factors are uncorrelated and have unit variance, leaving the risk-premia $m$ as free parameters to be estimated. MacKinlay and Pástor (2000) consider a different identification assumption. For $p = 1$ they estimate $\alpha$ without distinguishing between $A$ and $m$, implying that the contribution of the single missing factor to the return variance equals $\alpha\alpha'/(SR^h)^2$, where $SR^h$ is the Sharpe ratio of the missing factor.}

Importantly, using the Sherman-Morrison-Woodbury formula, it follows that:

$$\alpha'\Sigma^{-1} \alpha = m' A \Sigma^{-1} A m = m' (I_r + A' C^{-1} A)^{-1} (A' C^{-1} A) m.$$ 

Therefore, $\alpha'\Sigma^{-1} \alpha$ converges to $m'm$ as $N \to \infty$ because $(I_r + A' C^{-1} A)^{-1} (A' C^{-1} A)$ converges to the identify matrix given that the missing factors are pervasive implying that $(A' C^{-1} A)$ is increasing without bound. This means that the APT restriction is always satisfied for the case of only missing factors (that is, the case where $a = 0$), once we recognize that $\Sigma$ contains the loadings of the missing factors, $A$. 

33
However, for the general unbounded-variation case where the pricing error consists of both missing factors and a component that is unrelated to factors, $\alpha = A m + a$, the APT restriction is not automatically satisfied. Therefore, when estimating the model we need to impose the additional constraint: $a' a \leq \delta < \infty$ for any $N$. Under the same assumptions as above concerning the $K$ observed factors $f_t$, the true unconditional means and covariances of returns now satisfy the equations below, where $C_0$ has bounded maximum eigenvalue, in contrast to $\Sigma_0$:

$$E(r_t - r_f e) = \mu_0 - r_f e = A_0 m_0 + a_0 + B_0 \lambda_0,$$

$$\text{var}(r_t) = V_0 = B_0 \Omega_0 B_0' + A_0 A_0' + C_0.$$

As in the previous case of bounded variation, the joint log-likelihood function $L(\theta)$ can be decomposed as follows:

$$L(\theta) = -\frac{1}{2} \log(\det(\Sigma + C)) - \frac{1}{2T} \sum_{t=1}^{T} \left( r_t - r_f e - A m - a - B(f_t - r_f K) \right)' \left( AA' + C \right)^{-1} \left( r_t - r_f e - A m - a - B(f_t - r_f K) \right)$$

$$- \frac{1}{2} \log(\det(\Omega)) - \frac{1}{2T} \sum_{t=1}^{T} \left( f_t - r_f K - \lambda \right)' \Omega^{-1} \left( f_t - r_f K - \lambda \right).$$

Without loss of generality, one can assume that the missing factors are uncorrelated with each other and have unit variance, achieving identification of $A_0$.\(^{34}\) However, $m_0$ and $\alpha_0$ cannot be identified separately unless the APT restriction is imposed, as shown below.

**Theorem 4.6** (Parameter identification by imposing asset-pricing restriction: Unbounded-variation case). *Suppose that the vector of asset returns, $r_t$, satisfies Assumption 3.1. Given any $\kappa \geq 0$, then:*

$$\hat{\theta}_{MLC} = \arg \max_{\theta} L(\theta) - \kappa(a' \Sigma^{-1} a - \delta),$$

where

$$\hat{\theta}_{MLC} = (\hat{a}'_{MLC}, \hat{m}_{MLC}, \text{vec}(\hat{A}_{MLC})', \text{vec}(\hat{B}_{MLC})', \text{vech}(\hat{C}_{MLC})', \hat{\lambda}'_{MLC}, \text{vech}(\hat{\Omega}_{MLC})').$$

(i) *For $\kappa > 0$, if the APT restriction holds exactly, that is, $a_0' \Sigma_0^{-1} a_0 = \delta$, and $\Sigma_{ff} - \tilde{f} \tilde{f}'$ is nonsingular, then*

$$\text{vec}(\hat{B}_{MLC}) = \left( (\Sigma_{ff} \otimes I) + (\tilde{f} \tilde{f}' \otimes (GG - 2G)) \right)^{-1} \text{vec} \left( \Sigma_{rf} - (2G - GG)\tilde{r} \tilde{f}' \right),$$

\(^{34}\)We are implicitly assuming that observed and missing factors are mutually uncorrelated. Extension to the correlated case does not change the main implications of our method.
\[ \hat{m}_{\text{MLC}} = (\hat{A}'_{\text{MLC}} \Sigma^{-1}_{\text{MLC}} \hat{A}_{\text{MLC}})^{-1} \hat{A}'_{\text{MLC}} \Sigma^{-1}_{\text{MLC}} \left( \bar{r} - r_f e - \hat{B}_{\text{MLC}} (\bar{f} - r_f e_K) \right), \]

\[ \hat{a}_{\text{MLC}} = \frac{1}{\kappa + 1} (\bar{r} - r_f e - \hat{B}_{\text{MLC}} (\bar{f} - r_f e_K) - \hat{A}_{\text{MLC}} \hat{m}_{\text{MLC}}), \]

in which \( \Sigma_{\text{MLC}} = \hat{A}_{\text{MLC}} \hat{A}'_{\text{MLC}} + \hat{C}_{\text{MLC}} \), \( \Sigma_{rf} = \frac{1}{T} \sum_{t=1}^{T} \tilde{r}_t \tilde{f}_t \), \( \Sigma_{ff} = \frac{1}{T} \sum_{t=1}^{T} \tilde{f}_t \tilde{f}'_t \), \( \tilde{r}_t = (r_t - r_f e) \), \( \hat{r} = (\bar{r} - r_f e) \), \( \hat{f}_t = (f_t - r_f e_K) \), \( \bar{f} = (\bar{f} - r_f e_K) \), and

\[ G = \frac{1}{(\kappa + 1)} I + \frac{\kappa}{(\kappa + 1)} \hat{A}_{\text{MLC}} (\hat{A}'_{\text{MLC}} \Sigma^{-1}_{\text{MLC}} \hat{A}_{\text{MLC}})^{-1} \hat{A}'_{\text{MLC}} \Sigma^{-1}_{\text{MLC}}. \]

Note that \( \hat{A}_{\text{MLC}} \) and \( \hat{C}_{\text{MLC}} \) do not admit a closed-form solution, and as before, \( \hat{\lambda}_{\text{MLC}} \) and \( \hat{\Omega}_{\text{MLC}} \) coincide with the sample mean and sample covariance of the factors \( f_t \).

(ii) For \( \kappa = 0 \) and the APT restriction not holding exactly, one can identify only \( \alpha_0 = A_0 m_0 + a_0 \) (but not the components separately) and one obtains

\[ \hat{\alpha}_{\text{MLC}} = (\bar{r} - r_f e) - \hat{B}_{\text{MLC}} (\bar{f} - r_f e), \]

and the expression for \( \text{vec}(\hat{B}_{\text{MLC}}) \) can be obtained by setting \( \kappa = 0 \) in the expression above.

The expressions for \( \hat{\lambda}_{\text{MLC}} \) and \( \hat{\Omega}_{\text{MLC}} \) are unchanged, and, as before, the expressions for the estimators of the other parameters, \( \hat{A}_{\text{MLC}} \) and \( \hat{C}_{\text{MLC}} \) do not admit a closed-form solution.

Remark 4.6.1. The theorem above shows that the APT restriction does not afford just more precise estimation but also allows for identification of the model parameters.

5 Simulation Results

In this section, we evaluate the theoretical results described above in Section 4. This section has two parts. In the first part, we evaluate the gains from treating misspecification in the alpha component of returns, which is done by imposing the APT restriction. In the second part, we investigate the gains from treating misspecification in the beta component of returns, which is done by using the insights from our asymptotic analysis as the number of assets increases.

5.1 Model misspecification in the alpha component of returns

The design of our simulation analysis is similar to that in MacKinlay and Pástor (2000). We consider the case where the true process generating the \( N = 100 \) monthly stock returns is a two-factor model, but the investor assumes, and therefore estimates, a single factor model:

\[ r_t = \alpha + \beta f_t + \varepsilon_t. \]
Throughout the exercise, we assume that the risk-free interest rate is 0 and that the observed factor $f_t$ is IID and has Gaussian distribution. For the “base case” of our simulation exercise, we assume that the observed factor has a monthly mean equal to $\lambda = \frac{8}{12 \times 100}$ and monthly volatility equal to $\omega = \frac{16}{\sqrt{12 \times 100}}$. We generate the pricing error $\alpha$ from an IID multivariate Gaussian distribution with mean $0$ and covariance matrix equal to $\sigma_\alpha^2 I_N$, with $\sigma_\alpha = \frac{5}{\sqrt{12 \times 100}}$.

We consider two environments, one with bounded residual variation ($\alpha = a$) and the other with a special form of unbounded residual variation ($\alpha = Am$). In both cases, $\varepsilon_t$ is IID with a multivariate Gaussian distribution with monthly mean of $0$. In the bounded-variation case, the monthly covariance matrix is $\Sigma = \sigma_\varepsilon^2 I_N$ with $\sigma_\varepsilon = \frac{20}{\sqrt{12 \times 100}}$. In contrast, in the unbounded-variation case, the monthly covariance matrix is $\Sigma = \frac{1}{sr_\alpha} \alpha \alpha' + \sigma_\varepsilon^2 I_N$, with $sr_\alpha = \frac{1.20}{\sqrt{12}}$ and $\sigma_\varepsilon = \frac{20}{\sqrt{12 \times 100}}$.

In addition to the “base case” described above, we look at three variations. In the first, we look at the case where the risk premium $\lambda$ on the observed factor, which one may interpret as the market, is half of its base-case value; this allows us to study the performance of the various portfolios in a low-return environment. In the second variation, we look at the case where $\sigma_\varepsilon$ is half of its base-case value, which corresponds to the case where there is less residual risk in returns, and hence, estimation error is likely to be smaller. In the third variation, we look at the case where $\sigma_\alpha$ is a one-third of its base-case value, which corresponds to the case where the alphas have smaller dispersion. This allows us to identify the condition under which the portfolio strategy we develop in this paper is likely to perform poorly because the expected return of the alpha strategy is increasing in the dispersion of the alphas (in the bounded-residual-variation case).

For the base cases corresponding to each of the two environments described above, and the three variations of each base case, we compare using the out-of-sample Sharpe ratio the performance of the following portfolio strategies: (1) the equal-weighted (EW) portfolio where the weights are equal to $1/N$ and no estimation is required; (2) the global-minimum-variance (GMV) portfolio based on the sample covariance matrix; (3) the mean-variance (MV) portfolio based on plugging in the sample mean and sample covariance matrix; (4) the maximum-likelihood-based unconstrained (MLU) portfolio that is based on the sample mean and covariance matrix implied by the factor model, $\beta \beta' \omega^2 + \sigma^2 I$; and (5) the maximum-likelihood-based constrained (MLC) portfolio based on the sample mean but factor covariance matrix of $\beta \beta' \omega^2 + \Sigma$ and imposing the APT constraint in which $\delta$ is obtained using ten-fold cross validation.
We now explain how we compute the out-of-sample portfolio Sharpe ratios. First, using the above parameter values we simulate \( M = 100 \) Monte Carlo paths of length \( T = 300 \) months. For each path, we estimate the moments of asset returns using a rolling window of 120 monthly observations, based on which we construct the portfolio weight for each of the five strategies described above. We then compute the return on this portfolio based on the realized return in the 121st month following the estimation window. Based on the series of realized returns for the next 179 months, we construct the Sharpe ratio for this particular path of the Monte Carlo simulation. We repeat this for each of the Monte Carlo paths, and report the average Sharpe ratio across paths. We also report the \( t \)-statistic for the difference in the Sharpe ratio of the equally weighted portfolio and the portfolio based on the APT restriction where the standard errors are from the Monte Carlo paths.

Table 1 gives the annualized Sharpe ratio for the five portfolio strategies listed above. Examining Panel A of the table, which reports the results for the case where the pricing errors are unrelated to factors, we see from the first row of the table that for the base case the Sharpe ratio of the EW portfolio is 0.55 p.a. The GMV portfolio has a Sharpe ratio of only 0.20 p.a.; this lower Sharpe ratio, despite optimizing with respect to portfolio risk, is a result of the error in estimating the sample covariance matrix of returns. The MV portfolio has an even lower Sharpe ratio of 0.03 p.a. because this strategy relies on estimates of both the sample covariance matrix and the sample mean, and it is well-known that it is difficult to estimate the sample mean with precision. The MLU portfolio that relies on the sample mean but uses the covariance matrix \( \beta \beta' \omega^2 + \sigma^2 I \), performs much better: its Sharpe ratio of 1.06 p.a. is much higher than that of the EW portfolio. The MLC portfolio, which relies on the sample mean but uses the covariance matrix \( \beta \beta' \omega^2 + \Sigma \) along with the APT constraint, performs even better: its Sharpe ratio of 1.35 p.a. is about 2.5 times that of the EW portfolio and the difference between these Sharpe ratios has a \( t \)-statistic of 8.61.

Based on the results in the first row of Panel A, we deduce that the EW portfolio achieves a high Sharpe ratio relative to the GMV and MV portfolios because it does not suffer from estimation error; however, because it is a portfolio with only naive diversification, it fails to take advantage of the dispersed alphas and earns only the average alpha, which is zero; on the other hand, MLC can exploit fully the presence of pricing errors. Furthermore, the superior performance of the MLC portfolio relative to the MLU portfolio highlights the importance of the APT condition.

The second row of Panel A considers the case where the risk premium on the observed factor, \( \lambda \), is half the base-case value. This corresponds to a low-return environment. While the GMV portfolio has the same Sharpe ratio as that in the base case, the Sharpe ratios
of all the other portfolios decrease. However, the Sharpe ratio of the MLC strategy is now more than 2.5 times better than the EW strategy and about 1.5 times better than the MLU strategy.

The third row of Panel A considers the case where the residual risk is half of its base-case value. In this case, the Sharpe ratio of the EW portfolio does not change at all, but Sharpe ratios of all the other strategies that rely on estimated return moments improve. The Sharpe ratio of the MLC strategy is now about 6.5 times that of the EW portfolio and about 1.4 times that of the MLU strategy.

One might wonder when the EW strategy will outperform the MLC strategy: this happens when the alphas are small. In the last row of Panel A, we consider the case where the dispersion of alphas is only a third of the base-case value. In this case, the Sharpe ratio of the EW strategy is still the same as its base case value, but now the Sharpe ratio of MLC is significantly smaller.

Panel B of Table 1 shows that the insights described above are similar for the case where the pricing errors are related to factors. The EW portfolio continues to perform well. However, because the pricing error now shows up in both the mean and volatility of portfolio returns, the differences between the Sharpe ratio of the EW portfolio and the Sharpe ratios of the MLU and MLC strategies are lower than what they were before. In fact, the EW portfolio outperforms the MLU portfolio for all four cases reported in this panel. However, the MLC portfolio still outperforms the EW portfolio, except for the case reported in the last row of Panel B, where the dispersion of the pricing errors is a third of the base-case value.

Based on these simulations results, we conclude that the APT restriction leads to a significant improvement in Sharpe ratios, relative to the GMV, MV, and MLU strategies that do not impose this condition, and also relative to the EW portfolio that does not suffer from estimation risk. The MLC portfolio fails to outperform the EW portfolio when the mean of the alphas is zero and when the dispersion of alphas is low—in which case an optimizing portfolio that takes long and short positions relative to the market benchmark does not do much better than the EW portfolio that earns the average rate of return by being long all assets.

Note also that the covariance matrix for the MLU case is misspecified in the unbounded-residual-variation case.
5.2 Model misspecification in the beta component of returns

In the simulation analysis reported above, we considered various sources of model misspecification that are reflected in the alpha component of returns. In this section, we consider model misspecification in the beta component of returns; in particular, misspecification of the risk premia, which are given by the mean of the factors in excess of the risk-free rate, $\lambda$, and/or the covariance matrix of the factors, $\Omega$. Misspecification in these parameters can arise, for instance, if investors have particular views.\[36\]

In this analysis, we use the same data-generating model as in the base-case of the previous section. However, when we estimate the model, we impose the condition that the investor has the view that $\lambda$ is double or half of its sample mean estimator. Similarly, for $\Omega$ we consider the situation where the investor has the view that it is double or half of its sample covariance estimator. Then, we consider two versions of the MLC strategy (strategy (5)) described in the previous section. Recall, that this is the maximum-likelihood-based constrained portfolio based on the sample mean but factor covariance matrix of $\beta'\omega^2 + \Sigma$ and imposing the APT constraint in which $\delta$ is obtained using ten-fold cross validation. The first version of this strategy is based on the incorrect views of either the factor risk premia, $\lambda$, or the factor covariances, $\Omega$. The second version of the strategy is based on our asymptotic analysis in Theorem 4.4, which are immune to misspecification of $\lambda$ and $\Omega$.

Our numerical analysis leads to the following insights. First, the Sharpe ratio from using the MLC strategy based on the incorrect views regarding $\Omega$ performs substantially worse than the equal-weighted portfolio, and hence, also the MLC strategy based on the correct views. For example, the per annum Sharpe ratio of the MLC strategy based on the correct views is 0.81 (see the first row of Panel B of Table 1), the Sharpe ratio of the equal-weighted portfolio is 0.54 (first row of Panel B of Table 1), but the Sharpe ratio of the MLC strategy based on the incorrect view is much lower: in the case where the incorrect view is half (double) of the MLE estimate of $\Omega$, the Sharpe ratio is only 0.35 (0.52). For the case of misspecification in the factor risk premium, the cost is smaller: if the incorrect view is half (double) of the MLE estimate of $\lambda$, the Sharpe ratio is only 0.75 (0.69).

Second, we find that the strategy that takes into account the possibility of beta misspecification by replacing the beta portfolio that is based on the incorrect views about $\lambda$ and $\Omega$ with the benchmark portfolio, which in our case is the equal-weighted ($1/N$) portfolio, the resulting Sharpe ratio is substantially higher when the number of risky assets $N$.

\[36\]Alternatively, one could imagine a world in which the true risk premia are conditionally time varying but the investor models them as constant through time.
is sufficiently large. For example, if $N = 100$ and the investor believes that the value of $\Omega$ is half (double) its estimated value, then the per annum Sharpe ratio of the strategy based on the incorrect view is 0.52 (0.35); on the other hand, if the incorrect view is half (double) of the MLE estimate of $\lambda$, the Sharpe ratio is 0.75 (0.69). In contrast, the Sharpe ratio for the strategy that accounts for beta misspecification is 1.12; however, if $N = 5$, then as predicted by our theorem, the asymptotic results do not apply, and the strategy that accounts of beta misspecification achieves a Sharpe ratio of only 0.28.

Finally, using the insights from Theorem 4.4, and in particular the result for the special case in which $K = 1$, we find that one can achieve mean-variance efficiency by combining the $w^a$ portfolio with any benchmark portfolio satisfying the necessary conditions required for the theorem to hold that are listed in (30). That is, when $N$ is sufficiently large, the portfolio strategy that takes into account beta misspecification achieves a Sharpe ratio that is very close to that achieved by the strategy that is based on knowing the true parameter values. For example, if $N = 100$, then the Sharpe ratio of the strategy that accounts for beta misspecification is 1.12, whereas the Sharpe ratio one would achieve if one knew the true parameters is 1.26.

6 Conclusion

In this paper, we have provided a rigorous foundation and characterization, based on the APT, for alpha and beta portfolios, where the “alpha” portfolio is one that depends only on pricing errors and the “beta” portfolio depends only on factor risk premia. We then show how these properties can be exploited to mitigate the effects of model misspecification for portfolio choice.

Our first result is to explain that one can extend the interpretation of the APT so that the alpha in it represents not just small pricing errors that are independent of factors but also large pricing errors arising from mismeasured or missing factors. We also show how the APT model can capture misspecification in the beta component of returns. We then use the mathematical structure underlying the APT to study the mitigation of model misspecification for the family of mean-variance portfolios, including the mean-variance portfolio, the global minimum-variance portfolio, and the Markowitz frontier portfolios.

Our key insight is that instead of treating model misspecification directly in the mean-variance portfolios, it is better to first decompose mean-variance portfolios into two components, an “alpha” portfolio and a “beta” portfolio, and then to treat misspecification in these two components using different methods. Misspecification in the alpha component of
returns is treated by imposing the APT restriction on the weighted sum of squares of the pricing errors when estimating the return-generating model. Misspecification in the beta component of returns, on the other hand, is treated utilizing properties of the alpha and beta portfolios as the number of assets increases asymptotically. In particular, we use the property that the weights of the alpha portfolio dominate the corresponding weights in the beta portfolio asymptotically.

The results described above have important implications for both empirical and theoretical work. On the empirical side, our results allow one to improve the estimation of mean-variance portfolios in the presence of pricing errors. In particular, our theoretical results based on asymptotic analysis allow one to reduce the effect of model risk associated with misspecification of the beta component of returns. We also show how to minimize the impact of estimation risk arising from the alpha component of returns on portfolio performance by exploiting the restriction imposed by the APT. Using simulations, we demonstrate that our approach for treating model misspecification leads to mean-variance portfolio weights that have significantly higher out-of-sample Sharpe ratios compared to other benchmark strategies, such as the equal-weighted portfolio, the sample-based mean-variance portfolio, and the mean-variance portfolio that takes advantage of the factor structure of returns but does not impose the APT restriction.

On the theoretical front, the APT restriction on the weighted sum of the squared pricing errors can be interpreted as a bound on the Sharpe ratio of the alpha portfolio, which is similar to the discount factor volatility restriction in Hansen and Jagannathan (1991) and the “good-deal” bound studied in Cochrane and Saa-Requejo (2001); imposing this constraint is analogous to the approach adopted in Garlappi, Uppal, and Wang (2007), where one accounts for parameter uncertainty in portfolio choice using the minmax approach originally proposed in Gilboa and Schmeidler (1989). Our analysis can also be used to investigate dynamic mean-variance portfolios using the insights in Brandt and Santa-Clara (2006) and Basak and Chabakauri (2010).
A Proofs for Theorems

Note that Theorems 3.1 and 3.2 are derived in both Huberman (1982) and Ingersoll (1984), and so we do not include the proofs for them. The proofs for all other theorems in the main text of the manuscript are given below, starting with a preliminary lemma. For theorems stated in Appendices B.1, B.2, and B.3, the proofs follow directly after the statement of the theorem.

Lemma on Sharpe-ratio Decomposition

Lemma A.1. Consider the portfolio weights \( w = w_1 + w_2 \) such that

\[ w_1' V w_2 = 0. \]

Then, setting \( \text{SR}_i = w_i' (\mu - \text{er}_f) / (w_i' V w_i)^{1/2} \),

\[ \text{SR}^2 = \frac{(w' (\mu - \text{er}_f))^2}{w' V w} \leq (\text{SR}_1)^2 + (\text{SR}_2)^2. \]

Finally, equality holds if and only if:

\[ \frac{w_1' (\mu - \text{er}_f)}{w_1' V w_1} = \frac{w_2' (\mu - \text{er}_f)}{w_2' V w_2}. \]

Proof. Setting for simplicity \( \mu_i - r_f = w_i' (\mu - \text{er}_f), \sigma_i^2 = w_i' V w_i \), we have:

\[ \text{SR}^2 = \frac{(\mu_1 - r_f)^2}{\sigma_1^2} \frac{\sigma_1^2}{w' V w} + \frac{(\mu_2 - r_f)^2}{\sigma_2^2} \frac{\sigma_2^2}{w' V w} + 2 \frac{(\mu_1 - r_f)(\mu_2 - r_f)}{w' V w} \]

\[ = \frac{(\mu_1 - r_f)^2}{\sigma_1^2} + \frac{(\mu_2 - r_f)^2}{\sigma_2^2} \]

\[ + \left[ \frac{(\mu_1 - r_f)^2}{\sigma_1^2} \left( -1 + \frac{\sigma_1^2}{w' V w} \right) + \frac{(\mu_2 - r_f)^2}{\sigma_2^2} \left( -1 + \frac{\sigma_2^2}{w' V w} \right) + 2 \frac{(\mu_1 - r_f)(\mu_2 - r_f)}{w' V w} \right]. \]

Recalling that \( w' V w = w_1' V w_1 + w_2' V w_2 \), the term in square-brackets can be rewritten as

\[ - \frac{(\mu_1 - r_f)^2}{\sigma_1^2} \frac{\sigma_2^2}{w' V w} - \frac{(\mu_2 - r_f)^2}{\sigma_2^2} \frac{\sigma_1^2}{w' V w} + 2 \frac{(\mu_1 - r_f)(\mu_2 - r_f)}{w' V w} \]

\[ = \frac{1}{w' V w} \left( - \mu_1^2 \frac{\sigma_2^2}{\sigma_1^2} - \mu_2^2 \frac{\sigma_1^2}{\sigma_2^2} + 2 (\mu_1 - r_f) \frac{\sigma_2}{\sigma_1} (\mu_2 - r_f) \frac{\sigma_1}{\sigma_2} \right) \]

\[ = - \frac{1}{w' V w} \left( (\mu_1 - r_f) \frac{\sigma_2}{\sigma_1} - (\mu_2 - r_f) \frac{\sigma_1}{\sigma_2} \right)^2. \]
Hence,
\[
SR^2 = \left(\frac{\mu_1 - r_f}{\sigma_1^2}\right)^2 + \left(\frac{\mu_2 - r_f}{\sigma_2^2}\right)^2 - \frac{1}{w'w}((\mu_1 - r_f)\frac{\sigma_2}{\sigma_1} - (\mu_2 - r_f)\frac{\sigma_1}{\sigma_2})^2 \\
\leq \left(\frac{\mu_1 - r_f}{\sigma_1^2}\right)^2 + \left(\frac{\mu_2 - r_f}{\sigma_2^2}\right)^2 = (SR_1)^2 + (SR_2)^2.
\]

Equality holds if and only if
\[
\left(\frac{\mu_1 - r_f}{\sigma_2} - (\mu_2 = r_f)\frac{\sigma_1}{\sigma_2}\right)^2 = 0,
\]
which, in turn, can be rearranged as
\[
\frac{\mu_1 - r_f}{\sigma_1^2} = \frac{\mu_2 - r_f}{\sigma_2^2}.
\]

**Proof of Theorem 3.3**

By Chamberlain and Rothschild (1983, Theorem 4) the residual covariance matrix satisfies
\[
\Sigma_N = A_NA_N' + C_N
\]
where \(C_N\) is a positive definite matrix with eigenvalues uniformly bounded by \(g_{p+1N}(\Sigma_N)\).

By the Sherman decomposition
\[
\Sigma^{-1} = C_N^{-1} - C_N^{-1}A_N(I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1}.
\]

Therefore, by substitution,
\[
\alpha_N' \Sigma^{-1} \alpha_N = \alpha_N' C_N^{-1} \alpha_N - \alpha_N' C_N^{-1} A_N(I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} \alpha_N
\]
\[
= (A_N' m + a_N)' C_N^{-1} (A_N' m + a_N)
\]
\[
- (A_N' m + a_N)' C_N^{-1} A_N(I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} (A_N' m + a_N)
\]
\[
= m' A_N' C_N^{-1} A_N m - m' A_N' C_N^{-1} A_N(I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} A_N m
\]
\[
+ a_N' C_N^{-1} A_N - a_N' C_N^{-1} A_N(I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} a_N
\]
\[
+ 2a_N' C_N^{-1} A_N m - a_N' C_N^{-1} A_N(I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} A_N m.
\]

We now show that \(\alpha_N' \Sigma^{-1} \alpha_N\) is bounded even as \(N\) diverges. We look each of the term on the right hand side of the last equality sign, one by one. Thus:
\[
m' A_N' C_N^{-1} A_N m - m' A_N' C_N^{-1} A_N(I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} A_N m
\]
\[
= m'(I_N - A_N' C_N^{-1} A_N(I_p + A_N' C_N^{-1} A_N)^{-1}) A_N' C_N^{-1} A_N m
\]
\[
= m'(I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} A_N m \leq m'm,
\]

41
since $I_p - (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} A_N$ is positive semi definite. Next, for the second term

$$a_N' C_N^{-1} a_N \leq a_N' a_N g_N^{-1}(C_N).$$

Now, the $j$th element of $a_N' C_N^{-1} A_N$, obtained by considering the $j$th column of $A_N$, for every $1 \leq j \leq p$, satisfies

$$|a_N' C_N^{-1} g_j N v_j N| \leq g_j N (a_N' C_N^{-1} a_N)^{1/2} (v_j N C_N^{-1} v_j N)^{1/2},$$

where for simplicity we set $v_j N = v_j N(\Sigma N), g_j N = g_j N(\Sigma N)$. Since

$$a_N' C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} a_N$$

$$\leq g_1 p ((I_p + A_N' C_N^{-1} A_N)^{-1} (a_N' C_N^{-1} a_N)(A_N' C_N^{-1} a_N)$$

$$= g_1 p (a_N' C_N^{-1} A_N)(a_N' C_N^{-1} A_N)(A_N' C_N^{-1} a_N) \leq g_1 p (A_N' C_N^{-1} A_N) \leq \delta < \infty.$$

since the $(i, j)$th element, for every $1 \leq i, j \leq p$, of $(A_N' C_N^{-1} A_N)$ is equal to $g_1 N g_j N v_i N C_N^{-1} v_j N$. Concerning the third term, it turns out that it will converges to zero. In fact:

$$|2a_N' C_N^{-1} A_N m - 2a_N' C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} A_N m|$$

$$= 2|a_N' C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} m|$$

$$\leq (a_N' C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} a_N)^{1/2} (m'(I_p + A_N' C_N^{-1} A_N)^{-1} m)^{1/2} \leq \delta g_p^{-1} \rightarrow 0.$$

**Proof of Theorem 4.1**

Given that $\mu - e r_f = \bar{\alpha} + B \bar{\lambda}$

$$w^{mv} = \frac{1}{\gamma} V^{-1}(\mu - e r_f)$$

$$= \frac{1}{\gamma} V^{-1} \bar{\alpha} + \frac{1}{\gamma} V^{-1} B \bar{\lambda}$$

$$= \frac{\gamma^\alpha}{\gamma} \bar{\alpha} + \frac{\gamma^\beta}{\gamma} V^{-1} B \bar{\lambda}$$

$$= \delta^\alpha w^\alpha + \delta^\beta w^\beta.$$

Moreover

$$w^{n'} V w^\beta = \frac{1}{\gamma^\beta} w^{n'} V V^{-1} B \bar{\lambda}$$

$$= \frac{1}{\gamma^\beta} w^{n'} B \bar{\lambda}$$

$$= \frac{1}{\gamma^\alpha \gamma^\beta} \bar{\alpha}' V B \bar{\lambda}$$

42
because \( \tilde{V} \) is orthogonal to \( B \). Similarly,

\[
\mathbf{w}^\alpha \Sigma \mathbf{w}^\beta = \frac{1}{\gamma^\beta} \mathbf{w}^\alpha \Sigma \mathbf{V}^{-1} B \hat{\lambda} = \frac{1}{\gamma^\beta} \mathbf{w}^\alpha B (I_K - (\Omega^{-1} + B' \Sigma^{-1} B)^{-1} B' \Sigma^{-1} B) \hat{\lambda} = \frac{1}{\gamma^\alpha \gamma^\beta} \alpha' \tilde{V} B (I_K - (\Omega^{-1} + B' \Sigma^{-1} B)^{-1} B' \Sigma^{-1} B) \hat{\lambda} = 0.
\]

We now show that the \( \mathbf{w}^\alpha \) and the \( \mathbf{w}^\beta \) portfolios are one the minimum-variance orthogonal portfolio of the other. This is accomplished by showing that these portfolio weights satisfy the result of Theorem B.11. In particular, we need to verify that \( \mathbf{w}^\beta \) satisfy

\[
\mathbf{w}^\beta = (\mathbf{w}^\alpha, \mathbf{V}^{-1}(\mu - r_f \mathbf{e})) \left( \begin{array}{c} 0 \\ \mathbf{0} \end{array} \right) \left( \begin{array}{cc} (\sigma^\alpha)^2 & \mu^\alpha - r_f \\ \mu^\alpha - r_f & (\text{SR}^\text{mv})^2 \end{array} \right)^{-1} \left( \begin{array}{c} \mu^\beta - r_f \\ 0 \end{array} \right).
\]

Simple calculations lead to

\[
\left( \begin{array}{c} (\sigma^\alpha)^2 & \mu^\alpha - r_f \\ \mu^\alpha - r_f & (\text{SR}^\text{mv})^2 \end{array} \right)^{-1} \left( \begin{array}{c} \mu^\beta - r_f \\ 0 \end{array} \right)
= \frac{1}{(\text{SR}^\text{mv})^2 \alpha''_N \bar{V} \alpha_N (\gamma^\alpha)^2 - (\alpha''_N \bar{V} \alpha_N (\gamma^\alpha)^2)} \left( \begin{array}{c} (\text{SR}^\text{mv})^2 - \frac{\alpha''_N \bar{V} \alpha_N}{\gamma^\alpha} \\ -\frac{\alpha''_N \bar{V} \alpha_N}{\gamma^\alpha} \end{array} \right) \left( \begin{array}{c} 0 \\ 1 \end{array} \right) = \frac{\lambda'_N B'_N \mathbf{V}^{-1} B_N \hat{\lambda}_N}{\gamma^\beta}.
\]

Given that

\[
\left( \begin{array}{c} \Sigma^{-1} \alpha_N \alpha_N \mathbf{V}^{-1} (\alpha_N + B_N \hat{\lambda}_N) \\ -\frac{\alpha''_N \bar{V} \alpha_N}{\gamma^\alpha} \end{array} \right) = \frac{\alpha''_N \bar{V} \alpha_N}{(\gamma^\alpha)^2} \mathbf{V}^{-1} B_N \hat{\lambda}_N,
\]

and

\[
\left( \begin{array}{c} (\text{SR}^\text{mv})^2 \bar{\alpha}'_N \bar{V} \alpha_N (\gamma^\alpha)^2 - (\alpha''_N \bar{V} \alpha_N (\gamma^\alpha)^2) \\ -\frac{\alpha''_N \bar{V} \alpha_N}{\gamma^\alpha} \end{array} \right) = \frac{1}{(\gamma^\alpha)^2} \left( \alpha'_N \bar{V} \alpha_N \right) \left( \lambda'_N B'_N \mathbf{V}^{-1} B_N \hat{\lambda}_N \right),
\]

one finally obtains

\[
\left( \mathbf{w}^\alpha, \mathbf{V}^{-1}(\mu - r_f \mathbf{e}) \right) \left( \begin{array}{c} (\sigma^\alpha)^2 & \mu^\alpha - r_f \\ \mu^\alpha - r_f & (\text{SR}^\text{mv})^2 \end{array} \right)^{-1} \left( \begin{array}{c} \mu^\beta - r_f \\ 0 \end{array} \right)
= \frac{1}{(\gamma^\alpha)^2} \frac{\alpha''_N \bar{V} \alpha_N}{\gamma^\beta} \mathbf{V}^{-1} B_N \hat{\lambda}_N \frac{\lambda'_N B'_N \mathbf{V}^{-1} B_N \hat{\lambda}_N}{\gamma^\beta} = \mathbf{w}^\beta.
\]
Proof of Theorem 4.2

Because $\hat{V}V\hat{V} = \hat{V}\Sigma\hat{V} = \hat{V}$, it follows that

$$\mu^\alpha = w^\alpha' \mu + (1 - w^\alpha')r_f = w^\alpha' (\mu - er_f) + r_f = \frac{1}{\gamma^\alpha} \hat{\alpha}' \hat{V}(\hat{\alpha} + B\hat{\lambda}) + r_f = \frac{1}{\gamma^\alpha} \hat{\alpha}' \hat{V}\hat{\alpha} + r_f,$$

$$\mu^\beta = w^\beta' \mu + (1 - w^\beta')r_f = w^\beta' (\mu - er_f) + r_f = \frac{1}{\gamma^\beta} \hat{\lambda}' B'\hat{V}^{-1}(\hat{\alpha} + B\hat{\lambda}) + r_f$$

Then use $\hat{\alpha}' \hat{V}\hat{\alpha} = \hat{\alpha}' \Sigma^{-1} \hat{\alpha}$.

Because $B'S^{-1}\hat{\alpha} = 0$ implies $B'V^{-1}\hat{\alpha} = 0$, one gets

$$\mu^\beta = w^\beta' \mu + (1 - w^\beta')r_f = w^\beta' (\mu - er_f) + r_f = \frac{1}{\gamma^\beta} \hat{\lambda}' B'\hat{V}^{-1}(\hat{\alpha} + B\hat{\lambda}) + r_f$$

The boundedness of $\hat{\alpha}' \hat{V}\hat{\alpha} = \hat{\alpha}' \Sigma^{-1} \hat{\alpha}$ follows from Theorem 3.1 (APT). Next,

$$B'V^{-1}B = B'(S^{-1} - S^{-1}B(\Omega^{-1} + B'S^{-1}B)^{-1}B'S^{-1}B)B$$

$$= B'S^{-1}B(\Omega^{-1} + B'S^{-1}B)^{-1}\Omega^{-1},$$

$$= (B'S^{-1}B + \Omega^{-1} - \Omega^{-1})(\Omega^{-1} + B'S^{-1}B)^{-1}\Omega^{-1},$$

$$= \Omega^{-1} - \Omega^{-1}(\Omega^{-1} + B'S^{-1}B)^{-1}\Omega^{-1}.$$ 

Therefore the positive definite matrix $B'V^{-1}B$ is bounded above by the constant matrix $\Omega^{-1}$, implying boundedness of the former. Premultiplying and postmultiplying the above expression by $\hat{\lambda}$ yields the result.

Finally,

$$(SR^{mv})^2 = (\mu - er_f)' V^{-1}(\mu - er_f) = (\hat{\alpha} + B\hat{\lambda})' V^{-1}(\hat{\alpha} + B\hat{\lambda})$$

$$= \hat{\alpha}' V^{-1}\hat{\alpha} + \hat{\lambda}' B' V^{-1} B\hat{\lambda}$$

$$= \hat{\alpha}' \Sigma^{-1} \hat{\alpha} + \hat{\lambda}' B' V^{-1} B\hat{\lambda}$$

$$= (SR^{\alpha})^2 + (SR^{\beta})^2,$$

where the third equality follows by the orthogonality.

Proof of Theorem 4.3

Parti (i). By the orthogonality $$(\mu_N - r_f e) V^{-1} (\mu_N - r_f e) = \hat{\alpha}' N^{-1} \hat{\alpha} N + \hat{\lambda}' B' N^{-1} B N \hat{\lambda}.$$ This implies $0 \leq \delta^\alpha \leq 1, 0 \leq \delta^\beta \leq 1.$
Recall that now $\Sigma_N = A_N A_N^T + C_N$ and $\alpha_N = A_N m + a_N$. Consider first $w^*_i$, where its $i$-th component satisfies: $w^*_i = e_i^T w^*_N = \frac{1}{\gamma} e_i^T \Sigma_N^{-1} \alpha_N - \frac{1}{\gamma} e_i^T \Sigma_N^{-1} B_N (B_N^T \Sigma_N^{-1} B_N)^{-1} B_N^T \Sigma_N^{-1} \alpha_N$. We deal with the two terms on the right hand side of $w^*_i$ separately. By the Sherman-Morrison-Woodbury formula $\Sigma_N^{-1} = C_N^{-1} - C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} A_N$, obtaining

$$e_i^T \Sigma_N^{-1} \alpha_N = e_i^T C_N^{-1} \alpha_N - e_i^T C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} \alpha_N$$

$$= e_i^T C_N^{-1} A_N m - e_i^T C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} A_N m$$

$$+ e_i^T C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} a_N$$

$$= e_i^T C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} m$$

By Holder inequality, taking the norm and using the relation between norm and maximum eigenvalue one obtains

$$|e_i^T \Sigma_N^{-1} \alpha_N| = O \left( \frac{\|e_i^T C_N^{-1} A_N\|}{g_{pp}(A_N' C_N^{-1} A_N)} + \|e_i^T C_N^{-1} a_N\| + \|a_N\| \frac{\|e_i^T C_N^{-1} \alpha_N\|}{g_{pp}(A_N' C_N^{-1} A_N)} \right)$$

$$= O \left( \frac{\|e_i^T C_N^{-1} \alpha_N\|}{f(N)} + \|e_i^T C_N^{-1} a_N\| + \|a_N\| \frac{\|e_i^T C_N^{-1} \alpha_N\|}{f(N)} \right).$$

Along the same lines

$$e_i^T \Sigma_N^{-1} B_N = e_i^T C_N^{-1} B_N - e_i^T C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} B_N,$$

$$B_N' \Sigma_N^{-1} B_N = B_N' C_N^{-1} B_N - B_N' C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} B_N,$$

$$B_N' \Sigma_N^{-1} \alpha_N = B_N' C_N^{-1} \alpha_N - B_N' C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} \alpha_N$$

$$= B_N' C_N^{-1} A_N m - B_N' C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} A_N m$$

$$+ B_N' C_N^{-1} a_N - B_N' C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} a_N$$

$$= B_N' C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} m$$

$$+ B_N' C_N^{-1} a_N - B_N' C_N^{-1} A_N (I_p + A_N' C_N^{-1} A_N)^{-1} A_N' C_N^{-1} a_N.$$

Under our assumptions, the eigenvalues of $A_N' C_N^{-1} A_N$ and $B_N' C_N^{-1} B_N$ have the same behavior. The same applies to $e_i^T C_N^{-1} A_N$ and $e_i^T C_N^{-1} B_N$. Therefore, by using the same arguments above, one obtains

$$|e_i^T \Sigma_N^{-1} B_N (B_N' \Sigma_N^{-1} B_N)^{-1} B_N' \Sigma_N^{-1} \alpha_N| = O \left( \|a_N\| \frac{\|e_i^T C_N^{-1} A_N\|}{f(N)} + \|e_i^T C_N^{-1} B_N\| \right).$$

For the $w^*_N$ portfolio, its $i$-th component satisfies:

$$w^*_i = \frac{1}{\gamma} e_i^T \Sigma_N^{-1} B_N \lambda_N - \frac{1}{\gamma} e_i^T \Sigma_N^{-1} B_N (\Omega^{-1} + B_N' \Sigma_N^{-1} B_N)^{-1} B_N' \Sigma_N^{-1} B_N \lambda_N$$

$$= \frac{1}{\gamma} e_i^T \Sigma_N^{-1} B_N (\Omega^{-1} + B_N' \Sigma_N^{-1} B_N)^{-1} \Omega^{-1} \lambda_N,$$
and using the above formulae for $e_j^i \Sigma^{-1}_N B_N$ and $B_N^i \Sigma^{-1}_N B_N$ concludes, where we use $\hat{\lambda}_N \to \lambda$.

Regarding part (ii), $w^\alpha w^\alpha \leq \| \Sigma^{-1}_N \| (\hat{\alpha}_N^i \Sigma^{-1}_N \alpha_N)/\gamma^\alpha \leq \delta < \infty$. Moreover, $\hat{\lambda}_N B_N^i V^{-1} V^{-1} B_N \hat{\lambda}_N = \hat{\lambda}_N \Omega^{-1}(\Omega^{-1} + B_N^i \Sigma^{-1}_N B_N)^{-1} B_N \Sigma^{-1}_N \Sigma^{-1}_N B_N (\Omega^{-1} + B_N^i \Sigma^{-1}_N B_N)^{-1} \Omega^{-1} \hat{\lambda}_N = O(1/f(N))$ because $\|B_N^i \Sigma^{-1}_N \Sigma^{-1}_N B_N \| \leq \| \Sigma^{-1}_N \| \| B_N^i \Sigma^{-1}_N B_N \|$. This implies $w^\alpha w^\alpha = O(1/f(N))$.

Part (iii) follows from $|e_N^i w^\alpha| \leq \frac{1}{\gamma^\alpha} (\hat{e}_N^i \Sigma^{-1}_N e_N)^{\frac{1}{2}} (\hat{\alpha}_N \Sigma^{-1}_N \hat{\alpha}_N)^{\frac{1}{2}}$ with $e_N^i \Sigma^{-1}_N e_N \to \infty$. On the other hand, $|e_N^i w^\beta| = \frac{1}{\gamma^\alpha} (|\hat{e}_N^i \Sigma^{-1}_N B_N| (\Omega^{-1} + B_N^i \Sigma^{-1}_N B_N)^{-1} \Omega^{-1} \hat{\lambda}| \leq \delta < \infty$.

**Proof of Theorem 4.4**

For the $w^\alpha$ portfolio the boundedness follows by no-arbitrage and positivity by non-singularity of $\Sigma_N$. For the $w^\beta$ portfolio the result follows from

$$\hat{\lambda}_N B_N^i V^{-1} V^{-1} B_N \hat{\lambda}_N = \hat{\lambda}_N (B_N^i \Sigma^{-1}_N B_N) (\Omega^{-1} + B_N^i \Sigma^{-1}_N B_N)^{-1} \hat{\lambda}_N,$$

by then taking the limit for $N \to \infty$. By Theorem 4.2, the variances and Sharpe ratios have the same expression for the $w^\alpha$ and $w^\beta$ portfolios, respectively, so the same limits above apply.

**Proof of Theorem 4.5**

The formulae for $\hat{\alpha}_{MLC}$, $\hat{B}_{MLC}$ and $\hat{\Sigma}_{MLC}$ easily follow by solving the first order conditions. For $\hat{\lambda}_{MLC}$ and $\hat{\Omega}_{MLC}$ one obtains precisely the sample mean and sample covariance matrix of the $f_t$.

**Proof of Theorem 4.6**

Differentiating the Lagrangian with respect to $m$ and $\alpha$ one obtains the $K + N$ equations (after some algebra):

$$
\begin{pmatrix}
A^i \Sigma^{-1}_N \\
I_N
\end{pmatrix}
\begin{pmatrix}
\bar{r} - r_\gamma e - \tilde{B}(\bar{f} - r_\gamma e_K)
\end{pmatrix}
=
\begin{pmatrix}
A^i \Sigma^{-1}_N A \\
A^i \Sigma^{-1}_N (1 + \kappa) I_N
\end{pmatrix}
\begin{pmatrix}
\tilde{m} \\
\tilde{a}
\end{pmatrix}.
$$

It is straightforward to see that, because of the APT restriction, $m_0$ and $a_0$ can now be identified separately, as long as $\kappa > 0$. In fact, the above system of linear equations can be solved since the matrix pre-multiplying $\tilde{m}$ and $\tilde{a}$ is non-singular for every $\kappa > 0$, leading to the closed-form solution:

$$
\tilde{m} = (A^i \Sigma^{-1}_N A)^{-1} A^i \Sigma^{-1}_N \left( \bar{r} - r_\gamma e - B(\bar{f} - r_\gamma e_K) \right),
$$

$$
\tilde{a} = \frac{1}{\kappa + 1} \left( \bar{r} - r_\gamma e - B(\bar{f} - r_\gamma e_K) - A \tilde{m} \right).
$$
Substituting \( \tilde{m} \) and \( \tilde{a} \) into the log-likelihood function and re-arranging gives:

\[
L(\theta) = -\frac{1}{2} \log(\det(AA' + C))
\]

\[
- \frac{1}{2T} \sum_{t=1}^{T} (\tilde{r}_t - G\tilde{r} - B\tilde{f}_t + GB\tilde{f})' (AA' + C)^{-1} (\tilde{r}_t - G\tilde{r} - B\tilde{f}_t + GB\tilde{f})
\]

\[
- \frac{1}{2} \log(\det(\Omega))
\]

\[
- \frac{1}{2T} \sum_{t=1}^{T} (f_t - r_f e - \lambda)' \Omega^{-1} (f_t - r_f e - \lambda),
\]

where we defined

\[
G = \frac{1}{(\kappa + 1)} I + \frac{\kappa}{(\kappa + 1)} A(A'\Sigma^{-1} A)^{-1} A'\Sigma^{-1},
\]

and

\[
\tilde{r}_t = (r_t - r_f e), \quad \tilde{r} = (\tilde{r} - r_f e),
\]

\[
\tilde{f}_t = (f_t - r_f e_K), \quad \tilde{f} = (\tilde{f} - r_f e_K).
\]

We also define, for simplicity,

\[
g_t = (\tilde{r}_t - G\tilde{r} - B\tilde{f}_t + GB\tilde{f}), \quad \tilde{g} = \frac{1}{T} \sum_{t=1}^{T} g_t.
\]

Differentiating with respect to \( B \) and solving gives:

\[
\frac{1}{T} \sum_{t=1}^{T} (\Sigma^{-1} g_t \otimes (-)\tilde{f}_t) + \frac{1}{T} \sum_{t=1}^{T} (G'\Sigma^{-1} g_t \otimes \tilde{f}) = 0,
\]

which can be re-written as

\[
1/T \sum_{t=1}^{T} \tilde{f}_t g'_t = \tilde{g}' \Sigma^{-1} G \Sigma.
\]

Next, recalling that \( \Sigma_{rf} = \frac{1}{T} \sum_{t=1}^{T} \tilde{r}_t \tilde{f}_t', \quad \Sigma_{ff} = \frac{1}{T} \sum_{t=1}^{T} \tilde{f}_t \tilde{f}_t', \) with \( \Sigma_{fr} = \Sigma_{rf}^\prime, \) one obtains

\[
\Sigma^{-1} G \Sigma = \frac{1}{(\kappa + 1)} I + \frac{\kappa}{(\kappa + 1)} \Sigma^{-1} \tilde{A} (\tilde{A}'\Sigma^{-1} \tilde{A})^{-1} \tilde{A}' = G',
\]

and rearranging gives

\[
\Sigma_{fr} - \tilde{f} \tilde{r}' G' - \Sigma_{ff} B' + \tilde{f} \tilde{f}' B' G' = (\tilde{f} \tilde{r}' - \tilde{f} \tilde{r}' B')(I - G') G',
\]

which can be written as

\[
\Sigma_{fr} - \tilde{f} \tilde{r}' (2G' - G' G') = \Sigma_{ff} B' - \tilde{f} \tilde{f}' B' (2G' - G' G') = \Sigma_{ff} B' + \tilde{f} \tilde{f}' B' (G' G' - 2G'),
\]

47
Transposing, taking the vec and re-arranging:

$$\text{vec}(B) = \left( (\Sigma_{ff} \otimes I) + (\tilde{f} f' \otimes (GG - 2G)) \right)^{-1} \text{vec}(\Sigma_{rf} - (2G - GG)\tilde{f} f').$$

We need to show that a solution for $\tilde{B}$ exists. This requires one to establish that the matrix

$$\left( (\Sigma_{ff} \otimes I) + (\tilde{f} f' \otimes (GG - 2G)) \right)$$

is invertible. This matrix can be rewritten as

$$\left( (\Sigma_{ff} \otimes I) + (\tilde{f} f' \otimes (IGG - 2G)) \right) = \left( (\Sigma_{ff} + \tilde{f} f') \otimes (I + GG - 2G) \right).$$

The first matrix on the right hand side is non-singular. One then just needs to show that the second matrix is also non-singular. This follows because

$$I + GG - 2G = I - \frac{(1 + 2\kappa)}{(\kappa + 1)^2} I - \left( \frac{\kappa}{1 + \kappa} \right)^2 (A^T \Sigma^{-1} A) - 1 A^T \Sigma^{-1}$$

$$= \left( \frac{\kappa}{1 + \kappa} \right)^2 (I - A (A^T \Sigma^{-1} A) - 1 A^T \Sigma^{-1})$$

$$= \left( \frac{\kappa}{1 + \kappa} \right)^2 (I - \Sigma^{-1} A (A^T \Sigma^{-1} A) - 1 A^T \Sigma^{-1})$$

$$= \left( \frac{\kappa}{1 + \kappa} \right)^2 (I - \Sigma^{-1} A (A^T \Sigma^{-1} A) - 1 A^T \Sigma^{-1} \Sigma^{-1} A (A^T \Sigma^{-1} A) - 1 A^T \Sigma^{-1} \Sigma^{-1} A).$$

The right-hand side is the product of positive definite matrices and of the semi-positive definite matrix $I - \Sigma^{-1/2} A (A^T \Sigma^{-1} A) - 1 A^T \Sigma^{-1/2}$, which is a projection matrix orthogonal to $\Sigma^{-1/2} A$, and therefore positive semi-definite. Therefore, one obtains that

$$\tilde{m} = \tilde{m}(A, C, \kappa), \quad \tilde{a} = \tilde{a}(A, C, \kappa), \quad \tilde{B} = \tilde{B}(A, C, \kappa).$$

Substituting them $L(\theta)$ gives the concentrated likelihood function, which is a function of only $A$ and $C$, where the solutions for $\tilde{m}, \tilde{a}, \tilde{B}$ can be found by an iterative procedure. Notice that all the estimates depend ultimately on $\kappa$ that can be chosen by cross-validation methods.

Consider now case $\kappa = 0$. Differentiating $L(\theta)$ with respect to $m$ and $a$ and re-arranging:

$$\left( \begin{array}{c} A^T \Sigma^{-1} \\ I_N \end{array} \right) (\tilde{r} - r f e - B(\tilde{f} - r f e_k)) = \left( \begin{array}{c} A^T \Sigma^{-1} \\ I_N \end{array} \right) \left( \begin{array}{c} \tilde{m} \\ \tilde{a} \end{array} \right),$$

where, for simplicity, we set $\tilde{m} = \tilde{m}_{ML}, \tilde{a} = \tilde{a}_{ML}$ and recall that $\Sigma = AA' + C$. One can clearly obtain a unique solution for $(A, I_N)(\tilde{m} \tilde{a}) = A\tilde{m} + \tilde{a}$. However, to solve for $\tilde{m}$ and $\tilde{a}$ separately, one needs to invert the matrix:

$$\left( \begin{array}{c} A^T \Sigma^{-1} \\ I_N \end{array} \right) (A, I_N) = \left( \begin{array}{c} A^T \Sigma^{-1} A \\ A \end{array} \right),$$

which is not possible because it is of dimension $(N + K) \times (N + K)$ but of rank $N$, as the left hand side shows that it is obtained as the product of two matrices of dimension $(N + K) \times N$. All the other parameters are identified separately together with $\alpha_0 = A_0 m_0 + a_0$, and their expressions easily follow by differentiating $L(\theta)$ and solving the first-order conditions. QED
B Appendix with Additional Results

B.1 Mitigating Model Misspecification in Global-Minimum-Variance Portfolios

The proofs for all the results in this appendix are available from the authors upon request.

B.1.1 Decomposing the global-minimum-variance portfolio

The global minimum-variance portfolio is given by:

\[
\mathbf{w}_{\text{gmv}} = \frac{\mathbf{V}^{-1}\mathbf{e}}{\mathbf{e}'\mathbf{V}^{-1}\mathbf{e}}.
\]

We present our analysis of the decomposition of this portfolio in two parts: first we consider the case in which the number of risky assets is finite and then the case in which the number of risky assets is asymptotically large.

The following theorem, which is valid for any finite \( N > K \), establishes the relations that exist across the global-minimum-variance portfolio, \( \mathbf{w}_{\text{gmv}} \), and two portfolios of which one is associated with idiosyncratic risk, \( \mathbf{w}_{\Sigma} \), and the other with factor risk, \( \mathbf{w}_{\Omega} \). The portfolios \( \mathbf{w}_{\Sigma} \) and \( \mathbf{w}_{\Omega} \), which can be interpreted as the “alpha” and “beta” components of \( \mathbf{w}_{\text{gmv}} \), are defined in the theorem below; see Figure 3.

**Theorem B.1** (Decomposing weights of global minimum-variance portfolio). Suppose that the vector of asset returns, \( \mathbf{r}_t \), satisfies Assumptions 3.1 and 3.2. Then for any finite \( N > K \), the minimum-variance portfolio weights satisfies

\[
\mathbf{w}_{\text{gmv}} = \delta_{\Sigma} \mathbf{w}_{\Sigma} + \delta_{\Omega} \mathbf{w}_{\Omega},
\]

where:

\[
\mathbf{w}_{\Sigma} = \frac{\tilde{\mathbf{V}}\mathbf{e}}{\mathbf{e}'\tilde{\mathbf{V}}\mathbf{e}}, \quad \mathbf{w}_{\Omega} = \frac{(\mathbf{V}^{-1} - \tilde{\mathbf{V}})\mathbf{e}}{\mathbf{e}'(\mathbf{V}^{-1} - \tilde{\mathbf{V}})\mathbf{e}},
\]

with:

\[
\tilde{\mathbf{V}} = \left[ \mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{\Sigma}^{-1} \right], \quad \delta_{\Sigma} = \frac{\mathbf{e}'\tilde{\mathbf{V}}\mathbf{e}}{\mathbf{e}'\mathbf{V}^{-1}\mathbf{e}}, \quad \delta_{\Omega} = 1 - \delta_{\Sigma}(B1)
\]

and \( 0 \leq \delta_{\Sigma} \leq 1 \). Furthermore, portfolios \( \mathbf{w}_{\Sigma} \) and \( \mathbf{w}_{\Omega} \) satisfy the orthogonality condition, \( \mathbf{w}_{\Sigma}'\mathbf{V}\mathbf{w}_{\Omega} = \mathbf{w}_{\Sigma}'\mathbf{\Sigma}\mathbf{w}_{\Omega} = 0 \), and \( \mathbf{w}_{\Sigma} \) is the minimum-variance portfolio that is orthogonal to \( \mathbf{w}_{\Omega} \), and vice versa.

**Remark B.1.1.** The inefficient portfolios \( \mathbf{w}_{\Sigma} \) and \( \mathbf{w}_{\Omega} \) are not just a decomposition of \( \mathbf{w}_{\text{gmv}} \) based on the decomposition of the return variance. Despite being inefficient, the portfolios \( \mathbf{w}_{\Sigma} \) and \( \mathbf{w}_{\Omega} \) can also be obtained as the result of an optimization problem, as in Roll (1980). Moreover, we have three-fund separation; that is, the entire efficient frontier can be obtained by a linear combination of any efficient portfolio together with the inefficient \( \mathbf{w}_{\Sigma} \) and \( \mathbf{w}_{\Omega} \) portfolios.
We now characterize the returns of the two components of the global minimum-variance portfolio.

**Theorem B.2** (Decomposing returns of global minimum-variance portfolio). Suppose that the vector of asset returns, \( r_t \), satisfies Assumptions 3.1 and 3.2. Then for any finite \( N > K \): the returns on the portfolios \( w^\Sigma \) and \( w^\Omega \) have mean, volatility (standard deviation), and Sharpe ratio given by the following expressions:

\[
\mu^\Sigma - r_f = \frac{e'\tilde{V}\hat{\alpha}}{e'\tilde{V}e}; \quad \mu^\Omega - r_f = \frac{e'(V^{-1} - \tilde{V})\hat{B}\hat{\lambda}}{e'(V^{-1} - \tilde{V})e};
\]

\[
\sigma^\Sigma = (e'\tilde{V}e)^{-\frac{1}{2}}; \quad \sigma^\Omega = \left(e'(V^{-1} - \tilde{V})e\right)^{-\frac{1}{2}};
\]

\[
SR^\Sigma = \frac{e'\tilde{V}\hat{\alpha}}{(e'\tilde{V}e)^{\frac{1}{2}}}; \quad SR^\Omega = \frac{e'(V^{-1} - \tilde{V})\hat{B}\hat{\lambda}}{\left(e'(V^{-1} - \tilde{V})e\right)^{\frac{1}{2}}},
\]

where \( \tilde{V} \) is defined in (B1). Furthermore, these Sharpe ratios satisfy the following bounds:

\[ |SR^\Sigma| < \infty, \quad |SR^\Omega| < \infty, \quad \text{and} \quad (SR^{gmv})^2 \leq (SR^\Sigma)^2 + (SR^\Omega)^2. \]

**Remark B.2.1.** Observe that the return on the alpha component \( w^\Sigma \) is independent of the distribution of the factors, that is, \( \hat{\lambda} \) and \( \Omega \).

**Remark B.2.2.** For the global minimum-variance portfolio and its components there is no guarantee that their Sharpe ratios will be positive.

### B.1.2 Mitigating misspecification in the beta component of returns

We report the global-minimum-variance portfolio weights for the cases with bounded and unbounded residual variation below and we show how to use this result to alleviate the problem of beta misspecification for the global-minimum-variance portfolio.

**Theorem B.3** (Weights of components of the global minimum-variance portfolio for large \( N \) with bounded residual variation). Suppose that the vector of asset returns, \( r_t \), satisfies Assumptions 3.1 and 3.2, and the loadings \( B_N \) and \( e_N \) are \( \Sigma_N \)-regular with the same scaling factor \( f(N) \). Moreover, \( B_N \) and \( e_N \) are not asymptotically collinear. As \( N \to \infty \):

(i) Then \( \delta^\Sigma \to 1 \) and \( \delta^\Omega \to 0 \). Furthermore, the absolute value of the components of the global-minimum-variance portfolio vectors, \( w^\Sigma_N \) and \( w^\Omega_N \), decrease at most at the rate:

\[
\delta^\Sigma |w^\Sigma_i| = O\left(\frac{|e_i'\Sigma_N^{-1}e_N|}{f(N)} + \frac{\|e_i'\Sigma_N^{-1}B_N\|}{f(N)}\right),
\]

\[
\delta^\Omega |w^\Omega_i| = O\left(\frac{\|e_i'\Sigma_N^{-1}B_N\|}{f^2(N)}\right).
\]
(ii) The sum of the squared components of the global-minimum-variance portfolio vectors $(\delta^\Sigma)^2 w_N^\Sigma w_N^\Sigma$ and $(\delta^\Omega)^2 w_N^\Omega w_N^\Omega$ converge to zero.

(iii) The sum of the components of the global-minimum-variance portfolio vectors $\delta^\Sigma |e^\epsilon w_N^\Sigma|$ converges to one, whereas $\delta^\Omega |e^\epsilon w_N^\Omega|$ converges to zero.

In the corollary below, we look at a special case of part (i) of the above theorem to get a better understanding of this result. The other parts of the theorem are unchanged.

**Corollary B.3.1** (Weights of components of the global-minimum-variance portfolio for large $N$ with bounded residual variation and $f(N) = N$). Suppose that the assumptions of Theorem B.3 are satisfied and that the rows of $B_N$ and $\Sigma_N^{-1}$, are uniformly bounded. Then:

$$\delta^\Sigma |w_i^\Sigma| = \mathcal{O}\left(\frac{1}{N}\right),$$

$$\delta^\Omega |w_i^\Omega| = \mathcal{O}\left(\frac{1}{N^2}\right).$$

**Theorem B.4** (Weights of global-minimum-variance portfolio for large $N$ with unbounded residual variation). Suppose that the vector of asset returns, $r_t$, satisfies Assumptions 3.1 and 3.2 and $\alpha_N \neq 0$. Suppose also that $A_N$, $B_N$, $e_N$ are $C_N$-regular with the same scaling factor $f(N)$. Moreover, neither $B_N$ nor $e_N$ are asymptotically collinear with $A_N$. As $N \to \infty$:

(i) Then $\delta^\Sigma \to 1$, $\delta^\Omega \to 0$ and

$$\frac{\delta^\Omega |w_i^\Omega|}{\delta^\Sigma |w_i^\Sigma|} \to 0.$$

(ii) The sum of the squared components of the global-minimum-variance portfolio vectors $(\delta^\Sigma)^2 w_N^\Sigma w_N^\Sigma$ and $(\delta^\Omega)^2 w_N^\Omega w_N^\Omega$ converge to zero.

(iii) The sum of the components of the global-minimum-variance portfolio vectors $\delta^\Sigma |e^\epsilon w_N^\Sigma|$ converges to one, whereas $\delta^\Omega |e^\epsilon w_N^\Omega|$ converges to zero.

(iv) The vector of weights for the global-minimum-variance portfolio are asymptotically equivalent element-by-element to the weights of $w^\Sigma$:

$$w^{gmv} \sim \delta^\Sigma w^\Sigma \sim w^\Sigma = \frac{\check{e}v}{e^\epsilon.} \quad (B2)$$

**Remark B.4.1.** The third result in Part (i) of the above theorem follows from the fact that under Assumptions 3.1 and 3.2, the absolute value of the components of the global-minimum-variance portfolio vectors decrease at most at the rate:

$$\delta^\Sigma |w_i^\Sigma| = \mathcal{O}\left(\frac{||e^\epsilon C^{-1}_N e_N|| + ||e^\epsilon C^{-1}_N A_N||}{f(N)}\right), \quad \delta^\Omega |w_i^\Omega| = \mathcal{O}\left(\frac{||e^\epsilon C^{-1}_N B_N|| + ||e^\epsilon C^{-1}_N A_N||}{f^2(N)}\right).$$

In the corollary below, we look at a special case of part (i) of the above theorem to get a better understanding of this result. The other parts of the theorem are unchanged.

51
Corollary B.4.1 (Weights of global-minimum-variance portfolio for large \(N\) with unbounded residual variation and \(f(N) = N\)). Suppose that the assumptions of Theorem B.4 are satisfied and that the rows of \(A_N, B_N, \) and \(\Sigma_N^{-1}\), are uniformly bounded, then, as \(N \to \infty\), the absolute value of the components of the global-minimum-variance portfolio vectors decrease at most at the rate:

\[
\delta^\Sigma |w_i^\Sigma| = O\left(\frac{1}{N}\right) \quad \text{and} \quad \delta^\Omega |w_i^\Omega| = O\left(\frac{1}{N^2}\right).
\]

Remark B.4.2. We now explain the intuition underlying the differences in the rates at which the two components of the global-minimum-variance portfolio decline. Let us start by looking at the \(\delta^\Sigma w^\Sigma\) component of the \(w_{gmv}\) portfolio. In order to minimize the portfolio variance, this portfolio will eliminate the variance arising from factor exposure by being orthogonal to \(B\). But, this is accomplished irrespective of the rate at which the weights decrease because of orthogonality of the portfolio weights to \(B\). In order to reduce the portfolio variance arising from idiosyncratic risk, the rate of change in portfolio weights has to be faster than \(1/N^{1/2}\), but the weight constraint requires that this rate be exactly equal to \(1/N\).\(^{37}\)

Let us now look at the \(\delta^\Omega w^\Omega\) component of the \(w_{gmv}\) portfolio. In order to minimize portfolio variance, this component needs to eliminate the variance arising from idiosyncratic exposure, \(\Sigma\). If the terms \(\delta^\Omega |w_i^\Omega|\) decrease at any rate slower than \(1/N\), then the systematic exposure explodes; and, if the rate is exactly \(1/N\), then the systematic exposure remains (and is equal to a constant). So, the only way to eliminate the systematic exposure is to have the terms \(\delta^\Omega |w_i^\Omega|\) decline faster than \(1/N\). Note that this faster rate implies that asymptotically these components sum to zero, which then allows the \(\delta^\Sigma w^\Sigma\) component to satisfy the weight constraint. For the special case considered in the corollary, this rate is exactly \(1/N^2\); in general, it could be any rate faster than \(1/N\).

Remark B.4.3. Notice that Part (iv) of the theorem illustrates the treatment of beta misspecification for the global-minimum-variance portfolio. From equation (B2), \(w^\Sigma\) depends on only \(\tilde{V}\), which is a function of just \(\Sigma\) and \(B\), unlike \(w^\Omega\), which depends also on \(V^{-1}\), and hence, \(\Omega\). Given that \(\delta^\Sigma\) goes to one, and hence \(\delta^\Omega\) goes to zero, the asymptotic expression for \(w_{gmv}\) is immune to the form of misspecification described in Case 3 of Section 3.5; that is, when \(\Omega\) is parameterized incorrectly in the estimation.

We now study the properties of the returns of the global-minimum-variance portfolio for the case in which the number of assets is large.

Theorem B.5 (Returns of global-minimum-variance portfolio for large \(N\)). For portfolio weights \(w^\Sigma\) and \(w^\Omega\) that satisfy the assumptions of Theorem B.4, we have the following

\(^{37}\)To understand why the rate of change in portfolio weights has to be faster than \(1/N^{1/2}\), suppose the covariance matrix is the identity matrix, \(I_N\). Then, if the portfolio weights are exactly equal to \(1/N^{1/2}\), the portfolio variance is \(\frac{1}{N^2} I_N = \frac{1}{N^2} = 1\), so as \(N\) increases it does not go to zero.
Remark B.5.1. From the expressions above the mean excess return and volatility of the global-minimum-variance portfolio depend on neither the factor risk premia, $\lambda$, nor the factor covariance matrix, $\Omega$ asymptotically.

Remark B.5.2. For $N \to \infty$, the returns on the components of the global-minimum-variance portfolios $\delta^\Sigma w^\Sigma$ and $\delta^\Omega w^\Omega$ satisfy:
\[ \lim_{N \to \infty} \delta_N^\Sigma \mu_N^\Sigma = r_f; \quad \lim_{N \to \infty} \delta_N^\Omega \mu_N^\Omega = 0. \]

Furthermore, the asymptotic variances of the returns on the components of the global-minimum-variance portfolios $\delta^\Sigma w^\Sigma$ and $\delta^\Omega w^\Omega$ converge to zero. The Sharpe ratios
\[ \text{SR}_\infty^\Sigma = \lim_{N \to \infty} \text{SR}_N^\Sigma = \lim_{N \to \infty} \frac{e' \tilde{V} \hat{\alpha}}{(e' \tilde{V} \tilde{e})^{\frac{1}{2}}} \quad \text{and} \quad \text{SR}_\infty^\Omega = \lim_{N \to \infty} \text{SR}_N^\Omega = \lim_{N \to \infty} \frac{e'(V^{-1} - \tilde{V}) B \hat{\lambda}}{(e'(V^{-1} - \tilde{V}) e)^{\frac{1}{2}}} \]
satisfy $|\text{SR}_\infty^\Sigma| < \infty$ and $|\text{SR}_\infty^\Omega| < \infty$, implying that $|\text{SR}_\infty^{gmv}| < \infty$. Observe that we have stated the above boundedness condition in terms of the absolute value of the Sharpe ratio because the expression for the mean excess return is not a quadratic form, and hence, it is not guaranteed to be positive.

### B.2 Mitigating Model Misspecification in Markowitz Frontier Portfolios

In this section, we analyze the portfolio weights and returns of the efficient frontier portfolios for the case where the risk-free asset is not available. The frontier portfolios (fp) are portfolios that are mean-variance efficient; that is, they are portfolios that have the minimum variance for a given target mean, $\mu^{fp}$, and their weights sum to unity. The minimum-variance frontier portfolio weights $w^{fp}$ satisfy the following optimization:
\[ w^{fp} = \arg\min_w \frac{1}{2} w' V w \quad \text{such that} \quad w' \mu = \mu^{fp} \quad \text{and} \quad w' e = 1. \]

Following Huang and Litzenberger (1988), the solution to the above optimization problem can be written as:
\[ w^{fp} = g + h \mu^{fp} \quad \text{(B3)} \]

---

38Huang and Litzenberger (1988) provide a nice intuitive interpretation of $g$ and $h$. They explain that $g$ is the vector of portfolio weights of a frontier portfolio having a zero expected rate of return ($\mu^{fp} = 0$), while $g + h$ is the vector of portfolio weights of a frontier portfolio having an expected rate of return equal to 1 ($\mu^{fp} = 1$). We see from (B3) that by varying $\mu^{fp}$ the entire portfolio frontier can be generated, which is a special case of two-fund separation. The frontier portfolio with the highest Sharpe ratio is the tangency portfolio, $w^{tan}$. 

---

53
where
\[ g = \frac{1}{d}(bV^{-1}e - aV^{-1}\mu), \quad h = \frac{1}{d}(cV^{-1}\mu - aV^{-1}e), \]
and
\[ a = e'V^{-1}\mu, \quad b = \mu'V^{-1}\mu, \quad c = e'V^{-1}e, \quad d = bc - a^2. \]

Our analysis of the weights of frontier portfolios is presented in two parts: first we consider the case in which the number of risky assets is finite and then the case in which the number of risky assets is asymptotically large. Similar to our analysis for mean-variance and global-minimum-variance portfolios, we show below that portfolios on the Markowitz efficient frontier are linear combinations of two component portfolios, with the first component \((w^\alpha\Sigma)\) itself consisting of a combination of \(w^\alpha\) and \(w^\Sigma\), the second component \((w^\beta\Omega)\) consisting of \(w^\beta\) and \(w^\Omega\), with \(w^\alpha\Sigma\) orthogonal to \(w^\beta\Omega\). We also show that as the number of assets increases asymptotically, the weights and returns of the portfolio that depends on the pricing errors, \(w^\alpha\Sigma\), go to zero at a slower rate than the weights \(w^\beta\Omega\), which depend on the factor exposure and their risk premia.

### B.2.1 Decomposing frontier portfolios

The following theorem, which is valid for any finite \(N > K\), establishes the relations that exist across the frontier portfolio and its two components, one that depends on the pricing errors and the other that does not. A particular frontier portfolio along with its two components is displayed in Figure B1.

**Theorem B.6 (Decomposing weights of frontier portfolios).** Suppose that the vector of asset returns, \(r_t\), satisfies Assumptions 3.1 and 3.2. Then for any finite \(N > K\), the frontier portfolio weights satisfies the following decomposition:

\[ w_{fp} = \delta^{\alpha\Sigma}w^{\alpha\Sigma} + \delta^{\beta\Omega}w^{\beta\Omega}, \quad (B4) \]

where
\[ w^{\alpha\Sigma} = \frac{e w^\alpha + f w^\Sigma}{e'(e w^\alpha + f w^\Sigma)}, \quad w^{\beta\Omega} = \frac{e w^\beta + f w^\Omega}{e'(e w^\beta + f w^\Omega)}, \]

and \(^{39}\)
\[ e = \frac{c\mu^* - a}{d}, \quad f = \frac{c(b - a\mu^* + y_j(c\mu^* - a))}{d}, \]
\[ \delta^{\alpha\Sigma} = e'(e w^\alpha + f w^\Sigma), \quad \delta^{\beta\Omega} = e'(e w^\beta + f w^\Omega) = 1 - \delta^{\alpha\Sigma}. \]

\(^{39}\)Note that \(e\) and \(e\) denote different objects: \(e\) denotes the scalar quantity given by \(e = \frac{2\gamma - a}{d}\) while \(e\) denotes the vector of ones.
The portfolios \( w^{\alpha\Sigma} \) and \( w^{\beta\Omega} \) satisfy the orthogonality condition:

\[
(w^{\alpha\Sigma})' V w^{\beta\Omega} = (w^{\alpha\Sigma})' \Sigma w^{\beta\Omega} = 0,
\]

and \( w^{\alpha\Sigma} \) is the minimum-variance portfolio that is orthogonal to \( w^{\beta\Omega} \), and vice versa.

**Proof of Theorem B.6**

Recall that 

\[
w^{fp} = g + h \mu^{fp},
\]

where

\[
g = \frac{1}{d} (b V^{-1} e - a V^{-1} \mu), \quad h = \frac{1}{d} (c V^{-1} \mu - a V^{-1} e),
\]

and

\[
a = e' V^{-1} \mu, \quad b = \mu' V^{-1} \mu, \quad c = e' V^{-1} e, \quad d = bc - a^2.
\]

Replacing \( \mu \) by \( e r_f + \hat{\alpha} + B \hat{\lambda} \) and \( V^{-1} \) by \( \tilde{V} + V^{-1} - \tilde{V} \) and re-arranging gives:

\[
\frac{1}{d} (b V^{-1} e - a V^{-1} \mu) = \frac{1}{d} \left( b V^{-1} e - a V^{-1} (e r_f + \hat{\alpha} + B \hat{\lambda}) \right)
\]

\[
= \frac{1}{d} \left( (b - r_f a) \tilde{V} e + (b - r_f a) (V^{-1} - \tilde{V}) e - a V^{-1} (\hat{\alpha} + B \hat{\lambda}) \right)
\]

\[
= \frac{1}{d} \left( c(b - r_f a) w^\Sigma + c(b - r_f a) w^{\Omega} - a \gamma w^\alpha - a \gamma w^\beta \right)
\]

and

\[
\frac{1}{d} (-a V^{-1} e + c V^{-1} \mu) = \frac{1}{d} \left( -a V^{-1} e + c V^{-1} (e r_f + \hat{\alpha} + B \hat{\lambda}) \right)
\]

\[
= \frac{1}{d} \left( (r_f c - a) \tilde{V} e + (r_f c - a) (V^{-1} - \tilde{V}) e + c V^{-1} (\hat{\alpha} + B \hat{\lambda}) \right)
\]

\[
= \frac{1}{d} \left( c(r_f c - a) w^\Sigma + c(r_f c - a) w^{\Omega} + c \gamma w^\alpha + c \gamma w^\beta \right).
\]

Multiplying the second expression by \( \mu^{fp} \) and collecting terms gives the formulae for the \( w^{\alpha\Sigma} \) and \( w^{\beta\Omega} \) portfolios, where we multiply and divide \( c w^\alpha + f w^\Sigma \) and \( c w^\beta + f w^{\Omega} \) by \( \delta^{\alpha\Sigma} \) and \( \delta^{\beta\Omega} \), respectively, so that the resultant weights sum up to unity. Orthogonality of the \( w^{\alpha\Sigma} \) and \( w^{\beta\Omega} \) follows.

**Remark B.6.1.** Equation (B4) is a manifestation of the two-fund separation because it can be shown that \( w^{\alpha\Sigma} \) and \( w^{\beta\Omega} \) are efficient portfolios.

**Remark B.6.2.** The orthogonality condition in (B5) above says that the two portfolios \( w^{\alpha\Sigma} \) and \( w^{\beta\Omega} \) are uncorrelated, both conditional on the factors and also unconditionally.

Next, we characterize the returns of the two components of the minimum-variance frontier portfolios when the number of assets is finite.
Figure B1: Frontier portfolios and their decomposition

In this figure, we plot the minimum-variance frontier portfolios, \( w^{fp} \), and show the decomposition of an arbitrary frontier portfolio into two portfolios, one that depends only on the pricing errors, \( w^{a\Sigma} \), and another that depends only on the factor exposures and their premia, \( w^{B\Omega} \).

Theorem B.7 (Decomposing returns of frontier portfolios). Suppose that the vector of asset returns, \( r_t \), satisfies Assumptions 3.1 and 3.2. Then for any finite \( N > K \), the returns on the portfolios \( w^{a\Sigma} \) and \( w^{B\Omega} \) have means, volatilities (standard deviations), and Sharpe ratios given by the following expressions:

\[
\mu^{a\Sigma} - r_f = \frac{1}{\delta^{a\Sigma}} (e\hat{\alpha} + \frac{f}{c}e')\tilde{V}\hat{\alpha};
\]

\[
\sigma^{a\Sigma} = \frac{1}{\delta^{a\Sigma}} \left( \left( \frac{f}{c} \right)^2 e'\tilde{V}e + c^2 \hat{\alpha}'\tilde{V}\hat{\alpha} + 2c\frac{f}{c}e'e'\tilde{V}\hat{\alpha} \right)^{\frac{1}{2}};
\]

\[
\mu^{B\Omega} - r_f = \frac{1}{\delta^{B\Omega}} (e\hat{\lambda}B' + \frac{f}{c}e')(V^{-1} - \tilde{V})B\hat{\lambda} = \frac{1}{\delta^{B\Omega}} (e\hat{\lambda}B' + \frac{f}{c}e')V^{-1}B\hat{\lambda};
\]

\[
\sigma^{B\Omega} = \frac{1}{\delta^{B\Omega}} \left( \left( \frac{f}{c} \right)^2 e'(V^{-1} - \tilde{V})e + c^2 \hat{\lambda}'B'V^{-1}B\hat{\lambda} + 2c\frac{f}{c}e'e'V^{-1}B\hat{\lambda} \right)^{\frac{1}{2}}.
\]

Moreover, when \( \hat{\alpha} \) is not asymptotically collinear with \( e/(e'V^{-1}e)^{\frac{1}{2}} \), the levels of the Sharpe ratios satisfy the following bounds:

\[0 \leq |SR^{a\Sigma}| < \infty \quad \text{and} \quad 0 \leq |SR^{B\Omega}| < \infty,\]

and their squares satisfy

\[(SR^{p})^2 \leq (SR^{a\Sigma})^2 + (SR^{B\Omega})^2.\]

Proof of Theorem B.7

The formulae for \( \mu^{a\Sigma} - r_f, \mu^{B\Omega} - r_f, \sigma^{a\Sigma} \) and \( \sigma^{B\Omega} \) can be easily derived by substituting in
\( \mu = \mathbf{e}_f + \hat{\alpha} + \mathbf{B}\lambda \). The term \(-r_f\) arises since, for example, \( \mu^{\alpha\Sigma} = (\mathbf{w}^{\alpha\Sigma})' \mu = (\mathbf{w}^{\alpha\Sigma})'(\mathbf{e}_f + \hat{\alpha} + \mathbf{B}\lambda) = ((\mathbf{w}^{\alpha\Sigma})' \mathbf{e}) r_f + (\mathbf{w}^{\alpha\Sigma})'(\hat{\alpha} + \mathbf{B}\lambda) = r_f + (\mathbf{w}^{\alpha\Sigma})'(\hat{\alpha} + \mathbf{B}\lambda) \).

Consider now the Sharpe ratios. For any finite \( N \) these are always bounded so one needs only to focus on the case of arbitrarily large \( N \). It is useful to consider the following identities:

\[
\begin{align*}
b &= (\mu - \mathbf{e}_f)' \mathbf{V}^{-1}(\mu - \mathbf{e}_f) + r_f^2 c + 2r_f e' \mathbf{V}^{-1}(\mu - \mathbf{e}_f), \\
d &= c(\mu - \mathbf{e}_f)' \mathbf{V}^{-1}(\mu - \mathbf{e}_f) - ((\mu - \mathbf{e}_f)' \mathbf{V}^{-1} \mathbf{e})^2, \\
e &= c(\mu - \mathbf{e}_f)' \mathbf{V}^{-1}(\mu - \mathbf{e}_f), \\
f &= c(\mu - \mathbf{e}_f)' \mathbf{V}^{-1}(\mu - \mathbf{e}_f) - (\mu' \mathbf{p} - r_f) e' \mathbf{V}^{-1}(\mu - \mathbf{e}_f).
\end{align*}
\]

One can clearly see that, for every

\[
|e| \leq \delta < \infty \text{ implying } |e|(\hat{\alpha}' \mathbf{V}^{-1} \hat{\alpha})^{\frac{1}{2}} \leq \delta < \infty,
\]

where the positive constant \( \delta \) is not always the same one. Therefore, the Sharpe ratio \( SR^{\alpha\Sigma} \) satisfies

\[
|SR^{\alpha\Sigma}| = \frac{(e' \hat{\alpha}' + \frac{L}{c} e' \mathbf{V} \hat{\alpha})}{\left(\left(\frac{L}{c}\right)^2 e' \mathbf{V} e + e'^2 \hat{\alpha}' \mathbf{V} \hat{\alpha} + 2e \frac{L}{c} e' \mathbf{V} \hat{\alpha}\right)^{\frac{1}{2}}} \leq \frac{(e(\hat{\alpha}' \mathbf{V} \hat{\alpha})^{\frac{1}{2}} + \frac{L}{c} (e' \mathbf{V} e)^{\frac{1}{2}}) (\hat{\alpha}' \mathbf{V} \hat{\alpha})^{\frac{1}{2}}}{\left(\left(\frac{L}{c}\right)^2 e' \mathbf{V} e + e'^2 \hat{\alpha}' \mathbf{V} \hat{\alpha} + 2e \frac{L}{c} e' \mathbf{V} \hat{\alpha}\right)^{\frac{1}{2}}} \leq \delta \frac{|A + B|}{(A^2 + B^2 - 2AB)^{\frac{1}{2}}} = \delta \frac{|A + B|}{|A - B|} \leq \delta < \infty,
\]

setting for simplicity \( A = e(\hat{\alpha}' \mathbf{V} \hat{\alpha})^{\frac{1}{2}}, B = \frac{L}{c} (e' \mathbf{V} e)^{\frac{1}{2}} \). We have shown above that \( A \) and \( B \) are bounded. We now show that for any \( N \) sufficiently large \( A \) will always differ from \( B \).

Taking the difference \( A - B \) it follows that for large \( N \)

\[
A - B \sim \frac{(\mu' \mathbf{p} - r_f)(e' \mathbf{V}^{-1} \mathbf{e})}{d} (\hat{\alpha}' \mathbf{V} \hat{\alpha} + (e' \mathbf{V} \hat{\alpha}) (e' \mathbf{V} e)).
\]

Ruling out that the pricing errors \( \hat{\alpha} \) and \( e/(e' \mathbf{V}^{-1} \mathbf{e})^{\frac{1}{2}} \) are asymptotically collinear, concludes the proof.

### B.2.2 Mitigating misspecification in the beta component of returns

In this section, we study the properties of mean-variance frontier portfolios in the absence of a risk-free asset; for conciseness, we will refer to these portfolios as “frontier portfolios”.
in the earlier sections, we first characterize the portfolio weights and then the properties of
the returns of frontier portfolios. Given that the frontier portfolios are combinations of the
global-minimum-variance portfolio and the tangency portfolio, which is a particular case
of the mean-variance portfolio, the characterization of the weights and returns of frontier
portfolios follow from the previous results.

In our analysis of the weights of frontier portfolios, we first report the results for the
case with bounded residual variation considered in Theorem 3.2 and then for the case with
unbounded residual variation considered in Theorem 3.3.

**Theorem B.8** (Weights of components of frontier portfolios for large \( N \)). Suppose that the vector of asset returns, \( \mathbf{r}_t \), satisfies Assumptions 3.1 and 3.2, and the loadings \( \mathbf{B}_N \) and \( \mathbf{e}_N \) are \( \mathbf{\Sigma}_N \)-regular with the same scaling factor \( f(N) \). Moreover, \( \mathbf{B}_N \) and \( \mathbf{e}_N \) are not asymptotically collinear. Finally, assume that \( \mufp \neq rf \) and
\[
(c_N/d_N)\mathbf{e}_N\mathbf{V}_N^{-1}\mathbf{B}_N\lambda \neq (\mufp - rf)^{-1}
\]
for any large \( N \). Then, as \( N \to \infty \), the absolute value of the components of the frontier portfolio vectors, \( \mathbf{w}_{N}^{\alpha} \) and \( \mathbf{w}_{N}^{\beta} \), decrease at most at the rate:

\[
|w_{i}^{\alpha\Sigma}| = O \left( \frac{|e_i^{\prime}\mathbf{\Sigma}_N^{-1}\alpha_N| + \frac{|e_i^{\prime}\mathbf{\Sigma}_N^{-1}\mathbf{e}_N|}{f^{\frac{3}{2}}(N)} + \frac{\|e_i^{\prime}\mathbf{\Sigma}_N^{-1}\mathbf{B}_N\|}{f^{\frac{3}{2}}(N)}}{f(N)} \right), \tag{B6}
\]

\[
|w_{i}^{\beta\Omega}| = O \left( \frac{\|e_i^{\prime}\mathbf{\Sigma}_N^{-1}\mathbf{B}_N\|}{f(N)} \right) = O \left( \frac{\|e_i^{\prime}\mathbf{\Sigma}_N^{-1}\mathbf{B}_N\|}{f(N)} \right). \tag{B7}
\]

**Proof of Theorem B.8**

We first need to study the behavior of the normalizing constants \( \delta^{\beta\Omega} \) and \( \delta^{\alpha\Sigma} \). It turns out that they are both bounded and bounded away from zero. Therefore they can be ignored when studying the behavior of the portfolio weights for large \( N \). In turn, this means that to establish the asymptotic behavior of the portfolio weights for the frontier portfolio, one simply needs to add up the rates obtained for the mean-variance and for the global-minimum variance portfolios.

Concerning \( \delta^{\alpha\Sigma} \), a few algebraic steps lead to

\[
\delta^{\alpha\Sigma} = \frac{(\mu - \mathbf{e}\mathbf{r})^{\prime}\mathbf{V}^{-1}(\mu - \mathbf{e}\mathbf{r}) - (\mu fp - rf)}{d} \mathbf{e}^{\prime}\mathbf{V}\mathbf{e} + \frac{(\mu fp - rf)}{d} (\mathbf{e}^{\prime}\mathbf{V}^{-1}\alpha)(c - \mathbf{e}^{\prime}\mathbf{V}\mathbf{e})
\]

\[- \frac{(\mu fp - rf)}{d} (\mathbf{e}^{\prime}\mathbf{V}^{-1}\mathbf{B}\lambda)\mathbf{e}^{\prime}\mathbf{V}\mathbf{e} - (\frac{\|\mathbf{e}^{\prime}\mathbf{V}^{-1}\alpha\|}{d} - (\frac{\|\mathbf{e}^{\prime}\mathbf{V}^{-1}\mathbf{B}\lambda\|}{d} - (\frac{\|\mathbf{e}^{\prime}\mathbf{V}^{-1}\alpha\|}{d} - (\frac{\|\mathbf{e}^{\prime}\mathbf{V}^{-1}\mathbf{B}\lambda\|}{d} - (\frac{\|\mathbf{e}^{\prime}\mathbf{V}^{-1}\alpha\|}{d} - (\frac{\|\mathbf{e}^{\prime}\mathbf{V}^{-1}\mathbf{B}\lambda\|}{d})\mathbf{e}^{\prime}\mathbf{V}\mathbf{e}
\]

\[\sim \frac{(\mu fp - rf)}{d} (\mathbf{e}^{\prime}\mathbf{V}^{-1}\mathbf{B}\lambda)\mathbf{e}^{\prime}\mathbf{V}\mathbf{e}, \]
where the symbol $\sim$ means asymptotic equivalence (the other terms are of smaller order) for large $N$; see footnote 34. Similarly, one can show that

$$
\delta^{\beta\Omega} \sim \frac{(\mu^f p - r_f)}{d} (e'V^{-1}\hat{B}\hat{\lambda}) e'^{\hat{V}} e
$$

for large $N$. Therefore, both expressions show that $\delta^{\alpha\Sigma}$ and $\delta^{\beta\Omega}$ are bounded. Moreover, as expected, note that $\delta^{\alpha\Sigma} + \delta^{\beta\Omega} \sim 1$ also using these asymptotic equivalence results. Finally, our assumptions imply that both $\delta^{\alpha\Sigma}$ and $\delta^{\beta\Omega}$ are bounded away from zero. Therefore, these normalizing constants can be ignored for the purpose of evaluating the limiting behavior of the weights.

The proof is concluded by recalling that, from the proof of the previous theorem, $|e|$ is bounded and that $|f/c|$ decays to zero at rate $f^{-\frac{1}{2}}(N)$, and by using the results of Theorems 4.3 and B.4.

From equations (B6) and (B7), we see that $w^{\alpha\Sigma}_i$ always dominates $w^{\beta\Omega}_i$ as the number of assets increases. In the corollary below, we look at a special case to get a better understanding of this result.

**Corollary B.8.1** (Weights of components of frontier portfolios for large $N$ with bounded residual variation and $f(N) = N$). Suppose that the assumptions of Theorem B.8 are satisfied and that the rows of $B_N$ and $\Sigma^{-1}_N$ are uniformly bounded. Then, as $N \to \infty$, the absolute value of the components of the frontier portfolio vectors, $w^{\alpha\Sigma}$ and $w^{\beta\Omega}$, decrease at the rate:

$$
|w^{\alpha\Sigma}_i| = O\left(\frac{1}{N^2} + |e'\Sigma^{-1}_N \alpha_N|\right);
$$

$$
|w^{\beta\Omega}_i| = O\left(\frac{1}{N}\right).
$$

Comparing (B8) and (B9), we see that $w^{\alpha\Sigma}_i$ will always go to zero at a slower rate than $w^{\beta\Omega}_i$. If the pricing error $\alpha_N$ is zero, then (B8) simplifies to $1/N$, which is the rate at which the frontier portfolio weights decay to zero as we increase the number of assets under the traditional setting of exact pricing.

We now study the case of unbounded residual variation, that is, the case where $\Sigma_N$ satisfies the conditions listed in Theorem 3.3, which implies that

$$
\Sigma_N = A_N A'_N + C_N.
$$

As discussed before, the following specification of $\alpha_N$ is consistent with no arbitrage:

$$
\alpha_N = A_N m + a_N.
$$
Theorem B.9 (Weights of components of frontier portfolio for large \( N \) with unbounded residual variation). Suppose that the vector of asset returns, \( \mathbf{r}_t \), satisfies Assumptions 3.1 and 3.2. Suppose also that \( \mathbf{A}_N \) and \( \mathbf{B}_N \) are \( \mathbf{C}_N \)-regular with the same scaling factor \( f(N) \) and \( \mathbf{A}_N \) and \( \mathbf{B}_N \) are not asymptotically collinear. Finally, assume that \( \mu^{fp} \neq r_f \) and \( (c_N/d_N)\mathbf{e}_N^\prime \mathbf{V}_N^{-1} \mathbf{B}_N \hat{\lambda} \neq (\mu^{fp} - r_f)^{-1} \) for any large \( N \). Then, as \( N \to \infty \), the absolute value of the components of the frontier portfolio vectors, \( \mathbf{w}_N^{\alpha} \) and \( \mathbf{w}_N^{\beta} \), decrease at most at the rate:

\[
|w_i^{\alpha}| = O \left( \left| \mathbf{e}_N^\prime \mathbf{C}_N^{-1} \mathbf{a}_N \right| + \left| \mathbf{e}_N^\prime \mathbf{C}_N^{-1} \mathbf{e}_N \right| \frac{1}{f(N)} + \left| \mathbf{a}_N \right| \left( \left| \mathbf{e}_N^\prime \mathbf{C}_N^{-1} \mathbf{B}_N \right| \frac{1}{f(N)} + \left| \mathbf{e}_N^\prime \mathbf{C}_N^{-1} \mathbf{A}_N \right| \frac{1}{f(N)} \right)
\]

\[
|w_i^{\beta}| = O \left( \left| \mathbf{e}_N^\prime \mathbf{C}_N^{-1} \mathbf{B}_N \right| \frac{1}{f(N)} + \left| \mathbf{e}_N^\prime \mathbf{C}_N^{-1} \mathbf{A}_N \right| \frac{1}{f(N)} \right)
\]

Proof of Theorem B.9

The result follows along the same lines used for the bounded variation case.

Corollary B.9.1 (Weights of frontier portfolios for large \( N \) with unbounded residual variation and \( f(N) = N \)). Suppose that the assumptions of Theorem B.9 are satisfied and that the rows of \( \mathbf{A}_N, \mathbf{B}_N \) and \( \mathbf{C}_N^{-1} \) are uniformly bounded. Then, as \( N \to \infty \), the absolute value of the components of the frontier portfolio vectors, \( \mathbf{w}_N^{\alpha} \) and \( \mathbf{w}_N^{\beta} \), decrease at the rate:

\[
|w_i^{\alpha}| = O \left( \left| \mathbf{e}_N^\prime \mathbf{C}_N^{-1} \mathbf{a}_N \right| + \left| \mathbf{m} \right| \frac{1}{N} + \left| \mathbf{a}_N \right| \frac{1}{N^2} + \frac{1}{N^3} \right)
\]

\[
|w_i^{\beta}| = O \left( \frac{1}{N} \right)
\]

Finally, we study the properties of the returns of frontier portfolios for the case in which the number of assets is large. The results for both the case of bounded and the case of unbounded residual variation are presented in a single theorem below.

Theorem B.10 (Returns of frontier portfolios and its components for large \( N \)). Suppose that the vector of asset returns, \( \mathbf{r}_t \), satisfies Assumptions 3.1 and 3.2 and \( \mathbf{a}_N \neq 0 \). Then, for portfolio weights \( \mathbf{w}_N^{\alpha} \) and \( \mathbf{w}_N^{\beta} \) that satisfy the assumptions of Theorem 4.3 for the case of unbounded residual variation, we have the following results. For \( N \to \infty \), the asymptotic contribution to the mean of the frontier portfolios of the returns on the portfolios \( \mathbf{w}_N^{\alpha} \) and \( \mathbf{w}_N^{\beta} \) is defined by the expressions below:

\[
\lim_{N \to \infty} (\mu_N^{\alpha} - r_f) \delta_N^{\alpha} = (\mu^\ast - r_f) \kappa^{\alpha}, \quad \lim_{N \to \infty} (\mu_N^{\beta} - r_f) (1 - \delta_N^{\alpha}) = (\mu^\ast - r_f) (1 - \kappa^{\alpha})
\]
where
\[ \kappa^{\alpha\Sigma} = \lim_{N \to \infty} \frac{c_N \hat{\alpha}'_N \hat{\beta}_N - (e'_N \hat{\beta}_N)^2}{c_N b_N - a^2_N} \]

with \( 0 \leq \kappa^{\alpha\Sigma} \leq 1 \). Furthermore, the contribution to the asymptotic variances of the frontier portfolios are:
\[
\lim_{N \to \infty} (\delta^{\alpha\Sigma}_N \sigma^{\alpha\Sigma}_N)^2 = (\mu^* - r_f)^2 \kappa^{\alpha\Sigma} \lim_{N \to \infty} \frac{c_N}{d_N},
\]
\[
\lim_{N \to \infty} ((1 - \delta^{\alpha\Sigma}_N) \sigma^{\beta\Omega}_N)^2 = (\mu^* - r_f)^2 (1 - \kappa^{\alpha\Sigma}) \lim_{N \to \infty} \frac{c_N}{d_N}.
\]

Asymptotically, the Sharpe ratio of the contribution of the portfolios \( \mathbf{w}^{\alpha\Sigma} \) and \( \mathbf{w}^{\beta\Omega} \) are given by:
\[
\text{SR}^{\alpha\Sigma}_\infty = \lim_{N \to \infty} \text{SR}^{\alpha\Sigma} = \sqrt{\kappa^{\alpha\Sigma}} \lim_{N \to \infty} \sqrt{\frac{d_N}{c_N}},
\]
\[
\text{SR}^{\beta\Omega}_\infty = \lim_{N \to \infty} \text{SR}^{\beta\Omega} = \sqrt{1 - \kappa^{\alpha\Sigma}} \lim_{N \to \infty} \sqrt{\frac{d_N}{c_N}}.
\]

and
\[
(\text{SR}_\infty^f)^2 = (\text{SR}_\infty^{\alpha\Sigma})^2 + (\text{SR}_\infty^{\beta\Omega})^2 = \lim_{N \to \infty} \frac{d_N}{c_N}.
\]

**Proof of Theorem B.10**

The result follows by using the formulae of Theorem B.7. For instance for \( \mu^{\alpha\Sigma} \) one gets
\[
\delta^{\alpha\Sigma}_N (\mu^{\alpha\Sigma}_N - r_f) \sim (\mu^f - r_f) \left( \frac{c (\hat{\alpha}' V^{-1} \hat{\alpha}) - (\hat{\alpha}' V^{-1} e)^2}{d} \right) \sim (\mu^f - r_f) \kappa^{\alpha\Sigma},
\]
where one uses the fact that
\[
b \sim (\mu^f - r_f) \frac{c}{d}, \quad \frac{f}{c} \sim -(\mu^f - r_f) \frac{\hat{\alpha}' V^{-1} e}{d}.
\]

Similarly, for \( \mu^{\alpha\Sigma}_N \) one gets
\[
(1 - \delta^{\alpha\Sigma}_N) (\mu^{\alpha\Sigma}_N - r_f) \sim (\mu^f - r_f) \left( \frac{c (\hat{\lambda}' B' V^{-1} B \hat{\lambda}) - (\hat{\lambda}' V^{-1} e) (\hat{\lambda}' V^{-1} e)}{d} \right) \sim (\mu^f - r_f) (1 - \kappa^{\alpha\Sigma}).
\]

The same reasoning applies for the limit of the portfolio variances and Sharpe ratios.

**Remark B.10.1.** Both the \( \mathbf{w}^{\alpha\Sigma} \) and \( \mathbf{w}^{\beta\Omega} \) portfolios have a positive contribution to the target mean of the frontier portfolio. The weight \( \kappa^{\alpha\Sigma} \) is asymptotically proportional to the contribution of \( \hat{\alpha}_N \) to \( d_N \) and the weight \( 1 - \kappa^{\alpha\Sigma} \) is asymptotically proportional to the contribution of \( \hat{\lambda}_N \) to \( d_N \).
Remark B.10.2. The Sharpe ratios of the frontier portfolios are asymptotically equivalent to one another, implying that they are independent of the target means. This implies that the frontier will tend to the capital-market lines, whose slopes are equal to $\frac{dN}{cN}$ and $-\frac{dN}{cN}$ in the limit. This, together with the convergence of the global-minimum-variance-portfolio return to the risk-free rate, implies that the weights in all the frontier portfolios are non-negative asymptotically and satisfy the weight constraint. The reason for this is that, for a finite number of assets, all the portfolios on the capital market line that lie on the segment between the risk-free asset and the tangency portfolio have non-negative weights. As the number of assets increases, this segment will cover the entire capital market line. Moreover, all portfolios on the frontier satisfy the weight constraint, implying that as the number of assets increases, the same must apply to the portfolios on the capital market line.

B.3 Extension of Roll (1980)

Roll (1980) shows that in the absence of a risk-free rate, for any inefficient portfolio one can identify the subspace of portfolios that are orthogonal to this portfolio with minimum variance. That is, corresponding to any inefficient portfolio, there are an infinite number of zero-beta portfolios—one for each level of target mean. If the portfolio is efficient, then the subspace shrinks to a single point; that is, there is a unique zero-beta portfolio. In order to interpret our findings, we extend the result in Roll (1980) to the case where investors can invest also in a risk-free asset.

**Theorem B.11** (Extension of Roll (1980) to the case with a risk-free asset). Let $w^x$ be any, possibly inefficient, portfolio. Let $w^z$ be the portfolio that satisfies

$$\min \frac{1}{2}(w^z)'Vw^z \quad s.t. \quad (w^x)'Vw^z = 0,$$

and

$$\mu'w^z + (1 - e'w^z)r_f = \mu^z,$$

for a given target mean $\mu^z$. Then,

$$w^z = \left( w^x, V^{-1}(\mu - r_f e) \right) \begin{pmatrix} (\sigma^x)^2 & \mu^x - r_f \\ \mu^x - r_f & (\text{SR}_{\text{mv}})^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \mu^z - r_f \end{pmatrix},$$

where $\left( w^x, V^{-1}(\mu - r_f e) \right)$ is the $N \times 2$ matrix obtained by joining the $N \times 1$ vector of portfolio weights $w^x$ with the $N \times 1$ vector $V^{-1}(\mu - r_f e)$.

**Proof of Theorem B.11**

We adapt Roll’s (1980) proof of the main theorem. The Lagrangian for our problem is:

$$L(w^x, \lambda_1, \lambda_2) = (w^z)'Vw^z - \lambda_1((w^x)'Vw^z) - \lambda_2(\mu'w^z + (1 - e'w^z)r_f - \mu^z),$$
with first-order conditions:

$$2Vw^z = \begin{pmatrix} Vw^x, (\mu - er_f) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}.$$ 

Pre-multiplying both sides by $2^{-1}\begin{pmatrix} Vw^x, (\mu - er_f) \end{pmatrix}^T V^{-1}$ gives:

$$\begin{pmatrix} 0 \\ \mu^2 - rf \end{pmatrix} = \frac{1}{2} \begin{pmatrix} (\sigma^2)^2 \\ \mu^2 - rf \end{pmatrix} \begin{pmatrix} (SR_{mv})^2 \\ (SR^x)^2 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}.$$ 

Substituting out for $\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}$ concludes the proof.

**Remark B.11.1.** When $w^x$ is efficient, then $w^z = 0$, which implies that the zero-beta portfolio to $w^x$ is the portfolio that invests 100% in the risk-free asset.

Recall the well-known result that the entire efficient frontier, which in the presence of a risk-free asset is the capital market line, can be generated from holding any two efficient portfolios. However, one can show that the efficient frontier can be generated also by holding two *inefficient* portfolios, as long as one is the minimum-variance orthogonal portfolio of the other. This leads to the following result.

**Corollary B.11.1** (Extension of Corollary 3 of Roll (1980) to the case with a risk-free asset). There is a weighted average of, possibly inefficient, portfolio $w^x$ with a corresponding minimum-variance orthogonal portfolio $w^z$ that produces an efficient portfolio.

**Remark B.11.2.** The above theorem implies that the subspace of minimum-variance portfolios orthogonal to $w^x$ is given by the two lines described by the expression below:

$$\mu^z = rf \pm \sigma^z \sqrt{(SR_{mv})^2 - (SR^x)^2}.$$ 

Notice from the equation above and the dashed and dotted lines in Figure 4 that the slopes of the two lines are smaller (in absolute value) than the slopes of the capital market lines.\(^{40}\) For portfolios that are efficient, the subspace shrinks to a single point, which is the risk-free rate of return, as one can see from setting the Sharpe ratio of portfolio $w^x$ equal to the Sharpe ratio of the mean-variance portfolio $w^m$ in the equation above.

### B.4 Some Useful Results on the APT

In this section, we show that the APT restriction in terms of $\hat{\alpha}$ has an equivalent representation in terms of $\alpha_N$. Note that from (4) and (5) we have:

$$\hat{\alpha}_N = (\mu_N - rf e) - B_N \hat{\lambda}$$

\(^{40}\)Huang and Litzenberger (1988) show that, depending on the level of the risk-free rate relative to the mean of the global minimum-variance portfolio, the capital market line can be sloping up or down.
\[ \begin{align*}
= (I_N - B_N (B'_N \Sigma^{-1}_N B_N)^{-1} B'_N \Sigma^{-1}_N) (\alpha_N + B_N \lambda) \\
= (I_N - B_N (B'_N \Sigma^{-1}_N B_N)^{-1} B'_N \Sigma^{-1}_N) \alpha_N.
\end{align*} \]

Because
\[ \Sigma^{-1}_N \hat{\alpha}_N = \Sigma^{-1}_N (I_N - B_N (B'_N \Sigma^{-1}_N B_N)^{-1} B'_N \Sigma^{-1}_N) \alpha_N = \tilde{V}_N \alpha_N, \]

it follows that
\[ \hat{\alpha}'_N \Sigma^{-1}_N \hat{\alpha}_N = \hat{\alpha}'_N \Sigma^{-1}_N \Sigma^{-1}_N \hat{\alpha}_N \\
= \alpha'_N \tilde{V}_N \Sigma \tilde{V}_N \alpha_N \\
= \alpha'_N \tilde{V}_N \alpha_N, \tag{B10} \]

where \( \tilde{V} \) is defined in (B1). Therefore, the condition in (B11) below,
\[ \alpha'_N \Sigma^{-1}_N \hat{\alpha}_N \leq \delta < \infty, \tag{B11} \]

which is the central result of the APT (see, for instance, Ingersoll (1984, Theorem 1), Huberman (1982, Theorem 1), and Chamberlain and Rothschild (1983, Theorem 3')), implies from (B10) that
\[ \alpha'_N \tilde{V}_N \alpha_N = \alpha'_N \Sigma^{-1}_N \alpha_N - \alpha'_N \Sigma^{-1}_N B_N (B'_N \Sigma^{-1}_N B_N)^{-1} B'_N \Sigma^{-1}_N \alpha_N \leq \delta < \infty. \tag{B12} \]

Furthermore, because \( \alpha'_N \tilde{V}_N \alpha_N \geq 0 \), therefore (B12) implies that
\[ 0 \leq \alpha'_N \Sigma^{-1}_N B_N (B'_N \Sigma^{-1}_N B_N)^{-1} B'_N \Sigma^{-1}_N \alpha_N \leq \alpha'_N \Sigma^{-1}_N \alpha_N \leq \delta < \infty, \]

implying that (7), as well as the equation below, hold by no-arbitrage.
\[ \alpha'_N \Sigma^{-1}_N B_N (B'_N \Sigma^{-1}_N B_N)^{-1} B'_N \Sigma^{-1}_N \alpha_N \leq \delta < \infty. \]

### B.5 Examples of Estimation of Models with Misspecification

In this section, we provide two examples of the estimation procedure described in the main text. The first example is for the bounded-variation case and the second for the unbounded-variation case.

As an example of bounded-variation case, consider a model with one observed factor \( (K = 1) \) with residuals uncorrelated across assets and with the same variance \( \sigma^2_0 \). The observed factor has mean excess return equal to \( \lambda_0 \) and variance \( \omega^2_0 \). The true unconditional means and covariances of excess returns satisfy
\[ \mu_0 - r_f e = \alpha_0 + \beta_0 \lambda_0, \quad V_0 = \omega^2_0 \beta_0 \beta'_0 + \sigma^2_0 I. \]

The ML estimator for this case is
\[ \hat{\theta}_{ML} = \arg\max_{\theta} L(\theta), \]

where
\[ L(\theta) = -\frac{N}{2} \log(\sigma^2) - \frac{1}{2T\sigma^2} \sum_{t=1}^{T} \left( r_t - r_f e - \alpha - \beta(f_t - f_f) \right)' \left( r_t - r_f e - \alpha - \beta(f_t - f_f) \right) \]

\[ - \frac{1}{2} \log(\omega^2) - \frac{1}{2T\omega^2} \sum_{t=1}^{T} \left( f_t - r_f - \lambda \right)^2, \]

with \( \theta = (\alpha', \beta', \sigma^2, \lambda, \omega^2)' \). Clearly the ML estimator coincides with the OLS estimator.\(^{41}\)

To impose the APT restriction, one should instead consider the constrained estimator:

\[ \hat{\theta}_{MLC} = \arg\max_{\theta} L(\theta) \text{ such that } \frac{\alpha'\alpha}{\sigma^2} \leq \delta. \]

With respect the penalized log-likelihood \( L(\theta) - \kappa(\alpha'\alpha/\sigma^2 - \delta) \) one obtains the ridge-regression estimator for \( \alpha_0 \):

\[ \hat{\alpha}_{MLC} = \frac{1}{(1 + \hat{\kappa}_{MLC})} \left[ \bar{r} - r_f e - \hat{\beta}_{MLC}(\bar{f} - \bar{r}) \right], \]

and for the other parameters one obtains:

\[ (1 + \hat{\kappa}_{MLC}) = \frac{\left[ (\bar{r} - r_f e - \hat{\beta}_{MLC}(\bar{f} - \bar{r}))(\bar{r} - r_f e - \hat{\beta}_{MLC}(\bar{f} - \bar{r})) \right]^{1/2}}{\hat{\sigma}_{MLC}^2 \delta^{1/2}}, \]

\[ \hat{\sigma}_{MLC}^2 = \frac{1}{TN} \sum_{t=1}^{T} \left( \bar{r}_t - \hat{\beta}_{MLC} \bar{f}_t \right)' \left( \bar{r}_t - \hat{\beta}_{MLC} \bar{f}_t \right), \]

\[ \hat{\beta}_{MLC} = \frac{\sum_{t=1}^{T} \bar{r}_t \bar{f}_t}{\sum_{t=1}^{T} \bar{f}_t^2}, \]

and, as before, \( \hat{\lambda}_{MLC} = \frac{1}{T} \sum_{t=1}^{T} (f_t - r_f) \) and \( \hat{\omega}_{MLC}^2 = \frac{1}{T} \sum_{t=1}^{T} (f_t - \bar{f})^2 \), with \( \bar{f}_t \) and \( \bar{r}_t \) as defined in Theorem 4.5.

Consider now the unbounded-variation case. The true unconditional means and covariances of returns in this case satisfy

\[ \mu_0 = r_f e = A_0 m_0 + a_0 + \beta_0 \lambda_0, \quad \nu_0 = \omega_0^2 \beta_0^2 + A_0 A_0' + \epsilon_0^2 I. \]

\(^{41}\)The OLS estimator is:

\[ \hat{\alpha}_{ML} = \bar{r} - r_f e - \hat{\beta}_{ML}(\bar{f} - \bar{r}), \quad \hat{\beta}_{ML} = \frac{\sum_{t=1}^{T} (r_t - \bar{r})(f_t - \bar{f})}{\sum_{t=1}^{T} (f_t - \bar{f})^2}, \]

\[ \hat{\sigma}_{ML}^2 = \frac{1}{TN} \sum_{t=1}^{T} (r_t - r_f e - \hat{\alpha}_{ML} - \hat{\beta}_{ML}(f_t - f_f))' (r_t - r_f e - \hat{\alpha}_{ML} - \hat{\beta}_{ML}(f_t - f_f)), \]

\[ \hat{\lambda}_{ML} = \frac{1}{T} \sum_{t=1}^{T} (f_t - r_f), \quad \hat{\omega}_{ML}^2 = \frac{1}{T} \sum_{t=1}^{T} (f_t - \bar{f})^2. \]
The ML estimator is now
\[
\hat{\theta}_{ML} = \arg\max_{\theta} L(\theta), \quad \text{where}
\]
\[
L(\theta) = -\frac{1}{2} \log(\det(\mathbf{A}' + c^2 \mathbf{I}))
\]
\[
- \frac{1}{2T} \sum_{t=1}^{T} \left( \mathbf{r}_t - r_f \mathbf{e} - \mathbf{A} \mathbf{m}_t - \mathbf{a} - \beta(f_t - r_f) \right)' \left( \mathbf{A}' + c^2 \mathbf{I} \right)^{-1} \left( \mathbf{r}_t - r_f \mathbf{e} - \mathbf{A} \mathbf{m}_t - \mathbf{a} - \beta(f_t - r_f) \right)
\]
\[
- \frac{1}{2} \log(\omega) - \frac{1}{2T\omega^2} \sum_{t=1}^{T} (f_t - r_f - \lambda)^2,
\]
with \( \theta = (\mathbf{A}', \mathbf{a}', \beta', c^2, \lambda, \omega^2)'. \) However, \( \mathbf{a}_0 \) and \( \mathbf{m}_0 \) are not identified, as discussed above. One needs to impose the APT restriction, leading to the constrained estimator:
\[
\hat{\theta}_{MLC} = \arg\max_{\theta} L(\theta) \quad \text{such that} \quad \frac{\mathbf{a}' \mathbf{a}}{c^2} \leq \delta,
\]
which for the case \( L(\theta) - \kappa(\frac{\mathbf{a}' \mathbf{a}}{c^2} - \delta) \), gives:
\[
\hat{\mathbf{m}}_{MLC} = \frac{1}{\mathbf{A}'_{MLC} \mathbf{A}_{MLC}} \hat{\mathbf{A}}'_{MLC}(\mathbf{r} - r_f \mathbf{e} - \hat{\beta}_{MLC} \hat{\lambda}_{MLC}),
\]
\[
\hat{\mathbf{a}}_{MLC} = \frac{1}{(1 + \hat{\kappa}_{MLC})} (\mathbf{r} - r_f \mathbf{e} - \hat{\beta}_{MLC} \hat{\lambda}_{MLC} - \hat{\mathbf{A}}_{MLC} \hat{\mathbf{m}}_{MLC}),
\]
\[
\hat{\beta}_{MLC} = \left( (\hat{\omega}^2_{MLC} - \hat{\lambda}^2_{MLC}) \mathbf{I} + \hat{\lambda}^2_{MLC} (\frac{\hat{\kappa}_{MLC}}{1 + \hat{\kappa}_{MLC}}) ^2 (\mathbf{I} - \frac{\mathbf{A}_{MLC} \hat{\mathbf{A}}'_{MLC}}{\mathbf{A}'_{MLC} \mathbf{A}_{MLC}}) \right)^{-1}
\]
\[
\times \left( \Sigma_{rf} - \hat{\mathbf{r}} \hat{\lambda}_{MLC} + \hat{\lambda}_{MLC} \hat{\mathbf{r}} \left( \frac{\hat{\kappa}_{MLC}}{1 + \hat{\kappa}_{MLC}} \right)^2 (\mathbf{I} - \frac{\mathbf{A}_{MLC} \hat{\mathbf{A}}'_{MLC}}{\mathbf{A}'_{MLC} \mathbf{A}_{MLC}}) \right),
\]
where
\[
(1 + \hat{\kappa}_{MLC}) =
\]
\[
\left[ (\mathbf{r} - r_f \mathbf{e} - \hat{\beta}_{MLC} \hat{\lambda}_{MLC} - \hat{\mathbf{A}}_{MLC} \hat{\mathbf{m}}_{MLC})' \Sigma_{MLC}^{-1} (\mathbf{r} - r_f \mathbf{e} - \hat{\beta}_{MLC} \hat{\lambda}_{MLC} - \hat{\mathbf{A}}_{MLC} \hat{\mathbf{m}}_{MLC}) \right]^{\frac{1}{2}}
\]
\[
\frac{\delta^{\frac{1}{2}}}{},
\]
\[\text{tr}(\hat{\mathbf{A}}_{MLC} \hat{\mathbf{A}}'_{MLC} + \hat{\omega}^2_{MLC} \mathbf{I})^{-1}) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{g}_t' (\hat{\mathbf{A}}_{MLC} \hat{\mathbf{A}}'_{MLC} + \hat{\omega}^2_{MLC} \mathbf{I})^{-1} (\hat{\mathbf{A}}_{MLC} \hat{\mathbf{A}}'_{MLC} + \hat{\omega}^2_{MLC} \mathbf{I})^{-1} \mathbf{g}_t,
\]
where \( \mathbf{g}_t = (\mathbf{r}_t - r_f \mathbf{e} - \hat{\mathbf{A}}_{MLC} \hat{\mathbf{m}}_{MLC} - \hat{\mathbf{a}}_{MLC} - \hat{\beta}_{MLC} (f_t - r_f)). \)
with the APT restriction coincide and are equal to the sample variance and sample mean of $f_t - r_f$, respectively: $\tilde{\lambda}_{MLC} = \frac{1}{T} \sum_{t=1}^{T} (f_t - r_f)$ and $\tilde{\omega}^2_{MLC} = \frac{1}{T} \sum_{t=1}^{T} (f_t - \bar{f})^2$. 
Table 1: Out-of-Sample Sharpe Ratios With Alpha Misspecification

This table reports the annualized Sharpe ratios for five portfolios: (1) the equal-weighted (EW) portfolio; (2) the global-minimum-variance (GMV) portfolio, based on the sample covariance matrix; (3) the mean-variance (MV) portfolio based on the sample mean and sample covariance matrix; (4) the maximum-likelihood mean-variance unconstrained (MLU) portfolio based on the sample mean but factor covariance matrix implied by the factor model, \( \beta \omega^2 + \sigma^2 I \); and (5) the constrained maximum-likelihood mean-variance (MLC) portfolio based on the sample mean but factor covariance matrix of \( \beta \omega^2 + \Sigma \) and imposing the APT constraint in which \( \delta \) is obtained using ten-fold cross validation. The \( t \)-statistic is for the difference in the Sharpe ratio of the MLC portfolio with the APT restriction and the equally weighted (EW) portfolio, and the last two columns give the ratio of the Sharpe ratios for the MLC and EW portfolios and for the MLC and MLU portfolios.

We assume that the observed factor \( f_t \) is IID and has Gaussian distribution. For the “base case” of our simulation exercise, we assume that the observed factor has a monthly mean equal to \( \lambda = \frac{8}{12 \times 100} \) and monthly volatility equal to \( \omega = \frac{16}{\sqrt{12 \times 100}} \). The pricing error \( \alpha \) is from an IID multivariate Gaussian distribution with mean 0 and covariance matrix equal to \( \frac{5}{\sqrt{12 \times 100}} I_N \). We consider two environments, one with bounded residual variation (\( \alpha = a \)) in Panel A and the other with a special form of unbounded residual variation (\( \alpha = Am \)) in Panel B. In both cases, \( \epsilon_t \) is IID with a multivariate Gaussian distribution with monthly mean of 0. In the bounded-variation case, the monthly covariance matrix is \( \Sigma = \sigma^2 I_N \) with \( \sigma = \frac{20}{\sqrt{12 \times 100}} \). In the unbounded-variation case, the monthly covariance matrix is \( \Sigma = \frac{1}{sr^2 \alpha} \alpha \alpha' + \sigma^2 I_N \), with \( sr^2 \alpha = \frac{1}{\sqrt{12}} \) and \( \sigma = \frac{20}{\sqrt{12 \times 100}} \).

<table>
<thead>
<tr>
<th></th>
<th>EW</th>
<th>GMV</th>
<th>MV</th>
<th>MLU</th>
<th>MLC</th>
<th>t-stat</th>
<th>MLC/W</th>
<th>MLC/MLU</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base case</td>
<td>0.55</td>
<td>0.20</td>
<td>0.03</td>
<td>1.06</td>
<td>1.35</td>
<td>8.61</td>
<td>2.47</td>
<td>1.27</td>
</tr>
<tr>
<td>Low ( \lambda ) (1/2 of Base)</td>
<td>0.32</td>
<td>0.20</td>
<td>-0.03</td>
<td>0.57</td>
<td>0.84</td>
<td>5.04</td>
<td>2.65</td>
<td>1.48</td>
</tr>
<tr>
<td>Low ( \sigma_\epsilon ) (1/2 of Base)</td>
<td>0.55</td>
<td>0.24</td>
<td>0.09</td>
<td>2.57</td>
<td>3.58</td>
<td>17.16</td>
<td>6.50</td>
<td>1.39</td>
</tr>
<tr>
<td>Low ( \sigma_\alpha ) (1/3 of Base)</td>
<td>0.52</td>
<td>0.15</td>
<td>0.04</td>
<td>0.19</td>
<td>0.31</td>
<td>-5.03</td>
<td>0.60</td>
<td>1.67</td>
</tr>
<tr>
<td><strong>Panel B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base case</td>
<td>0.54</td>
<td>0.11</td>
<td>-0.01</td>
<td>0.47</td>
<td>0.81</td>
<td>5.89</td>
<td>1.50</td>
<td>1.72</td>
</tr>
<tr>
<td>Low ( \lambda ) (1/2 of Base)</td>
<td>0.29</td>
<td>0.08</td>
<td>0.02</td>
<td>0.25</td>
<td>0.39</td>
<td>2.23</td>
<td>1.33</td>
<td>1.59</td>
</tr>
<tr>
<td>Low ( \sigma_\epsilon ) (1/2 of Base)</td>
<td>0.55</td>
<td>0.07</td>
<td>0.01</td>
<td>0.38</td>
<td>0.89</td>
<td>7.83</td>
<td>1.64</td>
<td>2.35</td>
</tr>
<tr>
<td>Low ( \sigma_\alpha ) (1/3 of Base)</td>
<td>0.47</td>
<td>0.08</td>
<td>-0.05</td>
<td>0.21</td>
<td>0.46</td>
<td>-4.79</td>
<td>0.97</td>
<td>2.21</td>
</tr>
</tbody>
</table>
Figure 1: Mean-variance and minimum-variance frontier portfolios
In this figure, we plot the three kinds of portfolios we wish to study: (1) the global minimum-variance portfolio, \( w_{\text{gmv}} \); (2) the mean-variance efficient portfolio when a risk-free asset is available, \( w_{\text{mv}} \); (3) the mean-variance efficient portfolios in the absence of a risk-free asset, \( w_{\text{fp}} \). The figure also shows the tangency portfolio, \( w_{\text{tan}} \), which is a special case of the mean-variance portfolios where all the wealth is invested in risky assets, with nothing invested in the risk-free asset.
Figure 2: Mean-variance and minimum-variance portfolios as $N$ increases
In this figure, we plot the mean-variance and minimum-variance portfolios as we increase the number of assets, $N$. We plot three cases: $N = \{10, 50, 3000\}$. 

![Diagram showing the relationship between expected return and volatility for different values of N.](image)
Figure 3: Global-minimum-variance portfolio and its decomposition
In this figure, we plot the global-minimum-variance portfolio, $w_{\text{mv}}$, and its decomposition into two portfolios, one that depends only on the pricing errors, $\Sigma$, and another that depends only on the factor exposure and their premia, $w^\Omega$. 

![Graph showing the global-minimum-variance portfolio and its decomposition](image-url)
Figure 4: Mean-variance portfolio and its decomposition
In this figure, we plot the mean-variance portfolio in the presence of risk-free asset, $w^{\text{inv}}$, and its decomposition into two portfolios, one that depends only on the pricing errors, $w^{\alpha}$, and another that depends only on the factor exposure and their premia, $w^{\beta}$.
References


