

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

Raman Uppal

Paolo Zaffaroni

January 8, 2017

Abstract

Our objective is to investigate the effect of model misspecification on mean-variance portfolios and to show how asset-pricing theory and asymptotic analysis (for large number of assets) can be used to mitigate misspecification. The starting point of our analysis is the Arbitrage Pricing Theory (APT), which allows for pricing errors. We 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. Our key insight is that, instead of treating misspecification directly in the mean-variance portfolio, it is better to first decompose the portfolio into components that correspond to the two components of returns in the APT: a “beta” portfolio that depends on factor risk premia and an “alpha” portfolio that depends only on pricing errors. We show that even though both the beta and alpha portfolios are inefficient, they can be used to generate the entire mean-variance efficient portfolio frontier. For the beta portfolio, we treat misspecification using asymptotic analysis: as the number of assets increases, we show that the weights of the alpha portfolio dominate those of the beta portfolio, leading to an expression for mean-variance portfolio weights that is immune to beta misspecification. 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. Finally, we demonstrate that our approach achieves an out-of-sample Sharpe ratio that is more than double that of the equally weighted portfolio.

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 EDHEC Business School and CEPR; Email: Raman.Uppal@edhec.edu. Zaffaroni is affiliated with Imperial College Business School; Email: P.Zaffaroni@imperial.ac.uk. We thank the Mathematics Department at Imperial College London for allowing us to use their computer cluster. We gratefully acknowledge comments from Dante Amengual, Turan Bali, Harjoat Bhamra, John Cochrane, Rama Cont, Victor DeMiguel, Serge Darolles, Rene Garcia, Lorenzo Garlappi, Christian Gourieéroux, Amit Goyal, Christian Julliard, Andrew Karolyi, Ralph Koijen, Michelle Lee, Xioaji Lin, Lionel Martellini, Vasant Naik, Stefan Nagel, Ľuboš Pástor, Hashem Pesaran, Cesare Robotti, Lucio Sarno, Jay Shanken, Yazid Sharaiha, Eric Shirbini, Claudio Tebaldi, Nikolaos Tassaromatis, Allan Timmermann, Motohiro Yogo, Guofu Zhou, and seminar participants at: Imperial College Business School, Judge Business School, Norges Bank Investment Management, Financial Econometrics Conference at Toulouse School of Economics, NBER Summer Institute Workshop on Forecasting & Empirical Methods, UBC Summer Finance Conference, Advances in the Analysis of Hedge Fund Strategies Conference at Imperial College London, 8th Annual Hedge Fund Conference, and CEPR First Annual Spring Symposium in Financial Economics.

1 Introduction

In this paper, our objective is to study the effect of model misspecification on mean-variance portfolios and to show how asset-pricing theory and asymptotic analysis (as the number of assets increases) 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 *pricing error*, often referred to as alpha.¹ 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) of Ross (1976). The APT is particularly well-suited for our purpose because it allows for the possibility of model misspecification, and hence, mispricing (alpha), 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.²

Our work makes three contributions. Our first contribution is to extend the notion of pricing errors in the APT and to show that the APT applies much more broadly than typically assumed. Traditionally, the APT is interpreted as applying to pricing errors that are small and unrelated to factors (see, for example, Cochrane (2005, Ch. 9.4)), implying that the covariance matrix of residuals has bounded eigenvalues. We show that the APT

¹As 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).

²For 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).

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 may 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 interpretation of the APT can be used to capture, and then 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 depends only on pricing errors with zero exposure to common risk, and a “beta” portfolio, which depends on factor risk premia and their loadings.³ We also show that the alpha and beta portfolios, both of which are *inefficient*, span the entire mean-variance efficient frontier, thus leading to two-fund separation.⁴

In the second step, we treat misspecification in the beta portfolio using asymptotic analysis as the number of assets increases.⁵ In the environment with an asymptotically large number of assets, we show that the weights of the alpha portfolio typically dominate the weights of the beta portfolio. Given the secondary role played by the beta portfolio, we show that, under a set of mild conditions, it can be replaced, without any loss of efficiency, by a class of benchmark portfolios that by construction are independent of the mean vector and covariance matrix of the observed factors, and hence, immune to beta misspecification. This asymptotic analysis of portfolio weights also shows the dominant role played by the

³This decomposition is 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.

⁴Even though the alpha and beta portfolios themselves are not on the Markowitz efficient frontier, 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.

⁵The asymptotic analysis for a large number of assets is not just an abstract mathematical device, 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. Our asymptotic results bite even when the number of assets is as small as 100.

pricing errors (alphas) 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 restriction on the pricing errors when estimating the model, which reduces the error in the estimated 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. In the case where we have both errors that are unrelated and those that are related to factors, 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. We also show that the APT restriction on alphas is equivalent to imposing a bound on the square of the Sharpe ratio of the alpha portfolio, which turns out to be identical 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.⁶

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. Using simulations, we 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 can be twice as large as those of the equally weighted portfolio.

In summary, traditional models of portfolio choice either have no alpha or do not distinguish between the alpha and beta components of returns; and hence, when correcting for misspecification they shrink only the beta component of returns. Our model, founded on the APT, distinguishes between the alpha and beta components of returns. Our portfolio strategy exploits alpha to improve portfolio performance, while it shrinks completely

⁶The bound on the square of the Sharpe ratio of the alpha portfolio can also 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).

the beta components of returns, so that the beta portfolio does not depend at all on the distribution of the factor returns.

The rest of the paper is organized as follows. In Section 2, we discuss the literature related to our work. Our three main 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 the APT to the case of unbounded eigenvalues for the residual covariance matrix. 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 Markowitz efficient-frontier portfolios, are collected in the Online Appendix.

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 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 mildly correlated across assets. 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.⁷ 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 residual covariance matrix is 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 pervasive factors.

The reason that Chamberlain (1983) and Chamberlain and Rothschild (1983) do not consider the case where the maximum eigenvalue of the covariance matrix is 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 for the residuals left over after extracting the principal components has necessarily bounded maximum eigenvalue. However, in contrast to Chamberlain (1983) and Chamberlain and Rothschild (1983), 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.⁸ 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 as 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,

⁷See Reisman (1988, 1992a,b) and Shanken (1992) for further extensions of the APT.

⁸For work showing how beliefs about an asset-pricing model can influence portfolio choice, see Pástor (2000) and Pástor and Stambaugh (2000).

should one invest in only a diversified “passive” portfolio 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,⁹ 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. 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 residual covariance matrix to be diagonal, 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 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 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.¹⁰ We complement this result by showing that, in the presence of pricing errors, the long-short alpha portfolio weights dominate the 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

⁹See, 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. For surveys of this literature, see Kothari and Shanken (2002), Aragon and Ferson (2006), and Ferson (2013).

¹⁰Green (1986b) provides necessary and sufficient conditions for portfolios on the minimum-variance frontier to have positive weights in all assets.

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 optimization problem, a robust regression problem, and a Bayesian problem.

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 identity matrix, using the identity matrix as the covariance matrix, which is implicitly 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 naturally from the APT.

As highlighted by the literature described above, an investor choosing optimal portfolio weights faces both model and parameter uncertainty.¹¹ Two features of the APT allow us

¹¹Model 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

to deal with both sources of uncertainty. The first feature is *allowing for the presence of the pricing error, α* . The pricing error in the model tells us how the incorrectly-specified mean and variance of returns are to be adjusted in the presence of misspecification.

The second feature of the APT is the *restriction on the magnitude of the pricing error: $\alpha' \Sigma^{-1} \alpha \leq \delta < \infty$* , where α is the vector of pricing errors, Σ^{-1} is the inverse of the covariance matrix of residuals, and δ 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.¹² This restriction limits the magnitude of estimates for α , and hence, allows one to deal also with parameter uncertainty associated with α .¹³ Interestingly, imposing the APT restriction is analogous to the approach adopted in Garlappi, Uppal, and Wang (2007), where investors are assumed to be averse to their uncertainty (ambiguity) about the model, and one accounts for parameter uncertainty using the minmax approach originally proposed in Gilboa and Schmeidler (1989).¹⁴ 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 estimating the parameters subject to the constraint that the maximum-likelihood (ML) 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 its Sharpe ratio. Thus, the key difference between our approach and that of the minmax approach is that the minmax approach constrains total expected returns, while we constrain only the alpha component

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.

¹²Adopting 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. Hansen and Sargent (2008) also provide a comprehensive discussion of how one can analyze decision makers who regard their model as an approximation and who desire decision rules that work over a set of models in the neighborhood of the approximating model.

¹³The 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.

¹⁴There is an extensive literature modeling decision making in the presence of ambiguity: see, for example, Chen and Epstein (2002), Ghirardato, Maccheroni, and Marinacci (2004), Klibanoff, Marinacci, and Mukerji (2005), Ju and Miao (2012), and Maccheroni, Marinacci, and Ruffino (2013). For surveys of this literature, see Hansen and Sargent (2008) and Epstein and Schneider (2010).

of expected returns and account for misspecification in the beta component of returns using asymptotic analysis.¹⁵

3 Generalizing the APT

In this section, we start by describing our notation and assumptions. Our analysis is founded on precisely the same assumptions as the ones underlying the APT. After stating the main result of the APT, we show how it can be extended to incorporate 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. We conclude this section by describing how the APT model can capture misspecification arising from the beta and alpha components of returns.

3.1 Notation

The number of risky assets is denoted by N . Just like in Chamberlain and Rothschild (1983) and Ingersoll (1984), we study a market with an infinite number of assets. To make clear the dependence on the number of assets, we index quantities that are N -dimensional by the subscript N , except for random variables, such as the returns on risky assets, which have the subscript t . 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. Therefore, in the N th step, as a new asset is added to the first $N - 1$ assets, the parameters of the first $N - 1$ stay unchanged. These unchanging parameters can be interpreted as the parameters one would get in the limit as the number of assets becomes asymptotically large.

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

Given an arbitrary portfolio strategy a with weights $\mathbf{w}_N^a = (w_1^a, w_2^a, \dots, w_N^a)'$ of N risky assets, and using $\mathbf{1}_N$ to denote an N -dimensional vector of ones, we define the associated

¹⁵Garlappi, 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).

portfolio return as

$$r_t^a = \mathbf{r}_t' \mathbf{w}_N^a + r_f(1 - \mathbf{1}'_N \mathbf{w}_N^a),$$

with finite conditional mean, standard deviation, and Sharpe ratio defined by

$$\begin{aligned} \mu^a &= E(r_t^a) = E(\mathbf{r}_t)' \mathbf{w}_N^a + r_f(1 - \mathbf{1}'_N \mathbf{w}_N^a), \\ \sigma^a &= \sqrt{\text{var}(r_t^a)}, \\ \text{SR}^a &= \frac{\mu^a - r_f}{\sigma^a}. \end{aligned}$$

3.2 Linear factor model for asset returns with misspecification

We start our analysis with the assumption of a linear latent-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, Σ_N , 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 Σ_N are restricted to be bounded when N is large.

Assumption 3.1 (Linear factor model). *We assume the N -dimensional vector \mathbf{r}_t of asset returns can be characterized by:*

$$\mathbf{r}_t = \boldsymbol{\mu}_N + \mathbf{B}_N \mathbf{z}_t + \boldsymbol{\varepsilon}_t, \tag{1}$$

where $\mathbf{z}_t = (z_{1t}, z_{2t}, \dots, z_{Kt})'$ is the $K \times 1$ vector of common unobserved factors, assumed to be with zero mean, without loss of generality, as explained below; $\mathbf{B}_N = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_N)'$ is an $N \times K$ full-rank matrix of factor loadings with i th row $\boldsymbol{\beta}'_i$; $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Nt})'$ is an $N \times 1$ vector of residuals, and the $N \times 1$ vector $\boldsymbol{\mu}_N$ represents the mean of the vector of returns, \mathbf{r}_t .¹⁶ At any time t , \mathbf{z}_t is distributed with zero mean and $K \times K$ covariance matrix $\boldsymbol{\Omega}$ and $\boldsymbol{\varepsilon}_t$ is distributed with zero mean and $N \times N$ covariance matrix Σ_N , with $\boldsymbol{\Omega}$ and Σ_N being positive definite. Moreover, $\boldsymbol{\varepsilon}_t$ and \mathbf{z}_t are uncorrelated; that is, $E(\boldsymbol{\varepsilon}_t \mathbf{z}'_t) = \mathbf{0}$.

It is important to note that the APT is a model about the random component of returns, $\mathbf{r}_t - \boldsymbol{\mu}_N = \mathbf{B}_N \mathbf{z}_t + \boldsymbol{\varepsilon}_t$, and is silent about expected returns. Black (1995, p. 168)

¹⁶Throughout the paper, it will be assumed that $K < N$.

recognizes this explicitly and states that the ‘‘Arbitrage Pricing Theory (APT) is a model of variance. It says that the number of independent factors influencing return is limited, but it is silent on the pricing of these factors, so it is silent on expected return.’’ For instance, Assumption 3.1 implies that the conditional variance-covariance matrix for asset returns is:

$$E[(\mathbf{r}_t - \boldsymbol{\mu}_N)(\mathbf{r}_t - \boldsymbol{\mu}_N)'] = \mathbf{V}_N = \mathbf{B}_N \boldsymbol{\Omega} \mathbf{B}_N' + \boldsymbol{\Sigma}_N, \quad (2)$$

regardless of the vector of expected returns, $\boldsymbol{\mu}_N$.

3.3 Arbitrage Pricing Theory (APT)

We now describe the APT result in the literature. In the definition below, as well as throughout the paper, we use δ 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' :*

$$\text{var}(\mathbf{r}'_t \mathbf{w}_{N'}^a) \rightarrow 0 \text{ as } N' \rightarrow \infty \text{ and } (\boldsymbol{\mu}_{N'} - r_f \mathbf{1}_{N'})' \mathbf{w}_{N'}^a \geq \delta > 0 \text{ for all } N'.$$

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

$$\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^{-1} \check{\boldsymbol{\alpha}}_N \leq \delta < \infty, \quad (3)$$

where the vector of pricing errors is

$$\check{\boldsymbol{\alpha}}_N = (\boldsymbol{\mu}_N - r_f \mathbf{1}_N) - \mathbf{B}_N \check{\boldsymbol{\lambda}}, \quad (4)$$

and the vector of risk premia is

$$\check{\boldsymbol{\lambda}} = (\mathbf{B}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}_N' \boldsymbol{\Sigma}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N), \quad (5)$$

with the $\check{\cdot}$ symbol is used to denote that the quantity is obtained from a projection, and hence, every component of it will change with N .¹⁷

¹⁷Note that $\check{\boldsymbol{\alpha}}_N$ has the subscript N because it is N dimensional, while $\check{\boldsymbol{\lambda}}$ is not subscripted by N because it not N dimensional.

Remark 3.1.1. Observe that $\check{\lambda}$ is the projection coefficient when projecting $(\boldsymbol{\mu}_N - r_f \mathbf{1}_N)$ on \mathbf{B}_N and $\check{\boldsymbol{\alpha}}_N$ is the projection residual (pricing error) that satisfies

$$\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \check{\boldsymbol{\alpha}}_N = \mathbf{0}. \quad (6)$$

3.4 Extending the APT

Above, we described the standard APT. In this section, we explain our first contribution, which is to extend the standard 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.

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

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 $g_{1K}((\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1}) \rightarrow 0$ as $N \rightarrow \infty$. If the limit is positive instead of zero, then the factor representation is irregular. Equivalently, in a regular economy, the minimum eigenvalue $g_{KK}(\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)$ is diverging as $N \rightarrow \infty$.*

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 $\check{\lambda}$ as $N \rightarrow \infty$, which we label $\boldsymbol{\lambda}$, and the associated vector of pricing errors, $\boldsymbol{\alpha}_N = ((\boldsymbol{\mu}_N - r_f \mathbf{1}_N) - \mathbf{B}_N \boldsymbol{\lambda})$.¹⁸ Under Assumptions 3.1 and 3.2 and the assumption of a regular factor economy, Ingersoll (1984, Theorem 3 and ftn. 10) shows that $\boldsymbol{\lambda}$ is unique and prices assets with bounded squared error:

$$\boldsymbol{\alpha}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N \leq \delta < \infty, \quad (7)$$

¹⁸Note that $\boldsymbol{\alpha}_N$ is an N -dimensional vector that is a component of the infinite-dimensional vector $\boldsymbol{\alpha}$; as N increases, the number of elements in $\boldsymbol{\alpha}_N$ increases, but the elements themselves do not change.

where δ represents some positive arbitrary scalar. Observe that the APT restriction in (3) is expressed in terms of $\check{\alpha}_N$, while the restriction in (7) is expressed in terms of α_N ; in Lemma A.4, we show the equivalence between these two conditions.

Observe that, as discussed in Ingersoll (1984, footnote 10), in (7) the weighting of the squared pricing error uses Σ_N^{-1} . The intuition for why the squared-error bound in (7) is weighted by (the inverse of) the matrix of residual variances, Σ_N , is that an asset’s usefulness in arbitrage trades 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 restriction in (7), which is a consequence of asymptotic no arbitrage, links the pricing error α_N to the residual covariance matrix, Σ_N . There are two possible, complementary, cases for Σ_N as $N \rightarrow \infty$ that we will examine. In the first case, for large N all the eigenvalues of Σ_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 Σ_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 α_N by no arbitrage for case with bounded eigenvalues; Huberman (1982); Ingersoll (1984)). *If Σ_N has bounded eigenvalues for large N , then the restriction in (7) requires the elements of the pricing-error vector α_N to become small for large N in the following sense:*

$$\alpha'_N \alpha_N \leq g_{1N}(\Sigma_N)(\alpha'_N \Sigma_N^{-1} \alpha_N) \leq \delta < \infty,$$

but without α_N being tied down to Σ_N .

Remark 3.2.1. Notice that in the original formulation of the APT in (1), no factor is assumed to be observed; instead, the error term $\mathbf{r}_t - \boldsymbol{\mu}_N = \mathbf{B}_N \mathbf{z}_t + \boldsymbol{\varepsilon}_t$ has a latent factor

structure through the common innovation \mathbf{z}_t . In contrast, nowadays factor models are typically used assuming the existence of a given number of traded observed factors. This means that rather than the model in (1), one considers

$$\mathbf{r}_t - r_f \mathbf{1}_N = \boldsymbol{\alpha}_N + \mathbf{B}_N(\mathbf{f}_t - r_f \mathbf{1}_K) + \boldsymbol{\varepsilon}_t, \quad (8)$$

for the K -dimensional factors \mathbf{f}_t , all of which are observed. Notice that (1) and (8) are equivalent if $\mathbf{f}_t - r_f \mathbf{1}_K = \boldsymbol{\lambda} + \mathbf{z}_t$ and $\boldsymbol{\mu}_N - r_f \mathbf{1}_N = \boldsymbol{\alpha}_N + \mathbf{B}_N \boldsymbol{\lambda}$.

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 $\boldsymbol{\Sigma}_N$ are *not* bounded. This typically occurs when there are missing or mismeasured factors, as explained below in Section 3.5.

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

$$\boldsymbol{\alpha}_N = \mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{a}_N, \quad (9)$$

where \mathbf{A}_N is a $N \times p$ matrix whose j th column equals $g_{jN}^{\frac{1}{2}}(\boldsymbol{\Sigma}_N) \mathbf{v}_{jN}(\boldsymbol{\Sigma}_N)$, where $1 \leq j \leq p$, $\mathbf{v}_{jN}(\boldsymbol{\Sigma}_N)$ is the eigenvector of $\boldsymbol{\Sigma}_N$ associated with the eigenvalue $g_{jN}(\boldsymbol{\Sigma}_N)$, $\boldsymbol{\lambda}_{miss}$ is some $p \times 1$ vector, and \mathbf{a}_N is some non-zero $N \times 1$ vector that satisfies

$$\mathbf{a}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{a}_N \leq \delta < \infty.$$

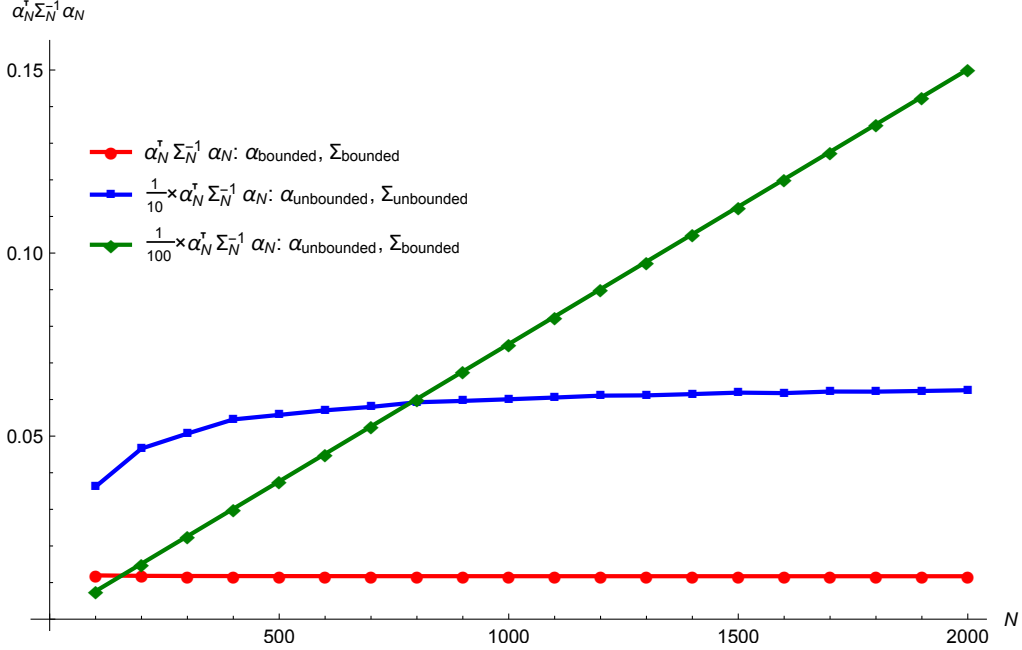
Remark 3.3.1. One can interpret the two components of $\boldsymbol{\alpha}_N$ in (9) in a variety of ways. The first term, $\mathbf{A}_N \boldsymbol{\lambda}_{miss}$, could be associated with p latent or missing pervasive factors, where \mathbf{A}_N are the factor loadings and $\boldsymbol{\lambda}_{miss}$ are the risk premia for the missing factors.¹⁹ The second term, \mathbf{a}_N , is the idiosyncratic part of the pricing error $\boldsymbol{\alpha}_N$; for instance, \mathbf{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:

$$E(\mathbf{r}_t - r_f \mathbf{1}_N) = \boldsymbol{\mu}_N - r_f \mathbf{1}_N = \boldsymbol{\alpha}_N + \mathbf{B}_N \boldsymbol{\lambda} = (\mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{a}_N) + \mathbf{B}_N \boldsymbol{\lambda}. \quad (10)$$

¹⁹Pervasiveness of the latent missing factors with loadings \mathbf{A}_N follows from the assumption that g_{jN} for $1 \leq j \leq p$ are diverging for large N .

Figure 1: Weighted sum of the squared pricing errors

In this figure, we plot the weighted sum of the squared pricing errors, $\alpha_N' \Sigma_N^{-1} \alpha_N$, as the number of assets N increases for three different cases. In the first case, which is the one studied in the existing literature, the elements of the pricing-error vector, α_N , become small as N increases and the eigenvalues of Σ_N are bounded; in this case, the weighted sum of the squared pricing errors is bounded, as shown by the red line with circle markers. In the second case, we allow for large pricing errors that are related to factors. In this case, some of the eigenvalues of Σ_N are *unbounded*. As the blue line with square markers shows, even in this case the weighted sum of the squared pricing errors is bounded, demonstrating that the pricing errors can be large without violating the APT restriction. The third case, illustrated by the green line with diamond markers, shows that if the pricing errors were large but the eigenvalues of Σ_N were *incorrectly* restricted to be bounded, only then would the APT restriction be violated.



Observe that, compared to equation (8), equation (10) contains the decomposition of the extra term $\alpha_N = (\mathbf{A}_N \lambda_{miss} + \mathbf{a}_N)$, which represents the effect of misspecification of the traded observed factors.

Remark 3.3.2. More importantly, Theorem 3.3 shows that the common perception that the pricing error α_N needs 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 APT restriction given in (7); this is illustrated in Figure 1. What Theorem 3.3 states is that if the maximum eigenvalue of Σ_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, Σ_N , has the following approximate p -factor structure:

$$\Sigma_N = \mathbf{A}_N \mathbf{A}'_N + \mathbf{C}_N,$$

where \mathbf{C}_N is a $N \times N$ positive semi-definite matrix with bounded eigenvalues.²⁰ Therefore, the residual variance of the return on any portfolio weights \mathbf{w}_N is given by

$$\mathbf{w}'_N \Sigma_N \mathbf{w}_N = \mathbf{w}'_N \mathbf{A}_N \mathbf{A}'_N \mathbf{w}_N + \mathbf{w}'_N \mathbf{C}_N \mathbf{w}_N.$$

Whereas the second term, $\mathbf{w}'_N \mathbf{C}_N \mathbf{w}_N$, goes to zero for well-spread portfolios (that is, for portfolios with $\mathbf{w}'_N \mathbf{w}_N \rightarrow 0$), there is no guarantee that the same occurs for the first term, $\mathbf{w}'_N \mathbf{A}_N \mathbf{A}'_N \mathbf{w}_N$.

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*.

3.5 Different forms of model misspecification

In this section, we study the different forms of model misspecification of the true model. Throughout this section, we assume that the true model satisfies

$$\mathbf{r}_t - r_f \mathbf{1}_N = \mathbf{a}_N^* + \mathbf{B}_N^* (\mathbf{f}_t^* - r_f \mathbf{1}_{K^*}) + \boldsymbol{\varepsilon}_t, \quad (11)$$

for K^* factors, of which $0 \leq K \leq K^*$ are observed. In the above equation, \mathbf{f}_t^* denotes all the common factors affecting returns of which \mathbf{f}_t are the factors that are observed, \mathbf{B}_N^* are the true loadings on these factors, and \mathbf{a}_N^* are the true firm-specific components of expected returns that satisfy the APT restriction. The true mean and covariance matrix of \mathbf{f}_t^* are denoted by $\boldsymbol{\lambda}^*$ and $\boldsymbol{\Omega}^*$, respectively. We describe four forms of model misspecification that are captured by the framework we have described in the paper. The first is related to the “beta” component of returns. The second, third, and fourth are related to the “alpha” component of returns, arising from the presence of: a pricing error that is unrelated to factors (that is, $\mathbf{a}_N^* \neq \mathbf{0}$), missing factors (that is, for $K < K^*$), and mismeasured factors. For expositional ease, we will assume that $\mathbf{a}_N^* = \mathbf{0}$ in Cases 1, 3, and 4 below.

²⁰Under the assumptions of Chamberlain and Rothschild (1983, Theorem 4), \mathbf{C}_N will be non-singular for a sufficiently large N , which implies that Σ_N is also nonsingular for a sufficiently large N .

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 in this case is to study misspecification in the “beta” component of returns; that is, misspecification in the assumed distribution of the factors, namely $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}^*$ or $\boldsymbol{\Omega} \neq \boldsymbol{\Omega}^*$ because investors have incorrect views.²¹ 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 consistent estimates of \mathbf{B}_N^* and the residual covariance matrix, $\boldsymbol{\Sigma}_N^*$.

Case 2: Pure pricing errors (unrelated to factors)

Next, consider the case where all factors are observed without error, but expected returns depend also on non-factor-related characteristics, given by $\mathbf{a}_N^* \neq \mathbf{0}$ in (11). Thus, in this case misspecification will arise if the investor erroneously sets $\mathbf{a}_N^* = \mathbf{0}$. In contrast to the previous case, in this setting \mathbf{a}_N^* does not influence the variance-covariance matrix of returns.

Case 3: Missing factors

Suppose now that of the K^* factors, only K are observed and $p = K^* - K > 0$ are missing and suppose for simplicity that the observed and missing factors are uncorrelated. For simplicity, we assume $\mathbf{a}_N^* = \mathbf{0}$. Then, the model in (11) can be re-written as follows:

$$\mathbf{r}_t - r_f \mathbf{1}_N = \mathbf{A}_N (\mathbf{f}_{miss,t} - r_f \mathbf{1}_p) + \mathbf{B}_N (\mathbf{f}_t - r_f \mathbf{1}_K) + \boldsymbol{\varepsilon}_t,$$

where $\mathbf{f}_{miss,t}$ are the missing factors,

$$\mathbf{f}_t^* = \begin{pmatrix} \mathbf{f}_{miss,t} \\ \mathbf{f}_t \end{pmatrix} \quad \text{and} \quad \mathbf{B}_N^* = \begin{pmatrix} \mathbf{A}_N & \mathbf{B}_N \end{pmatrix}.$$

²¹For example, one could imagine a world in which the true risk premia are conditionally time varying but the investor models them as constant through time.

Rewriting $\mathbf{f}_{miss,t} - r_f \mathbf{1}_p = \boldsymbol{\lambda}_{miss} + \mathbf{z}_{miss,t}$, it follows that

$$\mathbf{r}_t - r_f \mathbf{1}_N = \mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{B}_N (\mathbf{f}_t - r_f \mathbf{1}_K) + (\boldsymbol{\varepsilon}_t + \mathbf{A}_N \mathbf{z}_{miss,t}),$$

where $E[\mathbf{z}_{miss,t}] = \mathbf{0}$ and $E[\mathbf{z}_{miss,t} \mathbf{z}'_{miss,t}] = \mathbf{I}_p$ to achieve identification.²²

Case 4: Mismeasured factors

Finally, consider the case where all K^* factors are measured with error. In particular, the observed factors satisfy $\mathbf{f}_t = \mathbf{f}_t^* + \boldsymbol{\eta}_t$, where the measurement error $\boldsymbol{\eta}_t$ has mean $E[\boldsymbol{\eta}_t] = \boldsymbol{\mu}_\eta$ and covariance matrix $E[(\boldsymbol{\eta}_t - \boldsymbol{\mu}_\eta)(\boldsymbol{\eta}_t - \boldsymbol{\mu}_\eta)'] = \boldsymbol{\Sigma}_\eta$.²³ Recall that for simplicity we assume $\mathbf{a}_N^* = \mathbf{0}$. Then, (11) can be re-written as follows:

$$\mathbf{r}_t - r_f \mathbf{1}_N = -\mathbf{B}_N \boldsymbol{\mu}_\eta + \mathbf{B}_N (\mathbf{f}_t - r_f \mathbf{1}_K) + (\boldsymbol{\varepsilon}_t - \mathbf{B}_N (\boldsymbol{\eta}_t - \boldsymbol{\mu}_\eta)). \quad (12)$$

The econometric problem in estimating (12) is akin to an errors-in-variables problem: the residual is correlated with the observed factors through the measurement error, $\boldsymbol{\eta}_t$. For econometric identification one typically sets $\boldsymbol{\mu}_\eta = \mathbf{0}$. Observe also that there is a link between $\boldsymbol{\alpha}_N$ and \mathbf{B}_N (in fact, \mathbf{B}_N appears in three terms of the model in (12)); this restriction can be used to test for the presence of mismeasurement and can be exploited also in the estimation.

4 Mitigating Model Misspecification

We study the weights and returns for the family of mean-variance portfolios. In particular, we study: (1) the mean-variance efficient portfolio when a risk-free asset is available, \mathbf{w}_N^{mv} ; (2) the global minimum-variance portfolio when a risk-free asset is not available, $\mathbf{w}_N^{\text{gmV}}$; and (3) the mean-variance efficient portfolios in the absence of a risk-free asset, \mathbf{w}_N^{fp} , which are the Markowitz frontier portfolios that have the smallest variance for a given target mean.

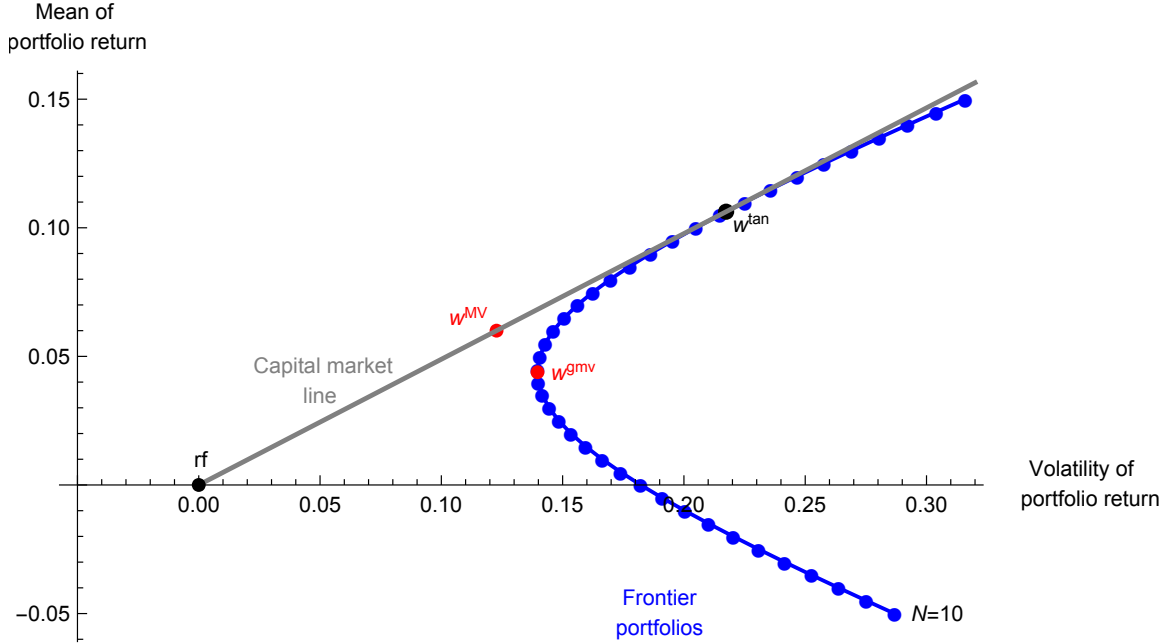
These three portfolio are displayed in Figure 2.

²²There are two possible cases for the set of missing factors, $\mathbf{f}_{miss,t}$. Some of them, for instance $\mathbf{f}_{miss_1,t}$ could possibly be pervasive while others, $\mathbf{f}_{miss_2,t}$, are non-pervasive, where in turn we can split the columns of $\mathbf{A}_N = (\mathbf{A}_{N,1}, \mathbf{A}_{N,2})$ and analogously for $\boldsymbol{\lambda}_{miss}$ and $\mathbf{z}_{miss,t}$. In particular, using our definition of regularity, one then obtains that as $N \rightarrow \infty$ that $g_{1p_1} (\mathbf{A}'_{N,1} (E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t])^{-1} \mathbf{A}_{N,1}) \rightarrow \infty$ and $g_{1p_2} (\mathbf{A}'_{N,2} (E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t])^{-1} \mathbf{A}_{N,2}) \leq \delta < \infty$, where $E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t]$ has bounded maximum eigenvalue and $p_1 + p_2 = p$.

²³Of course, it is possible that some of the factors are measured without any error; in that case the means and variances of the $\boldsymbol{\eta}_t$ associated with these factors are zero.

Figure 2: Mean-variance and minimum-variance frontier portfolios

In this figure, we plot three kinds of mean-variance portfolios one can study: (1) the global minimum-variance portfolio, \mathbf{w}^{gmV} ; (2) the mean-variance efficient portfolio when a risk-free asset is available, \mathbf{w}^{mV} , which lies on the Capital Market Line; and, (3) the mean-variance efficient frontier portfolios in the absence of a risk-free asset. The figure also shows the tangency portfolio, \mathbf{w}^{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.



For each of these three portfolios, we treat model misspecification in three steps. First, 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 a component that depends on the pricing error (alpha). Second, we show how misspecification in the beta component of returns can be mitigated. We demonstrate that the mean-variance portfolio weights are dominated by the alpha portfolio weights as the number of assets increases asymptotically. Given the secondary role played by the beta portfolio, we show that it can be replaced, without any loss of efficiency, by a class of benchmark portfolios that by construction are independent of the distribution of the observed factors, $\boldsymbol{\lambda}$ and $\boldsymbol{\Omega}$, and hence, immune to beta misspecification. Third, 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 econometric identification condition of the model for asset returns.

In the rest of this section, 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 and the frontier portfolios is relegated to the Online Appendix.

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:

$$\mathbf{w}_N^{\text{mv}} = \underset{\mathbf{w}_N}{\operatorname{argmax}} \left((\mathbf{w}'_N \boldsymbol{\mu}_N + (1 - \mathbf{w}'_N \mathbf{1}_N) r_f) - \frac{\gamma}{2} \mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N \right), \quad (13)$$

where $0 < \gamma < \infty$ is the coefficient of risk aversion; $\mathbf{w}_N^{\text{mv}} = (w_1^{\text{mv}}, \dots, w_N^{\text{mv}})'$ is the vector of portfolio weights in the N risky assets; and, the investment in the risk-free asset is given by $1 - \mathbf{1}'_N \mathbf{w}_N^{\text{mv}}$. Alternatively, if one wished to formulate the above problem in terms of a constraint that required the portfolio to achieve a target mean of μ^* , one needs to set

$$\gamma = \frac{(\boldsymbol{\mu}_N - r_f \mathbf{1}_N)' \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N)}{\mu^* - r_f}. \quad (14)$$

The solution to the optimization problem in (13) is:²⁴

$$\mathbf{w}_N^{\text{mv}} = \frac{1}{\gamma} \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N).$$

By standard arguments, the return on portfolio \mathbf{w}_N^{mv} has conditional mean, standard deviation, and Sharpe ratio given by the following three expressions:

$$\mu^{\text{mv}} - r_f = \gamma^{-1} \left((\boldsymbol{\mu}_N - r_f \mathbf{1}_N)' \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \right), \quad (15)$$

$$\sigma^{\text{mv}} = \gamma^{-1} \left((\boldsymbol{\mu}_N - r_f \mathbf{1}_N)' \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \right)^{1/2}, \quad (16)$$

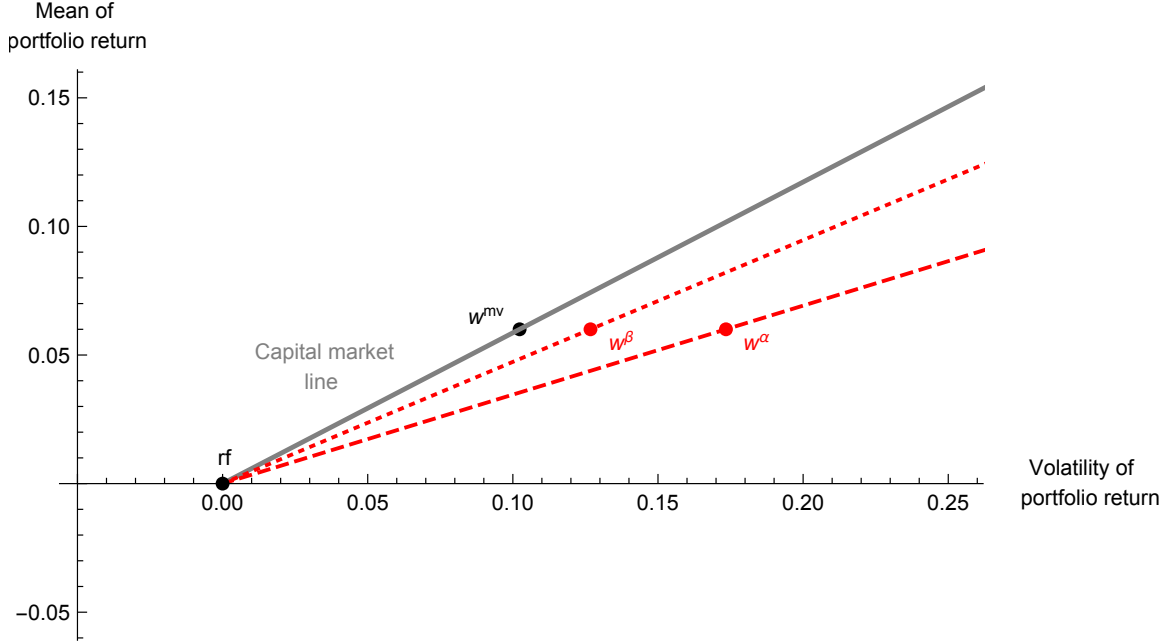
$$\text{SR}^{\text{mv}} = \left((\boldsymbol{\mu}_N - r_f \mathbf{1}_N)' \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \right)^{1/2}. \quad (17)$$

²⁴A special case of the mean-variance portfolio is the *tangency portfolio*, which has zero wealth invested in the risk-free asset:

$$\mathbf{w}_N^{\text{tan}} = \frac{\mathbf{w}_N^{\text{mv}}}{\mathbf{1}'_N \mathbf{w}_N^{\text{mv}}} = \frac{\mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N)}{\mathbf{1}'_N \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N)}.$$

Figure 3: Decomposition of the mean-variance portfolio

In this figure, we plot the mean-variance portfolio in the presence of risk-free asset, \mathbf{w}^{mv} , and its decomposition into two *inefficient* portfolios, one that depends only on the pricing errors, \mathbf{w}^α , and another that depends only on the factor exposure and their premia, \mathbf{w}^β .



The following theorem, which is valid for any finite N , establishes the relations that exist across the mean-variance portfolio, \mathbf{w}_N^{mv} , and the two portfolios that depend on the alpha and beta components of returns, \mathbf{w}_N^α and \mathbf{w}_N^β , respectively. The mean-variance portfolio and its decomposition into the “alpha” and “beta” portfolios, is displayed in Figure 3.

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 $\mu^* > r_f$, the mean-variance portfolio weights satisfies the following decomposition:*

$$\mathbf{w}_N^{mv} = \phi^\alpha \mathbf{w}_N^\alpha + \phi^\beta \mathbf{w}_N^\beta,$$

where:

$$\mathbf{w}_N^\alpha = \frac{1}{\gamma^\alpha} \Sigma_N^{-1} \check{\alpha}_N, \tag{18}$$

$$\mathbf{w}_N^\beta = \frac{1}{\gamma^\beta} \mathbf{V}_N^{-1} \mathbf{B}_N \check{\lambda}, \tag{19}$$

with $\gamma^\alpha = \frac{\check{\alpha}'_N \Sigma_N^{-1} \check{\alpha}_N}{\mu^* - r_f}$, $\gamma^\beta = \gamma - \gamma^\alpha = \frac{\check{\lambda}' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\lambda}}{\mu^* - r_f}$, $\phi^\alpha = \frac{\gamma^\alpha}{\gamma} = \frac{\check{\alpha}'_N \Sigma_N^{-1} \check{\alpha}_N}{(\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N)}$, $\phi^\beta = \frac{\gamma^\beta}{\gamma} = 1 - \phi^\alpha$, and γ defined in (14).²⁵ Furthermore, the portfolios \mathbf{w}_N^α and \mathbf{w}_N^β satisfy the orthogonality condition,

$$(\mathbf{w}_N^\alpha)' \mathbf{V}_N \mathbf{w}_N^\beta = (\mathbf{w}_N^\alpha)' \Sigma_N \mathbf{w}_N^\beta = 0. \quad (20)$$

Moreover, \mathbf{w}_N^β is the minimum-variance portfolio that is orthogonal to \mathbf{w}_N^α and vice versa.

Finally, we have two-fund separation: the inefficient portfolios \mathbf{w}_N^α and \mathbf{w}_N^β can generate all the portfolios on the efficient mean-variance frontier.

Remark 4.1.1. Note that the portfolio \mathbf{w}_N^α , defined in (18), depends only on the pricing error but not on the risk premia, $\check{\lambda}$, or the factor-covariance matrix, $\mathbf{\Omega}$, which is why we label this portfolio as the “alpha” portfolio. Analogously, the portfolio strategy \mathbf{w}_N^β , defined in (19), depends on factor exposures and their risk premia, but not on the pricing errors, $\check{\alpha}_N$. For practical portfolio construction and risk management, what this implies is that if the alpha portfolio is constructed using the expression in (18), then it will not have any exposure to factor risk because $\mathbf{w}_N^{\alpha'} \mathbf{B}_N = \mathbf{0}$.

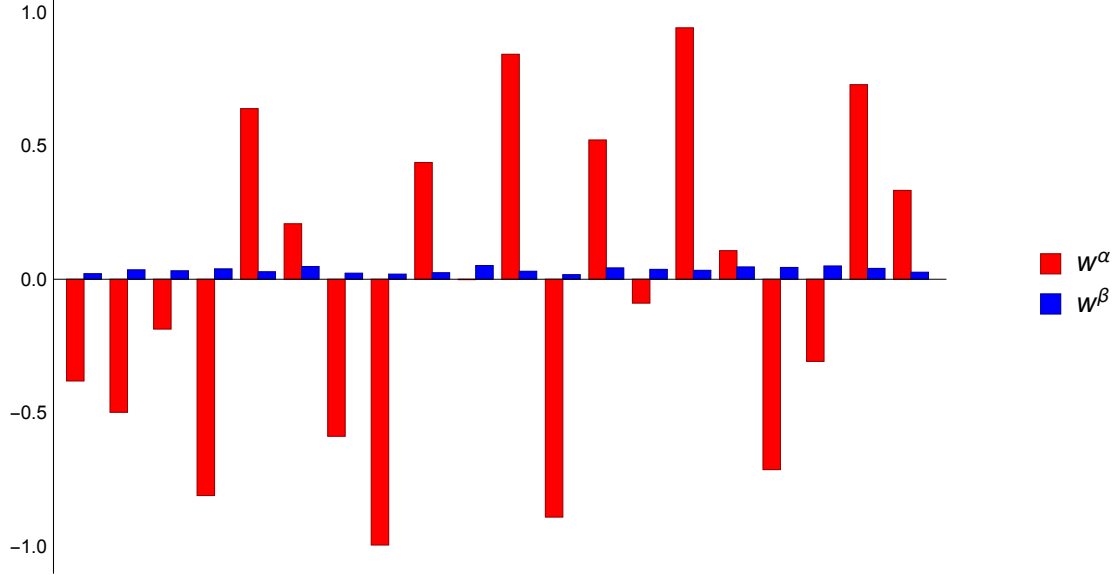
Moreover, as illustrated in Figure 4, the weights of the \mathbf{w}^β portfolio are typically small and positive, while the weights of the \mathbf{w}^α portfolio are large and take both positive and negative values.

Remark 4.1.2. The orthogonality condition in (20) above says that the two portfolios \mathbf{w}_N^α and \mathbf{w}_N^β are uncorrelated, both conditional on the factors and also unconditionally. In addition to \mathbf{w}_N^α and \mathbf{w}_N^β being orthogonal to each other, if one searched for the *minimum-variance* portfolio that is orthogonal to \mathbf{w}_N^α , the resulting portfolio would be \mathbf{w}_N^β , and vice versa. That is, even though the \mathbf{w}_N^α and \mathbf{w}_N^β portfolios are obtained simply by relying on the APT decomposition of the total mean return, these portfolios can be characterized also as being the result of an optimization, which is described in Lemma A.2 and that extends Roll (1980) to the case where, in addition to investing in risky assets, one can invest also

²⁵One can interpret γ^α as the ratio of the share of the contribution of \mathbf{w}_N^α to the expected return on the mean-variance portfolio with unit risk aversion, $(\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N)$, over the target excess mean return, $\mu^* - r_f$. The role of the ϕ^α and ϕ^β coefficients is to ensure that the \mathbf{w}_N^α and \mathbf{w}_N^β portfolios achieve the same target mean return as \mathbf{w}_N^{mv} , which is μ^* .

Figure 4: Typical weights of the \mathbf{w}^α and \mathbf{w}^β portfolios

In this bar chart, we plot the typical weights of the \mathbf{w}^α (red bars) and \mathbf{w}^β portfolios (blue bars) for the case where the number of assets is $N = 20$. The figure shows that the weights of the \mathbf{w}^β portfolio are small and positive. In contrast, the weights of the \mathbf{w}^α portfolio are large and take both positive and negative values.



in a risk-free asset. This mutual optimality property of the \mathbf{w}_N^α and \mathbf{w}_N^β portfolios drives the two-fund separation result: the alpha and beta portfolios, both of which are *inefficient*, span the entire efficient frontier.

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, \mathbf{r}_t , satisfies Assumptions 3.1 and 3.2 and $\boldsymbol{\alpha}_N \neq \mathbf{0}$. Then for any finite $N > K$, and assuming $\mu^* > r_f$, the returns on the inefficient portfolios \mathbf{w}_N^α and \mathbf{w}_N^β have mean, volatility (standard deviation), and Sharpe ratio that have the same quadratic form as the corresponding expressions for the efficient \mathbf{w}_N^{mv} given in (15), (16), and (17):*

$$\begin{aligned} \mu^\alpha - r_f &= \frac{1}{\gamma^\alpha} \check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^{-1} \check{\boldsymbol{\alpha}}_N = \mu^* - r_f; & \mu^\beta - r_f &= \frac{1}{\gamma^\beta} \check{\boldsymbol{\lambda}}' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}} = \mu^* - r_f; \\ \sigma^\alpha &= \frac{1}{\gamma^\alpha} (\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^{-1} \check{\boldsymbol{\alpha}}_N)^{\frac{1}{2}}; & \sigma^\beta &= \frac{1}{\gamma^\beta} (\check{\boldsymbol{\lambda}}' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}})^{\frac{1}{2}}; \end{aligned}$$

$$\text{SR}^\alpha = (\check{\alpha}'_N \Sigma_N^{-1} \check{\alpha}_N)^{\frac{1}{2}}; \quad \text{SR}^\beta = (\check{\lambda}' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\lambda})^{\frac{1}{2}}; \quad (21)$$

with the Sharpe ratios satisfying: $0 \leq \text{SR}^\alpha < \infty$; $0 \leq \text{SR}^\beta < \infty$; and

$$(\text{SR}^{mv})^2 = (\text{SR}^\alpha)^2 + (\text{SR}^\beta)^2. \quad (22)$$

Remark 4.2.1. An important insight is that the quantity on the left-hand side of the APT restriction in (3) is exactly the same as the square of the Sharpe ratio for the \mathbf{w}_N^α portfolio, SR^α , in (21). 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 \mathbf{w}_N^α portfolio. The decomposition of the square of the Sharpe ratio of the mean-variance portfolio in (22) 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 $\check{\alpha}'_N \Sigma_N^{-1} \check{\alpha}_N = (\text{SR}^{mv})^2 - (\text{SR}^\beta)^2$, but do not interpret the left-hand side as the square of the Sharpe ratio of a portfolio, in particular, the alpha portfolio. In Lemma A.1, we provide the general conditions that are needed for decomposing the squared Sharpe ratio of any portfolio that can be written as the sum of two orthogonal components.

4.2 Mitigating misspecification in the 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.

For this analysis, we need to extend Definition 3.2 of a regular economy in two different dimensions. The earlier definition of regularity was applied to \mathbf{B}_N and Σ_N . One, we extend the definition to any arbitrary matrix of dimension $N \times K$, such as \mathbf{D}_N , and an arbitrary positive-definite $N \times N$ matrix, \mathbf{C}_N . Two, we impose that all the eigenvalues are diverging at precisely the same rate.²⁶

Definition 4.1 (\mathbf{C}_N -regularity). *A matrix \mathbf{D}_N is \mathbf{C}_N -regular if there exists an increasing function of N , $f(N)$, such that for any $1 \leq j \leq K$, the eigenvalues $g_{jK}(\frac{1}{f(N)} \mathbf{D}'_N \mathbf{C}_N^{-1} \mathbf{D}_N) \rightarrow \delta_j > 0$, where δ_j is some finite positive constant.*

²⁶A 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.

We now state our result about the asymptotic properties of mean-variance portfolio weights.

Theorem 4.3 (Weights of alpha, beta, and mean-variance portfolios for large N). *Suppose that the vector of asset returns, \mathbf{r}_t , satisfies Assumptions 3.1 and 3.2 and $\boldsymbol{\alpha}_N \neq 0$. Suppose also that \mathbf{A}_N , \mathbf{B}_N , and $\mathbf{1}_N$ are \mathbf{C}_N -regular with the same scaling factor $f(N)$ and \mathbf{A}_N and \mathbf{B}_N are not asymptotically collinear.²⁷*

As $N \rightarrow \infty$, then:

(i) $0 \leq \phi^\alpha \leq 1$, $0 \leq \phi^\beta \leq 1$, and element-by-element

$$\frac{w_{N,i}^\beta}{w_{N,i}^\alpha} \rightarrow 0. \quad (23)$$

(ii) The sum of the squared components of the mean-variance portfolio vectors $\mathbf{w}_N^{\alpha'} \mathbf{w}_N^\alpha$ is always bounded, whereas $\mathbf{w}_N^{\beta'} \mathbf{w}_N^\beta$ always converges to zero.

(iii) The sum of the components of the mean-variance portfolio vectors $|\mathbf{1}'_N \mathbf{w}_N^\alpha|$ can diverge to infinity, whereas $|\mathbf{1}'_N \mathbf{w}_N^\beta|$ is always bounded.

(iv) The vector of weights for the mean-variance portfolio are asymptotically equivalent element-by-element to the weights of $\phi^\alpha \mathbf{w}_N^\alpha$:

$$w_{N,i}^{mv} = (1 - \phi^\beta)w_{N,i}^\alpha + \phi^\beta w_{N,i}^\beta \sim (1 - \phi^\beta)w_{N,i}^\alpha = \phi^\alpha w_{N,i}^\alpha, \quad (24)$$

where the symbol \sim denotes asymptotic equivalence.²⁸

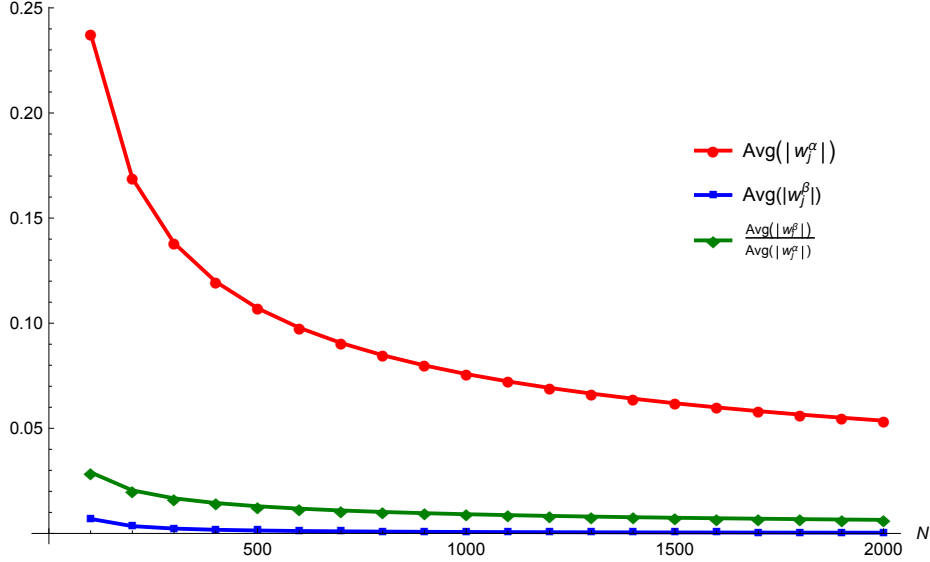
Remark 4.3.1. The \mathbf{w}_N^α portfolio dominates the \mathbf{w}_N^β portfolio across all three norms considered in the theorem above; this dominance is illustrated in Figure 5. The notion of diversification used in part (ii) of the theorem is the sum of the squares, which is the same notion adopted in Chamberlain (1983). Because $\mathbf{w}_N^{\beta'} \mathbf{w}_N^\beta \geq \sup_i |w_{N,i}^\beta|$, it follows that the \mathbf{w}_N^β portfolio is diversified according to the sup norm criterion, which is the norm used

²⁷By asymptotic collinearity we mean that either $\mathbf{A}'_N \mathbf{M}_{B_N} \mathbf{A}_N \rightarrow 0$ or $\mathbf{B}'_N \mathbf{M}_{A_N} \mathbf{B}_N \rightarrow 0$ or both, as N diverges, depending on whether the number of unobserved factors $p \leq K$, $p \geq K$ or $p = K$, where $\mathbf{M}_C = \mathbf{I}_N - \mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'$ is the matrix that spans the space orthogonal to any full-column-rank matrix \mathbf{C} . When $p \leq K$, a sufficient condition for this is $\mathbf{A}_N = \mathbf{B}_N \boldsymbol{\delta} + \mathbf{G}_N$ for some constant $K \times p$ matrix $\boldsymbol{\delta}$ and some residual matrix \mathbf{G}_N satisfying $\mathbf{G}'_N \mathbf{G}_N \rightarrow 0$.

²⁸We say that $a_n \sim b_n$, if $a_n/b_n \rightarrow 1$ as $n \rightarrow \infty$.

Figure 5: Relative average magnitude of weights of \mathbf{w}^α and \mathbf{w}^β portfolios

In this figure, we plot three quantities as the number of assets, N , increases. The three quantities are: (1) the average magnitude of the weights of the \mathbf{w}^α portfolio, given by the red line with circle markers; (2) the average magnitude of the weights of the \mathbf{w}^β portfolio, given by the blue line with square markers; and, (3) the ratio of the average magnitude of the weights of the \mathbf{w}^β portfolio to the corresponding weights of the \mathbf{w}^α portfolio, given by the green line with diamond markers. The figure shows that as N increases, the average magnitude of the weights of the \mathbf{w}^β portfolio declines faster than the average magnitude of the weights of the \mathbf{w}^α portfolio.



in Green and Hollifield (1992). In contrast, the \mathbf{w}_N^α portfolio is not necessarily diversified according to the squared norm.²⁹ Part (iii) of the theorem studies how the total investment in risky assets is allocated between the \mathbf{w}_N^α and \mathbf{w}_N^β portfolios. The result in the theorem shows that $\mathbf{1}'_N \mathbf{w}^\alpha$ 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 in the \mathbf{w}_N^α portfolio.³⁰ On the other hand, the investment in the \mathbf{w}_N^β portfolio is bounded, and hence, is associated with a finite amount of leverage.

²⁹For 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 \mathbf{w}_N^α portfolio can go to zero at a sufficiently slow rate, as slow as $1/\sqrt{N}$, as shown in Corollary 4.3.1.

³⁰To see that $\mathbf{1}'_N \mathbf{w}_N^\alpha$ can diverge, consider the following example in which $\boldsymbol{\alpha}_N = \mathbf{1}_N/\sqrt{N}$, $\boldsymbol{\Sigma}_N = \sigma^2 \mathbf{I}_N$, with \mathbf{I}_N the $N \times N$ identity matrix, and there is a single factor with β_i with iid distribution having mean 1 and variance $\sigma_\beta^2 > 0$. Then, $\mathbf{1}'_N \mathbf{w}_N^\alpha \sim \sqrt{N} \sigma_\beta^2 / (1 + \sigma_\beta^2)$ which goes to infinity with N .

In order to get a better understanding of the result in (23), in the corollary below we look at a special case where $f(N) = N$. We consider only part (i) of the theorem, because the other parts of the theorem are unchanged under the special case. Then, we provide the intuition for the result in the theorem above and the corollary.

Corollary 4.3.1 (Weights of alpha and beta portfolios for large N : Special case). *Suppose that the assumptions of Theorem 4.3 are satisfied and that the row sums of $\mathbf{A}_N, \mathbf{B}_N$ and \mathbf{C}_N^{-1} are uniformly bounded.³¹ Suppose also that $f(N) = N$. Then, as $N \rightarrow \infty$, for the case of bounded-residual variation the absolute value of the components of the mean-variance portfolio vectors, \mathbf{w}_N^α and \mathbf{w}_N^β decrease at most at the rate:*

$$|w_{N,i}^\alpha| = \mathcal{O} \left(|\mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N| + \frac{1}{N^{\frac{1}{2}}} \right) \quad \text{and} \quad |w_{N,i}^\beta| = \mathcal{O} \left(\frac{1}{N} \right),$$

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

$$|w_{N,i}^\alpha| = \mathcal{O} \left(|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{a}_N| + \frac{\|\mathbf{a}_N\|}{N^{\frac{1}{2}}} + \frac{\|\boldsymbol{\lambda}_{miss}\|}{N} \right) \quad \text{and} \quad |w_{N,i}^\beta| = \mathcal{O} \left(\frac{1}{N} \right).$$

To understand the intuition for the results about the dominance of the \mathbf{w}_N^α portfolio weights, recall that the objective of mean-variance portfolio optimization is to maximize the portfolio Sharpe ratio, which entails increasing the mean of the portfolio return and/or reducing the volatility of the portfolio return. There are two sources of risk: factor exposure and idiosyncratic exposure. The factor exposure of the \mathbf{w}_N^α portfolio is zero—*irrespective* of the rate at which the weights decrease—because of the orthogonality of \mathbf{w}_N^α to \mathbf{B}_N . Regarding exposure to idiosyncratic risk, the elements of \mathbf{w}_N^α cannot decrease faster than $1/N^{\frac{1}{2}}$ because then the idiosyncratic risk of the portfolio goes to zero; but, the idiosyncratic risk of the alpha portfolio coincides with the Sharpe ratio, implying that the Sharpe ratio also goes to zero. On the other hand, the APT restriction does not allow the rate at which the weights decrease to be slower than $1/N^{\frac{1}{2}}$. To understand this, consider the simple case where $\boldsymbol{\Sigma}_N$ is the identity matrix. Recall that for the sum of N positive terms to be bounded, it suffices that each term declines not slower than $1/N$. The square-root rate follows from the fact that in our case each term is in fact the square of $w_{N,i}^\alpha$. Thus, the rate of $1/N^{\frac{1}{2}}$ strikes just the correct balance between optimizing the risk and return of the \mathbf{w}_N^α portfolio.

³¹Given an $N \times M$ matrix \mathbf{D} , we say its row sums are uniformly bounded when $\sup_{1 \leq j \leq N} \sum_{i=1}^M |d_{ji}| \leq \delta < \infty$, for some arbitrary δ .

Let us now look at the \mathbf{w}_N^β portfolio. If the weights decrease at any rate slower than $1/N$, then the systematic exposure explodes because the factors are pervasive (see Definition 3.2). On the other hand, if the weights decrease faster than $1/N$, then the portfolio risk declines to zero, leading to a Sharpe ratio of zero because the expression for the Sharpe ratio is exactly the same as the one for portfolio risk. So, the rate of $1/N$ strikes the correct balance between optimizing the risk and return of the portfolio.³² Notice that the rate $1/N$ makes the \mathbf{w}_N^β portfolio well diversified even with respect to idiosyncratic exposure, enhancing its Sharpe ratio even further.

Above, we have demonstrated that the mean-variance portfolio weights are dominated by the alpha portfolio weights as the number of assets increases asymptotically. Given the secondary role played by the beta portfolio, we show below that it can be replaced, without any loss of efficiency, by a class of benchmark portfolios that by construction are independent of the distribution of the observed factors, $\boldsymbol{\lambda}$ and $\boldsymbol{\Omega}$, and hence, immune to beta misspecification. This demonstrates that one can construct optimal portfolios that are *independent* of risk premia.

Theorem 4.4 (Weight and Sharpe ratio of mean-variance portfolio for large N). *Suppose that the vector of asset returns, \mathbf{r}_t , satisfies Assumptions 3.1 and 3.2 and $\boldsymbol{\alpha}_N \neq 0$. Suppose further that the investor holds a well-diversified benchmark portfolio $\mathbf{w}_N^{\text{bench}}$ with target mean μ^* satisfying the following properties:*

$$(\mathbf{w}_N^{\text{bench}})' \boldsymbol{\alpha}_N \rightarrow 0, \quad \mathbf{B}'_N \mathbf{w}_N^{\text{bench}} \rightarrow \mathbf{c}^{\text{bench}}, \quad (\mathbf{w}_N^{\text{bench}})' \boldsymbol{\Sigma}_N \mathbf{w}_N^{\text{bench}} \rightarrow 0, \quad (25)$$

where $\mathbf{c}^{\text{bench}}$ is a $K \times 1$ vector of constants, different from the zero vector, satisfying $\boldsymbol{\lambda}' \mathbf{c}^{\text{bench}} \neq 0$. Set $\phi^{\text{bench}} = \frac{\gamma^{\text{bench}}}{\gamma}$, where $\gamma^{\text{bench}} = \frac{(SR^{\text{bench}})^2}{\mu^* - r_f}$ and γ is given in (14).

(i) If $\mathbf{c}^{\text{bench}}$ is perfectly proportional to the vector $\boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}$, then $w_{N,i}^{mv}$ is dominated by $w_{N,i}^\alpha$:

$$w_{N,i}^{mv} \sim (1 - \phi^{\text{bench}}) w_{N,i}^\alpha + \phi^{\text{bench}} w_{N,i}^{\text{bench}} \sim (1 - \phi^{\text{bench}}) w_{N,i}^\alpha,$$

but both SR^α and SR^{bench} contribute to SR^{mv} :

$$(SR^{mv})^2 \sim (SR^\alpha)^2 + (SR^{\text{bench}})^2.$$

³²We can see the above argument also in the expression of the portfolio weight: $\mathbf{w}_N^\beta = \mathbf{V}_N^{-1} (\mathbf{B}_N \tilde{\boldsymbol{\lambda}}_N) = (\mathbf{B}_N \boldsymbol{\Omega} \mathbf{B}'_N + \boldsymbol{\Sigma}_N)^{-1} (\mathbf{B}_N \tilde{\boldsymbol{\lambda}}_N)$. Because \mathbf{B}_N appears twice in denominator of \mathbf{w}_N^β , it causes its faster decay to zero. On the other hand, $\mathbf{w}_N^\alpha = \boldsymbol{\Sigma}_N^{-1} \tilde{\boldsymbol{\alpha}}_N$. However, only the $\mathbf{A}_N \boldsymbol{\lambda}_{\text{miss}}$ part of $\tilde{\boldsymbol{\alpha}}_N$ can appear in the denominator of \mathbf{w}_N^α , implying that \mathbf{w}_N^α decays to zero slowly whenever \mathbf{a}_N is not zero.

Moreover, if $K = 1$, then \mathbf{c}^{bench} is always perfectly proportional to the vector $\mathbf{\Omega}^{-1}\boldsymbol{\lambda}$.

(ii) If \mathbf{c}^{bench} is not proportional to the vector $\mathbf{\Omega}^{-1}\boldsymbol{\lambda}$, then w_N^{mv} is not asymptotically equivalent to $(1 - \phi^{bench})w_{N,i}^\alpha + \phi^{bench}w_{N,i}^{bench}$ and

$$(SR^{mv})^2 > (SR^\alpha)^2 + (SR^{bench})^2.$$

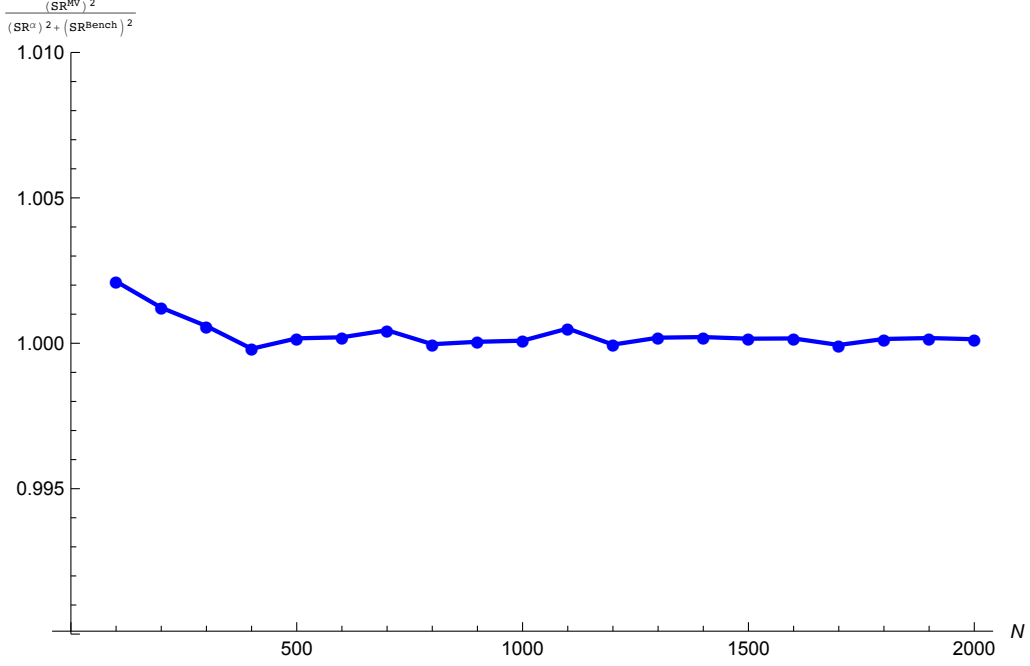
Remark 4.4.1. The first assumption in (25) implies that the benchmark portfolio is asymptotically orthogonal to $\boldsymbol{\alpha}_N$. The second assumption rules out that the benchmark portfolio return is equal to the risk-free return in the limit. The third assumption requires that the benchmark portfolio be well diversified. Note that for the unbounded-variation case, the first assumption is satisfied whenever $(\mathbf{w}_N^{bench})'\mathbf{a}_N \rightarrow 0$ and $(\mathbf{w}_N^{bench})'\mathbf{A}_N \rightarrow 0$, where the latter condition ensures that \mathbf{w}_N^{bench} diversifies away the contribution of the latent factors, \mathbf{A}_N , to $\boldsymbol{\Sigma}_N$.

Remark 4.4.2. To construct a valid benchmark portfolio, 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 \mathbf{w}_N^β can be approximated by a portfolio that is similar to the market portfolio, \mathbf{w}_N^{mkt} , suitably normalized to achieve a target mean of μ^* . Alternatively, the empirical findings of DeMiguel, Garlappi, and Uppal (2009) suggest that one could hold an equally weighted portfolio suitably normalized, implying that $\mathbf{w}_N^{bench} = \frac{\mu^* - r_f}{(\mathbf{1}_N/N)'(\boldsymbol{\mu} - r_f\mathbf{1}_N)} \frac{\mathbf{1}_N}{N}$. In the theorem above, we formalize these two proposals by reporting the results for any arbitrary benchmark portfolio with target mean μ^* . 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 when there is only a single factor, that is, $K = 1$.

Remark 4.4.3. The assumptions in (25) imply that the return on the benchmark portfolio is asymptotically equivalent to the return on the portfolio of factors with weight \mathbf{c}^{bench} ; that is, $(\mathbf{w}_N^{bench})'(\mathbf{r}_t - r_f\mathbf{1}_N) \approx (\mathbf{c}^{bench})'(\mathbf{f}_t - r_f\mathbf{1}_K)$. Therefore, asymptotic optimality of the benchmark portfolio requires that \mathbf{c}^{bench} equals the mean-variance portfolio constructed using the K factors. This choice guarantees that the benchmark portfolio achieves the largest possible Sharpe ratio, as stated in part (i) of the theorem. Figure 6 shows that as the number of risky assets increases, the ratio of $(SR^{mv})^2$ to $(SR^\alpha)^2 + (SR^{bench})^2$ quickly approaches 1.

Figure 6: Convergence of \mathbf{w}^α and $\mathbf{w}^{\text{bench}}$ portfolios to \mathbf{w}^{mv} portfolio

This figure shows that as the number of assets, N , increases, the sum of the squares of the Sharpe ratios of the \mathbf{w}^α and $\mathbf{w}^{\text{bench}}$ portfolios converge to the square of the Sharpe ratio of the mean-variance portfolio.



Remark 4.4.4. In striking contrast to the portfolio weights \mathbf{w}_N^α and \mathbf{w}_N^β , where the components of \mathbf{w}_N^α asymptotically dominate those of \mathbf{w}_N^β , their Sharpe ratios are, in general, of the same order of magnitude even for large N . The reason is that the return on the portfolio \mathbf{w}_N^β is $r_t^\beta - r_f = (\check{\boldsymbol{\lambda}}_N' + \mathbf{z}_t') \mathbf{B}'_N \mathbf{w}_N^\beta + \boldsymbol{\varepsilon}'_t \mathbf{w}_N^\beta$. Thus, even though the variance of the idiosyncratic component of the return, $\boldsymbol{\varepsilon}'_t \mathbf{w}_N^\beta$, is diversified away as we increase the number of assets, as long as some of the factors are pervasive, the term $\mathbf{B}'_N \mathbf{w}_N^\beta$ will not be diversified away. Hence, the excess mean return and Sharpe ratio of \mathbf{w}_N^β will not go to zero.

Having addressed the problem of misspecification arising from the beta component of returns, we now consider misspecification in the alpha component of returns.

4.3 Mitigating misspecification in the alpha component of returns

Section 3.5 explains how the vector of pricing errors, $\boldsymbol{\alpha}_N$, captures model misspecification arising from missing factors and asset-specific characteristics unrelated to factors. The

analysis in the previous section demonstrates that the vector of pricing errors, α_N , plays a dominant role in the choice of optimal portfolio weights and their returns. Therefore, it is vital to estimate α_N precisely. The APT restriction, which we reproduce below:

$$\alpha_N' \Sigma_N^{-1} \alpha_N \leq \delta < \infty. \quad (26)$$

provides exactly the condition that must be imposed in the estimation of the factor model generating returns. In principle, the constraint in (26) binds when $N \rightarrow \infty$ because the theory does not specify a particular value for δ (though, from the result in (22) we know that δ is less than the square of the Sharpe ratio of the market portfolio, if the market is an efficient portfolio). Following the common practice in applied econometrics and statistics 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.

The way in which we impose the APT restriction given in (26) depends on whether we are in the case of bounded or unbounded residual variation. We propose a multi-step procedure to 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.³³ 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 $\hat{\Sigma}_N$, where the $\hat{\cdot}$ denotes an estimated quantity. This part uses conventional principal-component analysis of $\hat{\Sigma}_N$ and it allows one to determine the number of latent pervasive factors, p ; see, for example, Anderson (1984). In the next two subsections, we provide the details of how to estimate the model when $p = 0$ (bounded-variation case) and when $p > 0$ (unbounded-variation case), after imposing the APT constraint. To conserve space, the case where the factors themselves are measured with error is discussed in Appendix A.12.

³³Some of the observed factors could be mismeasured, requiring one to use an estimation procedure that accounts for the possibility of errors in variables.

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

In the bounded-residual-variation case (that is, $\boldsymbol{\alpha}_N = \mathbf{a}_N$), the true unconditional means and covariances of returns satisfy

$$E(\mathbf{r}_t - r_f \mathbf{1}_N) = \boldsymbol{\mu}_0 - r_f \mathbf{1}_N = \boldsymbol{\alpha}_0 + \mathbf{B}_0 \boldsymbol{\lambda}_0, \quad \text{var}(\mathbf{r}_t) = \mathbf{V}_0 = \mathbf{B}_0 \boldsymbol{\Omega}_0 \mathbf{B}_0' + \boldsymbol{\Sigma}_0,$$

where the subscript “0” indicates the true value of a parameter (and we do not use the subscript “ N ” for the true value of parameters in order to limit the number of subscripts), $\boldsymbol{\lambda}_0 = E(\mathbf{f}_t) - r_f \mathbf{1}_K$ is the vector of risk premia and $\boldsymbol{\Omega}_0 = \text{var}(\mathbf{f}_t)$ is the covariance matrix of the factors, assuming stationarity of the K factors, \mathbf{f}_t , and assuming that the factors are traded, without loss of generality. In order to identify $\boldsymbol{\lambda}_0$ and $\boldsymbol{\Omega}_0$, one needs to consider also the information stemming from the sample observations for \mathbf{f}_t . Although our argument applies to virtually any estimation procedure, we will illustrate it with respect to the (pseudo) Gaussian 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; hence, the use of pseudo ML.

The (pseudo) ML estimator, based on the unconditional joint distribution of $\begin{pmatrix} \mathbf{r}_t \\ \mathbf{f}_t \end{pmatrix} - r_f \mathbf{1}_{N+K}$ and assuming i.i.d. residuals for simplicity, is:³⁴

$$\begin{aligned} L(\boldsymbol{\theta}) &= -\frac{1}{2} \log(\det(\boldsymbol{\Sigma}_N)) \\ &\quad - \frac{1}{2T} \sum_{t=1}^T \left(\mathbf{r}_t - r_f \mathbf{1}_N - \boldsymbol{\alpha}_N - \mathbf{B}_N (\mathbf{f}_t - r_f \mathbf{1}_K) \right)' \boldsymbol{\Sigma}_N^{-1} \left(\mathbf{r}_t - r_f \mathbf{1}_N - \boldsymbol{\alpha}_N - \mathbf{B}_N (\mathbf{f}_t - r_f \mathbf{1}_K) \right) \\ &\quad - \frac{1}{2} \log(\det(\boldsymbol{\Omega})) - \frac{1}{2T} \sum_{t=1}^T \left(\mathbf{f}_t - r_f \mathbf{1}_K - \boldsymbol{\lambda} \right)' \boldsymbol{\Omega}^{-1} \left(\mathbf{f}_t - r_f \mathbf{1}_K - \boldsymbol{\lambda} \right), \end{aligned} \quad (27)$$

where $\boldsymbol{\theta} = (\boldsymbol{\alpha}'_N, \text{vec}(\mathbf{B}_N)', \text{vech}(\boldsymbol{\Sigma}_N)', \boldsymbol{\lambda}', \text{vech}(\boldsymbol{\Omega})')'$.³⁵ Therefore, the ML estimator for $\boldsymbol{\alpha}_0, \mathbf{B}_0, \boldsymbol{\Sigma}_0$ coincide with the OLS estimator, conditional on the realization of the factors.

³⁴Notice 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 time-varying conditional means, conditional variances, and conditional covariances.

³⁵Note 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.

On the other hand, the ML estimators for $\boldsymbol{\lambda}_0$ and $\boldsymbol{\Omega}_0$ are the sample mean and sample covariance of the factors \mathbf{f}_t .

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

$$\hat{\boldsymbol{\theta}}_{MLC} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} L(\boldsymbol{\theta}) \text{ such that } \boldsymbol{\alpha}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N \leq \delta. \quad (28)$$

Because the parameter $\boldsymbol{\alpha}_0$ is constrained only by the APT restriction, imposing this constraint may lead $\hat{\boldsymbol{\theta}}_{MLC}$ to be a more precise estimator of the true parameter values compared to the unconstrained estimator, $\hat{\boldsymbol{\theta}}_{ML}$. The theory does not tell us what δ should be in (28). As discussed earlier, we know that the upper bound of δ must be less than the square of the Sharpe ratio of the mean-variance efficient portfolio; in our empirical application, we will choose δ 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). *Suppose that the vector of asset returns, \mathbf{r}_t , satisfies Assumption 3.1. Given any $\kappa > 0$,*

$$\hat{\boldsymbol{\theta}}_{MLC} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left\{ L(\boldsymbol{\theta}) - \kappa(\boldsymbol{\alpha}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N - \delta) \right\},$$

where $L(\boldsymbol{\theta})$ is defined in (27). If $\left(\sum_{t=1}^T \dot{\mathbf{f}}_t \dot{\mathbf{f}}_t'\right)$ is nonsingular, then $\hat{\boldsymbol{\theta}}_{MLC} = (\hat{\boldsymbol{\alpha}}'_{N,MLC}, \operatorname{vec}(\hat{\mathbf{B}}_{N,MLC})', \operatorname{vech}(\hat{\boldsymbol{\Sigma}}_{N,MLC})', \hat{\boldsymbol{\lambda}}'_{MLC}, \operatorname{vech}(\hat{\boldsymbol{\Omega}}_{MLC})')'$ exists, where:

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_{N,MLC} &= \frac{1}{1+\kappa} \left[\bar{\mathbf{r}} - r_f \mathbf{1}_N - \hat{\mathbf{B}}_{N,MLC} (\bar{\mathbf{f}} - r_f \mathbf{1}_K) \right], \\ \hat{\mathbf{B}}_{N,MLC} &= \left(\sum_{t=1}^T \dot{\mathbf{r}}_t \dot{\mathbf{f}}_t' \right) \left(\sum_{t=1}^T \dot{\mathbf{f}}_t \dot{\mathbf{f}}_t' \right)^{-1}, \\ \hat{\boldsymbol{\Sigma}}_{N,MLC} &= \frac{1}{T} \sum_{t=1}^T (\dot{\mathbf{r}}_t - \hat{\mathbf{B}}_{N,MLC} \dot{\mathbf{f}}_t) (\dot{\mathbf{r}}_t - \hat{\mathbf{B}}_{N,MLC} \dot{\mathbf{f}}_t)', \end{aligned}$$

where $\dot{\mathbf{f}}_t = \mathbf{f}_t - r_f \mathbf{1}_K - \frac{1}{(1+\kappa)}(\bar{\mathbf{f}} - r_f \mathbf{1}_K)$, $\dot{\mathbf{r}}_t = \mathbf{r}_t - r_f \mathbf{1}_N - \frac{1}{(1+\kappa)}(\bar{\mathbf{r}} - r_f \mathbf{1}_N)$, and the MLC estimators $\hat{\boldsymbol{\lambda}}'_{MLC}$ and $\operatorname{vech}(\hat{\boldsymbol{\Omega}}_{MLC})'$ coincide with the sample mean and covariance of the factors \mathbf{f}_t .

Remark 4.5.1. The constrained estimator $\hat{\alpha}_{N,MLC}$ turns out to be precisely the ridge estimator for α_0 , because $\kappa > 0$.³⁶ The estimators of $\hat{\mathbf{B}}_{N,MLC}$ and $\hat{\Sigma}_{N,MLC}$ are functions also of κ 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 \mathbf{f}_t because the APT constraint does not affect the distribution of the factors \mathbf{f}_t .

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

The second case of alpha misspecification is of unbounded residual variation, which arises when there are $p > 0$ missing pervasive factors. For the case where the pricing error unrelated to the missing factors, \mathbf{a}_N is zero in (9), we get that $\alpha_N = \mathbf{A}_N \lambda_{miss}$ and $\Sigma_N = \mathbf{A}_N \mathbf{A}'_N + \mathbf{C}_N$, where λ_{miss} is the risk premia corresponding to the missing factors, and \mathbf{C}_N 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 α_N is a component of the expected return, μ_N ; likewise, Σ_N is a component of the return-covariance matrix, \mathbf{V}_N . Hence, \mathbf{A}_N appears in both the mean and covariance matrix of returns. MacKinlay and Pástor (2000) use this insight to improve the precision of the estimated \mathbf{A}_N parameters, which, in turn, improves substantially the performance of the estimated portfolio.³⁷ Importantly, using the Sherman-Morrison-Woodbury formula, it follows that:

$$\alpha'_N \Sigma_N^{-1} \alpha_N = \lambda'_{miss} \mathbf{A}'_N \Sigma_N^{-1} \mathbf{A}_N \lambda_{miss} = \lambda'_{miss} (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} (\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N) \lambda_{miss}.$$

Thus, $\alpha'_N \Sigma_N^{-1} \alpha_N$ converges to $\lambda'_{miss} \lambda_{miss}$ as $N \rightarrow \infty$ because $(\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} (\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)$ converges to the identity matrix given that the missing factors are pervasive implying that $(\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)$ is increasing without bound. This means that the APT restriction is always

³⁶One approach is to solve for the Karush-Kuhn-Tucker multiplier $\hat{\kappa}_{MLC}$. An iterative procedure is required to solve for $\hat{\kappa}_{MLC}$ and $\hat{\mathbf{B}}_{N,MLC}$ jointly; then the estimators $\hat{\alpha}_{N,MLC}$ and $\hat{\Sigma}_{N,MLC}$ follow. Instead, we choose κ by cross-validation, which is computationally simpler. The formulae for the estimators of the other parameters remain the same. The cross-validation approach is closer to a lasso (least absolute shrinkage and selection operator) formulation where one considers the penalized log-likelihood $L(\theta) - \kappa(\alpha'_N \Sigma_N^{-1} \alpha_N)$; however, in contrast to lasso, our constraint is quadratic.

³⁷Note that because we are interpreting the missing factors as unobserved, without loss of generality one can assume that $\mathbf{A}_N \mathbf{A}'_N$ represents the contribution of the missing factors to the residual variance Σ_N because the missing factors are assumed to be uncorrelated and have unit variance, leaving the risk-premia λ_{miss} as free parameters to be estimated. MacKinlay and Pástor (2000) consider a different identification assumption. For $p = 1$ they estimate α_N without distinguishing between \mathbf{A}_N and λ_{miss} , implying that the contribution of the single missing factor to the return variance equals $\alpha_N \alpha'_N / (\text{SR}^h)^2$, where SR^h is the Sharpe ratio of the missing factor.

satisfied for the case of only missing pervasive factors (that is, the case where $\mathbf{a}_N = \mathbf{0}$), once we recognize that $\boldsymbol{\Sigma}_N$ contains the loadings of the missing factors, \mathbf{A}_N .

However, for the general unbounded-variation case where the pricing error consists of *both* missing factors and a component that is unrelated to factors, $\boldsymbol{\alpha}_N = \mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{a}_N$, the APT restriction is not automatically satisfied. Therefore, when estimating the model we need to impose the additional constraint: $\mathbf{a}'_N \mathbf{a}_N \leq \delta < \infty$ for any N . Under the same assumptions as above concerning the K observed factors \mathbf{f}_t , the true unconditional means and covariances of returns now satisfy the equations below, where \mathbf{C}_0 has bounded maximum eigenvalue, in contrast to $\boldsymbol{\Sigma}_0$:

$$E(\mathbf{r}_t - r_f \mathbf{1}_N) = \boldsymbol{\mu}_0 - r_f \mathbf{1}_N = \mathbf{A}_0 \boldsymbol{\lambda}_{miss0} + \mathbf{a}_0 + \mathbf{B}_0 \boldsymbol{\lambda}_0, \quad \text{var}(\mathbf{r}_t) = \mathbf{V}_0 = \mathbf{B}_0 \boldsymbol{\Omega}_0 \mathbf{B}'_0 + \mathbf{A}_0 \mathbf{A}'_0 + \mathbf{C}_0.$$

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

$$\begin{aligned} L(\boldsymbol{\theta}) = & -\frac{1}{2} \log(\det(\mathbf{A}_N \mathbf{A}'_N + \mathbf{C}_N)) - \frac{1}{2T} \sum_{t=1}^T \left(\mathbf{r}_t - r_f \mathbf{1}_N - \mathbf{A}_N \boldsymbol{\lambda}_{miss} - \mathbf{a}_N - \mathbf{B}_N (\mathbf{f}_t - r_f \mathbf{1}_K) \right)' \\ & \times (\mathbf{A}_N \mathbf{A}'_N + \mathbf{C}_N)^{-1} \left(\mathbf{r}_t - r_f \mathbf{1}_N - \mathbf{A}_N \boldsymbol{\lambda}_{miss} - \mathbf{a}_N - \mathbf{B}_N (\mathbf{f}_t - r_f \mathbf{1}_K) \right) \\ & - \frac{1}{2} \log(\det(\boldsymbol{\Omega})) - \frac{1}{2T} \sum_{t=1}^T \left(\mathbf{f}_t - r_f \mathbf{1}_K - \boldsymbol{\lambda} \right)' \boldsymbol{\Omega}^{-1} \left(\mathbf{f}_t - r_f \mathbf{1}_K - \boldsymbol{\lambda} \right). \end{aligned} \quad (29)$$

Without loss of generality, one can assume that the missing factors are uncorrelated with each other and have unit variance, achieving identification of \mathbf{A}_0 .³⁸ However, $\boldsymbol{\lambda}_{miss0}$ and $\boldsymbol{\alpha}_0$ *cannot* be identified separately unless the APT restriction is imposed, as shown below.

Theorem 4.6 (Unbounded-variation case). *Suppose that the vector of asset returns, \mathbf{r}_t , satisfies Assumption 3.1. Given any $\kappa \geq 0$, then:*

$$\hat{\boldsymbol{\theta}}_{MLC} = \underset{\boldsymbol{\theta}}{\text{argmax}} \left\{ L(\boldsymbol{\theta}) - \kappa (\mathbf{a}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{a}_N - \delta) \right\},$$

where $L(\boldsymbol{\theta})$ is defined in (29), and $\hat{\boldsymbol{\theta}}_{MLC} = (\hat{\mathbf{a}}'_{N,MLC}, \hat{\boldsymbol{\lambda}}_{miss,MLC}, \text{vec}(\hat{\mathbf{A}}_{N,MLC})', \text{vec}(\hat{\mathbf{B}}_{N,MLC})', \text{vech}(\hat{\mathbf{C}}_{N,MLC})', \hat{\boldsymbol{\lambda}}'_{MLC}, \text{vech}(\hat{\boldsymbol{\Omega}}_{MLC})')'$.

³⁸We are implicitly assuming that observed and missing factors are mutually uncorrelated.

(i) For $\kappa > 0$, if the APT restriction holds exactly, that is, $\mathbf{a}'_0 \boldsymbol{\Sigma}_0^{-1} \mathbf{a}_0 = \delta$, and $\boldsymbol{\Sigma}_{ff} - \ddot{\mathbf{f}}\mathbf{f}'$ is nonsingular, then

$$\text{vec}(\hat{\mathbf{B}}_{N,MLC}) = \left((\boldsymbol{\Sigma}_{ff} \otimes \mathbf{I}) - (\ddot{\mathbf{f}}\mathbf{f}' \otimes (2\mathbf{G}_N - \mathbf{G}_N \mathbf{G}_N)) \right)^{-1} \text{vec} \left(\boldsymbol{\Sigma}_{rf} - (2\mathbf{G}_N - \mathbf{G}_N \mathbf{G}_N) \ddot{\mathbf{r}} \right),$$

$$\hat{\boldsymbol{\lambda}}_{missMLC} = (\hat{\mathbf{A}}'_{N,MLC} \hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1} \hat{\mathbf{A}}_{N,MLC})^{-1} \hat{\mathbf{A}}'_{N,MLC} \hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1} (\ddot{\mathbf{r}} - \hat{\mathbf{B}}_{N,MLC} \ddot{\mathbf{f}}),$$

$$\hat{\boldsymbol{\alpha}}_{N,MLC} = \frac{1}{\kappa + 1} \left(\ddot{\mathbf{r}} - \hat{\mathbf{B}}_{N,MLC} \ddot{\mathbf{f}} - \hat{\mathbf{A}}_{N,MLC} \hat{\boldsymbol{\lambda}}_{missMLC} \right),$$

in which $\hat{\boldsymbol{\Sigma}}_{N,MLC} = \hat{\mathbf{A}}_{N,MLC} \hat{\mathbf{A}}'_{N,MLC} + \hat{\mathbf{C}}_{N,MLC}$, $\boldsymbol{\Sigma}_{rf} = \frac{1}{T} \sum_{t=1}^T \ddot{\mathbf{r}}_t \mathbf{f}'_t$, $\boldsymbol{\Sigma}_{ff} = \frac{1}{T} \sum_{t=1}^T \mathbf{f}_t \mathbf{f}'_t$, $\ddot{\mathbf{r}}_t = (\mathbf{r}_t - r_f \mathbf{1}_N)$, $\ddot{\mathbf{r}} = (\bar{\mathbf{r}} - r_f \mathbf{1}_N)$, $\mathbf{f}_t = (\mathbf{f}_t - r_f \mathbf{1}_K)$, $\ddot{\mathbf{f}} = (\bar{\mathbf{f}} - r_f \mathbf{1}_K)$, and

$$\mathbf{G}_N = \frac{1}{(\kappa + 1)} \mathbf{I}_N + \frac{\kappa}{(\kappa + 1)} \hat{\mathbf{A}}_{N,MLC} (\hat{\mathbf{A}}'_{N,MLC} \hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1} \hat{\mathbf{A}}_{N,MLC})^{-1} \hat{\mathbf{A}}'_{N,MLC} \hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1}.$$

Note that $\hat{\mathbf{A}}_{N,MLC}$ and $\hat{\mathbf{C}}_{N,MLC}$ do not admit a closed-form solution and, as before, $\hat{\boldsymbol{\lambda}}_{MLC}$ and $\hat{\boldsymbol{\Omega}}_{MLC}$ coincide with the sample mean and sample covariance of the factors \mathbf{f}_t .

(ii) For $\kappa = 0$ and the APT restriction not holding exactly, one can identify only $\boldsymbol{\alpha}_0 = \mathbf{A}_0 \boldsymbol{\lambda}_{miss0} + \mathbf{a}_0$ but not the two components separately, and one obtains

$$\hat{\boldsymbol{\alpha}}_{N,MLC} = \ddot{\mathbf{r}} - \hat{\mathbf{B}}_{N,MLC} \ddot{\mathbf{f}},$$

and the expression for $\text{vec}(\hat{\mathbf{B}}_{N,MLC})$ can be obtained by setting $\kappa = 0$ in the expression above. The expressions for $\hat{\boldsymbol{\lambda}}_{MLC}$ and $\hat{\boldsymbol{\Omega}}_{MLC}$ are unchanged, and, as before, the expressions for the estimators of $\hat{\mathbf{A}}_{N,MLC}$ and $\hat{\mathbf{C}}_{N,MLC}$ do not admit a closed-form solution.

5 Evaluating Out-of-Sample Portfolio Performance

In this section, we illustrate using simulated data the theoretical results described above in Section 4. This section is divided into three parts. In the first part, we explain the design of our simulation experiment and how we evaluate the out-of-sample performance of a variety of portfolio strategies. In the second part, we investigate the gains from treating misspecification in the beta component of returns, which is done by using our insights in Section 4.2 from the asymptotic analysis of portfolio weights as the number of assets increases. In the third part, we evaluate the gains from treating misspecification in the

alpha component of returns, which is done by imposing the APT restriction, as discussed in Section 4.3.

5.1 Simulation design and performance evaluation

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

$$\mathbf{r}_t = \boldsymbol{\alpha}_N + \boldsymbol{\beta}_N f_t + \boldsymbol{\varepsilon}_t.$$

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 variance equal to $\Omega = (\frac{16}{\sqrt{12 \times 100}})^2$; both λ and Ω are scalars because the investor assumes there is only a single factor.

We consider two environments, one with bounded residual variation ($\boldsymbol{\alpha} = \mathbf{a}$) and the other with a special form of unbounded residual variation ($\boldsymbol{\alpha} = \mathbf{A}\boldsymbol{\lambda}_{miss}$ with $\mathbf{a} = \mathbf{0}$). Both \mathbf{a} and \mathbf{A} are generated from an IID multivariate Gaussian distribution with mean $\mathbf{0}$ and covariance matrix equal to $\sigma_\alpha^2 \mathbf{I}_N$, with $\sigma_\alpha = \frac{1}{4} \frac{5}{\sqrt{12 \times 100}}$, which, in order to be conservative and consistent with the empirical data for individual stocks, is one quarter of the value used in MacKinlay and Pástor (2000). In both cases, $\boldsymbol{\varepsilon}_t$ is IID with a multivariate Gaussian distribution with monthly mean of $\mathbf{0}$. In the bounded-variation case, the monthly covariance matrix is $\boldsymbol{\Sigma} = \sigma_\varepsilon^2 \mathbf{I}_N$; in contrast, in the unbounded-variation case, the monthly covariance matrix is $\boldsymbol{\Sigma} = \boldsymbol{\alpha}\boldsymbol{\alpha}' + \sigma_\varepsilon^2 \mathbf{I}_N$. In both cases, $\sigma_\varepsilon = \frac{20}{\sqrt{12 \times 100}}$. Observe that to ensure identification one must set the variance of the missing factor equal to one, which implies that $\boldsymbol{\lambda}_{miss} = \frac{0.75}{\sqrt{12}}$ is the Sharpe ratio for the missing factor.

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 portfolio strategies described below. We then compute the return on each portfolio based

on the realized return in the 121st month following the estimation window. Using 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 compare using the out-of-sample Sharpe ratio the performance of the following portfolio strategies: (1) EW, the equal-weighted portfolio where the weights are equal to $1/N$ and no estimation is required; (2) GMV, the global-minimum-variance portfolio based on the sample covariance matrix; (3) MV, the mean-variance portfolio based on plugging in the sample mean and sample covariance matrix; (4) MLU, the ML-based unconstrained portfolio that is based on the sample mean and the wrong covariance matrix implied by the factor model, $\beta\beta'\Omega + \sigma_\varepsilon^2\mathbf{I}$; and (5) MLC, the ML-based constrained portfolio based on the sample mean but factor covariance matrix of $\beta\beta'\Omega + \Sigma$ and imposing the APT constraint in which δ is obtained using ten-fold cross validation.

5.2 Model misspecification in the beta component of returns

In this section, we study the effect on portfolio performance of model misspecification in the beta component of returns; in particular, misspecification of the risk premium, which is given by the mean of the factor in excess of the risk-free rate, λ .

In this analysis, we use the data-generating model described above, in which the sample excess mean return on the single factor is $\hat{\lambda}$ and its variance is $\hat{\Omega}$. In our experiment, we assume that the investor has the incorrect view that λ is either half or double of its sample estimator, $\hat{\lambda}$. Then, we compare two versions of the MLC strategy described above (strategy (5)).³⁹ The first version of this strategy, MLC_{views} is based on the (incorrect) view of the factor risk premia, λ . The second version of the strategy, MLC_{robust} , is based on our asymptotic analysis in Theorem 4.4, which is robust to misspecification of λ . The results of this comparison are given in Table 1.

In Panel A of Table 1, we consider the case where the pricing error has bounded variation; that is, $\alpha = \mathbf{a}$. The first row of this table gives the base case, where there are

³⁹Recall, that this is the ML-based constrained portfolio based on the sample mean but factor covariance matrix of $\beta\beta'\Omega + \Sigma$ and imposing the APT constraint in which δ is obtained using ten-fold cross validation.

Table 1: Out-of-Sample Sharpe Ratios With Beta Misspecification

This table reports the annualized Sharpe ratios averaged across $M = 100$ Monte Carlo simulations for five portfolios: (1) EW, the equal-weighted portfolio; (2) GMV, the global-minimum-variance portfolio, based on the sample covariance matrix; (3) MV, the mean-variance portfolio based on the sample mean and sample covariance matrix; (4) MLC_{views} , the ML-based constrained mean-variance portfolio based on the (incorrect) views of the investor regarding λ or Ω . (5) MLC_{robust} , the ML-based constrained mean-variance portfolio that is robust to model misspecification based on the results from our asymptotic analysis. The t -statistic in the penultimate column is for the difference in the Sharpe ratio of the MLC_{robust} portfolio and the equally weighted (EW) portfolio, and the t -statistic in the last column is for the difference in the Sharpe ratio of the MLC_{robust} portfolio and the MLC_{views} portfolio.

	EW	GMV	MV	MLC <i>views</i>	MLC <i>robust</i>	t -stat wrt EW	t -stat wrt MLC_{views}
<i>Panel A: Pricing errors unrelated to factors ($\alpha = \mathbf{a}$)</i>							
Base case ($\lambda = \hat{\lambda}, N = 100$)	0.48	0.14	0.02	1.96	1.96	15.74	—
Low λ (half of base case, $N = 100$)	0.48	0.14	0.02	0.99	1.96	16.47	8.60
High λ (double of base case, $N = 100$)	0.48	0.14	0.02	1.72	1.96	16.47	3.28
Low λ (half of base case, $N = 5$)	0.42	0.40	0.45	0.29	0.42	0.01	2.53
High λ (double of base case, $N = 5$)	0.42	0.40	0.45	0.30	0.42	0.01	2.32
<i>Panel B: Pricing errors related to factors ($\alpha = \mathbf{Am}$)</i>							
Base case ($\lambda = \hat{\lambda}, N = 100$)	0.48	0.06	0.00	0.54	0.54	2.75	—
Low λ (half of base case, $N = 100$)	0.48	0.06	0.00	0.50	0.54	2.57	2.53
High λ (double of base case, $N = 100$)	0.48	0.06	0.00	0.51	0.54	2.57	2.23
Low λ (half of base case, $N = 5$)	0.42	0.36	0.26	0.41	0.33	-2.48	-2.35
High λ (double of base case, $N = 5$)	0.42	0.36	0.26	0.47	0.33	-2.48	-4.69

$N = 100$ risky assets and the investor has the correct view about $\hat{\lambda}$, which implies that $MLC_{views} = MLC_{robust}$, which for our simulated data is 1.96. In contrast, the Sharpe ratio of the EW portfolio is only 0.55 p.a.; the GMV portfolio has an even lower Sharpe ratio of 0.20 p.a., which is a consequence of the error in estimating the sample covariance matrix of returns. The MV portfolio has the lowest 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.

Now, we study the effect of having an incorrect view about λ . We see that, for the case of $N = 100$, the Sharpe ratio from using the strategy based on the incorrect views regarding λ , denoted by MLC_{views} , is significantly smaller than that from the strategy based on our asymptotic analysis, MLC_{robust} , which is immune to misspecification in the beta component of returns. Moreover, we see that to achieve these gains based on our “large N ” analysis,

it is sufficient to have $N = 100$ risky assets. 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 0.42, which is the same as that for the equally weighted (EW) portfolio.

The insights that emerge from analyzing Panel B of Table 1, where we study the case where the pricing error has unbounded variation, $\boldsymbol{\alpha} = \mathbf{A} \boldsymbol{\lambda}_{miss}$, are similar to those from Panel A.

5.3 Model misspecification in the alpha component of returns

In this section, we study the gains from mitigating model misspecification in the alpha component of returns. In addition to the “base case” of the simulations described above, we look at three variations. In the first, we look at the case where the risk premium λ on the observed factor, which one may interpret as the return on the market portfolio, 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 σ_ε is half of its base-case value, which corresponds to the case where there is lower residual risk in returns, and hence, estimation error is likely to be smaller. In the third variation, we look at the case where σ_α is a one-half of its base-case value, which corresponds to the case where the alphas have smaller dispersion; this experiment 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).

Table 2 gives the annualized Sharpe ratio for the five portfolio strategies listed above. We start by looking at the base case in Panel A. As discussed in the previous section, the EW, GMV, and MV portfolios perform poorly. Examining the MLU portfolio, which relies on the sample mean but uses the covariance matrix $\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\Omega} + \sigma_\varepsilon^2\mathbf{I}$, we see that it performs much better than the EW portfolio: its Sharpe ratio is higher than that of the EW portfolio—1.41 p.a. instead of 0.48 p.a. The MLC portfolio, which relies on the sample mean but uses the covariance matrix $\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\Omega} + \boldsymbol{\Sigma}$ along with the APT constraint, performs even better: it has

Table 2: Out-of-Sample Sharpe Ratios With Alpha Misspecification

This table reports the annualized Sharpe ratios averaged across $M = 100$ Monte Carlo simulations for five portfolios: (1) EW, the equal-weighted portfolio; (2) GMV, the global-minimum-variance portfolio, based on the sample covariance matrix; (3) MV, the mean-variance portfolio based on the sample mean and sample covariance matrix; (4) MLU, the ML-based unconstrained mean-variance portfolio based on the sample mean and covariance matrix implied by the factor model, $\beta\beta'\Omega + \sigma^2\mathbf{I}$; and (5) MLC, the ML-based constrained mean-variance portfolio based on the sample mean but factor covariance matrix of $\beta\beta'\Omega + \Sigma$ and imposing the APT constraint in which δ 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.

	EW	GMV	MV	MLU	MLC	t -stat wrt EW	t -stat wrt MLU
<i>Panel A: Pricing errors unrelated to factors ($\alpha = \mathbf{a}$)</i>							
Base case	0.48	0.14	0.02	1.41	1.95	15.74	7.77
Low λ (half of base case)	0.23	0.11	-0.01	0.84	1.17	9.81	5.18
Low σ_ϵ (half of base case)	0.48	0.21	0.03	4.04	4.85	25.34	2.27
Low σ_α (half of base case)	0.48	0.09	-0.01	0.44	0.60	3.17	4.18
<i>Panel B: Pricing errors related to factors ($\alpha = \mathbf{A}\lambda_{miss}$)</i>							
Base case	0.48	0.06	0.00	0.19	0.57	2.75	7.31
Low λ (half of base case)	0.23	0.02	-0.03	0.17	0.28	1.38	1.88
Low σ_ϵ (half of base case)	0.48	0.02	-0.03	0.11	0.60	3.54	8.36
Low σ_α (half of base case)	0.48	0.06	0.01	0.26	0.56	2.40	6.27

a Sharpe ratio of 1.95 p.a., which is significantly higher than that of the EW and MLU strategies.

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 dispersion in alpha and earns only the average alpha, which is zero. On the other hand, MLC can exploit fully the presence of pricing errors, which explains why it outperforms the EW portfolio. Furthermore, the superior performance of the MLC portfolio relative to the MLU portfolio highlights the importance of the APT restriction.

The second row of Panel A considers the case where the risk premium on the observed factor, λ , 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, but that of the MLC strategy decreases less. 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 the Sharpe ratios of all the other strategies that rely on estimated return moments improve. One might wonder when the EW strategy will outperform the MLC strategy: this happens when the alphas are small in absolute value. In the last row of Panel A, we consider the case where the dispersion of alphas is only a half 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, though still significantly greater than that of the EW and MLU portfolios.

Panel B of Table 2 shows that the insights described above are similar for the case where the pricing errors are related to factors.

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.

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. We use simulations to demonstrate that these theoretical findings lead to an improvement in out-of-sample portfolio performance that is both economically and statistically significant.

A Proofs for Theorems

Note that Theorems 3.1 and 3.2 are derived in both Huberman (1982) and Ingersoll (1984); therefore, we do not include the proofs for these theorems. The proofs for all other theorems in the main text of the manuscript are given below, starting with some preliminary lemmas. For the theorems stated in the Online Appendix, the proofs follow directly after the statement of each theorem.

A.1 Lemma on decomposition of the Sharpe ratio

Lemma A.1. *Consider the portfolio weights $\mathbf{w}_N = \mathbf{w}_{N,1} + \mathbf{w}_{N,2}$ such that $\mathbf{w}_{N,1}$ is orthogonal to $\mathbf{w}_{N,2}$:*

$$\mathbf{w}'_{N,1} \mathbf{V}_N \mathbf{w}_{N,2} = 0.$$

Then, defining $\text{SR}_i = \mathbf{w}'_{N,i}(\boldsymbol{\mu}_N - r_f \mathbf{1}_N) / (\mathbf{w}'_{N,i} \mathbf{V}_N \mathbf{w}_{N,i})^{1/2}$ and letting SR denote the Sharpe ratio of the portfolio \mathbf{w}_N , we have:

$$\text{SR}^2 = \frac{(\mathbf{w}'_N (\boldsymbol{\mu}_N - r_f \mathbf{1}_N))^2}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} \leq (\text{SR}_1)^2 + (\text{SR}_2)^2.$$

Finally, equality holds if and only if:

$$\frac{\mathbf{w}'_{N,1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N)}{\mathbf{w}'_{N,1} \mathbf{V}_N \mathbf{w}_{N,1}} = \frac{\mathbf{w}'_{N,2} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N)}{\mathbf{w}'_{N,2} \mathbf{V}_N \mathbf{w}_{N,2}}$$

Proof. Defining for simplicity

$$\mu_i - r_f = \mathbf{w}'_{N,i} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N), \quad \text{and} \quad \sigma_i^2 = \mathbf{w}'_{N,i} \mathbf{V}_N \mathbf{w}_{N,i},$$

we have:

$$\begin{aligned} \text{SR}^2 &= \frac{(\mu_1 - r_f)^2}{\sigma_1^2} \frac{\sigma_1^2}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} + \frac{(\mu_2 - r_f)^2}{\sigma_2^2} \frac{\sigma_2^2}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} + 2 \frac{(\mu_1 - r_f)(\mu_2 - r_f)}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} \\ &= \frac{(\mu_1 - r_f)^2}{\sigma_1^2} + \frac{(\mu_2 - r_f)^2}{\sigma_2^2} \\ &\quad + \left[\frac{(\mu_1 - r_f)^2}{\sigma_1^2} \left(-1 + \frac{\sigma_1^2}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} \right) + \frac{(\mu_2 - r_f)^2}{\sigma_2^2} \left(-1 + \frac{\sigma_2^2}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} \right) + 2 \frac{(\mu_1 - r_f)(\mu_2 - r_f)}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} \right]. \end{aligned}$$

Using the orthogonality of $\mathbf{w}_{N,1}$ and $\mathbf{w}_{N,2}$, we have that $\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N = \mathbf{w}'_{N,1} \mathbf{V}_N \mathbf{w}_{N,1} + \mathbf{w}'_{N,2} \mathbf{V}_N \mathbf{w}_{N,2}$, so that the term in square-brackets can be rewritten as

$$- \frac{(\mu_1 - r_f)^2}{\sigma_1^2} \frac{\sigma_1^2}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} - \frac{(\mu_2 - r_f)^2}{\sigma_2^2} \frac{\sigma_2^2}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} + 2 \frac{(\mu_1 - r_f)(\mu_2 - r_f)}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N}$$

$$\begin{aligned}
&= \frac{1}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} \left(-\mu_1^2 \frac{\sigma_2^2}{\sigma_1^2} - \mu_2^2 \frac{\sigma_1^2}{\sigma_2^2} + 2(\mu_1 - r_f) \frac{\sigma_2}{\sigma_1} (\mu_2 - r_f) \frac{\sigma_1}{\sigma_2} \right) \\
&= -\frac{1}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} \left((\mu_1 - r_f) \frac{\sigma_2}{\sigma_1} - (\mu_2 - r_f) \frac{\sigma_1}{\sigma_2} \right)^2.
\end{aligned}$$

Hence,

$$\begin{aligned}
\text{SR}^2 &= \frac{(\mu_1 - r_f)^2}{\sigma_1^2} + \frac{(\mu_2 - r_f)^2}{\sigma_2^2} - \frac{1}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} \left((\mu_1 - r_f) \frac{\sigma_2}{\sigma_1} - (\mu_2 - r_f) \frac{\sigma_1}{\sigma_2} \right)^2 \\
&\leq \frac{(\mu_1 - r_f)^2}{\sigma_1^2} + \frac{(\mu_2 - r_f)^2}{\sigma_2^2} = (\text{SR}_1)^2 + (\text{SR}_2)^2.
\end{aligned}$$

Equality holds if and only if

$$\left((\mu_1 - r_f) \frac{\sigma_2}{\sigma_1} - (\mu_2 - r_f) \frac{\sigma_1}{\sigma_2} \right)^2 = 0,$$

which, in turn, can be rearranged as

$$\frac{(\mu_1 - r_f)}{\sigma_1^2} = \frac{(\mu_2 - r_f)}{\sigma_2^2}.$$

■

A.2 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.

Lemma A.2 (Extension of Roll (1980) to the case with a risk-free asset). *Let \mathbf{w}^x be any, possibly inefficient, portfolio. Let \mathbf{w}^z be the portfolio that satisfies*

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

and

$$\boldsymbol{\mu}' \mathbf{w}^z + (1 - \mathbf{1}'_N \mathbf{w}^z) r_f = \mu^z,$$

for a given target mean μ^z . Then,

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

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

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

$$L(\mathbf{w}^z, \lambda_1, \lambda_2) = (\mathbf{w}^z)' \mathbf{V} \mathbf{w}^z - \lambda_1 ((\mathbf{w}^x)' \mathbf{V} \mathbf{w}^z) - \lambda_2 (\boldsymbol{\mu}' \mathbf{w}^z + (1 - \mathbf{1}'_N \mathbf{w}^z) r_f - \mu^z),$$

with first-order conditions:

$$2\mathbf{V}\mathbf{w}^z = \left(\mathbf{V}\mathbf{w}^x, (\boldsymbol{\mu} - \mathbf{1}_N r_f) \right) \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}.$$

Pre-multiplying both sides by $2^{-1} \left(\mathbf{V}\mathbf{w}^x, (\boldsymbol{\mu} - \mathbf{1}_N r_f) \right)' \mathbf{V}^{-1}$ gives:

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

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

Remark A.2.1. When \mathbf{w}^x is efficient, then $\mathbf{w}^z = \mathbf{0}$, which implies that the zero-beta portfolio to \mathbf{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 A.2.1 (Extension of Corollary 3 of Roll (1980) to the case with a risk-free asset). *There is a weighted average of, possibly inefficient, portfolio \mathbf{w}^x with a corresponding minimum-variance orthogonal portfolio \mathbf{w}^z that produces an efficient portfolio.*

Remark A.2.2. The above theorem implies that the subspace of minimum-variance portfolios orthogonal to \mathbf{w}^x is given by the two lines described by the expression below:

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

Notice from the equation above and the dashed and dotted lines in Figure 3 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 \mathbf{w}^x equal to the Sharpe ratio of the mean-variance portfolio \mathbf{w}^{mv} in the equation above.

⁴⁰Huang 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.

A.3 Equivalent representations for the portfolio \mathbf{w}^α

The portfolio \mathbf{w}^α in (18) has four equivalent representations, which we will use throughout the paper to gain insights for the portfolio weights when the number of assets is large. These four representations are given in the next lemma.

Lemma A.3 (Equivalent representations for \mathbf{w}^α).

$$\begin{aligned}\mathbf{w}^\alpha &= \boldsymbol{\Sigma}^{-1}\check{\boldsymbol{\alpha}} \\ &= \mathbf{V}^{-1}\check{\boldsymbol{\alpha}} \\ &= \check{\mathbf{V}}\check{\boldsymbol{\alpha}} \\ &= \check{\mathbf{V}}\check{\boldsymbol{\mu}},\end{aligned}$$

where

$$\check{\mathbf{V}} = [\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{B}(\mathbf{B}'\boldsymbol{\Sigma}^{-1}\mathbf{B})^{-1}\mathbf{B}'\boldsymbol{\Sigma}^{-1}]. \quad (\text{A1})$$

Proof. The first equality is the definition of the \mathbf{w}^α portfolio in (18); the second equality follows from the orthogonality of the projection in (6); the third and fourth equalities follow from the definition of \mathbf{V} in (2), the definition of $\check{\mathbf{V}}$ (A1), and the fact that

$$\check{\mathbf{V}}\mathbf{B} = \mathbf{0}.$$

Remark A.3.1. It is useful to discuss the relation between \mathbf{V}^{-1} and $\check{\mathbf{V}}$. Note that the Sherman-Morrison-Woodbury formula implies:

$$\mathbf{V}^{-1} = [\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{B}(\boldsymbol{\Omega}^{-1} + \mathbf{B}'\boldsymbol{\Sigma}^{-1}\mathbf{B})^{-1}\mathbf{B}'\boldsymbol{\Sigma}^{-1}].$$

Setting $\boldsymbol{\Omega}^{-1} = \mathbf{0}$ in the expression above then leads to the expression for $\check{\mathbf{V}}$ in (A1); that is, when $\boldsymbol{\Omega}$ takes arbitrarily large value, then \mathbf{V}^{-1} tends toward $\check{\mathbf{V}}$. ■

A.4 APT restriction in terms of projection errors

In the lemma below, we show that the APT restriction can be expressed as a function of either the projection errors $\check{\boldsymbol{\alpha}}_N$ or their (element by element) limit $\boldsymbol{\alpha}_N$.

Lemma A.4 (Equivalence of APT constraint in terms of $\check{\boldsymbol{\alpha}}_N$ and $\boldsymbol{\alpha}_N$).

$$\check{\boldsymbol{\alpha}}_N \boldsymbol{\Sigma}_N^{-1} \check{\boldsymbol{\alpha}}_N \leq \delta < \infty \text{ implies } \boldsymbol{\alpha}_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N \leq \delta < \infty.$$

Proof. Note that from (4) and (5) we have:

$$\begin{aligned}\check{\alpha}_N &= (\boldsymbol{\mu}_N - r_f \mathbf{1}_N) - \mathbf{B}_N \check{\boldsymbol{\lambda}} \\ &= (\mathbf{I}_N - \mathbf{B}_N (\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1}) (\boldsymbol{\alpha}_N + \mathbf{B}_N \boldsymbol{\lambda}) \\ &= (\mathbf{I}_N - \mathbf{B}_N (\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1}) \boldsymbol{\alpha}_N.\end{aligned}$$

Because

$$\boldsymbol{\Sigma}_N^{-1} \check{\boldsymbol{\alpha}}_N = \boldsymbol{\Sigma}_N^{-1} (\mathbf{I}_N - \mathbf{B}_N (\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1}) \boldsymbol{\alpha}_N = \check{\mathbf{V}}_N \boldsymbol{\alpha}_N,$$

it follows that

$$\begin{aligned}\check{\boldsymbol{\alpha}}'_N \boldsymbol{\Sigma}_N^{-1} \check{\boldsymbol{\alpha}}_N &= \check{\boldsymbol{\alpha}}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^{-1} \check{\boldsymbol{\alpha}}_N \\ &= \check{\boldsymbol{\alpha}}'_N \check{\mathbf{V}}_N \boldsymbol{\Sigma}_N \check{\mathbf{V}}_N \boldsymbol{\alpha}_N \\ &= \check{\boldsymbol{\alpha}}'_N \check{\mathbf{V}}_N \boldsymbol{\alpha}_N,\end{aligned}\tag{A2}$$

where $\check{\mathbf{V}}$ is defined in (A1). Therefore, the condition in (A3) below,

$$\check{\boldsymbol{\alpha}}'_N \boldsymbol{\Sigma}_N^{-1} \check{\boldsymbol{\alpha}}_N \leq \delta < \infty,\tag{A3}$$

implies from (A2) that

$$\check{\boldsymbol{\alpha}}'_N \check{\mathbf{V}}_N \boldsymbol{\alpha}_N = \check{\boldsymbol{\alpha}}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N - \check{\boldsymbol{\alpha}}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N \leq \delta < \infty.\tag{A4}$$

■

Remark A.4.1. Because $\check{\boldsymbol{\alpha}}'_N \check{\mathbf{V}}_N \boldsymbol{\alpha}_N \geq 0$, therefore (A4) implies that

$$0 \leq \check{\boldsymbol{\alpha}}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N \leq \check{\boldsymbol{\alpha}}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N \leq \delta < \infty,$$

implying that (7), as well as the equation below, hold by no-arbitrage.

$$\check{\boldsymbol{\alpha}}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N \leq \delta < \infty.$$

A.5 Proof of Theorem 3.3

By Chamberlain and Rothschild (1983, Theorem 4) the residual covariance matrix satisfies

$$\boldsymbol{\Sigma}_N = \mathbf{A}_N \mathbf{A}'_N + \mathbf{C}_N$$

where \mathbf{C}_N is a positive definite matrix with eigenvalues uniformly bounded by $g_{p+1N}(\boldsymbol{\Sigma}_N)$.

By the Sherman-Morrison-Woodbury decomposition

$$\boldsymbol{\Sigma}_N^{-1} = \mathbf{C}_N^{-1} - \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1}.$$

Therefore, by substitution,

$$\begin{aligned}
\alpha'_N \Sigma_N^{-1} \alpha_N &= \alpha'_N \mathbf{C}_N^{-1} \alpha_N - \alpha'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \alpha_N \\
&= (\mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{a}_N)' \mathbf{C}_N^{-1} (\mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{a}_N) \\
&\quad - (\mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{a}_N)' \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} (\mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{a}_N) \\
&= \boldsymbol{\lambda}'_{miss} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} - \boldsymbol{\lambda}'_{miss} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} \\
&\quad + \mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{a}_N - \mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{a}_N \\
&\quad + 2\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} - \mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss}.
\end{aligned}$$

We now show that $\alpha'_N \Sigma_N^{-1} \alpha_N$ 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:

$$\begin{aligned}
&\boldsymbol{\lambda}'_{miss} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} - \boldsymbol{\lambda}'_{miss} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} \\
&= \boldsymbol{\lambda}'_{miss} (\mathbf{I}_N - \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1}) \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} \\
&= \boldsymbol{\lambda}'_{miss} (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} \leq \boldsymbol{\lambda}'_{miss} \boldsymbol{\lambda}_{miss},
\end{aligned}$$

because $\mathbf{I}_p - (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N$ is positive semidefinite. Next, for the second term

$$\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{a}_N \leq \mathbf{a}'_N \mathbf{a}_N g_{NN}^{-1}(\mathbf{C}_N).$$

Now, the j th element of $\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{a}_N$, obtained by considering the j th column of \mathbf{A}_N , for every $1 \leq j \leq p$, satisfies

$$|\mathbf{a}'_N \mathbf{C}_N^{-1} g_{jN}^{\frac{1}{2}} \mathbf{v}_{jN}| \leq g_{jN}^{\frac{1}{2}} (\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{a}_N)^{\frac{1}{2}} (\mathbf{v}'_{jN} \mathbf{C}_N^{-1} \mathbf{v}_{jN})^{\frac{1}{2}},$$

where for simplicity we set $\mathbf{v}_{jN} = \mathbf{v}_{jN}(\Sigma_N)$, $g_{jN} = g_{jN}(\Sigma_N)$. Because

$$\begin{aligned}
&\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{a}_N \\
&\leq g_{1p} ((\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1}) (\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{a}_N) (\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{a}_N) \\
&\leq g_{1p} ((\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1}) (\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{a}_N) (\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{a}_N) \\
&= g_{pp}^{-1} (\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N) (\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{a}_N) (\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{a}_N) \leq \delta < \infty.
\end{aligned}$$

since the (i, j) th element, for every $1 \leq i, j \leq p$, of $(\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)$ is equal to $g_{iN}^{\frac{1}{2}} g_{jN}^{\frac{1}{2}} \mathbf{v}'_{iN} \mathbf{C}_N^{-1} \mathbf{v}_{jN}$.

Concerning the third term, it turns out that it will converges to zero. In fact:

$$\begin{aligned}
&|2\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} - 2\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss}| \\
&= 2|\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \boldsymbol{\lambda}_{miss}| \\
&\leq (\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{a}_N)^{\frac{1}{2}} (\boldsymbol{\lambda}'_{miss} (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \boldsymbol{\lambda}_{miss})^{\frac{1}{2}} \leq \delta g_{pN}^{-\frac{1}{2}} \rightarrow 0.
\end{aligned}$$

■

A.6 Proof of Theorem 4.1

Given that $\boldsymbol{\mu}_N - r_f \mathbf{1}_N = \check{\boldsymbol{\alpha}} + \mathbf{B}_N \check{\boldsymbol{\lambda}}$

$$\begin{aligned}
\mathbf{w}_N^{\text{mv}} &= \frac{1}{\gamma} \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - \mathbf{1}_N r_f) \\
&= \frac{1}{\gamma} \mathbf{V}_N^{-1} \check{\boldsymbol{\alpha}}_N + \frac{1}{\gamma} \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}} \\
&= \frac{\gamma^\alpha}{\gamma} \frac{1}{\gamma^\alpha} \check{\mathbf{V}}_N \check{\boldsymbol{\alpha}}_N + \frac{\gamma^\beta}{\gamma} \frac{1}{\gamma^\beta} \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}} \\
&= \phi^\alpha \mathbf{w}_N^\alpha + \phi^\beta \mathbf{w}_N^\beta.
\end{aligned}$$

Moreover

$$\begin{aligned}
\mathbf{w}_N^{\alpha'} \mathbf{V}_N \mathbf{w}_N^\beta &= \frac{1}{\gamma^\beta} \mathbf{w}_N^{\alpha'} \mathbf{V}_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}} \\
&= \frac{1}{\gamma^\beta} \mathbf{w}_N^{\alpha'} \mathbf{B}_N \check{\boldsymbol{\lambda}} \\
&= \frac{1}{\gamma^\alpha \gamma^\beta} \check{\boldsymbol{\alpha}}_N' \check{\mathbf{V}}_N \mathbf{B}_N \check{\boldsymbol{\lambda}} \\
&= 0,
\end{aligned}$$

because $\check{\mathbf{V}}$ is orthogonal to \mathbf{B} . Similarly,

$$\begin{aligned}
\mathbf{w}_N^{\alpha'} \boldsymbol{\Sigma}_N \mathbf{w}_N^\beta &= \frac{1}{\gamma^\beta} \mathbf{w}_N^{\alpha'} \boldsymbol{\Sigma}_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}} \\
&= \frac{1}{\gamma^\beta} \mathbf{w}_N^{\alpha'} \mathbf{B}_N (\mathbf{I}_K - (\boldsymbol{\Omega}^{-1} + \mathbf{B}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N) \check{\boldsymbol{\lambda}} \\
&= \frac{1}{\gamma^\alpha \gamma^\beta} \check{\boldsymbol{\alpha}}_N' \check{\mathbf{V}}_N \mathbf{B}_N (\mathbf{I}_K - (\boldsymbol{\Omega}^{-1} + \mathbf{B}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N) \check{\boldsymbol{\lambda}} \\
&= 0.
\end{aligned}$$

We now show that the \mathbf{w}_N^α and the \mathbf{w}_N^β 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 A.2. In particular, we need to verify that \mathbf{w}^β satisfy

$$\mathbf{w}_N^\beta = \left(\mathbf{w}_N^\alpha, \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \right) \begin{pmatrix} (\sigma^\alpha)^2 & \mu^\alpha - r_f \\ \mu^\alpha - r_f & (\text{SR}^{\text{mv}})^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \mu^\beta - r_f \end{pmatrix}.$$

Simple calculations leads to

$$\begin{pmatrix} (\sigma^\alpha)^2 & \mu^\alpha - r_f \\ \mu^\alpha - r_f & (\text{SR}^{\text{mv}})^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \mu^\beta - r_f \end{pmatrix}$$

$$\begin{aligned}
&= \frac{1}{\left((\text{SR}^{\text{mv}})^2 \frac{\check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N}{(\gamma^\alpha)^2} - \frac{(\check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N)^2}{(\gamma^\alpha)^2} \right)} \begin{pmatrix} (\text{SR}^{\text{mv}})^2 & -\frac{\check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N}{\gamma^\alpha} \\ -\frac{\check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N}{\gamma^\alpha} & \frac{\check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N}{(\gamma^\alpha)^2} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \frac{\check{\lambda}'_N \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\lambda}_N}{\gamma^\beta} \\
&= \frac{1}{\left((\text{SR}^{\text{mv}})^2 \frac{\check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N}{(\gamma^\alpha)^2} - \frac{(\check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N)^2}{(\gamma^\alpha)^2} \right)} \begin{pmatrix} -\frac{\check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N}{\gamma^\alpha} \\ \frac{\check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N}{(\gamma^\alpha)^2} \end{pmatrix} \frac{\check{\lambda}'_N \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\lambda}_N}{\gamma^\beta}.
\end{aligned}$$

Given that

$$\left(\frac{\check{\Sigma}_N^{-1} \check{\alpha}_N}{\gamma^\alpha}, \mathbf{V}_N^{-1} (\check{\alpha}_N + \mathbf{B}_N \check{\lambda}_N) \right) \begin{pmatrix} -\frac{\check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N}{\gamma^\alpha} \\ \frac{\check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N}{(\gamma^\alpha)^2} \end{pmatrix} = \frac{\check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N}{(\gamma^\alpha)^2} \mathbf{V}_N^{-1} \mathbf{B}_N \check{\lambda}_N,$$

and

$$\left((\text{SR}^{\text{mv}})^2 \frac{\check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N}{(\gamma^\alpha)^2} - \frac{(\check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N)^2}{(\gamma^\alpha)^2} \right) = \frac{1}{(\gamma^\alpha)^2} (\check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N) (\check{\lambda}'_N \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\lambda}_N),$$

one finally obtains

$$\begin{aligned}
&\left(\mathbf{w}_N^\alpha, \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \right) \begin{pmatrix} (\sigma^\alpha)^2 & \mu^\alpha - r_f \\ \mu^\alpha - r_f & (\text{SR}^{\text{mv}})^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \mu^\beta - r_f \end{pmatrix} \\
&= \frac{1}{\left(\frac{1}{(\gamma^\alpha)^2} (\check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N) (\check{\lambda}'_N \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\lambda}_N) \right)} \frac{\check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N}{(\gamma^\alpha)^2} \mathbf{V}_N^{-1} \mathbf{B}_N \check{\lambda}_N \frac{\check{\lambda}'_N \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\lambda}_N}{\gamma^\beta} = \mathbf{w}_N^\beta.
\end{aligned}$$

■

A.7 Proof of Theorem 4.2

Because $\check{\mathbf{V}}_N \mathbf{V}_N \check{\mathbf{V}}_N = \check{\mathbf{V}}_N \check{\Sigma}_N \check{\mathbf{V}}_N = \check{\mathbf{V}}_N$, it follows that

$$\begin{aligned}
\boldsymbol{\mu}_N^\alpha &= \mathbf{w}_N^{\alpha'} \boldsymbol{\mu}_N + (1 - \mathbf{w}_N^{\alpha'} \mathbf{1}_N) r_f \\
&= \mathbf{w}_N^{\alpha'} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N) + r_f \\
&= \frac{1}{\gamma^\alpha} \check{\alpha}'_N \check{\mathbf{V}}_N (\check{\alpha}_N + \mathbf{B}_N \check{\lambda}) + r_f \\
&= \frac{1}{\gamma^\alpha} \check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N + r_f, \\
(\sigma^\alpha)^2 &= \mathbf{w}_N^{\alpha'} \mathbf{V}_N \mathbf{w}_N^\alpha = \frac{1}{(\gamma^\alpha)^2} \check{\alpha}'_N \check{\mathbf{V}}_N \mathbf{V}_N \check{\mathbf{V}}_N \check{\alpha}_N = \frac{1}{(\gamma^\alpha)^2} \check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N.
\end{aligned}$$

Then use $\check{\alpha}'_N \check{\mathbf{V}}_N \check{\alpha}_N = \check{\alpha}'_N \check{\Sigma}_N^{-1} \check{\alpha}_N$.

Because $\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \check{\boldsymbol{\alpha}}_N = \mathbf{0}$ implies $\mathbf{B}'_N \mathbf{V}_N^{-1} \check{\boldsymbol{\alpha}}_N = \mathbf{0}$, one gets

$$\begin{aligned} \boldsymbol{\mu}_N^\beta &= \mathbf{w}_N^{\beta'} \boldsymbol{\mu}_N + (1 - \mathbf{w}_N^{\beta'} \mathbf{1}_N) r_f = \mathbf{w}_N^{\beta'} (\boldsymbol{\mu}_N - \mathbf{1}_N r_f) + r_f = \frac{1}{\gamma^\beta} \check{\boldsymbol{\lambda}}' \mathbf{B}'_N \mathbf{V}_N^{-1} (\check{\boldsymbol{\alpha}}_N + \mathbf{B}_N \check{\boldsymbol{\lambda}}) + r_f \\ &= \frac{1}{\gamma^\beta} \check{\boldsymbol{\lambda}}' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}} + r_f, \\ (\boldsymbol{\sigma}^\beta)^2 &= \mathbf{w}_N^{\beta'} \mathbf{V}_N \mathbf{w}_N^\beta = \frac{1}{(\gamma^\beta)^2} \check{\boldsymbol{\lambda}}' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{V}_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}} = \frac{1}{(\gamma^\beta)^2} \check{\boldsymbol{\lambda}}' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}}. \end{aligned}$$

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

$$\begin{aligned} \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N &= \mathbf{B}'_N (\boldsymbol{\Sigma}_N^{-1} - \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1}) \mathbf{B}_N \\ &= \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \boldsymbol{\Omega}^{-1}, \\ &= (\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N + \boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}) (\boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \boldsymbol{\Omega}^{-1} \\ &= \boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1} (\boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \boldsymbol{\Omega}^{-1}. \end{aligned}$$

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

Finally,

$$\begin{aligned} (\text{SR}^{\text{mv}})^2 &= (\boldsymbol{\mu}_N - r_f \mathbf{1}_N)' \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \\ &= (\check{\boldsymbol{\alpha}}_N + \mathbf{B}_N \check{\boldsymbol{\lambda}})' \mathbf{V}_N^{-1} (\check{\boldsymbol{\alpha}}_N + \mathbf{B}_N \check{\boldsymbol{\lambda}}) \\ &= \check{\boldsymbol{\alpha}}_N' \mathbf{V}_N^{-1} \check{\boldsymbol{\alpha}}_N + \check{\boldsymbol{\lambda}}_N' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}} \\ &= \check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^{-1} \check{\boldsymbol{\alpha}}_N + \check{\boldsymbol{\lambda}}_N' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}} \\ &= (\text{SR}^\alpha)^2 + (\text{SR}^\beta)^2, \end{aligned}$$

where the third equality follows by the orthogonality. ■

A.8 Proof of Theorem 4.3

Parti (i) By orthogonality $(\boldsymbol{\mu}_N - r_f \mathbf{1}_N)' \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N) = \check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^{-1} \check{\boldsymbol{\alpha}}_N + \check{\boldsymbol{\lambda}}_N' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}}_N$. This implies $0 \leq \phi^\alpha \leq 1, 0 \leq \phi^\beta \leq 1$.

We now consider the unbounded variation case. The bounded variance case will follow by setting $\mathbf{A}_N = \mathbf{0}$. Recall that now $\boldsymbol{\Sigma}_N = \mathbf{A}_N \mathbf{A}'_N + \mathbf{C}_N$ and $\boldsymbol{\alpha}_N = \mathbf{A}_N \boldsymbol{\lambda}_{\text{miss}} + \mathbf{a}_N$. Consider first \mathbf{w}_N^α where its i -th component satisfies:

$$w_{N,i}^\alpha = \mathbf{1}'_{N_i} \mathbf{w}_N^\alpha = \frac{1}{\gamma^\alpha} \mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N - \frac{1}{\gamma^\alpha} \mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N.$$

We deal with the two terms on the right hand side of $w_{N,i}^\alpha$ separately. By the Sherman-Morrison-Woodbury formula $\Sigma_N^{-1} = \mathbf{C}_N^{-1} - \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1}$, obtaining

$$\begin{aligned} \mathbf{1}'_{N_i} \Sigma_N^{-1} \alpha_N &= \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \alpha_N - \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \alpha_N \\ &= \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N \lambda_{miss} - \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \lambda_{miss} \\ &\quad + \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{a}_N - \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{a}_N \\ &= \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \lambda_{miss} \\ &\quad + \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{a}_N - \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{a}_N. \end{aligned}$$

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

$$|\mathbf{1}'_{N_i} \Sigma_N^{-1} \alpha_N| = \mathcal{O} \left(\|\lambda_{miss}\| \frac{\|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N\|}{f(N)} + |\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{a}_N| + \|\mathbf{a}_N\| \frac{\|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N\|}{f^{\frac{1}{2}}(N)} \right),$$

since, under our assumptions, the eigenvalues of $\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N$ and $\mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{B}_N$ have the same behavior.

Along the same lines

$$\begin{aligned} \mathbf{1}'_{N_i} \Sigma_N^{-1} \mathbf{B}_N &= \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{B}_N - \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{B}_N, \\ \mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N &= \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{B}_N - \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{B}_N, \\ \mathbf{B}'_N \Sigma_N^{-1} \alpha_N &= \mathbf{B}'_N \mathbf{C}_N^{-1} \alpha_N - \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \alpha_N \\ &= \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \lambda_{miss} - \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \lambda_{miss} \\ &\quad + \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{a}_N - \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{a}_N \\ &= \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \lambda_{miss} \\ &\quad + \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{a}_N - \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{a}_N. \end{aligned}$$

The same applies to $\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N$ and $\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{B}_N$. Therefore, by using the same arguments above, one obtains

$$|\mathbf{1}'_{N_i} \Sigma_N^{-1} \mathbf{B}_N (\mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \Sigma_N^{-1} \alpha_N| = \mathcal{O} \left(\|\mathbf{a}_N\| \frac{\|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N\| + \|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{B}_N\|}{f^{\frac{1}{2}}(N)} \right).$$

For the \mathbf{w}_N^β portfolio, its i -th component satisfies:

$$\begin{aligned} w_{addN,i}^\beta &= \frac{1}{\gamma^\beta} \mathbf{1}'_{N_i} \Sigma_N^{-1} \mathbf{B}_N \check{\lambda}_N - \frac{1}{\gamma^\beta} \mathbf{1}'_{N_i} \Sigma_N^{-1} \mathbf{B}_N (\Omega^{-1} + \mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N \check{\lambda}_N \\ &= \frac{1}{\gamma^\beta} \mathbf{1}'_{N_i} \Sigma_N^{-1} \mathbf{B}_N (\Omega^{-1} + \mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N)^{-1} \Omega^{-1} \check{\lambda}_N, \end{aligned}$$

and using the above formulae for $\mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N$ and $\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N$ concludes, where we use $\check{\boldsymbol{\lambda}}_N \rightarrow \boldsymbol{\lambda} \neq \mathbf{0}$.

Regarding part (ii), $\mathbf{w}^{\alpha'} \mathbf{w}^\alpha \leq \|\boldsymbol{\Sigma}_N^{-1}\| (\check{\boldsymbol{\alpha}}'_N \boldsymbol{\Sigma}_N^{-1} \check{\boldsymbol{\alpha}}_N) / (\gamma^\alpha)^2 \leq \delta < \infty$. Moreover, $\check{\boldsymbol{\lambda}}'_N \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}}_N = \check{\boldsymbol{\lambda}}'_N \boldsymbol{\Omega}^{-1} (\boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \boldsymbol{\Omega}^{-1} \check{\boldsymbol{\lambda}}_N = \mathbf{O}(1/f(N))$ because $\|\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N\| \leq \|\boldsymbol{\Sigma}_N^{-1}\| \|\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N\|$. This implies $\mathbf{w}^{\lambda'} \mathbf{w}^\lambda = \mathbf{O}(1/f(N)) = o(1)$.

Part (iii) follows from $|\mathbf{1}'_N \mathbf{w}_N^\alpha| \leq \frac{1}{\gamma^\alpha} (\mathbf{1}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{1}_N)^{\frac{1}{2}} (\check{\boldsymbol{\alpha}}_N \boldsymbol{\Sigma}_N^{-1} \check{\boldsymbol{\alpha}}_N)^{\frac{1}{2}}$ with $\mathbf{1}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{1}_N \rightarrow \infty$. In fact $\mathbf{1}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{1}_N \geq N1/(g_{1N}(\boldsymbol{\Sigma}_N)) = N\delta > 0$. On the other hand, $|\mathbf{1}'_N \mathbf{w}_N^\beta| = \frac{1}{\gamma^\beta} |(\mathbf{1}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N) (\boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \boldsymbol{\Omega}^{-1} \check{\boldsymbol{\lambda}}| \leq \delta < \infty$.

Part (iv) follows from part (i) once we show that the coefficients are bounded away from zero. By assumption the limit of $\check{\boldsymbol{\alpha}}'_N \boldsymbol{\Sigma}_N^{-1} \check{\boldsymbol{\alpha}}_N$ is bounded away from zero. On the other hand, $\check{\boldsymbol{\lambda}}'_N \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}}_N \rightarrow \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda} > 0$. Therefore the normalizing constants ϕ^β and ϕ^α remains bounded away from zero also asymptotically and . can be ignored for the purpose of evaluating the limiting behavior of the weights. \blacksquare

Remark A.4.2. For the bounded-residual-variation case, the result in (23) 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:

$$|w_{N,i}^\alpha| = \mathcal{O} \left(|\mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N| + \frac{\|\mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N\|}{f^{\frac{1}{2}}(N)} \right), \quad (\text{A5})$$

$$|w_{N,i}^\beta| = \mathcal{O} \left(\frac{\|\mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N\|}{f(N)} \right), \quad (\text{A6})$$

where $\mathbf{1}_{N_i}$ is an N -dimensional vector in which the i th element is one and the rest of the elements are zero. From equations (A5) and (A6), we see that $w_{N,i}^\alpha$ can dominate $w_{N,i}^\beta$ as the number of assets increases. In particular, $w_{N,i}^\alpha$ dominates $w_{N,i}^\beta$ when the pricing-error term, $|\mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N|$, goes to zero slowly as the number of assets increases. The weights $w_{N,i}^\alpha$ dominate $w_{N,i}^\beta$ also because the second term on the right-hand side of (A5), $\frac{\|\mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N\|}{f^{\frac{1}{2}}(N)}$, dominates the term on the right-hand side of (A6), $\frac{\|\mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N\|}{f(N)}$; this dominance arises because the two terms have different denominators: $f^{\frac{1}{2}}(N)$ instead of $f(N)$.

Recall that the APT bounds the pricing error from above; that is, $\boldsymbol{\alpha}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N \leq \delta < \infty$. But, the APT is silent about whether $\boldsymbol{\alpha}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N$ is bounded away from zero. When this expression is bounded away from zero, then one can show that the ratio $w_{N,i}^\beta/w_{N,i}^\alpha$ always decreases at a rate that is equal or faster than $1/f^{\frac{1}{2}}(N)$.

Remark A.4.3. For the unbounded-residual-variation case (that is, where $\boldsymbol{\alpha}_N = \mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{a}_N$ and $\boldsymbol{\Sigma}_N = \mathbf{A}_N \mathbf{A}'_N + \mathbf{C}_N$), the result in (23) 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:

$$|w_{N,i}^\alpha| = \mathcal{O}\left(|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{a}_N| + \|\mathbf{a}_N\| \frac{\|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N\| + \|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{B}_N\|}{f^{\frac{1}{2}}(N)} + \|\boldsymbol{\lambda}_{miss}\| \frac{\|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N\|}{f(N)}\right); \quad (\text{A7})$$

$$|w_{N,i}^\beta| = \mathcal{O}\left(\frac{\|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N\| + \|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{B}_N\|}{f(N)}\right). \quad (\text{A8})$$

Recall that $\boldsymbol{\lambda}_{miss}$ can be interpreted as the risk-premia on the unobserved factors with loadings \mathbf{A}_N , whereas the vector \mathbf{a}_N represents the pure pricing error that is not associated with a factor structure. The \mathbf{a}_N component dominates the behavior of the portfolio weights $w_{N,i}^\alpha$ in (A7), whereas the risk premia $\boldsymbol{\lambda}_{miss}$ component declines to zero faster. In general, the portfolio weight $w_{N,i}^\beta$ in (A8) declines at the same, fast, rate as the risk-premia $\boldsymbol{\lambda}_{miss}$ component of $w_{N,i}^\alpha$. When \mathbf{a}_N is non-zero then, as before, the \mathbf{w}_N^α portfolio dominates the \mathbf{w}_N^β portfolio across all three norms considered in the theorem above.

Remark A.4.4. Equation (24) shows that the portfolio \mathbf{w}_N^α dominates \mathbf{w}_N^β for large N . Observe that \mathbf{w}_N^α is functionally independent of the factor risk premia, $\boldsymbol{\lambda}$, and the factor covariance matrix, $\boldsymbol{\Omega}$, making it robust to misspecification in the beta component of returns by construction. In contrast, portfolio \mathbf{w}_N^β depends on both $\boldsymbol{\lambda}$ and $\boldsymbol{\Omega}$. Observe that ϕ_N^α is the ratio of $\check{\boldsymbol{\alpha}}'_N \boldsymbol{\Sigma}_N^{-1} \check{\boldsymbol{\alpha}}_N$ to the variance of the return on the market portfolio, $(\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N)$, and hence, can be estimated from the data without using the beta portfolio.

A.9 Proof of Theorem 4.4

Define \mathbf{w}_N^{bench*} to be the benchmark portfolio constructed without imposing the target mean, which satisfies

$$\mathbf{B}' \mathbf{w}_N^{bench*} \rightarrow \mathbf{c}^{bench*} \neq \mathbf{0},$$

and recall that $\mathbf{w}_N^{bench} = \frac{\mu^* - r_f}{(\mathbf{w}_N^{bench*})' (\boldsymbol{\mu} - r_f \mathbf{1}_N)} \mathbf{w}_N^{bench*}$ is the benchmark portfolio normalized to have portfolio expected return equal to μ^* , such that $\mathbf{B}' \mathbf{w}_N^{bench} \rightarrow \mathbf{c}^{bench}$.

Consider first the case when $\mathbf{c}^{bench*} = \delta \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}$, for some scalar δ . Then $\mathbf{B}' \mathbf{w}_N^{bench} \rightarrow \frac{\delta(\mu^* - r_f)}{\delta(\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})} \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda} = \frac{(\mu^* - r_f)}{(\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})} \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}$. Moreover, since $(\mathbf{w}_N^{bench*})' \mathbf{V} \mathbf{w}_N^{bench*} \rightarrow (\mathbf{c}^{bench*})' \boldsymbol{\Omega} \mathbf{c}^{bench*} =$

$\delta^2 \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}$, then $(\mathbf{w}_N^{bench})' \mathbf{V} \mathbf{w}_N^{bench} \rightarrow \frac{(\mu^* - r_f)^2}{\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}}$, yielding

$$(SR_N^{bench})^2 \rightarrow (SR_\infty^{bench})^2 = \frac{(\mu^* - r_f)^2}{(\mu^* - r_f)^2 / (\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})} = \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda} = (SR_\infty^\beta)^2.$$

Therefore,

$$\phi_N^{bench} \sim \phi_N^\beta \sim \frac{\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}}{\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda} + \boldsymbol{\alpha}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\alpha}}$$

and Lemma 1 is satisfied because

$$\begin{aligned} \frac{\phi_N^{bench} (\mathbf{w}_N^{bench})' (\boldsymbol{\mu} - r_f \mathbf{1}_N)}{(\phi_N^{bench})^2 (\mathbf{w}_N^{bench})' \mathbf{V} \mathbf{w}_N^{bench}} &\rightarrow \frac{\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda} + \boldsymbol{\alpha}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\alpha}}{\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}} \frac{\mu^* - r_f}{(\mu^* - r_f)^2 / \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}} \\ &= \frac{(\boldsymbol{\mu} - r_f \mathbf{1}_N)' \mathbf{V}^{-1} (\boldsymbol{\mu} - r_f \mathbf{1}_N)}{\mu^* - r_f}, \end{aligned}$$

and

$$\frac{\phi_N^\alpha \mathbf{w}^{\alpha'} (\boldsymbol{\mu} - r_f \mathbf{1}_N)}{(\phi_N^\alpha)^2 \mathbf{w}^{\alpha'} \mathbf{V} \mathbf{w}^\alpha} \rightarrow \frac{(\boldsymbol{\mu} - r_f \mathbf{1}_N)' \mathbf{V}^{-1} (\boldsymbol{\mu} - r_f \mathbf{1}_N)}{\mu^* - r_f},$$

implying $(SR_N^{mv})^2 \sim (SR_N^\alpha)^2 + (SR_N^{bench})^2$.

If $K = 1$ then $\mathbf{c}^{bench} = \delta \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda} = \phi \frac{\boldsymbol{\lambda}}{\boldsymbol{\Omega}}$ is always satisfied and any benchmark portfolio \mathbf{w}^{bench} satisfying (25) can be combined with \mathbf{w}^α to obtain mean-variance efficiency.

Now consider the case when \mathbf{c}^{bench} is not proportional to $\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1}$. Then,

$$\mathbf{B}' \mathbf{w}^{bench} \rightarrow \frac{(\mu^* - r_f)}{(\mathbf{c}^{bench})' \boldsymbol{\lambda}} \mathbf{c}^{bench} \quad \text{and} \quad (\mathbf{w}^{bench})' \mathbf{V} \mathbf{w}^{bench} \rightarrow \left(\frac{\mu^* - r_f}{(\mathbf{c}^{bench})' \boldsymbol{\lambda}} \right)^2 (\mathbf{c}^{bench})' \boldsymbol{\Omega} \mathbf{c}^{bench}$$

yielding $(SR_N^{bench})^2 \rightarrow \frac{(\mathbf{c}^{bench})' \boldsymbol{\lambda}}{(\mathbf{c}^{bench})' \boldsymbol{\Omega} \mathbf{c}^{bench}} = (SR_\infty^{bench})^2$. However, note that Lemma 1 does not hold, implying that

$$(SR_\infty^\alpha)^2 + (SR_\infty^{bench})^2 = (\boldsymbol{\alpha}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\alpha}) + \frac{((\mathbf{c}^{bench})' \boldsymbol{\lambda})^2}{(\mathbf{c}^{bench})' \boldsymbol{\Omega} \mathbf{c}^{bench}} < (\boldsymbol{\alpha}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\alpha}) + (\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}) = (SR_\infty^{mv})^2,$$

because

$$\frac{((\mathbf{c}^{bench})' \boldsymbol{\lambda})^2}{(\mathbf{c}^{bench})' \boldsymbol{\Omega} \mathbf{c}^{bench}} = \frac{((\mathbf{c}^{bench})' \boldsymbol{\Omega}^{\frac{1}{2}} \boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{\lambda})^2}{(\mathbf{c}^{bench})' \boldsymbol{\Omega} \mathbf{c}^{bench}} < \frac{((\mathbf{c}^{bench})' \boldsymbol{\Omega} \mathbf{c}^{bench})(\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})}{(\mathbf{c}^{bench})' \boldsymbol{\Omega} \mathbf{c}^{bench}} = (\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}).$$

The strict inequality is implied whenever $\boldsymbol{\Omega}^{\frac{1}{2}} \mathbf{c}^{bench}$ and $\boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{\lambda}$ are not proportional, which is implied by \mathbf{c}^{bench} being not being proportional to $\boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}$. ■

Remark A.4.5. For portfolio weights \mathbf{w}^α and \mathbf{w}^β that satisfy the assumptions of Theorem 4.3 we have the following results. Defining $\check{\mathbf{V}}$ as

$$\check{\mathbf{V}} = [\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{B}(\mathbf{B}'\boldsymbol{\Sigma}^{-1}\mathbf{B})^{-1}\mathbf{B}'\boldsymbol{\Sigma}^{-1}],$$

we see that for $N \rightarrow \infty$, the asymptotic means of the excess returns on the portfolios \mathbf{w}^α and \mathbf{w}^β :

$$\mu_\infty^\alpha - r_f = \lim_{N \rightarrow \infty} (\mu_N^\alpha - r_f) = \lim_{N \rightarrow \infty} \frac{1}{\gamma^\alpha} \boldsymbol{\alpha}'_N \check{\mathbf{V}}_N \boldsymbol{\alpha}_N, \quad \mu_\infty^\beta - r_f = \lim_{N \rightarrow \infty} (\mu_N^\beta - r_f) = \frac{1}{\gamma^\beta} \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda},$$

satisfy $0 < (\mu_\infty^\alpha - r_f) < \infty$ and $0 < (\mu_\infty^\beta - r_f) < \infty$. Furthermore, the asymptotic variances and squared Sharpe ratios of the returns on the portfolios \mathbf{w}^α and \mathbf{w}^β are given by the same expressions as the means and satisfy the same properties.

We now interpret the inequality $\mu_\infty^\alpha - r_f > 0$ in terms of the asymptotic variance of the portfolio. Recall that μ_∞^α represents not just the asymptotic mean return but also the asymptotic variance of the portfolio, which is equal to the limit of the idiosyncratic variance of the portfolio because for any N

$$\text{var}(\mathbf{r}'_t \mathbf{w}_N^\alpha) = \text{var}(\boldsymbol{\varepsilon}'_t \mathbf{w}_N^\alpha) = \boldsymbol{\alpha}'_N \check{\mathbf{V}}_N \boldsymbol{\alpha}_N.$$

We will have exact equality, $\mu_\infty^\alpha - r_f = 0$, in two cases. One, where each element of $\boldsymbol{\alpha}_N$ is zero; that is, the case of exact pricing (see Chamberlain (1983, Corollary 1)). Two, where $\boldsymbol{\alpha}_N$ is asymptotically collinear with \mathbf{B}_N ; see footnote 22.

A.10 Proof of Theorem 4.5

The formulae for $\hat{\boldsymbol{\alpha}}_{N,\text{MLC}}$, $\hat{\mathbf{B}}_{N,\text{MLC}}$ and $\hat{\boldsymbol{\Sigma}}_{N,\text{MLC}}$ follow from solving the first order conditions. For $\hat{\boldsymbol{\lambda}}'_{\text{MLC}}$ and $\hat{\boldsymbol{\Omega}}_{\text{MLC}}$, one obtains precisely the sample mean and sample covariance matrix of the \mathbf{f}_t . ■

A.11 Proof of Theorem 4.6

Differentiating the Lagrangian with respect to $\boldsymbol{\lambda}_{\text{miss}}$ and $\boldsymbol{\alpha}$ one obtains the $K+N$ equations (after some algebra):

$$\begin{pmatrix} \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \\ \mathbf{I}_N \end{pmatrix} (\bar{\mathbf{r}} - r_f \mathbf{1}_N - \hat{\mathbf{B}}_N (\bar{\mathbf{f}} - r_f \mathbf{1}_K)) = \begin{pmatrix} \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{A}_N & \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \\ \mathbf{A}_N & (1 + \kappa) \mathbf{I}_N \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda}_{\text{miss}} \\ \hat{\boldsymbol{\alpha}}_N \end{pmatrix}.$$

It is straightforward to see that, because of the APT restriction, $\boldsymbol{\lambda}_{miss0}$ and \mathbf{a}_0 can now be identified separately, as long as $\kappa > 0$. In fact, the above system of linear equations can be solved because the matrix pre-multiplying $\boldsymbol{\lambda}_{miss}$ and $\hat{\mathbf{a}}_N$ is non-singular for every $\kappa > 0$, leading to the closed-form solution:

$$\begin{aligned}\boldsymbol{\lambda}_{miss}^{\hat{}} &= (\mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} (\bar{\mathbf{r}} - r_f \mathbf{1}_N - \mathbf{B}_N (\bar{\mathbf{f}} - r_f \mathbf{1}_K)), \\ \hat{\mathbf{a}}_N &= \frac{1}{\kappa + 1} (\bar{\mathbf{r}} - r_f \mathbf{1}_N - \mathbf{B}_N (\bar{\mathbf{f}} - r_f \mathbf{1}_K) - \mathbf{A}_N \boldsymbol{\lambda}_{miss}^{\hat{}}).\end{aligned}$$

Turning now to the first-order condition with respect to the generic (a, b) th element of \mathbf{B} , denoted by B_{ab} , one obtains:

$$-\frac{1}{T} \sum_{t=1}^T \mathbf{g}'_t \boldsymbol{\Sigma}_N^{-1} \left(-\frac{\partial \mathbf{B}_N}{\partial B_{ab}} \ddot{\mathbf{f}}_t + \mathbf{G}_N \frac{\partial \mathbf{B}_N}{\partial B_{ab}} \ddot{\mathbf{f}} \right) = 0, \quad \text{with } 1 \leq a \leq N, 1 \leq b \leq K,$$

setting, for simplicity,

$$\mathbf{g}_t = (\ddot{\mathbf{r}}_t - \mathbf{G}_N \ddot{\mathbf{r}} - \hat{\mathbf{B}}_N \ddot{\mathbf{f}}_t + \mathbf{G}_N \hat{\mathbf{B}}_N \ddot{\mathbf{f}}) \quad \text{and} \quad \mathbf{g} = \frac{1}{T} \sum_{t=1}^T \mathbf{g}_t,$$

with

$$\mathbf{G}_N = \frac{1}{(\kappa + 1)} \mathbf{I} + \frac{\kappa}{(\kappa + 1)} \mathbf{A}_N (\mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1},$$

and

$$\begin{aligned}\ddot{\mathbf{r}}_t &= (\mathbf{r}_t - r_f \mathbf{1}_N), & \ddot{\mathbf{r}} &= (\bar{\mathbf{r}} - r_f \mathbf{1}_N), \\ \ddot{\mathbf{f}}_t &= (\mathbf{f}_t - r_f \mathbf{1}_K), & \ddot{\mathbf{f}} &= (\bar{\mathbf{f}} - r_f \mathbf{1}_K).\end{aligned}$$

Taking the vec operator for both sides of the first-order condition above gives

$$\frac{1}{T} \sum_{t=1}^T (\ddot{\mathbf{f}}'_t \otimes \mathbf{g}'_t \boldsymbol{\Sigma}_N^{-1}) \text{vec} \left(\frac{\partial \mathbf{B}_N}{\partial B_{ab}} \right) = (\ddot{\mathbf{f}}' \otimes \mathbf{g}' \boldsymbol{\Sigma}_N^{-1} \mathbf{G}_N) \text{vec} \left(\frac{\partial \mathbf{B}_N}{\partial B_{ab}} \right), \quad \text{with } 1 \leq a \leq N, 1 \leq b \leq K,$$

which can be more succinctly re-written as

$$\frac{1}{T} \sum_{t=1}^T \ddot{\mathbf{f}}'_t \mathbf{g}'_t = \ddot{\mathbf{f}}' \boldsymbol{\Sigma}_N^{-1} \mathbf{G}_N \boldsymbol{\Sigma}_N.$$

Next, recalling that $\boldsymbol{\Sigma}_{rf} = \frac{1}{T} \sum_{t=1}^T \ddot{\mathbf{r}}_t \ddot{\mathbf{f}}'_t$, and $\boldsymbol{\Sigma}_{ff} = \frac{1}{T} \sum_{t=1}^T \ddot{\mathbf{f}}_t \ddot{\mathbf{f}}'_t$, with $\boldsymbol{\Sigma}_{fr} = \boldsymbol{\Sigma}'_{rf}$, one obtains $\boldsymbol{\Sigma}_N^{-1} \mathbf{G}_N \boldsymbol{\Sigma}_N = \frac{1}{(\kappa+1)} \mathbf{I} + \frac{\kappa}{(\kappa+1)} \boldsymbol{\Sigma}_N^{-1} \hat{\mathbf{A}}_N (\hat{\mathbf{A}}'_N \hat{\boldsymbol{\Sigma}}_N^{-1} \hat{\mathbf{A}}_N)^{-1} \hat{\mathbf{A}}'_N = \mathbf{G}'_N$ and rearranging the above first order-condition gives

$$\boldsymbol{\Sigma}_{fr} - \ddot{\mathbf{f}}' \mathbf{G}'_N - \boldsymbol{\Sigma}_{ff} \hat{\mathbf{B}}'_N + \ddot{\mathbf{f}}' \hat{\mathbf{B}}'_N \mathbf{G}'_N - (\ddot{\mathbf{f}}' - \ddot{\mathbf{f}}' \mathbf{B}'_N) (\mathbf{I} - \mathbf{G}'_N) \mathbf{G}'_N =$$

$$\boldsymbol{\Sigma}_{fr} - \ddot{\mathbf{r}}'(2\mathbf{G}'_N - \mathbf{G}'_N\mathbf{G}'_N) - \boldsymbol{\Sigma}_{ff}\hat{\mathbf{B}}'_N + \ddot{\mathbf{f}}'\hat{\mathbf{B}}'_N(2\mathbf{G}'_N - \mathbf{G}'_N\mathbf{G}'_N) = \mathbf{0}.$$

Transposing both sides, taking the vec and solving for $\text{vec}(\hat{\mathbf{B}}_N)$ gives:

$$\text{vec}(\hat{\mathbf{B}}_N) = \left((\boldsymbol{\Sigma}_{ff} \otimes \mathbf{I}) - (\ddot{\mathbf{f}}' \otimes (2\mathbf{G}_N - \mathbf{G}_N\mathbf{G}_N)) \right)^{-1} \text{vec} \left(\boldsymbol{\Sigma}_{rf} - (2\mathbf{G}_N - \mathbf{G}_N\mathbf{G}_N)\ddot{\mathbf{r}}' \right).$$

We need to show that a solution for $\hat{\mathbf{B}}_N$ exists. This requires one to establish that the matrix $\left((\boldsymbol{\Sigma}_{ff} \otimes \mathbf{I}) - (\ddot{\mathbf{f}}' \otimes (2\mathbf{G}_N - \mathbf{G}_N\mathbf{G}_N)) \right)$ is invertible. This matrix can be written as

$$\left((\boldsymbol{\Sigma}_{ff} \otimes \mathbf{I}) - (\ddot{\mathbf{f}}' \otimes (2\mathbf{G}_N - \mathbf{G}_N\mathbf{G}_N)) \right) = \left(((\boldsymbol{\Sigma}_{ff} - \ddot{\mathbf{f}}') \otimes \mathbf{I}) + (\ddot{\mathbf{f}}' \otimes (\mathbf{I}_N - (2\mathbf{G}_N - \mathbf{G}_N\mathbf{G}_N))) \right).$$

The first matrix on the right hand side is non-singular. One then just needs to show that the second matrix is positive semi-definitive. This follows because $\mathbf{I}_N - (2\mathbf{G}_N - \mathbf{G}_N\mathbf{G}_N) = (\mathbf{I}_N - \mathbf{G}_N)(\mathbf{I}_N + \mathbf{G}_N)$. We now show that $(\mathbf{I}_N - \mathbf{G}_N)$ is positive semi-definite. In fact

$$\begin{aligned} \mathbf{I}_N - \mathbf{G}_N &= \mathbf{I}_N - \frac{1}{(\kappa + 1)}\mathbf{I}_N - \left(\frac{\kappa}{1 + \kappa}\right)\mathbf{A}_N(\mathbf{A}'_N\boldsymbol{\Sigma}_N^{-1}\mathbf{A}_N)^{-1}\mathbf{A}'_N\boldsymbol{\Sigma}_N^{-1} \\ &= \left(\frac{\kappa}{1 + \kappa}\right)(\mathbf{I}_N - \mathbf{A}_N(\mathbf{A}'_N\boldsymbol{\Sigma}_N^{-1}\mathbf{A}_N)^{-1}\mathbf{A}'_N\boldsymbol{\Sigma}_N^{-1}) \\ &= \left(\frac{\kappa}{1 + \kappa}\right)\boldsymbol{\Sigma}_N(\boldsymbol{\Sigma}_N^{-1} - \boldsymbol{\Sigma}_N^{-1}\mathbf{A}_N(\mathbf{A}'_N\boldsymbol{\Sigma}_N^{-1}\mathbf{A}_N)^{-1}\mathbf{A}'_N\boldsymbol{\Sigma}_N^{-1}) \\ &= \left(\frac{\kappa}{1 + \kappa}\right)\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^{-1/2}(\mathbf{I}_N - \boldsymbol{\Sigma}_N^{-1/2}\mathbf{A}_N(\mathbf{A}'_N\boldsymbol{\Sigma}_N^{-1}\mathbf{A}_N)^{-1}\mathbf{A}'_N\boldsymbol{\Sigma}_N^{-1/2})\boldsymbol{\Sigma}_N^{-1/2}. \end{aligned}$$

The right-hand side is the product of positive definite matrices, namely $\boldsymbol{\Sigma}_N$ and $\boldsymbol{\Sigma}_N^{-1/2}$, and of the matrix $\mathbf{I}_N - \boldsymbol{\Sigma}_N^{-1/2}\mathbf{A}_N(\mathbf{A}'_N\boldsymbol{\Sigma}_N^{-1}\mathbf{A}_N)^{-1}\mathbf{A}'_N\boldsymbol{\Sigma}_N^{-1/2}$, which is a projection matrix orthogonal to $\boldsymbol{\Sigma}_N^{-1/2}\mathbf{A}_N$, and therefore, positive semidefinite. Therefore, plugging $\hat{\mathbf{B}}_N$ into $\boldsymbol{\lambda}_{miss}$ and $\hat{\mathbf{a}}_N$, one obtains that

$$\boldsymbol{\lambda}_{miss}^{\hat{}} = \boldsymbol{\lambda}_{miss}^{\hat{}}(\mathbf{A}_N, \mathbf{C}_N, \kappa), \quad \hat{\mathbf{a}}_N = \hat{\mathbf{a}}_N(\mathbf{A}_N, \mathbf{C}_N, \kappa),$$

since $\hat{\mathbf{B}}_N = \hat{\mathbf{B}}_N(\mathbf{A}_N, \mathbf{C}_N, \kappa)$. Substituting them into $L(\boldsymbol{\theta}) - \kappa(\mathbf{a}'_N\boldsymbol{\Sigma}_N^{-1}\mathbf{a}_N - \delta)$, gives the concentrated likelihood function, which is a function of only \mathbf{A}_N and \mathbf{C}_N . Notice that all the estimates depend ultimately on κ that can be chosen by cross-validation methods.

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

$$\begin{pmatrix} \mathbf{A}'_N\boldsymbol{\Sigma}_N^{-1} \\ \mathbf{I}_N \end{pmatrix} \left(\bar{\mathbf{r}} - r_f\mathbf{1}_N - \mathbf{B}_N(\bar{\mathbf{f}} - r_f\mathbf{1}_K) \right) = \begin{pmatrix} \mathbf{A}'_N\boldsymbol{\Sigma}_N^{-1} \\ \mathbf{I}_N \end{pmatrix} (\mathbf{A}_N, \mathbf{I}_N) \begin{pmatrix} \boldsymbol{\lambda}_{miss}^{\hat{}} \\ \hat{\mathbf{a}}_N \end{pmatrix},$$

where, for simplicity, we set $\boldsymbol{\lambda}_{miss}^{\hat{}} = \boldsymbol{\lambda}_{miss}^{\hat{}}_{ML}$, $\hat{\mathbf{a}}_N = \hat{\mathbf{a}}_{N,ML}$ and recall that $\boldsymbol{\Sigma}_N = \mathbf{A}_N\mathbf{A}'_N + \mathbf{C}_N$. One can clearly obtain a unique solution for $(\mathbf{A}_N, \mathbf{I}_N) \begin{pmatrix} \boldsymbol{\lambda}_{miss}^{\hat{}} \\ \hat{\mathbf{a}}_N \end{pmatrix} = \mathbf{A}_N\boldsymbol{\lambda}_{miss}^{\hat{}} + \hat{\mathbf{a}}_N$.

However, to solve for λ_{miss} and $\hat{\mathbf{a}}_N$ separately, one needs to invert the matrix:

$$\begin{pmatrix} \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \\ \mathbf{I}_N \end{pmatrix} (\mathbf{A}_N, \mathbf{I}_N) = \begin{pmatrix} \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{A}_N & \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \\ \mathbf{A}_N & \mathbf{I}_N \end{pmatrix},$$

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 from the product of two matrices of dimension $(N + K) \times N$. All the other parameters are identified separately, along with $\boldsymbol{\alpha}_0 = \mathbf{A}_0 \mathbf{m}_0 + \mathbf{a}_0$, and their expressions follow from differentiating $L(\boldsymbol{\theta})$ and solving the resulting first-order conditions. \blacksquare

A.12 Estimation for unbounded residual variation case with mismeasured factors

In this section, we consider the estimation of the model for the case where *all* factors are observed but are measured with error: $\mathbf{f}_t = \mathbf{f}_t^* + \boldsymbol{\eta}_t$, where $\boldsymbol{\eta}_t$ has mean $\boldsymbol{\mu}_{\eta_0}$ and covariance matrix $\boldsymbol{\Sigma}_{\eta_0}$. Then, $\boldsymbol{\alpha}_0 = -\mathbf{B}_0 \boldsymbol{\mu}_{\eta_0}$ and $\boldsymbol{\Sigma}_0 = \mathbf{B}_0 \boldsymbol{\Sigma}_{\eta_0} \mathbf{B}'_0 + \mathbf{C}_0$, where \mathbf{C}_0 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.⁴¹

The case of mis-measured factors turns out to be a particular case of the so-called multivariate errors-in-variables model; see (Fuller, 1987, Chapter 4) for a classical reference.⁴² Although several approaches for estimation of multivariate errors-in-variables models exist (see Fuller (1987)) in analogy with the other forms of misspecification discussed above, we illustrate here ML estimation. Every estimation method of errors-in-variables models requires some further information, beyond a sample of observations of the \mathbf{r}_t and \mathbf{f}_t . For ML estimation we assume that we have the availability of a preliminary estimate, $\mathbf{S}_{C\eta}$, of the matrix $\boldsymbol{\Sigma}_{C\eta_0} = \begin{pmatrix} \mathbf{C}_0 & \mathbf{0} \\ \mathbf{0}' & \boldsymbol{\Sigma}_{\eta_0} \end{pmatrix}$ with certain characteristics described in the statement of the theorem below and in Amemiya and Fuller (1984, Theorem 3).

⁴¹Observe that \mathbf{B}_0 appears in both the intercept and residual covariance matrix of returns. Recognizing this allows one to improve the precision of the estimated \mathbf{B}_0 , as discussed in the previous section. Using the Sherman-Morrison-Woodbury formula, we see that:

$$\boldsymbol{\alpha}'_0 \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\alpha}_0 = \boldsymbol{\mu}'_{\eta_0} \mathbf{B}'_0 \mathbf{C}_0^{-1} \mathbf{B}_0 (\boldsymbol{\Sigma}_{\eta_0}^{-1} + \mathbf{B}'_0 \mathbf{C}_0^{-1} \mathbf{B}_0)^{-1} \boldsymbol{\Sigma}_{\eta_0}^{-1} \boldsymbol{\mu}_{\eta_0}.$$

Therefore, $\boldsymbol{\alpha}'_0 \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\alpha}_0$ converges to $\boldsymbol{\mu}'_{\eta_0} \boldsymbol{\Sigma}_{\eta_0}^{-1} \boldsymbol{\mu}_{\eta_0}$ implying that the APT restriction is always satisfied and there is no need to impose it in the estimation.

⁴²Typically, errors-in-variables models are classified as being either functional or structural, depending on whether the \mathbf{f}_t^* are assumed constant or instead random variables with a known distribution. Here we adopt the latter formulation.

To derive the log-likelihood function for estimation, it is convenient to adopt a slightly different parameterization from Case 4 in Section 3.5, rewriting the model as

$$\begin{pmatrix} \mathbf{r}_t - r_f \mathbf{1}_N \\ \mathbf{f}_t - r_f \mathbf{1}_K \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_0 \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{B}_0 \\ \mathbf{I}_K \end{pmatrix} (\mathbf{f}_t^* + \boldsymbol{\mu}_{\eta 0} - r_f \mathbf{1}_K) + \begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\eta}_t - \boldsymbol{\mu}_{\eta 0} \end{pmatrix},$$

with $\boldsymbol{\alpha}_0 = -\mathbf{B}_0 \boldsymbol{\mu}_{\eta 0}$. Then setting

$$\mathbf{Z}_t = \begin{pmatrix} \mathbf{r}_t - r_f \mathbf{1}_N \\ \mathbf{f}_t - r_f \mathbf{1}_K \end{pmatrix},$$

one obtains:

$$\text{cov}(\mathbf{Z}_t) = \boldsymbol{\Sigma}_{Z0} = \begin{pmatrix} \mathbf{B}_0 \\ \mathbf{I}_K \end{pmatrix} \boldsymbol{\Omega}_0^* \begin{pmatrix} \mathbf{B}_0 \\ \mathbf{I}_K \end{pmatrix}' + \boldsymbol{\Sigma}_{C\eta 0}; \quad E(\mathbf{Z}_t) = \boldsymbol{\mu}_{Z0} = \begin{pmatrix} \boldsymbol{\alpha}_0 \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{B}_0 \\ \mathbf{I}_K \end{pmatrix} \boldsymbol{\lambda}_0^*,$$

where we set $\boldsymbol{\lambda}_0^* = E(\mathbf{f}_t^* + \boldsymbol{\mu}_{\eta} - r_f \mathbf{1}_K)$ and $\boldsymbol{\Omega}^* = \text{cov}(\mathbf{f}_t^*)$. We have implicitly assumed that $\boldsymbol{\varepsilon}_t, \boldsymbol{\eta}_s$ and \mathbf{f}_t^* are uncorrelated for any t, s, l in the evaluation of $\boldsymbol{\Sigma}_{Z0}$. Therefore, when $(\mathbf{f}_t^*, \boldsymbol{\varepsilon}_t', \boldsymbol{\eta}_t)'$ are iid and jointly normal, with a block-diagonal covariance matrix,⁴³ the log likelihood (here adjusted for the degrees of freedom) of model (12), with respect to \mathbf{Z}_t as the observable vector, is:

$$\begin{aligned} L(\boldsymbol{\theta}) = & -\frac{T-1}{2T} \log(\det(\boldsymbol{\Sigma}_Z)) - \frac{T-1}{2T^2} \sum_{t=1}^T (\mathbf{Z}_t - \boldsymbol{\mu}_Z)' \boldsymbol{\Sigma}_Z^{-1} (\mathbf{Z}_t - \boldsymbol{\mu}_Z) \\ & - \frac{d}{2T} \log(\det(\boldsymbol{\Sigma}_{C\eta})) - \frac{d}{2T} \text{tr}(\boldsymbol{\Sigma}_{C\eta} \mathbf{S}_{C\eta}), \end{aligned} \quad (\text{A9})$$

setting $\boldsymbol{\theta} = (\boldsymbol{\mu}'_{\eta}, \text{vech}(\boldsymbol{\Sigma}_{\eta})', \text{vec}(\mathbf{B})', \text{vech}(\mathbf{C})', \text{vech}(\boldsymbol{\Omega}^*)', \boldsymbol{\lambda}^{*'})'$ with $\boldsymbol{\mu}_Z = \begin{pmatrix} \boldsymbol{\alpha} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{B} \\ \mathbf{I}_K \end{pmatrix} \boldsymbol{\lambda}^*$

and $\boldsymbol{\Sigma}_Z = \begin{pmatrix} \mathbf{B} \\ \mathbf{I}_K \end{pmatrix} \boldsymbol{\Omega}^* \begin{pmatrix} \mathbf{B} \\ \mathbf{I}_K \end{pmatrix}' + \boldsymbol{\Sigma}_{C\eta}$, where we have dropped the N subscript in order to simplify notation. All the parameters' estimators are highly nonlinear, except for the estimator for $\boldsymbol{\lambda}^*$, but their closed-form expression has been established, as described in the next theorem.⁴⁴

The following theorem, exploits known results from Fuller (1987, Section 4.1.1., p.298) adapted to our framework, and hence, is stated without proof.

Theorem A.5 (Mismeasured-factors case). *Suppose that the vector of asset returns, \mathbf{r}_t , satisfies Assumption 3.1. Assume that $\boldsymbol{\Sigma}_{C\eta 0} > 0$ and that its estimator $\mathbf{S}_{C\eta}$ is an unbiased*

⁴³Mutual independence of $\boldsymbol{\varepsilon}_t$ with $\boldsymbol{\eta}_s$ is not necessary but is assumed here for the sake of simplicity.

⁴⁴Distributional assumptions (normally distributed $(\mathbf{f}_t^*, \boldsymbol{\varepsilon}_t', \boldsymbol{\eta}_t)'$ and Wishart $\mathbf{S}_{C\eta}$) are not required for consistency nor for asymptotic normality although the latter are often made when inference for all parameters is desired. Alternative conditions are bounded fourth moment of $(\mathbf{f}_t^*, \boldsymbol{\varepsilon}_t', \boldsymbol{\eta}_t)'$ and asymptotic normality of $d^{\frac{1}{2}}(\mathbf{S}_{C\eta} - \boldsymbol{\Sigma}_{C\eta 0})$; see Amemiya and Fuller (1984, comments to Theorem 3, p.507).

estimator for $\Sigma_{C\eta_0}$, distributed as Wishart with parameter $\Sigma_{C\eta_0}$ and degrees of freedom d satisfying $T/d \rightarrow c > 0$, some constant c and independent of $\{\mathbf{r}_t, \mathbf{f}_t\}$. Set $\mathbf{S}_Z = (T - 1)^{-1} \sum_{t=1}^T (\mathbf{Z}_t - \bar{\mathbf{Z}})(\mathbf{Z}_t - \bar{\mathbf{Z}})'$ and let $\hat{g}_{1,N+K} \geq \dots \geq \hat{g}_{N+K,N+K}$ be the eigenvalues of $\mathbf{S}_{C\eta}^{-\frac{1}{2}} \mathbf{S}_Z \mathbf{S}_{C\eta}^{-\frac{1}{2}}$ with corresponding eigenvectors $\hat{\mathbf{v}}_{i,N+K}$, with $1 \leq i \leq N + K$, the first K of which we collect as

$$\hat{\mathbf{G}} = \begin{pmatrix} \hat{g}_{1,N+K} & 0 & \dots \\ \dots & \hat{g}_{i,N+K} & \dots \\ 0 & \dots & \hat{g}_{K,N+K} \end{pmatrix} \quad \text{and} \quad \hat{\mathbf{V}} = (\hat{\mathbf{v}}_{1,N+K} \dots \hat{\mathbf{v}}_{K,N+K}).$$

If $\hat{g}_{K,N+K} > 1$, then $\hat{\boldsymbol{\theta}}_{MLC} = (\hat{\boldsymbol{\mu}}'_{\eta,MLC}, \text{vech}(\hat{\boldsymbol{\Sigma}}_{\eta,MLC})', \text{vec}(\hat{\mathbf{B}}_{MLC})', \text{vech}(\hat{\mathbf{C}}_{MLC})', \text{vech}(\hat{\boldsymbol{\Omega}}^*_{MLC})', \hat{\boldsymbol{\lambda}}^*{}'_{MLC})'$ satisfies

$$\hat{\boldsymbol{\theta}}_{MLC} = \underset{\boldsymbol{\theta}}{\text{argmax}} L(\boldsymbol{\theta}),$$

where $L(\boldsymbol{\theta})$ is defined in (A9),

$$\hat{\mathbf{B}}_{MLC} = \mathbf{P}_{NK} \mathbf{P}_{KK}^{-1},$$

$$\hat{\boldsymbol{\mu}}_{\eta,MLC} = -\left(\hat{\mathbf{B}}'_{MLC} \hat{\mathbf{B}}_{MLC}\right)^{-1} \hat{\mathbf{B}}'_{MLC} \left(T^{-1} \sum_{t=1}^T \mathbf{r}_t - r_f \mathbf{1}_N - \hat{\mathbf{B}}_{MLC} \hat{\boldsymbol{\lambda}}^*_{MLC}\right),$$

$$\hat{\boldsymbol{\Sigma}}_{C\eta,MLC} = (T - 1 + d)^{-1} [(T - 1)(\mathbf{S}_Z - \hat{\boldsymbol{\Sigma}}_z) + d\mathbf{S}_{C\eta}],$$

$$\hat{\boldsymbol{\lambda}}^*_{MLC} = T^{-1} \sum_{t=1}^T \mathbf{f}_t - r_f \mathbf{1}_K,$$

$$\hat{\boldsymbol{\Omega}}^*_{MLC} = \mathbf{P}_{KK} (\hat{\mathbf{G}} - \mathbf{I}_K) \mathbf{P}'_{KK},$$

setting $\hat{\boldsymbol{\Sigma}}_z = \begin{pmatrix} \hat{\mathbf{B}}_{MLC} \\ \mathbf{I}_K \end{pmatrix} \hat{\boldsymbol{\Omega}}^*_{MLC} \begin{pmatrix} \hat{\mathbf{B}}_{MLC} \\ \mathbf{I}_K \end{pmatrix}$, $\mathbf{P} = \mathbf{S}_{C\eta}^{\frac{1}{2}} \hat{\mathbf{V}} = (\mathbf{P}'_{NK}, \mathbf{P}_{KK})'$. Note that the upper-left and the lower-right blocks of $\hat{\boldsymbol{\Sigma}}_{C\eta,MLC}$ provide the estimators $\hat{\mathbf{C}}_{MLC}$ and $\hat{\boldsymbol{\Sigma}}_{\eta,MLC}$, respectively.

On the other hand, if $\hat{g}_{K,N+K} \leq 1$, then \mathbf{P}_{KK} is singular and $\hat{\mathbf{B}}_{MLC}$ is not well defined any longer.

Remark A.5.1. The unobserved factors \mathbf{f}_t^* , which determine the asset returns, can be estimated as

$$\hat{\mathbf{f}}_t^* = \left(T^{-1} \sum_{s=1}^T \mathbf{f}_s - \hat{\boldsymbol{\mu}}_{\eta,MLC}\right) + \hat{\mathbf{H}}(\mathbf{Z}_t - T^{-1} \sum_{s=1}^T \mathbf{Z}_s),$$

setting $\hat{\mathbf{H}} = \left((\mathbf{0}, \mathbf{I}_K) - \hat{\boldsymbol{\Sigma}}_{\eta v} \hat{\boldsymbol{\Sigma}}_{vv}^{-1} (\mathbf{I}_N, -\hat{\mathbf{B}}_{MLC})\right)$ with $\hat{\boldsymbol{\Sigma}}_{vv} = (\mathbf{I}_N, -\hat{\mathbf{B}}_{MLC}) \hat{\boldsymbol{\Sigma}}_{C\eta,MLC} (\mathbf{I}_N, -\hat{\mathbf{B}}_{MLC})'$ and $\hat{\boldsymbol{\Sigma}}_{\eta v} = \hat{\boldsymbol{\Sigma}}_{\eta,MLC} \hat{\mathbf{B}}'_{MLC}$.

Remark A.5.2. For portfolio construction purposes, having estimated the model's parameters, we obtain the mean-variance weights as:

$$\hat{\mathbf{w}}_N^{\text{mv}} = \frac{1}{\gamma} \left(\hat{\mathbf{B}}_{N,MLC} \hat{\mathbf{\Omega}}^* \hat{\mathbf{B}}'_{N,MLC} + \hat{\mathbf{C}}_{N,MLC} \right)^{-1} \left(-\hat{\mathbf{B}}_{N,MLC} \hat{\boldsymbol{\mu}}_{\eta,MLC} + \hat{\mathbf{B}}_{N,MLC} \hat{\boldsymbol{\lambda}}^*_{MLC} \right).$$

Remark A.5.3. As an alternative approach to the one described in the theorem above, one could estimate $\boldsymbol{\alpha}_0$, \mathbf{B}_0 , and the residual variance, $\boldsymbol{\Sigma}_0 = \mathbf{B}_0 \boldsymbol{\Sigma}_{\eta_0} \mathbf{B}'_0 + \mathbf{C}_0$, using an instrumental-variable type estimator, because of the correlation between the mismeasured factors, \mathbf{f}_t and the residual, $(\boldsymbol{\varepsilon}_t - \mathbf{B}_0(\boldsymbol{\eta}_t - \boldsymbol{\mu}_{\eta_0}))$. However, doing so would not allow one to exploit the fact that \mathbf{B}_0 appears both in the intercept term of expected returns and in the residual-covariance matrix, implying a loss of precision in the estimates. Note that the MLEs derived above requires the preliminary estimate $\mathbf{S}_{C\eta}$ of $\boldsymbol{\Sigma}_{C\eta_0}$, which can be obtained by the aforementioned instrumental-variable estimator, by means of an independent sample of observations.

Remark A.5.4. For expositional simplicity, we assumed above that all observed factors were measured with error. In practice, it is possible that some factors are measured without error and, in fact, the above estimation procedure holds also when $\boldsymbol{\Sigma}_{\eta_0}$ is singular; see Fuller (1987, Section 4.1.1., p.298).⁴⁵

Remark A.5.5. In practice, all the various forms of misspecification discussed above are likely to arise at the same time. Therefore, expected excess returns (conditional on the factors) are $\boldsymbol{\mu}_0 - r_f \mathbf{1}_N = \mathbf{a}_0 + \mathbf{A}_0 \boldsymbol{\lambda}_{\text{miss}0} - \mathbf{B}_{K_10} \boldsymbol{\mu}_{K_10} + \mathbf{B}_0 \boldsymbol{\lambda}_0^*$ and the residual covariance-matrix is $\boldsymbol{\Sigma}_0 = \mathbf{C}_0 + \mathbf{B}_{K_10} \boldsymbol{\Sigma}_{K_10} \mathbf{B}'_{K_10} + \mathbf{A}_0 \mathbf{A}'_0$. Using the same arguments of Theorem 4.6, the APT restriction allows one to separately identify \mathbf{A}_0 from $\boldsymbol{\lambda}_{\text{miss}0}$ and $\boldsymbol{\mu}_{K_10}$. Using a procedure similar to the one described in the remarks above, one can then identify the number of missing and/or mismeasured factors and then estimate the full model using ML that takes into account the various restrictions arising from misspecification.

⁴⁵Anderson (1984) provides a test for the null hypothesis that the rank of $\boldsymbol{\Sigma}_{\eta_0}$ is less or equal to K against the alternative that this rank is greater than K , without requiring one to identify which factor is affected by measurement error.

References

- AMEMIYA, Y., AND A. W. FULLER (1984): “Estimation for the Multivariate Errors-in-Variables Model with Estimated Error Covariance Matrix,” *The Annals of Statistics*, 12(2), 497–509.
- ANDERSON, T. M. (1984): *An Introduction to Multivariate Statistical Analysis*. John Wiley, New York.
- ANG, A. (2014): *Asset Management: A Systematic Approach to Factor Investing*. Oxford University Press.
- ARAGON, G. O., AND W. E. FERSON (2006): “Portfolio Performance Evaluation,” *Foundations and Trends*, 2(2), 83–190.
- BAWA, V. S., S. BROWN, AND R. KLEIN (1979): *Estimation Risk and Optimal Portfolio Choice*. North Holland, Amsterdam.
- BLACK, F. (1995): “Estimating Expected Return,” *Financial Analysts Journal*, pp. 168–171.
- BLACK, F., AND R. LITTERMAN (1990): “Asset Allocation: Combining Investor Views with Market Equilibrium,” Discussion paper, Goldman, Sachs & Co.
- (1992): “Global Portfolio Optimization,” *Financial Analysts Journal*, 48(5), 28–43.
- BROWN, K. C., AND C. TIU (2010): “Do Endowment Funds Select The Optimal Mix of Active and Passive Risk?,” *Journal of Investment Management*, 8(1), 1–25.
- CHAMBERLAIN, G. (1983): “Funds, Factors and Diversification in Arbitrage Pricing Models,” *Econometrica*, 51, 1305–1324.
- CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): “Arbitrage, Factor Structure and Mean-Variance Analysis on Large Asset Markets,” *Econometrica*, 51, 1281–1304.
- CHEN, Z., AND L. G. EPSTEIN (2002): “Ambiguity, Risk, and Asset Returns in Continuous Time,” *Econometrica*, 70(4), 1403–1443.
- COCHRANE, J., AND J. SAA-REQUEJO (2001): “Beyond Arbitrage: Good-deal Asset Price Bounds in Incomplete Markets,” *Journal of Political Economy*, 108, 79–119.
- COCHRANE, J. H. (2005): *Asset Pricing (revised edition)*. Princeton University Press.
- CONNOR, G., L. GOLDBERG, AND R. A. KORAJCZYK (2010): *Portfolio Risk Analysis*. Princeton University Press.
- CONNOR, G., AND R. A. KORAJCZYK (1986): “Performance Measurement with the Arbitrage Pricing Theory,” *Journal of Financial Economics*, 15(3), 373–394.
- DEMIGUEL, V., L. GARLAPPI, F. J. NOGALES, AND R. UPPAL (2009): “A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms,” *Management Science*, 55(5).

- DEMIGUEL, V., L. GARLAPPI, AND R. UPPAL (2009): “Optimal Versus Naïve Diversification: How Inefficient is the 1/N Portfolio Strategy?,” *Review of Financial Studies*, 22, 1915–1953.
- DYBVIK, P. H., AND S. A. ROSS (1985a): “The Analytics of Performance Measurement Using a Security Market Line,” *The Journal of Finance*, 40(2), 401–416.
- DYBVIK, P. H., AND S. A. ROSS (1985b): “Differential Information and Performance Measurement Using a Security Market Line,” *Journal of Finance*, 40(2), 383–399.
- EPSTEIN, L. G., AND M. SCHNEIDER (2010): “Ambiguity and Asset Markets,” *Annual Review of Financial Economics*, 2, 315–346.
- FERSON, W. E. (2013): *Investment Performance: A Review and Synthesis*, vol. 2A-2B. Elsevier Inc.
- FROST, P., AND E. SAVARINO (1986): “An Empirical Bayes Approach to Efficient Portfolio Selection,” *Journal of Financial and Quantitative Analysis*, 21, 293–305.
- (1988): “For Better Performance: Constrain Portfolio Weights,” *Journal of Portfolio Management*, 15, 29–34.
- FULLER, W. A. (1987): *Measurement Error Models*. John Wiley & Sons, New York.
- GARLAPPI, L., R. UPPAL, AND T. WANG (2007): “Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach,” *Review of Financial Studies*, 20(1), 41–81.
- GHIRARDATO, P., F. MACCHERONI, AND M. MARINACCI (2004): “Differentiating Ambiguity and Ambiguity Attitude,” *Journal of Economic Theory*, 118, 133–173.
- GIBBONS, M. R., S. A. ROSS, AND J. SHANKEN (1989): “A Test of the Efficiency of a Given Portfolio,” *Econometrica*, 57(5), 1121–1152.
- GILBOA, I., AND D. SCHMEIDLER (1989): “Maxmin Expected Utility Theory with Non-Unique Prior,” *Journal of Mathematical Economics*, 18, 141–153.
- GREEN, R., AND B. HOLLIFIELD (1992): “When Will Mean-Variance Efficient Portfolios be Well Diversified?,” *Journal of Finance*, 47, 1785–1809.
- GREEN, R. C. (1986a): “Benchmark Portfolio Inefficiency and Deviations from the Security Market Line,” *The Journal of Finance*, 41(2), 295–312.
- GREEN, R. C. (1986b): “Positively Weighted Portfolios on the Minimum-Variance Frontier,” *The Journal of Finance*, 41(5), 1051–1068.
- GRINBLATT, M., AND S. TITMAN (1989): “Portfolio Performance Evaluation: Old Issues and New Insights,” *The Review of Financial Studies*, 2(3), 393–421.
- HANSEN, L. P., AND R. JAGANNATHAN (1997): “Assessing Specification Errors in Stochastic Discount Factor Models,” *Journal of Finance*, 52(2), 557–590.
- HANSEN, L. P., AND T. J. SARGENT (2008): *Robustness*. Princeton University Press, Princeton, NJ.

- HE, Z. L. (2007): “Incorporating Alpha Uncertainty into Portfolio Decisions : A Bayesian Revisit of the Treynor-Black Model,” *Journal of Asset Management*, 8(3), 161–175.
- HUANG, C.-F., AND R. H. LITZENBERGER (1988): *Foundations of Financial Economics*. Elsevier Science.
- HUBERMAN, G. (1982): “A Simple Approach to the Arbitrage Pricing Theory,” *Journal of Economic Theory*, 28, 183–191.
- INGERSOLL, J. (1984): “Some Results in the Theory of Arbitrage Pricing,” *Journal of Finance*, 39, 1021–1039.
- JAGANNATHAN, R., AND T. MA (2003): “Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps,” *Journal of Finance*, 58, 1651–1684.
- JOBSON, J. D., AND R. KORKIE (1980): “Estimation for Markowitz Efficient Portfolios,” *Journal of the American Statistical Association*, 75, 544–554.
- JORION, P. (1986): “Bayes-Stein Estimation for Portfolio Analysis,” *The Journal of Financial and Quantitative Analysis*, 21(3), 279–292.
- JU, N., AND J. MIAO (2012): “Ambiguity, Learning, and Asset Returns,” *Econometrica*, 80(2), 559–591.
- KANE, A., T.-H. KIM, AND H. WHITE (2003): “Active Portfolio Management: The Power of the Treynor-Black Model,” Working Paper.
- KLIBANOFF, P., M. MARINACCI, AND S. MUKERJI (2005): “A Smooth Model of Decision Making Under Ambiguity,” *Econometrica*, 73(6), 1849–1892.
- KOSOWSKI, R., A. TIMMERMANN, R. WERMERS, AND H. WHITE (2006): “Can Mutual Fund Stars Really Pick Stocks? New Evidence from a Bootstrap Analysis,” *Journal of Finance*, 61(6), 2551–2595.
- KOTHARI, S., AND J. SHANKEN (2002): *Anomalies and Efficient Portfolio Formation*. CFA Institute.
- LEHMANN, B., AND D. MODEST (1988): “The Empirical Foundations of the Arbitrage Pricing Theory,” *Journal of Financial Economics*, 21, 213–254.
- MACCHERONI, F., M. MARINACCI, AND D. RUFFINO (2013): “Alpha as Ambiguity: Robust Mean-Variance Portfolio Analysis,” *Econometrica*, 81(3), 1075–1113.
- MACKINLAY, A. C., AND Ľ. PÁSTOR (2000): “Asset Pricing Models: Implications for Expected Returns and Portfolio Selection,” *Review of Financial Studies*, 13(4), 883–916.
- MICHAUD, R. (1989): “The Markowitz Optimization Enigma: Is “Optimized” Optimal?,” *Financial Analysts Journal*, pp. 31–42.
- OLIVARES-NADAL, A. V., AND V. DEMIGUEL (2015): “A Data-Driven Perspective on Transaction Costs in Portfolio Selection,” Working Paper, London Business School.

- PÁSTOR, Ľ. (2000): “Portfolio Selection and Asset Pricing Models,” *Journal of Finance*, 55, 179–223.
- PÁSTOR, Ľ., AND R. F. STAMBAUGH (2000): “Comparing Asset Pricing Models: An Investment Perspective,” *Journal of Financial Economics*, pp. 335–381.
- PESARAN, M., AND P. ZAFFARONI (2009): “Optimality and Diversifiability of Mean Variance and Arbitrage Pricing Portfolios,” Working Paper, University of Cambridge and Imperial College London.
- PETTENUZZO, D., A. G. TIMMERMANN, AND R. VALKANOV (2014): “Forecasting Stock Returns under Economic Constraints,” *Journal of Financial Economics*, 114(3), 517–553.
- REISMAN, H. (1988): “A General Approach to the Arbitrage Pricing Theory (APT),” *Econometrica*, 56(2), 473–476.
- REISMAN, H. (1992a): “Intertemporal Arbitrage Pricing Theory,” *Review of Financial Studies*, 5(1), 105–122.
- REISMAN, H. (1992b): “Reference Variables, Factor Structure, and the Approximate Multi-beta Representation,” *Journal of Finance*, 47(4), 1303–1314.
- ROLL, R. (1977): “A Critique of the Asset Pricing Theory’s Tests Part I: On Past and Potential Testability of the Theory,” *Journal of Financial Economics*, 4(2), 129–176.
- (1980): “Orthogonal Portfolios,” *Journal of Financial and Quantitative Analysis*, 15(5), 1005–1023.
- ROSS, S. (1976): “The Arbitrage Theory of Capital Asset Pricing,” *Journal of Economic Theory*, 13, 341–360.
- ROSS, S. A. (1977): “Return, Risk, and Arbitrage,” in *Risk and Return in Finance*, ed. by I. Friend, and J. Bicksler. Ballinger, Cambridge, MA.
- SHANKEN, J. (1992): “The Current State of the Arbitrage Pricing Theory,” *Journal of Finance*, 47(4), 1569–1574.
- SHARPE, W. F. (1963): “A Simplified Model for Portfolio Analysis,” *Management Science*, 9, 277–293.
- TIBSHIRANI, R. (1996): “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288.
- TREYNOR, J. L., AND F. BLACK (1973): “How to Use Security Analysis to Improve Portfolio Selection,” *The Journal of Business*, 46(1), 66.