Robust Portfolio Choice*

Valentina Raponi Raman Uppal Paolo Zaffaroni

September 9, 2022

Abstract

We develop a normative theory for constructing mean-variance portfolios robust to model misspecification. We identify two *inefficient* portfolios—an "alpha" portfolio, representing latent asset demand, that depends only on pricing errors and a "beta" portfolio that depends on observed factor risk premia—which, when combined, give mean-variance *efficient* portfolios. We show that the alpha and beta portfolios have different economic properties and therefore misspecification in these portfolios should be treated using different methods. Our theoretical insights lead to an economically substantial and statistically significant improvement in out-of-sample portfolio performance, with latent asset demand playing a dominant role.

JEL classification: C58, G11, G12.

Keywords: Alpha, asset-specific risk, factor models, factor investing, latent asset demand, model misspecification.

*Raponi is affiliated with IESE Business School; email: `vraponi@iese.edu`. 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 the use of its computer cluster and Mengmeng Ao, Yingying Li, and Xinghua Zheng for sharing their data with us. We gratefully acknowledge comments from Dante Amengual (discussant), Turan Bali, Harjoat Bhamra (discussant), Tim Bollerslev, Svetlana Bryzgalova (discussant), Timothy Christensen, John Cochrane, Rama Cont, Massimo Dello Preite, Victor DeMiguel, Serge Darolles, Patrick Gagliardini, René Garcia, Lorenzo Garlappi, Christian Gourieéroux (discussant), Amit Goyal, Christian Julliard, Andrew Karolyi, Ralph Koijen, Hugues Langlois, Michelle Lee, Xioaji Lin (discussant), Lionel Martellini, Alberto Martín-Utrera, Stefan Nagel (discussant), Vasant Naik, Javier Nogales, Ľuboš Pástor (discussant), Markus Pelger, Hashem Pesaran, Eric Renault, Cesare Robotti, Lucio Sarno, Olivier Scaillet, Jay Shanken, Yazid Sharaiha, Eric Shirbini, Claudio Tebaldi, Nikolaos Tessaromatis, Allan Timmermann, Fabio Trojani, Tan Wang, Raina Uppal, Bas Werker, Michael Wolf, Motohiro Yogo, Guofu Zhou, and Irina Zviadadze and seminar participants at the 8th Annual Hedge Fund Conference, Advances in the Analysis of Hedge Fund Strategies Conference (Imperial College London), American Finance Association meetings, CEPR First Annual Spring Symposium in Financial Economics, European Finance Association meetings, Financial Econometrics Conference (Toulouse School of Economics), Inquire UK Conference, Judge Business School, NBER Summer Institute Workshop on Forecasting & Empirical Methods, Norges Bank Investment Management, UBC Summer Finance Conference, Vienna-Copenhagen Conference on Financial Econometrics, and Western Finance Association meetings.

1 Introduction

The mean-variance portfolio choice model is the cornerstone of financial economics because it captures the fundamental trade-off between expected return and risk. To implement this model, an investor needs to know the means, variances, and covariances of asset returns. This is particularly challenging when the number of assets is large, which is increasingly the case given the availability of data on individual stocks. This challenge motivates the use of a beta-pricing model, such as the Sharpe (1964) Capital Asset Pricing Model or the Fama and French (2015) five-factor model. However, as highlighted by Hansen and Sargent (1999), "with tractability comes misspecification," because when specifying a particular factor model for asset returns, factors could be mismeasured (Roll, 1977) or missing (e.g., MacKinlay and Pástor (2000)).

In this paper, we develop a theory to resolve model misspecification when using a candidate beta-pricing model for constructing mean-variance portfolios and demonstrate its effectiveness empirically. The candidate beta-pricing model may be misspecified because it omits pervasive risk factors, asset-specific components, or both. These omissions can also arise when considering misspecified conditional asset-pricing models. We show how to use the Arbitrage Pricing Theory (APT) of Ross (1976) to address these sources of misspecification by extending the conventional view of the APT. We demonstrate that the APT allows for not just "small" asset-specific pricing errors but also "large" pricing errors related to omitted pervasive risk factors while still satisfying the no-arbitrage condition that bounds the Sharpe ratio of the overall portfolio.

Typically, correcting a candidate beta-pricing model for misspecification has been done by augmenting the model with additional (pervasive) risk factors. However, purely asset-specific pricing errors have been ignored because, being idiosyncratic, they are deemed diversifiable and hence not priced. But, this view is correct only if the beta-pricing model, after it is augmented with the missing risk factors, becomes correctly specified so that there are zero pricing errors and exact pricing holds. In the absence of exact pricing, purely asset-specific pricing errors cannot be diversified by augmenting the candidate beta-pricing model with omitted pervasive factors. We explain that asset-specific pricing errors represent compensation for bearing asset-specific risk. Our perspective, which allows for asset-specific pricing errors, is a crucial departure from the standard framework with exact pricing and leads to novel insights.

These insights have important implications for portfolio selection, which we present in four propositions. In Proposition 4.1, we show that under the APT the entire set of mean-variance *efficient* portfolios is generated by two *inefficient* portfolios—the "beta" portfolio, which depends on observed factor risk premia, and the "alpha" portfolio, which has exposure only to omitted risk factors and asset-specific risk. The alpha portfolio, because it depends on pricing errors that are spanned by latent pervasive risk factors and latent asset-specific risk, represents the latent asset demand studied by Koijen and Yogo (2019) but, in our case, founded on the APT. We show that if one were to ignore latent asset demand generated by the purely asset-specific pricing errors, the resulting portfolio would not be on the efficient frontier *regardless of how many risk factors were included in the model.* Moreover, we establish in Proposition 4.2 that, as the number of assets N increases, the elements of the alpha portfolio can dominate, in terms of magnitude, the corresponding elements of the beta portfolio.[1] This result motivates us to use different approaches for dealing with misspecification in the alpha and beta portfolios.

In Proposition 4.3, we establish a set of conditions under which, as the number of assets increases, the beta portfolio can be replaced, without loss of performance, by a benchmark portfolio (such as the equal- or value-weighted portfolio) that by construction is functionally independent of the mean vector (risk premia) and covariance matrix of the observed factors, and hence, is immune to misspecification of these quantities. Finally, we show in Proposition 4.4 that, under the APT, the alpha portfolio, which represents latent asset demand, is precisely the "robust" portfolio one would obtain when using the max-min approach of Gilboa and Schmeidler (1989) and Hansen and Sargent (2007) to treat misspecification in the alpha component of returns.

To illustrate the substantial improvement in the out-of-sample performance of the *robust mean-variance* (RMV) portfolios resulting from our theoretical insights, we study the same two data sets considered in Ao, Li, and Zheng (2019).[2] The first data set consists of $N = 30$ stock constituents of the Dow Jones Industrial Average (DJIA), and the second of $N = 100$ randomly selected stock constituents of the S&P 500 index, in each case augmented with the Fama-French three factors. We show that our RMV strategy achieves an increase in Sharpe ratio of 125% to 150% relative to the equally weighted portfolio and an increase of 50% to 100% relative to that of the "maximum-Sharpe-ratio estimated and sparse regression strategy" (MAXSER) developed in Ao et al. (2019).[3] We demonstrate the importance of the latent purely asset-specific risk relative to that of observed and pervasive risk factors: for both data sets, the squared correlation between the optimal demand for an asset and its latent asset-specific component is over 95% and the latent component contributes to more than 85% of the total squared Sharpe ratio of the optimal portfolio. Furthermore, we

[1] The analysis for a large number of assets is not just an abstract mathematical exercise but also corresponds to practice. Hedge funds and sovereign-wealth funds hold a large number of assets; for instance, the portfolio of Norges Bank has over 9,000 assets. Moreover, our results have a bite even when the number of assets is as small as 30.

[2] In Section IA.4 of the Internet Appendix, we also use simulated data to evaluate performance.

[3] This improvement is noteworthy given that DeMiguel, Garlappi, and Uppal (2009b, Table 1) and Ao et al. (2019, Table 1) show, respectively, that the $1/N$ and MAXSER portfolios outperform a large number of other strategies.

show that the performance of our strategy improves with the number of risky assets, in contrast to traditional portfolio-selection models.

To summarize, traditional portfolio choice models ignore asset-specific risk and, when correcting for misspecification, focus on addressing misspecification by including additional risk factors; moreover, traditional models find it challenging to handle the case of large N. In contrast, our analysis is founded on the APT, leading to three significant benefits. One, it allows us to distinguish between the alpha and beta components of returns. Two, within the alpha component of returns, it enables us to account for the presence of pricing errors associated with omitted risk factors and omitted asset-specific risk. Three, it rules out arbitrage opportunities; so, unbounded Sharpe ratios are ruled out even when N is large. Theoretically, this leads to the critical insight that, asymptotically, the optimal portfolio weight is dominated by the purely asset-specific pricing errors in the alpha portfolio and that the specification of the beta portfolio can be made immune to missing risk factors. Empirically, we demonstrate that allowing for asset-specific pricing errors leads to a substantial improvement in portfolio performance.

We now explain how our work is related to the existing literature. Just like in Koijen and Yogo (2019), a key quantity that we study is latent asset demand. Their work is *positive* and shows the importance of latent asset demand in explaining empirically observed asset holdings. Thus, in their work, latent asset demand is the wedge between the asset demand implied by a candidate factor model and the *observed* asset demand. In contrast, our work is *normative*: in our APT-based model, latent asset demand is the wedge between the asset demand implied by the candidate factor model and the *optimal* asset demand. The theory we develop explains why, in the presence of model misspecification, an optimal portfolio must include a latent component and why, as the number of assets increases, this component will dominate the one that depends on observed factors. Our empirical results confirm the dominant role of latent asset demand, in particular, the demand arising from the latent purely-asset-specific component.

There is an extensive literature that studies the portfolio-choice problem in the presence of model misspecification using the max-min approach; see, for instance, Trojani and Vanini (2002), Uppal and Wang (2003), Garlappi, Uppal, and Wang (2007), Guidolin and Rinaldi (2009), and Christensen (2017). In contrast to these papers, our work adopts as the "approximating model" the APT, decomposes the mean-variance portfolio into alpha and beta components, and shows that treating misspecification in only the alpha-portfolio component with the max-min approach, which we show is equivalent to imposing the APT no-arbitrage restriction, leads to substantial improvement in empirical out-of-sample performance.

Our work is also related to the extensive empirical literature that aims to improve the performance of mean-variance portfolios by imposing portfolio constraints. Jagannathan and Ma (2003) demonstrate that imposing short-sale constraints is equivalent to shrinking the covariance matrix. DeMiguel, Garlappi, Nogales, and Uppal (2009a) show that using a more general form of the short-sale constraint improves performance further. Economically motivated constraints improve the estimation of time-series forecasts of the equity risk premium (Pettenuzzo, Timmermann, and Valkanov, 2014) and recovery of the minimum variance stochastic discount factor (SDF) under a conservative max-min Sharpe-ratio criterion (Schneider and Trojani, 2019).[4] In our work, the constraint imposed when estimating asset returns follows from the APT's no-arbitrage restriction. Ao et al. (2019) also study estimation of the mean-variance portfolios directly, using a regression approach with a lasso constraint to handle a large number of assets; moreover, they exploit a decomposition of the portfolio weights based on statistical assumptions, whereas our theory is founded on the APT, and thus, on the no-arbitrage restriction.

The notion that mean-variance portfolios can be decomposed into two inefficient portfolios was pioneered by Treynor and Black (1973), who labeled them "active" and "passive" portfolios. Roll (1980) formalizes this idea by introducing "orthogonal portfolios" (i.e. the active portfolios). He shows that for a given inefficient portfolio (i.e. the passive portfolio), there is a continuum of corresponding zero-beta portfolios, in contrast to the case of an efficient portfolio, for which there is only a single orthogonal portfolio.[5] While in Roll (1980) the orthogonal portfolios are a consequence of optimization, we show that they also result from the APT's no-arbitrage restriction.

MacKinlay (1995, page 6) exploits the notion of orthogonal portfolios of Roll (1980) to quantify the behavior of the Gibbons et al. (1989) test when the candidate beta-pricing model is misspecified, stating that "...for the nonrisk-based alternatives the maximum squared Sharpe measure ...can, in principle, be unbounded." Consequently, in subsequent work, the nonrisk-based alternative (i.e. the asset-specific component) was ignored when constructing optimal portfolios. We show that the APT provides the exact condition that must be satisfied in order to include the nonrisk-based alternative in the construction of the optimal portfolio while still satisfying no arbitrage.

Pástor and Stambaugh (1999, 2000) and Pástor (2000), building on MacKinlay (1995), use a Bayesian approach to study the effect of model misspecification on the cost of equity and asset

[4]Brandt (1999) and Aït-Sahalia and Brandt (2001) propose a nonparametric approach for estimating portfolio weights. Ghosh, Julliard, and Taylor (2019) propose an alternative nonparametric approach that builds on an entropy-based estimator of the SDF.

[5]Gibbons, Ross, and Shanken (1989) use the notion of orthogonal portfolios developed by Roll (1980) to show that the GRS test statistic can be represented as the squared t-ratio for testing that the orthogonal portfolio has zero intercept when regressed against the set of candidate factors.

allocation. They show that the posterior mean of the pricing error coincides with a shrinkage version of the OLS estimator of the intercept. Our analysis of misspecification, founded on the APT, also has a shrinkage element, but only for the component associated with asset-specific risk.

MacKinlay and Pástor (2000), again building on MacKinlay (1995), study the portfolio-choice problem when one of the risk factors from the candidate asset-pricing model is omitted. In contrast, we allow not just for omitted risk factors but also for purely asset-specific pricing errors unrelated to common risk while still satisfying the APT no-arbitrage restriction. Moreover, we show analytically and empirically that, as the number of assets increases, the portfolio weights are dominated by asset-specific pricing errors rather than observed and omitted risk factors.

2 A Preview of the Results in a Simplified Setting

In this section, we provide an informal introduction in a simple setting to our main theoretical results to prepare the ground for the formal analysis of the general model that follows.

Suppose that a single-period mean-variance investor with unit risk aversion can invest in N risky assets, with return r_i, for $i = 1, \ldots, N$, and the risk-free asset with return r_f. Then, her optimal portfolio is given by the well-known Markowitz portfolio:

$$w_i^{mv} = \sum_{j=1}^{N} \sigma_{ij}^{(-1)} \mu_j, \tag{1}$$

where $\sigma_{ij}^{(-1)}$ is the (i, j)th element of the inverse of the covariance matrix for the N risky-asset returns and $\mu_j = E(r_j) - r_f$ for $j = 1, \ldots, N$ are the mean excess returns. The expression in (1) highlights the problem we face: because the portfolio weights depend on the N expected excess returns and $N(N + 1)/2$ variances and covariances of returns, a portfolio formed using sample moments performs poorly out of sample, especially when N is large.[6]

To deal with this problem, suppose that the investor postulates a candidate beta-pricing model

$$r_i - r_f = \alpha_i + \beta_i f + \epsilon_i, \quad \text{for } i = 1, \ldots, N, \tag{2}$$

where f is an observed, for simplicity, tradable factor (for instance, the market excess return), with risk premium $\lambda = E(f)$ and finite variance $\mathrm{var}(f) > 0$, β_i is the corresponding beta, ϵ_i is the asset-specific innovation with zero mean and finite variance $\mathrm{var}(\epsilon_i)$ that is independent of f_t, and α_i are the asset-specific pricing errors corresponding to the candidate beta-pricing model in (2),

[6] Note that it is difficult to estimate precisely the expected return even for a single risky asset (Merton, 1980).

which, to rule out arbitrage opportunities, need to satisfy the APT no-arbitrage restriction

$$\sum_{i,j=1}^{N} \alpha_i \alpha_j \operatorname{cov}^{(-1)}(\epsilon_i, \epsilon_j) < \delta_{\text{apt}}, \tag{3}$$

for some finite positive quantity δ_{apt}, where $\operatorname{cov}^{(-1)}(\epsilon_i, \epsilon_j)$ denotes the (i, j)th element of the inverse of the covariance matrix of the ϵ_i.[7] Without loss of generality, we assume that α_i and β_i in (2) are orthogonal. Our portfolio-decomposition results below hold even if the α_i and β_i are correlated across assets, as we demonstrate in Section IA.1 of the Internet Appendix.

Suppose first that there is no misspecification so that the excess returns on the risky assets obey exactly the one-factor model in (2). In this case, α_i is zero for all assets, and one can assume for simplicity that ϵ_i are independent and identically distributed (IID) across assets. Then, the ith weight in the mean-variance portfolio is equal to the weight in the so-called beta portfolio

$$w_i^{\beta} = \frac{1}{N} \left(\frac{\lambda}{\operatorname{var}(f)} \right) \left(\frac{\beta_i}{\bar{\beta^2}} \right), \quad \text{when } N \text{ is large}, \tag{4}$$

assuming pervasiveness of the observed risk factor f, namely $0 < \bar{\beta^2} = N^{-1} \sum_{i=1}^{N} \beta_i^2 < \infty$. In this case, the beta portfolio is efficient and attains the maximum Sharpe ratio $\sqrt{\lambda^2/\operatorname{var}(f)}$, while diversifying away asset-specific risk: $\operatorname{var}\big(\sum_{i=1}^{N} w_i^{\beta} \epsilon_i\big) = \big(\lambda/\bar{\beta^2}\operatorname{var}(f)\big)^2/N^2 \sum_{i=1}^{N} \beta_i^2 \operatorname{var}(\epsilon_i) \to 0$.

Now consider the case where the candidate model (2) is misspecified. Model misspecification can arise for several reasons. Here we consider misspecification because of the potential omission of a pervasive factor, the presence of asset-specific errors (unrelated to the pervasive factor), or both. With misspecification, the beta portfolio continues to satisfy (4) and still diversifies what the investor perceives to be asset-specific risk: $\operatorname{var}\big(\sum_{i=1}^{N} w_i^{\beta} \epsilon_i\big) \to 0$. However, the portfolio w_i^{β} in (4) is no longer on the efficient frontier and $\sqrt{\lambda^2/\operatorname{var}(f)}$ is not the maximum attainable Sharpe ratio.

To see the effect of model misspecification, we can use the APT no-arbitrage restriction to show that in (2) the pricing error α_i, i.e. the asset-specific nonrandom component, and the innovation ϵ_i, i.e. the asset-specific random component, can be decomposed as

$$\alpha_i = A_i \lambda_{miss} + a_i \qquad \text{and} \qquad \epsilon_i = A_i z + \eta_i, \tag{5}$$

where A_i is the loading on the omitted pervasive risk factor z, which has a risk premium λ_{miss}, and a_i is the asset-specific pricing error unrelated to pervasive risk.[8] The η_i represent purely asset-

[7] One has to allow for the asset-specific innovation ϵ_i to be correlated across assets because the investor's candidate beta-pricing model (2) could be misspecified leading to $\operatorname{cov}(\epsilon_i, \epsilon_j) \neq 0$ for different assets i and j.

[8] To simplify the exposition, we have assumed that the investor omits only one pervasive risk factor z, for which the assumption that $E(z) = 0$ and $\operatorname{var}(z) = 1$ is without loss of generality because z is latent. Pervasiveness of the omitted factor z follows when $0 < N^{-1} \sum_{i=1}^{N} A_i^2 < \infty$.

specific risks, for simplicity assumed IID with zero mean and finite variance, $\mathrm{var}(\eta_i) = \mathrm{var}(\eta) > 0$. The α_i component generates the latent asset demand, as shown below.

Consider first the special case of (5), in which the investor omits the pervasive risk factor z, but there is no asset-specific error (i.e. $a_i = 0$ for all assets). This is the case typically considered in the literature. In this case, the mean-variance portfolio weights w_i^{mv} can be expressed as the sum of two portfolios, $w_i^{mv} = w_i^A + w_i^\beta$, with

$$w_i^A = \frac{1}{N}\lambda_{miss}\left(\frac{A_i}{\bar{A^2}}\right), \quad \text{when } N \text{ is large,} \tag{6}$$

where $\bar{A^2} = N^{-1}\sum_{i=1}^N A_i^2$ and w_i^β given in (4). The weight w_i^A in (6) is analogous to w_i^β in (4). Observe that, in contrast to the common belief, in this case the pricing errors do not need to be small to satisfy the APT restriction; the APT restriction (3) is always satisfied, i.e. $\sum_{i=1}^N A_i^2 / \big(\sum_{i=1}^N A_i^2 + \mathrm{var}(\eta)\big)$ remains bounded even when the A_i are very large. Given latency of the omitted factor z, when N is large the Sharpe ratio of the w_i^A portfolio equals $\sqrt{\lambda_{miss}^2} = \sqrt{\lambda_{miss}^2/\,\mathrm{var}(z)}$.

We now consider the general case of misspecification (5) with non-zero asset-specific pricing errors a_i, in addition to an omitted risk factor. This case has received much less attention in the literature, and we show that it leads to portfolios with strikingly different properties. The mean-variance portfolio in this case is

$$w_i^{mv} = w_i^\alpha + w_i^\beta = (w_i^a + w_i^A) + w_i^\beta, \tag{7}$$

with w_i^β given in (4) and where the alpha portfolio is $w_i^\alpha = w_i^a + w_i^A$, with w_i^A given in (6), and

$$w_i^a = \frac{a_i}{\mathrm{var}(\eta)}, \quad \text{when } N \text{ is large.} \tag{8}$$

Equation (7) shows that latent asset demand, represented by the alpha portfolio, is spanned by omitted pervasive risk factors and asset-specific risk. Thus, the APT provides a formal foundation for the composition of latent asset demand.[9] Observe also that the Sharpe ratio of the portfolio with weights w_i^a equals $\sqrt{\sum_{i=1}^N a_i^2/\mathrm{var}(\eta)}$, and thus, is bounded (because of no arbitrage) and strictly positive as long as the asset-specific pricing error a_i is non-zero for at least one asset. This holds regardless of whether the a_i terms are negative or positive. By no-arbitrage and ruling out the pathological case of perfect correlation between the A_i and β_i, the squared Sharpe ratio of the tangency portfolio is

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

[9]The expression in (8) follows from the APT no-arbitrage restriction (3), in particular, from the fact that $\sum_{i=1}^N a_i^2$ must be bounded even for large N. For a formal proof, see Proposition 4.1.

$$= (\mathrm{SR}^a)^2 + (\mathrm{SR}^A)^2 + (\mathrm{SR}^\beta)^2$$

$$= \sum_{i=1}^{N} a_i^2/\mathrm{var}(\epsilon_i) + \lambda_{miss}^2 + \lambda^2/\mathrm{var}(f).$$

The above decomposition of the portfolio squared Sharpe ratio holds even if α_i and β_i are not orthogonal.

To implement the w_i^{mv} portfolio, when we *estimate* (7) subject to the APT no-arbitrage restriction in (3), we obtain

$$\hat{w}_i^{mv} = \sqrt{\delta_{\mathrm{apt}}}\,\hat{w}_i^\alpha + \hat{w}_i^\beta, \tag{9}$$

where $\hat{w}_i^\alpha$ and $\hat{w}_i^\beta$ are functions of the estimated parameters.[10] The above expression shows that as we decrease δ_{apt} we effectively "shrink" the alpha portfolio; in the case where $\delta_{\mathrm{apt}} = 0$, the mean-variance portfolio is given by just the beta portfolio, and exact pricing holds.

Based on the above, we collect our key insights regarding the alpha and beta portfolios for the case of general misspecification. These results represent a major departure from both the case of a correctly specified model with weights (4) and the case in which one allows for only omitted pervasive factors with weights (6).

1. Two-fund separation holds for the mean-variance portfolio weight, $w_i^{mv} = w_i^\alpha + w_i^\beta$, where $w_i^\alpha = w_i^a + w_i^A$, because the alpha and beta portfolio each span orthogonal sources of risk. Consequently, the squared Sharpe ratio of the mean-variance portfolio equals the sum of the squared Sharpe ratios of the alpha and beta portfolios, implying that the alpha and beta portfolios are inefficient. The general case of this result is given in Proposition 4.1.

2. The component of latent asset demand given by the portfolio w_i^a dominates the beta portfolio weights; i.e. $w_i^\beta/w_i^a \to 0$ when N diverges (w_i^a also dominates w_i^A). The general case of this result is given in Proposition 4.2.

3. The beta portfolio weights in (4), depend on the parameters λ and $\mathrm{var}(f)$. However, the parameter-free equally weighted portfolio weights, $w_i^{ew} = 1/N$, also achieve the same Sharpe ratio as that of w_i^β, $\lambda/\sqrt{\mathrm{var}(f)}$, under some conditions that we identify; thus, the beta portfolio can be replaced by the equally weighted portfolio without any loss of performance, implying that the beta portfolio is immune to errors in estimating λ and $\mathrm{var}(f)$. The general case of this result is given in Proposition 4.3.

[10] Note that $\hat{w}_i^\alpha = \hat{w}_i^\alpha(\delta_{\mathrm{apt}}) = \sqrt{\delta_{\mathrm{apt}}}\hat{w}_i^\alpha(1)$.

4. If a mean-variance investor is averse to model misspecification (in the sense of Hansen and Sargent (2007)), the investor's optimal estimated "robust" portfolio weights, labeled $\hat{w}_i^{rmv}$, obtained by solving a max-min optimization problem and estimated by imposing the APT restriction (3) for a given δ^*_{apt}, is

$$\hat{w}_i^{rmv} = \Big(\frac{\sqrt{\delta^*_{apt}}}{\phi}\Big)\hat{w}_i^{\alpha} + \hat{w}_i^{\beta}, \tag{10}$$

   for some $\phi \geq 1$ representing the investor's aversion to misspecification, with $\phi = 1$ the case of neutrality to misspecification. It follows that there exists a $\delta_{apt} = \sqrt{\delta^*_{apt}}/\phi$ so that the (estimated) portfolio weights in (9) are identical to the robust portfolio weights in (10) for some $\phi > 1$. That is, the APT leads to the *same* portfolio as the robust portfolio of an investor averse to model misspecification. The general case of this result is given in Proposition 4.4.

In summary, whenever some of the w_i^a weights in (8) are non-zero, then purely asset-specific risk is *not* diversified away. In fact, it is optimal for to bear purely asset-specific risk to earn the associated asset-specific premium and thereby achieve mean-variance efficiency.

3 Candidate Beta Models and the APT

In this section, we describe the formal framework used for our analysis. We assume that the investor starts with a candidate model, namely an asset-pricing model linear in K observed factors $\mathbf{f}_t$ (either tradable or nontradable), referred to as the beta-pricing model and aims to use this model to quantify the means, variances, and covariances of asset returns for asset allocation with a large number N of risky assets.[11] However, the investor's candidate beta-pricing model may be misspecified. The key idea of this paper is to use the APT to identify and mitigate the effect of pricing errors caused by the use of the candidate beta-pricing model.

The classical APT of Chamberlain (1983) has no observed factors. Instead, all factors are assumed to be latent, about which the investor can obtain full knowledge using population principal-component analysis.[12] Consequently, only small pricing errors could potentially be present. In contrast, in our framework, because the investor is assumed to start with a possibly incomplete set

[11] We study a market with a countably infinite number of assets, just like in the classical APT (see Chamberlain and Rothschild (1983)). Gagliardini, Ossola, and Scaillet (2016), building on the work of Al-Najjar (1998), extend the APT to allow for an uncountable number of assets and also relax the boundedness assumption of the maximum eigenvalue of the residual covariance matrix.

[12] It is important to note that the APT is a model of the random component of returns, and is silent about expected returns. Black (1995, p. 168) recognizes this 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."

of candidate observed factors $\mathbf{f}_t$ driving the beta model, the pricing errors can also include omitted pervasive risk factors. There is a misperception that the APT framework allows only for small pricing errors but we explain below that the APT applies even when the pricing errors include omitted pervasive risk factors.

Let r_{ft} denote the return on the risk-free asset, and let $\mathbf{r}_t = (r_{1t}, r_{2t}, \ldots, r_{Nt})'$ denote the N-dimensional vector of one-period returns on the risky 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 $\mathbf{r}_t$, which have the subscript t. Given an arbitrary portfolio strategy s with weights $\mathbf{w}_N^s = (w_1^s, w_2^s, \ldots, w_N^s)'$, and using $\mathbf{1}_N$ to denote an N-dimensional vector of ones, we define the associated portfolio return as

$$r_t^s = \mathbf{r}_t'\mathbf{w}_N^s + r_{ft}(1 - \mathbf{1}_N'\mathbf{w}_N^s),$$

with finite mean, standard deviation, and Sharpe ratio defined, respectively, as

$$\mu^s = E(\mathbf{r}_t)'\mathbf{w}_N^s + r_{ft}(1 - \mathbf{1}_N'\mathbf{w}_N^s), \qquad \sigma^s = \sqrt{\mathrm{var}(r_t^s)}, \quad \text{and} \quad \mathrm{SR}^s = \frac{\mu^s - r_{ft}}{\sigma^s}.$$

We now state our main assumptions, adapting the APT assumptions to our setup where the investor starts with an initial, possibly misspecified, candidate beta-pricing model.

Assumption 3.1 (Candidate beta-pricing model)**.** *We assume that the N-dimensional vector $\mathbf{r}_t$ of asset returns can be characterized by*

$$\mathbf{r}_t = \boldsymbol{\mu}_N + \mathbf{B}_N\left(\mathbf{f}_t - E(\mathbf{f}_t)\right) + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\mu}_N$ is the $N \times 1$ vector of expected returns and $\mathbf{B}_N$ is the $N \times K$ full-rank matrix of factor loadings, with $K < N$. At any time t, the $K \times 1$ vector of common observed factors, $\mathbf{f}_t$, is distributed with mean $E(\mathbf{f}_t)$ and $K \times K$ covariance matrix $\boldsymbol{\Omega}$, and the $N \times 1$ vector of innovations $\boldsymbol{\varepsilon}_t$ is distributed with zero mean and the $N \times N$ covariance matrix $\boldsymbol{\Sigma}_N$, with $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}_N$ being positive definite. Moreover, $\boldsymbol{\varepsilon}_t$ and $\mathbf{f}_t$ are uncorrelated.

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

$$E\left[(\mathbf{r}_t - \boldsymbol{\mu}_N)(\mathbf{r}_t - \boldsymbol{\mu}_N)'\right] = \mathbf{V}_N = \mathbf{B}_N\,\boldsymbol{\Omega}\,\mathbf{B}_N' + \boldsymbol{\Sigma}_N. \tag{11}$$

In the assumption stated below and throughout the paper, we use δ to denote an arbitrary positive scalar, not necessarily having the same value.

Assumption 3.2 (No asymptotic arbitrage). *As $N \to \infty$, there is no sequence of portfolios for which, along some subsequence $\tilde{N}$,*

$$\mathrm{var}(\mathbf{r}_t' \mathbf{w}_{\tilde{N}}^p) \to 0 \quad \textit{and} \quad (\boldsymbol{\mu}_{\tilde{N}} - r_{ft}\mathbf{1}_{\tilde{N}})' \mathbf{w}_{\tilde{N}}^p \geq \delta > 0 \quad \textit{for all} \quad \tilde{N}.$$

When Assumptions 3.1 and 3.2 hold, with the largest eigenvalue[13] of $(\mathbf{B}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1}$ satisfying $g_{1K}((\mathbf{B}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1}) \to 0$ as $N \to \infty$, then there exists some positive number δ_{apt} and a unique vector of risk premia $\boldsymbol{\lambda}$ such that the vector of pricing errors

$$\boldsymbol{\alpha}_N = (\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N) - \mathbf{B}_N \boldsymbol{\lambda}, \tag{12}$$

satisfies the APT restriction

$$\sup_N \boldsymbol{\alpha}_N' \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N \leq \delta_{\text{apt}} < \infty. \tag{13}$$

In equation (12), the vector of true risk premia $\boldsymbol{\lambda}$ is the limit of $\left(\mathbf{B}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N\right)^{-1} \mathbf{B}_N' \boldsymbol{\Sigma}_N^{-1} (\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N)$, which exists and is well defined under Assumptions 3.1 and 3.2. When the observed risk factors are tradable, then $\boldsymbol{\lambda} = E(\mathbf{f}_t) - r_{ft}\mathbf{1}_K$. The APT theory is silent about the magnitude of δ_{apt} in (13). Ross (1977) suggests choosing δ_{apt} as a multiple of the Sharpe ratio of the market. In practice, one can estimate it using cross-validation methods commonly used in statistics, as explained in Internet Appendix IA.3.

As is typical in the APT literature (Chamberlain, 1983, 1987; Huberman, 1982; Ingersoll, 1984) and the portfolio-choice literature (Daniel, Mota, Rottke, and Santos, 2020; Kelly, Pruitt, and Su, 2020; Kim, Korajczyk, and Neuhierl, 2021), one could assume that $\boldsymbol{\alpha}_N$ and $\boldsymbol{\beta}_N$ in (12) are orthogonal. As we demonstrate in Section IA.1 of the Internet Appendix, this would be without loss of generality for our results on portfolio choice.

The restriction in (13), a consequence of asymptotic no-arbitrage, links the pricing error $\boldsymbol{\alpha}_N$ to the residual covariance matrix, $\boldsymbol{\Sigma}_N$. Two complementary cases are possible. In the first case, the pricing errors are unrelated to pervasive factors and, therefore, must be small on average, i.e. all the eigenvalues of $\boldsymbol{\Sigma}_N$ are bounded: $\sup_N \boldsymbol{\alpha}_N' \boldsymbol{\alpha}_N \leq \sup_N g_{1N}(\boldsymbol{\Sigma}_N) \boldsymbol{\alpha}_N' \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N \leq g_{1N}(\boldsymbol{\Sigma}_N) \delta_{\text{apt}} < \infty$. The classical APT literature (Huberman, 1982; Ingersoll, 1984) has focused on studying this case. In the second case, which is the focus of our framework, the pricing errors also contain omitted pervasive risk factors and, therefore, can be large on average (i.e. at least one of the eigenvalues of $\boldsymbol{\Sigma}_N$ diverges with N). However, as we now explain, even in this case, the APT no-arbitrage restriction (13) is satisfied.

[13] Consider a symmetric $M \times M$ matrix $\mathbf{A}$. Let $g_{iM}(\mathbf{A})$ denote the ith eigenvalue of $\mathbf{A}$ in decreasing order for $1 \leq i \leq M$. Then, the maximum eigenvalue is $g_{1M}(\mathbf{A})$, and the minimum eigenvalue is $g_{MM}(\mathbf{A})$.

To understand this, observe that once one obtains $\boldsymbol{\lambda}$, given Assumptions 3.1–3.2 and applying Chamberlain (1983) to the returns *net* of the observed risk factors, i.e. to $\mathbf{r}_t - \mathbf{B}_N(\mathbf{f}_t + \boldsymbol{\lambda} - E(\mathbf{f}_t))$, which equals $\boldsymbol{\alpha}_N + \boldsymbol{\varepsilon}_t$, its covariance matrix equals

$$\boldsymbol{\Sigma}_N = \mathbf{A}_N \mathbf{A}_N' + \mathbf{C}_N, \tag{14}$$

where $\mathbf{A}_N$ is an $N \times p$ matrix of loadings corresponding to the p omitted pervasive factors, and $\mathbf{C}_N$ is an $N \times N$ positive semi-definite matrix with uniformly bounded eigenvalues. The expression in (14) shows that the residual covariance matrix $\boldsymbol{\Sigma}_N$ has two terms: the first term, $\mathbf{A}_N \mathbf{A}_N'$, represents the risk from the missing factors and the second term, $\mathbf{C}_N$, represents purely asset-specific risk.

The pricing error $\boldsymbol{\alpha}_N$ in (12), corresponding to the candidate beta-pricing model, satisfies

$$\boldsymbol{\alpha}_N = \mathbf{A}_N \boldsymbol{\lambda}_{\text{miss}} + \mathbf{a}_N \text{ such that } \sup_N \mathbf{a}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{a}_N \leq \delta < \infty. \tag{15}$$

The first term, $\mathbf{A}_N \boldsymbol{\lambda}_{\text{miss}}$, of the pricing error $\boldsymbol{\alpha}_N$ in (15) is associated with the p omitted pervasive risk factors, where $\boldsymbol{\lambda}_{\text{miss}}$ is the vector of risk premia for the missing factors.[14] The second term, $\mathbf{a}_N$, is a non-zero $N \times 1$ vector associated with the purely asset-specific pricing errors and is the compensation for purely asset-specific risk. The recognition that purely asset-specific risk is compensated is a crucial departure from the standard asset-pricing framework with exact pricing, where there is no compensation for pure idiosyncratic risk, and will play a vital role in the optimal portfolio.

To show that the APT restriction continues to hold, despite the possibly large pricing errors because of the omitted pervasive risk factors, we use (14) and (15) to obtain

$$\boldsymbol{\alpha}_N' \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N = (\mathbf{A}_N \boldsymbol{\lambda}_{\text{miss}} + \mathbf{a}_N)' (\mathbf{A}_N \mathbf{A}_N' + \mathbf{C}_N)^{-1} (\mathbf{A}_N \boldsymbol{\lambda}_{\text{miss}} + \mathbf{a}_N). \tag{16}$$

To show that $\boldsymbol{\alpha}_N' \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N$ is bounded, we need to demonstrate that both $(\mathbf{A}_N \boldsymbol{\lambda}_{\text{miss}})' \boldsymbol{\Sigma}_N^{-1} (\mathbf{A}_N \boldsymbol{\lambda}_{\text{miss}}) = (\mathbf{A}_N \boldsymbol{\lambda}_{\text{miss}})' (\mathbf{A}_N \mathbf{A}_N' + \mathbf{C}_N)^{-1} (\mathbf{A}_N \boldsymbol{\lambda}_{\text{miss}})$ and $\mathbf{a}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{a}_N$ are bounded. The former is always bounded because it contains $\mathbf{A}_N$ both in the denominator and numerator; therefore, an arbitrarily large $\mathbf{A}_N$ would still leave this term bounded. Boundedness of the latter follows from (15).

4 The Portfolio Choice Problem under the APT

This section contains our main theoretical results, presented in four propositions.

[14] Pervasiveness of the omitted latent risk factors, with loadings $\mathbf{A}_N$, follows from the assumption that $g_{jN}(\boldsymbol{\Sigma}_N)$ for $1 \leq j \leq p$ are increasing for large N. Notice that, because of the latency of the omitted risk factors, $\mathbf{A}_N$ and $\boldsymbol{\lambda}_{\text{miss}}$ are identified up to an unknown nonsingular rotation.

4.1 The Alpha and Beta Portfolios

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

$$\mathbf{w}_N^{\mathrm{mv}} = \underset{\mathbf{w}_N}{\mathrm{argmax}} \left((\mathbf{w}_N'\boldsymbol{\mu}_N + (1 - \mathbf{w}_N'\mathbf{1}_N) r_{ft}) - \frac{\gamma}{2}\mathbf{w}_N'\mathbf{V}_N\mathbf{w}_N \right), \tag{17}$$

where $0 < \gamma < \infty$ is the coefficient of risk aversion, $\mathbf{w}_N^{\mathrm{mv}} = (w_1^{\mathrm{mv}}, \ldots, w_N^{\mathrm{mv}})'$ is the vector of portfolio weights in the N risky assets, and the investment in the risk-free asset is given by $1 - \mathbf{1}_N'\mathbf{w}_N^{\mathrm{mv}}$. The well-known solution to the optimization problem in (17) is

$$\mathbf{w}_N^{\mathrm{mv}} = \frac{1}{\gamma}\mathbf{V}_N^{-1}(\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N). \tag{18}$$

The return on portfolio $\mathbf{w}_N^{\mathrm{mv}}$ has Sharpe ratio $\mathrm{SR}^{\mathrm{mv}} = \left((\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N)'\mathbf{V}_N^{-1}(\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N)\right)^{1/2}$ with mean and standard deviation given by, respectively, $\mu^{\mathrm{mv}} - r_{ft} = \gamma^{-1}(\mathrm{SR}^{\mathrm{mv}})^2$ and $\sigma^{\mathrm{mv}} = \gamma^{-1}\mathrm{SR}^{\mathrm{mv}}$.

The following proposition establishes the relations that exist between the mean-variance portfolio, $\mathbf{w}_N^{\mathrm{mv}}$, and the two inefficient portfolios, $\mathbf{w}_N^{\alpha}$ and $\mathbf{w}_N^{\beta}$, that depend on the alpha and beta components of returns, respectively. Throughout the analysis, when we write $\mathbf{w}_N^{x} = \mathbf{w}_N^{y} + \mathcal{O}(1)$, we mean that, as N increases, the returns of portfolios x and y have the same mean, variance, and Sharpe ratio. A more detailed explanation of this notation is given at the start of Appendix A.

Proposition 4.1 (Two-fund separation under the APT)**.** *Suppose that the vector of asset returns,* $\mathbf{r}_t$*, satisfies Assumptions 3.1 and 3.2,* $\boldsymbol{\alpha}_N \neq \mathbf{0}_N$ *and* $\boldsymbol{\lambda} \neq \mathbf{0}_K$*,*[15] *and that* $\mathbf{A}_N$*,* $\mathbf{B}_N$*, and* $\mathbf{1}_N$ *are* $\mathbf{C}_N$*-regular at rate* N*,*[16] $\mathbf{A}_N$ *and* $\mathbf{B}_N$ *are not asymptotically collinear,*[17] *and that the row sums of* $\mathbf{A}_N, \mathbf{B}_N$ *and* $\mathbf{C}_N^{-1}$ *are uniformly bounded.*[18] *Then:*

[15] We denote by $\mathbf{0}_n$ the n-dimensional vector of zeros. Proposition 4.1 is trivial when either $\boldsymbol{\alpha}_N = \mathbf{0}_N$ or $\boldsymbol{\lambda} = \mathbf{0}_K$, so we rule this out without loss of generality.

[16] Building on Ingersoll (1984, eq. (21)), a matrix $\mathbf{D_N}$ is $\mathbf{C}_N$-regular at rate $f(N)$ if there exists an increasing function of N, $f(N)$, such that for any $1 \leq j \leq K$, the eigenvalues $g_{jK}(\frac{1}{f(N)}\mathbf{D}_N'\mathbf{C}_N^{-1}\mathbf{D}_N) \to \delta_j > 0$, where δ_j is some finite positive constant. Note that this condition imposes that all the eigenvalues diverge at precisely the same rate. Suppose $\mathbf{C}_N$ has eigenvalues that are both bounded above and away from zero for every N. In that case, this property depends only on $\mathbf{D_N}$, but adopting our definition of $\mathbf{C}_N$-regularity simplifies the proof.

[17] By asymptotic collinearity we mean that either $\mathbf{A}_N'\mathbf{M}_{B_N}\mathbf{A}_N \to 0$ or $\mathbf{B}_N'\mathbf{M}_{A_N}\mathbf{B}_N \to 0$ 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 $\mathbf{A}_N = \mathbf{B}_N\mathbf{H} + \mathbf{G}_N$, for some constant $K \times p$ matrix $\mathbf{H}$ and some residual matrix $\mathbf{G}_N$ satisfying $\mathbf{G}_N'\mathbf{G}_N \to 0$. The pathological case where the loadings to the latent factors are *perfectly* collinear to the loadings on the observed factors needs to be ruled out for identification. For example, if $\mathbf{A}_N = \mathbf{B}_N$, the model is linear in the aggregate factors $\mathbf{F}_t = \mathbf{f}_t + \mathbf{f}_{t,miss}$ with risk premia $\boldsymbol{\Lambda} = \boldsymbol{\lambda} + \boldsymbol{\lambda}_{miss}$ so one would not be able to disentangle observed and latent factors.

[18] Given an $N \times M$ matrix $\mathbf{D}_N$, we say its row sums are uniformly bounded when $\sup_N \sum_{i=1}^{M} |d_{ji}| \leq \delta < \infty$, for some arbitrary positive scalar δ.

(i) For any $N > K$ the mean-variance portfolio weights satisfy the following decomposition:

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

$$\text{with} \qquad \mathbf{w}_N^\alpha = \frac{1}{\gamma^\alpha} \mathbf{\Sigma}_N^+ \boldsymbol{\alpha}_N + \mathcal{O}(1) \qquad \text{and} \qquad \mathbf{w}_N^\beta = \frac{1}{\gamma^\beta} \mathbf{V}_N^{-1} \mathbf{B}_N \boldsymbol{\lambda}, \tag{19}$$

where γ^α and γ^β are some positive constants, $\phi^\alpha = \frac{\gamma^\alpha}{\gamma}$, $\phi^\beta = \frac{\gamma^\beta}{\gamma} = 1 - \phi^\alpha$,[19] and[20]

$$\mathbf{\Sigma}_N^+ = \left[\mathbf{\Sigma}_N^{-1} - \mathbf{\Sigma}_N^{-1} \mathbf{B}_N (\mathbf{B}_N' \mathbf{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}_N' \mathbf{\Sigma}_N^{-1} \right]. \tag{20}$$

(ii) As $N \to \infty$, $\mathbf{w}_N^\beta$ ($\mathbf{w}_N^\alpha$) is the minimum-variance portfolio *that is conditionally and unconditionally (with respect to the factors) orthogonal to $\mathbf{w}_N^\alpha$ ($\mathbf{w}_N^\beta$): $(\mathbf{w}_N^\alpha)' \mathbf{\Sigma}_N \mathbf{w}_N^\beta \to 0$ and $(\mathbf{w}_N^\alpha)' \mathbf{V}_N \mathbf{w}_N^\beta \to 0$.*

(iii) As $N \to \infty$, we have two-fund separation: the inefficient *portfolios $\mathbf{w}_N^\alpha$ and $\mathbf{w}_N^\beta$ can generate all the portfolios on the* efficient *mean-variance frontier of risky assets, namely,*

$$(\mathrm{SR}^{mv})^2 - \left((\mathrm{SR}^\alpha)^2 + (\mathrm{SR}^\beta)^2 \right) \to 0, \tag{21}$$

where the Sharpe ratios of the $\mathbf{w}_N^\alpha$ and $\mathbf{w}_N^\beta$ portfolios satisfy $0 \le \mathrm{SR}^\alpha < \infty$, $0 \le \mathrm{SR}^\beta < \infty$, with

$$\mathrm{SR}^\alpha = (\boldsymbol{\alpha}_N' \mathbf{\Sigma}_N^+ \boldsymbol{\alpha}_N)^{\frac{1}{2}} + \mathcal{O}(1), \qquad \text{and} \qquad \mathrm{SR}^\beta = (\boldsymbol{\lambda}' \mathbf{\Omega}^{-1} \boldsymbol{\lambda})^{\frac{1}{2}} + \mathcal{O}(1). \tag{22}$$

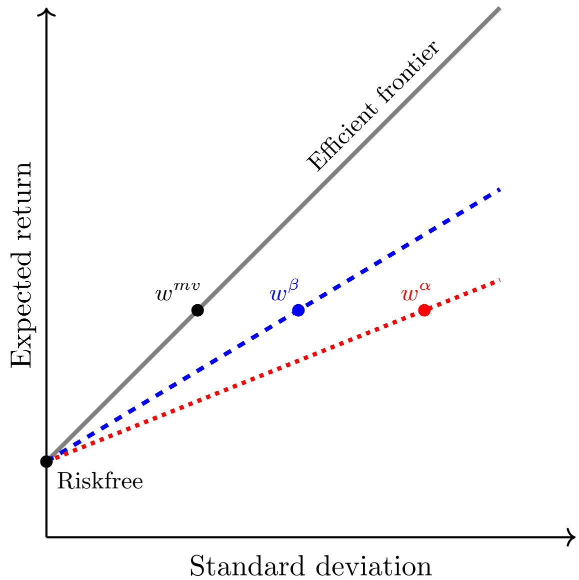
Note from (19) that, for large N, portfolio $\mathbf{w}_N^\alpha$ is given by $\mathbf{\Sigma}_N^+ \boldsymbol{\alpha}_N / \gamma^\alpha$, implying that (asymptotically) it depends only on the pricing errors but not on the risk premia, $\boldsymbol{\lambda}$, nor on the factor-covariance matrix, $\mathbf{\Omega}$, which is why we label this the "alpha" portfolio. This portfolio represents latent asset demand because it depends on the pricing errors, which are spanned by latent pervasive risk factors and latent asset-specific risk. On the other hand, the portfolio $\mathbf{w}_N^\beta$ depends on factor exposures and their risk premia, but not on the pricing errors, $\boldsymbol{\alpha}_N$. Moreover, because $\mathbf{\Sigma}_N^+ \mathbf{B}_N = \mathbf{0}$, the $\mathbf{w}_N^\alpha$ portfolio diversifies away the observed risk factors, and thus, is beta-factor neutral.

The (asymptotic) orthogonality property, $(\mathbf{w}_N^\alpha)' \mathbf{\Sigma}_N \mathbf{w}_N^\beta \to 0$, says that the two portfolios $\mathbf{w}_N^\alpha$ and $\mathbf{w}_N^\beta$ are uncorrelated, conditional on the factors, as N gets large. This property is true *regardless* of whether one assumes that $\boldsymbol{\alpha}_N$ and $\boldsymbol{\beta}_N$ in (12) are orthogonal for any finite N, as

[19] The choice of γ^α and γ^β depends on which type of mean-variance strategy is considered. For example, for mean-variance portfolios with target mean μ^*, one selects γ^α as the ratio of the share of the contribution of $\mathbf{w}_N^\alpha$ to the expected return on the mean-variance portfolio with unit risk aversion, $(\boldsymbol{\mu}_N - r_{ft} \mathbf{1}_N) \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_{ft} \mathbf{1}_N)$, over the target excess mean return, $\mu^* - r_{ft}$. Then the role of the ϕ^α and ϕ^β coefficients is to ensure that the $\mathbf{w}_N^\alpha$ and $\mathbf{w}_N^\beta$ portfolios achieve the same target mean return μ^* as $\mathbf{w}_N^{\mathrm{mv}}$.

[20] Note the Sherman-Morrison-Woodbury formula implies $\mathbf{V}_N^{-1} = \left[\mathbf{\Sigma}_N^{-1} - \mathbf{\Sigma}_N^{-1} \mathbf{B}_N (\mathbf{\Omega}^{-1} + \mathbf{B}_N' \mathbf{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}_N' \mathbf{\Sigma}_N^{-1} \right]$. Setting $\mathbf{\Omega}^{-1} = \mathbf{0}$ (a $K \times K$ matrix of zeros) in the expression for $\mathbf{V}_N^{-1}$ leads to the result for $\mathbf{\Sigma}_N^+$ in (20) because when N increases $\mathbf{B}_N' \mathbf{\Sigma}_N^{-1} \mathbf{B}_N$ increases without bound, dominating $\mathbf{\Omega}^{-1}$, implying that $\mathbf{V}_N^{-1}$ behaves like $\mathbf{\Sigma}_N^+$.

Figure 1: Decomposition of the mean-variance portfolio
In this figure, we plot the mean-variance portfolio, $\mathbf{w}_N^{\text{mv}}$, and its decomposition into two *inefficient* portfolios: one that depends only on the pricing errors, $\mathbf{w}_N^{\alpha}$, and another that depends only on the factor exposure and their premia, $\mathbf{w}_N^{\beta}$.



shown in Section IA.1 of the Internet Appendix. This explains why we have not needed to assume orthogonality between $\boldsymbol{\alpha}_N$ and $\boldsymbol{\beta}_N$ when describing the data generating process for returns.

In addition to this, if one constructed the *minimum-variance* portfolio that is orthogonal to $\mathbf{w}_N^{\alpha}$, the resulting portfolio would be $\mathbf{w}_N^{\beta}$, and vice versa. That is, even though the $\mathbf{w}_N^{\alpha}$ and $\mathbf{w}_N^{\beta}$ portfolios are obtained simply by relying on the APT decomposition of the total mean return, these portfolios can also be characterized as the result of an optimization that is described in Proposition IA.2 of the Internet Appendix, which extends Roll (1980) to the case where, in addition to investing in risky assets, one can also invest in a risk-free asset.[21] This mutual optimality property of the $\mathbf{w}_N^{\alpha}$ and $\mathbf{w}_N^{\beta}$ portfolios drives the two-fund separation result: the alpha and beta portfolios, despite being inefficient (because they have a smaller slope than the capital market line), span the entire efficient frontier of risky assets, as illustrated in the left-hand-side plot of Figure 1.

The result in equation (22) leads to some new interpretations, especially given the representation $\text{SR}^{\alpha} = (\boldsymbol{\alpha}_N' \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N)^{\frac{1}{2}} + o(1)$, which holds by slightly strengthening our assumptions.[22] First,

[21] The (asymptotic) optimality of $\mathbf{w}_N^{\alpha}$ and $\mathbf{w}_N^{\beta}$ also emerges from recognizing that the mean and variance of the corresponding portfolio returns are proportional to one another, just like for efficient mean-variance portfolios, implying that the associated Sharpe ratios can be expressed as a quadratic form. In fact, $\mu^{\alpha} - r_{ft} = \frac{1}{\gamma^{\alpha}} \boldsymbol{\alpha}_N' \boldsymbol{\Sigma}_N^{+} \boldsymbol{\alpha}_N + o(1)$, $(\sigma^{\alpha})^2 = \frac{1}{(\gamma^{\alpha})^2} \boldsymbol{\alpha}_N' \boldsymbol{\Sigma}_N^{+} \boldsymbol{\alpha}_N + o(1)$, $\mu^{\beta} - r_{ft} = \frac{1}{\gamma^{\beta}} \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda} + o(1)$, and $(\sigma^{\beta})^2 = \frac{1}{(\gamma^{\beta})^2} \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda} + o(1)$.

[22] When $\boldsymbol{\alpha}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N = o(N^{\frac{1}{2}})$, then $\boldsymbol{\alpha}_N' \boldsymbol{\Sigma}_N^{+} \boldsymbol{\alpha}_N = \boldsymbol{\alpha}_N' \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N - \boldsymbol{\alpha}_N \mathbf{B}_N' \boldsymbol{\Sigma}_N^{-1} (\mathbf{B}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N \boldsymbol{\alpha}_N = \boldsymbol{\alpha}_N' \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N + o(N^{\frac{1}{2}}) \mathcal{O}(N^{-1}) o(N^{\frac{1}{2}}) = \boldsymbol{\alpha}_N' \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N + o(1)$.

the quantity on the left-hand side of the APT restriction in (13), $\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\alpha}_N$, is the same (asymptotically) as the square of the Sharpe ratio of the alpha portfolio, $(\mathrm{SR}^{\alpha})^2$. Thus, the APT restriction in (13), which is typically interpreted as a bound on the pricing errors, can instead be interpreted as a bound on the Sharpe ratio of the alpha portfolio.[23] Second, the bound on the square of the Sharpe ratio of the alpha portfolio can also be seen as providing a theoretical rationalization for the use of Sharpe ratios in the "no-good-deal bound" of Cochrane and Saá-Requejo (2001).[24]

Proposition 4.2 (Weights of alpha and beta portfolios for large N)**.** *Under the assumptions of Proposition 4.1, the alpha and beta portfolios weights are of the order* $|w_{N,i}^{\alpha}| = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$ *and* $|w_{N,i}^{\beta}| = \mathcal{O}\left(\frac{1}{N}\right)$*, where* $x = \mathcal{O}(y)$ *means that* $|x|/y$ *is bounded, implying that, element by element, for some constant* δ*, as* $N \to \infty$*,*

$$\frac{w_{N,i}^{\beta}}{w_{N,i}^{\alpha}} \to 0 \quad \textit{if} \quad \mathbf{a}_N \neq \mathbf{0}_N, \qquad \textit{and} \qquad \frac{w_{N,i}^{\beta}}{w_{N,i}^{\alpha}} \to \delta \quad \textit{if} \quad \mathbf{a}_N = \mathbf{0}_N. \tag{23}$$

To understand the intuition for the dominance of the weights of the alpha portfolio if $\mathbf{a}_N \neq \mathbf{0}_N$, recall that mean-variance portfolio optimization aims to maximize the portfolio Sharpe ratio. There are two sources of risk in the APT: factor exposure and asset-specific exposure. The factor exposure $\mathbf{w}_N^{\alpha}$ is zero—*irrespective* of the rate at which the weights decrease—because of the orthogonality of $\mathbf{w}_N^{\alpha}$ to $\mathbf{B}_N$ for any N. Regarding exposure to asset-specific risk, the elements of $\mathbf{w}_N^{\alpha}$ cannot decrease faster than $1/\sqrt{N}$ because then the asset-specific risk of the portfolio goes to zero; however, the asset-specific risk of the alpha portfolio coincides with its Sharpe ratio, implying that the Sharpe ratio would also go to zero. On the other hand, the APT restriction in (13) does not allow the rate at which the weights decrease to be slower than $1/\sqrt{N}$. Thus, the rate of $1/\sqrt{N}$ strikes the correct balance between optimizing the risk and return of the $\mathbf{w}_N^{\alpha}$ portfolio. On the other hand, if $\mathbf{a}_N = \mathbf{0}_N$, we face exposure only to latent factors (if $\mathbf{A}_N$ is not the null matrix), which affects both the mean and variance of asset returns so that the APT restriction is redundant and, therefore, the weights of the alpha portfolio behave like those for the beta portfolio.[25]

---

[23]The decomposition of the square of the Sharpe ratio of the mean-variance portfolio in (21) is also obtained in Treynor and Black (1973) for the case of the single-index model with a diagonal covariance matrix for the residuals. Our insight that $\mathbf{a}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{a}_N$ represents the squared Sharpe ratio of the $\mathbf{a}_N$-portfolio leads to a sharper bound on $\mathbf{a}_N'\mathbf{a}_N$ than the one established in Chamberlain and Rothschild (1983, Theorem 4 and Corollary 2), which instead uses the squared Sharpe ratio of the overall efficient portfolio: $\mathbf{a}_N'\mathbf{a}_N \leq g_{1N}(\mathbf{C}_N)\mathbf{a}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{a}_N \leq g_{1N}(\mathbf{C}_N)\boldsymbol{\mu}_N'\mathbf{V}_N^{-1}\boldsymbol{\mu}_N$.

[24]Korsaye, Quaini, and Trojani (2020) show how the "no-good-deal-bounds" can be formalized through the notion of "smart" stochastic discount factors.

[25]The $\mathbf{w}_N^{\alpha}$ portfolio dominates the $\mathbf{w}_N^{\beta}$ portfolio across other norms besides the sup norm criterion, which is the norm used in Green and Hollifield (1992). In particular, the sum of the squared portfolio weights $\mathbf{w}_N^{\alpha\prime}\mathbf{w}_N^{\alpha}$, which is the same notion adopted in Chamberlain (1983), is always bounded, whereas $\mathbf{w}_N^{\beta\prime}\mathbf{w}_N^{\beta}$ always converges to zero. Moreover, the sum of the portfolio weights $|\mathbf{1}_N'\mathbf{w}_N^{\alpha}|$ can diverge to infinity, whereas $|\mathbf{1}_N'\mathbf{w}_N^{\beta}|$ is always bounded.

Let us now look at the weights of the beta portfolio, $\mathbf{w}_N^\beta$. If the weights decrease at any rate slower than $1/N$, then the systematic exposure explodes because the factors are pervasive. 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 the same as that for the risk of the portfolio. So the rate of $1/N$ strikes the correct balance between optimizing the risk and return of the $\mathbf{w}_N^\beta$ portfolio. Observe that the rate $1/N$ makes the $\mathbf{w}_N^\beta$ portfolio well diversified, even with respect to idiosyncratic exposure, enhancing its Sharpe ratio even further.

The dominance of the elements of $\mathbf{w}_N^\alpha$ relative to those of $\mathbf{w}_N^\beta$ characterized in (23) highlights the importance of studying latent asset demand in optimal portfolios. Moreover, their disparate asymptotic behavior suggests that it may be advantageous to use different approaches for addressing model misspecification in the alpha and beta portfolios. We describe these approaches below.

4.2 Mitigating Misspecification in the Beta Component of Returns

In the proposition below, we establish the conditions under which the beta portfolio can be replaced, without any loss of performance, by a benchmark portfolio independent of $\boldsymbol{\lambda}$ and $\boldsymbol{\Omega}$, and hence, immune to misspecification. This demonstrates that, under suitable assumptions, one can construct the correctly specified beta portfolio that is *independent* of risk premia.

Proposition 4.3 (Weight and Sharpe ratio of benchmark portfolio for large N)**.** *Suppose that the vector of asset returns, $\mathbf{r}_t$, satisfies Assumptions 3.1 and 3.2 and $\boldsymbol{\alpha}_N \neq \mathbf{0}_N$. Suppose further that there exists a benchmark portfolio $\mathbf{w}_N^{bench}$, with associated Sharpe ratio SR^{bench}, satisfying the following properties:*

$$(\mathbf{w}_N^{bench})'\boldsymbol{\alpha}_N \to 0, \qquad \mathbf{B}_N'\mathbf{w}_N^{bench} \to \mathbf{c}^{bench}, \qquad (\mathbf{w}_N^{bench})'\boldsymbol{\Sigma}_N\mathbf{w}_N^{bench} \to 0, \tag{24}$$

where $\mathbf{c}^{bench}$ is a $K \times 1$ vector of constants satisfying $\boldsymbol{\lambda}'\mathbf{c}^{bench} \neq 0$. Then,

(i) If $K = 1$,

$$(SR^{mv})^2 - \Big((SR^\alpha)^2 + (SR^{bench})^2\Big) \to 0.$$

(ii) If $K > 1$ and $\mathbf{c}^{bench}$ is perfectly proportional to the vector $\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda}$,

$$(SR^{mv})^2 - \Big((SR^\alpha)^2 + (SR^{bench})^2\Big) \to 0.$$

(iii) If $K > 1$ and $\mathbf{c}^{bench}$ is not proportional to the vector $\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda}$, then for some positive δ,

$$(SR^{mv})^2 - \Big((SR^\alpha)^2 + (SR^{bench})^2\Big) \to \delta > 0.$$

The first assumption in (24) requires that the benchmark portfolio be 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 the benchmark portfolio to be well diversified. Note that for the case of large pricing errors, the first assumption is satisfied whenever $(\mathbf{w}_N^{\text{bench}})'\mathbf{a}_N \to 0$ and $\mathbf{A}_N'\mathbf{w}_N^{\text{bench}} \to \mathbf{0}_p$, where the latter condition ensures that $\mathbf{w}_N^{\text{bench}}$ diversifies away the contribution of the latent factors $\mathbf{A}_N$ in $\boldsymbol{\Sigma}_N$.

The assumptions in (24) imply that the return on the benchmark portfolio is asymptotically equivalent to the return on the portfolio of factors with weight $\mathbf{c}^{\text{bench}}$; that is, $(\mathbf{w}_N^{\text{bench}})'(\mathbf{r}_t - r_{ft}\mathbf{1}_N) = (\mathbf{c}^{\text{bench}})'(\mathbf{f}_t - r_{ft}\mathbf{1}_K) + o_p(1)$, implying that $\mathbf{c}^{\text{bench}}$ equals the mean-variance portfolio constructed using just the K factors. This choice guarantees that the benchmark portfolio achieves the largest possible Sharpe ratio, as stated in parts (i) and (ii) of the proposition.

In the proposition above, we established the condition under which a benchmark portfolio, combined with the alpha portfolio, will coincide with the optimal mean-variance portfolio asymptotically. Interestingly, this condition is *always* satisfied when one postulates a single factor (i.e. $K = 1$) for the beta portfolio. When $K > 1$, it requires an additional condition, which can be verified empirically, using the estimated loadings, the sample mean, and the sample covariance matrix of the K factors, and comparing the sample equivalent of $\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda}$ to $\mathbf{B}_N'\mathbf{w}_N^{\text{bench}}$.

To construct a valid benchmark portfolio, one can use the insights of Treynor and Black (1973) and DeMiguel et al. (2009b). Treynor and Black (1973) suggest that $\mathbf{w}_N^{\beta}$ can be approximated by a portfolio that is similar to the market portfolio, $\mathbf{w}_N^{\text{mkt}}$. Alternatively, the findings of DeMiguel et al. (2009b) suggest that one could use an equally weighted portfolio. Therefore, our result provides theoretical underpinnings for these empirical approaches in the literature, demonstrating their optimality (in the sense of Proposition 4.3) when a single-factor beta portfolio is considered. The benchmark could also be based on multiple factors, such as the Fama and French factors.

4.3 Mitigating Misspecification in the Alpha Component of Returns

In this section, we show that for the alpha portfolio, the presence of pricing errors $\boldsymbol{\alpha}_N$ in the APT naturally allows one to mitigate model misspecification in the sense of Hansen and Sargent (2007). In particular, we establish that the alpha portfolio $\mathbf{w}_N^{\alpha}$ identified in equation (19) is equivalent to the portfolio constructed by an investor who mitigates model misspecification in the alpha component of returns using robust-control theory by finding the solution to:

$$\max_{\mathbf{w}_N} \min_{\boldsymbol{\alpha}_N} \left\{ \mathbf{w}_N'(\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N) - \frac{\gamma}{2}\mathbf{w}_N'\mathbf{V}_N\mathbf{w}_N \right\}, \tag{25}$$

subject to the relative-entropy constraint

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\ln \frac{f_{\boldsymbol{\alpha}_N}(\mathbf{r}_t^e)}{f_{\hat{\boldsymbol{\alpha}}_N}(\mathbf{r}_t^e)} \right) f_{\boldsymbol{\alpha}_N}(\mathbf{r}_t^e)\, d\mathbf{r}_t^e \leq \delta_{\text{entropy}}, \tag{26}$$

where the constant δ_{entropy} measures aversion to alpha misspecification, $\mathbf{r}_t^e = \mathbf{r}_t - r_{ft}\mathbf{1}_N$ is the vector of risky-asset returns in excess of the risk-free rate, and, given $\boldsymbol{\lambda}$, $\mathbf{B}_N$, and $\mathbf{V}_N$, $f_{\boldsymbol{\alpha}_N}(\cdot)$ is the probability density function of a random vector distributed as $\mathcal{N}(\boldsymbol{\alpha}_N + \mathbf{B}_N\boldsymbol{\lambda}, \mathbf{V}_N)$, with $\hat{\boldsymbol{\alpha}}_N$ being an estimator of $\boldsymbol{\alpha}_N$. Our distributional assumptions allow one to relate the entropy constraint to a constraint on the Sharpe ratio of the alpha portfolio.[26]

Observe that the objective function in (25) entails both a maximization over the vector of portfolio weights, $\mathbf{w}_N$, standard in portfolio-choice, and a minimization over the pricing errors, $\boldsymbol{\alpha}_N$, which are constrained to lie in a neighborhood of the estimate $\hat{\boldsymbol{\alpha}}_N$. Standard optimal-control theory assumes that the decision-maker knows the true model. In contrast, robust-control theory treats the decision-maker's model as an approximation and seeks a single rule that works over a set of nearby models that might govern the data. The minimization over $\boldsymbol{\alpha}_N$ is a consequence of the investor's preference for robustness; the role of the constraint in (26) is to ensure that the chosen $\boldsymbol{\alpha}_N$ are statistically indistinguishable (for a given significance level) from the estimate $\hat{\boldsymbol{\alpha}}_N$, which represents the so-called *approximating model*.[27] Using a set of perturbed models, which, given the available data, are difficult to distinguish statistically from the approximating model, protects the max-min investor from choosing weights that are too extreme.

We show (in the proof for Proposition 4.4) that the solution to this *robust mean-variance* problem, $\mathbf{w}_N^{\text{rmv}}$, is:

$$\mathbf{w}_N^{\text{rmv}} = \phi^\alpha \mathbf{w}_N^{r\alpha} + \phi^\beta \mathbf{w}_N^\beta + \mathcal{O}(1), \tag{27}$$

where ϕ^α, ϕ^β, and $\mathbf{w}_N^\beta$ are defined in Proposition 4.1, the robust alpha portfolio, $\mathbf{w}_N^{r\alpha}$, is

$$\mathbf{w}_N^{r\alpha} = \frac{1}{\phi}\frac{1}{\gamma^\alpha}\boldsymbol{\Sigma}_N^+ \hat{\boldsymbol{\alpha}}_N, \tag{28}$$

[26] Note that under our distributional assumption for $\mathbf{r}_t^e$, the relative entropy constraint in (26) simplifies to $\frac{1}{2}(\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N)'\boldsymbol{\Sigma}_N^{-1}(\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N) \leq \delta_{\text{entropy}}$, where a larger δ_{entropy} represents an increase in the investor's degree of aversion to model misspecification. Note also that when $\delta_{\text{entropy}} = \frac{1}{2}\frac{(1+\boldsymbol{\lambda}'\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda})}{T}\chi^2_{N,x\%}$, where $\chi^2_{N,x\%}$ corresponds to the x^{th} $(0 \leq x \leq 1)$ quantile of a χ^2_N distribution, then the constraint above defines the set of $\boldsymbol{\alpha}_N$ that are statistically indistinguishable from $\hat{\boldsymbol{\alpha}}_N$ at $(1-x)\%$ for the case of $\sqrt{T}(\hat{\boldsymbol{\alpha}}_N - \boldsymbol{\alpha}_N)$ normally distributed with mean 0 and covariance matrix equal to $(1 + \boldsymbol{\lambda}'\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda})\boldsymbol{\Sigma}_N$.

[27] Following Hansen and Sargent (2007), we postulate that the agent has estimated the approximating model, in particular, through the APT. However, the result of Proposition 4.4 is unchanged if an approximating model, dictated by a perceived pricing error $\boldsymbol{\alpha}_N^*$, is used to solve the max-min optimization problem and only subsequently estimated using the APT. In principle, but at the cost of greater complexity, one could also consider the case where the investor solves the max-min problem taking into account the uncertainty about the number of missing factors.

$\hat{\boldsymbol{\alpha}}_N = \hat{\boldsymbol{\alpha}}_N(\delta_{\text{apt}})$ is the constrained ML (Maximum Likelihood) estimator of $\boldsymbol{\alpha}_N$, which is described in Section 5.2, and $\phi = \phi(\delta_{\text{entropy}}) \geq 1$ is the shrinkage parameter (given in equation (A6)), implying that $\mathbf{w}_N^{r\alpha} = \mathbf{w}_N^{r\alpha}(\delta_{\text{apt}}, \delta_{\text{entropy}})$. Observe from (28) that in the absence of a concern for model misspecification, the expression for $\mathbf{w}_N^{r\alpha}$ would have $\phi = 1$, which is the solution to the classical mean-variance portfolio problem in (17). In contrast, the investor's concern for misspecification implies that the alpha portfolio weights are scaled by $1/\phi$, with $\phi > 1$. Moreover, we show that the magnitude of ϕ increases as we increase the size of the set of models over which the investor is uncertain, that is, as we increase δ_{entropy}.

Importantly, in the proposition below, we demonstrate that, under the appropriate conditions, the mean-variance optimal portfolio estimated under the APT is precisely the same that one would obtain from the max-min optimization (25)–(26). This is because both δ_{apt} in the APT constraint and ϕ (which depends positively on δ_{entropy}) in the max-min optimization enter multiplicatively into the expressions for $\mathbf{w}_N^{\alpha}$ and $\mathbf{w}_N^{r\alpha}$, respectively; hence, a smaller δ_{apt} and a larger δ_{entropy} induce an *equivalent* reduction in the demand for risky assets in that portfolio. Therefore, there exist a δ_{apt} and a δ_{entropy} that lead to two identical vectors of alpha-portfolio weights.[28]

Proposition 4.4 (Equivalence between alpha portfolio weights)**.** *Under the assumptions of Proposition 4.1, as $N \to \infty$, the estimated alpha portfolio $\mathbf{w}_N^{\alpha}(\delta_{apt})$ in equation (19) corresponding to a given δ_{apt}, is asymptotically equivalent to the robust mean-variance alpha portfolio $\mathbf{w}_N^{r\alpha}(\delta^*_{apt}, \delta_{entropy})$ in equation (28), corresponding to some given δ^*_{apt} and $\delta_{entropy}$:*

$$\mathbf{w}_N^{r\alpha}(\delta^*_{apt}, \delta_{entropy}) = \mathbf{w}_N^{\alpha}(\delta_{apt}) + \mathcal{O}(1),$$

$$\textit{whenever} \quad \delta^*_{apt}\left(1 - \frac{(2\delta_{entropy})^{\frac{1}{2}}}{(\hat{\boldsymbol{\alpha}}'_N \boldsymbol{\Sigma}_N^{+} \hat{\boldsymbol{\alpha}}_N)^{\frac{1}{2}}}\right)^2 = \delta_{apt}.$$

Thus, the proposition highlights how the APT is precisely the appropriate framework for accounting for model misspecification in the alpha portfolio.

4.4 Extension to Conditional Asset-Pricing Setting

We now discuss how our methodology can be extended to address misspecification in a *conditional* asset-pricing setting. Considering conditional asset pricing models has the advantage of allowing us to study misspecification associated with time-variation in the model's parameters. This is

[28] In practice, δ_{apt} and δ_{entropy} are unknown. In Internet Appendix IA.3, we explain our cross-validation procedure to select these constants; our empirical results confirms that if one fixes either δ_{apt} or δ_{entropy} and determines the other one optimally by cross-validation, it leads to the same robust mean-variance portfolio.

important because of the increasing focus on individual securities and the evidence of time variation in their betas and risk premia (see, for instance, Gagliardini et al. (2016)).

4.4.1 Generalizing the APT theory to a conditional setting

To formulate the APT in a conditional setting, one needs to replace Assumptions 3.1 and 3.2 with the following.[29]

Assumption 4.1 (Candidate beta-pricing model—conditional)**.** *We assume that the N-dimensional vector $\mathbf{r}_t$ of asset returns is given by*

$$\mathbf{r}_t = \boldsymbol{\mu}_{N,t-1} + \mathbf{B}_{N,t-1}\left(\mathbf{f}_t - E_{t-1}(\mathbf{f}_t)\right) + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\mu}_{N,t-1}$ is the $N \times 1$ vector of conditional *expected returns and $\mathbf{B}_{N,t-1}$ is the $N \times K$ full-rank matrix of conditional factor loadings, with $K < N$. At any time t, the $K \times 1$ vector of common observed factors, $\mathbf{f}_t$, is distributed with conditional mean $E_{t-1}(\mathbf{f}_t)$ and $K \times K$ conditional covariance matrix $\boldsymbol{\Omega}_{t-1}$, and the $N \times 1$ vector of innovations $\boldsymbol{\varepsilon}_t$ is distributed with zero conditional mean and the $N \times N$ conditional covariance matrix $\boldsymbol{\Sigma}_{N,t-1}$, with $\boldsymbol{\Omega}_{t-1}$ and $\boldsymbol{\Sigma}_{N,t-1}$ being positive definite almost surely. Moreover, $\boldsymbol{\varepsilon}_t$ and $\mathbf{f}_t$ are conditionally uncorrelated.*

Assumption 4.2 (No asymptotic arbitrage—conditional)**.** *As $N \to \infty$, there is no sequence of portfolios for which, along some subsequence $\tilde{N}$,*

$$\mathrm{var}_{t-1}(\mathbf{r}_t'\mathbf{w}^a_{\tilde{N},t-1}) \to 0 \quad \textit{and} \quad (\boldsymbol{\mu}_{\tilde{N}_{t-1}} - r_{ft}\mathbf{1}_{\tilde{N}})'\mathbf{w}^a_{\tilde{N},t-1} \geq \delta > 0 \quad \textit{a.s.} \quad \textit{for all} \quad \tilde{N}.$$

Under Assumptions 4.1 and 4.2, together with a limited degree of cross-sectional dependence and a sufficient degree of smoothness of the loadings (see Stambaugh (1983, thms. 1 and 2) and Zaffaroni (2020, prop. 4)), a conditional version of the APT holds. That is, there exists a unique vector of risk premia $\boldsymbol{\lambda}_{t-1}$ such that the pricing errors $\boldsymbol{\alpha}_{N,t-1} = (\boldsymbol{\mu}_{N,t-1} - r_{ft}\mathbf{1}_N) - \mathbf{B}_{N,t-1}\boldsymbol{\lambda}_{t-1}$ satisfy the conditional-APT restriction:[30]

$$\sup_N \boldsymbol{\alpha}'_{N,t-1}\boldsymbol{\Sigma}^{-1}_{N,t-1}\boldsymbol{\alpha}_{N,t-1} \leq \delta_{t-1} < \infty \;\; a.s. \text{ for some finite non-random } \delta_{t-1}. \tag{29}$$

Importantly, under the assumptions of Proposition 4.1, extended to the case of time-varying parameters, the pricing errors satisfy the decomposition into the conditional latent common component, with loadings $\mathbf{A}_{N,t-1}$ and risk premia $\boldsymbol{\lambda}_{\text{miss}t-1}$, and the conditional purely asset-specific

[29] $E_{t-1}(.)$ and $\mathrm{var}_{t-1}(.)$ denote the operators conditional with respect to the information available at time $t-1$.

[30] Establishing the conditional APT requires one to ensure invariance to asset repackaging, which is relevant when agents construct portfolios based on different information sets (see Hansen and Richard (1987, sec. 5)). For proofs of the conditional APT in various settings, see Stambaugh (1983), Reisman (1992), Al-Najjar (1999), Gagliardini et al. (2016), and Zaffaroni (2020).

component with conditional mean $\mathbf{a}_{N,t-1}$:

$$\boldsymbol{\alpha}_{N,t-1} = \mathbf{A}_{N,t-1}\boldsymbol{\lambda}_{\text{miss}t-1} + \mathbf{a}_{N,t-1} \quad \text{such that} \quad \sup_N \mathbf{a}'_{N,t-1}\boldsymbol{\Sigma}^{-1}_{N,t-1}\mathbf{a}_{N,t-1} \leq \delta_{t-1} < \infty.$$

4.4.2 Modeling time variation and estimation of the conditional model

When estimating the model with the conditional APT, we need to specify the form of time-variation of the model's parameters and then the appropriate estimation procedure. Two approaches for modeling time variation in the existing literature on empirical asset pricing are: (i) a model-free "nonparametric" approach, under which the form of time-variation is unspecified (Lewellen and Nagel, 2006; Ang and Kristensen, 2012; Zaffaroni, 2020); and (ii) a "parametric" approach, where one specifies a particular functional form, typically linear, along with a set of observed state variables that drive the time-variation, assuming that the state variables are either *common* across assets (for instance, as in Ferson and Harvey, 1991) or *asset specific* (for instance, as in Gagliardini et al., 2016). Below, we explain how to extend our model to handle both the nonparametric and parametric approaches.

Nonparametric approach

A simple and successful estimation approach for the model-free formulation of the conditional APT consists of using short rolling windows of data to capture the time variation nonparametrically, as popularized in empirical finance with the two-pass methodology of Fama and MacBeth (1973).[31] The nonparametric approach has the advantage of avoiding misspecification arising from a potentially incorrect specification of the mappings between the time-varying parameters and the state variables driving their dynamics. The rolling-window estimation we undertake in the empirical application in Section 5 is in line with this approach to conditioning. However, one of the drawbacks of the nonparametric approach is that the estimates obtained exhibit larger standard errors than those from the parametric approach, which we describe next.

Parametric approach

We consider the parametric approach for two settings: first, in which the state variables are common across assets, and second, in which the state variables are asset specific.

For the parametric approach where time variation is driven by *common* state variables, such as risk factors or macroeconomic variables, the methodology and estimation procedure developed

[31] A rolling window estimate can be formally interpreted as a nonparametric kernel estimator, corresponding to a particular bandwidth, determining the size of the window, and a rectangular kernel function.

above are still valid. In this setting, omitting relevant state variables is observationally equivalent to omitting common risk factors, which, as we explain below, is fully accounted for by our methodology.

For simplicity, let us consider the case where the loadings on the observed candidate factors are time varying, i.e. $\mathbf{B}_{N,t-1}$, keeping all other parameters constant. However, one can extend our approach to allow also for time-varying risk premia, pricing errors, and covariance matrices. For illustrative purposes, assume that there exists a single $K_g = 1$ (lagged) state variable g_{t-1} driving the time variation in $\mathbf{B}_{N,t-1}$ such that:

$$\mathbf{B}_{N,t-1} = \mathbf{B}_N + g_{t-1}\mathbf{B}_{N,h}, \tag{30}$$

for some constant $N \times K$ matrices $\mathbf{B}_N$ and $\mathbf{B}_{N,h}$, and where, without loss of generality, one can assume that g_t has zero mean.[32] Assuming, for simplicity, that there are no latent factors, i.e. $\mathbf{A}_{N,t-1} = \mathbf{0}_{N\times K^{mis}}$, then excess returns satisfy

$$\mathbf{r}_{N,t} - r_{ft}\mathbf{1}_{\tilde{N}} = \boldsymbol{\alpha}_N + \mathbf{B}_{N,t-1}\,\mathbf{f}_t + \boldsymbol{\varepsilon}_t = \boldsymbol{\alpha}_N + \mathbf{B}_N\,\mathbf{f}_t + \mathbf{B}_{N,h}\mathbf{h}_t + \boldsymbol{\varepsilon}_t,$$

where $\mathbf{h}_t = g_{t-1}\mathbf{f}_t$. That is, one obtains a model with $2K$ risk factors, where $\mathbf{h}_t$ can be interpreted as the *scaled* factors in Gagliardini, Ossola, and Scaillet (2016, 2019a).

Expected excess returns can now be represented as $\boldsymbol{\mu}_N = \boldsymbol{\alpha}_N + \mathbf{B}_N\,\boldsymbol{\lambda} + \mathbf{B}_{N,h}\boldsymbol{\lambda}_h$, implying an asset-pricing model with an asset-specific component and the $K + K$ risk factors $\mathbf{f}_t$ and $\mathbf{h}_t$ with premia $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}_h$, respectively. If g_t is observed, one obtains an APT model with only observed risk factors, together with the asset-specific component $\mathbf{a}_N$. On the other hand, if g_t is unknown, then one obtains an APT model with K observed risk factors $\mathbf{f}_t$ and K latent risk factors $\mathbf{h}_t$.[33] Therefore, by setting $\mathbf{A}_N = \mathbf{B}_{N,h}$ and $\boldsymbol{\lambda}_{miss} = \boldsymbol{\lambda}_h$, one is back to the APT model we considered earlier with constant parameters that are a combination of latent and observed risk factors. The estimation strategy, and the associated empirical portfolio construction, follow the same steps as for the unconditional setting.[34]

For the second parametric setting, where time variation is driven by *asset-specific* state variables such as firm characteristics, our methodology for the unconditional setting needs to be extended, as we explain below.

[32] In fact, when $Eg_{t-1} = \mu_g \neq 0$, one obtains $\mathbf{B}_N + g_{t-1}\mathbf{B}_{N,h} = \mathbf{B}_N + (g_{t-1} - \mu_g + \mu_g)\mathbf{B}_{N,h} = \tilde{\mathbf{B}}_N + \mathbf{B}_{N,h}(g_{t-1} - \mu_g)$, setting $\tilde{\mathbf{B}}_N = \mathbf{B}_N + \mu_g\mathbf{B}_{N,h}$.

[33] Note that h_t and $\mathbf{f}_t$ are mutually orthogonal whenever g_{t-1} and $\mathbf{f}_t$ are assumed independent with $Eg_{t-1} = 0$, despite $\mathbf{h}_t$ being a function of $\mathbf{f}$. In fact, straightforward calculations show that $\mathrm{cov}(\mathbf{h}_t, \mathbf{f}_t') = E(g_{t-1})\,\mathrm{var}(\mathbf{f}_t) = \mathbf{0}_{K\times K}$.

[34] When $K_g > 1$, equation (30) is replaced by $\mathbf{B}_{N,t-1} = \mathbf{B}_N + (\mathbf{I}_N \otimes \mathbf{g}_{t-1}')\mathbf{B}_{N,h}$ where now $\mathbf{B}_{N,h}$ is an $NK_g \times K$ constant matrix. Solving the model, one obtains the APT with constant parameters corresponding to $K + KK_g$ risk factors, with KK_g of them being latent for unspecified $\mathbf{g}_{t-1}$. One can then readily apply the methodology outlined in the previous sections for empirical portfolio construction.

Studying again the simple case of a $K_g = 1$ asset-specific state variable $g_{i,t}$ so that there is one for each of the assets, the time-varying loadings of the observed risk factors $\mathbf{f}_t$ can be expressed as $\mathbf{B}_{N,t-1} = \mathbf{B}_N + \text{diag}(\mathbf{g}_{t-1})\mathbf{B}_{N,h}$, where $\text{diag}(\mathbf{g}_{t-1})$ is the diagonal matrix consisting of the elements of the vector $\mathbf{g}_{t-1} = (g_{1,t-1}, \cdots, g_{N,t-1})'$.

To ensure identification, some restrictions must be imposed because, for instance, the $g_{i,t-1}$ and $\mathbf{f}_t$ cannot *both* be treated as latent if the model is to be identified (for a discussion of this, see Gagliardini et al. (2019a)). Following the empirical asset-pricing literature, we focus on the case of *observed* asset-specific state variables, in which case the model can be written as a function of K observed risk factors $\mathbf{f}_t$ and NK observed asset-specific variables $\mathbf{h}_t = (\mathbf{g}_{t-1} \otimes \mathbf{f}_t)$. Therefore, in contrast to the previous case, asset-specific state variables do not induce latent risk factors.[35] However, latent risk factors are still permitted, for example because of an incomplete specification of the priced risk factors $\mathbf{f}_t$, now coupled with their time-varying loadings driven by the observed $g_{i,t-1}$, as, for instance, $\mathbf{A}_{N,t-1} = \mathbf{A}_N + \text{diag}(\mathbf{g}_{t-1})\mathbf{A}_{N,h}$, for constant $N \times p$ matrices $\mathbf{A}_N$ and $\mathbf{A}_{N,h}$. Thus, the case of asset-specific state variables does not pose any additional problems for our methodology, except for a slight modification of the maximum-likelihood estimator.[36]

4.4.3 Portfolio choice in a conditional setting

Under Assumptions 4.1 and 4.2 given above, the conditional versions of Propositions 4.1–4.4 hold. For instance, the conditional mean-variance portfolio weights satisfy

$$\mathbf{w}^{\text{mv}}_{N,t-1} = \phi^{\alpha}_{t-1}\, \mathbf{w}^{\alpha}_{N,t-1} + \phi^{\beta}_{t-1}\, \mathbf{w}^{\beta}_{N,t-1},$$

$$\text{in which} \qquad \mathbf{w}^{\alpha}_{N,t-1} = \frac{1}{\gamma^{\alpha}_{t-1}} \mathbf{\Sigma}^{+}_{N,t-1} \boldsymbol{\alpha}_{N,t-1} + \mathcal{O}(1) \quad \text{and} \quad \mathbf{w}^{\beta}_{N} = \frac{1}{\gamma^{\beta}_{t-1}} \mathbf{V}^{-1}_{N,t-1} \mathbf{B}_{N,t-1} \boldsymbol{\lambda}_{t-1},$$

where $\phi^{\alpha}_{t-1} = \frac{\gamma^{\alpha}_{t-1}}{\gamma_{t-1}}$, $\phi^{\beta}_{t-1} = \frac{\gamma^{\beta}_{t-1}}{\gamma_{t-1}} = 1 - \phi^{\alpha}_{t-1}$ for some positive constants, $\gamma_{t-1}, \gamma^{\alpha}_{t-1}, \gamma^{\beta}_{t-1}$, and where $\mathbf{V}_{N,t-1}$ and $\mathbf{\Sigma}^{+}_{N,t-1}$ follow from (11) and (20) after replacing $\mathbf{B}_N$, $\mathbf{\Omega}$, and $\mathbf{\Sigma}_N$ with the corresponding quantities $\mathbf{B}_{N,t-1}$, $\mathbf{\Omega}_{t-1}$, and $\mathbf{\Sigma}_{N,t-1}$.

[35] In particular, when the $N^{-1}\sum_{i=1}^{N} g_{i,t-1} \to_p 0$, then $h_{i,t} = g_{i,t-1} f_t$ cannot be disentangled from the asset-specific error term $\epsilon_{i,t}$, hence the identification problem highlighted in Gagliardini et al. (2019a).

[36] Gagliardini et al. (2016) provide the econometric analysis for the case of time-varying parameters driven by observable asset-specific state variables when the risk factors are assumed to be observed in the context of a continuum of assets and imposing exact pricing. Connor and Linton (2007); Connor, Hagmann, and Linton (2012), Fan, Liao, and Wang (2016), Kelly et al. (2020), Pelger and Xiong (2019), and Kim et al. (2021) provide methodologies that allow for the estimation of the latent risk factors, retaining observability of the asset-specific state variables that drive the dynamics of the parameters, although some further conditions are necessary. For example, Kelly et al. (2020) restrict $\mathbf{A}_{N,h} = (\mathbf{1}_N \otimes \mathbf{A}_h)$ for a $K_g \times K$ constant matrix $\mathbf{A}_h$ whereas Connor and Linton (2007), Connor et al. (2012), Fan et al. (2016), and Kim et al. (2021) require that the asset-specific state variables be constant over time, i.e. $g_{i,t-1} = g_i$, although time variation can still be allowed through the coefficients on g_i when estimated with short rolling time windows of data.

4.5 Comparison with Recent Papers on Portfolio Choice

We conclude this section by explaining how our work is related to recent work on portfolio choice.

4.5.1 Comparison with Koijen and Yogo (2019)

Koijen and Yogo (2019) highlight the importance of studying the asset allocation of institutional and household investors using observed portfolio holdings as opposed to asset-returns data. Although the aim of their *positive* analysis, understanding the role of institutions in asset-market movements, volatility, and predictability, differs from our *normative* objective, namely, how to construct mean-variance efficient portfolios robust to model misspecification, strong analogies exist between the two approaches.

To better understand the similarities and differences between our normative approach and the positive approach of Koijen and Yogo (2019), consider the simple candidate beta-pricing model in equation (2) of Section 2 for a generic institutional investor leading to the beta portfolio w_i^{β} given in (4). Our methodology then allows us to construct the orthogonal alpha portfolio,

$$w_i^{\alpha} = w_i^{\text{mv}} - w_i^{\beta}. \tag{31}$$

Suppose now that for a particular investor one *observes* the portfolio holdings w_i^{obs} for each asset i, as in Koijen and Yogo (2019). Given the beta portfolio weights in (4), one then obtains

$$w_i^{\alpha\text{-AKY}} = w_i^{\text{obs}} - \theta w_i^{\beta}, \tag{32}$$

with $w_i^{\alpha\text{-AKY}}$ being the alpha asset demand orthogonal to the beta portfolio, with the projection coefficient $\theta = (\mathbf{w}^{\beta\prime}\mathbf{w}^{obs})/(\mathbf{w}^{obs\prime}\mathbf{w}^{obs})$. The superscript "AKY" on $w_i^{\alpha\text{-AKY}}$ indicates that this corresponds to the Koijen and Yogo (2019) portfolio under an *additive* portfolio decomposition, instead of the multiplicative decomposition they adopt, which is convenient for their empirical analysis.

Following the ideas in Koijen and Yogo (2019), $w_i^{\alpha\text{-AKY}}$ in (32) can be interpreted as the investor's latent asset-demand, representing the wedge between *observed* holdings w_i^{obs} and the mean-variance portfolio weights associated with a candidate mean-variance beta-model w_i^{β}. On the other hand, w_i^{α} in (31) is the *normative* latent asset demand in the sense that it represents the wedge between the mean-variance *efficient* portfolio that an investor should hold and the beta portfolio weights associated with a candidate beta-model. It is in this sense that the role played by w_i^{α} in our model is similar to that of $w_i^{\alpha\text{-AKY}}$ in Koijen and Yogo (2019).

Of course, one can also use our result in (32) in a *positive* way. Essentially, one can project the beta portfolio based on an investor's candidate factor model on the observed holdings of the investor to obtain the alpha portfolio corresponding to the observed portfolio weights. Then, one can study the properties of this alpha portfolio, for instance, to see which part is related to latent factors omitted in the beta portfolio and which to purely asset-specific returns.

4.5.2 Comparison with Daniel, Mota, Rottke, and Santos (2020)

Some recent papers on portfolio choice have been motivated by the observation that factor models (i.e. what we refer to as "candidate beta models," such as the one based on the Fama and French risk factors), are not mean-variance efficient.

Daniel et al. (2020) argue that beta models with characteristic-based risk factors, such as the Fama and French factors, do not span the mean-variance frontier because they contain exposure to unpriced factors, i.e. factors with zero risk premia. In other words, misspecification arises from the beta model containing *too many factors*. Their goal is to mitigate this problem when constructing characteristics-sorted portfolios. To address this problem, one needs to quantify these unpriced risk factors through the so-called hedging portfolios and net them out from the initial factor beta model. Specifically, an orthogonal representation is assumed whereby

$$\mathbf{w}_N^{\beta,cp} = \mathbf{w}_N^{hedge}\boldsymbol{\Delta} + \mathbf{w}_N^{mv}, \tag{33}$$

in which $\mathbf{w}_N^{\beta,cp}$ refers to the portfolio weights of the characteristic-based beta model, with the mean-variance portfolio weights $\mathbf{w}_N^{mv}$ and the hedging portfolio weights $\mathbf{w}_N^{hedge}$ being orthogonal to one another, and where $\boldsymbol{\Delta}$ is the corresponding matrix of hedge ratios, i.e. the regression coefficients of the $\mathbf{w}_N^{\beta,cp}$ returns onto $\mathbf{w}_N^{hedge}$ returns. Notice that the hedge portfolio weights $\mathbf{w}_N^{hedge}$ are constructed as being maximally correlated with $\mathbf{w}_N^{\beta,cp}$ but with a null contribution to their mean return (Daniel et al., 2020, def. 2.3 and prop. 2.4). The mean-variance portfolio weights $\mathbf{w}_N^{mv}$, which are spanned by the so-called efficient characteristics portfolios (Daniel et al., 2020, prop. 2.2), represent the (orthogonal) *residuals* in the projection given in equation (33).

Just like Daniel et al. (2020), our work is also based on an orthogonal decomposition involving the mean-variance portfolio weights. However, in contrast to Daniel et al. (2020), our beta portfolio contains *too few risk factors*, in addition to missing the asset-specific component. By augmenting the beta portfolio with the (orthogonal) alpha portfolio, we achieve two-fund separation and mean-variance efficiency, i.e. $\mathbf{w}_N^{mv} = \phi^\alpha \mathbf{w}_N^\alpha + \phi^\beta \mathbf{w}_N^\beta$. Instead, in Daniel et al. (2020) the beta portfolio contains too many factors; therefore, to obtain the mean-variance portfolios, one needs to net out the unpriced sources of risks (by algebraic difference).

A second difference between our approach and that in Daniel et al. (2020) is that our methodology takes advantage of the benefits of portfolio optimization, whereas Daniel et al. (2020) propose an empirical approach based on sorting arguments. Dominant factors with a small or even zero risk premium can arise in practice, but *mean-variance optimization automatically accounts for this.* Note that the analysis in Daniel et al. (2020) crucially assumes that the risk exposures to the characteristics-based factors are proportional to a set of observed characteristics and that no priced factors have been erroneously omitted from the beta model.

A final noteworthy difference is that, in contrast to our modeling framework, Daniel et al. (2020) do not allow for asset-specific sources of risk.[37]

4.5.3 Comparison with Kim, Korajczyk, and Neuhierl (2021)

In a conditional asset-pricing setting, Kim et al. (2021) argue that *arbitrage portfolios*, namely portfolios spanned by a set of asset-specific characteristics orthogonal to the risk factors driving the beta part of the asset-pricing model, lead to superior portfolio performance, suggesting the inadequacy of factor-only models. Strong analogies arise between Kim et al. (2021) and our methodology, although with some important differences, which we describe below.

Simplifying arguments and abstracting from estimation issues, Kim et al. (2021) assume that expected asset excess returns satisfy, just like our equation (12),

$$(\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N) = \boldsymbol{\alpha}_N + \mathbf{B}_N\boldsymbol{\lambda},$$

where $\boldsymbol{\alpha}_N = \boldsymbol{\alpha}_N(\mathbf{X}) = \mathbf{X}\boldsymbol{\theta}$ for a matrix of observed asset-specific characteristic $\mathbf{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_N)'$ and a constant vector of parameters $\boldsymbol{\theta}$.[38] The crucial assumption in Kim et al. (2021) is that, for some positive constant δ_{KKN},

$$\frac{\boldsymbol{\alpha}_N'\boldsymbol{\alpha}_N}{N} \rightarrow_p \delta_{\text{KKN}} > 0. \tag{34}$$

This means that their "arbitrage portfolios" $\mathbf{w}_N^{\text{KKN}} = \boldsymbol{\alpha}_N/N$ lead to (under their assumptions, including orthogonality between $\boldsymbol{\alpha}_N$ and $\mathbf{B}_N$) a portfolio return satisfying $\mathbf{w}^{\text{KKN}\prime}(\mathbf{r}_t - r_{ft}\mathbf{1}_N) =$

[37] In this respect, our work is more akin to Daniel and Titman (1997), whose main goal is to show empirically the importance of firm-specific characteristics to explain the cross-section of expected returns. However, because Daniel and Titman (1997) is not founded on the APT, it allows for the possibility of arbitrage.

[38] The loadings on the risk factors are also assumed to be function of the $\mathbf{X}$, i.e. $\mathbf{B}_N = \mathbf{B}_N(\mathbf{X})$ and the pricing errors are assumed to satisfy $\boldsymbol{\alpha}_N(\mathbf{X}) = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\Gamma}_\alpha$ for a vector of constants $\boldsymbol{\Gamma}_\alpha$. Our simplifications do not compromise the comparison of their methodology with ours.

$\boldsymbol{\alpha}_N'\boldsymbol{\alpha}_N/N + \boldsymbol{\alpha}_N'\boldsymbol{\varepsilon}_t/N$ such that

$$E(\mathbf{w}^{\text{KKN}\prime}(\mathbf{r}_t - r_{ft}\mathbf{1}_N)) \to \delta_{\text{KKN}}, \qquad \text{var}(\mathbf{w}^{\text{KKN}\prime}(\mathbf{r}_t - r_{ft}\mathbf{1}_N)) \to 0, \qquad \text{SR}^{\text{KKN}} \to \infty. \tag{35}$$

That is, the arbitrage portfolio return earns a positive mean and a zero variance, leading its Sharpe ratio to diverge to infinity, implying a violation of no-arbitrage. Therefore, while the methodology in Kim et al. (2021) satisfies the first APT assumption, i.e. a factor structure, it violates the second APT assumption, i.e. the absence of arbitrage.

The arbitrage portfolio of Kim et al. (2021) is closely related to our alpha portfolio, in particular, when we consider the conditional version of our methodology in Section 4.4, because both portfolios are linear in the vector of pricing errors, $\boldsymbol{\alpha}_N$, and both are factor neutral. However, there is a crucial difference stemming from the APT no-arbitrage assumption. In our case, because of the APT restriction (29), and in stark contrast to (34),

$$\frac{\boldsymbol{\alpha}_N'\boldsymbol{\alpha}_N}{N} \to 0.$$

Indeed, assuming the simplified setting considered in Section 2, the alpha portfolio weights in our APT setting equals $\mathbf{w}_N^{\alpha} = \boldsymbol{\alpha}_N/\sigma_\epsilon^2$, with the normalization by $1/N$ being redundant. Given $\mathbf{w}_N^{\alpha\prime}(\mathbf{r}_t - r_{ft}\mathbf{1}_N) = \boldsymbol{\alpha}_N'\boldsymbol{\alpha}_N/\sigma_\epsilon^2 + \boldsymbol{\alpha}_N'\boldsymbol{\varepsilon}_t/\sigma_\epsilon^2$, by Proposition 4.1,

$$E(\mathbf{w}^{\alpha\prime}(\mathbf{r}_t - r_{ft}\mathbf{1}_N)) \to \frac{\sum_{i=1}^{\infty}\alpha_i^2}{\sigma_\epsilon^2},$$

$$\text{var}(\mathbf{w}^{\alpha\prime}(\mathbf{r}_t - r_{ft}\mathbf{1}_N)) \to \frac{\sum_{i=1}^{\infty}\alpha_i^2}{\sigma_\epsilon^2},$$

$$\text{SR}^{\alpha} \to \left(\frac{\sum_{i=1}^{\infty}\alpha_i^2}{\sigma_\epsilon^2}\right)^{\frac{1}{2}} < \delta_{apt}^{\frac{1}{2}} < \infty,$$

i.e. a bounded Sharpe ratio, in stark contrast to (35), where the Sharpe ratio goes to infinity.

Therefore, just like Kim et al. (2021), our alpha portfolio is factor neutral, exposed only to asset-specific risk, with weights linear in $\boldsymbol{\alpha}_N$. However, in contrast to Kim et al. (2021), our alpha portfolio has a positive but *bounded* Sharpe ratio because arbitrage opportunities are ruled out. Moreover, while we establish a two-fund separation result that shows how to recover mean-variance efficient portfolios by combining the inefficient alpha and beta portfolios, Kim et al. (2021) focus only on the alpha portfolio, without investigating the beta portfolio and two-fund separation.

Summarizing, recent advances in portfolio choice and empirical asset pricing point out the deficiency of factor asset-pricing models both in terms of fitting the cross-section of returns and achieving mean-variance efficiency, regardless of whether the factors are assumed to be latent or observed.[39] The asset-specific component of returns resolves this conundrum, as has been known since the empirical evidence provided in Daniel and Titman (1997). However, as described in MacKinlay (1995) and Daniel and Titman (1997), the asset-specific component opens the possibility of violation of no-arbitrage, as one can see in Kim et al. (2021). Our paper resolves this dilemma by showing how the APT provides the foundations that allow for the presence of an asset-specific component along with omitted pervasive risk factors within a no-arbitrage framework.

5 Evaluating Out-of-Sample Performance

In this section, we illustrate the improvement in the out-of-sample portfolio performance that results from our theoretical insights.

5.1 Data and Experiment Design

The design of our analysis follows the approach adopted in Ao et al. (2019). Just as in their paper, we study portfolios based on monthly returns for the two empirical data sets they study. The first data set consists of monthly returns on a small number $N = 30$ stock constituents of the Dow Jones Industrial Average (DJ30). The second data set consists of monthly returns for a larger number $N = 100$ randomly selected stocks from the S&P 500, in each case, as in Ao et al. (2019), augmented with the Fama-French $K = 3$ factors.[40] Thus, when reporting our empirical results, all tables and figures have two panels, with Panel A for DJ30 constituents and Panel B for S&P 500 constituents.[41] Both data sets span the period 1977 to 2016. In all cases, we estimate the model using a rolling window of $T = 120$ months.[42]

[39] This inadequacy of factor models to tackle misspecification can be traced back to MacKinlay (1995). Various efforts have been made to tackle this problem, such as the suggestion in Kozak, Nagel, and Santosh (2018) to consider higher-order principal components and, in the same spirit, the suggestion of Lettau and Pelger (2020) to account for weak factors. However, the difficulty is that if latent, then the weak factors cannot be estimated consistently (Lettau and Pelger, 2020) and if observed, the weak factors give rise to biased risk-premia estimates. Our methodology allows for weak factors in that our alpha portfolio allows weak factors to earn compensation while bypassing the need to identify and estimate them and their corresponding risk premia.

[40] Our decision to work with individual stock returns as opposed to portfolio returns is motivated by the results in Ang, Liu, and Schwarz (2010) and because the APT theory is designed to handle a large number of assets.

[41] Note that the stocks comprising the DJ30 and S&P 500 change over the sample period.

[42] In principle, one could consider data that includes information about stock characteristics because the APT allows for this in a conditional setting. However, to make our results comparable to those of Ao et al. (2019), we do not use stock-specific characteristics, though we benchmark our model against the Fama-French three-factor model.

5.2 Estimating the APT

To estimate the APT, the key quantity to be estimated is $\boldsymbol{\alpha}_N$. We will describe the (pseudo) Gaussian maximum-likelihood constrained (MLC) estimator, although other estimation procedures could be used. For $\boldsymbol{\theta} = (\mathbf{a}_N', \boldsymbol{\lambda}_{\text{miss}}', \text{vec}(\mathbf{A}_N)', \text{vech}(\mathbf{C}_N)', \boldsymbol{\lambda}', \text{vec}(\mathbf{B}_N)', \text{vech}(\boldsymbol{\Omega})')'$, we consider

$$\hat{\boldsymbol{\theta}}_{\text{MLC}} = \underset{\tilde{\boldsymbol{\theta}}}{\text{argmax}}\, L(\tilde{\boldsymbol{\theta}}) \quad \text{subject to the APT constraint} \quad \tilde{\mathbf{a}}_N' \tilde{\boldsymbol{\Sigma}}_N^{-1} \tilde{\mathbf{a}}_N \leq \delta_{\text{apt}}, \tag{36}$$

where $L(\tilde{\boldsymbol{\theta}})$ defines the log-likelihood function, $\tilde{\boldsymbol{\theta}}$ is the set of feasible parameters, and the $\hat{\cdot}$ denotes an estimated quantity. Propositions IA.1 and IA.2 in the appendix give the formulas for $\hat{\boldsymbol{\theta}}_{\text{MLC}}$ for the cases of small pricing errors ($p = 0$ implying that $\mathbf{A}_N = \mathbf{0}$) and large pricing errors ($p > 0$ implying that $\mathbf{A}_N \neq \mathbf{0}$), respectively. These results also show that, thanks to the APT constraint in (13), the estimator $\hat{\mathbf{a}}_{N,\text{MLC}}$ turns out to be precisely the ridge estimator for $\mathbf{a}_N$. Moreover, the MLC estimator for the latent factors risk premia $\boldsymbol{\lambda}_{\text{miss}}$ coincides with the GLS two-pass estimator, which one would have obtained if the latent factors were nontradable but observed.

A multistep procedure is used to estimate the model. In the first step, one estimates the parameters of the factor model conditional on the factor realizations without imposing the APT restriction. The second step is to analyze the possibility of pervasive missing factors. In particular, this part uses conventional principal-component analysis of $\hat{\boldsymbol{\Sigma}}_N$ to estimate the number of latent pervasive factors, p; see, for example, Anderson (1984) and, more recently, Gagliardini et al. (2019a). In the third step, one estimates the model when either $\hat{p} = 0$ (i.e. small pricing-error case) or $\hat{p} > 0$ (i.e. large pricing-error case) while imposing the APT restriction. For the case where $\hat{p} = 0$, we have $\hat{\boldsymbol{\alpha}}_N = \hat{\mathbf{a}}_N$ and $\hat{\boldsymbol{\Sigma}}_N = \hat{\mathbf{C}}_N$, whereas for the case where $\hat{p} > 0$, we have $\hat{\boldsymbol{\alpha}}_N = \hat{\mathbf{A}}_N \hat{\boldsymbol{\lambda}}_{\text{miss}} + \hat{\mathbf{a}}_N$ and $\hat{\boldsymbol{\Sigma}}_N = \hat{\mathbf{A}}_N \hat{\mathbf{A}}_N' + \hat{\mathbf{C}}_N$. Further details of the estimation are provided in Internet Appendix IA.3.

Observe that to overcome the curse of dimensionality, we must impose sparsity assumptions for estimating $\mathbf{C}_N$. Motivated by the findings in MacKinlay and Pástor (2000), we assume that $\mathbf{C}_N = \sigma^2 \mathbf{I}_N$. This sparsity assumption on $\mathbf{C}_N$ is less stringent than what one may think because the APT allows for a factor structure that captures most of the cross-sectional return dependence through both observed and unobserved factors. However, one could make other, more general, sparsity assumptions, as explained in Gagliardini, Ossola, and Scaillet (2016, 2019b).

5.3 Portfolios From the Existing Literature

We compare the performance of our strategies to five strategies from the literature.[43] The first strategy is the mean-variance (MV) efficient portfolio, which, is given by (18) and where we plug in the sample estimates of the mean and covariance matrix of returns. To match Ao et al. (2019),[44] we require the portfolio to achieve a target volatility of σ^* and hence set $\gamma = \frac{\sqrt{(\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N)'\mathbf{V}_N^{-1}(\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N)}}{\sigma^*}$.

There is an extensive literature (see, for example, DeMiguel et al., 2009b) that shows the out-of-sample performance of the MV portfolio with plug-in sample moments is poor, mainly because the sample mean stock returns are estimated with substantial error (Kan and Zhou, 2007). Motivated by this observation, an alternative to the MV portfolio proposed is the global minimum-variance (GMV) portfolio, which has the advantage that it requires estimation of only the variance-covariance matrix and not mean returns. Furthermore, we can improve the GMV strategy's performance by estimating the covariance matrix using the Ledoit and Wolf (2003) shrinkage method. Thus, the GMV-LW portfolio serves as our second benchmark.

We consider the equally weighted (EW) portfolio as a third reference strategy. DeMiguel et al. (2009b) show that the MV and GMV portfolios often do not outperform the EW portfolio, which does not require the estimation of either mean returns or the variance-covariance matrix of returns. Moreover, we have shown in Proposition 4.3 that, under certain conditions, the EW portfolio mimics the beta portfolio.

Our fourth reference strategy is based on Principal Component Analysis (PCA) of returns. In particular, using the $N \times T$ matrix of returns we extract from one up to ten principal components for each rolling-window, and treat the principal components as observed factors in the factor-model specification.

Finally, the fifth benchmark we consider is the MAXSER strategy in Ao et al. (2019).[45]

[43] See footnote 3.

[44] The results we obtain are similar if one chooses to match a particular level of mean return in excess of the risk-free rate, $\mu^* - r_{ft}$, in which case one needs to set $\gamma = \frac{(\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N)'\mathbf{V}_N^{-1}(\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N)}{\mu^* - r_{ft}}$.

[45] We estimate the parameters needed for computing the MAXSER portfolio using our approach, instead of that in Ao et al. (2019), which explains the small differences with the numbers reported in their tables. Because we use the same data as in Ao et al. (2019), one could, of course, study the results reported in their paper if one wanted to see what one would get if one used their estimation approach. Moreover, because we are using the same universe of stocks as Ao et al. (2019), one could compare the performance of our proposed strategies to that for the 14 other portfolio strategies they report.

5.4 Robust-Mean-Variance Portfolios

We now describe the set of strategies based on the theoretical results developed in our paper and whose performance we compare to the benchmark strategies described above.

We report the performance of four *robust mean-variance* (RMV) strategies that rely on the results in Propositions 4.1–4.4. In our first strategy, labeled "RMV using $\mathbf{V}$," γ is specified to match the target volatility of 5% per month or $5\% \times \sqrt{12} = 17.32\%$ per year. The second strategy, "RMV using $\mathbf{V}$: OptComb," combines optimally the alpha portfolio and the beta portfolio, recognizing that for any sample with finite N the alpha and beta portfolios are not necessarily orthogonal; hence, when combining them optimally we exploit their cross-correlation.[46] The third strategy, "RMV using $\mathbf{\Omega}$," is obtained by recognizing that for large N, $\mathbf{B}_N' \mathbf{V}_N^{-1} \mathbf{B}_N \to \mathbf{\Omega}^{-1}$, implying that the beta portfolio return obtained from investing in the N assets is equivalent to the return from investing in only the $K = 3$ observed risk factors (i.e. the three Fama-French factors). Finally, the fourth strategy, "RMV using $\mathbf{\Omega}$: OptComb," combines optimally the alpha portfolio for the N assets and the beta-equivalent portfolio using only the $K = 3$ Fama-French factors.

Given that returns may not be distributed normally, we compute the t-statistic for the difference between two Sharpe ratios using the heteroskedasticity and autocorrelation robust (HAC) kernel estimation approach in Ledoit and Wolf (2008, sec. 3.1).

5.5 Evaluating Out-of-Sample Performance

Table 1 reports the out-of-sample performance for the portfolio strategies described above. Panel A of Table 1, which is for the case with $N = 30$ stock constituents of the Dow Jones Industrial Average (DJ30), shows that the annualized Sharpe ratio of the MV portfolio is 0.287, while that of the GMV-LW portfolio is 0.181.[47] Even though the number of risky assets for this data set is relatively small, $N = 30$, the EW portfolio with a Sharpe ratio of 0.334 outperforms these optimizing strategies. For the first data set, the best performing strategy based on the principal components of returns is the one with two principal components. However, none of the strategies based on one to ten principal components of returns outperform EW, so to save space, we display the results only for

[46] Let $r_t^\alpha = \mathbf{r}_t' \hat{\mathbf{w}_N}^\alpha$ and $r_t^\beta = \mathbf{r}_t' \hat{\mathbf{w}_N}^\beta$ be the time series of the realized portfolio returns associated with the (estimated) alpha and beta portfolios, respectively. Then the return to the strategy that optimally combines these two portfolios is $r_t^{\text{strategy}} = \hat{\gamma}^{-1}(\bar{r}^\alpha, \bar{r}^\beta)\widehat{\text{cov}}(r^\alpha, r^\beta)^{-1}(r_t^\alpha, r_t^\beta)'$, where $\widehat{\text{cov}}(r^\alpha, r^\beta)$ and $\bar{r}^\alpha, \bar{r}^\beta$ are the sample covariance matrix and the sample means of the r_t^α and r_t^β, respectively, and $\hat{\gamma} = (\sigma^*)^{-1}\left((\bar{r}^\alpha, \bar{r}^\beta)\widehat{\text{cov}}(r^\alpha, r^\beta)^{-1}(\bar{r}^\alpha, \bar{r}^\beta)'\right)^{\frac{1}{2}}$.

[47] There are two reasons for the poor performance of the GMV-LW portfolio relative to the MV portfolio: the number of assets is small, which helps the performance of the MV portfolio, and because we are using monthly data, estimates of the variances and covariances are not as precise as they would be with higher-frequency data.

the strategies based on two, three, four, and ten principal components. The MAXSER strategy of Ao et al. (2019) outperforms all these strategies: it has a Sharpe ratio of 0.426, which is 27.6% higher than the Sharpe ratio of the EW portfolio.[48]

We now evaluate the performance of our strategies, reported in the last four rows of Panel A of Table 1. The strategy "RMV using $\mathbf{V}$" achieves a Sharpe ratio of 0.556, which is 65.7% greater than that of the EW portfolio, consistent with the theoretical result in Propositions 4.1, and 29.8% higher than that of MAXSER. The Sharpe ratio of the "RMV using $\mathbf{V}$: OptComb" strategy, which combines the alpha and beta portfolios optimally, is even higher, 0.872, which is 160.9% larger than that of the EW portfolio and 104.5% higher than that of MAXSER, an increase that is both economically and statistically significant. The results for the other two strategies, "RMV using $\mathbf{\Omega}$" and "RMV using $\mathbf{\Omega}$: OptComb," are similar: the Sharpe ratios are greater than those of the EW by 72.6% and 100.3%, respectively, and greater than those of MAXSER by 35.2% and 57%, respectively, though in this case, the differences in Sharpe ratios are not significant statistically.

To evaluate our methodology when one has a larger number of assets, we report in Panel B of Table 1 the results for $N = 100$ stocks randomly selected from the S&P 500. We observe that the Sharpe ratio of the MV portfolio is now negative because estimation error increases with N (DeMiguel et al., 2009b). For this data set, the best performing strategy based on the principal components of returns is the one with four principal components. As for the previous data set, none of the strategies based on principal components outperform EW. The EW portfolio outperforms also MV and GMV-LW, while MAXSER does even better: it achieves a Sharpe ratio of 0.672, which is 36% higher than that of EW. However, our strategies perform even better: all four RMV strategies achieve Sharpe ratios substantially higher than those of EW and MAXSER. In particular, consistent with the result in Proposition 4.3 that, when N is large, one can replace the beta portfolio with a portfolio based only on the $K = 3$ Fama-French factors, using $\mathbf{\Omega}$ instead of $\mathbf{V}$, the Sharpe ratio is 1.222, which is 147.2% higher than that of EW and 81.8% higher than that of MAXSER, with significant t-statistics for both differences.

[48] Note that Ao et al. (2019) report in their Table 3 that their strategy, MAXSER, achieves an annual Sharpe ratio of 0.556; our estimate of the Sharpe ratio for their strategy is 0.471, with the difference being a consequence of differences in estimation methods. Also, for this data set, they use an estimation window of $T = 60$ months, while we use 120 months for both data sets. For the second data set, the Sharpe ratio we estimate for the MAXSER strategy of Ao et al. (2019) exceeds that in their paper.

Table 1: Out-of-sample portfolio performance

Panel A of this table reports, using monthly returns on the $N = 30$ stock constituents of the Dow Jones Industrial Average (DJ30), the performance of five strategies from the literature and four strategies developed in this paper when the investor targets a volatility of $0.05 \times \sqrt{12} = 0.1732$ per year. Panel B reports the same quantities for $N = 100$ randomly selected stocks from the S&P 500 index. The parameters for all strategies are estimated using a rolling window of $T = 120$ monthly observations. For each strategy, the table reports its per annum return's mean and Sharpe ratio. The table also reports the improvement in the Sharpe ratio of each strategy with respect to the EW and MAXSER strategies; for instance, when comparing strategy k to EW, SR wrt EW is $(\text{SR}_k - \text{SR}_{\text{EW}})/\text{SR}_{\text{EW}}$. Finally, the table reports the t-statistics for the difference between each strategy's Sharpe ratio and the EW and MAXSER strategies, computed as in Ledoit and Wolf (2008).

	Mean	SR	SR wrt		t-stat wrt	
	p.a.	p.a.	EW	MAXSER	EW	MAXSER
Panel A: For DJIA 30 constituents						
MV	0.045	0.287	−0.140	−0.326	−0.136	−0.714
GMV-LW	0.030	0.181	−0.458	−0.575	−0.611	−0.855
EW	0.058	0.334	0.000	−0.216	—	−0.303
PCA2	0.026	0.161	−0.517	−0.621	−0.770	−1.179
PCA3	0.013	0.087	−0.739	−0.795	−1.101	−1.510
PCA4	−0.019	−0.121	−1.362	−1.284	−2.030	−2.437
PCA10	−0.032	−0.194	−1.583	−1.457	−2.360	−2.767
MAXSER	0.061	0.426	0.276	0.000	0.303	—
RMV using $\mathbf{V}$	0.080	0.556	0.657	0.298	0.656	0.860
RMV using $\mathbf{V}$: OptComb	0.173	0.872	1.609	1.045	1.949	1.682
RMV using $\mathbf{\Omega}$	0.079	0.576	0.726	0.352	0.729	1.194
RMV using $\mathbf{\Omega}$: OptComb	0.129	0.669	1.003	0.570	0.600	0.519
Panel B: For S&P 500 constituents						
MV	−0.024	−0.190	−1.385	−1.283	−2.052	−3.112
GMV-LW	0.019	0.146	−0.704	−0.782	−1.176	−1.816
EW	0.070	0.494	0.000	−0.265	—	−0.495
PCA2	−0.001	−0.001	−1.003	−1.002	−2.206	−2.985
PCA3	0.056	0.334	−0.324	−0.502	−0.712	−1.497
PCA4	0.075	0.460	−0.068	−0.314	−0.150	−0.937
PCA10	0.033	0.203	−0.587	−0.696	−1.292	−2.075
MAXSER	0.094	0.672	0.360	0.000	0.495	—
RMV using $\mathbf{V}$	0.116	0.763	0.546	0.137	0.703	0.406
RMV using $\mathbf{V}$: OptComb	0.114	0.642	0.298	−0.046	0.952	0.875
RMV using $\mathbf{\Omega}$	0.137	1.016	1.055	0.511	1.959	1.623
RMV using $\mathbf{\Omega}$: OptComb	0.206	1.222	1.472	0.818	2.169	2.203

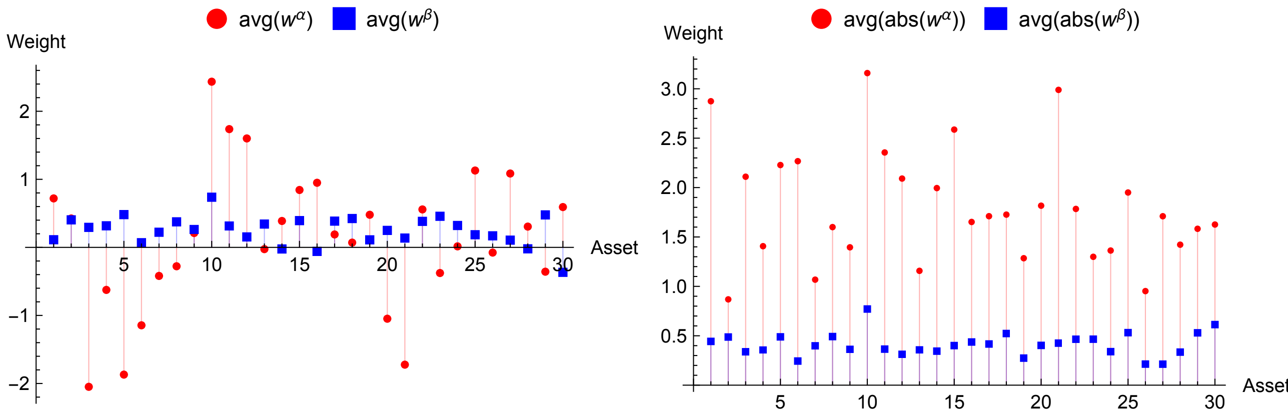
5.6 Importance of Latent Asset Demand

To understand the importance of latent asset demand, represented by the alpha portfolio, we examine the alpha- and beta-portfolio components, $\mathbf{w}_N^{\alpha}$ and $\mathbf{w}_N^{\beta}$, of the robust-mean-variance portfolio, $\mathbf{w}_N^{rmv}$ (i.e. RMV using $\mathbf{V}$). We also examine the components $\mathbf{w}_N^{a}$ and $\mathbf{w}_N^{A}$ that comprise

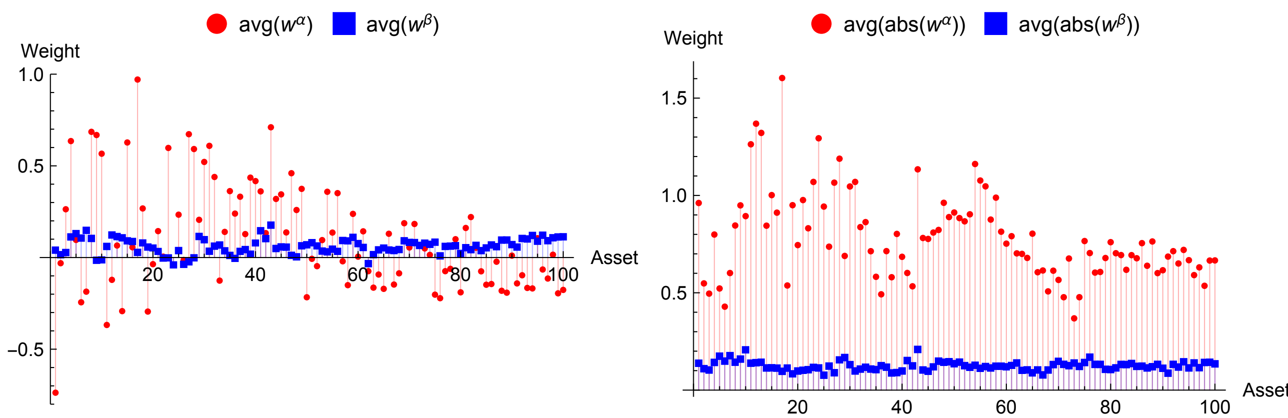
Figure 2: Weights for $\mathbf{w}_N^\alpha$ and $\mathbf{w}_N^\beta$ portfolios

The plots on the left Panels A and B report the time-series average of the w_i^α (red circles) and w_i^β (blue squares) components of the robust mean-variance portfolio for each of the N assets. The plots on the right report the time-series average of the *absolute* value of the w_i^α and w_i^β. Panel A reports these quantities for the $N = 30$ stock constituents of the Dow Jones Industrial Average (DJ30), while Panel B reports the same quantities for $N = 100$ randomly selected stocks from the S&P 500 index.

Panel A: For DJIA 30 constituents



Panel B: For S&P 500 constituents



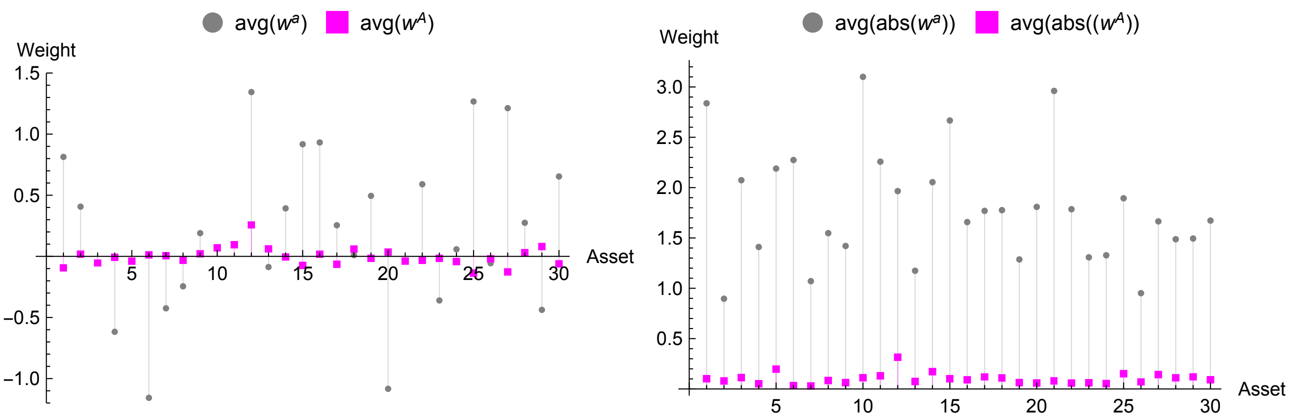
the alpha portfolio $\mathbf{w}_N^\alpha$. Recall that we have 240 out-of-sample months, so for each month and each of the portfolio components, we have a vector of, depending on the dataset we are considering, either $N = 30$ or $N = 100$ weights. We characterize these weights in a variety of ways.

In the plots on the left side of Panels A and B of Figure 2, we report for each of the N assets the time-series average of the w_i^α (in red circles) and w_i^β (in blue squares) components of the robust mean-variance portfolio. These plots show that the w_i^β are mostly positive. In contrast, the w_i^α are both positive and negative, implying long and short positions. The plots on the right side of Panels A and B of Figure 2 display the time-series average for each of the N assets of the *absolute* value of the elements of the alpha and beta portfolios. These plots show that the w_i^α weights, which

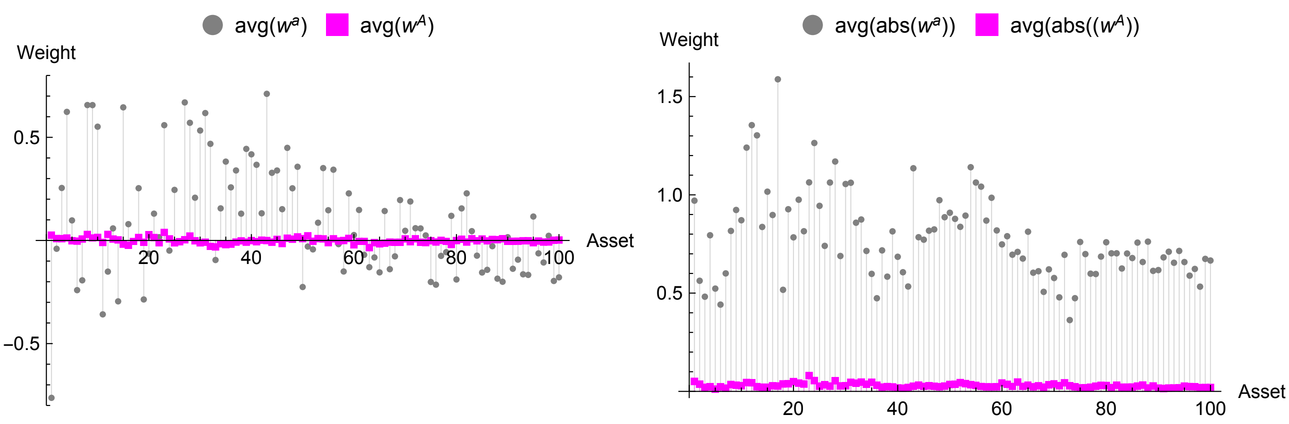
Figure 3: Weights for $\mathbf{w}_N^a$ and $\mathbf{w}_N^A$ portfolios

The plots on the left of Panels A and B report the time-series average of the w_i^a (in gray circles) and w_i^A (in magenta squares) components of the w_i^α portfolio for each of the N assets. The plots on the right report the time-series average of the *absolute* value of the w_i^a and w_i^A components for each of the N assets. Panel A reports these quantities for the case of the $N = 30$ stock constituents of the Dow Jones Industrial Average (DJ30), while Panel B reports the same quantities for $N = 100$ randomly selected stocks from the S&P 500 index.

Panel A: For DJIA 30 constituents



Panel B: For S&P 500 constituents



represent latent asset demand dominate the corresponding w_i^β weights, confirming the theoretical result in Proposition 4.2.

In Figure 3, we decompose the $\mathbf{w}_N^\alpha$ portfolio further into its $\mathbf{w}_N^a$ and $\mathbf{w}_N^A$ components, where the first weight is associated with the asset-specific component of returns and the second with the returns of omitted pervasive risk factors. Just as in the previous figure, the plots on the left side of Panels A and B show the time-series average of the w_i^a (in gray circles) and w_i^A (in magenta squares) weights. These plots show that the w_i^A are mostly positive. In contrast, the w_i^a are both positive and negative. The plots on the right side of Panels A and B of Figure 3 display the time-series average of the *absolute* values of w_i^a and w_i^A. These plots show that the w_i^a components dominate

Table 2: Squared-correlations of weights of portfolio components

This table reports the time-series average over the out-of-sample period of the squared correlations between the portfolio components of the robust mean-variance portfolio. Panel A reports the squared correlations between the portfolio components for the case of the $N = 30$ stocks that are part of the Dow Jones Industrial Average (DJ30), while Panel B reports the same quantities for the case of $N = 100$ randomly selected stocks from the S&P 500.

	$\mathbf{w}_N^{rmv}$	$\mathbf{w}_N^{\alpha}$	$\mathbf{w}_N^{\beta}$	$\mathbf{w}_N^{a}$	$\mathbf{w}_N^{A}$
Panel A: For DJIA 30 constituents					
$\mathbf{w}_N^{rmv}$	1.0000	0.9630	0.0391	0.9570	0.0224
$\mathbf{w}_N^{\alpha}$	0.9630	1.0000	0.0063	0.9930	0.0217
$\mathbf{w}_N^{\beta}$	0.0391	0.0063	1.0000	0.0058	0.0459
$\mathbf{w}_N^{a}$	0.9570	0.9930	0.0058	1.0000	0.0138
$\mathbf{w}_N^{A}$	0.0224	0.0217	0.0459	0.0138	1.0000
Panel B: For S&P 500 constituents					
$\mathbf{w}_N^{rmv}$	1.0000	0.9760	0.0311	0.9680	0.0067
$\mathbf{w}_N^{\alpha}$	0.9760	1.0000	0.0124	0.9920	0.0049
$\mathbf{w}_N^{\beta}$	0.0311	0.0124	1.0000	0.0124	0.0543
$\mathbf{w}_N^{a}$	0.9680	0.9920	0.0124	1.0000	0.0011
$\mathbf{w}_N^{A}$	0.0067	0.0049	0.0543	0.0011	1.0000

the corresponding w_i^A components, thus highlighting the importance of the purely asset-specific component of returns in driving latent asset demand.

In Table 2, we report the time-series average of the squared correlations between the different portfolio components. Panels A and B show that for both data sets, the squared correlations of the robust mean-variance portfolio $\mathbf{w}_N^{rmv}$ with the alpha-portfolio component $\mathbf{w}_N^{\alpha}$ is over 95%, with most of this coming from the weight on the purely asset-specific component, $\mathbf{w}_N^{a}$. In contrast, the squared correlations of $\mathbf{w}_N^{rmv}$ with the weight on the portfolio of observed and latent factors, $\mathbf{w}_N^{\beta}$ and $\mathbf{w}_N^{A}$, respectively, are substantially lower—only about 6%. These results, again, highlight the crucial role played by the latent-asset-demand component of the robust mean-variance portfolio.

Next we investigate the *returns* of the portfolio components. Panels A and B of Table 3 show that for both data sets, the squared correlations between the returns of the robust mean-variance portfolio r^{rmv} and alpha portfolio r^{α} are over 90%, with most of this coming from the purely asset-specific component, r^{a}. In sharp contrast, the squared correlations between r^{rmv} and the returns of the portfolios associated with observed and latent factors, r^{β} and r^{A}, are substantially smaller.

Finally, to examine the importance of latent asset demand in driving the performance of the robust mean-variance portfolio, in Table 4 we report the annualized means, standard deviations,

Table 3: Squared-correlations of returns of portfolio components

This table reports the squared correlations between the returns of the portfolio components of the robust mean-variance portfolio. Panel A reports the squared correlations between the returns of the portfolio components for the case of the $N = 30$ stocks that are part of the Dow Jones Industrial Average (DJ30), while Panel B reports the same quantities for the case of $N = 100$ randomly selected stocks from the S&P 500.

	r^{rmv}	r^{α}	r^{β}	r^{a}	r^{A}
Panel A: For DJIA 30 constituents					
r^{rmv}	1.0000	0.9060	0.0528	0.8900	0.1240
r^{α}	0.9060	1.0000	0.0065	0.9850	0.1320
r^{β}	0.0528	0.0065	1.0000	0.0068	0.0001
r^{a}	0.8900	0.9850	0.0068	1.0000	0.0612
r^{A}	0.1240	0.1320	0.0001	0.0612	1.0000
Panel B: For S&P 500 constituents					
r^{rmv}	1.0000	0.9610	0.0106	0.9490	0.0010
r^{α}	0.9610	1.0000	0.0091	0.9920	0.0001
r^{β}	0.0106	0.0091	1.0000	0.0110	0.0121
r^{a}	0.9490	0.9920	0.0110	1.0000	0.0063
r^{A}	0.0010	0.0001	0.0121	0.0063	1.0000

and Sharpe ratios of returns of the $\mathbf{w}_N^{rmv}$ portfolio (i.e. RMV using $\mathbf{V}$) and those of its components, $\mathbf{w}_N^{\alpha}$ and $\mathbf{w}_N^{\beta}$, along with the components of $\mathbf{w}_N^{\alpha}$, which are denoted by $\mathbf{w}_N^{a}$ and $\mathbf{w}_N^{A}$.

The first rows of Panels A and B of Table 4 show that more than 90% of the mean return of the robust mean-variance portfolio comes from the mean return of the alpha portfolio, which represents latent-asset demand, with the beta portfolio contributing only a small fraction. And most of the alpha portfolio's mean return comes from the purely asset-specific component, with the pervasive risk factors contributing a negligible amount. The second rows of Panels A and B show that the alpha portfolio also contributes to most of the risk of the robust mean-variance portfolio, with a large proportion of this coming from the asset-specific component. The Sharpe ratios, reported in the last rows of Panels A and B, show that the return-to-risk tradeoff for the asset-specific component is attractive: the Sharpe ratios of the alpha-portfolio component and, in particular, the part associated with the purely asset-specific return, is much larger than that of the observed and latent risk factors. Thus, consistent with the result in Proposition 4.1, if one were to ignore the latent demand from purely asset-specific pricing errors, the resulting portfolio would not be on the efficient frontier *regardless of how many risk factors were included in the model.*

Table 4: Moments of returns of portfolio components

This table reports the moments of returns of the portfolio components of the robust mean-variance portfolio (RMV using $\mathbf{V}$). The components of the $\mathbf{w}_N^{rmv}$ portfolio are $\mathbf{w}_N^{\alpha}$ and $\mathbf{w}_N^{\beta}$, with $\mathbf{w}_N^{\alpha}$ comprising of $\mathbf{w}_N^{a}$ and $\mathbf{w}_N^{A}$. The moments reported are the annualized mean return, standard deviation of return, and Sharpe ratio. Panel A reports the moments for the case of the $N = 30$ stocks that are part of the Dow Jones Industrial Average (DJ30), while Panel B reports the moments for the case of $N = 100$ randomly selected stocks from the S&P 500 index.

	r^{rmv}	r^{α}	r^{β}	r^{a}	r^{A}
Panel A: For DJIA 30 constituents					
Mean	0.0792	0.0760	0.0032	0.0710	0.0050
Standard deviation	0.1430	0.1390	0.0440	0.1340	0.0176
Sharpe ratio	0.5560	0.5460	0.0731	0.5300	0.2880
Panel B: For S&P 500 constituents					
Mean	0.1160	0.1070	0.0088	0.1070	0.0001
Standard deviation	0.1520	0.1520	0.0302	0.1520	0.0136
Sharpe ratio	0.7630	0.7050	0.2920	0.7020	0.0090

5.7 Source of Performance Gains Relative to EW and MAXSER Portfolios

The reason our strategies outperform the EW portfolio is explained by Propositions 4.1 and 4.3. Proposition 4.1 shows that the squared Sharpe ratio of the optimal mean-variance portfolio is the sum of the squared Sharpe ratios of the alpha and beta portfolios, while Proposition 4.3 shows that the beta portfolio can be replaced by a benchmark portfolio, such as the EW portfolio or a portfolio based on the three Fama-French factors, without any performance loss. That is, our beta portfolio, based on the three Fama-French factors, performs similar to the EW portfolio, so the performance gain of our strategies is the additional Sharpe ratio generated by the alpha portfolio. The alpha portfolio exploits the cross-sectional dispersion of the components of $\mathbf{a}_N$, a signal orthogonal to the observed factors, which is what the beta portfolio relies on. In other words, our methodology exploits what is missing from factor-based asset-pricing models, namely, asset-specific risk and return. Finally, the superior performance even with a large number of assets is a consequence of our strategy being founded on the APT, which is a theory that bites when N is large.

The differences between our portfolio strategy and MAXSER are more subtle. Both strategies rely on decomposing the mean-variance portfolio into alpha and beta components. In both strategies, the beta portfolio is the same—investing in the three Fama-French factors. The difference, therefore, arises from the alpha portfolio. The MAXSER strategy imposes sparsity of the alpha portfolio weights—see Ao et al. (2019, p. 2905, assumption C2)—which is achieved by imposing an ℓ_1 constraint that limits the number of assets included in the portfolio, as explained in Ao et al.

(2019, sec. 1.5.3). In contrast, because our methodology is founded on the APT, it works for large N regardless of T; therefore, our alpha portfolio does not restrict the number of assets in which it invests and so can take full advantage of all N assets available.

6 Conclusion

In this paper, our objective is to address model misspecification in mean-variance portfolios. We list below our novel results, each representing a significant departure from the existing literature.

First, in contrast to the classical interpretation of the APT, we show that the APT allows not just for small pricing errors but also large pricing errors related to missing pervasive risk factors. Second, we show that one can generate the entire set of mean-variance *efficient* portfolios from two *inefficient* portfolios: the "beta" portfolio, which depends on factor risk premia but not on pricing errors, and the "alpha" portfolio, representing latent asset demand, which depends only on latent pervasive and asset-specific risk, with zero exposure to observed risk factors. Third, in contrast to the traditional approaches for treating misspecification in mean-variance portfolios, we apply different methods to treat misspecification in the beta and alpha portfolios. For the beta portfolio, we identify a set of conditions under which it can be replaced, without any performance loss, by a benchmark portfolio (such as the equal- or value-weighted portfolio) that by construction is immune to misspecification. For the alpha portfolio, we treat misspecification using the robust-control approach of Hansen and Sargent (2007), which we show is equivalent to imposing the APT no-arbitrage restriction.

Finally, we demonstrate using two data sets (for monthly returns on stocks from DJ30 and S&P 500) that our theoretical results lead to a substantial improvement in out-of-sample portfolio performance relative to strategies in the existing literature. We also show the dominant role of the latent-asset-demand component of the optimal portfolio: for both data sets, the squared correlation between the optimal demand for an asset and its latent-demand component is over 95%, with the latent-demand component contributing more than 85% of the total (squared) Sharpe ratio of the optimal portfolio. These results support the findings in Koijen and Yogo (2019), who show the substantial role of latent asset demand in explaining empirically observed asset holdings.

The key takeaway from our work is that, what is usually regarded as mispricing, i.e. pricing errors, should instead be viewed as an integral part of the asset-pricing model. And, rather than searching for missing pervasive risk factors, our theoretical and empirical results highlight the importance of accounting for the purely asset-specific component of the pricing errors.

A Proofs for Propositions

In this appendix, we provide the proofs for the propositions in the main text of the manuscript.

Throughout the analysis, we need to study the limiting behavior of quantities of interest, such as portfolio weights and Sharpe ratios, as N diverges. This is conveniently achieved by the $\mathcal{O}(\cdot)$ and $\mathcal{o}(\cdot)$ notation. Specifically, we write $a_N = \mathcal{O}(b_N)$ for two generic sequences a_N and $b_N > 0$ if, as N increases, the ratio $|a_N|/b_N < \infty$; that is, the ratio is bounded. Similarly, we write $a_N = \mathcal{o}(b_N)$ if, as N increases, the ratio $|a_N|/b_N \to 0$; that is, the ratio goes to zero or, equivalently, a_N increases (decreases) at a rate slower (faster) than b_N. As special cases, when $b_N = 1$, then $a_N = \mathcal{O}(1)$ and $a_N = \mathcal{o}(1)$ simply mean that $|a_N| < \infty$ and $|a_N| \to 0$, respectively.[49] This notation is unambiguous for scalar quantities and easily extended to finite-dimensional vectors. However, we will also adopt it for vectors of portfolio weights, whose dimension grows with N, with the following meaning: when two generic portfolios satisfy $\mathbf{w}_N^a = \mathbf{w}_N^b + \mathcal{o}(1)$, we mean that, as N increases, the returns of portfolios a and b have an identical mean, variance, and Sharpe ratio.

A.1 Proof of Proposition 4.1

(i) Given that $\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N = \boldsymbol{\alpha}_N + \mathbf{B}_N\boldsymbol{\lambda}$,

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

However, in view of (20) and the Sherman-Morrison-Woodbury theorem applied to $\mathbf{V}_N$, which implies $\mathbf{V}_N^{-1} = \left[\boldsymbol{\Sigma}_N^{-1} - \boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N(\boldsymbol{\Omega}^{-1} + \mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1}\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\right]$, we get

$$\begin{aligned}
\mathbf{V}_N^{-1}\boldsymbol{\alpha}_N &= (\boldsymbol{\Sigma}_N^{+} + \mathbf{V}_N^{-1} - \boldsymbol{\Sigma}_N^{+})\boldsymbol{\alpha}_N \\
&= \boldsymbol{\Sigma}_N^{+}\boldsymbol{\alpha}_N + \boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N(\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1}\boldsymbol{\Omega}^{-1}(\boldsymbol{\Omega}^{-1} + \mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1}\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\alpha}_N \\
&= \boldsymbol{\Sigma}_N^{+}\boldsymbol{\alpha}_N + \boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\boldsymbol{\delta} \times \mathcal{O}(N^{-\frac{3}{2}}) = \boldsymbol{\Sigma}_N^{+}\boldsymbol{\alpha}_N + \mathcal{o}(1),
\end{aligned}$$

given

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

[49] For example, if $a_N = \mathcal{O}(N^{-\frac{1}{2}})$, it means that a_N goes to zero at most at (no faster than) the rate $N^{-\frac{1}{2}}$, whereas $a_N = \mathcal{o}(N^{-\frac{1}{2}})$ means that a_N goes to zero faster than $N^{-\frac{1}{2}}$. Alternatively, if $a_N = \mathcal{O}(N^{\frac{1}{2}})$, it means that a_N diverges at most at (no faster than) the rate $N^{\frac{1}{2}}$, whereas $a_N = \mathcal{o}(N^{\frac{1}{2}})$ means that a_N diverges slower than $N^{\frac{1}{2}}$.

$$
\begin{aligned}
&= \boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\left[-(\boldsymbol{\Omega}^{-1} + \mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1} + (\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1}\right]\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1} \\
&= \boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N(\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1}\left[-(\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N) + (\boldsymbol{\Omega}^{-1} + \mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)\right](\boldsymbol{\Omega}^{-1} + \mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1}\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1} \\
&= \boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N(\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1}\boldsymbol{\Omega}^{-1}(\boldsymbol{\Omega}^{-1} + \mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1}\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1},
\end{aligned}
$$

together with $(\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1} = \mathcal{O}(N^{-1})$, $(\boldsymbol{\Omega}^{-1} + \mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1} = \mathcal{O}(N^{-1})$, and $\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\alpha}_N = \mathcal{O}(N^{\frac{1}{2}})$, where $N^{\frac{3}{2}}(\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1}\boldsymbol{\Omega}^{-1}(\boldsymbol{\Omega}^{-1} + \mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1}\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\alpha}_N = \mathcal{O}(\boldsymbol{\delta})$, for some finite vector $\boldsymbol{\delta} \neq \mathbf{0}_K$.

(ii) The portfolios $\mathbf{w}_N^\alpha$ and $\mathbf{w}_N^\beta$ are, conditionally and unconditionally, orthogonal as N diverges.

$$
\begin{aligned}
\mathbf{w}_N^{\alpha\prime}\mathbf{V}_N\mathbf{w}_N^\beta &= \frac{1}{\gamma^\beta}\mathbf{w}_N^{\alpha\prime}\,\mathbf{V}_N\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda} = \frac{1}{\gamma^\beta}\mathbf{w}_N^{\alpha\prime}\,\mathbf{B}_N\boldsymbol{\lambda} \\
&= \frac{1}{\gamma^\alpha\gamma^\beta}\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^{+}\mathbf{B}\boldsymbol{\lambda} + \mathcal{O}(N^{-\frac{3}{2}})\frac{1}{\gamma^\alpha\gamma^\beta}\boldsymbol{\delta}'\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda} \\
&= \mathcal{O}(N^{-\frac{1}{2}}),
\end{aligned}
$$

because $\boldsymbol{\Sigma}_N^{+}\mathbf{B}_N = \mathbf{0}_K$ and $\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N = \mathcal{O}(N)$. Similarly, given $\boldsymbol{\Sigma}_N\mathbf{V}_N^{-1} = (\mathbf{I}_K - \mathbf{B}_N(\boldsymbol{\Omega}^{-1} + \mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1}\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1})$, we have

$$
\begin{aligned}
\mathbf{w}_N^{\alpha\prime}\boldsymbol{\Sigma}_N\mathbf{w}_N^\beta &= \frac{1}{\gamma^\beta}\mathbf{w}_N^{\alpha\prime}\boldsymbol{\Sigma}_N\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda} \\
&= \frac{1}{\gamma^\beta}\mathbf{w}_N^{\alpha\prime}\mathbf{B}_N(\mathbf{I}_K - (\boldsymbol{\Omega}^{-1} + \mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1}\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)\boldsymbol{\lambda} \\
&= \mathcal{O}(N^{-\frac{3}{2}})\boldsymbol{\delta}'(\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)(\boldsymbol{\Omega}^{-1} + \mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1}\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda} \\
&= \mathcal{O}(N^{-\frac{3}{2}}).
\end{aligned}
$$

We now show that, when N becomes large, $\mathbf{w}_N^\alpha$ and $\mathbf{w}_N^\beta$ are the minimum-variance portfolios orthogonal to one another. This is accomplished by showing that these portfolio weights satisfy, asymptotically, Proposition IA.2. In particular, we first need to verify that $\mathbf{w}_N^\beta$ is (asymptotically) the minimum variance, orthogonal, portfolio to $\mathbf{w}_N^\alpha$, by verifying

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

Analogously, one needs to verify that $\mathbf{w}_N^\alpha$ is (asymptotically) the minimum variance, orthogonal, portfolio to $\mathbf{w}_N^\beta$ but this part is omitted as it follows the same steps.

In view of $\boldsymbol{\Sigma}_N^{+}\mathbf{V}_N\boldsymbol{\Sigma}_N^{+} = \boldsymbol{\Sigma}_N^{+}\boldsymbol{\Sigma}_N\boldsymbol{\Sigma}_N^{+} = \boldsymbol{\Sigma}_N^{+}$, simple calculations lead to

$$
\begin{aligned}
\mu^\alpha &= \mathbf{w}_N^{\alpha\prime}\boldsymbol{\mu}_N + (1 - \mathbf{w}_N^{\alpha\prime}\mathbf{1}_N)r_{ft} = \mathbf{w}_N^{\alpha\prime}(\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N) + r_{ft} \\
&= \frac{1}{\gamma^\alpha}(\boldsymbol{\Sigma}_N^{+}\boldsymbol{\alpha}_N + \boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\boldsymbol{\delta} \times \mathcal{O}(N^{-\frac{3}{2}}))'(\boldsymbol{\alpha}_N + \mathbf{B}_N\boldsymbol{\lambda}) + r_{ft}
\end{aligned}
$$

$$= \frac{1}{\gamma^\alpha}\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N + r_{ft} + \mathcal{O}(N^{-\frac{1}{2}}), \tag{A1}$$

$$(\sigma^\alpha)^2 = \mathbf{w}_N^{\alpha\prime}\mathbf{V}_N\mathbf{w}_N^\alpha = \frac{1}{(\gamma^\alpha)^2}\boldsymbol{\alpha}_N'\mathbf{V}_N^{-1}\mathbf{V}_N\mathbf{V}_N^{-1}\boldsymbol{\alpha}_N = \frac{1}{(\gamma^\alpha)^2}\boldsymbol{\alpha}_N'\mathbf{V}_N^{-1}\boldsymbol{\alpha}_N$$

$$= \frac{1}{(\gamma^\alpha)^2}\boldsymbol{\alpha}_N'(\boldsymbol{\Sigma}_N^+ + \mathbf{V}_N^{-1} - \boldsymbol{\Sigma}_N^+)\boldsymbol{\alpha}_N = \frac{1}{(\gamma^\alpha)^2}\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N + \mathcal{O}(N^{-1}), \tag{A2}$$

and

$$\mu^\beta = \mathbf{w}_N^{\beta\prime}\boldsymbol{\mu}_N + (1 - \mathbf{w}_N^{\beta\prime}\mathbf{1}_N)r_{ft} = \mathbf{w}_N^{\beta\prime}(\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N) + r_{ft}$$

$$= \frac{1}{\gamma^\beta}\boldsymbol{\lambda}'\mathbf{B}_N'\mathbf{V}_N^{-1}(\boldsymbol{\alpha}_N + \mathbf{B}_N\boldsymbol{\lambda}) + r_{ft} = \frac{1}{\gamma^\beta}\boldsymbol{\lambda}'\mathbf{B}_N'\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda} + r_{ft} + \mathcal{O}(N^{-\frac{1}{2}}), \tag{A3}$$

$$(\sigma^\beta)^2 = \mathbf{w}_N^{\beta\prime}\mathbf{V}_N\mathbf{w}_N^\beta = \frac{1}{(\gamma^\beta)^2}\boldsymbol{\lambda}'\mathbf{B}_N'\mathbf{V}_N^{-1}\mathbf{V}_N\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda} = \frac{1}{(\gamma^\beta)^2}\boldsymbol{\lambda}'\mathbf{B}_N'\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda}, \tag{A4}$$

yielding

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

$$= \frac{1}{\left((\mathrm{SR}^{\mathrm{mv}})^2[\frac{\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N}{(\gamma^\alpha)^2} + \mathcal{O}(N^{-1})] - [\frac{\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N}{(\gamma^\alpha)} + \mathcal{O}(N^{-\frac{1}{2}})]^2\right)}$$

$$\times \begin{pmatrix} -\frac{\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N}{(\gamma^\alpha)} + \mathcal{O}(N^{-\frac{1}{2}}) \\ \frac{\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N}{(\gamma^\alpha)^2} + \mathcal{O}(N^{-1}) \end{pmatrix} \frac{\boldsymbol{\lambda}'\mathbf{B}_N'\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda} + \mathcal{O}(N^{-\frac{1}{2}})}{\gamma^\beta}.$$

Premultiplying the numerator of the latter expression by $\left(\mathbf{w}_N^\alpha, \mathbf{V}_N^{-1}(\boldsymbol{\alpha}_N + \mathbf{B}_N\boldsymbol{\lambda})\right)$ gives

$$\left(\frac{\mathbf{V}_N^{-1}\boldsymbol{\alpha}_N}{\gamma^\alpha}, \mathbf{V}_N^{-1}(\boldsymbol{\alpha}_N + \mathbf{B}_N\boldsymbol{\lambda})\right) \times \begin{pmatrix} -\frac{\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N}{(\gamma^\alpha)} + \mathcal{O}(N^{-\frac{1}{2}}) \\ \frac{\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N}{(\gamma^\alpha)^2} + \mathcal{O}(N^{-1}) \end{pmatrix} \frac{\boldsymbol{\lambda}'\mathbf{B}_N'\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda} + \mathcal{O}(N^{-\frac{1}{2}})}{\gamma^\beta}$$

$$= \left(\frac{\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N}{\gamma^\alpha} + \mathcal{O}(N^{-\frac{3}{2}}), \boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N + \mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda} + \mathcal{O}(N^{-\frac{3}{2}})\right)$$

$$\times \begin{pmatrix} -\frac{\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N}{(\gamma^\alpha)} + \mathcal{O}(N^{-\frac{1}{2}}) \\ \frac{\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N}{(\gamma^\alpha)^2} + \mathcal{O}(N^{-1}) \end{pmatrix} \frac{\boldsymbol{\lambda}'\mathbf{B}_N'\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda} + \mathcal{O}(N^{-\frac{1}{2}})}{\gamma^\beta}$$

$$= \frac{\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N}{(\gamma^\alpha)^2}\frac{(\boldsymbol{\lambda}'\mathbf{B}_N'\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda})}{\gamma^\beta}\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda} + \frac{(\boldsymbol{\lambda}'\mathbf{B}_N'\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda})}{\gamma^\beta(\gamma^\alpha)^2}(\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N)\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N$$

$$- \frac{(\boldsymbol{\lambda}'\mathbf{B}_N'\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda})}{\gamma^\beta(\gamma^\alpha)^2}(\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N)\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N + o(1)$$

$$= \frac{\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N}{(\gamma^\alpha)^2}\frac{(\boldsymbol{\lambda}'\mathbf{B}_N'\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda})}{\gamma^\beta}\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda} + o(1).$$

Given that, using (21), the denominator satisfies

$$\Big((\mathrm{SR}^{\mathrm{mv}})^2[\frac{\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N}{(\gamma^\alpha)^2}+\mathcal{O}(N^{-1})]-[\frac{\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N}{(\gamma^\alpha)}+\mathcal{O}(N^{-\frac{1}{2}})]^2\Big)$$
$$=\frac{1}{(\gamma^\alpha)^2}(\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N)(\boldsymbol{\lambda}'\mathbf{B}_N'\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda})+\mathcal{O}(N^{-\frac{1}{2}}),$$

by combining terms, one finally obtains

$$\Big(\mathbf{w}_N^\alpha,\mathbf{V}_N^{-1}(\boldsymbol{\mu}_N-r_{ft}\mathbf{1}_N)\Big)\begin{pmatrix}(\sigma^\alpha)^2 & \mu^\alpha-r_{ft}\\ \mu^\alpha-r_{ft} & (\mathrm{SR}^{\mathrm{mv}})^2\end{pmatrix}^{-1}\begin{pmatrix}0\\ \mu^\beta-r_{ft}\end{pmatrix}$$
$$=\frac{1}{\frac{1}{(\gamma^\alpha)^2}(\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N)(\boldsymbol{\lambda}'\mathbf{B}_N'\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda})}\frac{(\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N)}{(\gamma^\alpha)^2}\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda}\frac{(\boldsymbol{\lambda}'\mathbf{B}_N'\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda})}{\gamma^\beta}+o(1)$$
$$=\frac{\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda}}{\gamma^\beta}+o(1)=\mathbf{w}_N^\beta+o(1).$$

Along the same lines, it follows that $\mathbf{w}_N^\alpha$ satisfies

$$\Big(\mathbf{w}_N^\beta,\mathbf{V}_N^{-1}(\boldsymbol{\mu}_N-r_{ft}\mathbf{1}_N)\Big)\begin{pmatrix}(\sigma^\beta)^2 & \mu^\beta-r_{ft}\\ \mu^\beta-r_{ft} & (\mathrm{SR}^{\mathrm{mv}})^2\end{pmatrix}^{-1}\begin{pmatrix}0\\ \mu^\alpha-r_{ft}\end{pmatrix}=\mathbf{w}_N^\alpha+o(1),$$

implying that $\mathbf{w}_N^\alpha$ is (asymptotically) the minimum variance, orthogonal, portfolio to $\mathbf{w}_N^\beta$.

(iii) Given the result in (i) and (ii), two-fund separation holds by recognizing that $\mathbf{w}_N^\alpha$ (and thus $\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N/\gamma^\alpha$) and $\mathbf{w}_N^\beta$ are inefficient. This follows as long as both $\boldsymbol{\alpha}_N\neq\mathbf{0}_N$ and $\boldsymbol{\lambda}\neq\mathbf{0}_K$. In fact, given that one is the minimum-variance orthogonal portfolio to the other, if one was efficient, then by Proposition IA.2, its minimum-variance orthogonal portfolio exposure to the risky assets must be the zero vector, which is ruled out by our assumptions. However, by part (i), their linear combination always spans the efficient frontier.

With respect to Sharpe ratios, given (A1)–(A2), $\mathrm{SR}^\alpha=(\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N)^{\frac{1}{2}}+\mathcal{O}(N^{-\frac{1}{2}})$. Similarly, given (A3)–(A4), $\mathrm{SR}^\beta=(\boldsymbol{\lambda}'\mathbf{B}_N'\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda})^{\frac{1}{2}}+\mathcal{O}(N^{-\frac{1}{2}})$. The term $\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N\le\boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\alpha}_N$ is bounded as explained in (16), whereas

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

Moreover, $\mathbf{B}_N'\mathbf{V}_N^{-1}\mathbf{B}_N\to\boldsymbol{\Omega}^{-1}$, implying $\mathrm{SR}^\beta=(\boldsymbol{\lambda}'\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda})^{\frac{1}{2}}+\mathcal{O}(N^{-\frac{1}{2}})$. Finally,

$$(\mathrm{SR}^{\mathrm{mv}})^2=(\boldsymbol{\mu}_N-r_{ft}\mathbf{1}_N)'\mathbf{V}_N^{-1}(\boldsymbol{\mu}_N-r_{ft}\mathbf{1}_N)$$

$$
\begin{aligned}
&= (\boldsymbol{\alpha}_N + \mathbf{B}_N\boldsymbol{\lambda})'\mathbf{V}_N^{-1}(\boldsymbol{\alpha}_N + \mathbf{B}_N\boldsymbol{\lambda}) \\
&= \boldsymbol{\alpha}_N'\mathbf{V}_N^{-1}\boldsymbol{\alpha}_N + \boldsymbol{\lambda}_N'\mathbf{B}_N'\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda} + 2\boldsymbol{\alpha}_N'\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda} \\
&= \boldsymbol{\alpha}_N'\boldsymbol{\Sigma}_N^{+}\boldsymbol{\alpha}_N + \boldsymbol{\lambda}'\mathbf{B}_N'\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda} + \mathcal{O}(N^{-\frac{1}{2}}) \\
&= (\mathrm{SR}^{\alpha})^2 + (\mathrm{SR}^{\beta})^2 + \mathcal{O}(N^{-\frac{1}{2}}),
\end{aligned}
$$

implying, when $N \to \infty$, $(\mathrm{SR}^{\mathrm{mv}})^2 - (\mathrm{SR}^{\alpha})^2 - (\mathrm{SR}^{\beta})^2 \to 0$ because the conditions of Proposition IA.1 are satisfied. ■

A.2 Proof of Proposition 4.2

Assumption $\boldsymbol{\alpha}_N \neq \mathbf{0}_N$ rules out that $\mathbf{w}_N^{\alpha} = \mathbf{0}_N$ for any N. Recall that now $\boldsymbol{\Sigma}_N = \mathbf{A}_N\mathbf{A}_N' + \mathbf{C}_N$ and $\boldsymbol{\alpha}_N = \mathbf{A}_N\boldsymbol{\lambda}_{\mathrm{miss}} + \mathbf{a}_N$. Consider first $\mathbf{w}_N^{\alpha}$, where its ith component satisfies

$$
\begin{aligned}
w_{N,i}^{\alpha} &= \mathbf{1}_{N_i}'\mathbf{w}_N^{\alpha} = \mathbf{1}_{N_i}'\boldsymbol{\Sigma}_N^{+}\boldsymbol{\alpha}_N + o(1) \\
&= \frac{1}{\gamma^{\alpha}}\mathbf{1}_{N_i}'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\alpha}_N - \frac{1}{\gamma^{\alpha}}\mathbf{1}_{N_i}'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N(\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1}\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\alpha}_N + o(1),
\end{aligned}
$$

where $\mathbf{1}_{N_i}$ is an N-dimensional vector in which the ith element is one and the rest of the elements are zero. We deal with the two terms on the right-hand side of $w_{N,i}^{\alpha}$ separately. By the Sherman-Morrison-Woodbury theorem $\boldsymbol{\Sigma}_N^{-1} = \mathbf{C}_N^{-1} - \mathbf{C}_N^{-1}\mathbf{A}_N(\mathbf{I}_p + \mathbf{A}_N'\mathbf{C}_N^{-1}\mathbf{A}_N)^{-1}\mathbf{A}_N'\mathbf{C}_N^{-1}$, obtaining

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

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

$$
\begin{aligned}
|\mathbf{1}_{N_i}'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\alpha}_N| &= \mathcal{O}\Big(\|\boldsymbol{\lambda}_{\mathrm{miss}}\|\frac{\|\mathbf{1}_{N_i}'\mathbf{C}_N^{-1}\mathbf{A}_N\|}{N} + |\mathbf{1}_{N_i}'\mathbf{C}_N^{-1}\mathbf{a}_N| + \|\mathbf{a}_N\|\frac{\|\mathbf{1}_{N_i}'\mathbf{C}_N^{-1}\mathbf{A}_N\|}{N^{\frac{1}{2}}}\Big) \\
&= \mathcal{O}\Big(\frac{\|\boldsymbol{\lambda}_{\mathrm{miss}}\|}{N} + |\mathbf{1}_{N_i}'\mathbf{C}_N^{-1}\mathbf{a}_N| + \frac{\|\mathbf{a}_N\|}{N^{\frac{1}{2}}}\Big).
\end{aligned}
$$

Along the same lines,

$$
\begin{aligned}
\mathbf{1}_{N_i}'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N &= \mathbf{1}_{N_i}'\mathbf{C}_N^{-1}\mathbf{B}_N - \mathbf{1}_{N_i}'\mathbf{C}_N^{-1}\mathbf{A}_N(\mathbf{I}_p + \mathbf{A}_N'\mathbf{C}_N^{-1}\mathbf{A}_N)^{-1}\mathbf{A}_N'\mathbf{C}_N^{-1}\mathbf{B}_N, \\
\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N &= \mathbf{B}_N'\mathbf{C}_N^{-1}\mathbf{B}_N - \mathbf{B}_N'\mathbf{C}_N^{-1}\mathbf{A}_N(\mathbf{I}_p + \mathbf{A}_N'\mathbf{C}_N^{-1}\mathbf{A}_N)^{-1}\mathbf{A}_N'\mathbf{C}_N^{-1}\mathbf{B}_N, \quad \text{and}
\end{aligned}
$$

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

Therefore, using the same arguments as above, one obtains

$$|\mathbf{1}_{N_i}'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N(\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1}\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\alpha}_N| = \mathcal{O}\Big(\frac{\|\boldsymbol{\lambda}_{\text{miss}}\|}{N} + \frac{\|\mathbf{a}_N\|}{N^{\frac{1}{2}}}\Big),$$

because, 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. In particular, the first term $\mathbf{1}_{N_i}'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N$ is $\mathcal{O}(\|\mathbf{1}_{N_i}'\mathbf{C}_N^{-1}\mathbf{A}_N\| + \|\mathbf{1}_{N_i}'\mathbf{C}_N^{-1}\mathbf{B}_N\|) = \mathcal{O}(1)$, the second term $(\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1}$ is $\mathcal{O}(N^{-1})$, and the third term $\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\alpha}_N$ is $\mathcal{O}(\|\boldsymbol{\lambda}_{\text{miss}}\| + N^{\frac{1}{2}}\|\mathbf{a}_N\|)$.

For the $\mathbf{w}_N^{\beta}$ portfolio, its ith component satisfies

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

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 $\boldsymbol{\lambda} \neq \mathbf{0}_K$. ■

A.3 Proof of Proposition 4.3

The mean and variance of the excess return for the benchmark portfolio satisfy $\mathbf{w}_N^{\text{bench}\,\prime}(\boldsymbol{\alpha}_N + \mathbf{B}_N\boldsymbol{\lambda}) \to \mathbf{c}^{\text{bench}\,\prime}\boldsymbol{\lambda}$ and $\mathbf{w}_N^{\text{bench}\,\prime}\mathbf{V}_N\mathbf{w}_N^{\text{bench}} \to \mathbf{c}^{\text{bench}\,\prime}\boldsymbol{\Omega}\mathbf{c}^{\text{bench}}$, respectively.

Proof for (i) and (ii). Consider first the case $K > 1$ in which $\mathbf{c}^{\text{bench}} = \delta\,\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda}$, for some scalar $\delta \neq 0$. Then

$$(\text{SR}^{\text{bench}})^2 \to \frac{\delta^2(\boldsymbol{\lambda}'\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda})^2}{\delta^2(\boldsymbol{\lambda}'\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda})} = \boldsymbol{\lambda}'\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda},$$

where the last equality follows by combining (A3) and (A4). By Proposition IA.1, recalling $\text{SR}^{\beta} = (\boldsymbol{\lambda}'\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda})^{\frac{1}{2}} + o(1)$, the result follows. When $K = 1$, then $\mathbf{c}^{\text{bench}} = \delta\,\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda}$ for *every* scalar $\delta \neq 0$, as all the previous quantities are scalar, and thus $(\text{SR}^{\text{bench}})^2 \to \boldsymbol{\lambda}'\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda}$ under (24).

Proof for (iii). Now consider the case in which $\mathbf{c}^{\text{bench}}$ is not proportional to $\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda}$. Then

$$(\text{SR}^{\text{bench}})^2 \to \frac{((\mathbf{c}^{\text{bench}})'\boldsymbol{\lambda})^2}{(\mathbf{c}^{\text{bench}})'\boldsymbol{\Omega}\mathbf{c}^{\text{bench}}} < (\boldsymbol{\lambda}'\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda}) = (\text{SR}^{\beta})^2 + o(1),$$

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

The strict inequality is implied whenever $\boldsymbol{\Omega}^{\frac{1}{2}}\mathbf{c}^{\text{bench}}$ and $\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{\lambda}$ are not proportional, which in turn is equivalent to $\mathbf{c}^{\text{bench}}$ not being proportional to $\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda}$, as stated above. ■

A.4 Proof of Proposition 4.4

We first establish the solution (27), and then prove the asymptotic equivalence between the alpha portfolio from the APT and that from robust-control theory. Start by rewriting the max-min optimization as

$$\max_{\mathbf{w}_N} \min_{\boldsymbol{\alpha}_N} \left\{ \mathbf{w}_N'(\boldsymbol{\alpha}_N + \mathbf{B}_N\boldsymbol{\lambda}) - \frac{\gamma}{2}\mathbf{w}_N'\mathbf{V}_N\mathbf{w}_N \right\}.$$

We first show that the relative entropy constraint satisfies

$$\begin{aligned}
&\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\ln \frac{f_{\boldsymbol{\alpha}_N}(\mathbf{r}_t^e)}{f_{\hat{\boldsymbol{\alpha}}_N}(\mathbf{r}_t^e)} \right) f_{\boldsymbol{\alpha}_N}(\mathbf{r}_t^e)\, d\mathbf{r}_t^e \\
&= \frac{1}{2}\Big(\hat{\boldsymbol{\alpha}}_N'\mathbf{V}_N^{-1}\hat{\boldsymbol{\alpha}}_N - \boldsymbol{\alpha}_N'\mathbf{V}_N^{-1}\boldsymbol{\alpha}_N - 2\hat{\boldsymbol{\alpha}}_N'\mathbf{V}_N^{-1}E(\mathbf{r}_t^e - \mathbf{B}_N\boldsymbol{\lambda}) + 2\boldsymbol{\alpha}_N'\mathbf{V}_N^{-1}E(\mathbf{r}_t^e - \mathbf{B}_N\boldsymbol{\lambda}) \Big) \\
&= \frac{1}{2}\Big(\hat{\boldsymbol{\alpha}}_N'\mathbf{V}_N^{-1}\hat{\boldsymbol{\alpha}}_N - \boldsymbol{\alpha}_N'\mathbf{V}_N^{-1}\boldsymbol{\alpha}_N - 2\hat{\boldsymbol{\alpha}}_N'\mathbf{V}_N^{-1}\boldsymbol{\alpha}_N + 2\boldsymbol{\alpha}_N'\mathbf{V}_N^{-1}\boldsymbol{\alpha}_N \Big) \\
&= \frac{1}{2}(\hat{\boldsymbol{\alpha}}_N - \boldsymbol{\alpha}_N)'\mathbf{V}_N^{-1}(\hat{\boldsymbol{\alpha}}_N - \boldsymbol{\alpha}_N) \leq \frac{1}{2}(\hat{\boldsymbol{\alpha}}_N - \boldsymbol{\alpha}_N)'\boldsymbol{\Sigma}_N^{-1}(\hat{\boldsymbol{\alpha}}_N - \boldsymbol{\alpha}_N) \leq \delta_{\text{entropy}},
\end{aligned}$$

where the first inequality follows from using $\mathbf{V}_N^{-1} = \boldsymbol{\Sigma}_N^{-1} - \boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N(\boldsymbol{\Omega}^{-1} + \mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N)^{-1}\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}$.

Considering first the minimization step, one gets the Lagrangian

$$\mathcal{L}(\boldsymbol{\alpha}_N, \epsilon) = \mathbf{w}_N'\boldsymbol{\alpha}_N + \frac{\epsilon}{2}\left[(\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N)'\boldsymbol{\Sigma}_N^{-1}(\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N) - 2\delta_{\text{entropy}} \right],$$

where we need to find the saddle-point satisfying $\mathcal{L}({\boldsymbol{\alpha}_N}^{\text{rmv}}, \epsilon^{\text{rmv}}) = \min_{\boldsymbol{\alpha}_N} \max_{\epsilon \geq 0} \mathcal{L}(\boldsymbol{\alpha}_N, \epsilon)$. First consider the case in which $\epsilon > 0$. Then the first-order condition with respect to $\boldsymbol{\alpha}_N^{\text{rmv}}$ is

$$0 = \mathbf{w}_N + \epsilon\, \boldsymbol{\Sigma}_N^{-1}(\boldsymbol{\alpha}_N^{\text{rmv}} - \hat{\boldsymbol{\alpha}}_N),$$

and rearranging gives

$$\boldsymbol{\alpha}_N^{\text{rmv}} = \hat{\boldsymbol{\alpha}}_N - \frac{\boldsymbol{\Sigma}_N}{\epsilon}\mathbf{w}_N. \tag{A5}$$

Substituting the latter into the constraint $(\boldsymbol{\alpha}_N^{\text{rmv}} - \hat{\boldsymbol{\alpha}}_N)'\boldsymbol{\Sigma}_N^{-1}(\boldsymbol{\alpha}_N^{\text{rmv}} - \hat{\boldsymbol{\alpha}}_N) = 2\delta_{\text{entropy}}$, which holds with equality as $\epsilon > 0$, solving for ϵ^{rmv}, and substuting back into (A5) yields the final solution,

$$\boldsymbol{\alpha}_N^{\text{rmv}} = \hat{\boldsymbol{\alpha}}_N - \frac{(2\delta_{\text{entropy}})^{\frac{1}{2}}}{(\mathbf{w}_N'\boldsymbol{\Sigma}_N\mathbf{w}_N)^{\frac{1}{2}}}\boldsymbol{\Sigma}_N\mathbf{w}_N, \qquad \epsilon^{\text{rmv}} = \sqrt{\frac{\mathbf{w}_N'\boldsymbol{\Sigma}_N\mathbf{w}_N}{2\delta_{\text{entropy}}}}.$$

We now show that the Lagrangian is globally minimized at this point, ruling out the case $\epsilon = 0$.

$$\begin{aligned}\mathcal{L}(\boldsymbol{\alpha}_N, 0) &= \mathbf{w}_N'\boldsymbol{\alpha}_N = \mathbf{w}_N'\hat{\boldsymbol{\alpha}}_N + \mathbf{w}_N'(\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N) \\ &\geq \mathbf{w}_N'\hat{\boldsymbol{\alpha}}_N - |\mathbf{w}_N'(\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N)| \geq \mathbf{w}_N'\hat{\boldsymbol{\alpha}}_N - (\mathbf{w}_N'\boldsymbol{\Sigma}_N\mathbf{w}_N)^{\frac{1}{2}}((\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N)\boldsymbol{\Sigma}_N^{-1}(\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N))^{\frac{1}{2}} \\ &> \mathbf{w}_N'\hat{\boldsymbol{\alpha}}_N - (\mathbf{w}_N'\boldsymbol{\Sigma}_N\mathbf{w}_N)^{\frac{1}{2}}(2\delta_{\text{entropy}})^{\frac{1}{2}} = \mathcal{L}(\boldsymbol{\alpha}_N^{\text{rmv}}, \epsilon^{\text{rmv}}),\end{aligned}$$

where the second inequality follows from the Cauchy-Schwarz inequality, and the last, strict, inequality follows from the slackness condition for $\epsilon = 0$. The last term on the right-hand side is the Lagrangian evaluated when the relative entropy constraint is binding; i.e. for $\epsilon = \epsilon^{\text{rmv}} > 0$.

Consider now the maximization step,

$$\begin{aligned}&\max_{\mathbf{w}_N}\left\{\mathbf{w}_N'\left(\boldsymbol{\alpha}_N^{\text{rmv}} + \mathbf{B}_N\boldsymbol{\lambda}\right) - \frac{\gamma}{2}\mathbf{w}_N'\mathbf{V}_N\mathbf{w}_N\right\} \\ &= \max_{\mathbf{w}_N}\left\{\mathbf{w}_N'\left(\hat{\boldsymbol{\alpha}}_N - \frac{(2\delta_{\text{entropy}})^{\frac{1}{2}}}{(\mathbf{w}_N'\boldsymbol{\Sigma}_N\mathbf{w}_N)^{\frac{1}{2}}}\boldsymbol{\Sigma}_N\mathbf{w}_N + \mathbf{B}_N\boldsymbol{\lambda}\right) - \frac{\gamma}{2}\mathbf{w}_N'\mathbf{V}_N\mathbf{w}_N\right\} \\ &= \max_{\mathbf{w}_N}\left\{\mathbf{w}_N'\left(\hat{\boldsymbol{\alpha}}_N + \mathbf{B}_N\boldsymbol{\lambda}\right) - (2\delta_{\text{entropy}})^{\frac{1}{2}}(\mathbf{w}_N'\boldsymbol{\Sigma}_N\mathbf{w}_N)^{1/2} - \frac{\gamma}{2}\mathbf{w}_N'\mathbf{V}_N\mathbf{w}_N\right\}.\end{aligned}$$

By rearranging the first-order condition,

$$0 = \hat{\boldsymbol{\alpha}}_N + \mathbf{B}_N\boldsymbol{\lambda} - (2\delta_{\text{entropy}})^{\frac{1}{2}}(\mathbf{w}_N^{\text{rmv}\prime}\boldsymbol{\Sigma}_N\mathbf{w}_N^{\text{rmv}})^{-1/2}\boldsymbol{\Sigma}_N\mathbf{w}_N^{\text{rmv}} - \gamma\mathbf{V}_N\mathbf{w}_N^{\text{rmv}},$$

one obtains $\hat{\boldsymbol{\alpha}}_N + \mathbf{B}_N\boldsymbol{\lambda} = \left(\left(\frac{2\delta_{\text{entropy}}}{(\mathbf{w}_N^{\text{rmv}\prime}\boldsymbol{\Sigma}_N\mathbf{w}_N^{\text{rmv}})}\right)^{\frac{1}{2}}\boldsymbol{\Sigma}_N + \gamma\mathbf{V}_N\right)\mathbf{w}_N^{\text{rmv}}$. Recalling that $\mathbf{V}_N = \boldsymbol{\Sigma}_N + \mathbf{B}_N\boldsymbol{\Omega}\mathbf{B}_N'$, one obtains

$$\begin{aligned}\mathbf{w}_N^{\text{rmv}} &= \left(\left[\gamma + \left(\frac{2\delta_{\text{entropy}}}{(\mathbf{w}_N^{\text{rmv}\prime}\boldsymbol{\Sigma}_N\mathbf{w}_N^{\text{rmv}})}\right)^{\frac{1}{2}}\right]\boldsymbol{\Sigma}_N + \gamma\mathbf{B}_N\boldsymbol{\Omega}\mathbf{B}_N'\right)^{-1}(\hat{\boldsymbol{\alpha}}_N + \mathbf{B}_N\boldsymbol{\lambda}) \\ &= \frac{1}{\gamma}\left(\underbrace{\left[1 + \left(\frac{2\delta_{\text{entropy}}}{\gamma^2(\mathbf{w}_N^{\text{rmv}\prime}\boldsymbol{\Sigma}_N\mathbf{w}_N^{\text{rmv}})}\right)^{\frac{1}{2}}\right]}_{=\phi}\boldsymbol{\Sigma}_N + \mathbf{B}_N\boldsymbol{\Omega}\mathbf{B}_N'\right)^{-1}(\hat{\boldsymbol{\alpha}}_N + \mathbf{B}_N\boldsymbol{\lambda}) \\ &= \frac{1}{\gamma}\left(\phi\boldsymbol{\Sigma}_N + \mathbf{B}_N\boldsymbol{\Omega}\mathbf{B}_N'\right)^{-1}(\hat{\boldsymbol{\alpha}}_N + \mathbf{B}_N\boldsymbol{\lambda}), \quad \text{where we define} \\ \phi &= \left[1 + \left(\frac{2\delta_{\text{entropy}}}{\gamma^2(\mathbf{w}_N^{\text{rmv}\prime}\boldsymbol{\Sigma}_N\mathbf{w}_N^{\text{rmv}})}\right)^{\frac{1}{2}}\right].\end{aligned}$$

Notice that the solution for $\mathbf{w}_N^{\text{rmv}}$ is an implicit solution because ϕ itself is a function of $\mathbf{w}_N^{\text{rmv}}$.

To establish the asymptotic behavior of the robust mean-variance portfolios, using the Sherman-Morrison-Woodbury theorem one obtains

$$\mathbf{w}_N^{\text{rmv}} = \frac{1}{\gamma\phi}\left(\boldsymbol{\Sigma}_N + \frac{1}{\phi}\mathbf{B}_N\boldsymbol{\Omega}\mathbf{B}_N'\right)^{-1}(\hat{\boldsymbol{\alpha}}_N + \mathbf{B}_N\boldsymbol{\lambda})$$

$$
\begin{aligned}
&= \frac{1}{\gamma\phi}\left(\boldsymbol{\Sigma}_N^{-1} - \boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\left(\phi\boldsymbol{\Omega}^{-1} + \mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\right)^{-1}\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\right)(\hat{\boldsymbol{\alpha}}_N + \mathbf{B}_N\boldsymbol{\lambda}) \\
&= \frac{1}{\gamma\phi}\left(\boldsymbol{\Sigma}_N^{-1}\hat{\boldsymbol{\alpha}}_N - \boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\left(\phi\boldsymbol{\Omega}^{-1} + \mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\right)^{-1}\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\hat{\boldsymbol{\alpha}}_N\right) \\
&\quad + \frac{1}{\gamma\phi}\left(\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N - \boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\left(\phi\boldsymbol{\Omega}^{-1} + \mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\right)^{-1}\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\right)\boldsymbol{\lambda} \\
&= \frac{1}{\gamma\phi}\left(\boldsymbol{\Sigma}_N^{-1}\hat{\boldsymbol{\alpha}}_N - \boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\left(\phi\boldsymbol{\Omega}^{-1} + \mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\right)^{-1}\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\hat{\boldsymbol{\alpha}}_N\right) \\
&\quad + \frac{1}{\gamma}\left(\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\left(\phi\boldsymbol{\Omega}^{-1} + \mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\right)^{-1}\right)\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda}.
\end{aligned}
$$

For large N, one obtains

$$
\begin{aligned}
\mathbf{w}_N^{\text{rmv}} &= \frac{1}{\gamma\phi}\left(\boldsymbol{\Sigma}_N^{-1} - \boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\left(\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\right)^{-1}\mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\right)\hat{\boldsymbol{\alpha}}_N \\
&\quad + \frac{1}{\gamma}\left(\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\left(\boldsymbol{\Omega}^{-1} + \mathbf{B}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{B}_N\right)^{-1}\right)\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda} + \mathcal{O}(1) \\
&= \frac{1}{\gamma\phi}\boldsymbol{\Sigma}_N^{+}\hat{\boldsymbol{\alpha}}_N + \frac{1}{\gamma}\mathbf{V}_N^{-1}\mathbf{B}_N\boldsymbol{\lambda} + \mathcal{O}(1) \\
&= \phi^{\alpha}\mathbf{w}_N^{r\alpha} + \phi^{\beta}\mathbf{w}_N^{\beta} + \mathcal{O}(1).
\end{aligned}
$$

We now establish an asymptotic equivalence of the shrinkage parameter ϕ. This follows from the fact that $\mathbf{w}_N^{\beta}$ diversifies asset-specific risk, namely, that $\mathbf{w}_N^{\beta}\boldsymbol{\Sigma}_N\mathbf{w}_N^{\beta} \to 0$ (unlike $\mathbf{w}_N^{\alpha}$, which diversifies away only common risk). Therefore, $\mathbf{w}_N^{\text{rmv}\prime}\boldsymbol{\Sigma}_N\mathbf{w}_N^{\text{rmv}} = \mathbf{w}_N^{r\alpha\prime}\boldsymbol{\Sigma}_N\mathbf{w}_N^{r\alpha} + \mathcal{O}(1) = (\gamma\phi)^{-2}\hat{\boldsymbol{\alpha}}_N'\boldsymbol{\Sigma}_N^{+}\hat{\boldsymbol{\alpha}}_N + \mathcal{O}(1)$, yielding

$$
\begin{aligned}
\phi &= \left[1 + \left(\frac{2\delta_{\text{entropy}}}{\gamma^2(\mathbf{w}^{\text{rmv}\prime}\boldsymbol{\Sigma}_N\mathbf{w}^{\text{rmv}})}\right)^{\frac{1}{2}}\right] \qquad \text{(A6)}\\
&= \left[1 + \left(\frac{2\delta_{\text{entropy}}}{\gamma^2(\gamma\phi)^{-2}\hat{\boldsymbol{\alpha}}_N'\boldsymbol{\Sigma}_N^{+}\hat{\boldsymbol{\alpha}}_N}\right)^{\frac{1}{2}}\right] + \mathcal{O}(1) \\
&= \left[1 + \left(\frac{(2\delta_{\text{entropy}})^{\frac{1}{2}}\,|\phi|}{(\hat{\boldsymbol{\alpha}}_N'\boldsymbol{\Sigma}_N^{+}\hat{\boldsymbol{\alpha}}_N)^{\frac{1}{2}}}\right)\right] + \mathcal{O}(1).
\end{aligned}
$$

This provides a nonlinear equation in ϕ (due to the modulus) but, by seeking only the positive solution $\phi > 0$, one obtains the (approximate) linear equation

$$
\phi = \left[1 + \left(\frac{(2\delta_{\text{entropy}})^{\frac{1}{2}}\phi}{(\hat{\boldsymbol{\alpha}}_N'\boldsymbol{\Sigma}_N^{+}\hat{\boldsymbol{\alpha}}_N)^{\frac{1}{2}}}\right)\right] + \mathcal{O}(1),
$$

with solution:

$$
\phi = \left[1 - (2\delta_{\text{entropy}})^{\frac{1}{2}}(\hat{\boldsymbol{\alpha}}_N'\boldsymbol{\Sigma}_N^{+}\hat{\boldsymbol{\alpha}}_N)^{-\frac{1}{2}}\right]^{-1} + \mathcal{O}(1), \qquad \text{(A7)}
$$

assuming $2\delta_{\text{entropy}} < (\hat{\boldsymbol{\alpha}}_N'\boldsymbol{\Sigma}_N^{+}\hat{\boldsymbol{\alpha}}_N)$.

We now establish the (asymptotic) equivalence between the two alpha portfolios, assuming that the robust alpha portolio $\mathbf{w}_N^{r\alpha}$ is constructed with respect to some given $\delta^*_{\text{apt}}, \delta_{\text{entropy}}$. Given that $\boldsymbol{\alpha}_N = \mathbf{a}_N + \mathbf{A}_N\boldsymbol{\lambda}_{\text{miss}}$, as $N \to \infty$ the ith element of $\mathbf{w}_N^{r\alpha}$ is dominated by the component in $\mathbf{a}_N$ by Proposition 4.2, namely

$$
\begin{aligned}
w_i^{r\alpha}(\delta^*_{\text{apt}}, \delta_{\text{entropy}}) &= \frac{1}{\gamma^\alpha \phi}\mathbf{1}'_{N_i}\boldsymbol{\Sigma}_N^+\hat{\mathbf{a}}_N + \mathcal{O}(1) \\
&= \frac{1}{\gamma^\alpha \phi}\mathbf{1}'_{N_i}\boldsymbol{\Sigma}_N^+\frac{1}{(\hat{\kappa}+1)}\left(\bar{\mathbf{r}}_N^e - \hat{\mathbf{B}}_{N,\text{MLC}}\,\bar{\mathbf{f}}^e - \hat{\mathbf{A}}_{N,\text{MLC}}\,\hat{\boldsymbol{\lambda}}_{\text{miss,MLC}}\right) + \mathcal{O}(1) \\
&= \delta^{*\frac{1}{2}}_{\text{apt}}\left(1 - \frac{(2\delta_{\text{entropy}})^{\frac{1}{2}}}{(\hat{\boldsymbol{\alpha}}'_N\boldsymbol{\Sigma}_N^+\hat{\boldsymbol{\alpha}}_N)^{\frac{1}{2}}}\right)c_i + \mathcal{O}(1) \\
&= \delta^{\frac{1}{2}}_{\text{apt}}c_i + \mathcal{O}(1) = w_i^\alpha(\delta_{\text{apt}}) + \mathcal{O}(1),
\end{aligned}
$$

setting $\delta^{\frac{1}{2}}_{\text{apt}} = \delta^{*\frac{1}{2}}_{\text{apt}}\left(1 - \frac{(2\delta_{\text{entropy}})^{\frac{1}{2}}}{(\hat{\boldsymbol{\alpha}}'_N\boldsymbol{\Sigma}_N^+\hat{\boldsymbol{\alpha}}_N)^{\frac{1}{2}}}\right)$, for some constant c_i, function of the data, of γ^α, and of $\boldsymbol{\Sigma}_N^+$, where the second equality follows from the formula for the estimator $\hat{\mathbf{a}}_N$ in Proposition IA.2 for a given δ^*_{apt}, based on

$$
(1+\hat{\kappa})^2 = \frac{\left[\bar{\mathbf{r}} - r_{ft}\mathbf{1}_N - \hat{\mathbf{B}}_N(\bar{\mathbf{f}} - r_{ft}\mathbf{1}_K) - \hat{\mathbf{A}}_N\hat{\boldsymbol{\lambda}}_{\text{miss}}\right]'\hat{\boldsymbol{\Sigma}}_N^{-1}\left[\bar{\mathbf{r}} - r_{ft}\mathbf{1}_N - \hat{\mathbf{B}}_N(\bar{\mathbf{f}} - r_{ft}\mathbf{1}_K) - \hat{\mathbf{A}}_N\hat{\boldsymbol{\lambda}}_{\text{miss}}\right]}{\delta^*_{\text{apt}}},
$$

and the third equality follows from (A7). Finally, note that $\delta^{\frac{1}{2}}_{\text{apt}}c_i$ is (asymptotically) equivalent to the (estimated) alpha portfolio weight w_i^α, based on $\hat{\boldsymbol{\alpha}}_N(\delta_{\text{apt}})$. ■

Corollary A.1. *Under the assumptions of Proposition 4.4 and* $\delta_{entropy} = \frac{1}{2}\frac{(1+\boldsymbol{\lambda}'\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda})}{T}\chi^2_{N,x\%}$, *with* $\chi^2_{N,x\%}$ *the* x^{th} $(0 \le x \le 1)$ *quantile of a* χ^2_N *distribution, the shrinkage parameter* ϕ *satisfies*

$$
\phi = \left[1 + \left(\frac{\chi^2_{N,x\%}}{\gamma^2 T}\frac{(1+\boldsymbol{\lambda}'\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda})}{(\mathbf{w}_N^{rmv\prime}\boldsymbol{\Sigma}_N\mathbf{w}_N^{rmv})}\right)^{\frac{1}{2}}\right] = \frac{1}{\left(1 - (\frac{\chi^2_{N,x\%}}{T})^{\frac{1}{2}}\frac{(1+\boldsymbol{\lambda}'\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda})^{\frac{1}{2}}}{(\breve{\boldsymbol{\alpha}}'_N\boldsymbol{\Sigma}_N^+\breve{\boldsymbol{\alpha}}_N)^{\frac{1}{2}}}\right)} + \mathcal{O}(1)\ as\, N \to \infty. \tag{A8}
$$

To interpret expression (A8), note that ϕ is increasing with $\chi^2_{N,x\%}$, which increases (approximately) linearly with the number of assets N (because $E[\chi^2_N] = N$), and $(\boldsymbol{\lambda}'\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda})^{\frac{1}{2}}$, which is the Sharpe ratio of the beta portfolio as $N \to \infty$, and decreases with the sample size T (because the investor gathers more precise informations about the parameters, and $(\boldsymbol{\alpha}'_N\boldsymbol{\Sigma}_N^+\boldsymbol{\alpha}_N)^{\frac{1}{2}}$, which is the Sharpe ratio of the alpha portfolio as $N \to \infty$. ■

References

Aït-Sahalia, Yacine, and Michael W. Brandt, 2001, Variable selection for portfolio choice, *Journal of Finance* 56, 1297–1351.

Al-Najjar, N., 1999, On the robustness of factor structures to asset repackaging, *Journal of Mathematical Economics* 31, 309–320.

Al-Najjar, Nabil I., 1998, Factor analysis and arbitrage pricing in large asset economies, *Journal of Economic Theory* 78, 231–262.

Anderson, T. M., 1984, *An Introduction to Multivariate Statistical Analysis* (John Wiley, New York).

Ang, A., and D. Kristensen, 2012, Testing conditional factor models, *Journal of Financial Economics* 106, 132–156.

Ang, Andrew, Jun Liu, and Krista Schwarz, 2010, Using stocks or portfolios in tests of factor models, *Journal of Financial and Quantitative Analysis* 00, 1–42.

Ao, Mengmeng, Yingying Li, and Xinghua Zheng, 2019, Approaching mean-variance efficiency for large portfolios, *Review of Financial Studies* 32, 2890–2919.

Black, Fischer, 1995, Estimating expected return, *Financial Analysts Journal* 168–171.

Black, Fischer, Michael C. Jensen, and Myron Scholes, 1972, The capital asset pricing model: Some empirical tests, in Michael C. Jensen, (ed.) *Studies in the Theory of Capital Markets* (Praeger Publishers, New York, NY).

Brandt, Michael W., 1999, Estimating portfolio and consumption choice: A conditional Euler equations approach, *Journal of Finance* 54, 1609–1646.

Chamberlain, Gary, 1983, Funds, factors and diversification in arbitrage pricing models, *Econometrica* 51, 1305–1324.

Chamberlain, Gary, 1987, Asymptotic efficiency in estimation with conditional moment restrictions, *Journal of Econometrics* 34, 305–334.

Chamberlain, Gary, and Michael Rothschild, 1983, Arbitrage, factor structure and mean-variance analysis on large asset markets, *Econometrica* 51, 1281–1304.

Christensen, Timothy M, 2017, Nonparametric stochastic discount factor decomposition, *Econometrica* 85, 1501–1536.

Cochrane, John H., and Jesús Saá-Requejo, 2001, Beyond arbitrage: Good-deal asset price bounds in incomplete markets, *Journal of Political Economy* 108, 79–119.

Connor, Gregory, Matthias Hagmann, and Oliver Linton, 2012, Efficient estimation of a semiparametric characteristic-based factor model of security returns, *Econometrica* 80, 713–754.

Connor, Gregory, and Oliver Linton, 2007, Semiparametric estimation of a characteristic-based factor model of common stock returns, *Journal of Empirical Finance* 14, 694–717.

Daniel, Kent, Lira Mota, Simon Rottke, and Tano Santos, 2020, The cross-section of risk and returns, *Review of Financial Studies* 33, 1927–1979.

Daniel, Kent, and Sheridan Titman, 1997, Evidence on the characteristics of cross sectional variation in stock returns, *Journal of Finance* 52, 1–33.

DeMiguel, Victor, Lorenzo Garlappi, Francisco J. Nogales, and Raman Uppal, 2009a, A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms, *Management Science* 55, 798–812.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2009b, Optimal versus naïve diversification: How inefficient is the 1/N portfolio strategy?, *Review of Financial Studies* 22, 1915–1953.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Fan, Jianqing, Yuan Liao, and Weichen Wang, 2016, Projected principal component analysis in factor models, *Annals of statistics* 44, 219.

Ferson, Wayne E, and Campbell R Harvey, 1991, The variation of economic risk premiums, *Journal of political economy* 99, 385–415.

Gagliardini, Patrick, Elisa Ossola, and Olivier Scaillet, 2016, Time-varying risk premium in large cross-sectional equity data sets, *Econometrica* 84, 985–1046.

Gagliardini, Patrick, Elisa Ossola, and Olivier Scaillet, 2019a, A diagnostic criterion for approximate factor structure, *Journal of Econometrics* 212, 503–521.

Gagliardini, Patrick, Elisa Ossola, and Olivier Scaillet, 2019b, Estimation of large dimensional conditional factor models in finance, in S. Durlauf, L. Hansen, J. Heckman, and R. Matzkin, (eds.) *Handbook of Econometrics*, volume 7 (Elsevier, Amsterdam).

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

Ghosh, Anisha, Christian Julliard, and Alex P. Taylor, 2019, An information-theoretic asset pricing model, Working Paper, London School of Economics.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.

Gilboa, Itzhak, and David Schmeidler, 1989, Maxmin expected utility theory with non-unique prior, *Journal of Mathematical Economics* 18, 141–153.

Green, Richard C., and Burton Hollifield, 1992, When will mean-variance efficient portfolios be well diversified?, *Journal of Finance* 47, 1785–1809.

Guidolin, Massimo, and Francesca Rinaldi, 2009, A simple model of trading and pricing risky assets under ambiguity: Any lessons for policy makers, working Paper, Federal Reserve Board, St. Louis.

Hansen, Lars Peter., and Scott F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 587–613.

Hansen, Lars Peter, and Thomas J. Sargent, 1999, Wanting robustness in macroeconomics, working Paper, University of Chicago.

Hansen, Lars Peter, and Thomas J. Sargent, 2007, *Robustness* (Princeton University Press, Princeton, NJ).

Hastie, Trevor, Robert Tibshirani, and Martin Wainwright, 2015, *Statistical Learning with Sparsity: The Lasso and Generalizations* (CRC Press).

Huang, Chi-Fu, and Robert H. Litzenberger, 1988, *Foundations of Financial Economics* (Elsevier Science, New York, NY).

Huberman, Gur, 1982, A simple approach to the Arbitrage Pricing Theory, *Journal of Economic Theory* 28, 183–191.

Ingersoll, Jonathan E., Jr., 1984, Some results in the theory of arbitrage pricing, *Journal of Finance* 39, 1021–1039.

Jagannathan, Ravi, and Tongshu Ma, 2003, Risk reduction in large portfolios: Why imposing the wrong constraints helps, *Journal of Finance* 58, 1651–1684.

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

Kelly, Bryan T, Seth Pruitt, and Yinan Su, 2020, Instrumented principal component analysis, *Available at SSRN 2983919* .

Kim, Soohun, Robert A Korajczyk, and Andreas Neuhierl, 2021, Arbitrage portfolios, *The Review of Financial Studies* 34, 2813–2856.

Koijen, Ralph S. J., and Motohiro Yogo, 2019, A demand system approach to asset pricing, *Journal of Political Economy* 127, 1475–1515.

Korsaye, Sofonias A, Alberto Quaini, and Fabio Trojani, 2020, Smart stochastic discount factors, SSRN Working Paper.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2018, Interpreting factor models, *The Journal of Finance* 73, 1183–1223.

Ledoit, Olivier, and Michael Wolf, 2003, Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, *Journal of Empirical Finance* 10, 603–621.

Ledoit, Olivier, and Michael Wolf, 2008, Robust performance hypothesis testing with the Sharpe ratio, *Journal of Empirical Finance* 15, 850–859.

Lettau, Martin, and Markus Pelger, 2020, Factors that fit the time-series and cross-section of stock returns, *Review of Financial Studies* 33, 2274–2325.

Lewellen, J., and S. Nagel, 2006, The conditional CAPM does not explain asset-pricing anomalies, *Journal of Financial Economics* 82, 289–314.

Lo, Andrew W., 2002, The statistics of Sharpe ratios, *Financial Analysts Journal* 58, 36–52.

MacKinlay, A. Craig, 1995, Multifactor models do not explain deviations from the capm, *Journal of Financial Economics* 38, 3–28.

MacKinlay, A. Craig, and Ľuboš Pástor, 2000, Asset pricing models: Implications for expected returns and portfolio selection, *Review of Financial Studies* 13, 883–916.

Merton, Robert C., 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* 8, 323–361.

Pástor, Ľuboš, 2000, Portfolio selection and asset pricing models, *Journal of Finance* 55, 179–223.

Pástor, Ľuboš, and Robert F. Stambaugh, 1999, Cost of equity capital and model mispricing, *Journal of Finance* 54, 67–121.

Pástor, Ľuboš, and Robert F. Stambaugh, 2000, Comparing asset pricing models: An investment perspective, *Journal of Financial Economics* 56, 335–381.

Pelger, M., and R. Xiong, 2019, State-varying factor models of large dimensions, *Working Paper, SSRN: https://ssrn.com/abstract=3109314* .

Pettenuzzo, Davide, Allan G. Timmermann, and Rossen Valkanov, 2014, Forecasting stock returns under economic constraints, *Journal of Financial Economics* 114, 517–553.

Reisman, Haim, 1992, Reference variables, factor structure, and the approximate multibeta representation, *Journal of Finance* 47, 1303–1314.

Roll, Richard, 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, 129–176.

Roll, Richard, 1980, Orthogonal portfolios, *Journal of Financial and Quantitative Analysis* 15, 1005–1023.

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Ross, Stephen A., 1977, Return, risk, and arbitrage, in Irwin Friend and J.L. Bicksler, (eds.) *Risk and Return in Finance* (Ballinger, Cambridge, MA).

Schneider, Paul, and Fabio Trojani, 2019, (Almost) model-free recovery, *Journal of Finance* 74, 323–370.

Sharpe, W., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 1919, 425–442.

Stambaugh, R.F., 1983, Asset pricing with information, *Journal of Financial Economics* 12, 357–369.

Treynor, Jack L., and Fischer Black, 1973, How to use security analysis to improve portfolio selection, *Journal of Business* 46, 66–86.

Trojani, Fabio, and P. Vanini, 2002, A note on robustness in Merton's model of intertemporal consumption and portfolio choice, *Journal of Economic Dynamics and Control* 26, 423–435.

Uppal, Raman, and Tan Wang, 2003, Model misspecification and underdiversification, *Journal of Finance* 58, 2465–2486.

Zaffaroni, P, 2020, Factor models for conditional asset pricing: Online appendix, *Working Paper, Imperial College London* .

Internet Appendix to

Robust Portfolio Choice

Section IA.1 of this internet appendix shows that for mean-variance portfolios and their decomposition into alpha and beta portfolios, when the APT assumptions hold and N is large, then one obtains the same results regardless of whether one starts from the orthogonal or non-orthogonal representation of returns. Section IA.2 contains auxiliary results about the decomposition of the Sharpe ratio and extends the results in Roll (1980) to the case in which investors can invest also in a risk-free asset. Section IA.3 explains how we estimate the APT model using maximum likelihood. Section IA.4 compares the performance of the strategies we develop in this paper to that of the benchmark strategies for simulated data, i.e. in a controlled environment.

IA.1 Orthogonal vs. Nonorthogonal Representation of Returns

As we explain in the main text, our model is designed to exploit exactly the kind of anomaly identified in Black, Jensen, and Scholes (1972) by constructing what Treynor and Black (1973) call "active" and "passive" portfolios, which we label "alpha" and "beta" portfolios. Below, we provide a specific model designed to capture exactly this feature of the data: "high-beta assets earn negative and low-beta assets earn positive alphas." We then solve this model analytically, and prove four results, some of which are well-known whereas others are novel. These four results allow us to conclude that working with the orthogonal representation of returns is without loss of generality. That is, from the perspective of mean-variance portfolios, one obtains the same results regardless of whether one starts from the orthogonal or non-orthogonal representation of returns, when the APT assumptions hold and N is large.

IA.1.1 Details of the Model

Set the total number of assets to be equal to $N = 2n + 1$ for some integer $n \geq 0$. Let the return of asset i be given by the market model,

$$r_i - r_f = \alpha_i + \beta_i f + \varepsilon_i,$$

where f denotes the excess return on the market.

Let the betas of the N assets range from $\beta_i = 1 + b(i - (n+1))/n$ for $i = 1, ..., 2n+1$, where $0 \leq b < 1$, with the vector of betas denoted by $\mathbf{B}$. Similarly, let the alphas of the N assets range from $\alpha_i = -a(i - (n+1))/n^{\frac{3}{2}}$ for $i = 1, ..., 2n+1$, with the vector of alphas denoted by $\boldsymbol{\alpha}$. Thus, the β_i are uniformly spread between $1-b$ and $1+b$, and the α_i are uniformly spread between $a/\sqrt{n}$ and $-a/\sqrt{n}$, and the alphas and betas have perfect negative correlation.[i]

[i]For instance, in the setting where $n = 1$ so that there are only $N = 2n + 1 = 3$ risky assets, the betas of the assets would be $\beta_1 = 1 - b$, $\beta_2 = 1$, and $\beta_3 = 1 + b$, and the alphas would be $\alpha_1 = a$, $\alpha_2 = 0$, and $\alpha_3 = -a$.

Let $\lambda = E(f)$ denote the expected market risk premium and $\text{var}(f)$ the variance of the market factor. Let $\varepsilon_i \sim iid(0, \text{var}(\varepsilon_i))$ so that the residuals are iid across assets and across time. Using $\check{}$ to denote orthogonalized quantities and $\mathbf{M}_B$ to denote the projection matrix $\mathbf{I} - \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$ that spans the space orthogonal to the columns of $\mathbf{B}$, we obtain:

$$\begin{aligned}
\check{\boldsymbol{\alpha}} &= \mathbf{M}_B \boldsymbol{\alpha} = \boldsymbol{\alpha} - \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\boldsymbol{\alpha}, \\
\check{\lambda} &= \lambda + (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\boldsymbol{\alpha}, \\
\check{f} &= f - E(f) + \check{\lambda}.
\end{aligned}$$

IA.1.2 Main Results

Result 1. Well-known: Any factor asset-pricing model can be expressed in terms of orthogonal or non-orthogonal components; that is, in terms of the observable returns, the orthogonal and non-orthogonal representations are equivalent, although the components clearly differ in the two representations.

One can express excess returns as

$$\begin{aligned}
\mathbf{r} - r_f \mathbf{1}_N &= \boldsymbol{\alpha} + \mathbf{B}f + \boldsymbol{\varepsilon} \\
&= \boldsymbol{\alpha} + \mathbf{B}\lambda + \mathbf{B}f + \boldsymbol{\varepsilon} \qquad \text{(IA1)} \\
&= \check{\boldsymbol{\alpha}} + \mathbf{B}\check{\lambda} + \mathbf{B}f + \boldsymbol{\varepsilon}. \qquad \text{(IA2)}
\end{aligned}$$

Thus, the same vector of excess returns, $\mathbf{r} - r_f \mathbf{1}_N$, has two equivalent representations: in terms of non-orthogonal alphas and betas, as given in equation (IA1), and in terms of orthogonalized quantities, as given in equation (IA2).

Result 2. Well-known: For any given N, the overall mean-variance portfolio (that is, the linear combination of the alpha and beta portfolios) is identical for the orthogonal and non-orthogonal representations of returns.

To see this, denote the vector of excess mean returns by $\boldsymbol{\mu}$, the variance-covariance matrix for asset returns by $\mathbf{V}$, the variance-covariance matrix for the innovations by $\boldsymbol{\Sigma}$, and define

$$\check{\boldsymbol{\Sigma}}^{-1} = \left[\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{B}(\mathbf{B}'\boldsymbol{\Sigma}^{-1}\mathbf{B})^{-1}\mathbf{B}'\boldsymbol{\Sigma}^{-1}\right].$$

Then, the overall mean-variance weights in the risky assets can be expressed as:

$$\begin{aligned}
\mathbf{w}^{mv} &= \mathbf{V}^{-1}\boldsymbol{\mu} \\
&= \mathbf{V}^{-1}\boldsymbol{\alpha} + \mathbf{V}^{-1}\mathbf{B}\lambda \qquad \text{(IA3)} \\
&= \mathbf{V}^{-1}\check{\boldsymbol{\alpha}} + \mathbf{V}^{-1}\mathbf{B}\check{\lambda} \qquad \text{(IA4)}
\end{aligned}$$

$$= \boldsymbol{\Sigma}^{-1}\check{\boldsymbol{\alpha}} + \mathbf{V}^{-1}\mathbf{B}\check{\lambda} \tag{IA5}$$

$$\approx \check{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\alpha} + \mathbf{V}^{-1}\mathbf{B}\lambda, \quad \text{for large } N. \tag{IA6}$$

This shows that the vector of optimal mean-variance portfolio weights, $\mathbf{w}^{mv}$, can be decomposed into an alpha component and a beta component either in terms of the non-orthogonal expressions for $\boldsymbol{\alpha}$ and λ, as given in equations (IA3) and (IA6), or in terms of the orthogonal expressions, $\check{\boldsymbol{\alpha}}$ and $\check{\lambda}$, as given in equations (IA4) and (IA5).

Result 3. New result: When N is large and the APT no-arbitrage assumption is satisfied, for both the orthogonal and non-orthogonal representations of returns: (i) the alpha and beta portfolio weights are the same (even though the alpha and beta returns differ across the two representations) and (ii) the returns of the alpha and beta portfolios are uncorrelated.

To see this, note that, in terms of the orthogonal decomposition of returns (IA2), when N is large the weights of the alpha and beta portfolios are, respectively:

$$\check{w}_i^{\alpha} = \frac{\check{\alpha}_i}{\mathrm{var}(\varepsilon_i)}; \qquad \check{w}_i^{\beta} = \frac{\beta_i}{(\frac{\mathrm{var}(\varepsilon_i)}{\mathrm{var}(f)} + \mathbf{B}'\mathbf{B})}\frac{\check{\lambda}}{\mathrm{var}(f)} \sim \frac{\beta_i}{2n(1+b^2/3)}\frac{\lambda}{\mathrm{var}(f)}. \tag{IA7}$$

In terms of the non-orthogonal decomposition of returns (IA1), when N is large the weights of the alpha and beta portfolios are, respectively:

$$w_i^{\alpha} = \frac{1}{\mathrm{var}(\varepsilon_i)}(\alpha_i - \beta_i\rho) \sim \check{w}_i^{\alpha}; \qquad w_i^{\beta} = \frac{\beta_i}{(\frac{\mathrm{var}(\varepsilon_i)}{\mathrm{var}(f)} + \mathbf{B}'\mathbf{B})}\frac{\lambda}{\mathrm{var}(f)} \sim \frac{\beta_i}{2n(1+b^2/3)}\frac{\lambda}{\mathrm{var}(f)}, \tag{IA8}$$

where $\rho = \frac{B'\alpha}{\mathrm{var}(\varepsilon_i)/\mathrm{var}(f)+\mathbf{B}'\mathbf{B}} \sim -(ab/\sqrt{n})/(3+b^2)$.

From the above expressions, we see that when N is large the alpha and beta portfolios are the same under the orthogonalized representation of returns (equation (IA7)) and the non-orthogonalized representation of returns (equation (IA8)). It follows that when N is large and the no-arbitrage restriction is satisfied, the returns on the alpha and beta portfolios are orthogonal to one another under both representations.

Result 4. New result: When N is large and the APT no-arbitrage assumption is satisfied, for both the orthogonal and non-orthogonal representations of returns, the Sharpe ratios of the alpha and beta portfolio returns are the same.

When N is large and the no-arbitrage restriction is satisfied, then in terms of the orthogonal decomposition of returns (IA2), the squared Sharpe ratios of the alpha (i.e. $r^{\alpha} - r_f = \mathbf{w}^{\alpha\prime}(\mathbf{r} - r_f\mathbf{1}_N)$) and beta portfolios (i.e. $r^{\beta} - r_f = \mathbf{w}^{\beta\prime}(\mathbf{r} - r_f\mathbf{1}_N)$) are, respectively:

$$(SR_{(\check{r}^{\alpha}-r_f)})^2 \sim \frac{2a^2}{\mathrm{var}(\varepsilon_i)(3+b^2)}; \qquad SR^2_{(\check{r}^{\beta}-r_f)} \sim \frac{\lambda^2}{\mathrm{var}(f)},$$

and in terms of the non-orthogonal decomposition of returns (IA1), the squared Sharpe ratios of the alpha and beta portfolios are, respectively:

$$(SR_{(r^{\alpha}-r_f)})^2 \sim \frac{2a^2}{\text{var}(\varepsilon_i)(3+b^2)}; \qquad (SR_{(r^{\beta}-r_f)})^2 \sim \frac{\lambda^2}{\text{var}(f)}.$$

Therefore, regardless of whether one starts from the orthogonal or non-orthogonal representation of asset returns, when N is large and one imposes the APT no-arbitrage restriction to ensure boundedness, one obtains the same Sharpe ratios for the alpha, beta, and mean-variance portfolios.

IA.1.3 Proofs for the above results

We start by deriving the following quantities, given the specification of the model in terms of uniformly distributed alphas and betas that are perfectly negatively correlated.

$$\begin{aligned}
\mathbf{B}'\mathbf{B} &= \sum_{i=1}^{2n+1} \beta_i^2 \\
&= (2n+1) + \frac{b^2}{n^2}[(2n+1)(n+1)(4n+3)/3 + (n+1)^2(2n+1) - 2(n+1)^2(2n+1)] \\
&= (2n+1) + \frac{b^2(n+1)(2n+1)}{3n} \\
&= (2n+1)(1+b^2(n+1)/3n) \sim 2n(1+b^2/3) \\
\boldsymbol{\alpha}'\boldsymbol{\alpha} &= \frac{a^2(n+1)(2n+1)}{3n^2} \sim \frac{2a^2}{3} \\
\mathbf{B}'\boldsymbol{\alpha} &= -\frac{ab(n+1)(2n+1)}{3n^{\frac{3}{2}}} \sim -\frac{2ab}{3}\sqrt{n} \\
\delta &= \frac{\mathbf{B}'\boldsymbol{\alpha}}{\mathbf{B}'\mathbf{B}} = \frac{-ab(n+1)(2n+1)/3n^{\frac{3}{2}}}{(2n+1)(1+b^2(n+1)/3n)} = \frac{-ab(n+1)/3n^{\frac{3}{2}}}{(1+b^2(n+1)/3n)} \sim -(ab/\sqrt{n})/(3+b^2) \\
\rho &= \frac{\mathbf{B}'\boldsymbol{\alpha}}{\frac{\text{var}(\varepsilon_i)}{\text{var}(f)} + \mathbf{B}'\mathbf{B}} = \frac{-ab(n+1)(2n+1)/3n}{\frac{\text{var}(\varepsilon_i)}{\text{var}(f)} + (2n+1)(1+b^2(n+1)/3n)} \sim \delta \sim -(ab/\sqrt{n})/(3+b^2).
\end{aligned}$$

Based on the above quantities, we define the orthogonalized alpha:

$$\begin{aligned}
\breve{\alpha}_i = \alpha_i - \beta_i\delta &= -a(i-(n+1))/n^{\frac{3}{2}} - [1+b(i-(n+1))/n]\delta \\
&= -\delta - (a/n^{\frac{1}{2}} + \delta b)(i-(n+1))/n.
\end{aligned}$$

So $\breve{\alpha}_i$ goes from the positive $\breve{\alpha}_1 = a/\sqrt{n} - \delta(1-b) \sim (a/\sqrt{n})(3+b)/(3+b^2) > a/\sqrt{n}$ to the negative $\breve{\alpha}_N = -a/\sqrt{n} - \delta(1+b) \sim -(a/\sqrt{n})(3-b)/(3+b^2) > -a/\sqrt{n}$, centred around $0 < -\delta = \breve{\alpha}_{n+1}$.

Then, for the orthogonalized representation, the alpha and beta portfolio weights are:

$$\breve{w}_i^{\alpha} = \frac{\breve{\alpha}_i}{\text{var}(\varepsilon_i)}$$

$$\breve{w}_i^\beta = \frac{\beta_i}{\left(\frac{\text{var}(\varepsilon_i)}{\text{var}(f)} + \mathbf{B}'\mathbf{B}\right)} \frac{\breve{\lambda}}{\text{var}(f)} \sim \frac{\beta_i}{2n(1+b^2/3)} \frac{\lambda}{\text{var}(f)}.$$

For the non-orthogonalized representation, the alpha and beta portfolio weights are:

$$w_i^\alpha = \frac{1}{\text{var}(\varepsilon_i)}(\alpha_i - \beta_i\rho) \sim \breve{w}_i^\alpha$$

$$w_i^\beta = \frac{\beta_i}{\left(\frac{\text{var}(\varepsilon_i)}{\text{var}(f)} + \mathbf{B}'\mathbf{B}\right)} \frac{\lambda}{\text{var}(f)} \sim \frac{\beta_i}{2n(1+b^2/3)} \frac{\lambda}{\text{var}(f)}.$$

Finally, for the orthogonalized representation of returns, one can obtain the squared Sharpe ratio of the beta portfolio as follows:

$$\begin{aligned}
\breve{r}^\beta - r_f &= \breve{\mathbf{w}}^{\beta\prime}(\mathbf{r} - r_f\mathbf{1}_N) \\
&= \frac{\breve{\lambda}}{\text{var}(f)} \frac{\mathbf{B}'\mathbf{B}}{(\text{var}(\varepsilon_i)/\text{var}(f) + \mathbf{B}'\mathbf{B})}(\breve{\lambda} + f - \lambda) + \frac{\breve{\lambda}}{\text{var}(f)} \frac{\mathbf{B}'\varepsilon}{(\text{var}(\varepsilon_i)/\text{var}(f) + \mathbf{B}'\mathbf{B})} \\
E(\breve{r}^\beta - r_f) &= \frac{\breve{\lambda}^2}{\text{var}(f)} \frac{\mathbf{B}'\mathbf{B}}{(\text{var}(\varepsilon_i)/\text{var}(f) + \mathbf{B}'\mathbf{B})} \\
\text{var}(\breve{r}^\beta - r_f) &= \sum_i (\breve{w}_i^\beta)^2 + \text{var}(f)(\sum_i \breve{w}_i^\beta \beta_i)^2 = \frac{\breve{\lambda}^2}{\text{var}(f)} \frac{\mathbf{B}'\mathbf{B}}{(\text{var}(\varepsilon_i)/\text{var}(f) + \mathbf{B}'\mathbf{B})} \\
\text{SR}^2_{(\breve{r}^\beta - r_f)} &= \frac{\breve{\lambda}^2}{\text{var}(f)} \frac{\mathbf{B}'\mathbf{B}}{(\text{var}(\varepsilon_i)/\text{var}(f) + \mathbf{B}'\mathbf{B})} \sim \frac{\breve{\lambda}^2}{\text{var}(f)} \sim \frac{\lambda^2}{\text{var}(f)},
\end{aligned}$$

and of the alpha portfolio as follows:

$$\begin{aligned}
\breve{r}^\alpha - r_f &= \breve{\mathbf{w}}^{\alpha\prime}(\mathbf{r} - r_f\mathbf{1}_N) = \frac{1}{\text{var}(\varepsilon_i)} \sum_i \breve{\alpha}_i^2 + \frac{1}{\text{var}(\varepsilon_i)} \sum_i \breve{\alpha}_i \varepsilon_i \\
E(\breve{r}^\alpha - r_f) &= \frac{1}{\text{var}(\varepsilon_i)} \sum_i \breve{\alpha}_i^2 \sim \frac{2a^2}{\text{var}(\varepsilon_i)(3+b^2)} \\
\text{var}(\breve{r}^\alpha - r_f) &= \frac{1}{\text{var}(\varepsilon_i)} \sum_i \breve{\alpha}_i^2 \sim \frac{2a^2}{\text{var}(\varepsilon_i)(3+b^2)} \\
\text{SR}^2_{(\breve{r}^\alpha - r_f)} &\sim \frac{2a^2}{\text{var}(\varepsilon_i)(3+b^2)}.
\end{aligned}$$

Observe that $\text{cov}(\breve{r}^\beta - r_f, \breve{r}^\alpha - r_f) = 0$ implying $\text{SR}^2_{mv} = \boldsymbol{\mu}'\mathbf{V}^{-1}\boldsymbol{\mu} = \text{SR}^2_{(\breve{r}^\alpha - r_f)} + \text{SR}^2_{(\breve{r}^\beta - r_f)}$.

Similarly, for the non-orthogonalized representation of returns, one can obtain the squared Sharpe ratio of the beta portfolio as follows:

$$\begin{aligned}
r^\beta - r_f &= \mathbf{w}^{\beta\prime}(\mathbf{r} - r_f\mathbf{1}_N) \\
&= \frac{\lambda}{\text{var}(f)} \frac{\boldsymbol{\alpha}'\mathbf{B}}{(\text{var}(\varepsilon_i)/\text{var}(f) + \mathbf{B}'\mathbf{B})}
\end{aligned}$$

$$+\frac{\lambda}{\operatorname{var}(f)}\frac{\mathbf{B}'\mathbf{B}}{(\operatorname{var}(\varepsilon_i)/\operatorname{var}(f)+\mathbf{B}'\mathbf{B})}(\lambda+f-\lambda)+\frac{\lambda}{\operatorname{var}(f)}\frac{\mathbf{B}'\varepsilon}{(\operatorname{var}(\varepsilon_i)/\operatorname{var}(f)+\mathbf{B}'\mathbf{B})}$$

$$E(r^{\beta}-r_f)=\frac{\lambda^2}{\operatorname{var}(f)}\frac{\mathbf{B}'\mathbf{B}}{(\operatorname{var}(\varepsilon_i)/\operatorname{var}(f)+\mathbf{B}'\mathbf{B})}+\frac{\lambda}{\operatorname{var}(f)}\frac{\boldsymbol{\alpha}'\mathbf{B}}{(\operatorname{var}(\varepsilon_i)/\operatorname{var}(f)+\mathbf{B}'\mathbf{B})}$$

$$\sim\frac{\lambda^2}{\operatorname{var}(f)}-\frac{\lambda}{\operatorname{var}(f)}\frac{ab}{(3+b^2)}$$

$$\operatorname{var}(r^{\beta}-r_f)=\sum_i(w_i^{\beta})^2+\operatorname{var}(f)(\sum_i w_i^{\beta}\beta_i)^2=\frac{\lambda^2}{\operatorname{var}(f)}\frac{\mathbf{B}'\mathbf{B}}{(\operatorname{var}(\varepsilon_i)/var(f)+\mathbf{B}'\mathbf{B})}\sim\frac{\lambda^2}{\operatorname{var}(f)}$$

$$\mathrm{SR}^2_{(r^{\beta}-r_f)}=\left(\frac{\lambda}{\sqrt{\operatorname{var}(f)}}\sqrt{\frac{\mathbf{B}'\mathbf{B}}{(\operatorname{var}(\varepsilon_i)/\operatorname{var}(f)+\mathbf{B}'\mathbf{B})}}+\frac{\boldsymbol{\alpha}'\mathbf{B}}{\sqrt{\operatorname{var}(f)}\sqrt{\mathbf{B}'\mathbf{B}}\sqrt{(\operatorname{var}(\varepsilon_i)/\operatorname{var}(f)+\mathbf{B}'\mathbf{B})}}\right)^2$$

$$\sim\left(\frac{\lambda}{\sqrt{\operatorname{var}(f)}}-\frac{1}{\sqrt{var(f)}}\frac{(ab/\sqrt{n})}{(3+b^2)}\right)^2\sim\frac{\lambda^2}{\operatorname{var}(f)},$$

and of the alpha portfolio as

$$r^{\alpha}-r_f=\mathbf{w}^{\alpha\prime}(\mathbf{r}-r_f\mathbf{1}_N)=\sum_i w_i^{\alpha}\alpha_i+(\sum_i w_i^{\alpha}\beta_i)(\lambda+f-\lambda)+\sum_i w_i^{\alpha}\varepsilon_i$$

$$E(r^{\alpha}-r_f)=\sum_i\alpha_i w_i^{\alpha_i}+\lambda\sum_i w_i^{\alpha}\beta_i\sim\frac{2a^2}{\operatorname{var}(\varepsilon_i)(3+b^2)}$$

$$\operatorname{var}(r^{\alpha}-r_f)=\operatorname{var}(\varepsilon_i)\sum_i(w_i^{\alpha})^2+\operatorname{var}(f)(\sum_i w_i^{\alpha}\beta_i)^2\sim\frac{2a^2}{\operatorname{var}(\varepsilon_i)(3+b^2)}$$

$$\mathrm{SR}^2_{(r^{\alpha}-r_f)}\sim\frac{2a^2}{\operatorname{var}(\varepsilon_i)(3+b^2)}.$$

IA.2 Auxiliary Results

In this section, we provide some auxiliary results and extend the result in Roll (1980) to the case in which investors can invest also in a risk-free asset.

IA.2.1 Decomposition of the Sharpe Ratio

Proposition IA.1 (Portfolio Sharpe ratio)**.** *Consider the portfolio weights* $\mathbf{w}_N=\mathbf{w}_{N,1}+\mathbf{w}_{N,2}$ *such that* $\mathbf{w}_{N,1}$ *is orthogonal to* $\mathbf{w}_{N,2}$*, as follows*

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

Then, defining $\mu_i - r_{ft} = \mathbf{w}'_{N,i}(\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N)$, $\sigma_i^2 = \mathbf{w}'_{N,i}\mathbf{V}_N\mathbf{w}_{N,i}$, *and* $\mathrm{SR}_i = \mathbf{w}'_{N,i}(\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N)/(\mathbf{w}'_{N,i}\mathbf{V}_N\mathbf{w}_{N,i})^{1/2}$, *for* $i = 1, 2$, *and letting* SR *denote the Sharpe ratio of the portfolio* $\mathbf{w}_N$, *we always have*

$$\begin{aligned}
\mathrm{SR}^2 &= \frac{(\mathbf{w}'_N(\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N))^2}{\mathbf{w}'_N\mathbf{V}_N\mathbf{w}_N} \\
&= (\mathrm{SR}_1)^2 + (\mathrm{SR}_2)^2 - \frac{1}{\mathbf{w}'_N\mathbf{V}_N\mathbf{w}_N}(\sigma_2\mathrm{SR}_1 - \sigma_1\mathrm{SR}_2)^2 \\
&\leq (\mathrm{SR}_1)^2 + (\mathrm{SR}_2)^2.
\end{aligned}$$

Finally, strict equality holds if and only if $\sigma_2\mathrm{SR}_1 - \sigma_1\mathrm{SR}_2 = 0$*; that is,*

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

Proof. We have

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

Using the orthogonality of $\mathbf{w}_{N,1}$ and $\mathbf{w}_{N,2}$, we have $\mathbf{w}'_N\mathbf{V}_N\mathbf{w}_N = \mathbf{w}'_{N,1}\mathbf{V}_N\mathbf{w}_{N,1} + \mathbf{w}'_{N,2}\mathbf{V}_N\mathbf{w}_{N,2} = \sigma_1^2 + \sigma_2^2$, so that the term in square brackets can we rewritten as

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

Hence,

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

Equality holds if and only if

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

which, in turn, can be rearranged as $\frac{(\mu_1 - r_{ft})}{\sigma_1^2} = \frac{(\mu_2 - r_{ft})}{\sigma_2^2}$. ■

This result clarifies that when a portfolio can be decomposed into two orthogonal components, that is

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

then $\mathrm{SR}^2 \leq (\mathrm{SR}_1)^2 + (\mathrm{SR}_2)^2$. However, it does not always follows that $\mathrm{SR}^2 = (\mathrm{SR}_1)^2 + (\mathrm{SR}_2)^2$, unless the additional condition $\sigma_2 \mathrm{SR}_1 - \sigma_1 \mathrm{SR}_2 = 0$ holds.

IA.2.2 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, the number of zero-beta portfolios is infinite—one for each level of the 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 obtain our theoretical results, we extend the result in Roll (1980) to the case in which investors can invest also in a risk-free asset.

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

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

and

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

for a given target mean μ^z. Then

$$\mathbf{w}_N^z = \Big(\mathbf{w}_N^x, \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_{ft} \mathbf{1}_N) \Big) \begin{pmatrix} (\sigma^x)^2 & \mu^x - r_{ft} \\ \mu^x - r_{ft} & (\mathrm{SR}^{mv})^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \mu^z - r_{ft} \end{pmatrix},$$

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

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

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

with first-order conditions

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

Premultiplying both sides by $2^{-1}\Big(\mathbf{V}_N\mathbf{w}_N^x, (\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N)\Big)'\mathbf{V}_N^{-1}$ gives

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

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

Observe that when $\mathbf{w}_N^x$ is efficient, then $\mathbf{w}_N^z = \mathbf{0}_N$, which implies that the zero-beta portfolio to $\mathbf{w}_N^x$ is the portfolio that invests 100% in the risk-free asset. In fact, substituting $\mathbf{w}_N^x = \gamma^{-1}\mathbf{V}_N^{-1}(\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N)$ into the first constraint gives $0 = (\mathbf{w}_N^x)'\mathbf{V}_N\mathbf{w}_N^z = \gamma^{-1}(\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N)'\mathbf{w}_N^z = \gamma^{-1}(\mu_z - r_{ft})$, where the last equality is due to the second constraint. Therefore, one obtains $\mu_z = r_{ft}$ which, by no-arbitrage, implies $\mathbf{w}_N^z = \mathbf{0}_N$.

Recall the well-known result that the entire efficient frontier of risky assets can be generated from holding any two efficient portfolios. However, one can show that the efficient frontier of risky assets can also be generated by holding two *inefficient* portfolios, as long as one is the minimum-variance orthogonal portfolio of the other, which leads to the following:

Proposition IA.3 (Extension of Corollary 3 of Roll (1980) to the case with a risk-free asset)**.** *There is a weighted average of, possibly inefficient, portfolio $\mathbf{w}_N^x$ with a corresponding minimum-variance orthogonal portfolio $\mathbf{w}_N^z$, which produces an efficient portfolio.*

The above proposition implies that the subspace of minimum-variance portfolios (possibly inefficient) orthogonal to $\mathbf{w}_N^x$ is given by the two lines described by the expression below:

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

Notice from the equation above and the dashed and dotted lines in Figure 1 that the slopes of the two lines are smaller (in absolute value) than the slopes of the capital market lines.[ii] 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 $\mathbf{w}_N^x$ equal to the Sharpe ratio of the mean-variance portfolio $\mathbf{w}_N^{\mathrm{mv}}$ in the equation above.

IA.3 Estimating the Approximating Model

We describe briefly how to estimate the APT model, which represents the approximating model considered by the misspecification-averse investor, by maximum likelihood. In the first part of this

[ii] Huang and Litzenberger (1988) show that, depending on the level of the risk-free rate relative to the mean of the global-minimum-variance portfolio, the capital market line could be sloping up or down.

section, we consider estimation of the APT model with only asset-specific pricing errors and with only tradable candidate factors, providing the main intuition for our estimator. In the second part, we consider estimation of the general model when pricing errors associated omitted risk factors are also present, and with both tradable and nontradable observed risk factors.

IA.3.1 Estimation of APT: Case of Asset-Specific Pricing Errors

In this section, we consider the case of asset-specific only pricing errors, i.e. pricing errors unrelated to omitted risk factors ($p = 0$), which we estimate using (pseudo) maximum likelihood (ML). The (pseudo) ML is a 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 except for efficiency. The (pseudo) ML estimator maximizes the unconditional joint distribution of $\begin{pmatrix} \mathbf{r}_t^e \\ \mathbf{f}_t^e \end{pmatrix} = \begin{pmatrix} \mathbf{r}_t \\ \mathbf{f}_t \end{pmatrix} - r_{ft}\mathbf{1}_{N+K}$ which, assuming i.i.d. error term $\boldsymbol{\varepsilon}_t$ for simplicity, equals[iii]

$$L(\tilde{\boldsymbol{\theta}}) = -\frac{1}{2}\log(\det(\tilde{\boldsymbol{\Sigma}}_N)) - \frac{1}{2T}\sum_{t=1}^{T}\left(\mathbf{r}_t^e - \tilde{\boldsymbol{\alpha}}_N - \tilde{\mathbf{B}}_N\mathbf{f}_t^e\right)'\tilde{\boldsymbol{\Sigma}}_N^{-1}\left(\mathbf{r}_t^e - \tilde{\boldsymbol{\alpha}}_N - \tilde{\mathbf{B}}_N\mathbf{f}_t^e\right)$$
$$-\frac{1}{2}\log(\det(\tilde{\boldsymbol{\Omega}})) - \frac{1}{2T}\sum_{t=1}^{T}\left(\mathbf{f}_t^e - \tilde{\boldsymbol{\lambda}}\right)'\tilde{\boldsymbol{\Omega}}^{-1}\left(\mathbf{f}_t^e - \tilde{\boldsymbol{\lambda}}\right), \qquad \text{(IA9)}$$

where $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\alpha}}_N', \text{vec}(\tilde{\mathbf{B}}_N)', \text{vech}(\tilde{\boldsymbol{\Sigma}}_N)', \tilde{\boldsymbol{\lambda}}', \text{vech}(\tilde{\boldsymbol{\Omega}})')'$.[iv] Therefore, the ML estimators for $\boldsymbol{\alpha}_N, \mathbf{B}_N$ and $\boldsymbol{\Sigma}_N$ coincide with the OLS estimators denoted here by $\hat{\boldsymbol{\theta}}$, conditional on the realization of the factors. The ML estimators for $\boldsymbol{\lambda}$ and $\boldsymbol{\Omega}$ are the sample mean and covariance of the factors $\mathbf{f}_t^e$.

However, because the APT restriction is not guaranteed to hold for an arbitrary δ_{apt}, one should consider the maximum-likelihood estimator subject to this restriction. Moreover, with the parameter $\boldsymbol{\alpha}_N$ constrained by the APT restriction, imposing such a constraint may lead to a more precise estimator of the true parameter values compared to the *unconstrained* estimator $\hat{\boldsymbol{\theta}}$. In the general case presented in the next section, , where one allows for both omitted pervasive risk and asset-specific risk, imposing the APT restriction leads to *identification* of the APT parameters, in particular for $\mathbf{a}_N$ and $\boldsymbol{\lambda}_{\text{miss}}$.

[iii] Notice that we have expressed the joint distribution as the product of a conditional distribution and a marginal distribution. Relaxing the i.i.d. assumption requires specification of time-varying conditional means, conditional variances, and conditional covariances.

[iv] Here det(·) denotes the determinant, vec(·) denotes the operator that stacks the columns of a matrix into a single column vector, and vech(·) denotes the operator that stacks the unique elements of the columns of a symmetric matrix into a single column vector.

Proposition IA.1 (Parameter estimation by imposing asset-pricing restriction: Case for asset-specific pricing error)**.** *Suppose that the vector of asset returns,* $\mathbf{r}_t$*, satisfies Assumption 3.1. Then*

$$\hat{\boldsymbol{\theta}}_{MLC} = \underset{\tilde{\boldsymbol{\theta}}}{\text{argmax}} \quad L(\tilde{\boldsymbol{\theta}}) \quad \textit{subject to} \quad \tilde{\boldsymbol{\alpha}}_N' \tilde{\boldsymbol{\Sigma}}_N^{-1} \tilde{\boldsymbol{\alpha}}_N \le \delta_{apt},$$

where $L(\tilde{\boldsymbol{\theta}})$ *is defined in* (IA9)*. If* $\left(\sum_{t=1}^T \dot{\mathbf{f}}_t \dot{\mathbf{f}}_t'\right)$ *is nonsingular, then* $\hat{\boldsymbol{\theta}}_{MLC} = (\hat{\boldsymbol{\alpha}}_{N,MLC}', vec(\hat{\mathbf{B}}_{N,MLC})', vech(\hat{\boldsymbol{\Sigma}}_{N,MLC})', \hat{\boldsymbol{\lambda}}_{MLC}', vech(\hat{\boldsymbol{\Omega}}_{MLC})')'$ *exists, where*

$$\hat{\boldsymbol{\alpha}}_{N,MLC} = \frac{1}{1+\hat{\kappa}_{MLC}} \Big[\bar{\mathbf{r}}_N^e - \hat{\mathbf{B}}_{N,MLC} \bar{\mathbf{f}}^e \Big], \tag{IA10}$$

$$\hat{\mathbf{B}}_{N,MLC} = \Big(\sum_{t=1}^T \dot{\mathbf{r}}_t \dot{\mathbf{f}}_t' \Big) \Big(\sum_{t=1}^T \dot{\mathbf{f}}_t \dot{\mathbf{f}}_t' \Big)^{-1}, \quad \textit{and}$$

$$\hat{\boldsymbol{\Sigma}}_{N,MLC} = \frac{1}{T} \sum_{t=1}^T (\dot{\mathbf{r}}_t - \hat{\mathbf{B}}_{N,MLC} \dot{\mathbf{f}}_t)(\dot{\mathbf{r}}_t - \hat{\mathbf{B}}_{N,MLC} \dot{\mathbf{f}}_t)',$$

in which $\hat{\kappa}_{MLC} \ge 0$ *is the optimal value of the Karush-Kuhn-Tucker multiplier,* $\bar{\mathbf{f}}^e = T^{-1} \sum_{t=1}^T \mathbf{f}_t^e, \bar{\mathbf{r}}_N^e = T^{-1} \sum_{t=1}^T \mathbf{r}_t^e$*,* $\dot{\mathbf{f}}_t = \mathbf{f}_t^e - \frac{1}{(1+\hat{\kappa}_{MLC})} \bar{\mathbf{f}}^e$*,* $\dot{\mathbf{r}}_t = \mathbf{r}_t^e - \frac{1}{(1+\hat{\kappa}_{MLC})} \bar{\mathbf{r}}_N^e$*, and the MLC estimators,* $\hat{\boldsymbol{\lambda}}_{MLC}$ *and* $vech(\hat{\boldsymbol{\Omega}}_{MLC})$*, coincide with the sample mean and sample covariance matrix of the factors* $\mathbf{f}_t$*.*

Proof. 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 associated with the Lagrangian problem:

$$\{\hat{\boldsymbol{\theta}}_{MLC}, \hat{\kappa}\} = \underset{\tilde{\boldsymbol{\theta}}}{\text{argmax}} \ \underset{\tilde{\kappa} \ge 0}{\text{argmax}} \ L(\tilde{\boldsymbol{\theta}}) - \tilde{\kappa} (\tilde{\boldsymbol{\alpha}}_N' \tilde{\boldsymbol{\Sigma}}_N^{-1} \tilde{\boldsymbol{\alpha}}_N - \delta_{\text{apt}}).$$

Start with $\hat{\kappa}_{MLC} = 0$. Then the MLC estimator for $\boldsymbol{\theta}$ coincides with the OLS estimator $\hat{\boldsymbol{\theta}}$, readily obtained by setting $\hat{\kappa}_{MLC} = 0$ in the above formulae for $\hat{\boldsymbol{\theta}}_{\text{MLC}}$, and one needs to evaluate whether $\hat{\boldsymbol{\alpha}}_N' \hat{\boldsymbol{\Sigma}}_N^{-1} \hat{\boldsymbol{\alpha}}_N > \delta_{\text{apt}}$. If the latter inequality holds, $\hat{\kappa}_{MLC} = 0$ violates the complementary slackness condition and can be ruled out. Alternatively, when $\hat{\boldsymbol{\alpha}}_N' \hat{\boldsymbol{\Sigma}}_N^{-1} \hat{\boldsymbol{\alpha}}_N \le \delta_{\text{apt}}$, then we evaluate $L(\hat{\boldsymbol{\theta}})$ and then consider the case in which $\hat{\kappa}_{MLC} > 0$. In particular, by solving the first-order equation for $\hat{\boldsymbol{\alpha}}_{N,\text{MLC}}$ and $\hat{\kappa}_{MLC}$ sequentially, one gets:

$$(1+\hat{\kappa}_{MLC})^2 = \frac{\Big[\bar{\mathbf{r}}_N^e - \hat{\mathbf{B}}_{N,\text{MLC}} \bar{\mathbf{f}}^e \Big]' \hat{\boldsymbol{\Sigma}}_{N,\text{MLC}}^{-1} \Big[\bar{\mathbf{r}}_N^e - \hat{\mathbf{B}}_{N,\text{MLC}} \bar{\mathbf{f}}^e \Big]}{\delta_{\text{apt}}}, \tag{IA11}$$

and now the case where $\hat{\kappa}_{MLC} > 0$ is feasible only when the right hand side of the above equation is bigger than one. When this occurs, then one evaluates $L(\hat{\boldsymbol{\theta}}_{\text{MLC}})$. Note that, by (IA10) and (IA11), $\hat{\boldsymbol{\alpha}}_{N,\text{MLC}}$ will satisfy the constraint exactly, that is $\hat{\boldsymbol{\alpha}}_{N,\text{MLC}}' \hat{\boldsymbol{\Sigma}}_{N,\text{MLC}}^{-1} \hat{\boldsymbol{\alpha}}_{N,\text{MLC}} = \delta_{\text{apt}}$, by construction. Incidentally, note that the estimates of $\hat{\mathbf{B}}_{N,\text{MLC}}$ and $\hat{\boldsymbol{\Sigma}}_{N,\text{MLC}}$ are function of $\hat{\kappa}_{MLC}$ so a (simple) iterative procedure is required. Alternatively, if the right hand of the above equation side is smaller than one, then the case $\hat{\kappa}_{MLC} > 0$ is ruled out by the complementary slackness condition, and one retains the OLS estimator $\hat{\boldsymbol{\theta}}$. Finally, when both cases $\hat{\kappa}_{MLC} = 0$ and $\hat{\kappa}_{MLC} > 0$ are feasible, one simply needs to compare $L(\hat{\boldsymbol{\theta}})$ with $L(\hat{\boldsymbol{\theta}}_{\text{MLC}})$, and select the corresponding estimate (either $\hat{\boldsymbol{\theta}}$ or $\hat{\boldsymbol{\theta}}_{\text{MLC}}$) that maximizes the log-likelihood function. However, in most cases, either $\hat{\kappa}_{MLC} = 0$ or $\hat{\kappa}_{MLC} > 0$ is feasible, but not both, simplifying the solution of the Karush-Kuhn-Tucker problem.

Finally, for $\hat{\boldsymbol{\lambda}}_{\text{MLC}}$ and $\hat{\boldsymbol{\Omega}}_{\text{MLC}}$, one obtains precisely the sample mean and sample covariance matrix of $\mathbf{f}^e_t$. ■

IA.3.2 Estimation of APT: The General Case

We now explain how to estimate the APT allowing for both tradable and nontradable factors, and for both asset-specific pricing errors and pricing errors arising from omitting pervasive risk factors. Assume that

$$\mathbf{r}^e_{t+1} = \boldsymbol{\alpha}_N + \mathbf{B}_{1N}(\boldsymbol{\lambda}_1 + \mathbf{f}_{1t+1} - E(\mathbf{f}_{1t+1})) + \mathbf{B}_{2N}\mathbf{f}^e_{2t+1} + \boldsymbol{\varepsilon}_{t+1}, \text{ with}$$
$$\boldsymbol{\alpha}_N = \mathbf{a}_N + \mathbf{A}_N\boldsymbol{\lambda}_{\text{miss}} \quad \text{and} \quad \text{var}(\boldsymbol{\varepsilon}_{t+1}) = \mathbf{A}_N\mathbf{A}'_N + \mathbf{C}_N,$$

where we set $\mathbf{B}_N = (\mathbf{B}_{1N}, \mathbf{B}_{2N})$, $\boldsymbol{\Omega} = \text{var}(\mathbf{f}_{t+1})$, $\mathbf{f}_{t+1} = (\mathbf{f}'_{1t+1}, \mathbf{f}^{e\prime}_{2t+1})'$, with $\mathbf{f}_{1t+1}$ denoting the set of K_1 nontradable observed factors and $\mathbf{f}^e_{2t+1}$ the set of K_2 tradable observed factors, expressed as excess returns, where $K = K_1 + K_2$. We assume that the missing factors are uncorrelated with the observed factors.[v] Given that $\mathbf{f}^e_{2t}$ are excess returns on tradable assets, their risk premia satisfy $\boldsymbol{\lambda}_2 = E(\mathbf{f}^e_{2t})$ and, to avoid confusion with the risk premia of the nontradable assets $\boldsymbol{\lambda}_1$, we will use the expectation formulation for $\boldsymbol{\lambda}_2$.

The joint log-likelihood function takes the following form:

$$\begin{aligned} L(\tilde{\boldsymbol{\theta}}) = &-\frac{1}{2}\log(\det(\tilde{\mathbf{A}}_N\tilde{\mathbf{A}}'_N + \tilde{\mathbf{C}}_N)) \\ &- \frac{1}{2T}\sum_{t=1}^{T}\Big(\mathbf{r}^e_t - \tilde{\mathbf{A}}_N\tilde{\boldsymbol{\lambda}}_{miss} - \tilde{\mathbf{a}}_N - \tilde{\mathbf{B}}_{1N}(\tilde{\boldsymbol{\lambda}}_1 + \mathbf{f}_{1t} - \tilde{E}(\mathbf{f}_{1t})) - \tilde{\mathbf{B}}_{2N}\mathbf{f}^e_{2t}\Big)' \\ &\qquad \times (\tilde{\mathbf{A}}_N\tilde{\mathbf{A}}'_N + \tilde{\mathbf{C}}_N)^{-1}\Big(\mathbf{r}^e_t - \tilde{\mathbf{A}}_N\tilde{\boldsymbol{\lambda}}_{miss} - \tilde{\mathbf{a}}_N - \tilde{\mathbf{B}}_{1N}(\tilde{\boldsymbol{\lambda}}_1 + \mathbf{f}_{1t} - \tilde{E}(\mathbf{f}_{1t})) - \tilde{\mathbf{B}}_{2N}\mathbf{f}^e_{2t}\Big) \\ &- \frac{1}{2}\log(\det(\tilde{\boldsymbol{\Omega}})) - \frac{1}{2T}\sum_{t=1}^{T}(\mathbf{f}_t - \tilde{E}(\mathbf{f}_t))'\tilde{\boldsymbol{\Omega}}^{-1}(\mathbf{f}_t - \tilde{E}(\mathbf{f}_t)). \end{aligned} \tag{IA12}$$

Without loss of generality, one can assume that the missing factors have unit variance, achieving identification of $\mathbf{A}_N$. However, it turns out that $\boldsymbol{\lambda}_{\text{miss}}$ and $\boldsymbol{\alpha}_N$ *cannot* be identified separately unless the APT restriction is imposed, as shown in Proposition IA.2 below.

[v] The estimator can be extended to the case of correlated observed and omitted risk factors; details are available upon request.

Proposition IA.2 (Parameter estimation of APT: General Case)**.** *Suppose that the vector of asset returns,* $\mathbf{r}_t$*, satisfies Assumption 3.1 and that* $\mathbf{\Sigma}_{f_2^e f_2^e} - \bar{\mathbf{f}}_2^{\,e}\bar{\mathbf{f}}_2^{\,e\prime}$ *is nonsingular, where* $\mathbf{\Sigma}_{f_2^e f_2^e} = T^{-1}\sum_{t=1}^{T}\mathbf{f}_{2t}^e\mathbf{f}_{2t}^{e\prime}$ *and* $\bar{\mathbf{f}}_2^{\,e} = T^{-1}\sum_{t=1}^{T}\mathbf{f}_{2t}^e$*. Then*

$$\hat{\boldsymbol{\theta}}_{MLC} = \underset{\tilde{\boldsymbol{\theta}}}{\operatorname{argmax}} \quad L(\tilde{\boldsymbol{\theta}}) \quad \text{subject to} \quad \tilde{\mathbf{a}}_N'\tilde{\mathbf{\Sigma}}_N^{-1}\tilde{\mathbf{a}}_N \leq \delta_{apt},$$

where $L(\tilde{\boldsymbol{\theta}})$ *is defined in (IA12), and* $\hat{\boldsymbol{\theta}}_{MLC} = (\hat{\mathbf{a}}_{N,MLC}', \hat{\boldsymbol{\lambda}}_{miss,MLC}', \hat{\boldsymbol{\lambda}}_{1,MLC}', \hat{E}(\mathbf{f}_{1t})_{MLC}', \hat{E}(\mathbf{f}_{2t}^e)_{MLC}', vec(\hat{\mathbf{A}}_{N,MLC})', vec(\hat{\mathbf{B}}_{N,MLC})', vech(\hat{\mathbf{C}}_{N,MLC})', vech(\hat{\mathbf{\Omega}}_{MLC})')'$.

(i) If the optimal value of the Karush-Kuhn-Tucker multiplier satisfies $\hat{\kappa}_{MLC} > 0$*, setting*

$$\mathbf{D}_N = (\mathbf{A}_N, \mathbf{B}_{1N}), \;\; \boldsymbol{\lambda} = (\boldsymbol{\lambda}_{miss}', \boldsymbol{\lambda}_1')',$$

then, using $\otimes$ *to denote the Kronecker product,*

$$vec(\hat{\mathbf{B}}_{2N,MLC}) = \Big((\mathbf{\Sigma}_{f_2^e f_2^e} \otimes \mathbf{I}_N) - (\bar{\mathbf{f}}_2^e\bar{\mathbf{f}}_2^{e\prime} \otimes \hat{\mathbf{G}}_N)\Big)^{-1} vec\Big(\mathbf{\Sigma}_{hf_2^e} - \hat{\mathbf{G}}_N\bar{\mathbf{h}}_N\bar{\mathbf{f}}_2^{e\prime}\Big), \tag{IA13}$$

$$\hat{\boldsymbol{\lambda}}_{MLC} = (\hat{\mathbf{D}}_{N,MLC}'\hat{\mathbf{\Sigma}}_{N,MLC}^{-1}\hat{\mathbf{D}}_{N,MLC})^{-1}\hat{\mathbf{D}}_{N,MLC}'\hat{\mathbf{\Sigma}}_{N,MLC}^{-1}\Big(\bar{\mathbf{h}}_N - \hat{\mathbf{B}}_{2N,MLC}\,\bar{\mathbf{f}}_2^e\Big),$$

$$\hat{\mathbf{a}}_{N,MLC} = \frac{1}{\hat{\kappa}_{MLC}+1}\Big(\bar{\mathbf{h}}_N - \hat{\mathbf{B}}_{2N,MLC}\bar{\mathbf{f}}_2^e - \hat{\mathbf{D}}_{N,MLC}\hat{\boldsymbol{\lambda}}_{MLC}\Big),$$

where $\hat{\mathbf{\Sigma}}_{N,MLC} = \hat{\mathbf{A}}_{N,MLC}\hat{\mathbf{A}}_{N,MLC}' + \hat{\mathbf{C}}_{N,MLC}$*,* $\mathbf{\Sigma}_{hf_2^e} = T^{-1}\sum_{t=1}^{T}\mathbf{h}_t\mathbf{f}_{2t}^{e\prime}$*,* $\bar{\mathbf{h}}_N = T^{-1}\sum_{t=1}^{T}\mathbf{h}_t$ *with* $\mathbf{h}_t = \mathbf{r}_t^e - \hat{\mathbf{B}}_{1N,MLC}(\mathbf{f}_{1t} - \bar{\mathbf{f}}_1)$ *and* $\bar{\mathbf{f}}_1 = T^{-1}\sum_{t=1}^{T}\mathbf{f}_{1t}$*, and*

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

Note that $\hat{\mathbf{D}}_{N,MLC} = (\hat{\mathbf{A}}_{N,MLC}, \hat{\mathbf{B}}_{1N,MLC})$ *and* $\hat{\mathbf{C}}_{N,MLC}$ *do not admit a closed-form solution and, as before,* $\hat{E}(\mathbf{f}_t)_{MLC}$ *and* $\hat{\mathbf{\Omega}}_{MLC}$ *coincide with the sample mean and sample covariance of the observed factors* $\mathbf{f}_t$*.*

(ii) If the optimal value of the Karush-Kuhn-Tucker multiplier satisfies $\hat{\kappa}_{MLC} = 0$ *one can estimate only* $\boldsymbol{\alpha}_N = \mathbf{a}_N + \mathbf{D}_N\boldsymbol{\lambda}$ *but not the two components separately, and one obtains*

$$\hat{\boldsymbol{\alpha}}_{N,MLC} = \bar{\mathbf{r}}_N^e - \hat{\mathbf{B}}_{2N,MLC}\bar{\mathbf{f}}_2^e,$$

and the expression for $vec(\hat{\mathbf{B}}_{2N,MLC})$ *can be obtained by setting* $\hat{\kappa}_{MLC} = 0$ *in the terms that appear in (IA13). The expressions for* $\hat{E}(\mathbf{f}_t)_{MLC}$ *and* $\hat{\mathbf{\Omega}}_{MLC}$ *are unchanged, and, as before, the expressions for the estimators of* $\hat{\mathbf{D}}_{N,MLC}$ *and* $\hat{\mathbf{C}}_{N,MLC}$ *do not admit a closed-form solution.*

Proof. Within this proof, for simplicity, we do not use the $\tilde{}$ notation to denote feasible parameter values.

Defining by $\hat{\boldsymbol{\theta}}$ the MLC corresponding to $\hat{\kappa}=0$, this is unfeasible whenever we have that $\hat{\mathbf{a}}_N'\hat{\boldsymbol{\Sigma}}_N^{-1}\hat{\mathbf{a}}_N > \delta$. Similarly, case $\hat{\kappa}_{MLC}>0$ is unfeasible whenever,

$$\left(\bar{\mathbf{r}}_N^e - \hat{\mathbf{B}}_{2N,\text{MLC}}\,\bar{\mathbf{f}}_2^e - \hat{\mathbf{D}}_{N,\text{MLC}}\,\hat{\boldsymbol{\lambda}}_{MLC}\right)'\hat{\boldsymbol{\Sigma}}_{N,\text{MLC}}^{-1}\left(\bar{\mathbf{r}}_N^e - \hat{\mathbf{B}}_{2N,\text{MLC}}\,\bar{\mathbf{f}}_2^e - \hat{\mathbf{D}}_{N,\text{MLC}}\,\hat{\boldsymbol{\lambda}}_{MLC}\right) < \delta_{apt},$$

because $(1+\hat{\kappa}_{MLC})^2 = \frac{\left[\bar{\mathbf{r}}_N^e - \hat{\mathbf{B}}_{2N,\text{MLC}}\,\bar{\mathbf{f}}_2^e - \hat{\mathbf{D}}_{N,\text{MLC}}\,\hat{\boldsymbol{\lambda}}_{MLC}\right]'\hat{\boldsymbol{\Sigma}}_{N,\text{MLC}}^{-1}\left[\bar{\mathbf{r}}_N^e - \hat{\mathbf{B}}_{2N,\text{MLC}}\,\bar{\mathbf{f}}_2^e - \hat{\mathbf{D}}_{N,\text{MLC}}\,\hat{\boldsymbol{\lambda}}_{MLC}\right]}{\delta_{apt}}$. When both cases are feasible, the optimal value for the Karush-Kuhn-Tucker multiplier will be greater than zero or equal to zero, depending on which case maximizes the log-likelihood, namely depending on whether $L(\hat{\boldsymbol{\theta}}_{\text{MLC}})$ or $L(\hat{\boldsymbol{\theta}})$ is largest, respectively. Note that when $\kappa_{MLC}>0$ then $\hat{\mathbf{a}}_{N,\text{MLC}}'\hat{\boldsymbol{\Sigma}}_{N,\text{MLC}}^{-1}\hat{\mathbf{a}}_{N,\text{MLC}} = \delta_{apt}$ by construction.

We now derive the formulae for the estimators. Assume for now that case $\hat{\kappa}>0$ holds. Differentiating the penalized log-likelihood with respect to $\boldsymbol{\lambda}$, $\mathbf{a}_N$, and the Lagrange multiplier κ, the first K^*+N equations, setting $K^*=p+K_1$, (after some algebra) are:

$$\begin{pmatrix} \mathbf{D}_N{}'\boldsymbol{\Sigma}_N^{-1} \\ \mathbf{I}_N \end{pmatrix}\left(\bar{\mathbf{r}}_N^e - \mathbf{B}_{2N}\bar{\mathbf{f}}_2^e\right) = \begin{pmatrix} \mathbf{D}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{D}_N & \mathbf{D}_N'\boldsymbol{\Sigma}_N^{-1} \\ \mathbf{D}_N & (1+\hat{\kappa}_{MLC})\mathbf{I}_N \end{pmatrix}\begin{pmatrix} \hat{\boldsymbol{\lambda}}_{\text{MLC,MLC}} \\ \hat{\mathbf{a}}_{N,\text{MLC}} \end{pmatrix},$$

where recall that $\boldsymbol{\Sigma}_N = \mathbf{A}_N\mathbf{A}_N' + \mathbf{C}_N$, and noting that all the expressions above and below are left as function of the feasible values for $\mathbf{C}_N$ and $\mathbf{D}_N$ (as opposed to their MLC values). It is straightforward to see that, because of the APT restriction, $\boldsymbol{\lambda}$ and $\mathbf{a}_N$ can now be identified separately, as long as $\hat{\kappa}_{MLC}>0$. In fact, the above system of linear equations can be solved because the matrix pre-multiplying $\hat{\boldsymbol{\lambda}}_{\text{MLC}}$ and $\hat{\mathbf{a}}_{N,\text{MLC}}$ is non-singular for every $\hat{\kappa}_{MLC}>0$, leading to the closed-form solution:

$$\hat{\boldsymbol{\lambda}}_{\text{MLC}} = (\mathbf{D}_N'\boldsymbol{\Sigma}_N^{-1}\mathbf{D}_N)^{-1}\mathbf{D}_N'\boldsymbol{\Sigma}_N^{-1}\left(\bar{\mathbf{r}}_N^e - \mathbf{B}_{2N,}\bar{\mathbf{f}}_2^e\right), \tag{IA14}$$

$$\hat{\mathbf{a}}_{N,\text{MLC}} = \frac{1}{\hat{\kappa}_{MLC}+1}\left(\bar{\mathbf{r}}_N^e - \mathbf{B}_{2N}\bar{\mathbf{f}}_2^e - \mathbf{D}_N\hat{\boldsymbol{\lambda}}\right). \tag{IA15}$$

Turning now to the first-order condition with respect to the generic (a,b)th element of $\mathbf{B}_{2N,\text{MLC}}$, denoted by B_{2ab} with $1\le a\le N,\ 1\le b\le K_2$, one obtains

$$-\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{r}_t^e - \mathbf{A}_N\boldsymbol{\lambda}_{miss} - \mathbf{a}_N - \mathbf{B}_{1N}(\boldsymbol{\lambda}_1 + \mathbf{f}_{1t} - \bar{\mathbf{f}}_1) - \hat{\mathbf{B}}_{2N,\text{MLC}}\mathbf{f}_{2t}^e\right)'\boldsymbol{\Sigma}_N^{-1}\left(-\frac{\partial\mathbf{B}_{2N,\text{MLC}}}{\partial B_{2ab}}\,\mathbf{f}_{2t}^e\right) = 0,$$

which can be re-arranged as

$$\Sigma_{r^e f_2^e} - (\mathbf{a}_N + \mathbf{A}_N\boldsymbol{\lambda}_{miss} + \mathbf{B}_{1N}\boldsymbol{\lambda}_1)\bar{\mathbf{f}}_2^{e\prime} - \mathbf{B}_{1N}(\Sigma_{f_1 f_2^e} - \bar{f}_1\bar{f}_2^{e\prime}) - \hat{\mathbf{B}}_{2N,\text{MLC}}\Sigma_{f_2^e f_2^e} = \mathbf{0}_{N\times K_2}.$$

Now, inserting (IA14) and (IA15) into the above expression, setting

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

and re-arranging terms yields

$$\hat{\mathbf{B}}_{2N,\text{MLC}}\Sigma_{f_2^e f_2^e} - \mathbf{G}_N\hat{\mathbf{B}}_{2N,\text{MLC}}\bar{f}_2^e\bar{f}_2^{e\prime} = \Sigma_{r^e f_2^e} - \mathbf{G}_N\bar{r}_N^e\bar{f}_2^{e\prime} - \mathbf{B}_{1N}(\Sigma_{f_1 f_2^e} - \bar{f}_1\bar{f}_2^{e\prime}),$$

which can be rewritten more succinctly as $\frac{1}{T}\sum_{t=1}^T \mathbf{f}_{2t}^e\mathbf{g}_t' = \mathbf{0}_{K_2\times N}$, with $\mathbf{g}_t = \Big(\mathbf{h}_t - \mathbf{G}_N\bar{\mathbf{h}}_N - \hat{\mathbf{B}}_{2N,\text{MLC}}\mathbf{f}_{2t}^e + \mathbf{G}_N\hat{\mathbf{B}}_{2N,\text{MLC}}\bar{\mathbf{f}}_2^e\Big)$. Taking the vec operator and solving for $\hat{\mathbf{B}}_{2N,\text{MLC}}$ gives the desired expression in (IA13).

We need to show that a solution for $\hat{\mathbf{B}}_{2N,\text{MLC}}$ exists. This requires one to establish that the matrix $\Big((\mathbf{\Sigma}_{f_2^e f_2^e}\otimes\mathbf{I}_N) - (\bar{\mathbf{f}}_2^e\,\bar{\mathbf{f}}_2^{e\prime}\otimes\mathbf{G}_N)\Big)$ is invertible. This matrix can be written as

$$\Big((\mathbf{\Sigma}_{f_2^e f_2^e}\otimes\mathbf{I}_N) - (\bar{\mathbf{f}}_2^e\,\bar{\mathbf{f}}_2^{e\prime}\otimes\mathbf{G}_N)\Big) = \Big(((\mathbf{\Sigma}_{f_2^e f_2^e} - \bar{\mathbf{f}}_2^e\,\bar{\mathbf{f}}_2^{e\prime})\otimes\mathbf{I}_N) + (\bar{\mathbf{f}}_2^e\,\bar{\mathbf{f}}_2^{e\prime}\otimes(\mathbf{I}_N - \mathbf{G}_N))\Big).$$

The first matrix on the right hand side is non-singular, given the assumptions made. One then just needs to show that the second matrix is positive semi-definitive. In turn, this follows because $(\mathbf{I}_N - \mathbf{G}_N)$ is positive semi-definite, given

$$\begin{aligned}
\mathbf{I}_N - \mathbf{G}_N &= \mathbf{I}_N - \frac{1}{(\hat{\kappa}_{MLC}+1)}\mathbf{I}_N - (\frac{\hat{\kappa}_{MLC}}{1+\hat{\kappa}_{MLC}})\mathbf{D}_N(\mathbf{D}_N'\mathbf{\Sigma}_N^{-1}\mathbf{D}_N)^{-1}\mathbf{D}_N'\mathbf{\Sigma}_N^{-1} \\
&= (\frac{\hat{\kappa}_{MLC}}{1+\hat{\kappa}_{MLC}})(\mathbf{I}_N - \mathbf{D}_N(\mathbf{D}_N'\mathbf{\Sigma}_N^{-1}\mathbf{D}_N)^{-1}\mathbf{D}_N'\mathbf{\Sigma}_N^{-1}) \\
&= (\frac{\hat{\kappa}_{MLC}}{1+\hat{\kappa}_{MLC}})\mathbf{\Sigma}_N(\mathbf{\Sigma}_N^{-1} - \mathbf{\Sigma}_N^{-1}\mathbf{D}_N(\mathbf{D}_N'\mathbf{\Sigma}_N^{-1}\mathbf{D}_N)^{-1}\mathbf{D}_N'\mathbf{\Sigma}_N^{-1}) \\
&= (\frac{\hat{\kappa}_{MLC}}{1+\hat{\kappa}_{MLC}})\mathbf{\Sigma}_N\mathbf{\Sigma}_N^{-1/2}(\mathbf{I}_N - \mathbf{\Sigma}_N^{-1/2}\mathbf{D}_N(\mathbf{D}_N'\mathbf{\Sigma}_N^{-1}\mathbf{D}_N)^{-1}\mathbf{D}_N'\mathbf{\Sigma}_N^{-1/2})\mathbf{\Sigma}_N^{-1/2}.
\end{aligned}$$

The right-hand side is the product of positive-definite matrices, namely $\mathbf{\Sigma}_N$ and $\mathbf{\Sigma}_N^{-1/2}$, and of the matrix $\mathbf{I}_N - \mathbf{\Sigma}_N^{-1/2}\mathbf{D}_N(\mathbf{D}_N'\mathbf{\Sigma}_N^{-1}\mathbf{D}_N)^{-1}\mathbf{D}_N'\mathbf{\Sigma}_N^{-1/2}$, which is the projection matrix orthogonal to $\mathbf{\Sigma}_N^{-1/2}\mathbf{D}_N$, and therefore, positive semidefinite.

Therefore, plugging $\hat{\mathbf{B}}_{2N,\text{MLC}} = \hat{\mathbf{B}}_{2N}(\mathbf{D}_N,\mathbf{C}_N)$ into $\hat{\boldsymbol{\lambda}}_{\text{MLC}}$ and $\hat{\mathbf{a}}_{N,\text{MLC}}$, and then $\hat{\boldsymbol{\lambda}}_{\text{MLC}}$ into $\hat{\mathbf{a}}_{N,\text{MLC}}$, one obtains that

$$\hat{\mathbf{B}}_{2N,\text{MLC}} = \hat{\mathbf{B}}_{2N}(\mathbf{D}_{N,},\mathbf{C}_N),\;\; \hat{\boldsymbol{\lambda}}_{\text{MLC}} = \hat{\boldsymbol{\lambda}}(\mathbf{D}_{N,},\mathbf{C}_N),\;\; \hat{\mathbf{a}}_{N,\text{MLC}} = \hat{\mathbf{a}}_N(\mathbf{D}_N,\mathbf{C}_N),\;\text{ and }\; \hat{\kappa} = \hat{\kappa}(\mathbf{D}_N,\mathbf{C}_N).$$

Substituting them into $L(\boldsymbol{\theta}) - \kappa(\mathbf{a}_N'\mathbf{\Sigma}_N^{-1}\mathbf{a}_N - \delta)$, gives the concentrated log-likelihood function, which is a function of only $\mathbf{D}_N$ and $\mathbf{C}_N$ and it will be maximized numerically to obtain $\hat{\mathbf{D}}_{N,\text{MLC}}$ and $\hat{\mathbf{C}}_{N,\text{MLC}}$. Observe that the penalization term vanishes for the concentrated log likelihood function for either $\hat{\kappa} = 0$ and $\hat{\kappa}_{MLC} > 0$.

(ii) Suppose now that $\hat{\kappa} = 0$. One can clearly obtain a unique solution for $(\mathbf{D}_N, \mathbf{I}_N)\begin{pmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\mathbf{a}}_{N,} \end{pmatrix} = \mathbf{D}_N \hat{\boldsymbol{\lambda}} + \hat{\mathbf{a}}_N$. However, to solve for $\hat{\boldsymbol{\lambda}}$ and $\hat{\mathbf{a}}_N$ separately, one needs to invert the matrix

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

which is not possible because it is of dimension $(N + K^*) \times (N + K^*)$ but of rank N, because the left-hand side shows that it is obtained from the product of two matrices of dimension $(N+K^*) \times N$. Thus, only the sum $\mathbf{D}_N \hat{\boldsymbol{\lambda}} + \hat{\mathbf{a}}_N$ can be estimated. All the other parameters are identified separately and their expressions follow from differentiating $L(\boldsymbol{\theta})$ and solving the resulting first-order conditions. For instance, the formula for $\hat{\mathbf{B}}_{2N}$ follows from setting $\hat{\mathbf{G}}_N = \mathbf{I}_N$ into (IA13). ■

IA.4 Evaluating Out-of-Sample Performance for Simulated Data

In Section 5 of the main text, we have compared the performance of the strategies developed in this paper to that of the benchmark strategies for two empirical data sets. In this section, we compare the performance for simulated data, which allows us to evaluate how the relative performance of our strategies is influenced by changes in the data-generating process.

The data and experiment design are explained in Section 5.1 of the main text. Details of how the model parameters are estimated are given in Section 5.2. The benchmark strategies we consider are described in Section 5.3 and the robust-mean-variance portfolios whose performance we evaluate are described in Section 5.4.

IA.4.1 Simulation Design

In our simulation analysis, to match the two empirical data sets we study (see Section 5.1 for details), we consider the case in which the number of assets is $N = 30$ and $N = 100$, which allows us to illustrate the effect of having a small and a large number of assets. Throughout all our experiments, the investor assumes, and therefore estimates, a three-factor model:

$$\mathbf{r}_t = \boldsymbol{\alpha}_N + \mathbf{B}_N \mathbf{f}_t + \boldsymbol{\varepsilon}_t.$$

We assume that the risk-free interest rate is zero and that the observed factors are uncorrelated with an IID Gaussian distribution. We calibrate the distribution of the three factors to match the mean and standard deviations of the three Fama-French factors, with the vector of means equal to $\{\frac{8}{12\times100}, \frac{3}{12\times100}, \frac{4}{12\times100}\}$ and the vector of standard deviations equal to $\{\frac{16}{\sqrt{12\times100}}, \frac{11}{\sqrt{12\times100}}, \frac{10}{\sqrt{12\times100}}\}$.

We consider an environment where the mispricing allows for a missing pervasive factor ($p = 1$) and also a component unrelated to the pervasive factor; that is, $\boldsymbol{\alpha}_N = \mathbf{A}_N \lambda_{\text{miss}} + \mathbf{a}_N$. Both $\mathbf{a}_N$ and

$\mathbf{A}_N$ are generated from an IID multivariate Gaussian distribution with mean $\mathbf{0}_N$ and covariance matrices equal to $\sigma_a^2\mathbf{I}_N$ and $\sigma_A^2\mathbf{I}_N$, respectively. We calibrate the values of these parameters using estimates based on the DJ30 dataset, so the value of σ_a is set equal to $\frac{1}{3} \times \frac{5}{\sqrt{12\times 100}}$ and the value of σ_A is equal to $\frac{6}{\sqrt{12\times 100}}$.[vi] The monthly covariance matrix is $\mathbf{\Sigma}_N = \mathbf{A}_N\mathbf{A}_N' + \sigma_\varepsilon^2\mathbf{I}_N$, where ε_t is IID with a multivariate Gaussian distribution with a monthly mean of 0 and a standard deviation of $\sigma_\varepsilon = \frac{20}{\sqrt{12\times 100}}$. Note that to ensure identification, just as in MacKinlay and Pástor (2000), we set the variance of the missing factor equal to one, which implies that the magnitude of the risk premium λ_{miss} coincides with the Sharpe ratio for the missing factor. Based on the estimate from the DJ30 data set, λ_{miss} is set equal to $\frac{65}{\sqrt{12\times 100}}$. Based on these parameter values, we consider 100 simulations and report results averaged across these simulations.

An important element of the estimation procedure is the choice of δ_{apt} that appears in the APT constraint in (36). Ross (1976, p. 354) suggests constraining δ_{apt} to be a multiple of the Sharpe ratio of the market portfolio. Instead, we use the 5-fold cross-validation procedure in Hastie, Tibshirani, and Wainwright (2015, Section 2.3) to identify δ_{apt}, by maximizing the Sharpe ratio of the holdout portfolio returns in the cross-validation approach. To ensure that there is no look-ahead bias, when choosing δ_{apt} we rely only on data in the observation window.

IA.4.2 Out-of-sample performance for simulated data sets

The performance metrics for the four benchmark strategies and the four strategies developed in this paper for the data set consisting of monthly simulated returns for $N = 30$ and $N = 100$ stocks are reported in Panels A and B of Table IA.1. Each panel reports, for each strategy, the mean and Sharpe ratio of the per annum returns in excess of the risk-free rate. The table also reports the improvement in Sharpe ratio of each strategy relative to the Sharpe ratio of the EW and MAXSER strategies; for instance, when comparing strategy k to EW, we report $(\text{SR}_k - \text{SR}_{\text{EW}})/\text{SR}_{\text{EW}}$. Finally, the table reports the t-statistic for the difference between each strategy's Sharpe ratio and the Sharpe ratios of the EW and MAXSER strategies. Given that the simulated data is generated using normal distributions, we compute the t-statistic using the approach in Lo (2002).[vii]

We start by looking at the performance of the benchmark strategies reported in the first four rows of Panel A.1 for the "base case" in Table IA.1. We see that the MV portfolio strategy achieves

[vi] The value for σ_a is one third of the value used in MacKinlay and Pástor (2000) and, as we show below, using a larger value of σ_a would lead to stronger results.

[vii] Specifically, we construct the test statistic for the null hypothesis that $\text{SR}_k = \text{SR}_0$ using the limiting distribution of $\widehat{\text{SR}}$ reported in Lo (2002, p. 38), where the standard error is given in his equation (9). Here SR_k means the Sharpe ratio associated with strategy k, while SR_0 is the Sharpe ratio associated with our benchmark (EW or MAXSER). Thus, the t-statistic is $(\text{SR}_k - \text{SR}_0)/\widehat{\text{SE}}_0$, where $\widehat{\text{SE}}_0$ is the standard error under the null hypothesis obtained using Lo (2002, eq. 38).

an annual Sharpe ratio of 0.329, while that of GMV-LW is slightly higher, 0.395. The GMV-LW portfolio is only slightly better than the MV portfolio because the number of assets is small. Even though the number of risky assets for this data set is relatively small, $N = 30$, the EW portfolio with a Sharpe ratio of 0.419 outperforms both of these optimizing strategies—as has been shown before in the literature (DeMiguel et al., 2009b). The MAXSER strategy, developed in Ao et al. (2019), outperforms all these strategies—it has a substantially higher Sharpe ratio of 0.835, which is 99.2% higher than the Sharpe ratio of the EW portfolio.

We now compare the performance of our strategies, reported in the last four rows of Panel A.1 of Table IA.1, to that of the benchmark strategies described above. The strategy "RMV using $\mathbf{V}$" achieves a Sharpe ratio of 0.918, which is 119% greater than that of the EW portfolio and 10% higher than that of MAXSER. The Sharpe ratio of the "RMV using $\mathbf{V}$: OptComb" strategy, which combines the alpha and beta portfolios optimally, is even higher, 0.932, which is 122.3% larger than that of the EW portfolio and 11.6% higher than that of MAXSER, an increase that is statistically significant (the t-statistic is greater than 4). The results for the other two strategies, "RMV using $\boldsymbol{\Omega}$" and "RMV using $\boldsymbol{\Omega}$: OptComb," are similar. In both cases, the Sharpe ratios are greater than those of the EW and MAXSER strategies, with the difference being statistically significant (the t-statistic is around 3); the difference is a bit smaller because replacing $\mathbf{V}_N$ with $\boldsymbol{\Omega}$ works better when N is large, as described below.

In addition to the "base case" in Panel A.1 described above, we look at five variations of the data-generating process in Panels B to F of Table IA.1. Panel A.2 considers the case in which the residual risk is 75% of its base-case value. In this case, the Sharpe ratios of the strategies that rely on estimated mean returns improve.[viii] The Sharpe ratio of the MV strategy improves from 0.329 in Panel A to 0.336 in Panel B. Much more striking is the improvement in the Sharpe ratio of MAXSER: from 0.835 to 1.157. However, even in this setting, the four strategies developed in this paper outperform MAXSER, with the t-statistic for the difference in Sharpe ratios ranging from 2.055 to 4.469.

Panel A.3 considers the case in which the standard deviation of $\mathbf{a}_N$, which represents the cross-section dispersion of the pure pricing errors across assets, is 1.25 times its base-case value in Panel A. In this case, the four portfolios developed in this paper outperform the EW portfolio substantially, but the improvement in their performance relative to MAXSER is not statistically significant.

Panel A.4 considers the case in which the standard deviation of $\mathbf{A}_N$, is 1.25 times it base-case value in Panel A. In this case, the performance of the MV portfolio deteriorates substantially, but

[viii]The Sharpe ratio of the unscaled EW portfolio does not change at all, but that of the EW strategy scaled to achieve a particular level of volatility increases.

that of MAXSER improves. However, the four strategies developed in this paper still outperform EW, with a Sharpe ratio that is about 90% higher with the t-statistic for the difference in Sharpe ratios being greater than 25. Our strategies also outperform MAXSER, with the t-statistic for the difference in Sharpe ratios ranging from 2.314 to 5.380.

Panels A.5 and A.6 consider the cases in which the true risk premium on the observed factor, λ and the missing factor, λ_{miss}, respectively, is set to 1.25 times its base-case value in Panel A.1. Even in this setting, "RMV using $\mathbf{V}$: OptComb" has a Sharpe ratio whose difference relative to all the other strategies is statistically significant.

To understand how the performance of our strategies changes with the number of risky assets, we now study the results in Panel B of Table IA.1, which is for the case of $N = 100$ stocks. Comparing the results for the base case in Panel B.1 to those in Panel A.1, we find the performance of the MV and GMV-LW strategies drops. As is well known, this is because the estimation problem becomes more severe as the number of assets increases. On the other hand, the performance of the EW and MAXSER strategies improves with N. However, the performance of the robust mean-variance portfolios developed in this paper improves even more, which is because our theory applies particularly well to the case of large N given that it is based on the APT. The robust-mean-variance portfolios achieve a Sharpe ratio more than 175% higher than that of the EW portfolio (with t-statistic greater than 49) and more than 30% higher than that of MAXSER (with t-statistic greater than 17).

In Panel B of Table IA.1, we can look at how the performance of the robust-mean-variance portfolios varies as we change the parameters of the data-generating process when we have $N = 100$ stocks, just as we did in Panel A for the case of $N = 30$ stocks. Panels B.2 to B.6 of Table IA.1 show that in each case the results are much stronger than in the corresponding cases considered in Panel A of Table IA.1. In all five panels, the Sharpe ratios of the robust-mean-variance portfolios are about 150% higher than of the EW portfolio, with the t-statistic for the difference ranging from 40 to 95. When compared to MAXSER, the Sharpe ratios of the robust-mean-variance portfolios are about 10% to 30% higher, with the t-statistic for the difference ranging from 6 to 33.

These results provide strong evidence that the robust-mean-variance portfolios deliver performance gains out of sample that are economically substantial and statistically significant relative to the strategies in the existing literature. These gains are particularly striking when the number of assets is large, which is a case that traditional methodologies find especially challenging to handle. In Section 5 of the main text, we show that the performance gains demonstrated in simulated data are also present in empirical data.

Table IA.1: Out-of-sample portfolio performance: Simulated returns

This table reports, for monthly returns of $N = 30$ assets (Panel A) and $N = 100$ assets (Panel B) generated via Monte Carlo simulations, the performance of four benchmark strategies, and four strategies developed in this paper when the investor targets volatility of $0.05 \times \sqrt{12} = 0.1732$ per year. The parameters for all strategies are estimated using a rolling window of $T = 120$ monthly observations. For each strategy, the table reports its per annum return's mean and Sharpe ratio. The table also reports the improvement in the Sharpe ratio of each strategy with respect to the EW and MAXSER strategies; for instance, when comparing strategy k to EW, SR wrt EW is $(\mathrm{SR}_k - \mathrm{SR}_{\mathrm{EW}})/\mathrm{SR}_{\mathrm{EW}}$. Finally, the table reports the t-statistics for the difference between each strategy's Sharpe ratio and the EW and MAXSER strategies.

	Mean	SR	SR wrt		t-stat wrt	
	p.a.	p.a.	EW	MAXSER	EW	MAXSER
Panel A: For DJIA 30 constituents						
Panel A.1: Base case						
MV	0.168	0.329	−0.215	−0.606	−4.646	−25.538
GMV-LW	0.096	0.395	−0.058	−0.527	−1.251	−22.215
EW	0.156	0.419	0.000	−0.498	NA	−20.990
MAXSER	0.204	0.835	0.992	0.000	21.443	NA
RMV using $\mathbf{V}$	0.168	0.918	1.190	0.100	25.818	4.283
RMV using $\mathbf{V}$: OptComb	0.156	0.932	1.223	0.116	26.446	4.898
RMV using $\mathbf{\Omega}$	0.168	0.894	1.132	0.071	24.488	2.980
RMV using $\mathbf{\Omega}$: OptComb	0.156	0.908	1.165	0.087	25.283	3.759
Panel A.2: Low σ_ϵ (0.75 × base case)						
MV	0.240	0.336	−0.260	−0.710	−6.068	−40.406
GMV-LW	0.144	0.385	−0.153	−0.668	−3.569	−38.019
EW	0.144	0.454	0.000	−0.608	NA	−34.610
MAXSER	0.228	1.157	1.550	0.000	36.229	NA
RMV using $\mathbf{V}$	0.228	1.247	1.748	0.078	40.954	4.513
RMV using $\mathbf{V}$: OptComb	0.216	1.247	1.748	0.078	40.908	4.469
RMV using $\mathbf{\Omega}$	0.228	1.199	1.641	0.036	38.380	2.055
RMV using $\mathbf{\Omega}$: OptComb	0.216	1.233	1.718	0.066	40.242	3.834
Panel A.3: High σ_a (1.25 × base case)						
MV	0.204	0.391	−0.252	−0.648	−6.763	−35.606
GMV-LW	0.144	0.464	−0.113	−0.583	−3.093	−32.076
EW	0.144	0.523	0.000	−0.530	NA	−29.101
MAXSER	0.216	1.112	1.126	0.000	30.255	NA
RMV using $\mathbf{V}$	0.216	1.136	1.172	0.022	31.559	1.254
RMV using $\mathbf{V}$: OptComb	0.192	1.129	1.159	0.016	31.071	0.784
RMV using $\mathbf{\Omega}$	0.228	1.109	1.119	−0.003	30.116	−0.134
RMV using $\mathbf{\Omega}$: OptComb	0.192	1.129	1.159	0.016	31.198	0.907
Panel A.4: High σ_A (1.25 × base case)						
MV	0.072	0.048	−0.919	−0.955	−28.025	−50.524
GMV-LW	0.096	0.513	−0.140	−0.519	−4.261	−27.499
EW	0.144	0.596	0.000	−0.442	NA	−23.371
MAXSER	0.204	1.067	0.791	0.000	24.123	NA
RMV using $\mathbf{V}$	0.204	1.133	0.901	0.062	27.421	3.195
RMV using $\mathbf{V}$: OptComb	0.204	1.174	0.971	0.101	29.677	5.380
RMV using $\mathbf{\Omega}$	0.204	1.112	0.866	0.042	26.511	2.314
RMV using $\mathbf{\Omega}$: OptComb	0.204	1.160	0.948	0.088	28.897	4.625
Panel A.5: High λ (1.25 × base case)						
MV	0.168	0.319	−0.471	−0.678	−14.190	−33.548
GMV-LW	0.120	0.565	−0.063	−0.430	−1.596	−21.270
EW	0.156	0.603	0.000	−0.392	NA	−19.714
MAXSER	0.180	0.991	0.644	0.000	20.221	NA
RMV using $\mathbf{V}$	0.180	1.015	0.684	0.024	21.397	1.147
RMV using $\mathbf{V}$: OptComb	0.192	1.105	0.833	0.115	26.019	5.653
RMV using $\mathbf{\Omega}$	0.180	0.998	0.655	0.007	20.562	0.332
RMV using $\mathbf{\Omega}$: OptComb	0.180	1.057	0.753	0.066	23.654	3.347
Panel A.6: High λ_{miss} (1.25 × base case)						
MV	0.192	0.436	−0.300	−0.564	−9.565	−28.164
GMV-LW	0.096	0.513	−0.178	−0.488	−5.672	−24.366
EW	0.144	0.624	0.000	−0.377	NA	−18.834
MAXSER	0.192	1.001	0.606	0.000	19.307	NA
RMV using $\mathbf{V}$	0.192	1.025	0.644	0.024	20.474	1.138
RMV using $\mathbf{V}$: OptComb	0.180	1.046	0.678	0.045	21.614	2.250
RMV using $\mathbf{\Omega}$	0.168	0.980	0.572	−0.021	18.287	−0.995
RMV using $\mathbf{\Omega}$: OptComb	0.180	1.039	0.667	0.038	21.251	1.896

	Mean p.a.	SR p.a.	SR wrt EW	SR wrt MAXSER	t-stat wrt EW	t-stat wrt MAXSER
Panel B: For S&P 500 constituents						
Panel B.1: Base case						
MV	0.144	0.048	−0.910	−0.957	−25.253	−53.635
GMV-LW	0.144	0.121	−0.776	−0.893	−21.525	−50.054
EW	0.120	0.540	0.000	−0.524	NA	−29.380
MAXSER	0.300	1.136	1.103	0.000	30.588	NA
RMV using $\mathbf{V}$	0.300	1.510	1.795	0.329	49.794	18.448
RMV using $\mathbf{V}$: OptComb	0.288	1.566	1.897	0.378	52.639	21.181
RMV using $\mathbf{\Omega}$	0.300	1.496	1.769	0.317	49.083	17.764
RMV using $\mathbf{\Omega}$: OptComb	0.288	1.562	1.891	0.375	52.462	21.010
Panel B.2: Low σ_ϵ (0.75 × base case)						
MV	0.192	0.042	−0.934	−0.976	−29.930	−78.455
GMV-LW	0.144	0.128	−0.796	−0.926	−25.502	−74.460
EW	0.180	0.627	0.000	−0.640	NA	−51.451
MAXSER	0.444	1.742	1.779	0.000	57.026	NA
RMV using $\mathbf{V}$	0.432	2.290	2.652	0.314	84.965	25.208
RMV using $\mathbf{V}$: OptComb	0.420	2.477	2.950	0.421	94.612	33.912
RMV using $\mathbf{\Omega}$	0.432	2.279	2.635	0.308	84.545	24.829
RMV using $\mathbf{\Omega}$: OptComb	0.420	2.456	2.917	0.410	93.541	32.945
Panel B.3: High σ_a (1.25 × base case)						
MV	0.108	0.028	−0.957	−0.985	−31.652	−82.029
GMV-LW	0.144	0.111	−0.829	−0.939	−27.406	−78.236
EW	0.120	0.648	0.000	−0.645	NA	−53.751
MAXSER	0.408	1.826	1.818	0.000	60.164	NA
RMV using $\mathbf{V}$	0.396	2.099	2.241	0.150	74.142	12.488
RMV using $\mathbf{V}$: OptComb	0.384	2.252	2.476	0.233	81.986	19.496
RMV using $\mathbf{\Omega}$	0.384	2.072	2.198	0.135	72.644	11.150
RMV using $\mathbf{\Omega}$: OptComb	0.372	2.224	2.433	0.218	80.582	18.241
Panel B.4: High σ_A (1.25 × base case)						
MV	0.168	0.031	−0.952	−0.979	−31.782	−70.052
GMV-LW	0.084	0.097	−0.852	−0.936	−28.374	−66.884
EW	0.120	0.655	0.000	−0.566	NA	−40.501
MAXSER	0.300	1.507	1.302	0.000	43.557	NA
RMV using $\mathbf{V}$	0.312	1.635	1.497	0.085	50.101	6.084
RMV using $\mathbf{V}$: OptComb	0.300	1.708	1.608	0.133	53.808	9.532
RMV using $\mathbf{\Omega}$	0.312	1.649	1.519	0.094	50.808	6.742
RMV using $\mathbf{\Omega}$: OptComb	0.288	1.690	1.582	0.122	52.947	8.731
Panel B.5: High λ (1.25 × base case)						
MV	0.016	0.038	−0.944	−0.975	−32.672	−69.312
GMV-LW	0.024	0.177	−0.740	−0.882	−25.580	−62.700
EW	0.144	0.679	0.000	−0.546	NA	−38.854
MAXSER	0.300	1.496	1.204	0.000	41.679	NA
RMV using $\mathbf{V}$	0.324	1.642	1.418	0.097	49.057	6.878
RMV using $\mathbf{V}$: OptComb	0.300	1.746	1.571	0.167	54.364	11.825
RMV using $\mathbf{\Omega}$	0.300	1.642	1.418	0.097	49.082	6.901
RMV using $\mathbf{\Omega}$: OptComb	0.300	1.725	1.541	0.153	53.257	10.793
Panel B.6: High λ_{miss} (1.25 × base case)						
MV	0.168	0.111	−0.831	−0.927	−27.714	−66.408
GMV-LW	0.084	0.201	−0.693	−0.867	−23.116	−62.134
EW	0.120	0.655	0.000	−0.567	NA	−40.649
MAXSER	0.324	1.510	1.307	0.000	43.734	NA
RMV using $\mathbf{V}$	0.312	1.673	1.556	0.108	51.981	7.666
RMV using $\mathbf{V}$: OptComb	0.312	1.808	1.762	0.197	58.969	14.161
RMV using $\mathbf{\Omega}$	0.312	1.663	1.540	0.101	51.556	7.270
RMV using $\mathbf{\Omega}$: OptComb	0.300	1.777	1.714	0.177	57.377	12.681