Statistical Arbitrage without Arbitrage *

Massimo Dello Preite Valentina Raponi Paolo Zaffaroni

March 15, 2025

Abstract

Statistical arbitrage strategies are quantitative trading strategies based on alpha, namely the signal extracted from the residuals when fitting an asset pricing model to the returns data. We develop a normative theory that combines the insights of mean-variance portfolio choice with statistical arbitrage in a no-arbitrage setting (*statistical arbitrage without arbitrage*). We establish a novel two-fund separation result that combines the statistical arbitrage portfolio and the beta portfolio (i.e., the portfolio stemming from factor asset pricing models), showing how their combination can span the *efficient* frontier despite both being *inefficient* . Under no-arbitrage, as the number of assets increases, the "statistical arbitrage portfolio" dominates the "beta portfolio" both in terms of the magnitude of its weights and in terms of its Sharpe ratio. Exploiting the insights from mean-variance portfolio choice, we show how to construct a type of statistical arbitrage portfolio that does not require estimating alpha, in contrast to the common view. When statistical arbitrage is combined with factor asset pricing modelling, alpha and the latent factors' risk premia are not identified. We derive a penalized estimator under a no-arbitrage constraint that allows for the econometric identification of the statistical arbitrage portfolio, leading to a shrinkage-type estimator of the asset pricing model's parameters. We demonstrate our theoretical insights by means of Monte Carlo experiments and empirical applications.

JEL classification: C58, G11, G12.
Keywords: Statistical Arbitrage Portfolios; APT; Alpha and Beta portfolios; Mean-Variance Efficient Portfolio; Global Minimum-Variance Portfolio; Identification; Regularized Estimation.

*Raponi is affiliated with IESE Business School; email: vraponi@iese.edu. Dello Preite is affiliated with Imperial College Business School; email: M.Dellopreite@imperial.ac.uk. Zaffaroni is affiliated with Imperial College Business School; email: P.Zaffaroni@imperial.ac.uk.

1 Introduction

Statistical Arbitrage (SA) porfolios constitute an important class of data-driven quantitative trading strategies, with a multitude of proprietary variants developed by the most sophisticated trading houses. SA strategies are by and large based on the signal extracted from the residuals when fitting an asset pricing model to the returns data, i.e., on alpha, and exploit temporal price differences between assets, i.e., asset-specific dynamics, using a variety of statistical methods. The *arbitrage* notion of the SA portfolios refers to them being portfolios that hedge out the systematic risk, i.e., the risk spanned by systematic risk factors.

However, the development of SA strategies has been by and large independent of asset pricing theory, prompting many unresolved questions of both normative and positive trait. First and foremost, we do not know how no-arbitrage, the cornerstone assumption of asset pricing, affects and constraints SA. Second, from a different angle, we do not know the extent to which SA contributes to achieving mean-variance efficiency when combined with more traditional beta strategies, i.e., strategies that rely on using a factor asset pricing model. Third, it has never been analyzed whether the insights stemming from mean-variance portfolio choice help elicit salient properties of SA. Addressing these questions is the goal of this paper which, more in general, can be interepreted as analyzing the role of SA, and its properties, from the lenses of mean-variance portfolio choice.

The mean-variance portfolio choice model of Markowitz is the cornerstone of financial economics because it captures the fundamental trade-off between expected return and risk.[1] Motivated by the key insight of the Capital Asset Pricing Model (Sharpe, 1964) that investors hold only the market portfolio in equilibrium, leads to the result that mean-variance portfolios are exposed only to a single source of systematic risk, i.e., the risk of the market portfolio, and thus alpha risk is not compensated. However, the poor empirical performance of the CAPM to explain the cross-section of stock returns has led the literature to explored a large number of other observed *systematic* risk factors, which Cochrane (2011) describes as the *factor zoo*. However, empirical evidence suggests that multi-factor models of systematic risk alone do not span the cross-section of stock returns, and thus, the efficient frontier (e.g., MacKinlay, 1995; Daniel and Titman, 1997). In other words, another source of risk, other than systematic risk, here referred to *alpha* risk, appears priced in the cross-section, the compensation for which is precisely what SA is harvesting from a porfolio choice perspective.

In this paper, we investigate theoretically and empirically the implications for portfolio choice when alpha risk is priced in the no-arbitrage setting of the APT (Ross, 1976, 1977; Chamberlain, 1983; Chamberlain and Rothschild, 1983), a crucial departure from the standard framework with exact-factor pricing, leading to novel insights for portfolio choice, by combining *alpha* (i.e., SA) portfolios with the traditional mean-variance *beta* portfolios. The APT is a convenient framework to study SA within the context of mean-variance portfolio choice because it allows for non-zero alphas, hence justifying a role for SA, whilst ruling out arbitrage opportunities. The existing portfolio choice theoretical literature has largely ignored SA portfolios because, if alpha risk is assumed to be uncompensated, then it is indeed optimal to diversify it away. But, if alpha risk is

[1] Cochrane (2014) provides a good description of the broad range of applications of mean-variance theory. 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. However, when choosing a portfolio, it is not possible for the investor to wait for an infinite amount of data when employing nonparametric approaches, such as using the sample moments of asset returns, estimating the portfolio weights directly (Aït-Sahalia and Brandt, 2001), or identifying the model-free stochastic discount factor (Schneider and Trojani, 2019; Ghosh, Julliard, and Taylor, 2019).

compensated, then it is not optimal to diversify it away and SA becomes important. This is what we address in this paper.

Our paper advances our understanding of SA from two view angles, namely in terms of portfolio choice theory and in terms of portfolio choice estimation. We provide a theory of optimal mean-variance portfolio choice when alpha risk is compensated, implying that the traditional systematic (ie., beta) portfolio is no longer an efficient portfolio, i.e., it is inside the MVE frontier. Our first theoretical result shows how to construct the mean-variance efficient frontier by combining the risk-free asset and the inefficient beta portfolio with the SA portfolio, also inefficient, which depends on alpha risk and its compensation, thus achieving three-fund separation. Therefore, SA plays a fundamental role when considered within the context of mean-variance portfolio choice. When analyzed jointly in a mean-variance setting, the SA and beta portfolios, despite being inefficient, are tight together and are shown to share a (sub-) optimality property, namely one is the *unique minimum-variance* portfolio orthogonal to the other.

Our second theoretical result establish the properties of the SA portfolio under the APT assumptions, in particular when arbitrage opportunities are ruled out. In particular, the SA portfolio becomes orthogonal to the beta portfolio, i.e., it is beta-neutral, as the number of assets increases. This holds regardless of whether the alphas and the systematic risk factors' exposures are themselves cross-sectionally orthogonal.[2]. Moreover, as the number of risky assets increases, the SA portfolio dominates the beta portfolio both in terms of portfolio weights and Sharpe ratio (SR).

Existing SA strategies are typically based on estimates of alpha, defined as the discrepancy between average return and the estimated asset pricing model-implied expected return. However, alpha, which can be interpreted as the asset-specific component of expected returns, can be small and often very un-precisely estimated, representing a serious challenge for the implementation of SA portfolios. Building on our mean-variance fund separation result, our third theoretical insight shows how one construct a special SA strategy that is functionally independent of the alphas, i.e. that does not require to estimate them. In particular, we show how to decompose the GMV portfolio into a beta and beta-neutral components, the latter representing a type of SA strategy but with the special feature of being *functionally independent* of the alphas, thanks to the GMV portfolio feature of not requiring expected returns for its formulation.[3]

Joint estimation of alpha (i.e., SA) and beta strategies jointly poses an *identification* problem, hitherto unknown as SA were never formally studied in the context of portfolio choice. In particular, we show that the risk premia to systematic latent risk factors and alpha are not jointly identified without further restrictions. We show how the APT main asset pricing insight, namely boundedness of the SR of the SA portfolio (hereafter referred to as the *APT constraint*), provides the required identification condition. We derive an estimator based on the principle of *penalized* (subject to the APT constraint) maximum likelihood (ML), able to estimate the SA and beta portfolios jointly. The APT constraint not only ensures model identification but leads to *shrinkage* estimator of the SA portfolio, in the same spirit to the recent advances in machine learning estimators (e.g., see Hastie, Tibshirani, and Friedman (2011)) but specifically applied to asset pricing. This shrinkage flavour is particularly relevant given that the number of parameters associated with SA, namely the asset-

[2]Slope of capital market line.

[3]As explained below, clearly the statistical properties (i.e., expected return and SR) of the beta-neutral component of the GMV portfolio will depend on the (population) alphas.

specific component (i.e., alphas) of expected returns and their covariances, are large-dimensional because they increase with N.[4]

As a by-product of our estimation procedure, we also show that the SA portfolio estimator is observationally equivalent to the portfolio obtained by an investor averse to model misspecification in the sense of Hansen and Sargent (2007), thus mitigating the concern about model misspecification because of the use of a fully parameterized specification. The estimated beta portfolio does *not* correspond to a portfolio robust to misspecification; however, we show how it can be made *perfectly immune* to misspecification of the estimated premia and covariances for systematic risk factors.

The single economic insight that underlies our analysis that the tradeoff between risk and return changes significantly when there is compensation for alpha risk. To better understand our results, recall that mean-variance portfolio optimization aims to maximize the portfolio SR. If alpha risk has zero compensation, as is typically assumed, then it is indeed optimal to fully diversify away the alpha risk by choosing portfolio weights that are small and that optimize the tradeoff between systematic risk and its compensation. We denote this portfolio as the "systematic-risk portfolio." In contrast, if the alpha s risk has nonzero compensation, then it is no longer optimal to fully diversify this risk by choosing weights that are small because this will reduce the return from bearing alpha risk. Instead, it is now optimal to choose a portfolio that has two components: the *systematic-risk portfolio*, which has the same properties as before, and the *SA portfolio*, whose weights are large but are long and short positions in order to diversify (i.e., orthogonalize) systematic risk, while optimizing the tradeoff between alpha risk and its compensation (i.e., alpha). Importantly, because of the above property, as the number of assets N increases, the weights comprising the systematic-risk portfolio decrease faster (at the rate $1/N$) than the weights of the SA portfolio (which decrease at the rate $1/\sqrt{N}$). As a result of this, the weights of the overall portfolio (MVE or GMV) are dominated by the weights of the SA portfolio. More importantly, the overall portfolio's SR is also dominated by the SA portfolio.

We demonstrate that our theoretical insights on SA portfolios lead to an economically and statistically significant improvement in out-of-sample performance of mean-variance efficient portfolios. Empirically, we demonstrate that allowing for compensation for alpha risk leads to a substantial improvement in portfolio performance.

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).[5] 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).[6] We demonstrate the importance of the latent purely unsystematic 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

[4]The portfolio choice analysis of SA portfolios developed in the first part of this paper does not depend on the estimation procedure proposed in the second part of the paper: the penalized estimator here proposed represents one of the many possible estimation procedures for our methodology.

[5]In Section **??** of the Internet Appendix, we also use simulated data to evaluate performance.

[6]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.

We now explain how the three facets of our work—portfolio choice theory, regularized estimation, and empirical results—are related to the existing literature. The notion that mean-variance portfolios can be decomposed into two inefficient portfolios was pioneered by Treynor and Black (1973), who labelled them "active" and "passive" portfolios. Roll (1980) formalizes this idea by introducing "orthogonal portfolios" (i.e. the active portfolios).[7] MacKinlay (1995, page 6) exploits the notion of orthogonal portfolios of Roll (1980) and states that "...for the nonrisk-based [what we refer to as unsystematic risk] alternatives the maximum squared Sharpe measure ...can, in principle, be unbounded." The poor performance of factor models to explain the cross-section of expected returns also emerges in Daniel and Titman (1997) who demonstrate how asset-specific characteristics (i.e., alpha) rather than the covariance structure of returns (i.e., betas) appear much more informative cross-sectionally. MacKinlay and Pástor (2000), building on MacKinlay (1995), study portfolio choice with a missing systematic risk factor but ignore any role for compensated alpha risk.[8] Daniel, Mota, Rottke, and Santos (2020) show 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 factors with zero risk premia. The evidence in these papers strongly supports our approach to building a mean-variance efficient portfolio that exploits alpha risk by means of an SA strategy. Because our theory is founded on the APT, we can analyze the implication of no-arbitrage. Moreover, the APT by construction allows us to include latent systematic risk factors that may be otherwise omitted. Moreover, our methodology highlights the importance of the APT constraint, another reflection of the important role of SA strategies within mean-variance theory.

On the estimation front, our estimation approach is related to the vast literature that use of variations of principal-component analysis (PCA) for estimating factor asset pricing models, often using a variety of machine-learning methods.[9] Kozak, Nagel, and Santosh (2018) show that the stochastic discount factor can be represented as a function of a few dominant sources of return variation, which can be accurately estimated using PCA. The high-frequency PCA in Pelger (2020) uses high-frequency data to estimate local time-varying latent risk factors in a conditional asset pricing model setting. Lettau and Pelger (2020b,a) generalize PCA by including a penalty on the pricing error in expected returns (coining the Risk Premium PCA) and show that they can identify "weak factors" (factors with small risk exposures) with high SRs. Giglio and Xiu (2021) propose a method to estimate the risk premium of an observable factor that is valid even when some systematic risk factors are assumed latent (captured by PCA), extended in Giglio, Xiu, and Zhang (2023) to allow for the case of weak observed risk factors. Zaffaroni (2025) studies local PCAs to estimate conditional asset pricing models applicable to short panels of asset returns, developing tests of pricing ability.

Whereas the previous papers focused on estimating systematic risk factors, and their risk premia, other PCA-type methodologies focused on SA investment strategies, in the sense of strategies

[7] 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.

[8] See also Pástor and Stambaugh (1999, 2000) and Pástor (2000).

[9] A related, but different, strand of empirical asset pricing develop regularization and machine learning methods to estimate asset pricing models (i.e., the SDF) using vast data made of test asset returns and observed risk factors without resorting to latent factor models (see, e.g., Freyberger, Neuhierl, and Weber (2020), Feng, Giglio, and Xiu (2020), Chen, Pelger, and Zhu (2024), Bryzgalova, Huang, and Julliard (2023a), Bryzgalova, Pelger, and Zhu (2023b), among others).

that are orthogonal to systematic risk and are constructed using the estimated residuals from fitting a factor asset pricing model to the data.[10] Estimation of arbitrage portfolios, where alphas are assumed linear in observed characteristics, is studied in Kelly, Pruitt, and Su (2019) although the main focus of their IPCA methodology is the estimation of latent risk factors with risk exposures spanned by the same characteristics used for capture alphas. Kozak, Nagel, and Santosh (2020) propose a methodology to regularize higher-order PCAs (i.e, PCAs with a small variance contribution) using an elastic net approach, which can be interpreted as as capturing the information stemming from the data once the low-order PCAs (i.e., that capture systematic risk) have been netted out. Kim, Korajczyk, and Neuhierl (2021) propose a methodology (building on the Projected-PCA of Fan, Liao, and Wang (2016)) for forming arbitrage portfolios that utilizes the information contained in firm characteristics for both abnormal returns and factor loadings. Lopez-Lira and Roussanov (2022) also find that empirically a substantial portion of expected returns is left unexplained by systematic risk factors even when their risk exposures are linear in stock characteristics, and construct an SA strategy with a good performance that uses machine learning methods to construct out-of-sample stock returns forecasts. Da, Nagel, and Xiu (2022) investigate the extent to which SA strategies, including those employing optimal machine learning techniques, can be successfully implemented when the signal provided by the alphas is rare and weak, which in turn imply a substantial adverse effect of the estimation error. He, He, Huang, and Zhou (2022) construct test of asset pricing models by exploiting the profitability of trading residuals based on high-minus-low sorting strategies of prior residual returns. Guijarro-Ordonez, Pelger, and Zanotti (2025) analyze what makes an SA strategy successfull, using machine-learning time-series solution that exploits price reversals at various data frequencies, demonstrating the substantial gain that can be obtained using U.S. equity markets data.[11]

Our novel estimation procedure, based on a penalized ML estimator, provides shrinkage-type estimators for all the parameters of the model. However, the penalization is derived from an *economic* constraint, i.e., the APT constraint, a reflection of asymptotic no-arbitrage, rather than from a statistical constraint. The APT constraint also serves as the condition required to resolve the lack of identification arising when estimating the SA and beta portolios jointly.

Finally, 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), Guidolin and Rinaldi (2009), and Christensen (2017). Also not the main focus of our methodology, we marginally contribute to this literature. In contrast to these papers, our work adopts as the *approximating model* the APT, decomposes the mean-variance portfolio into systematic-risk and SA components, and shows that treating misspecification in only the SA portfolio with the max-min approach, which we show is equivalent to imposing the APT restriction, leads to substantial improvement in empirical out-of-sample performance.

Finally, our work is also related to the extensive empirical literature that aims to improve the performance of mean-variance portfolios by imposing portfolio constraints or imposing mild (semi)-parametric features. 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

[10] Classical statistical arbitrage focused on designing algorithms for pair trading (e.g., see Gatev, Goetzmann, and Rouwenhorst (2006), and Avellaneda and Lee (2010) among many others).

[11] Some of the methodologies listed in the previous paragraph, about estimating systematic risk factors, can also be interpreted as belonging to the SA contributions listed here when referring to weak factors (see Lettau and Pelger (2020b,a), and Giglio et al. (2023)), as the latter can be intereprted as contributing to the residuals once the systematic risk factors are netted out.

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). 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. 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 rest of the paper is organized as follows. In Section 3, we specify the linear factor model for asset returns within the APT. Section 2.2 shows how to mitigate model misspecification in mean-variance portfolios by treating the alpha- and beta-portfolio components separately. We then explain how our methodology can be extended to address misspecification in a conditional asset-pricing setting. In Section 5, we demonstrate how our theoretical results lead to portfolio weights that achieve superior out-of-sample performance for two data sets. Finally, we conclude in Section 6. Proofs are collected in the appendix, while estimation details and additional results are collected in the Internet Appendix.

2 Portfolio Choice under the APT

In this section, we describe the framework used for our analysis, which is founded on the APT of Ross (1976, 1977) and formalized by Chamberlain (1983); Chamberlain and Rothschild (1983). The APT represents an ideal setting for our analysis of SA portfolios because non-zero alphas coexist with no-arbitrage. The classical APT has no observed factors. Instead, all factors are assumed to be latent, about which the investor can obtain full knowledge using PCA.

2.1 The APT Model

Let r_f denote the constant 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. Given an arbitrary portfolio strategy s with weights $\mathbf{w}^s = (w_1^s, w_2^s, \ldots, w_N^s)'$, and using $\mathbf{1}_N$ to denote an N-dimensional vector of ones, we define the associated portfolio return as

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

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

$$\mu^s \equiv \mathbb{E}(\mathbf{r}_t)' \mathbf{w}^s + r_f(1 - \mathbf{1}_N' \mathbf{w}^s), \qquad \sigma^s \equiv \sqrt{\mathbb{V}\text{ar}(r_t^s)}, \quad \text{and} \quad \text{SR}^s \equiv \frac{\mu^s - r_f}{\sigma^s}.$$

We now state our main assumptions. The first assumption of the APT says that the variance of asset returns is governed by a latent factor structure. It does *not* impose any conditions on the expected excess returns.[12]

[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."

Assumption 2.1 (Latent multifactor structure for asset returns)**.** *We assume that the N-dimensional vector $\mathbf{r}_t$ of gross asset returns can be characterized by*

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

where $\boldsymbol{\mu}$ is the $N \times 1$ vector of expected returns and $\mathbf{B}$ is the $N \times K$ full-rank matrix of risk exposures, with $K < N$. At any time t, the $K \times 1$ vector of latent systematic factors, $\mathbf{f}_t$, has mean $\mathbb{E}(\mathbf{f})$ and $K \times K$ covariance matrix $\boldsymbol{\Sigma}_f > 0$, and the $N \times 1$ vector of asset-specific innovations $\boldsymbol{\varepsilon}_t$ has zero mean and the $N \times N$ covariance matrix $\boldsymbol{\Sigma}_\varepsilon > 0$, both assumed to be positive definite. Moreover, $\boldsymbol{\varepsilon}_t$ and $\mathbf{f}_t$ are uncorrelated.

Assumption 2.1 implies that the variance-covariance matrix for asset returns is

$$\boldsymbol{\Sigma}_r \equiv \mathbb{E}\left((\mathbf{r}_t - \boldsymbol{\mu})(\mathbf{r}_t - \boldsymbol{\mu})'\right) = \mathbf{B}\,\boldsymbol{\Sigma}_f\,\mathbf{B}' + \boldsymbol{\Sigma}_\varepsilon.$$

The second assumption of the APT rules out asymptotic no-arbitrage.

Assumption 2.2 (No-asymptotic arbitrage)**.** *As $N \to \infty$, there is no sequence of portfolios $\mathbf{w}^*$ for which, along some subsequence N^* of the assets, $\mathbf{r}_t^*$, with expected returns $\boldsymbol{\mu}^*$,*

$$\mathbb{V}\mathrm{ar}(\mathbf{r}_t^{*\prime}\mathbf{w}^*) \to 0 \quad \textit{and} \quad (\boldsymbol{\mu}^* - r_f\mathbf{1}_N)'\mathbf{w}^* \geq \delta > 0 \quad \textit{for all} \quad N^* \quad \textit{and some constant} \quad \delta,$$

where $\mathbf{1}_N$ is the N-dimensional vector of ones.

Combining these two assumptions leads to an implication on expected excess returns, which represents the single most important result of the APT. That is, when Assumptions 2.1 and 2.2 hold, with the largest eigenvalue[13] of $(\mathbf{B}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B})^{-1}$ satisfying $g_{1K}((\mathbf{B}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B})^{-1}) \to 0$ as $N \to \infty$, then there exists some positive number δ_{apt} and a unique vector of risk premia $\boldsymbol{\lambda}$ (see (Chamberlain, 1983; Huberman, 1982; Ingersoll, 1984) among others) such that

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

where $\boldsymbol{\alpha}$ satisfies the so-called APT constraint[14]

$$\sup_N \boldsymbol{\alpha}'\boldsymbol{\Sigma}_\varepsilon^{-1}\boldsymbol{\alpha} \leq \delta_{\text{apt}} < \infty. \tag{2}$$

In (1), $\boldsymbol{\alpha}$ represents the deviation from exact pricing, the vector of risk premia $\boldsymbol{\lambda}$ of the risk factors $\mathbf{f}_t$ is the limit of $\left(\mathbf{B}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B}\right)^{-1}\mathbf{B}'\boldsymbol{\Sigma}_\varepsilon^{-1}(\boldsymbol{\mu} - r_f\mathbf{1}_N)$ as $N \to \infty$, which exists and is well defined under Assumptions 2.1 and 2.2 (see Ingersoll (1984)). Typically, $\boldsymbol{\alpha}$ has been interpreted in the existing literature as a pricing error; i.e., the deviation from exact pricing.[15]

[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 $g_{MM}(\mathbf{A})$.

[14] The APT theory is silent about the magnitude of δ_{apt} in (2). 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 Section 4.

[15] A common misperception is that the APT implies a multifactor structure for expected excess asset returns with zero pricing errors. However, the multifactor structure is an *assumption* (spelled out in Assumption 2.1), not a *result*, of the APT. More importantly, the elements of $\boldsymbol{\alpha}$ do not need to be zero. Instead, the *key* theoretical prediction of the APT is that exact pricing can fail, i.e., $\boldsymbol{\alpha}$ can be nonzero, as long as some discipline is imposed on its magnitude, which is exactly what the no-arbitrage condition in (2) does.

As is typical in the APT literature (Chamberlain, 1983, 1987; Huberman, 1982; Ingersoll, 1984) and the portfolio-choice literature (Daniel et al., 2020; Kelly, Pruitt, and Su, 2020; Kim et al., 2021), one could assume that $\boldsymbol{\alpha}$ and $\mathbf{B}$ in (1) are orthogonal across assets. As we demonstrate in Section **??** of the Internet Appendix, this would be without loss of generality for our results on portfolio choice: even when $\boldsymbol{\alpha}$ and $\mathbf{B}$ are cross-sectionally correlated, the portfolio returns corresponding to $\boldsymbol{\alpha}$ and $\mathbf{B}$, which we construct below, are orthogonal (asymptotically in N), because of linear projection arguments embedded in the analytics of mean-variance portfolios.

For expositional ease, in the rest of this section, we will focus on the special case of a one-factor APT model, i.e., $K = 1$, with risk exposures $\boldsymbol{\beta}$ and risk premium λ, respectively, and where the asset-specific shocks are uncorrelated across assets and with have a constant variance, i.e., $\boldsymbol{\Sigma}_\varepsilon = \sigma_\varepsilon^2 \mathbb{I}_N$, where $\mathbb{I}_N$ denotes the $N \times N$ identity matrix. In Section **??** of the Appendix, the proofs for all our results are provided for the general multifactor case where $K \geq 1$ and $\boldsymbol{\Sigma}_\varepsilon$ is not restricted to be $\sigma_\varepsilon^2 \mathbb{I}_N$. In Section 3, we will complement the results of this section by studying the case in which the investor, instead of assuming that there are K latent factors, specifies a particular candidate beta-pricing model with K_{can} systematic risk factors, but wrongly assumes $\boldsymbol{\alpha} = \mathbf{0}_N$ and omits K_{mis} systematic risk factors; that is, $K = K_{\text{can}} + K_{\text{mis}}$.[16]

For this special case of the APT model, where there is only one latent risk factor, excess returns satisfy

$$\begin{aligned} \mathbf{r}_t - r_f \mathbf{1}_N &= \boldsymbol{\alpha} + \boldsymbol{\beta}\lambda + \boldsymbol{\beta}(f_t - \mathbb{E}(f)) + \boldsymbol{\varepsilon}_t, \quad \text{with} \\ \boldsymbol{\Sigma}_r &= \sigma_f^2 \boldsymbol{\beta}\boldsymbol{\beta}' + \sigma_\varepsilon^2 \mathbb{I}_N, \qquad \boldsymbol{\mu} - r_f \mathbf{1}_N = \boldsymbol{\alpha} + \boldsymbol{\beta}\lambda, \qquad \sigma_f^2 \equiv \mathbb{V}\text{ar}(f), \end{aligned} \tag{3}$$

and the APT restriction in (2) simplifies to

$$\sup_N \frac{\boldsymbol{\alpha}'\boldsymbol{\alpha}}{\sigma_\varepsilon^2} \leq \delta_{\text{apt}} < \infty. \tag{4}$$

2.2 Mean-Variance Analysis with Statistical Arbitrage

This section contains our main theoretical results on the importance of alpha risk. In particular, we show how, under the APT, the mean-variance efficient (MVE) portfolio can be decomposed into two portfolios: the *SA portfolio*, which is *factor-neutral* and a function of the asset-specific components of returns, in particular, $\boldsymbol{\alpha}$ and σ_ε^2, and the *systematic-risk portfolio* which depends on the risk-premium, variance of the systematic risk factor, and its risk exposures. We then derive the following results for these two portfolios: (i) the SA portfolio and beta portfolio are *inefficient*, but together they span the mean-variance efficient frontier. (ii) the SA portfolio and beta portfolio are *orthogonal* to one another; this result is true even if $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are cross-sectionally correlated. Even if the contribution of $\boldsymbol{\alpha}$ to the overall expected return of each asset is small, (iii) the weights of the SA portfolio *dominate* the corresponding weights of the systematic-risk portfolio and (iv) in terms of performance, measured by their contribution to the overall Sharpe ratio of the MVE portfolio, the SA portfolio *dominates* the systematic-risk portfolio. (v) Finally, surprisingly, the GMV portfolio satisfies all the above properties even though it is functionally independent of expected returns and, therefore, of $\boldsymbol{\alpha}$. In other words, the importance of SA portfolios is much broader than typically thought, being a fundamental feature of all MVE portfolios, and in particular not just of portfolios that depend on $\boldsymbol{\alpha}$.

[16] We denote by $\mathbf{0}_N$ the N-dimensional vector of zeros.

2.2.1 MVE: Portfolios with Expected Returns

The MVE portfolio in the presence of a risk-free asset is defined as the solution to the following optimization problem:

$$\mathbf{w}^{\text{mve}} \equiv \underset{\mathbf{w}}{\text{argmax}} \left((\mathbf{w}'\boldsymbol{\mu} + (1 - \mathbf{w}'\mathbf{1}_N)r_f) - \frac{\gamma}{2}\mathbf{w}'\boldsymbol{\Sigma}_r\mathbf{w} \right), \tag{5}$$

where $0 < \gamma < \infty$ is the coefficient of risk aversion, $\mathbf{w}^{\text{mve}} = (w_1^{\text{mve}}, \ldots, w_N^{\text{mve}})'$ is the vector of MVE portfolio weights in the N risky assets, and the investment in the risk-free asset is given by $1 - \mathbf{1}_N'\mathbf{w}^{\text{mve}}$. The well-known solution to the optimization problem in (5) is

$$\mathbf{w}^{\text{mve}} \equiv \frac{1}{\gamma}\boldsymbol{\Sigma}_r^{-1}(\boldsymbol{\mu} - r_f\mathbf{1}_N). \tag{6}$$

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

The following proposition establishes the relations that exist between the MVE portfolio, $\mathbf{w}^{\text{mve}}$, and the two *inefficient* portfolios, the SA portfolio $\mathbf{w}^{\alpha}$ and the systematic-risk portfolio $\mathbf{w}^{\beta}$. Throughout the analysis, when we write $\mathbf{w}^x = \mathbf{w}^y + \mathcal{O}(1)$, we mean that, as N increases, the returns of portfolios x and y have the same mean, variance, and Sharpe ratio.[17]

Proposition 2.1 (Three-Fund Separation with SA Portfolio - MVE Case)**.** *Suppose that the vector of asset returns, $\mathbf{r}_t$, satisfies Assumptions 2.1 and 2.2, $\boldsymbol{\alpha} \neq \mathbf{0}_N$, and $\lambda \neq 0$,*[18] *that $\boldsymbol{\beta}'\boldsymbol{\beta}/N$ is converging to a finite constant, with the elements of $\boldsymbol{\beta}$ being uniformly bounded.*[19] *Then, for any $N > K = 1$, there exist two* inefficient *portfolios*

$$\mathbf{w}^{\alpha} \equiv \left(\frac{1}{\sigma_\varepsilon^2}\right)\left(\boldsymbol{\alpha} - \tfrac{(\boldsymbol{\beta}'\boldsymbol{\alpha})}{(\boldsymbol{\beta}'\boldsymbol{\beta})}\boldsymbol{\beta}\right) + \mathcal{O}(1) \qquad \textit{and} \qquad \mathbf{w}^{\beta} \equiv \left(\frac{\lambda}{\sigma_f^2}\right)\frac{\boldsymbol{\beta}}{(\boldsymbol{\beta}'\boldsymbol{\beta})} + \mathcal{O}(1) \tag{7}$$

where one portfolio is the unique minimum-variance orthogonal *portfolio to the other, i.e.,*[20]

$$\mathbf{w}^{\alpha} \equiv \underset{\substack{\mathbf{w} \ s.t. \\ \mathbf{w}'\boldsymbol{\Sigma}_r\mathbf{w}^{\beta}\to 0}}{\text{argmin}} \ \mathbf{w}'\boldsymbol{\Sigma}_r\mathbf{w} \qquad \textit{and} \qquad \mathbf{w}^{\beta} \equiv \underset{\substack{\mathbf{w} \ s.t. \\ \mathbf{w}'\boldsymbol{\Sigma}_r\mathbf{w}^{\alpha}\to 0}}{\text{argmin}} \ \mathbf{w}'\boldsymbol{\Sigma}_r\mathbf{w}, \tag{8}$$

and together with the risk-free asset they span *the mean-variance efficient frontier, i.e.,*[21]

$$\mathbf{w}^{mve} = \frac{1}{\gamma}\left(\mathbf{w}^{\alpha} + \mathbf{w}^{\beta}\right).$$

Remark 2.1.1. Defining

$$\boldsymbol{\Sigma}_\varepsilon^{+} \equiv \left(\frac{1}{\sigma_\varepsilon^2}\right)\left(\mathbb{I}_N - \frac{\boldsymbol{\beta}\boldsymbol{\beta}'}{(\boldsymbol{\beta}'\boldsymbol{\beta})}\right),$$

[17] A more detailed explanation of this notation is given at the beginning of Appendix **??**.

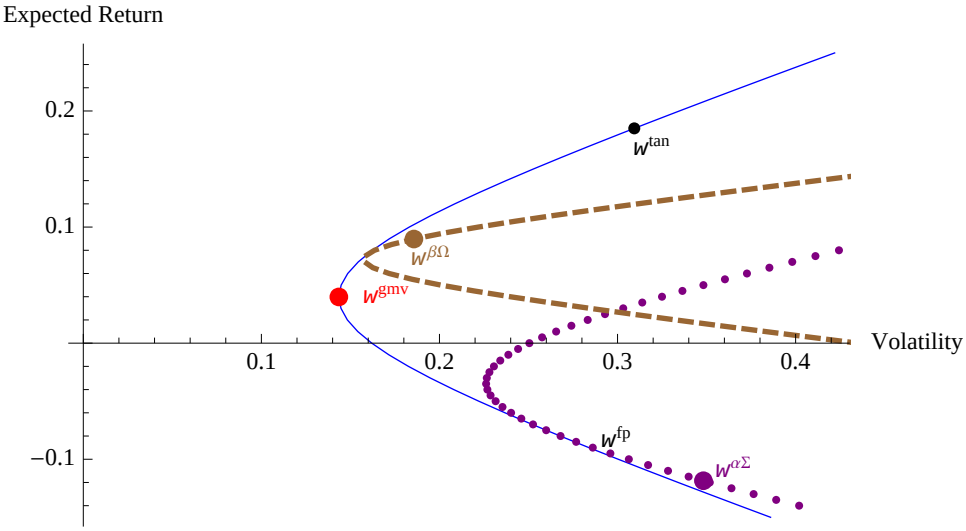
[18] Proposition 2.1 is trivial when either $\boldsymbol{\alpha} = \mathbf{0}_N$ or $\lambda = 0$, so we rule these cases out without loss of generality.

[19] We require that $\sup_{1\leq i\leq N}|\beta_i| \leq \delta < \infty$ for some arbitrary positive scalar δ.

[20] In practice, one does not need to solve the mathematical problems in (8) because $\mathbf{w}^{\alpha}$ and $\mathbf{w}^{\beta}$ portfolios are obtained simply by relying on the APT decomposition of the total mean return. This is shown in Proposition **??** 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] Note that it is straightforward to scale the weights of the $\mathbf{w}^{\alpha}$ and $\mathbf{w}^{\beta}$ portfolios so that the weights of each portfolio sum to one.

Figure 1: MVE portfolio and its decomposition in absence of the risk-free asset
In this figure, we plot the MVE portfolio, $\mathbf{w}^{\text{mve}}$, and its decomposition into two portfolios, $\mathbf{w}^{\alpha}$ and $\mathbf{w}^{\beta}$.



the SA portfolio (net of the small-order term $\mathcal{O}(1)$) can be written as $\mathbf{w}^{\alpha} = \mathbf{\Sigma}_{\varepsilon}^{+}\boldsymbol{\alpha}$. Because $\mathbf{\Sigma}_{\varepsilon}^{+}\boldsymbol{\beta} = \mathbf{0}_N$, it follows that the SA portfolio $\mathbf{w}^{\alpha}$ completely diversifies away the systematic risk for any N, and thus, is beta-factor neutral, namely it is *functionally independent* of the risk premium, λ, and the factor's variance, σ_f^2. On the other hand, the systematic-risk portfolio, which we denote as the "beta portfolio" $\mathbf{w}^{\beta}$, depends (asymptotically) only on the factor exposures $\boldsymbol{\beta}$, on the factor risk premium λ, and the factor variance σ_f^2, but is functionally independent of the compensation for alpha risk, $\boldsymbol{\alpha}$, and the idiosyncratic variance σ_{ε}^2.

Remark 2.1.2. The (asymptotic) orthogonality property $(\mathbf{w}^{\alpha})'\mathbf{\Sigma}_{\varepsilon}\mathbf{w}^{\beta} \to 0$, conditional on the systematic risk factor f_t, implies that the two portfolios $\mathbf{w}^{\alpha}$ and $\mathbf{w}^{\beta}$ are uncorrelated as N gets large. This property is true *regardless* of whether $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in (1) are cross-sectionally uncorrelated, as shown in Section **??** of the Internet Appendix.

We now characterize the asymptotic (in N) behavior of the weights of the SA and beta portfolios and their performance in terms of the Sharpe ratio. Hereafter, we use the notation $x = \mathcal{O}(y)$ to denote that $|x|/y$ is bounded above for generic non-zero quantities x, y.

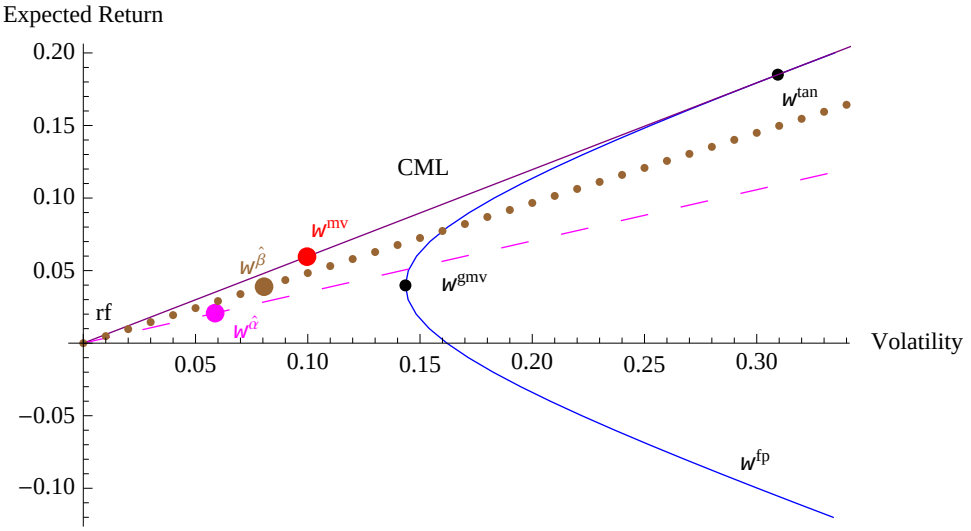
Proposition 2.2 (Dominance of the SA Portfolio - MVE case)**.** *Under the assumptions of Proposition 2.1 with* $\boldsymbol{\alpha} \neq \mathbf{0}_N$*:*

(i) The weights of the SA and beta portfolios satisfy

$$|w_i^{\alpha}| = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right) \quad \textit{and} \quad |w_i^{\beta}| = \mathcal{O}\left(\frac{1}{N}\right), \textit{implying} \quad |w_i^{\beta}|/|w_i^{\alpha}| \to 0.$$

(ii) The Sharpe ratios SR^{α} *and* SR^{β} *of the SA and beta portfolios satisfy, respectively,* $\text{SR}^{\alpha} = (\boldsymbol{\alpha}'\mathbf{\Sigma}_{\varepsilon}^{+}\boldsymbol{\alpha})^{\frac{1}{2}} + \mathcal{O}(1)$ *and* $\text{SR}^{\beta} = (\lambda^2/\sigma_f^2)^{\frac{1}{2}} + \mathcal{O}(1)$*, such that, for some constants* $0 < \underline{c}$ *and* $\bar{c} < 1$*,*

Figure 2: MVE portfolio and its decomposition with the risk-free asset
In this figure, we plot the MVE portfolio, $\mathbf{w}^{\text{mve}}$, and its decomposition into two portfolios, $\mathbf{w}^{\alpha}$ and $\mathbf{w}^{\beta}$.



the attributions of the SA and systematic-risk portfolios to the Sharpe ratio of the MVE portfolio satisfy

$$\bar{c} \leq \frac{(\text{SR}^{\alpha})^2}{(\text{SR}^{\alpha})^2 + (\text{SR}^{\beta})^2} \quad \textit{and} \quad \frac{(\text{SR}^{\beta})^2}{(\text{SR}^{\alpha})^2 + (\text{SR}^{\beta})^2} \leq \underline{c},$$

and

$$(\text{SR}^{mve})^2 - \left((\text{SR}^{\alpha})^2 + (\text{SR}^{\beta})^2\right) \to 0.$$

Remark 2.2.1. To understand the intuition for the dominance of the weights of the SA portfolio if $\boldsymbol{\alpha} \neq \mathbf{0}_N$, recall that mean-variance portfolio optimization aims to maximize the portfolio Sharpe ratio. There are two sources of risk in the APT: systematic risk and alpha risk. The exposure of $\mathbf{w}^{\alpha}$ to systematic risk is zero—irrespective of the rate at which the weights decrease—because of the orthogonality of $\mathbf{w}^{\alpha}$ to $\boldsymbol{\beta}$. Regarding exposure to alpha risk, the elements of $\mathbf{w}^{\alpha}$ cannot decrease faster than $1/\sqrt{N}$ because then the alpha risk of the portfolio goes to zero; however, the alpha risk of this portfolio coincides with its Sharpe ratio, implying that its Sharpe ratio would also go to zero. On the other hand, the APT constrain in (4) 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 $\mathbf{w}^{\alpha}$ portfolio.[22]

Remark 2.2.2. For the weights of the systematic-risk portfolio, $\mathbf{w}^{\beta}$, if the weights decrease at a rate slower than $1/N$, then the systematic exposure explodes because the factors are pervasive. On

[22]The $\mathbf{w}^{\alpha}$ 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 $\mathbf{w}^{\alpha\prime}\mathbf{w}^{\alpha}$, which is the same notion adopted in Chamberlain (1983), is always bounded, whereas $\mathbf{w}^{\beta\prime}\mathbf{w}^{\beta}$ always converges to zero. Moreover, the sum of the portfolio weights $|\mathbf{1}_N'\mathbf{w}^{\alpha}|$ can diverge to infinity, whereas $|\mathbf{1}_N'\mathbf{w}^{\beta}|$ is always bounded.

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 $\mathbf{w}^{\beta}$ portfolio. Observe that the rate $1/N$ makes the $\mathbf{w}^{\beta}$ portfolio diversified also with respect to alpha risk.

Remark 2.2.3. The expression for the SR^{β} in the proposition shows that it is independent of N. In contrast, SR^{α} increases monotonically in N, so that for sufficiently large N it can be the dominant component of $\mathrm{SR}^{\mathrm{mve}}$. For example, in the setting where $\boldsymbol{\alpha}'\boldsymbol{\beta}=0$, adding new assets does not change the Sharpe ratio of the systematic-risk portfolio but, because new independent bets are being added, increases the Sharpe ratio of the SA portfolio.

Remark 2.2.4. The result that $\mathrm{SR}^{\alpha}=(\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\varepsilon}^{+}\boldsymbol{\alpha})^{\frac{1}{2}}+\mathcal{O}(1)$ also leads to a new interpretation of the APT restriction in (4), which says that no-arbitrage implies boundedness of the SA portfolio's Sharpe because $\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\varepsilon}^{+}\boldsymbol{\alpha}\leq\boldsymbol{\alpha}'\boldsymbol{\alpha}/\sigma_{\varepsilon}^{2}$ (ignoring small order terms). Moreover, the bound on the square of the Sharpe ratio of the SA portfolio can also be seen as providing a theoretical rationalization for the "no-good-deal bound" in Cochrane and Saá-Requejo (2001).[23]

Example 2.2.1 (MVE Portfolio: Conditional Factor Asset Pricing Models (Kelly et al. (2019), Kim et al. (2021), Guijarro-Ordonez et al. (2025)))**.** *Although our theoretical analysis is presented in a traditional APT setting, with constant parameters (in particular, constant alphas and betas), it can be readily extended to a conditional APT setting, making it for suitable for applications with individual securities (see Zaffaroni (2025)[Proposition XXX] for a proof of the conditional APT), by replacing the single-factor version of Assumption 2.1, i.e., (3), with*

$$\mathbf{r}_t-r_{ft}\mathbf{1}_N=\boldsymbol{\alpha}_{t-1}+\boldsymbol{\beta}_{t-1}\lambda_{t-1}+\boldsymbol{\beta}(f_t-\mathbb{E}_{t-1}(f))+\boldsymbol{\varepsilon}_t,\quad \textit{where}$$
$$\boldsymbol{\Sigma}_{r,t-1}=\sigma_{f,t-1}^{2}\boldsymbol{\beta}_{t-1}\boldsymbol{\beta}_{t-1}'+\sigma_{\varepsilon,t-1}^{2}\mathbb{I}_N,\qquad \boldsymbol{\mu}_{t-1}-r_{ft}\mathbf{1}_N=\boldsymbol{\alpha}_{t-1}+\boldsymbol{\beta}_{t-1}\lambda_{t-1},\qquad \sigma_{f,t-1}^{2}\equiv\mathbb{V}\mathrm{ar}_{\text{t-1}}(f),$$

with the APT restriction in (2) becoming $\sup_N \frac{\boldsymbol{\alpha}_{t-1}'\boldsymbol{\alpha}_{t-1}}{\sigma_{\varepsilon,t-1}^{2}}\leq\delta_{apt,t-1}<\infty$*, for some (time-dependent) quantity* $\delta_{apt,t-1}$*, and replacing Assumption 2.2 by assuming that, as* $N\to\infty$*, there is no sequence of portfolios for which, along some subsequence* $\tilde{N}$*,*

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

Then, conditional versions of Propositions 2.1 and 2.2 hold, yielding

$$\mathbf{w}_{t-1}^{\alpha}\equiv\left(\frac{1}{\sigma_{\varepsilon,t-1}^{2}}\right)\left(\boldsymbol{\alpha}_{t-1}-\tfrac{(\boldsymbol{\beta}_{t-1}'\boldsymbol{\alpha}_{t-1})}{(\boldsymbol{\beta}_{t-1}'\boldsymbol{\beta}_{t-1})}\boldsymbol{\beta}_{t-1}\right)+\mathcal{O}(1)\qquad \textit{and}\qquad \mathbf{w}_{t-1}^{\beta}\equiv\left(\frac{\lambda_{t-1}}{\sigma_{f,t-1}^{2}}\right)\frac{\boldsymbol{\beta}_{t-1}}{(\boldsymbol{\beta}_{t-1}'\boldsymbol{\beta}_{t-1})}+\mathcal{O}(1),$$

and $\mathrm{SR}_{t-1}^{\alpha}=\frac{1}{\sigma_{\varepsilon,t-1}}\left(\boldsymbol{\alpha}_{t-1}'\left(\mathbb{I}_N-\frac{\boldsymbol{\beta}_{t-1}\boldsymbol{\beta}_{t-1}'}{(\boldsymbol{\beta}_{t-1}'\boldsymbol{\beta}_{t-1})}\right)\boldsymbol{\alpha}_{t-1}\right)^{\frac{1}{2}}+\mathcal{O}(1)$ *and* $\mathrm{SR}_{t-1}^{\beta}=(\lambda_{t-1}^{2}/\sigma_{f,t-1}^{2})^{\frac{1}{2}}+\mathcal{O}(1)$*, with* $(\mathrm{SR}_{t-1}^{mve})^2-\left((\mathrm{SR}_{t-1}^{\alpha})^2+(\mathrm{SR}_{t-1}^{\beta})^2\right)\to 0$*. such that, for some (time-dependent) constants* $0<\underline{c}_{t-1}$ *and* $\bar{c}_{t-1}<1$*, the attributions of the SA and systematic-risk portfolios to the Sharpe ratio of the MVE portfolio satisfy* $\bar{c}_{t-1}\leq\frac{(\mathrm{SR}_{t-1}^{\alpha})^2}{(\mathrm{SR}_{t-1}^{\alpha})^2+(\mathrm{SR}_{t-1}^{\beta})^2}$ *and* $\frac{(\mathrm{SR}_{t-1}^{\beta})^2}{(\mathrm{SR}_{t-1}^{\alpha})^2+(\mathrm{SR}_{t-1}^{\beta})^2}\leq\underline{c}_{t-1}$*. Therefore, our results on SA apply to the conditional factor asset pricing models of Kelly et al. (2019), Kim et al. (2021), and Guijarro-Ordonez et al. (2025), among others, when one imposes the required*

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

assumptions outlined above.[24] *In particular, Kelly et al. (2019) and Kim et al. (2021) specify (among other things) that alphas are linear in observed characteristics, i.e.,* $\boldsymbol{\alpha}_{t-1} = \mathbf{Z}_{t-1}\mathbf{a}$ *for an* $N \times K_z$ *matrix* $\mathbf{Z}_{t-1}$ *of observed characteristics and a constant* $K_z \times 1$ *vector of parameters* $\mathbf{a}$*. Guijarro-Ordonez et al. (2025) considers a semiparametric specification of the model, and of the SA portfolio weights, such as* $\mathbf{w}^{\alpha}_{t-1} = \mathbf{w}^{\alpha}(\boldsymbol{\theta}(\boldsymbol{\varepsilon}^{L}_{t-1}))$*, where* $\boldsymbol{\theta}(\cdot)$ *is the non-parametric signal function (the function that summarizes all the relevant information to construct the trading strategy),* $\mathbf{w}^{\alpha}(\cdot)$ *is the allocation function, and* $\boldsymbol{\varepsilon}^{L}_{t-1}$ *contains the lagged* $N \times 1$ *vectors of the idiosyncratic error from* $t-1$ *up to* $t-L$*, for some* $L \geq 1$*.*[25] *Guijarro-Ordonez et al. (2025) demonstrate the performance ability stemming from estimating* $\boldsymbol{\theta}(\cdot)$ *by means of a sophisticated deep neural network architecture.*

Example 2.2.2 (MVE Portfolio: Uncorrelated Alpha and Beta (Da et al. (2022), Example 1))**.** *Assume the one-factor model in (3), with* α_i *distributed iid and satisfying a symmetric distribution around 0,*

$$\alpha_i \sim \begin{cases} \mu & \text{with prob. } \pi/2, \\ -\mu & \text{with prob. } \pi/2, \\ 0 & \text{with prob. } 1-\pi, \end{cases}$$

for some $0 \leq \pi \leq 1, \mu \equiv AN^{-\psi} > 0$*, for some constants* $A > 0$*,* $\psi > 0$*, and* $N^{-1}\sum_{i=1}^{N}\beta_i^2 \to \mu_{\beta^2}$*, with* $0 < \mu_{\beta^2} < \infty$*, with* $\max_i |\beta_i| < \delta < \infty$*, for an arbitrary positive constant* δ*. Asymptotic no arbitrage, i.e., Assumption 2.2, requires* $\psi \geq 1/2$ *although to rule out a zero Sharpe ratio, the value of* ψ *cannot be strictly bigger than* $1/2$*, implying that the only economically meaningful value is* $\psi = 1/2$*, as discussed also in Da et al. (2022).*[26]

Then, the SA portfolio weights are

$$w_i^{\alpha} = \frac{1}{\sigma_{\varepsilon}^2}\left(\alpha_i - \beta_i \frac{\sum_{j=1}\beta_j\alpha_j}{\sum_{j=1}\beta_j^2}\right) + \mathcal{O}_p(1) = \frac{1}{\sigma_{\varepsilon}^2}\left(\alpha_i + \mathcal{O}_p\left(\beta_i \frac{A\pi^{\frac{1}{2}}}{N^{\frac{1}{2}}\mu_{\beta^2}^{\frac{1}{2}}}\right)\right) + \mathcal{O}_p(1) \quad = \frac{\alpha_i}{\sigma_{\varepsilon}^2} + \mathcal{O}_p(1),$$

with squared Sharpe ratio $(SR^{\alpha})^2 = A^2\pi/\sigma_{\varepsilon}^2 + \mathcal{O}_p(1)$*, whereas for the beta portfolio,* $(\mathrm{SR}^{\beta})^2 = (\lambda^2/\sigma_f^2) + \mathcal{O}(1)$*.*[27] *Just like* λ^2 *is the square of the risk-premium for the single factor and* σ_f^2 *is its variance,* $N^{-1}A^2\pi$ *is the square of the risk-premium for bearing alpha risk (i.e., a squared risk premium of* $N^{-1}A^2$ *with probability* π *and zero with probability* $1-\pi$*), with* $N^{-1}\sigma_{\varepsilon}^2$ *being its variance. Obviously, as the* α_i *and* β_i *are cross-sectionally uncorrelated, the weights of the SA and beta portfolios are orthogonal.*

Then, for any constants $0 < \underline{c} \leq \bar{c} < 1$*, when* $A^2/\sigma_{\varepsilon}^2$ *is large enough, the contribution of the SA and beta portfolios to the MVE Sharpe ratio satisfy*

$$\bar{c} \leq \frac{(\mathrm{SR}^{\alpha})^2}{(\mathrm{SR}^{\alpha})^2 + (\mathrm{SR}^{\beta})^2} = \frac{A^2\pi/\sigma_{\varepsilon}^2}{A^2\pi/\sigma_{\varepsilon}^2 + \lambda^2/\sigma_f^2} \quad \text{and} \quad \frac{(\mathrm{SR}^{\beta})^2}{(\mathrm{SR}^{\alpha})^2 + (\mathrm{SR}^{\beta})^2} = \frac{\lambda^2/\sigma_f^2}{A^2\pi/\sigma_{\varepsilon}^2 + \lambda^2/\sigma_f^2} \leq \underline{c}.$$

We now consider the same setting of a single-factor model, but where α_i and β_i are allowed to be correlated across assets; a detailed description of this example is provided in **??**. This example shows that the dominance result of the SA portfolio holds also for the case of correlated α_i and β_i.

[24] This might require a suitable modification of their models. For example, Kim et al. (2021) do do not impose boundedness of $(\mathrm{SR}^{\alpha}_{t-1})^2$, hence allowig for arbitrage opportunities, although this can be easily imposed.

[25] We adopt a different notation from the one used in Guijarro-Ordonez et al. (2025).

[26] One can generalize this model by allowing that the α_i are (arbitrarily) large for a *finite* number of assets without changing the implications of the model.

[27] The notation $\mathcal{o}_p(1)$ denotes that $|x|/y$ is bounded above in probability for a generic random-variable x and a positive constant y.

Example 2.2.3 (MVE Portfolio: Correlated Alpha and Beta (Section **??** of the Internet Appendix))**.** *Assume the single-factor model described in (3). Set the total number of assets to be $N = 2n+1$ for some integer $n \geq 0$, and assume that $\beta_i = 1+b(i-(n+1))/n$ for $i = 1, \ldots, 2n+1$, where $0 \leq b < 1$, and $\alpha_i = -a(i-(n+1))/n^{\frac{3}{2}}$ for $i = 1, \ldots, 2n+1$. 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 α_i and β_i have perfect negative (deterministic) correlation, in the sense that their sample correlation converges to -1.*[28]

Defining $\rho \equiv -(ab/\sqrt{n})/(3+b^2)$,[29] *the SA portfolio weights satisfy*

$$w_i^{\alpha} = \frac{1}{\sigma_\varepsilon^2}(\alpha_i - \beta_i \rho) + \mathcal{O}(1) = \mathcal{O}(\frac{1}{\sqrt{N}}),$$

with Sharpe ratio $(SR^{\alpha})^2 = 2a^2/\sigma_\varepsilon^2(3+b^2) + \mathcal{O}(1)$. The systematic-risk portfolio weights are

$$w_i^{\beta} = \frac{\beta_i}{2n(1+b^2/3)}\frac{\lambda}{\sigma_f^2}, + \mathcal{O}(1) = \mathcal{O}(\frac{1}{N}),$$

with Sharpe ratio $(SR^{\beta})^2 = \lambda^2/\sigma_f^2 + \mathcal{O}(1)$. Moreover, even though α_i and β_i are perfectly negatively correlated, the weights of the SA and beta portfolios are orthogonal.

Then, for any constants $0 < \underline{c} \leq \bar{c} < 1$, when a^2/σ_ε^2 is large enough, the contributions of the SA and beta portfolios to the MVE Sharpe ratio of the satisfy

$$\bar{c} \leq \frac{(\mathrm{SR}^{\alpha})^2}{(\mathrm{SR}^{\alpha})^2 + (\mathrm{SR}^{\beta})^2} = \frac{2a^2/\sigma_\varepsilon^2(3+b^2)}{2a^2/\sigma_\varepsilon^2(3+b^2) + \lambda^2/\sigma_f^2} \quad \textit{and}$$
$$\frac{(\mathrm{SR}^{\beta})^2}{(\mathrm{SR}^{\alpha})^2 + (\mathrm{SR}^{\beta})^2} = \frac{\lambda^2/\sigma_f^2}{2a^2/\sigma_\varepsilon^2(3+b^2) + \lambda^2/\sigma_f^2} \leq \underline{c}.$$

2.2.2 GMV: Portfolios without Expected Returns

The second portfolio strategy whose weights and returns we investigate is the *global-minimum-variance (GMV) portfolio* in the presence of a risk-free asset. In this section, we show that the GMV portfolio, despite being independent of expected returns, has the same properties as the ones established for the mean-variance efficient portfolios and its decomposition into alpha and beta components. This demonstrates that the properties of the SA portfolio weights are not simply driven by the compensation for alpha risk.

The GMV portfolio is defined by the solution to the following optimization problem:

$$\mathbf{w}^{\mathrm{gmv}} \equiv \underset{\mathbf{w}}{\operatorname{argmin}}\, \mathbf{w}'\mathbf{V}\mathbf{w} \quad \text{such that} \quad \mathbf{w}'\mathbf{1}_N = 1, \tag{9}$$

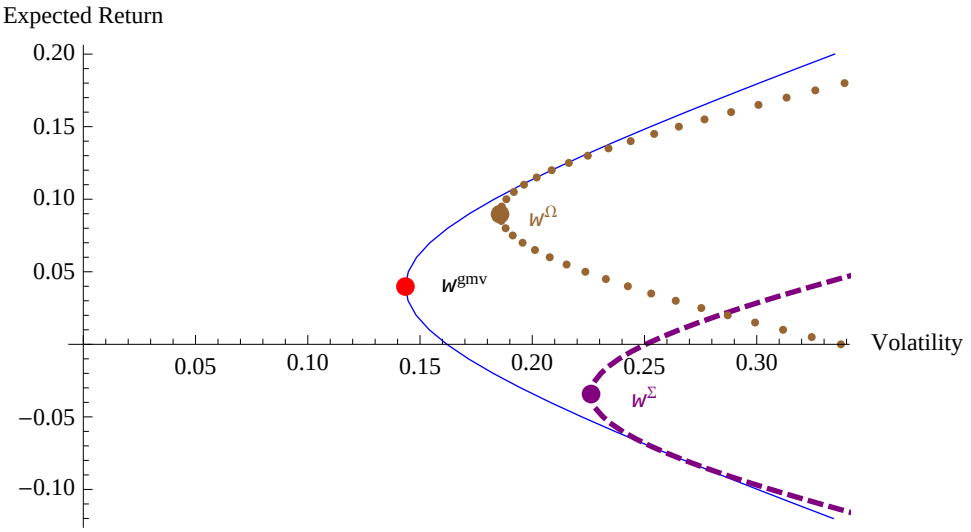
where $\mathbf{w}^{\mathrm{gmv}} = (w_1^{\mathrm{gmv}}, \ldots, w_N^{\mathrm{gmv}})'$ are the weights in the risky assets given by the well-known solution to the optimization problem in (9)

$$\mathbf{w}^{\mathrm{gmv}} \equiv \frac{\mathbf{V}^{-1}\mathbf{1}_N}{\mathbf{1}_N'\mathbf{V}^{-1}\mathbf{1}_N}. \tag{10}$$

[28] 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 $\alpha_3 = -a$.

[29] The quantity ρ is asymptotically equivalent to the ordinary-least squares (OLS) coefficient of α_i on the β_i without the intercept; this quantity is key for making the SA portfolio weights orthogonal to the beta portfolio weights.

Figure 3: Global-minimum-variance portfolio and its decomposition
In this figure, we plot the global-minimum-variance portfolio, $\mathbf{w}^{\text{gmv}}$, and its decomposition into two portfolios, $\mathbf{w}^{f}$ and $\mathbf{w}^{\varepsilon}$.



The return on portfolio $\mathbf{w}^{\text{gmv}}$ has Sharpe ratio $\text{SR}^{\text{gmv}} = \left(\mathbf{1}_N'\mathbf{V}^{-1}(\boldsymbol{\alpha}+\boldsymbol{\beta}\lambda)\right)/\left(\mathbf{1}_N'\mathbf{V}^{-1}\mathbf{1}_N\right)^{1/2}$ with mean and standard deviation given by, respectively, $\mu^{\text{gmv}} - r_f = \left(\mathbf{1}_N'\mathbf{V}^{-1}(\boldsymbol{\alpha}+\boldsymbol{\beta}\lambda)\right)/\left(\mathbf{1}_N'\mathbf{V}^{-1}\mathbf{1}_N\right)$ and $\sigma^{\text{gmv}} = 1/\left(\mathbf{1}_N'\mathbf{V}^{-1}\mathbf{1}_N\right)^{1/2}$.

The following proposition establishes the relations that exist between the GMV portfolio, $\mathbf{w}^{\text{gmv}}$, and two *inefficient* portfolios, the SA-GMV portfolio $\mathbf{w}^{\varepsilon}$ and the beta-GMV portfolio $\mathbf{w}^{f}$, that depend on the asset-specific shock (σ_ε^2) and factor (σ_f^2) variances, respectively.

Proposition 2.3 (Two-Fund Separation with SA Portfolio - GMV Case)**.** *Suppose that the vector of asset returns, $\mathbf{r}_t$, satisfies Assumptions 2.1 and 2.2, and that $\boldsymbol{\beta}'\boldsymbol{\beta}/N$ is converging to a finite constant, with the elements of $\boldsymbol{\beta}$ being uniformly bounded.*[30] *Then, for any $N > K = 1$, there exist two* inefficient *portfolios, i.e., inside the efficient frontier,*

$$\mathbf{w}^{\varepsilon} \equiv \frac{1}{N}\frac{(\overline{\beta^2}\mathbf{1}_N - \overline{\beta^1}\boldsymbol{\beta})}{\bar{\sigma}_\beta^2} + \mathcal{O}(1) \qquad \textit{and} \qquad \mathbf{w}^{f} \equiv \frac{1}{N}\frac{\boldsymbol{\beta}}{\overline{\beta^1}} + \mathcal{O}(1), \tag{11}$$

where $\overline{\beta^k} \equiv N^{-1}\sum_{i=1}^{N}\beta_i^k$, $\bar{\sigma}_\beta^2 \equiv (\overline{\beta^2} - (\overline{\beta^1})^2)$, where one portfolio is the unique minimum-variance orthogonal *portfolio to the other, i.e., they satisfy*

$$\mathbf{w}^{\varepsilon} = \underset{\substack{\mathbf{w}\ s.t.\\ \mathbf{w}'\boldsymbol{\Sigma}_r\mathbf{w}^{f}\to 0}}{\text{argmin}}\ \mathbf{w}'\boldsymbol{\Sigma}_r\mathbf{w} \qquad \textit{and} \qquad \mathbf{w}^{f} = \underset{\substack{\mathbf{w}\ s.t.\\ \mathbf{w}'\boldsymbol{\Sigma}_r\mathbf{w}^{\varepsilon}\to 0}}{\text{argmin}}\ \mathbf{w}'\boldsymbol{\Sigma}_r\mathbf{w},$$

[30] We mean that $\sup_{1\le i\le N}|\beta_i| \le \delta < \infty$, for some arbitrary positive scalar δ.

and together span *the GMV portfolio, i.e., defining* $\phi^{\varepsilon} \equiv 1 - \phi^{f}$, *with* $\phi^{f} \equiv \left(\frac{1}{N} \frac{\sigma_{\varepsilon}^2}{\sigma_f^2} \frac{(\overline{\beta^1})^2 / \overline{\beta^2}}{\bar{\sigma}_{\beta}^2} \right)$,

$$\mathbf{w}^{\text{gmv}} = \phi^{\varepsilon} \mathbf{w}^{\varepsilon} + \phi^{f} \mathbf{w}^{f}.$$

Remark 2.3.1. The SA-GMV portfolio $\mathbf{w}^{\varepsilon}$ diversifies away the risk factor f_t because $(\overline{\beta^2} \mathbf{1}_N - \overline{\beta^1} \boldsymbol{\beta})' \boldsymbol{\beta} = 0$ and, thus, is beta-factor neutral. The beta-GMV portfolio $\mathbf{w}^{f}$ is instead completely exposed to the systematic risk factor because $\boldsymbol{\beta}' \boldsymbol{\beta} > 0$, unless $\boldsymbol{\beta} = \mathbf{0}_N$.

Remark 2.3.2. Looking at the coefficients ϕ^{ε} and ϕ^{f}, that represent the contributions of the SA and beta portfolios, respectively, to the overall GMV portfolio, one can see that the share of the beta component is decreasing with systematic risk, represented by σ_f^2, with the variability of the risk exposures β_i, and increasing in the variability of alpha risk, i.e., σ_{ε}^2. The opposite holds for the share of the SA-GMV portfolio, given that $\phi^{\varepsilon} = 1 - \phi^{f}$.

We now characterize the the limiting behavior and performance properties of the SA- and Beta-GMV portfolio weights when N diverges.

Proposition 2.4 (Dominance of the SA Portfolio—GMV Case)**.** *Under the assumptions of Proposition 2.3:*

(i) The SA- and Beta-GMV portfolio weights are of the order

$$\begin{aligned} &\phi^{\varepsilon} |w_i^{\varepsilon}| = \mathcal{O}\left(\frac{1}{N}\right) && \text{and} && \phi^{f} |w_i^{f}| = \mathcal{O}\left(\frac{1}{N^2}\right), \\ &\text{with} \quad \phi^{\varepsilon} = \mathcal{O}(1), && \text{and} && \phi^{f} = \mathcal{O}\left(\frac{1}{N}\right). \end{aligned} \tag{12}$$

(ii) The Sharpe ratios SR^{ε} *and* SR^{f} *of the SA- and Beta-GMV portfolios are given by, respectively,*

$$SR^{\varepsilon} = \frac{1}{(\sigma_{\varepsilon}^2)^{\frac{1}{2}}} \left(\overline{\alpha^1} \left(1 - \frac{(\overline{\beta^1})^2}{\overline{\beta^2}} \right)^{\frac{1}{2}} - \left(\frac{\overline{\beta^1}}{(\overline{\beta^2})^{\frac{1}{2}}} \right) \bar{\rho}_{\alpha,\beta} \, \bar{\sigma}_{\alpha} \right) + \mathcal{O}(1) \text{ and } SR^{f} = sgn(\overline{\beta^1}) \frac{\lambda}{(\sigma_f^2)^{\frac{1}{2}}} + \mathcal{O}(1),$$

with the key result: $$(\text{SR}^{\text{gmv}})^2 - (SR^{\varepsilon})^2 \to 0,$$

where $sgn(\cdot)$ *gives the sign of its argument,* $\bar{\rho}_{\alpha,\beta}$ *is the sample correlation between* β_i *and* α_i, $\bar{\sigma}_{\alpha}^2 \equiv \overline{\alpha^2} - (\overline{\alpha^1})^2$, *and* $\overline{\alpha^k} \equiv N^{-1} \sum_{i=1}^{N} (N^{\frac{1}{2}} \alpha_i)^k$.[31]

Remark 2.4.1. We see from (12) that the SA-GMV portfolio *dominates* the GMV portfolio, i.e., the GMV portfolio is (when N becomes large) a beta-neutral SA-portfolio. This dominance emerges in terms of the portfolio weights and Sharpe ratios, because $w_i^{f} / w_i^{\varepsilon} \to 0$. Moreover, in contrast to the case of the MVE portfolio, the dominance is present also in terms of the portfolio *returns* themselves because

$$r_t^{\text{gmv}} - r_t^{\varepsilon} \to_p 0,$$

where $r_t^{\varepsilon} = \mathbf{w}^{\varepsilon \prime} (\mathbf{r}_t - r_f \mathbf{1}_N) + r_f$.[32]

[31] The no-arbitrage condition (4) implies, for all but a vanishing fraction of N' assets (i..e, $N'/N = o(1)$), that $\alpha_i = \mathcal{O}(N^{-\frac{1}{2}})$, requiring to consider $\alpha_i \times N^{\frac{1}{2}} = \mathcal{O}(1)$ when evaluating moments involving the α_i, such as $\bar{\rho}_{\alpha,\beta}$ and $\overline{\alpha^k}$.

[32] We would like to alert the reader that $r_t^{f} = \mathbf{w}^{f \prime} (\mathbf{r}_t - r_f \mathbf{1}_N) + r_f$, with the superscript f, denotes the return on the Beta-GMV portfolio, while r_f, with the subscript f, denotes the return on the risk-free asset.

Remark 2.4.2. Recall that for every parameter's values, the GMV portfolio converges to the risk-free rate, namely $r_t^{\text{gmv}} \to_p r_f$ where, because of the dominance of the SA-GMV portfolio, $r_t^{\varepsilon} \to_p r_f$ and $r_t^f \to_p 0$. However, the Sharpe ratio expressions reveal many insights.[33] For instance, SR^{ε} decreases when σ_{ε}^2 increases, when the dispersion of the risk exposures $\bar{\sigma}_{\beta}^2$ increases, and when $\overline{\alpha^1}$ increases. In contrast, SR^{ε} decreases when $\bar{\rho}_{\alpha,\beta}$ increases and when $\bar{\sigma}_{\alpha}^2$ increases.

Remark 2.4.3. Interestingly, when one compares the SR of the beta and SA components between the MVE and the GMV, it emerges that the SR of the beta components of either mean-variance strategy is identical (asymptotically - in absolute value), i.e., $(SR^f)^2 = (SR^{\beta})^2 + \mathcal{O}(1)$, whereas for the SA component, assuming $\bar{\rho}_{\alpha,\beta} = 0$ for simplicity, one obtains

$$(SR^{\varepsilon})^2 = \frac{1}{(\sigma_{\varepsilon}^2)} \left(\overline{\alpha^1} \left(1 - \frac{(\overline{\beta^1})^2}{\overline{\beta^2}} \right) \right)^2 + \mathcal{O}(1) \leq \frac{(\overline{\alpha^1})^2}{\sigma_{\varepsilon}^2} + \mathcal{O}(1) \leq \frac{\boldsymbol{\alpha}'\boldsymbol{\alpha}}{\sigma_{\varepsilon}^2} + \mathcal{O}(1) = (SR^{\alpha})^2,$$

that is the SA portfolio achieves a larger SR than the corresponding SA-GMV portfolio.[34]

The implication is that the GMV portfolio is essentially a beta-neutral strategy whose performance is driven by the asset-specific part of the model, i.e. alpha and the asset-specific volatility.

Example 2.4.1 (GMV Portfolio: Uncorrelated Alpha and Beta (cont.) (citeDaNagelXiu2022, Example 1))**.** *Under the assumptions of Example 2.2.2, the SA-GMV portfolio weights satisfy (11). However, recall that the Beta-GMV portfolio vanishes, i.e., $r_t^f \to_p 0$, implying that for this case $(SR^{gmv})^2 - (SR^{\varepsilon})^2 \to_p 0$. However, this implies that $(SR^{gmv})^2 = \mathcal{O}_p(1)$ because the α_i are centered around zero and are independent of the β_i, implying $\bar{\rho}_{\alpha,\beta} = \mathcal{O}_p(1)$ and $\overline{\alpha^1} = \mathcal{O}_p$. We conclude that for the case where α_i and β_i are uncorrelated, the squared SRs $(SR^{\varepsilon})^2$, and therefore, $(SR^{gmv})^2$ converge to zero.*

Example 2.4.2 (GMV Portfolio: Correlated Alpha and Beta (cont.))**.** *Under the assumptions of Example 2.2.3, the SA-GMV portfolio weights satisfy*

$$w_i^{\varepsilon} = \frac{1}{(2n+1)\sigma_{\varepsilon}^2} \frac{((2 + \frac{2b^2}{3}) - \beta_i)}{(\frac{2b^2}{3} + 1)} + \mathcal{O}(1) = \mathcal{O}(\frac{1}{N}), \quad \phi^{\varepsilon} = 1 + \mathcal{O}(1),$$

with $(SR^{\varepsilon})^2 = a^2/(\sigma_{\varepsilon}^2(3 + b^2)) + \mathcal{O}(1)$. For the Beta-GMV portfolio one obtains

$$w_i^f = \frac{\beta_i}{(2n+1)} + \mathcal{O}(1) = \mathcal{O}(\frac{1}{N}), \qquad \phi^f = \left(\frac{1}{(2n+1)} \frac{\sigma_{\varepsilon}^2}{\sigma_f^2} \frac{1/((2 + 2b^2/3))}{(2b^2/3 + 1)} \right) = \mathcal{O}(\frac{1}{N}),$$

with $(SR^f)^2 = \lambda^2/\sigma_f^2 + \mathcal{O}(1)$.

Recalling that the Beta-GMV portfolio vanishes, i.e., $r_t^f \to_p 0$, $(SR^{gmv})^2 - (SR^{\varepsilon})^2 \to_p 0$, implying $(SR^{gmv})^2 = a^2/(\sigma_{\varepsilon}^2(3 + b^2)) + \mathcal{O}_p(1)$. Finally, as a reflection of the sub-optimality of the GMV portfolio

$$(SR^{\varepsilon})^2 = \frac{1}{2}(SR^{\alpha})^2,$$

whereas the Beta-MVE and Beta-GMV portfolio have the identical SRs, a consequence of the single-factor structure of this section.

[33] The SR is well-defined and can be different from zero even though the excess return of the GMV portfolio converges to zero, i.e., $r_t^{\text{gmv}} - r_f \to_p 0$, because the SR reflects the behaviour of the ratio of the GMV mean excess return over its standard deviation. The latter also happens to vanish but at the same speed (as N increases) as the GMV mean excess return.

[34] The first inequality (from the left-hand side) holds because $\left(1 - \frac{(\overline{\beta^1})^2}{\overline{\beta^2}}\right) \leq 1$ and the second inequality follows by Holder's inequality, i.e., $(\sum_{i=1}^N c_i)^2 \leq N(\sum_{i=1}^N c_i^2)$ for every sequence of c_is.

3 Portfolio Choice with Candidate Factor Models

So far, we have considered the APT model in which the systematic risk factors are latent, and therefore, the model does not have any missing risk factors. However, investors often start with a model for asset returns that have observed risk factors rather than latent risk factors. We refer to these models as "candidate models," which are likely to be misspecified. Typical examples are the CAPM and Fama and French (1993, 2015) models. In this section, we explain how to use our methodology to construct mean-variance efficient portfolios even when starting with a (possibly misspecified) candidate model.

We assume that the investor starts with the candidate asset-pricing model linear in K_{can} observed factors $\mathbf{f}_{t,\text{can}}$ (either tradable or nontradable, the difference arising only in the estimation step). However, we assume that the true asset pricing model has further systematic risk factors, implying that the candidate model is missing K_{mis} latent factors $\mathbf{f}_{t,\text{mis}}$, i.e., the APT holds with $K = K_{\text{can}} + K_{\text{mis}}$ systematic risk factors.

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

$$\mathbf{r}_t = \boldsymbol{\mu} + \mathbf{B}_{can}\left(\mathbf{f}_{t,can} - \mathbb{E}(\mathbf{f}_{can})\right) + \boldsymbol{\varepsilon}_{t,can} \qquad \textit{where,} \qquad \boldsymbol{\varepsilon}_{t,can} = \mathbf{B}_{mis}\left(\mathbf{f}_{t,mis} - \mathbb{E}(\mathbf{f}_{mis})\right) + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\mu}$ is the $N \times 1$ vector of expected returns, $\mathbf{B}_{can}$ is the $N \times K_{can}$ full-rank matrix of risk exposures corresponding to the candidate risk factors $\mathbf{f}_{t,can}$, and $\mathbf{B}_{mis}$ is the $N \times K_{mis}$ full-rank matrix of factor loadings corresponding to the latent risk factors $\mathbf{f}_{t,mis}$, with $K = K_{can} + K_{mis} < N$. At any time t, the $K_{can} \times 1$ vector of systematic candidate factors, $\mathbf{f}_{t,can}$, is distributed with mean $\mathbb{E}(\mathbf{f}_{can})$ and covariance matrix $\boldsymbol{\Sigma}_{can} > 0$, the $K_{mis} \times 1$ vector of systematic latent factors, $\mathbf{f}_{t,mis}$, is distributed with mean $\mathbb{E}(\mathbf{f}_{mis})$ and covariance matrix $\boldsymbol{\Sigma}_{mis} > 0$, and the $N \times 1$ vector of shocks $\boldsymbol{\varepsilon}_t$ is distributed with zero mean and covariance matrix $\boldsymbol{\Sigma}_{\varepsilon} > 0$, with all the covariance matricies being positive definite. Moreover, $\boldsymbol{\varepsilon}_t$, $\mathbf{f}_{t,can}$, and $\mathbf{f}_{t,mis}$ are mutually uncorrelated at all leads and lags.

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

$$\mathbb{E}\left[(\mathbf{r}_t - \boldsymbol{\mu})(\mathbf{r}_t - \boldsymbol{\mu})'\right] = \boldsymbol{\Sigma}_r = \mathbf{B}_{\text{can}}\,\boldsymbol{\Sigma}_{\text{can}}\,\mathbf{B}_{\text{can}}' + \mathbf{B}_{\text{mis}}\,\boldsymbol{\Sigma}_{\text{mis}}\,\mathbf{B}_{\text{mis}}' + \boldsymbol{\Sigma}_{\varepsilon}.$$

The classical APT literature (Huberman, 1982; Ingersoll, 1984) has focused on the case where the compensations for alpha risk, $\boldsymbol{\alpha}$, satisfies $\sup_N \boldsymbol{\alpha}'\boldsymbol{\alpha} \leq \sup_N g_{1N}(\boldsymbol{\Sigma}_{\varepsilon})\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\boldsymbol{\alpha} \leq g_{1N}(\boldsymbol{\Sigma}_{\varepsilon})\delta_{\text{apt}} < \infty$. However, when positing a candidate model, the elements of $\boldsymbol{\alpha}$ can be large because of the presence of missing systematic risk factors $\mathbf{f}_{t,\text{mis}}$. Our novel insight is that, even in this case, the APT's main implication, i.e., the no-arbitrage restriction (2), continues to be satisfied.

To understand this, observe that given Assumptions 2.2 and 3.1, and applying the results of Chamberlain (1983) to the returns *net* of the candidate risk factors contribution, i.e., to $\mathbf{r}_t - \mathbf{B}_{\text{can}}(\mathbf{f}_{t,\text{can}} - \mathbb{E}(\mathbf{f}_{\text{can}}))$ leads to

$$\boldsymbol{\Sigma}_{\varepsilon_{\text{can}}} \equiv \mathbf{B}_{\text{mis}}\boldsymbol{\Sigma}_{\text{mis}}\mathbf{B}_{\text{mis}}' + \boldsymbol{\Sigma}_{\varepsilon}, \tag{13}$$

where $\mathbf{B}_{\text{mis}}\boldsymbol{\Sigma}_{\text{mis}}\mathbf{B}_{\text{mis}}'$ represents the covariance from the latent factors missing in the candidate model, and $\boldsymbol{\Sigma}_{\varepsilon}$, represents a alpha risk. The vector $\boldsymbol{\alpha}$ in (1) is now replaced by

$$\boldsymbol{\alpha}_{\text{can}} \equiv \mathbf{B}_{\text{mis}}\boldsymbol{\lambda}_{\text{mis}} + \boldsymbol{\alpha}, \quad \text{such that} \quad \sup_N \boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\boldsymbol{\alpha} \leq \delta_{\text{apt}} < \infty. \tag{14}$$

To show that the APT restriction continues to hold, despite the presence of the latent systematic risk factors, we use (13) and (14) to obtain

$$\boldsymbol{\alpha}_{\text{can}}'\boldsymbol{\Sigma}_{\varepsilon_{\text{can}}}^{-1}\boldsymbol{\alpha}_{\text{can}} = (\mathbf{B}_{\text{mis}}\boldsymbol{\lambda}_{\text{mis}} + \boldsymbol{\alpha})'(\mathbf{B}_{\text{mis}}\boldsymbol{\Sigma}_{\text{mis}}\mathbf{B}_{\text{mis}}' + \boldsymbol{\Sigma}_{\varepsilon})^{-1}(\mathbf{B}_{\text{mis}}\boldsymbol{\lambda}_{\text{mis}} + \boldsymbol{\alpha}).$$

To show that $\boldsymbol{\alpha}_{\text{can}}'\boldsymbol{\Sigma}_{\varepsilon_{\text{can}}}^{-1}\boldsymbol{\alpha}_{\text{can}}$ is bounded, we need to demonstrate that both $(\mathbf{B}_{\text{mis}}\boldsymbol{\lambda}_{\text{mis}})'\boldsymbol{\Sigma}_{\varepsilon_{\text{can}}}^{-1}(\mathbf{B}_{\text{mis}}\boldsymbol{\lambda}_{\text{mis}})$ and $\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\varepsilon_{\text{can}}}^{-1}\boldsymbol{\alpha}$ are bounded. The former is always bounded because it contains $\mathbf{B}_{\text{mis}}$ ("squared") both in the denominator and numerator; therefore, an arbitrarily large $\mathbf{B}_{\text{mis}}$ would still leave $(\mathbf{B}_{\text{mis}}\boldsymbol{\lambda}_{\text{mis}})'\boldsymbol{\Sigma}_{\varepsilon_{\text{can}}}^{-1}(\mathbf{B}_{\text{mis}}\boldsymbol{\lambda}_{\text{mis}})$ bounded (uniformly in N). Boundedness of the second term, $\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\varepsilon_{\text{can}}}^{-1}\boldsymbol{\alpha}$, follows from (14), given that $\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\varepsilon_{\text{can}}}^{-1}\boldsymbol{\alpha} \leq \boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\varepsilon}^{-1}\boldsymbol{\alpha}$ because $\mathbf{B}_{\text{mis}}\boldsymbol{\Sigma}_{\text{mis}}\mathbf{B}_{\text{mis}}'$ is positive semi-definite.

For expositional ease, in the rest of this section, we focus on the stylized two-factor model, i.e., $K = 2$, obtained when the investor posits a model with one candidate risk factor $f_{t,\text{can}}$ but omits a second risk factor $f_{t,\text{mis}}$, which is therefore latent, with risk exposure vectors $\boldsymbol{\beta}_{\text{can}}$ and $\boldsymbol{\beta}_{\text{mis}}$, and risk premia λ_{can} and λ_{mis}, respectively, yielding $K = K_{\text{can}} + K_{\text{mis}} = 2$. As in the previous section, we also assume that the asset-specific shock is uncorrelated across assets and has the same variance for all assets, i.e., $\boldsymbol{\Sigma}_{\varepsilon} = \sigma_{\varepsilon}^2\mathbb{I}_N$, so that the APT restriction in (4) continues to be

$$\sup_N \frac{\boldsymbol{\alpha}'\boldsymbol{\alpha}}{\sigma_{\varepsilon}^2} \leq \delta_{\text{apt}} < \infty,$$

and the asset excess returns covariance matrix and mean are

$$\boldsymbol{\Sigma}_r = \sigma_{f_{\text{can}}}^2\boldsymbol{\beta}_{\text{can}}\boldsymbol{\beta}_{\text{can}}' + \sigma_{f_{\text{mis}}}^2\boldsymbol{\beta}_{\text{mis}}\boldsymbol{\beta}_{\text{mis}}' + \sigma_{\varepsilon}^2\mathbb{I}_N \quad \text{and} \quad \boldsymbol{\mu} - r_f\mathbf{1}_N = \boldsymbol{\alpha} + \boldsymbol{\beta}_{\text{can}}\lambda_{\text{can}} + \boldsymbol{\beta}_{\text{mis}}\lambda_{\text{mis}},$$

where we made the further simplifying orthogonality assumptions $\boldsymbol{\beta}_{\text{can}}'\boldsymbol{\beta}_{\text{mis}} = 0$.

3.1 MVE: Portfolios with Expected Returns For Candidate Factor Models

The two propositions below extend our previous results for MVE portfolios to the case of candidate factor models. We start by deriving the expressions for the SA and Beta portfolios.

Proposition 3.1 (Candidate Model - Three-Fund Separation with SA and Beta Portfolios - MVE Case)**.** *Suppose that the vector of asset returns, $\mathbf{r}_t$, satisfies Assumptions 2.2 and 3.1, and that $\boldsymbol{\alpha} \neq \mathbf{0}_N, \lambda_{can} \neq 0, \lambda_{mis} \neq 0$, $\boldsymbol{\beta}_{can}'\boldsymbol{\beta}_{can}/N$ and $\boldsymbol{\beta}_{mis}'\boldsymbol{\beta}_{mis}/N$ are converging to some finite constants, with $\boldsymbol{\beta}_{can}'\boldsymbol{\beta}_{mis} = 0$, and the elements of $\boldsymbol{\beta}_{can}$ and $\boldsymbol{\beta}_{mis}$ are uniformly bounded.*

Then, for any $N > K_{can} + K_{mis} = 2$, there exist two inefficient *portfolios, i.e., inside the efficient frontier,*

$$\begin{aligned}
\mathbf{w}^{\alpha} &\equiv \frac{1}{\sigma_{\varepsilon}^2}\left(\boldsymbol{\alpha} - \tfrac{(\boldsymbol{\beta}_{can}'\boldsymbol{\alpha})}{(\boldsymbol{\beta}_{can}'\boldsymbol{\beta}_{can})}\boldsymbol{\beta}_{can} - \tfrac{(\boldsymbol{\beta}_{mis}'\boldsymbol{\alpha})}{(\boldsymbol{\beta}_{mis}'\boldsymbol{\beta}_{mis})}\boldsymbol{\beta}_{mis}\right) + \mathcal{O}(1), \qquad \textit{and} \\
\mathbf{w}^{\beta} &\equiv \frac{(\lambda_{can}/\sigma_{f,can}^2)}{(\boldsymbol{\beta}_{can}'\boldsymbol{\beta}_{can})}\boldsymbol{\beta}_{can} + \frac{(\lambda_{mis}/\sigma_{f,mis}^2)}{(\boldsymbol{\beta}_{mis}'\boldsymbol{\beta}_{mis})}\boldsymbol{\beta}_{mis} + \mathcal{O}(1),
\end{aligned} \tag{15}$$

where one portfolio is the unique minimum-variance orthogonal *portfolio to the other, i.e.,*

$$\mathbf{w}^{\alpha} = \underset{\substack{\mathbf{w}\ s.t. \\ \mathbf{w}'\boldsymbol{\Sigma}_r\mathbf{w}^{\beta}\to 0}}{\text{argmin}}\ \mathbf{w}'\boldsymbol{\Sigma}_r r\mathbf{w} \qquad \textit{and} \qquad \mathbf{w}^{\beta} = \underset{\substack{\mathbf{w}\ s.t. \\ \mathbf{w}'\boldsymbol{\Sigma}_r\mathbf{w}^{\alpha}\to 0}}{\text{argmin}}\ \mathbf{w}'\boldsymbol{\Sigma}_r\mathbf{w},$$

and, along with the risk-free asset, they span *the mean-variance-efficient frontier, i.e.,*

$$\mathbf{w}^{mve} = \frac{1}{\gamma}\left(\mathbf{w}^{\alpha} + \mathbf{w}^{\beta}\right)$$

We now characterize the limiting behavior of the weights of the SA and Beta portfolios, along with their performance, when N diverges.

Proposition 3.2 (Candidate Model - Dominance of the SA Portfolio - MVE case)**.** *Under the assumptions of Proposition 3.1:*

(i) The weights of the SA and Beta portfolios satisfy

$$|w_i^\alpha| = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right) \quad \textit{and} \quad |w_i^\beta| = \mathcal{O}\left(\frac{1}{N}\right), \quad \textit{implying} \quad |w_i^\beta|/|w_i^\alpha| \to 0.$$

(ii) The Sharpe ratios SR^α *and* SR^β *of the SA and Beta portfolio satisfy, respectively,* $\mathrm{SR}^\beta = (\frac{\lambda_{can}^2}{\sigma_{f,can}^2})^{\frac{1}{2}} + (\frac{\lambda_{mis}^2}{\sigma_{f,mis}^2})^{\frac{1}{2}} + \mathcal{O}(1)$ *and* $\mathrm{SR}^\alpha = (\boldsymbol{\alpha}'\boldsymbol{\Sigma}_\varepsilon^+\boldsymbol{\alpha})^{\frac{1}{2}} + \mathcal{O}(1)$ *for* N *large enough, setting* $\boldsymbol{\Sigma}_\varepsilon^+ \equiv \left(\frac{1}{\sigma_\varepsilon^2}\right)\left(\mathbb{I}_N - \frac{\boldsymbol{\beta}_{can}\boldsymbol{\beta}_{can}'}{(\boldsymbol{\beta}_{can}'\boldsymbol{\beta}_{can})} - \frac{\boldsymbol{\beta}_{mis}\boldsymbol{\beta}_{mis}'}{(\boldsymbol{\beta}_{mis}'\boldsymbol{\beta}_{mis})}\right)$*, such that, for some constants* $0 < \underline{c} \leq \bar{c} < 1$*, the contributions to the MVE Sharpe ratio of the SA and Beta portfolios satisfy*

$$\bar{c} \leq \frac{(\mathrm{SR}^\alpha)^2}{(\mathrm{SR}^\alpha)^2 + (\mathrm{SR}^\beta)^2} \quad \textit{and} \quad \frac{(\mathrm{SR}^\beta)^2}{(\mathrm{SR}^\alpha)^2 + (\mathrm{SR}^\beta)^2} \leq \underline{c},$$

and

$$(\mathrm{SR}^{mve})^2 - \left((\mathrm{SR}^\alpha)^2 + (\mathrm{SR}^\beta)^2\right) \to 0.$$

3.2 GMV: Portfolios without Expected Returns for Candidate Factor Models

The next two propositions extend our previous results for GMV portfolios to the case of candidate factor models.

Proposition 3.3 (Candidate Model - Two-Fund Separation of SA and Beta Portfolios - GMV Case)**.** *Suppose that the vector of asset returns,* $\mathbf{r}_t$ *satisfies Assumptions 2.2 and 3.1, that* $\boldsymbol{\beta}_{can}'\boldsymbol{\beta}_{can}/N$ *and* $\boldsymbol{\beta}_{mis}'\boldsymbol{\beta}_{mis}/N$ *are converging to some finite constants, with* $\boldsymbol{\beta}_{can}'\boldsymbol{\beta}_{mis} = 0$*, and with the elements of* $\boldsymbol{\beta}_{can}$ *and* $\boldsymbol{\beta}_{mis}$ *are uniformly bounded.*

Then, for any $N > K_{can} + K_{mis} = 2$*, there exist two* inefficient *portfolios, i.e., inside the efficient frontier,*

$$\mathbf{w}^\varepsilon \equiv \frac{1}{N}\frac{\left(\mathbf{1}_N - \frac{\overline{\beta_{can}^1}}{\overline{\beta_{can}^2}}\boldsymbol{\beta}_{can} - \frac{\overline{\beta_{mis}^1}}{\overline{\beta_{mis}^2}}\boldsymbol{\beta}_{mis}\right)}{(1 - \frac{(\overline{\beta_{can}^1})^2}{\overline{\beta_{can}^2}} - \frac{(\overline{\beta_{mis}^1})^2}{\overline{\beta_{mis}^2}})} + \mathcal{O}(1), \qquad \textit{and}$$

$$\mathbf{w}^f \equiv \frac{1}{N}\frac{\left(\frac{\overline{\beta_{can}^1}}{\sigma_{f,can}^2(\overline{\beta_{can}^2})^2}\boldsymbol{\beta}_{can} + \frac{\overline{\beta_{mis}^1}}{\sigma_{f,mis}^2(\overline{\beta_{mis}^2})^2}\boldsymbol{\beta}_{mis}\right)}{((\frac{\overline{\beta_{can}^1}}{\sigma_{f,can}\overline{\beta_{can}^2}})^2 + (\frac{\overline{\beta_{mis}^1}}{\sigma_{f,mis}\overline{\beta_{mis}^2}})^2)} + \mathcal{O}(1),$$

where $\overline{\beta_{can}^k} \equiv N^{-1}\sum_{i=1}^N \beta_{i,can}^k$, $\overline{\beta_{mis}^k} \equiv N^{-1}\sum_{i=1}^N \beta_{i,mis}^k$, $\phi^\varepsilon \equiv 1 - \phi^f$*, with*

$$\phi^f \equiv \frac{1}{N}\left(\frac{1 - \frac{(\overline{\beta_{can}^1})^2}{\overline{\beta_{can}^2}} - \frac{(\overline{\beta_{mis}^1})^2}{\overline{\beta_{mis}^2}}}{\frac{\sigma_\varepsilon^2}{\sigma_{f,can}}(\frac{\overline{\beta_{can}^1}}{\overline{\beta_{can}^2}})^2 + \frac{\sigma_\varepsilon^2}{\sigma_{f,mis}}(\frac{\overline{\beta_{mis}^1}}{\overline{\beta_{mis}^2}})^2}\right),$$

and where one portfolio is the unique minimum-variance orthogonal *portfolio to the other, i.e.,*

$$\mathbf{w}^{\varepsilon} = \underset{\substack{\mathbf{w} \ s.t. \\ \mathbf{w}'\mathbf{\Sigma}_r\mathbf{w}^f \to 0}}{\operatorname{argmin}} \ \mathbf{w}'\mathbf{\Sigma}_r\mathbf{w} \qquad \textit{and} \qquad \mathbf{w}^{f} = \underset{\substack{\mathbf{w} \ s.t. \\ \mathbf{w}'\mathbf{\Sigma}_r\mathbf{w}^{\varepsilon} \to 0}}{\operatorname{argmin}} \ \mathbf{w}'\mathbf{\Sigma}_r\mathbf{w},$$

and together these two portfolios span *the GMV portfolio, i.e.,*

$$\mathbf{w}^{\text{gmv}} = \phi^{\varepsilon}\mathbf{w}^{\varepsilon} + \phi^{f}\mathbf{w}^{f}.$$

Proposition 3.4 (Candidate Model - Dominance of the SA Portfolio - GMV Case)**.** *Under the assumptions of Proposition 2.3:*

(i) The SA- and Beta-GMV portfolio weights are of the order

$$|w_i^f| = \mathcal{O}\left(\frac{1}{N}\right) \qquad \textit{and} \qquad |w_i^{\varepsilon}| = \mathcal{O}\left(\frac{1}{N}\right)$$
$$\textit{with } \phi^f = \mathcal{O}\left(\frac{1}{N}\right) \qquad \textit{and} \qquad \phi^{\varepsilon} = 1 - \phi^f = \mathcal{O}\left(1\right),$$

(ii) The Sharpe ratios SR^{ε} and SR^f of the SA and Beta portfolios are given by, respectively,

$$SR^{\varepsilon} = \frac{1}{(\sigma_{\varepsilon}^2)^{\frac{1}{2}}}\left(\overline{\alpha^1}\left(1 - \frac{(\overline{\beta^1_{can}})^2}{\overline{\beta^2_{can}}} - \frac{(\overline{\beta^1_{mis}})^2}{\overline{\beta^2_{mis}}}\right)^{\frac{1}{2}} - \frac{\left(\left(\frac{\overline{\beta^1_{can}}}{\overline{\beta^2_{can}}}\right)\bar{\sigma}_{\alpha,\beta_{can}} + \left(\frac{\overline{\beta^1_{mis}}}{\overline{\beta^2_{mis}}}\right)\bar{\sigma}_{\alpha,\beta_{mis}}\right)}{\left(1 - \frac{(\overline{\beta^1_{can}})^2}{\overline{\beta^2_{can}}} - \frac{(\overline{\beta^1_{mis}})^2}{\overline{\beta^2_{mis}}}\right)^{\frac{1}{2}}}\right) + \mathcal{O}(1)$$

and

$$SR^f = \frac{\left(\left(\frac{\overline{\beta^1_{can}}}{\overline{\beta^2_{can}}}\right)\frac{\lambda_{can}}{\sigma^2_{f,can}} + \left(\frac{\overline{\beta^1_{mis}}}{\overline{\beta^2_{mis}}}\right)\frac{\lambda_{mis}}{\sigma^2_{f,mis}}\right)}{\left(\left(\frac{\overline{\beta^1_{can}}}{\sigma_{f,can}\overline{\beta^2_{can}}}\right)^2 + \left(\frac{\overline{\beta^1_{mis}}}{\sigma_{f,mis}\overline{\beta^2_{mis}}}\right)^2\right)^{\frac{1}{2}}} + \mathcal{O}(1),$$

where $\bar{\sigma}_{\alpha,\beta_{can}}$ and $\bar{\sigma}_{\alpha,\beta_{mis}}$ are the sample covariances between the $\beta_{i,can}$ and $\beta_{i,mis}$ with the $N^{\frac{1}{2}}\alpha_i$, respectively, $\overline{\beta^k} \equiv N^{-1}\sum_{i=1}^N(\beta_i)^k$, and $\overline{\alpha^k} \equiv N^{-1}\sum_{i=1}^N(N^{\frac{1}{2}}\alpha_i)^k$.[35] *More importantly,*

$$(\text{SR}^{\text{gmv}})^2 - (SR^{\varepsilon})^2 \to 0.$$

4 Estimating Mean-Variance Portfolios with SA Constraint

We explain how to estimate mean-variance portfolios into their two components, namely the SA and Beta portfolios, which in turn entails estimating the APT model. The novelty of our approach is that we formally impose the no-arbitrage constraint (2), which by remarks 2.2.4 and 2.4.2 is equivalent to impose a bounded SR condition on the SA portfolio, that is $(SR^{\alpha})^2 \leq \delta_{\text{apt}}$ for some constant δ_{apt}. For this reason, (2) will be hereafter referred to as the *SA constraint*. For expositional

[35] The no-arbitrage condition (4) implies, for all but a vanishing fraction of N' assets (i..e, $N'/N = o(1)$), that $\alpha_i = \mathcal{O}(N^{-\frac{1}{2}})$, requiring to consider $\alpha_i \times N^{\frac{1}{2}}$ when evaluating $\bar{\sigma}_{\alpha,\beta_{\text{can}}}$, $\bar{\sigma}_{\alpha,\beta_{\text{mis}}}$, and $\overline{\alpha^k}$.

ease, we maintain the same setting of the previous section, and assume that the investor starts with the candidate asset-pricing model linear in $K_{\text{can}} = 1$ observed factors $f_{t,\text{can}}$ (either tradable or nontradable, the only difference arising in the estimation step). However, we assume that the true asset pricing model has two systematic risk factors, implying that the candidate model is missing $K_{\text{mis}} = 1$ latent factor $f_{t,\text{mis}}$, implying that the APT holds with $K = 2$ systematic risk factors.

We advocate using the pseudo-maximum likelihood estimator, based on a Gaussian distribution, penalized by imposing the SA constraint (2), and optimized using Karush-Kuhn-Tucker. It turns out that imposing this constraint is crucially important. First, it represents an *identification* condition for some of the parameters of the asset pricing model, in particular for the risk premium of the systematic risk factor λ_{mis} and the asset-specific expected returns $\boldsymbol{\alpha}$, without which these parameters could not be estimated. Second, it leads to a *shrinkage* feature on the estimator of the various parameters, moreover with a clearcut economic interpretation. The effect of the shrinkage depends on the parameter under consideration, as we discuss below after presenting the closed-form solutions of the estimator for the candidate factor model.

The joint log-likelihood function, corresponding to a generic parameter value $\tilde{\boldsymbol{\theta}}$, takes the following form:[36]

$$L(\tilde{\boldsymbol{\theta}}) = -\frac{1}{2}\log(\det(\tilde{\boldsymbol{\beta}}_{\text{mis}}\tilde{\boldsymbol{\beta}}'_{\text{mis}} + \tilde{\sigma}_\varepsilon \mathbb{I}_N)) - \frac{1}{2T}\sum_{t=1}^{T}\Big(\mathbf{r}_t - r_f\mathbf{1}_N - \tilde{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\beta}}_{\text{mis}}\lambda_{\text{mis}} - \tilde{\boldsymbol{\beta}}_{\text{can}}f_t\Big)'(\tilde{\boldsymbol{\beta}}_{\text{mis}}\tilde{\boldsymbol{\beta}}'_{\text{mis}} + \tilde{\sigma}_\varepsilon \mathbb{I}_N)^{-1}\Big(\mathbf{r}_t - r_f\mathbf{1}_N - \tilde{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\beta}}_{\text{mis}}\lambda_{\text{mis}} - \tilde{\boldsymbol{\beta}}_{\text{can}}f_t\Big)$$

Without loss of generality, we assume that the latent risk factor has unit variance, i.e. $\sigma^2_{\text{mis}} = 1$, achieving identification of $\boldsymbol{\beta}_{\text{mis}}$.[37] Notice that, although possible with our methodology, we do not need to estimate the latent risk factor $f_{t,\text{mis}}$ itself to estimate the portfolio weights.[38] Moreover, should the candidate systematic risk factor $f_{t,\text{can}}$ be tradable portfolio excess return, which we now assume for simplicity, then its risk premium satisfies $\lambda_{\text{can}} = \mathbb{E}(f_{\text{can}})$.

Proposition 4.1 (Candidate Model - Estimation with SA Constraint - Benchmark Case)**.** *Suppose that the vector of asset returns, $\mathbf{r}_t$, satisfies Assumption 3.1. Then, for given δ_{apt},*

$$\hat{\boldsymbol{\theta}} = \underset{\tilde{\boldsymbol{\theta}}}{\text{argmax}} \;\; L(\tilde{\boldsymbol{\theta}}) \quad \textit{subject to} \quad \tilde{\boldsymbol{\alpha}}'(\tilde{\boldsymbol{\beta}}_{mis}\tilde{\boldsymbol{\beta}}'_{mis} + \tilde{\sigma}_\varepsilon \mathbb{I}_N)^{-1}\tilde{\boldsymbol{\alpha}} \leq \delta_{apt},$$

where $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}', \hat{\lambda}_{mis}, \hat{\boldsymbol{\beta}}'_{mis}, \hat{\boldsymbol{\beta}}'_{can}, \hat{\mathbb{E}}(f_{can}), \hat{\sigma}^2_{can}, \hat{\sigma}^2_\varepsilon)$ satisfy the following.

(i) If the optimal value of the Karush-Kuhn-Tucker multiplier satisfies $\hat{\kappa} = 0$, then $\boldsymbol{\alpha}$ and λ_{mis} are not separately identified*, and one can only estimate $\boldsymbol{\alpha}_{\varepsilon_{can}} = (\boldsymbol{\alpha} + \boldsymbol{\beta}_{mis}\lambda_{mis})$ but not the two components separately, whereas all the other parameters are identified. The MLE without constraint (denoted by the * superscript) is given as follows.*

$$\hat{\boldsymbol{\alpha}}^*_{\varepsilon_{can}} \equiv \bar{\mathbf{r}} - r_f\mathbf{1}_N - \bar{f}_{can}\hat{\boldsymbol{\beta}}^*_{can},$$

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

[37] By standard arguments in models with latent factors, these and their risk exposures are estimated up to an unknown rotation matrix. Such rotation matrix is irrelevant when computing mean-variance portfolio weights and the SDF, but fixing the rotation (achieved here by setting $\sigma^2_{\text{mis}} = 1$) makes the exposition of our analysis simpler.

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

$$\hat{\boldsymbol{\beta}}^*_{can} \equiv \frac{\hat{\boldsymbol{\Sigma}}_{r,f_{can}}}{\hat{\sigma}^2_{can}},$$
$$\hat{\sigma}^{*2}_{\varepsilon} \equiv \frac{1}{N} tr(\hat{\mathbb{M}}^*_{mis}\hat{\boldsymbol{\Sigma}}^*_{r-f_{can}\beta_{can}}) + \mathcal{O}(1),$$

and $\hat{\boldsymbol{\beta}}^*_{mis}$ *satisfies*

$$\left(\hat{\sigma}^{*2}_{\varepsilon} + \hat{\boldsymbol{\beta}}^{*\prime}_{mis}\hat{\boldsymbol{\beta}}^*_{mis}\right)\hat{\boldsymbol{\beta}}^*_{mis} = \hat{\boldsymbol{\Sigma}}^*_{r-f_{can}\beta_{can}}\hat{\boldsymbol{\beta}}^*_{mis} + \mathcal{O}(1).$$

where we set $\hat{\mathbb{M}}^*_{mis} \equiv (\mathbb{I}_N - \frac{\boldsymbol{\beta}^*_{mis}\boldsymbol{\beta}^{*\prime}_{mis}}{\boldsymbol{\beta}^{*\prime}_{mis}\boldsymbol{\beta}^*_{mis}})$, $\hat{\boldsymbol{\Sigma}}^*_{r-f_{can}\beta_{can}} \equiv T^{-1}\sum_{t=1}^T(\mathbf{r}_t - \bar{\mathbf{r}} - \hat{\boldsymbol{\beta}}^*_{can}(f_{t,can} - \bar{f}_{can}))(\mathbf{r}_t - \bar{\mathbf{r}} - \hat{\boldsymbol{\beta}}^*_{can}(f_{t,can} - \bar{f}_{can}))'$, $\hat{\boldsymbol{\Sigma}}_{r,f_{can}} \equiv T^{-1}\sum_{t=1}^T(\mathbf{r}_t - \bar{\mathbf{r}})(f_{t,can} - \bar{f}_{can})$, $\hat{\mathbb{E}}(f_{can}) \equiv \bar{f}_{can} \equiv T^{-1}\sum_{t=1}^t f_{t,can}$, *and* $\hat{\sigma}^2_{can} \equiv T^{-1}\sum_{t=1}^T(f_{t,can} - \bar{f}_{can})^2$.

(ii) If the optimal value of the Karush-Kuhn-Tucker multiplier satisfies $\hat{\kappa} > 0$*, all parameters are identifed, and the MLE with the constraint is given as follows.*

$$\begin{aligned}
\hat{\boldsymbol{\alpha}} &\equiv \tfrac{1}{(1+\hat{\kappa})}\hat{\mathbb{M}}_{mis}\hat{\mathbb{P}}_{mis}(\bar{\mathbf{r}} - r_f\mathbf{1}_N - \hat{\boldsymbol{\beta}}^*_{can}\bar{f}_{can}), \qquad (16)\\
\hat{\lambda}_{mis} &\equiv (\boldsymbol{\beta}'_{mis}\boldsymbol{\beta}_{mis})^{-1}\boldsymbol{\beta}'_{mis}\hat{\mathbb{P}}_{mis}(\bar{\mathbf{r}} - r_f\mathbf{1}_N - \hat{\boldsymbol{\beta}}^*_{can}\bar{f}_{can}),\\
\hat{\boldsymbol{\beta}}_{can} &\equiv \hat{\boldsymbol{\beta}}^*_{can} + (\tfrac{\hat{\kappa}}{1+\hat{\kappa}})(\tfrac{\bar{f}^2_{can}}{\hat{\sigma}^2_{\bar{f}_{can}}})\hat{\mathbb{P}}_{mis}\hat{\mathbb{M}}_{mis}\left(\frac{\bar{\mathbf{r}} - r_f\mathbf{1}_N}{\bar{f}_{can}} - \hat{\boldsymbol{\beta}}^*_{can}\right),\\
\hat{\sigma}^2_{\varepsilon} &\equiv \frac{1}{N}tr(\hat{\mathbb{M}}_{mis}\hat{\boldsymbol{\Sigma}}_{r-f_{can}\beta_{can}})\\
&\quad + (\frac{((\frac{\hat{\kappa}}{1+\hat{\kappa}}))^2}{N} + (\tfrac{\hat{\kappa}}{(1+\hat{\kappa})^2}))(\bar{\mathbf{r}} - r_f\mathbf{1}_N - \bar{f}_{can}\hat{\boldsymbol{\beta}}^*_{can})'\hat{\mathbb{P}}_{mis}\hat{\mathbb{M}}_{mis}\hat{\mathbb{P}}_{mis}(\bar{\mathbf{r}} - r_f\mathbf{1}_N - \bar{f}_{can}\hat{\boldsymbol{\beta}}^*_{can}) + \mathcal{O}(1),
\end{aligned}$$

and $\hat{\boldsymbol{\beta}}_{mis}$ *satisfies*

$$\left(\hat{\sigma}^2_{\varepsilon} + \hat{\boldsymbol{\beta}}'_{mis}\hat{\boldsymbol{\beta}}_{mis}\right)\hat{\boldsymbol{\beta}}_{mis} = \left(\hat{\boldsymbol{\Sigma}}_{r-f_{can}\beta_{can}} + (\tfrac{\hat{\kappa}}{1+\hat{\kappa}})\hat{\mathbb{M}}_{mis}\hat{\mathbb{P}}_{mis}(\bar{\mathbf{r}} - r_f\mathbf{1}_N - \bar{f}_{can}\hat{\boldsymbol{\beta}}^*_{can})(\bar{\mathbf{r}} - r_f\mathbf{1}_N - \bar{f}_{can}\hat{\boldsymbol{\beta}}^*_{can})'\hat{\mathbb{P}}_{mis}\right)\hat{\boldsymbol{\beta}}_{mis} + \mathcal{O}(1).$$

where we set $\hat{\mathbb{M}}_{mis} \equiv (\mathbb{I}_N - \frac{\boldsymbol{\beta}_{mis}\boldsymbol{\beta}'_{mis}}{\boldsymbol{\beta}'_{mis}\boldsymbol{\beta}_{mis}})$, $\hat{\mathbb{P}}_{mis} \equiv \left(\mathbb{I}_N + (\tfrac{\hat{\kappa}}{1+\hat{\kappa}})(\tfrac{\bar{f}^2_{can}}{\hat{\sigma}^2_{can}})\hat{\mathbb{M}}_{mis}\right)^{-1}$, $\hat{\boldsymbol{\Sigma}}_{r-f_{can}\beta_{can}} \equiv T^{-1}\sum_{t=1}^T(\mathbf{r}_t - \bar{\mathbf{r}} - \hat{\boldsymbol{\beta}}_{can}(f_{t,can} - \bar{f}_{can}))(\mathbf{r}_t - \bar{\mathbf{r}} - \hat{\boldsymbol{\beta}}_{can}(f_{t,can} - \bar{f}_{can}))'$.

Remark 4.1.1. We derive the closed-form expression for all the parameters of the asset-pricing factor model. Our result applies to the present case when a candidate factor model is assumed, but it is misspecified because the investor omits an additional systematic factors together with the compensation for alpha risk. However, as special cases, the result applies to both cases when the candidate factor model is not misspecified (setting $\boldsymbol{\beta}_{\text{mis}} = \mathbf{0}_N$) and when not candidate factor model is not pre-assumed by the investor (setting $\boldsymbol{\beta}_{\text{can}} = \mathbf{0}_N$), implying that the APT holds. For some parameters, the closed form is obtained when N is taken to be large, such as for the estimator of σ^2_{ε}.

Remark 4.1.2. The SA constraint affects (almost) all the parameters' estimates, providing a clearcut economic underpinning to the statistical properties of the MLE.

In particular, $\boldsymbol{\alpha}$ is *only econometrically identified* when the SA constraint holds, i.e., when the KKT multiplier satisfies $\hat{\kappa} > 0$, and is not identified otherwise. In the former case, $\hat{\boldsymbol{\alpha}}$ embeds two sources of shrinkages: the first is directly related to the constraint itself, namely the smaller is δ_{apt}

(i.e., the larger is $\hat{\kappa}$), the stronger is the shrinkage expressed by the size of the term $1/(1+\hat{\kappa})$, because the compensations for alpha risk need to be compressed in case δ_{apt} is taken to be small. Recall that the SA constraint involves the inverse of σ_ε^2, explaining why the constrained estimator for σ_ε^2 is bigger than the corresponding unconstrained estimator for $\hat{\kappa} > 0$, i.e., to increase the likelihood that the SA constraint is enforced; the second source of shrinkage is inversely related to $\bar{f}_{\text{can}}^2/\hat{\sigma}_{\text{can}}^2$, the SR of the candidate systematic risk factor, but it only arises whenever $\hat{\kappa}$ is not zero. This second channel is economically meaningful, as exposure to alpha risk is smaller the more profitable the candidate risk factor is because in this circumstance the investor seeks to be more exposed to the Beta portfolio. Interestingly, regardless of whether $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_{\text{mis}}$ are cross-sectionally correlated, the corresponding estimators are uncorrelated to one another, i.e., $\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\beta}}_{\text{mis}} = 0$, as $\hat{\boldsymbol{\alpha}}$ embeds $\mathbb{M}_{\text{mis}}$, the orthogonal projection to $\hat{\boldsymbol{\beta}}_{\text{mis}}$.

The risk premium of the latent risk factor λ_{mis} is jointly identifed (with $\boldsymbol{\alpha}$) whenever $\hat{\kappa} > 0$, with a shinkage effect that increases with $\bar{f}_{\text{can}}^2/\hat{\sigma}_{\text{can}}^2$, that is the exposure to the latent systematic factor diminishes when the candidate (observed) systematic factor is profitable.

The risk exposures to the candidate risk factor are captured by the traditional OLS estimator $\hat{\boldsymbol{\beta}}_{\text{can}}^*$ under no constraint but can differ substantially from it when the constraint is binding (i.e., $\hat{\kappa} > 0$). However, in this case the effect is not of the shrinkage-type and, instead, it appears to push the estimator down (up) whenever the OLS estimator is too large or too small, in particular compared with the quantity $(\bar{\mathbf{r}} - r_f\mathbf{1}_N)/\bar{f}_{\text{can}}$. As expected, now this effect is increasing with $\bar{f}_{\text{can}}^2/\hat{\sigma}_{\text{can}}^2$. The discrepancy term, $\hat{\mathbb{M}}_{\text{mis}}((\bar{\mathbf{r}} - r_f\mathbf{1}_N)/\bar{f}_{\text{can}} - \hat{\boldsymbol{\beta}}_{\text{can}}^*)$, can be interpreted as the degree of missspecification of the candidate model as it equals $\mathbb{M}_{\text{mis}}\boldsymbol{\alpha}/\mathbb{E}(f_{can})$ in population. Therefore, assuming a binding constraint, the exposure to the candidate factor increases with its SR, mitigating somehow the misspecification.

For the vector of risk exposures $\boldsymbol{\beta}_{\text{mis}}$ to the latent systematic risk factor, when there is no constraint, we obtain that its estimator coincides with the eigenvector corresponding to the largest eigenvalue of the sample covariance matrix $\hat{\boldsymbol{\Sigma}}_{r-f_{\text{can}}\beta_{\text{can}}}$ corresponding to the excess returns netting out the candidate systematic factor, essentially the traditonal PCA estimator. However, when the constraint binds, the estimator coincides with a *penalized version* of the PCA, as one considers the penalized-sample covariance matrix $\left(\hat{\boldsymbol{\Sigma}}_{r-f_{\text{can}}\beta_{\text{can}}} + (\frac{\hat{\kappa}}{1+\hat{\kappa}})\hat{\mathbb{M}}_{\text{mis}}(\bar{\mathbf{r}} - r_f\mathbf{1}_N - \bar{f}_{\text{can}}\hat{\boldsymbol{\beta}}_{\text{can}})(\bar{\mathbf{r}} - r_f\mathbf{1}_N - \bar{f}_{\text{can}}\hat{\boldsymbol{\beta}}_{\text{can}})'\right)$.

If one neglects the term $\hat{\mathbb{M}}_{\text{mis}}$, the penalization is perfectly *analogous* to the one adopted by the RP-PCA estimator of Lettau and Pelger (2020b), except that now the penalization parameter is bounded between zero and one, i.e., it equals $(\frac{\hat{\kappa}}{1+\hat{\kappa}})$ whereas in Lettau and Pelger (2020b) can take any positive value.[39] Without the presence of the matrix $\hat{\mathbb{M}}_{\text{mis}}$, the penalization induces an increment of the maximum eigenvalue of the target covariance matrix by a term $(\frac{\hat{\kappa}}{1+\hat{\kappa}})(\hat{\boldsymbol{\alpha}}+\hat{\boldsymbol{\beta}}_{\text{mis}})'(\hat{\boldsymbol{\alpha}}+\hat{\boldsymbol{\beta}}_{\text{mis}})$, whereas in our case the change in the maximum eigenvale equals $(\frac{\hat{\kappa}}{1+\hat{\kappa}})\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\alpha}}$, emphasizing the central role of alpha risk in our analysis.[40] However, the presence of the matrix $\hat{\mathbb{M}}_{\text{mis}}$ into the penalized-sample covariance matrix implies that the computation of the estimator $\hat{\boldsymbol{\beta}}_{\text{mis}}$ is recursive.

[39] The RP-PCA estimator, for the one-factor case, is obtained as the eigenvector corresponding to the largest eigenvector of $\hat{\boldsymbol{\Sigma}}_r + \gamma(\bar{\mathbf{r}} - r_f\mathbf{1}_N)(\bar{\mathbf{r}} - r_f\mathbf{1}_N)'$ where the tuning parameter γ is an unconstrained positive constant, coinciding with our penalized covariance matrix for the case of no-candidate systematic factor (i.e., $\boldsymbol{\beta}_{\text{can}}$ equal to the zero vector), when $\gamma = \hat{\kappa}/(1+\hat{\kappa})$ and there is no term $\hat{\mathbb{M}}_{\text{mis}}$.

[40] Another crucial difference with respect to Lettau and Pelger (2020b) is that they assume that the true $\boldsymbol{\alpha}$ is null, and all the focus is on the latent risk factors only. However, Lettau and Pelger (2020b) allow for some latent risk factors to be weak, hence capturing an important source of alpha risk, and establish that RP-PCA can accurately, albeit never perfectly so, estimate such weak factors, when compared with PCA.

Remark 4.1.3. By the KKT methodology,

$$(1+\hat{\kappa})^2 = \left[\bar{\mathbf{r}} - \bar{r_f}\mathbf{1}_N - \hat{\boldsymbol{\beta}}_{\text{can}}\bar{f}_{\text{can}} - \hat{\boldsymbol{\beta}}_{\text{mis}}\lambda_{\text{mis}}\right]' \hat{\boldsymbol{\Sigma}}_{\varepsilon}^{-1} \left[\bar{\mathbf{r}} - \bar{r_f}\mathbf{1}_N - \hat{\boldsymbol{\beta}}_{\text{can}}\bar{f}_{\text{can}} - \hat{\boldsymbol{\beta}}_{\text{mis}}\lambda_{\text{mis}}\right]' / \delta_{\text{apt}}$$

4.1 Robustifying Mean-Variance Portfolios

Any estimation procedure is potentially plagued by the possibility of model misspecification, let alone estimation uncertainty. However, thanks to the no-arbitrage penalization, it turns out that our regularized MLE estimator of the SA portfolio is *equivalent* to the estimator obtained by an investor, adverse to model misspecification, who uses robust control theory. Although a similar result does not hold for the systematic risk portfolio, we are able to show how to construct so-called benchmark portfolios that, without requiring specification and estimation of the systematic risk factors' risk premia covariances and /or risk premia, achieve the same performance as the systematic risk portfolios.

4.1.1 Robust Systematic MVE Portfolio $\mathbf{w}^{\beta}$

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/or $\boldsymbol{\Sigma}_f$, and hence, immune to misspecification. This demonstrates that, under suitable assumptions, one can construct the correctly specified beta portfolio that is *independent* of the systematic risk factors' premia and covariances. The result is illustrated for the APT case, i.e., when all risk factors are latent, with $1 \leq K \leq 2$, but it holds more generally for the candidate model case, setting $K = K_{\text{can}} + K_{\text{mis}} = 2$, implying in this case $\mathbf{B} = (\boldsymbol{\beta}_{\text{can}}, \boldsymbol{\beta}_{\text{mis}})$.

Proposition 4.2 (Approximating the Systematic MVE Portfolio for large N)**.** *Suppose that the vector of asset returns, $\mathbf{r}_t$, satisfies Assumptions 2.1 and 2.2 and $\boldsymbol{\alpha} \neq \mathbf{0}_N$. Suppose further that there exists a benchmark portfolio $\mathbf{w}^{bench}$, with associated Sharpe ratio SR^{bench}, satisfying the following properties:*

$$(\mathbf{w}^{bench})'\boldsymbol{\alpha} \to 0, \qquad \mathbf{B}'\mathbf{w}^{bench} \to \mathbf{c}^{bench}, \qquad (\mathbf{w}^{bench})'\boldsymbol{\Sigma}_{\varepsilon}\mathbf{w}^{bench} \to 0, \tag{17}$$

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

(i) If either $K = 1$, or $K = 2$ together with $\mathbf{c}^{bench}$ perfectly proportional to the vector $\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda}$,

$$(SR^{mve})^2 - \left((SR^{\alpha})^2 + (SR^{bench})^2\right) \to 0,$$

that is the beta portfolio can be well-approximated by the benchmark portfolio without any loss in performance.

(ii) If $K = 2$ and $\mathbf{c}^{bench}$ is not proportional to the vector $\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda}$, then for some positive δ,

$$(SR^{mve})^2 - \left((SR^{\alpha})^2 + (SR^{bench})^2\right) \to \delta > 0,$$

that is the beta portfolio cannot be well-approximated by the benchmark portfolio without any loss in performance.

The first assumption in (17) requires that the benchmark portfolio be asymptotically orthogonal to $\boldsymbol{\alpha}$. 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.

The assumptions in (17) 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,

$$(\mathbf{w}^{\text{bench}})'(\mathbf{r}_t - r_f \mathbf{1}_K) = (\mathbf{c}^{\text{bench}})'(\mathbf{f}_t - r_f \mathbf{1}_K) + \mathcal{O}(1),$$

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 $\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda}$ to $\mathbf{B}'\mathbf{w}^{\text{bench}}$.

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 $\mathbf{w}^{\beta}$ can be approximated by a portfolio that is similar to the market portfolio, $\mathbf{w}^{\text{mkt}} = XXX$. 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 4.2) when a single-factor beta portfolio is considered. The benchmark could also be based on multiple factors, such as the Fama and French factors.

In Proposition 4.2 we established the conditions under which the beta portfolio $\mathbf{w}^{\beta}$ can be replaced, without any loss of performance, by a benchmark portfolio independent of $\boldsymbol{\lambda}$ and $\boldsymbol{\Sigma}_f$, and hence, immune to misspecification. If one is willing to assume that $\boldsymbol{\Sigma}_f$ is well-specified, the systematic GMV portfolio $\mathbf{w}^f$ provides another concrete example of a benchmark portfolio that could be as efficient as the MVE portfolio without the need to specify and estimate the risk premia $\boldsymbol{\lambda}$. This is a remarkable result, given the great challenges in estimating $\boldsymbol{\lambda}$ accurately, providing further insights into the GMV portfolio.

Corollary 4.1 (Approximating the Systematic MVE Portfolio with GMV for large N)**.** *Suppose that the vector of asset returns, $\mathbf{r}_t$, satisfies Assumptions 2.1 and 2.2 with $1 \leq K \leq 2$. Then, the systematic GMV portfolio $\mathbf{w}^f$ of (10) satisfies conditions (17), one obtains the following.*

Whenever $\boldsymbol{\lambda}'\mathbf{c}^{bench} \neq 0$,

(i) If either $K = 1$, or $K > 1$ together with

$$\mathbf{c}^{bench} \equiv \frac{1}{\left(\left(\frac{\overline{\beta^1_{can}}}{\sigma_{f,can}\overline{\beta^2_{can}}}\right)^2 + \left(\frac{\overline{\beta^1_{mis}}}{\sigma_{f,mis}\overline{\beta^2_{mis}}}\right)^2\right)} \begin{pmatrix} \overline{\beta^1_{can}}/(\sigma^2_{f,can}\overline{\beta^2_{can}}) \\ \overline{\beta^1_{mis}}/(\sigma^2_{f,mis}\overline{\beta^2_{mis}}) \end{pmatrix} + \mathcal{O}(1), \tag{18}$$

perfectly proportional to the vector $\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda}$,

$$(SR^{mve})^2 - \left((SR^{\alpha})^2 + (SR^f)^2\right) \to 0,$$

that is the beta portfolio can be perfectly approximated by the systematic GMV portfolio without any loss in performance. Mean-variance efficiency is obtained by combining $\mathbf{w}^f$ *with* $\mathbf{w}^\alpha$.

(ii) If $K > 1$ *and* $\mathbf{c}^{bench}$ *in* (18) *is* not *proportional to the vector* $\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda}$, *then for some positive* δ,

$$(SR^{mve})^2 - \left((SR^\alpha)^2 + (SR^f)^2\right) \to \delta > 0,$$

that is the beta portfolio cannot be well-approximated by the systematic GMV portfolio without any loss in performance.

When the candidate model contains only one systematic risk factor, then the systematic GMV portfolio $\mathbf{w}^f$ achieves mean-variance efficiency, without the need to estimate the risk premia $\boldsymbol{\lambda}$, and we provide the sufficient condition to achieve mean-variance efficiency when $K > 1$.[41]

4.1.2 Robust Unsystematic MVE Portfolio $\mathbf{w}^\alpha$

Our regularized estimator belongs to the class of machine learning techniques that are becoming increasingly adopted in empirical asset pricing. In this section, we establish another insightful analogy, namely that the estimated alpha portfolio $\hat{\mathbf{w}}^\alpha$ using our regularized ML estimator (subject to SA constraint) is *equivalent* to the portfolio constructed by an investor who mitigates model misspecification in the alpha component of returns using robust-control theory, in the sense of Hansen and Sargent (2007). In particular, consider the problem:

$$\max_{\mathbf{w}} \min_{\boldsymbol{\alpha}} \left\{ \mathbf{w}'(\boldsymbol{\mu} - r_f \mathbf{1}_N) - \frac{\gamma}{2}\mathbf{w}'\boldsymbol{\Sigma}_r\mathbf{w} \right\}, \tag{19}$$

for a constant $\gamma > 0$, subject to the relative-entropy constraint

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\ln \frac{f_{\boldsymbol{\alpha}}(\mathbf{r}_t^e)}{f_{\hat{\boldsymbol{\alpha}}}(\mathbf{r}_t^e)} \right) f_{\boldsymbol{\alpha}}(\mathbf{r}_t^e)\, d\mathbf{r}_t^e \leq \delta_{\text{entropy}}, \tag{20}$$

where the constant δ_{entropy} measures aversion to alpha misspecification, $\mathbf{r}_t^e \equiv \mathbf{r}_t - r_f \mathbf{1}_N$ is the vector of excess asset returns, and $f_{\boldsymbol{\alpha}}(\cdot)$ is the probability density function of a random vector distributed as $\mathcal{N}(\boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\lambda}, \boldsymbol{\Sigma}_r)$, for given $\boldsymbol{\lambda}$, $\mathbf{B}$, and $\boldsymbol{\Sigma}_r$, with $\hat{\boldsymbol{\alpha}}$ being our estimator of $\boldsymbol{\alpha}$ from Proposition 4.1. We illustrate here the case when the investor starts from the one-factor candidate model (3.1), in which case $\mathbf{B} = (\boldsymbol{\beta}_{\text{can}}, \boldsymbol{\beta}_{\text{mis}})$ and $K_{\text{mis}} = K_{\text{can}} = 1$, yielding $K = K_{\text{mis}} + K_{\text{can}} = 2$. The multifactor extension is relegated to the Internet Appendix.

Our distributional assumptions allow one to relate the entropy constraint to a constraint on the Sharpe ratio of the alpha portfolio. In fact, under the above distributional assumption for $\mathbf{r}_t^e$, the relative entropy constraint in (20) simplifies to $\frac{1}{2}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})'\boldsymbol{\Sigma}_\varepsilon^{-1}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}) \leq \delta_{\text{entropy}}$, where a larger δ_{entropy} represents an increase in the investor's degree of aversion to model misspecification. Note also that when $\delta_{\text{entropy}} = \frac{1}{2}\frac{(1+\boldsymbol{\lambda}'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda})}{T}\chi^2_{N,x\%}$, 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 (20) defines the set of $\boldsymbol{\alpha}$ that are statistically indistinguishable from $\hat{\boldsymbol{\alpha}}$ at $(1-x)\%$ for the case of $\sqrt{T}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ normally distributed with mean $\mathbf{0}$ and covariance matrix equal to $(1 + \boldsymbol{\lambda}'\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\lambda})\boldsymbol{\Sigma}_\varepsilon$.

[41] A more careful examination shows that the $K = 1$ result holds when $\lambda > 0$ as the Sharpe ratio of the systematic GMV portfolio equals λ/σ_f, which differs from $(\lambda^2/\sigma_f^2)^{\frac{1}{2}}$, i.e., the Sharpe ratio of the MVE portfolio, when $\lambda < 0$.

We show (in the proof for Proposition 4.3) that the solution to this *robust mean-variance* problem, $\mathbf{w}^{\text{rmv}}$, is:

$$\mathbf{w}^{\text{rmv}} \equiv \frac{1}{\gamma}(\mathbf{w}^{r\alpha} + \mathbf{w}^{\beta}) + \mathcal{O}(1),$$

where the robust alpha portfolio, $\mathbf{w}_N^{r\alpha}$ is

$$\mathbf{w}^{r\alpha} \equiv \frac{1}{\phi}\boldsymbol{\Sigma}_{\varepsilon}^{+}\hat{\boldsymbol{\alpha}}, \tag{21}$$

$\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\alpha}}(\delta_{\text{apt}})$ is the constrained ML estimator of $\boldsymbol{\alpha}$ in (16), and $\phi = \phi(\delta_{\text{entropy}}) \geq 1$ is the shrinkage parameter (given in equation (**??**)), implying that $\mathbf{w}^{r\alpha} = \mathbf{w}^{r\alpha}(\delta_{\text{apt}}, \delta_{\text{entropy}})$ arising from the adversity to misspecification feared by the investor. Observe from (21) that in the absence of a concern for model misspecification, the expression for $\mathbf{w}^{r\alpha}$ would have $\phi = 1$, which is the solution to the classical mean-variance portfolio problem in (5). 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 SA constraint is *precisely* the same that one would obtain from the max-min optimization (19)–(20). 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 $\mathbf{w}^{\alpha}$ and $\mathbf{w}^{r\alpha}$, respectively; hence, a smaller δ_{apt} and a larger δ_{entropy} induce an *equivalent* reduction in the demand for risky assets in that portfolio. In other words, there exists a δ_{apt} and a δ_{entropy} that lead to two identical vectors of alpha-portfolio weights.[42]

Proposition 4.3 (Equivalence between SA Portfolio and Robust Portfolio)**.** *Under the assumptions of Proposition 2.1, as $N \to \infty$, the estimated alpha portfolio $\mathbf{w}^{\alpha}(\delta_{apt})$ in equation (15) corresponding to a given δ_{apt}, is asymptotically equivalent to the robust mean-variance alpha portfolio $\mathbf{w}^{r\alpha}(\delta_{apt}^{*}, \delta_{entropy})$ in equation (21), corresponding to some given δ_{apt}^{*} and $\delta_{entropy}$:*

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

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

Thus, the proposition highlights how our regularized estimator subject to the SA constraint is precisely the appropriate framework to account for model misspecification in the alpha portfolio.

5 Empirical Application: 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.

[42] In practice, δ_{apt} and δ_{entropy} are unknown. In the Internet Appendix 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.

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.[43] 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.[44] Both data sets span the period 1977 to 2016. In all cases, we estimate the model using a rolling window of $T = 120$ months.[45]

We compare the performance of our strategies to five strategies from the literature.[46] The first strategy is the mean-variance (MV) efficient portfolio, which, is given by (6) and where we plug in the sample estimates of the mean and covariance matrix of returns. To match Ao et al. (2019),[47] 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 4.2 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).[48]

[43] 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.

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

[45] 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.

[46] See footnote 6.

[47] 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 $\gamma = \frac{(\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N)'\mathbf{V}_N^{-1}(\boldsymbol{\mu}_N - r_{ft}\mathbf{1}_N)}{\mu^* - r_{ft}}$.

[48] 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

5.2 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 2.1–4.3. 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.[49] 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).

5.3 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.[50] 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 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.[51]

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

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.

[49] 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 $r_t^{\text{strategy}} = \hat{\gamma}^{-1}(\bar{r}^\alpha, \bar{r}^\beta)\widehat{\mathbb{C}\text{ov}}(r^\alpha, r^\beta)^{-1}(r_t^\alpha, r_t^\beta)'$, where $\widehat{\mathbb{C}\text{ov}}(r^\alpha, r^\beta)$ 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 $\hat{\gamma} = (\sigma^*)^{-1}\left((\bar{r}^\alpha, \bar{r}^\beta)\widehat{\mathbb{C}\text{ov}}(r^\alpha, r^\beta)^{-1}(\bar{r}^\alpha, \bar{r}^\beta)'\right)^{\frac{1}{2}}$.

[50] 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.

[51] 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

than that of the EW portfolio, consistent with the theoretical result in Propositions 2.1, 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 4.2 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.

5.4 Importance of Alpha Risk

To understand the importance of alpha risk, represented by the SA 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 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 4, 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 4 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 represent latent asset demand dominate the corresponding w_i^β weights, confirming the theoretical result in Proposition 2.2.

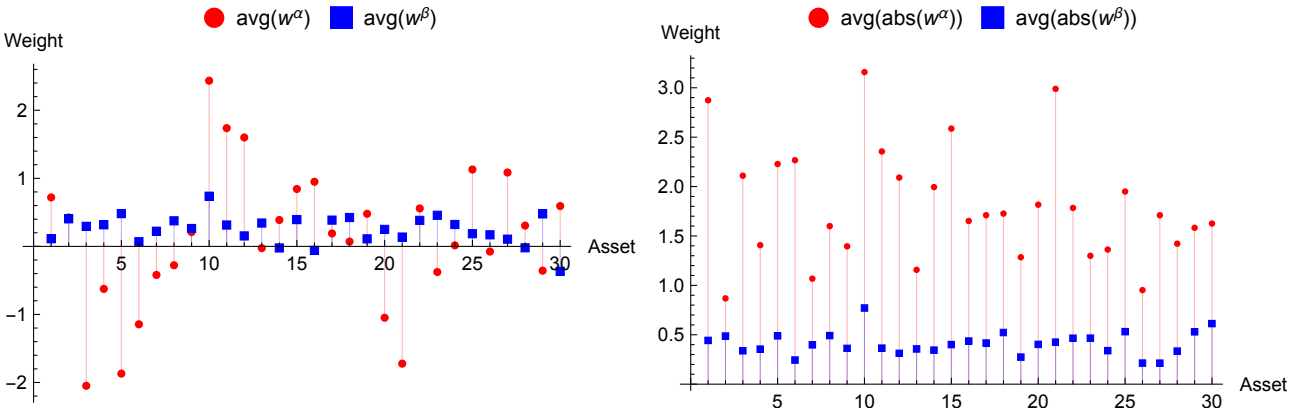
In Figure 5, we decompose the $\mathbf{w}_N^\alpha$ portfolio further into its $\mathbf{w}_N^a$ and $\mathbf{w}_N^A$ components, where the first weight is associated with the asset-specific component of returns and the second with the returns of omitted pervasive risk factors. Just as in the previous 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 5 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 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, $\mathbf{w}_N^\beta$

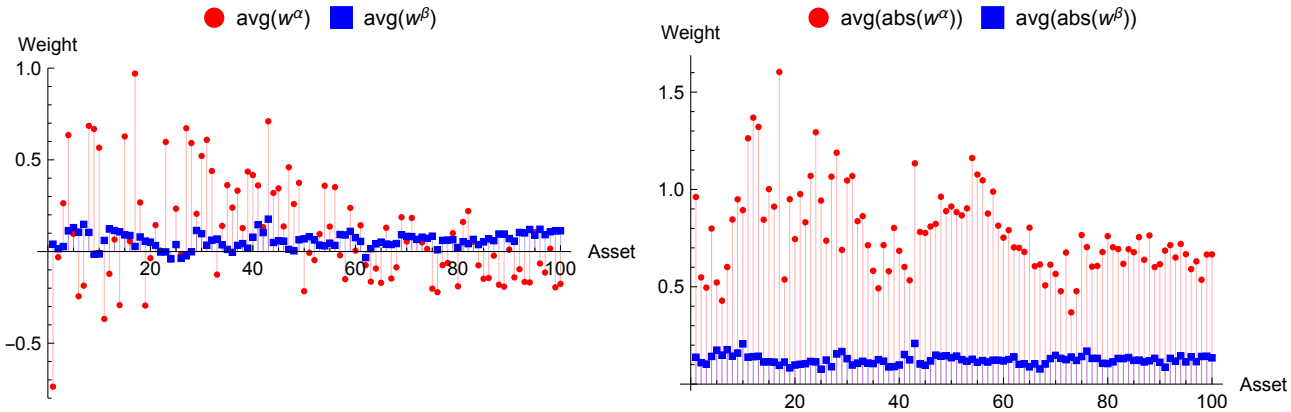
Figure 4: Weights for $\mathbf{w}_N^{\alpha}$ and $\mathbf{w}_N^{\beta}$ portfolios

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



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.

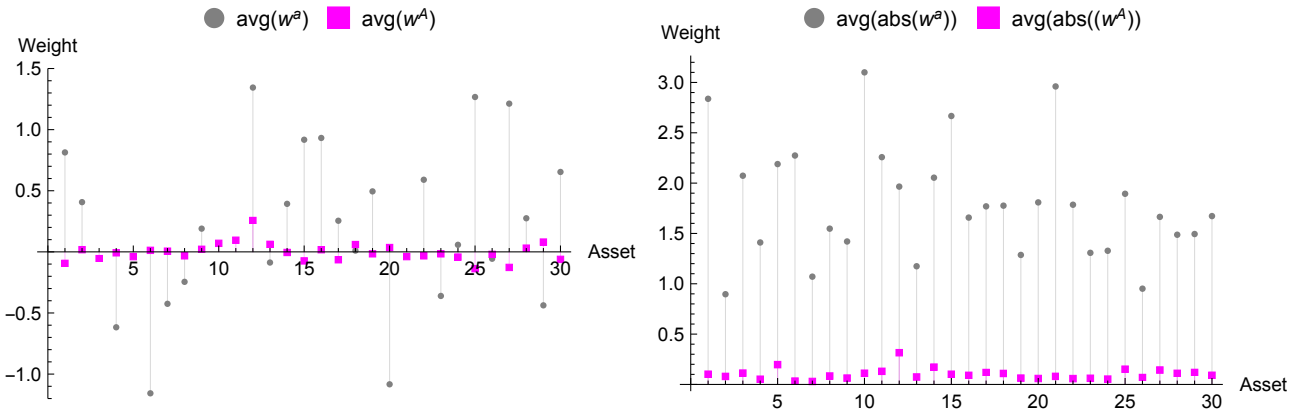
Next we investigate the *returns* of the portfolio components. Panels A and B of Table 3 show that for both data sets, the squared correlations between the returns of the robust mean-variance portfolio r^{rmv} and alpha portfolio r^{α} are over 90%, with most of this coming from the purely asset-specific 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, 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 $\mathbf{w}_N^A$.

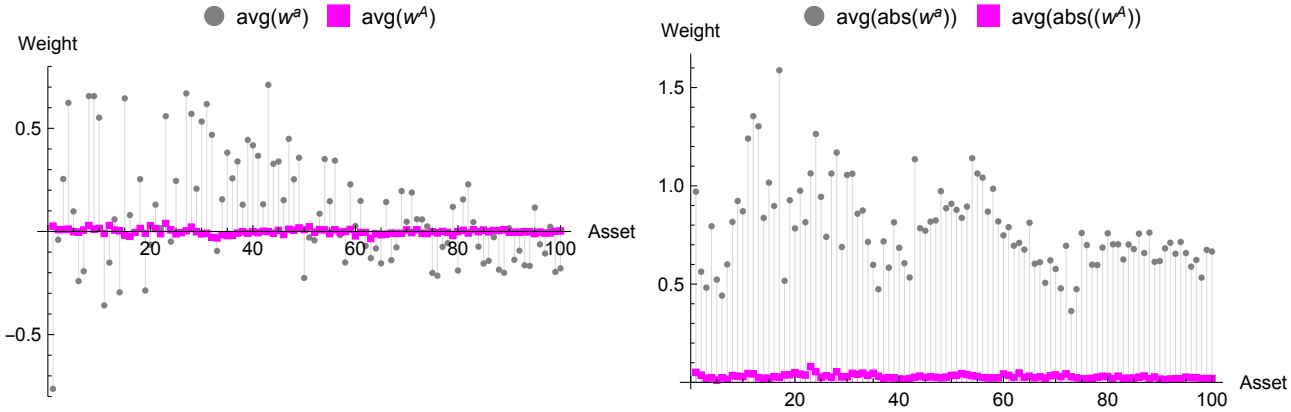
Figure 5: Weights for $\mathbf{w}_N^a$ and $\mathbf{w}_N^A$ portfolios

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



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 2.1, 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 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 reason our strategies outperform the EW portfolio is explained by Propositions 2.1 and 4.2. Proposition 2.1 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 4.2 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.

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

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.

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

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.

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.

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.

Avellaneda, M., and J.H. Lee, 2010, Statistical arbitrage in the U.S. equities market, *Quantitative Finance* 10, 761–782.

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.

Bryzgalova, S., J. Huang, and C. Julliard, 2023a, Bayesian solutions for the factor zoo: We just ran two quadrillion models, *Journal of Finance* 78, 487–557.

Bryzgalova, S., M. Pelger, and J. Zhu, 2023b, Forest through the trees: Building cross-sections of stock returns, *Journal of Finance (forthcoming)* .

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.

Chen, L., M. Pelger, and J. Zhu, 2024, Deep learning in asset pricing, *Management Science* 70, 714–750.

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

Cochrane, John H., 2011, Presidential address: Discount rates, *Journal of Finance* 66, 1047–1108.

Cochrane, John H., 2014, A mean-variance benchmark for intertemporal portfolio theory, *Journal of Finance* 69, 1–49.

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.

Da, Rui, Stefan Nagel, and Dacheng Xiu, 2022, The statistical limit of arbitrage, Working Paper, Chicago Booth.

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, E. F., and K. R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

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, J., Y. Liao, and W. Wang, 2016, Projected principal component analysis in factor models, *Review of Financial Studies* 44, 219–254.

Feng, G., S. Giglio, and D. Xiu, 2020, Taming the factor zoo: A test of new factors, *Journal of Finance* 75, 1327–1370.

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

Freyberger, J., A. Neuhierl, and M. Weber, 2020, Dissecting characteristics nonparametrically, *Review of Financial Studies* 33, 2326–2377.

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).

Gatev, E., W. N. Goetzmann, and K. Rouwenhorst, 2006, Pairs trading: Performance of a relative-value arbitrage rule, *Review of Financial Studies* 19, 797–827.

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.

Giglio, S., D. Xiu, and D. Zhang, 2023, Test assets and weak factors, *Chicago Booth Research Paper 21-04* .

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

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.

Guijarro-Ordonez, J., M. Pelger, and G. Zanotti, 2025, Deep learning statistical arbitrage, *Managent Science*, forthcoming.

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, 2007, *Robustness* (Princeton University Press, Princeton, NJ).

Hastie, T., R. Tibshirani, and J. Friedman, 2011, *The Elements of Statistical Learning*, Springer Series in Statistics (Springer New York, NY).

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

He, A., S. He, D. Huang, and G. Zhou, 2022, Testing asset pricing models using pricing error information, working Paper, U of Washington in St. Louis.

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, 2019, Characteristics are covariances: A unified model of risk and return, *Journal of Financial Economics* 134, 501–524.

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.

Kozak, Serhyi, Stefan Nagel, and Shrihari Santosh, 2020, Shrinking the cross-section, *Journal of Financial Economics* 135, 271–292.

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, 2020a, Estimating latent asset-pricing factors, *Journal of Econometrics* 281, 1–31.

Lettau, Martin, and Markus Pelger, 2020b, 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.

Lopez-Lira, Alejandro, and Nikolai Roussanov, 2022, Do common factors really explain the cross-section of stock returns?, working Paper.

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.

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., 2020, Understanding systematic risk: A high-frequency approach, *Journal of Finance* 74, 2179–2220.

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, 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* .

Zaffaroni, P., 2025, Factor models for conditional asset pricing, journal of Political Economy (forthcoming).