Hitotsubashi ICS-FS Working Paper Series

FS-2014-E-003

Asset Demand and Ambiguity Aversion

Chiaki Hara

Institute of Economic Research, Kyoto University

Toshiki Honda

Graduate School of International Corporate Strategy, Hitotsubashi University

First version: December 24, 2014

Asset Demand and Ambiguity Aversion

Chiaki Hara* Toshiki Honda†

December 24, 2014

Abstract

We study the optimal portfolio choice problem of an investor who is averse to both risk and ambiguity. Using the class of utility functions proposed by Klibanoff, Marinacci, and Mukerji (2005), we establish a generalized mutual fund theorem, which shows that there are a fixed number of mutual funds that cater for all investors, regardless of their ambiguity aversion. We prove that the optimal portfolio is decomposed into two, one remaining and the other vanishing as the degree of ambiguity aversion goes to infinity. We also introduce factor models with ambiguity and compare our results with the Bayesian portfolio approach.

JEL Classification Codes: C38, D81, G11.

Keywords: Ambiguity aversion, optimal portfolio, $1/N$ portfolio, mutual fund theorem, factor model, Bayesian portfolio choice problem.

*The paper is part of the academic projects on Economic Analysis of Intergenerational Issues: Searching for Further Development, funded by the Grant-in-Aid for Specially Promoted Research (grant number 22000001), and on Equilibrium Analysis on Financial Markets with Transaction costs, funded by the Grant-in-Aid (A) (grant number 25245046), both from Japan's Ministry of Education, Culture, Sports, Science and Technology. Hara's email address is `hara.chiaki.7x@kyoto-u.ac.jp`.

†The paper is part of the academic project funded by the Grant-in-Aid (grant number 25285097) from Japan's Ministry of Education, Culture, Sports, Science and Technology. Honda's email address is `thonda@ics.hit-u.ac.jp`. We are both grateful to Robert Anderson, Darrell Duffie, Yoichiro Higashi, Katsuhito Wakai, and participants of our presentations at Kyoto University, Hitotsubashi University, Meiji University, Yokohama National University, the 2012 Annual Meeting of the Nippon Finance Association, the Swiss-Kyoto Symposium, the 2013 Asian Meeting of the Econometric Society, and the 2014 Annual Meeting of the Korean Association of Financial Engineering.

1 Introduction

In many uncertain situations, it is extremely difficult or merely impossible to estimate the distributions of possible outcomes, because there are only few samples or the underlying mechanism is complex. When the asset returns depend on these uncertainties, it is unreasonable to assume that investors have expected utility functions, because, to calculate expected utilities, it would be necessary to know the distributions of asset returns, which are, in fact, unknown to them. In the presence of such ambiguous uncertainties, it is more reasonable to use utility functions that are averse not only to risk but also to ambiguity. In this paper, we study the optimal portfolio choice problem of an investor who exhibits ambiguity aversion.

More specifically, in the traditional single-period setting, we consider an investor who has a utility function in the class proposed by Klibanoff, Marinacci, and Mukerji (2005) (hereafter KMM). It has an advantage over the more commonly used class of ambiguity-averse preferences proposed by Gilboa and Schmeidler (1989), in that it can control the degree of ambiguity aversion simply by varying some parameter values. With this flexibility, we can investigate how the optimal portfolio changes as the degree of ambiguity aversion increases, while the degree of risk aversion remains fixed.

Our analysis specializes in the case where the degree of absolute risk aversion and the degree of ambiguity aversion are constant (independent of consumption levels) and the asset returns are normally distributed. The ambiguity lies in the means (expectations) of the asset returns, which are themselves random variables. As is well known, in the standard setting without ambiguity, the joint assumption of constant absolute risk aversion (CARA) and normally distributed asset returns would give rise to many interesting implications on equilibrium asset prices and risk allocations. Of particular interest is the mutual fund theorem, which claims that a single mutual fund is sufficient to cater for all investors, in the sense that every investor's optimal portfolio of risky assets is a positive multiple of the single mutual fund. Our main question is how the optimal portfolio of the ambiguity averse investor differs from the case of the mutual fund theorem.

Our first main result, Theorem 1, answers the question. This theorem is stated in terms of a matrix, denoted by Q, that roughly measures the ratio of the variance of asset returns due solely to ambiguity to the total variance of these asset returns. It represents the optimal portfolio as a linear combination of the eigenvectors of Q, with the associated coefficients depending on the degree of ambiguity aversion. Among other things, it implies that any single mutual fund is no longer sufficient to cater for all investors.

Our second main result, Theorem 2, characterizes the optimal portfolio by decomposing it into two portfolios: one remaining and the other vanishing as the degree of ambiguity aversion goes to infinity. Each of the two portfolios is characterized as an expected-utility maximizer's optimal portfolio for an appropriately chosen pair of a mean vector and a covariance matrix. In this sense, we decompose the expected returns and the covariance matrix, in addition to the optimal portfolio, into the unambiguous (purely risky) part and the ambiguous part.

Abstract and barren as they may seem, they are rich in implications, as presented in Section 4. Among them are a justification for choosing the so-called $1/N$ portfolio, in which the wealth is allocated equally among all the risky assets (N in number) based on an optimization behavior, and a sufficient condition under which the optimal holding of assets with ambiguous returns decreases as the degree of ambiguity aversion increases. The former is important, as choosing the $1/N$ portfolio has been regarded as a rule of thumb, rather than as a consequence of solving an optimization problem. The latter is significant, as it generalizes a main result by Maccheroni, Marinacci, and Ruffino (2013) (hereafter MMR), whom we shall repeatedly refer to.

Once we have a good grasp of the nature of optimal portfolios for ambiguity-averse investors, we can investigate how much the traditional finance theory has missed by ignoring ambiguity aversion. Of these, we take up two questions. The first one, to be discussed in Section 5, is how to extend the so-called factor model, which is a most commonly used model of asset prices as it allow us to avoid the curse of dimensionality, to accommodate ambiguity aversion. The second one, to be discussed in Section 6, is how to compare the ambiguity-averse investor's portfolio choice problem with the so-called Bayesian portfolio choice problem, a question that arises naturally because, in both problems, it is the lack of sufficient information on the return distributions that induces investors behave differently from those who know the return distributions and maximize expected utility.

In the factor model, the asset returns are linear combinations of a few factors and idiosyncratic shocks. Choosing an appropriate set of factors is, however, a highly nontrivial task. In fact, a recent study by Harvey, Liu, and Zhu (2014) catalogue 316 different factor candidates mainly from already published papers. Moreover, an inappropriate choice of factors results in the non-zero alphas, or the non-zero intercepts of the excess returns of risky assets when regressed on the returns of those factors. Instead of tirelessly searching for the right set of factors and assuming that the investors knows what they are, a more sensible modeling strategy is to assume that the investor is unsure of the validity of the factor model, by formulating that the factors and idiosyncratic risks are ambiguous. We shall do so, thereby generalizing, in some respects, the models by Pástor (2000), Pástor and Stambaugh (2000), and Wang (2005) of ambiguous alphas.

The Bayesian portfolio choice problem arises from the recognition that the means, variances, and covariances of asset returns are unknown and need to be estimated from the market data. It postulates prior distributions of these unknown quantities and assumes that the investor maximizes his expected utility calculated from the conditional distributions of asset returns given the past data (realized returns). It is, on the surface, similar to the maximization problem for our ambiguity-averse investor, since they both treat the expected returns as random quantities. Yet they are, in fact, different, since our ambiguity-averse investor has different attitudes towards estimation errors and risky asset returns, while they are implicitly assumed to be the same in the Bayesian portfolio choice problem. Garlappi, Uppal, and Wang (2007) already made this point, but we will make it more apparent, by decomposing the conditional covariance matrix of asset returns into the ambiguous part and the unambiguous part.

We also present, in Section 7, a numerical analysis on the US stock returns along the lines of Fama and French's three factor model, for which the parameter values are chosen to match the past return data. We take the sample covariance matrix of asset returns as consisting of the ambiguous part and the unambiguous part, and investigate how the optimal portfolio is affected by the way in which the sample covariance matrix is decomposed into the two parts and the investor's attitudes towards ambiguity. While the way in which the matrix is decomposed has no impact on the optimal portfolio for the ambiguity-neutral investors, the lesson to be learned from our numerical analysis is that it has a significant impact for the ambiguity-averse investor, even on the allocation of wealth only among the risky assets (excluding the risk-free asset).

There is a growing literature on the use of ambiguity-averse investors to investigate optimal portfolios and asset pricing in finance and macroeconomics. The most relevant work is MMR, to whom the extended notion of certainty equivalents for general KMM preferences is attributed. They then considered the same CARA-type utility functions as we do in this paper. Our contributions over and above their contributions are the mutual fund theorem and the generalization of their sufficient condition under which an increase in ambiguity aversion leads to a decrease in the optimal holding of an asset with ambiguous returns. Ruffino (2014) investigated under what conditions the single mutual fund is sufficient to cater for all ambiguity-averse investors. Her main result, Theorem 1, is identical to our Proposition 1, which is a special case of our Theorem 1 where the ratio between the degree of risk aversion and that of ambiguity aversion is assumed to be common across all investors. Under the same conditions, Wakai (2014) showed that if we (incorrectly) calculate the beta of an asset with ambiguous returns using the covariance of the purely risky parts of the returns of the asset and the market portfolio, then the alpha of the asset is positive if and only if the beta thus calculated is larger than the beta that we (correctly) obtain when the ambiguity in asset returns is taken into consideration. Ju and Miao (2012) introduced a generalized class of recursive ambiguity-averse utility functions in a discrete-time stochastic model to study asset-pricing implications. The literature on the factor model and the Bayesian portfolio analysis will be surveyed in Sections 5 and 6.

The rest of this paper is organized as follows. Section 2 sets up the model. The main results are presented in Section 3. Section 4 gathers some applications of the main results. The factor model with ambiguity is presented and explored in Section 5. Section 6 compares the ambiguity-averse investor's optimal portfolio choice problem and the Bayesian portfolio choice problem. Section 7 presents numerical applications of the factor model based on the U.S. equity data. Section 8 concludes and suggests directions of future research. All proofs and most lemmas are given in the appendix.

2 Model

Let $(\Omega, \mathscr{F}, P)$ be a probability space. We incorporate ambiguity in the CARA-Normal setting, often used in the study of rational expectations equilibrium, in the following manner. Let M be a random vector defined on Ω. We regard the ranges of M as the set of (names of) conceivable probabilistic models and may, though not formally necessary, take Ω to be the product of the range of M and some space representing physical uncertainties. For each $\theta > 0$, define $u_\theta : \boldsymbol{R} \to \boldsymbol{R}$ by letting $u_\theta(x) = -\exp(-\theta x)$ for every $x \in \boldsymbol{R}$. This felicity function exhibits constant absolute risk aversion (CARA) and its coefficient is equal to θ. For each $\gamma > 0$ and each $\theta > 0$, define a utility function $U_{\gamma,\theta}$ by letting

$$U_{\gamma,\theta}(Z) = E\left[u_\gamma\left(u_\theta^{-1}\left(E\left[u_\theta(Z)|M\right]\right)\right)\right]. \tag{1}$$

If we write $\varphi_{\gamma,\theta} = u_\gamma \circ u_\theta^{-1}$, then

$$\varphi_{\gamma,\theta}(z) = -(-z)^{\gamma/\theta} \tag{2}$$

for every $z < 0$, and

$$U_{\gamma,\theta}(Z) = E\left[\varphi_{\gamma,\theta}\left(E\left[u_\theta(Z)|M\right]\right)\right].$$

If $\gamma = \theta$, then $\varphi_{\gamma,\theta}$ is the identity map and, by the law of iterated expectation, $U_{\gamma,\theta}(Z) = E\left[u_\theta(Z)\right]$. In this case, therefore, $U_{\gamma,\theta}$ is an expected utility function with CARA coefficient θ. We then say that the investor is ambiguity-neutral. We say that an investor who has the utility function $U_{\gamma,\theta}$ with $\gamma > \theta$ is ambiguity-averse. If $\gamma < \theta$, then the investor is ambiguity-loving, though we will not pay any special attention to this case.

Assume that two types of assets are traded. The first one is N risky assets, whose gross returns are represented by an N-variate random vector X defined on Ω. The second one is the risk-free bond, whose gross return is equal to $R \in \boldsymbol{R}$. We assume also that M (as well as X) is an N-variate random vector, and M and X are jointly normally distributed. We further assume that $E[M] = E[X]$ and $\mathrm{Cov}[M, X] = \mathrm{Var}[M]$. We can thus write

$$\begin{pmatrix} M \\ X \end{pmatrix} \sim \mathscr{N}\left(\begin{pmatrix} \mu_M \\ \mu_M \end{pmatrix}, \begin{pmatrix} \Sigma_M & \Sigma_M \\ \Sigma_M & \Sigma_X \end{pmatrix} \right).$$

This assumption involves no loss of generality. It can indeed be shown that even if M did not satisfy this assumption (possibly with a dimension different from N), some linear transformation of M added by some (deterministic) vector of $\boldsymbol{R}^N$ would satisfy this assumption.

Then the conditional return of X given M is normally distributed:

$$X|M \sim \mathscr{N}\left(M, \Sigma_{X|M}\right),$$

where $\Sigma_{X|M} = \Sigma_X - \Sigma_M$. The interpretation along the lines of KMM and MMR would be that the investor believes that the expected returns of the risky assets are ambiguous and the covariance matrix $\Sigma_{X|M}$ is unambiguous, that, in model M, the expected returns are equal to M, and that these models are distributed according to mean μ_M and covariance matrix Σ_M. We take a more lax interpretation in this paper: the asset prices follow the distribution of $\mathscr{N}(\mu_M, \Sigma_X)$, but the investor has different coefficients of aversion to the part of randomness in prices which is due to (can be explained by) the random vector M and the part which is independent of (cannot be updated by) M. In fact, in the numerical examples of Section 7, we take Σ_X as fixed but vary Σ_M while keeping $\Sigma_X - \Sigma_M$ positive semidefinite to see how the optimal portfolios depend on the choice of Σ_M.

Denoted by $\mathscr{S}^N$ the set of all $N \times N$ symmetric matrices. Denote by $\mathscr{S}^N_{++}$ the set of all symmetric and positive definite $N \times N$ matrix, and by $\mathscr{S}^N_{+}$ the set of all symmetric positive semidefinite $N \times N$ matrix. Then $\mathscr{S}^N_{++} \subset \mathscr{S}^N_{+} \subset \mathscr{S}^N$. We assume that $\Sigma_X \in \mathscr{S}^N_{++}$ but allow for $\Sigma_M \in \mathscr{S}^N_{+} \setminus \mathscr{S}^N_{++}$ and $\Sigma_{X|M} \in \mathscr{S}^N_{+} \setminus \mathscr{S}^N_{++}$. That is, while we allow for perfect correlation between the linear combinations of X with respect to the ambiguity covariance matrix Σ_M or the unambiguous covariance matrix $\Sigma_{X|M}$, we exclude perfect correlation among the linear combinations of N random variable having covariance matrix Σ_X. Note that for every $\Sigma \in \mathscr{S}^N$, $\mathrm{Row}\,\Sigma = \mathrm{Col}\,\Sigma$ and $\mathrm{Ker}\,\Sigma = (\mathrm{Row}\,\Sigma)^\perp = (\mathrm{Col}\,\Sigma)^\perp$ and that, for every $\Sigma \in \mathscr{S}^N_{+}$ and every $v \in \boldsymbol{R}^N$, $v \in \mathrm{Ker}\,\Sigma$ if and only if $v^\top \Sigma v = 0$.

Denote by $(a,b) \in \boldsymbol{R}^N \times \boldsymbol{R}$ a portfolio of these $N+1$ assets, representing the monetary amounts invested in each of these assets. Once the state is realized, the portfolio pays out $a^\top X + bR$. Denote the initial wealth by $W \in \boldsymbol{R}$. Let $\mathbf{1}$ be the vector in $\boldsymbol{R}^N$ of which the N coordinates are all equal to one. Then the budget constraint on the portfolio $(a,b) \in \boldsymbol{R}^N \times \boldsymbol{R}$ is $\mathbf{1}^\top a + b \leq W$. The decision maker's utility maximization problem is given by

$$\begin{aligned} \max_{(a,b)\in \boldsymbol{R}^N \times \boldsymbol{R}} \quad & U_{\gamma,\theta}(a^\top X + bR) \\ \text{subject to} \quad & \mathbf{1}^\top a + b \leq W. \end{aligned} \tag{3}$$

Define $V_{\gamma,\theta} : \boldsymbol{R}^N \times \boldsymbol{R} \to \boldsymbol{R}$ by letting

$$V_{\gamma,\theta}(a,b) = \mu_M^\top a + Rb - \frac{1}{2} a^\top \left(\gamma \Sigma_M + \theta \Sigma_{X|M}\right) a$$

for every $(a,b) \in \boldsymbol{R}^N \times \boldsymbol{R}$. Since $\Sigma_{X|M} = \Sigma_X - \Sigma_M$, this can be rewritten as

$$V_{\gamma,\theta}(a,b) = \mu_M^\top a + Rb - \frac{\theta}{2} a^\top \Sigma_X a - \frac{\gamma - \theta}{2} a^\top \Sigma_M a.$$

Thus, it is a robust mean-variance utility function of MMR.

Lemma 1 *For every $(a,b) \in \boldsymbol{R}^N \times \boldsymbol{R}$, $U_{\gamma,\theta}(a^\top X + bR) = -\exp\left(-\gamma V_{\gamma,\theta}(a,b)\right)$.*

If (a,b) is a solution to the utility maximization problem (3), then $\mathbf{1}^\top a + b = W$. Hence, by Lemma 1, for every $(a,b) \in \boldsymbol{R}^N \times \boldsymbol{R}$, (a,b) is a solution to (3) if a is a solution to

$$\max_{a \in \boldsymbol{R}^N} V_{\gamma,\theta}(a, W - \mathbf{1}^\top a) \tag{4}$$

and $b = W - \mathbf{1}^\top a$. Since $\gamma \Sigma_M + \theta \Sigma_{X|M} \in \mathscr{S}^N_{++}$, the first-order condition gives the solution to the problem (3):

$$a = (\gamma \Sigma_M + \theta \Sigma_{X|M})^{-1} (\mu_M - R\mathbf{1}). \tag{5}$$

The equality (24) of MMR is an equivalent characterization of the optimal portfolio.

3 Main results

In this section, we give a generalized version of the mutual fund theorem and a decomposition of the optimal portfolio into two portfolios, one remaining and the other vanishing as the investor becomes extremely ambiguity-averse. These results characterize the way in which the original version of the theorem fails and the nature of portfolios held by ambiguity-averse investors.

We write

$$\eta \equiv \frac{\gamma}{\theta} - 1 \quad \text{and} \quad Q \equiv \Sigma_X^{-1} \Sigma_M.$$

A parameter η is a coefficient of ambiguity aversion.[1] A matrix Q roughly measures the ratio of the variance of asset returns due solely to ambiguity to the total variance of these asset returns. Both Σ_X^{-1} and Σ_M, but Q need not be symmetric.

We define $\zeta : (-1, \infty) \to \boldsymbol{R}^N$ by letting

$$\zeta(\eta) = (I + \eta Q)^{-1} \Sigma_X^{-1}(\mu_M - R\mathbf{1})$$

for every $\eta \in (-1, \infty)$. Then the solution (5) to the problem (3) satisfies $a = \theta^{-1}\zeta(\eta)$. In other words, the function ζ tells us how the investor's portfolio depends on the η. In particular, $\zeta(0) = \Sigma_X^{-1}(\mu_M - R\mathbf{1})$, and $\theta^{-1}\zeta(0)$ is the portfolio that the expected-utility maximizer would hold. Moreover, the portfolio of the investor whose coefficient of ambiguity aversion η can be obtained by transforming $\zeta(0)$ by the matrix $(I + \eta Q)^{-1}$. That is, $\zeta(\eta) = (I + \eta Q)^{-1} \zeta(0)$ for every $\eta > -1$. Our analysis is focused on how $\zeta(\eta)$ varies with η.

The following theorem is a generalized version of the mutual fund theorem, which is applicable to the utility functions $U_{\gamma,\theta}$ with $\gamma \neq \theta$. We eliminate the case where $\mu_M - R\mathbf{1} = 0$, because the portfolio demand is equal to zero for all values of γ and θ.

Theorem 1 (Generalized Mutual Fund Theorem) *Suppose that $\mu_M - R\mathbf{1} \neq 0$. Then there are a $K \in \{1, 2, \ldots, N\}$ and K eigenvectors $v_1, v_2, \ldots, v_K$ of Q with corresponding*

[1] Theorem 2 of KMM implies, in our setting, that the more concave the function $\varphi_{\gamma,\theta}$ is, the more ambiguity-averse the investor is. The function $\varphi_{\gamma,\theta}$ is more concave the larger the value of γ/θ. For this reason, it is appropriate to think of η as a coefficient of ambiguity aversion.

eigenvalues $\lambda_1 < \lambda_2 < \cdots < \lambda_K$ *such that*

$$\zeta(\eta) = \sum_{k=1}^{K} \frac{1}{1 + \lambda_k \eta} v_k. \tag{6}$$

This theorem is rich in interpretation. First, it is a generalized mutual fund theorem: there are K mutual funds, or portfolios of the N risky assets, $v_1, v_2, \ldots, v_K$, that cater for all investors who exhibit any degrees of ambiguity aversion. Second, if $K = 1$, that is, $\zeta(0)$ is an eigenvector of Q, then the original mutual fund theorem holds: a single mutual fund v_1 is sufficient to satisfy all investors' portfolio demands. Third, if $\lambda_k > 0$, then the demand for the k-th mutual fund v_k decreases and converges to zero as the coeffieint η of ambiguity aversion diverges to the infinity; but if $\lambda_k = 0$, which implies that $k = 1$, then the demand for the first mutual fund v_1 does not depend on η. This should come as no surprise because, then, $v_1 \in \operatorname{Ker} \Sigma_M$ and v_1 is a fund that involves no ambiguity. Finally, since

$$\frac{(1 + \lambda_k \eta)^{-1}}{(1 + \lambda_\ell \eta)^{-1}} = \frac{1 + \lambda_\ell \eta}{1 + \lambda_k \eta} = \frac{\lambda_\ell}{\lambda_k} + \left(1 - \frac{\lambda_\ell}{\lambda_k}\right)\left(\frac{1}{1 + \lambda_k \eta}\right),$$

if $k > \ell$, then $(1 + \lambda_k \eta)^{-1}/(1 + \lambda_\ell \eta)^{-1}$ is a strictly decreasing function of η and converges to λ_ℓ / λ_k. Therefore, as $\eta \to \infty$, $\zeta(\eta)$ converges to v_1 if $\lambda_1 = 0$, and $\zeta(\eta)$ converges to 0 but tends to be proportional to $(\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_K^{-1})$ if $\lambda_1 > 0$.

Theorem 1 suggests that it is important to distinguish between the mutual fund that corresponds to the zero eigenvalue and the mutual funds that correspond to the strictly positive eigenvalues. The former remains to be demanded but the demand for the latter vanishes as the coeffieint η of ambiguity aversion diverges to infinity. The next theorem represents the optimal portfolio as the sum of two portfolios, one corresponding to the zero eigenvalue and the other corresponding to the strictly positive eigenvalues. We shall do so by decomposing the total covariance matrix Σ_X into two parts, say Σ_1 and Σ_2, in a way that respects the ambiguity covariance matrix Σ_M.

To see how we should determine the way to decompose the total covariance matrix Σ_X, let's think about what properties the constituent matrices Σ_1 and Σ_2 ought to satisfy. First, we ought to require $\Sigma_i \in \mathscr{S}_+^N$ for each $i = 1, 2$, since they should themselves be covariance matrices. Second, we should of course require $\Sigma_X = \Sigma_1 + \Sigma_2$. With these two requirements, $\operatorname{rank}\Sigma_1 + \operatorname{rank}\Sigma_2 \geq \operatorname{rank}\Sigma_X = N$. The third requirement is, in fact, that $\operatorname{rank}\Sigma_1 + \operatorname{rank}\Sigma_2 = N$, that is, the ranks of the Σ_i's are minimal. This requirement is met if and only if $\operatorname{Row}\Sigma_1 \cap \operatorname{Row}\Sigma_2 = \{0\}$, that is, there is no overlap between Σ_1 and Σ_2 in the linear subspaces spanned by the eigenvectors corresponding to the strictly positive eigenvalues. The fourth requirement, which finally takes Σ_M into consideration, is that one of the $\operatorname{Ker}\Sigma_i$'s, say $\operatorname{Ker}\Sigma_1$, coincides with $\operatorname{Ker}\Sigma_M$. In other words, a portfolio has zero variance with respect to Σ_M (or, equivalently, a portfolio involves no ambiguity) if and only if it does so with Σ_1. The following definition formalizes the way we decompose Σ_X in a more general manner, which allows the matrix not to be of full rank, for future references.

Definition 1 Let $\Sigma \in \mathscr{S}_+^N$ and S be a linear subspace of $\boldsymbol{R}^N$. Let $(\Sigma_1, \Sigma_2) \in \mathscr{S}_+^N \times \mathscr{S}_+^N$. We say that (Σ_1, Σ_2) is an *S-based decomposition of* Σ if $\Sigma_1 + \Sigma_2 = \Sigma$, $\operatorname{rank}\Sigma_1 + \operatorname{rank}\Sigma_2 = \operatorname{rank}\Sigma$, and $\operatorname{Ker}\Sigma_1 = S$.

If $S = \boldsymbol{R}^N$, then $(0, \Sigma)$ is the unique S-based decomposition of Σ, while if $S = \operatorname{Ker}\Sigma$, then $(\Sigma, 0)$ is the unique S-based decomposition of Σ. If there is an S-based decomposition of Σ, then $S \supseteq \operatorname{Ker}\Sigma$, because $S = \operatorname{Ker}\Sigma_1 \supseteq \operatorname{Ker}\Sigma$ whenever (Σ_1, Σ_2) is an S-based decomposition of S. The following lemma shows the existence and uniqueness of the decomposition when $S \supseteq \operatorname{Ker}\Sigma$.

Lemma 2 *For every $\Sigma \in \mathscr{S}_+^N$ and every linear subspace S of $\boldsymbol{R}^N$ that includes $\operatorname{Ker}\Sigma$, there is a unique S-based decomposition of Σ.*

Our second main result can be stated as follows.

Theorem 2 (Risk-Ambiguity Decomposition Theorem) *There is a unique $(\operatorname{Ker}\Sigma_M)$-based decomposition $(\Sigma_{\mathrm{A}}, \Sigma_{\mathrm{R}})$ of Σ_X. Moreover, there exist a unique $(w_{\mathrm{A}}, w_{\mathrm{R}}) \in \operatorname{Row}\Sigma_{\mathrm{A}} \times \operatorname{Row}\Sigma_{\mathrm{R}}$ such that $\mu_M - R\mathbf{1} = w_{\mathrm{A}} + w_{\mathrm{R}}$, and a unique $(v_{\mathrm{R}}, v_{\mathrm{A}}) \in \operatorname{Ker}\Sigma_{\mathrm{A}} \times \operatorname{Ker}\Sigma_{\mathrm{R}}$ such that $\Sigma_{\mathrm{R}} v_{\mathrm{R}} = w_{\mathrm{R}}$ and $\Sigma_{\mathrm{A}} v_{\mathrm{A}} = w_{\mathrm{A}}$. Furthermore, $\zeta(0) = v_{\mathrm{R}} + v_{\mathrm{A}}$ and $\zeta(\eta) \to v_{\mathrm{R}}$ as $\eta \to \infty$.*

This theorem shows that the expected-utility maximizer's optimal portfolio $\zeta(0)$ can be decomposed into two portfolios v_{R} and v_{A}, where the return from v_{R} is unambiguous, or purely risky (whence the subscript R), while the second sub-portfolio v_{A} involves ambiguity (whence the subscript A). Since $\zeta(0) = \Sigma_X^{-1}(\mu_M - R\mathbf{1})$, the definition of $(w_{\mathrm{A}}, w_{\mathrm{R}})$ implies that

$$\zeta(0) = \Sigma_X^{-1} w_{\mathrm{R}} + \Sigma_X^{-1} w_{\mathrm{A}}.$$

But what this theorem implies is something more than this. Interpreting Σ_{A}^{-1} and Σ_{R}^{-1} as the inverse mappings defined on $\operatorname{Row}\Sigma_{\mathrm{A}}$ and $\operatorname{Row}\Sigma_{\mathrm{R}}$ and taking values in $\operatorname{Ker}\Sigma_{\mathrm{R}}$ and $\operatorname{Ker}\Sigma_{\mathrm{A}}$,[2] we can write $v_{\mathrm{R}} = \Sigma_{\mathrm{R}}^{-1} w_{\mathrm{R}}$ and $v_{\mathrm{A}} = \Sigma_{\mathrm{A}}^{-1} w_{\mathrm{A}}$. Then,

$$\zeta(0) = \Sigma_{\mathrm{R}}^{-1} w_{\mathrm{R}} + \Sigma_{\mathrm{A}}^{-1} w_{\mathrm{A}},$$

which represents the optimal portfolio as the sum of two portfolios, each represented as if a mean-variance-efficient portfolio with respect to Σ_{R} or Σ_{A}, which is, unlike Σ_X, positive definite only on $\operatorname{Row}\Sigma_{\mathrm{A}}$ or $\operatorname{Row}\Sigma_{\mathrm{R}}$. More specifically, Σ_{R} is the covariance matrix and w_{R} is the expected excess return for portfolios yielding purely risky returns, while Σ_{A} is the covariance matrix and w_{A} is the expected excess return for portfolios involving ambiguity. If $\Sigma_M \in \mathscr{S}_{++}^N$, then $\Sigma_{\mathrm{A}} = \Sigma_X \in \mathscr{S}_{++}^N$ and $v_{\mathrm{R}} = 0$. The theorem, then, implies that $\zeta(\eta) \to 0$ as $\eta \to \infty$.

4 Applications

4.1 Heterogeneous Investors

The Generalized Mutual Fund Theorem (Theorem 1) can probably be most clearly grasped if it is cast in a model of heterogeneous investors. Consider a model in which there are I investors, indexed by $i = 1, 2, \dots, I$ and each investor i has has utility function U_{γ^i,θ^i}. Write $\eta^i = \gamma^i/\theta^i - 1$. Then his demand a^i for the risky assets are equal to $(\theta^i)^{-1}\zeta(\eta^i)$. The following proposition is an immediate consequence of the definition of $\zeta(\cdot)$. Part 1 of the proposition

[2]Since $\operatorname{Ker}\Sigma_{\mathrm{A}} \cap \operatorname{Ker}\Sigma_{\mathrm{A}} = \{0\}$, Σ_{A} defines a one-to-one mapping from $\operatorname{Ker}\Sigma_{\mathrm{R}}$ into $\operatorname{Row}\Sigma_{\mathrm{A}}$. Since $\dim \operatorname{Row}\Sigma_{\mathrm{A}} = N - \dim \operatorname{Ker}\Sigma_{\mathrm{A}} = \dim \operatorname{Ker}\Sigma_{\mathrm{R}}$, this mapping is onto. Hence we can think of Σ_{A}^{-1} as the mapping of $\operatorname{Row}\Sigma_{\mathrm{A}}$ onto $\operatorname{Ker}\Sigma_{\mathrm{R}}$. For the same reason, we can think of Σ_{R}^{-1} as the mapping of $\operatorname{Row}\Sigma_{\mathrm{R}}$ onto $\operatorname{Ker}\Sigma_{\mathrm{A}}$.

is essentially the same as Theorem 1 of Ruffino (2014).

Proposition 1 *Suppose that $\eta^1 = \eta^2 = \cdots = \eta^I$. Then:*

1. *The demands a^i for risky assets are positive multiples of one another.*
2. *Denote the common value of the η^i by η. Define θ and γ by $\theta^{-1} = \sum_i (\theta^i)^{-1}$ and $\eta = \gamma/\theta - 1$. Then $\theta^{-1}\zeta(\eta) = \sum_i a_i$.*

This proposition deals with the case where all investors share a common coefficient η^i of ambiguity aversion. It includes not only the standard case of expected utility functions, where the η^i are all equal to zero, but also the case of ambiguity aversion, where the η^i are strictly positive. The first part implies that all investors can attain their demands for risky asset by buying a single mutual fund. The second part shows that the aggregate demand coincides with the demand of the representative investor whose risk tolerance (the reciprocal of the coefficient of absolute risk aversion) is equal to the sum of the investors' risk tolerances, and whose degree of ambiguity is equal to each individual investor's counterpart. This equality between the representative investor's risk tolerance and the the sum of the individual investors' counterparts is well known for expected utilities,[3] and the second part generalizes this relation to ambiguity-averse investors.

4.2 Asymptotically optimal wealth allocations

As an application of the Generalized Mutual Fund Theorem (Theorem 1), we consider how the allocation of wealth over the N risky asset will vary as the investor becomes unboundedly ambiguity-averse. The following proposition deals with the case where $\Sigma_M \in \mathscr{S}^N_{++}$ and hence $\lambda_k > 0$ for every k. Then $\zeta(\eta) \to 0$ as $\eta \to \infty$. That is, the portfolio demand converges to zero as the degree of ambiguity aversion diverges to infinity. But it is still interesting to see how the proportional allocation of wealth among the N risky assets, $(\mathbf{1}^\top \zeta(\eta))^{-1}\zeta(\eta)$, varies.

[3] This equality can be traced back at least to Wilson (1968).

Proposition 2 *If* $\Sigma_M \in \mathscr{S}_{++}^N$ *and* $\mathbf{1}^\top \zeta(\eta) > 0$ *for every sufficiently large* η*, then*

$$\frac{1}{\mathbf{1}^\top \zeta(\eta)} \zeta(\eta) \to \frac{1}{\mathbf{1}^\top \Sigma_M^{-1}(\mu_M - R\mathbf{1})} \Sigma_M^{-1}(\mu_M - R\mathbf{1})$$

as $\eta \to \infty$.

This proposition states that an extremely ambiguity averse investor would allocate his wealth among the N risky assets in the same way as the ambiguity-neutral investor would do when the gross returns of the risky assets follow a multivariate normal distribution with mean vector μ_M and covariance matrix Σ_M. Note that the covariance matrix that is relevant here is not $\Sigma_{X|M}$ but Σ_M.

4.3 Optimality of the $1/N$ portfolio

Some empirical studies report that the out-of-sample performance of the sample-based mean-variance model is no better than the $1/N$-portfolio, in which the total wealth is allocated equally over the N assets. For example, DeMiguel, Garlappi, and Uppal (2009) conclude that the out-of-sample performance of the models that are designed to reduce estimation errors is not consistently better than that of the $1/N$-portfolio. This result is shocking because, unlike the portfolio of the sample-based mean-variance model, the $1/N$-portfolio is not derived from solving the utility maximization problem, but regarded, rather, as a rule of thumb.

In this subsection, we show that the $1/N$-portfolio is, at least approximately and, in some cases, exactly, a solution to the utility maximization problem of an investor who is extremely ambiguity averse. This result provides a theoretical justification to the prevalent use of the $1/N$-portfolio. The key assumption is that the covariance matrix Σ_M, which represents ambiguity, has the following form:

$$\begin{pmatrix} \sigma^2 & & & & \\ & \sigma^2 & & \kappa & \\ & & \ddots & & \\ & \kappa & & \sigma^2 & \\ & & & & \sigma^2 \end{pmatrix}, \tag{7}$$

that is, all diagonal elements are equal to σ^2 and all off-diagonal elements are equal to κ. This class can be justified along the lines of the principle of insufficient reason of Savage (1954, Section 4.5).[4] Indeed, Σ_M must necessarily have the above form whenever the ambiguity regarding the expected returns of the N risky assets is represented by a distribution having a common variance and being symmetric with one another when it comes to calculating covariances. As we shall explain later in Subsection 6.2, this form of covariance matrices was used by Frost and Savarino (1986) to estimate an unknown covariance matrix of risky asset returns.

It can easily be shown that the matrix (7) has two eigenspaces. One is the line spanned by $\mathbf{1}$, which corresponds to eigenvalue $\sigma^2 + (N-1)\kappa$. The other is the hyperplane with normal vector $\mathbf{1}$, which corresponds to eigenvalue $\sigma^2 - \kappa$. From this fact, we can derive a couple of useful facts. First, since $\Sigma_M \in \mathscr{S}^N_+$,

$$-\frac{1}{N-1}\sigma^2 \leq \kappa \leq \sigma^2.$$

Second, if

$$-\frac{1}{N-1}\sigma^2 < \kappa < \sigma^2, \tag{8}$$

then $\Sigma_M = \mathscr{S}^N_{++}$. Third, if $\kappa = -(N-1)^{-1}\sigma^2$, then $\operatorname{Ker}\Sigma_M$ coincides with the line consisting of all scalar multiples of $\mathbf{1}$.

The following proposition gives the limiting proportional wealth distribution as $\eta \to \infty$

[4]But we do not claim that the principle is justifiable for the covariance matrix $\Sigma_{X|M}$, which represents risk.

when (8) is met. It also impose the assumption that μ_M is a scalar multiple of $\mathbf{1}$, which can also be justified by the principle of insufficient reason.

Proposition 3 *Suppose that there is a $\delta \in \boldsymbol{R}$ with $\delta \neq R$ such that $\mu_M = \delta \mathbf{1}$, and there are a $\sigma \in \boldsymbol{R}_{++}$, and a $\kappa \in \boldsymbol{R}$ satisfying* (8) *such that Σ_M satisfies* (7)*. Then*

$$\frac{1}{\zeta(\eta)^\top \mathbf{1}} \zeta(\eta) \to \frac{1}{N} \mathbf{1}$$

as $\eta \to \infty$.

The next proposition gives the limiting wealth distribution as $\eta \to \infty$ in the case of $\kappa = -(N-1)^{-1}\sigma^2$. It is different from the previous proposition in that it deals with the absolute, but not proportional, wealth distributions, it does not impose any assumption on μ_M, and it follows immediately from the Risk-Ambiguity Decomposition Theorem (Theorem 2) because $\operatorname{Ker} \Sigma_M$ coincides with the line spanned by $\mathbf{1}$.

Proposition 4 *Suppose that there are a $\sigma \in \boldsymbol{R}_{++}$, and a $\kappa \in \boldsymbol{R}$ satisfying $\kappa = -(N-1)^{-1}\sigma^2$ such that Σ_M satisfies* (7)*. Then $\zeta(\eta)$ converges to a scalar multiple of $\mathbf{1}$ as $\eta \to \infty$.*

4.4 Optimal portfolios with one or two mutual funds

In this subsection, we give sufficient condition for one or two mutual funds to cater for all investors. We also show how the optimal portfolio depends on the coefficient η of ambiguity aversion.

We start with the case of just one mutual fund.

Proposition 5 *If there is a $\lambda \geq 0$ such that $\lambda \Sigma_X = \Sigma_M$, then there is a $v \in \boldsymbol{R}^N$ such that*

$$\zeta(\eta) = \frac{1}{1 + \lambda \eta} v$$

for every $\eta > -1$.

If $\lambda\Sigma_X = \Sigma_M$, then $Q = \lambda I_N$, where I_N is the $N \times N$ identity matrix, and, in particular, $\mu_M - R\mathbf{1}$ is an eigenvector of Q. This proposition, thus, follows from the Generalized Mutual Fund Theorem (Theorem 1).

Under the assumption of Proposition 5,

$$\gamma\Sigma_M + \theta\Sigma_{X|M} = (\lambda\gamma + (1-\lambda)\theta)\,\Sigma_X.$$

Thus

$$V_{\gamma,\theta}(a,b) = V_{\lambda\gamma+(1-\lambda)\theta,\lambda\gamma+(1-\lambda)\theta}(a,b)$$

for every $(a,b) \in \boldsymbol{R}^N \times \boldsymbol{R}$. The utility function of an ambiguity-averse investor coincides with the utility function of the ambiguity-neutral investor with the degree of risk aversion equal to $\lambda\gamma + (1-\lambda)\theta$. Therefore the mean-variance utility function for the ambiguity-averse investor, where the mean and variance are calculated using μ_M and Σ_X, is well defined, and all the standard results for the mean-variance utility functions are valid in our setting. In fact, they can be obtained simply by assuming that $\mu_M - R\mathbf{1}$ is an eigenvector of Q, although the proof is much more complicated and thus skipped.

An immediate consequence of Proposition 5 can be obtained regarding how an increase in the coefficient η of ambiguity aversion affect optimal portfolios. Denote by $\zeta_n(\eta)$ the n-the coordinate of $\zeta(\eta)$.

Corollary 1 *Let $n \in \{1, 2, \ldots, N\}$. Under the assumption of Proposition 5, if $\zeta_n(\eta) > 0$ for some $\eta > -1$, then $\zeta_n(\eta) > 0$ for every $\eta > -1$. Moreover, then, $\zeta_n(\eta)$ is converges strictly decreasingly to 0 as $\eta \to \infty$.*

This corollary implies that there is an investor who holds a long position of an asset, then all investors hold long position of the asset, and that for any two investors having the same coefficient of risk aversion, the investor with a higher coefficient of ambiguity aversion holds less of it.

Next, we give a sufficient condition for two mutual funds to be sufficient, one of which remains demanded even when the coefficient η of ambiguity aversion becomes unboundedly large. The sufficient condition requires, roughly, that a scalar multiple of Σ_M should give a

($\operatorname{Ker}\Sigma_M$)-based decomposition of Σ_X.

Proposition 6 *If there is a $\lambda > 0$ such that $(\lambda^{-1}\Sigma_M, \Sigma_X - \lambda^{-1}\Sigma_M)$ is a $(\operatorname{Ker}\Sigma_M)$-based decomposition of Σ_X, then there are a $v_{\mathrm{R}} \in \boldsymbol{R}^N$ and a $v_{\mathrm{A}} \in \boldsymbol{R}^N$ such that*

$$\zeta(\eta) = v_{\mathrm{R}} + \frac{1}{1+\lambda\eta} v_{\mathrm{A}}$$

for every $\eta > -1$.

Now, in the case of two mutual funds, we consider how an increase in the coefficient η of ambiguity aversion affect the optimal portfolio. To start, note that even when $\zeta_n(\eta) > 0$ for some η, we cannot guarantee that the n-th coordinates of v_{R} and v_{A} are strictly positive. In general, therefore, we cannot conclude that $\zeta_n(\eta) > 0$ for every η or it is a strictly decreasing function of η. However, if there are some assets of which the returns are unambiguous, such predictions are possible for assets with ambiguous returns.

Corollary 2 *Under the assumption of Proposition 6, suppose in addition that there are an $L < N$ and a $\hat{\Sigma}_M \in \mathscr{S}_{+}^{N-L}$ such that*

$$\Sigma_M = \left(\begin{array}{c:c} 0 & 0 \\ \hdashline 0 & \hat{\Sigma}_M \end{array}\right) \tag{9}$$

Let $n > L$. If $\zeta_n(\eta) > 0$ for some $\eta > -1$, then $\zeta_n(\eta) > 0$ for every $\eta > -1$. Moreover, then, $\zeta_n(\eta)$ converges strictly decreasingly to 0 as $\eta \to \infty$.

Under the assumption of Proposition 6, if (9) is met, then the upper-left $L \times L$ submatrix of $\Sigma_{X|M}$ belongs to $\mathscr{S}_{++}^{L}$ and its first L row vectors span $\operatorname{Row}(\Sigma_X - \lambda^{-1}\Sigma_M)$.

Note that Corollary 2 generalizes Proposition 8 of MMR in two respects. First, the assumption is given only in terms of the positivity of optimal holdings for an risky asset. Second, the number of risky assets (and those of purely risky assets and ambiguous assets) is arbitrary.

5 Factor model with ambiguity

In this section, we develop a factor model with ambiguity. Although the model appears to be more complicated than the model we have been analyzing, the two models are, in fact, equivalent to each other, as we shall prove in subsection 5.1. The reduction in the dimension by factor model structure makes calculations much simpler in the portfolio selection problem. In subsection 5.2, we present a more specialized factor model, in which the factors are traded assets. In subsection 5.3, we present an even more specialized factor model, which still generalizes some aspects of the models of Pástor (2000) and Pástor, Stambaugh (2000), and Wang (2005), who investigated investors who believe the validity of the factor model probabilistically, rather than deterministically.

5.1 General theory

To give a factor model with ambiguity, we start with two L-variate random vectors G and Y, and two N-variate random vectors H and Z. We interpret Y as common factors, with an ambiguous (conditional) mean vector G, and Z as idiosyncratic shocks, with an ambiguous (conditional) mean vector H. We assume that G, Y, H, and Z are jointly normally distributed and

$$\operatorname{Var}[G] = \operatorname{Cov}[G, Y], \qquad E[G] = E[Y], \tag{10}$$

$$\operatorname{Var}[H] = \operatorname{Cov}[H, Z], \qquad E[H] = E[Z], \tag{11}$$

$$\operatorname{Cov}[G, H] = \operatorname{Cov}[Y, H], \tag{12}$$

$$\operatorname{Cov}[G, H] = \operatorname{Cov}[G, Z], \tag{13}$$

$$\operatorname{Cov}[Y, Z] = 0. \tag{14}$$

The first two constraints, (10) and (11), involve no loss of generality. As pointed out in Section 2, if these constraints are not met, then we can always replace G and H by some linear combinations of them added with some (deterministic) vectors. The next constraint (12) implies, as we will soon see, that once mean G of factor Y are known, knowing mean H of idiosyncratic shock Z does not help us to further update the distribution of factors

Y. Similarly, (13) implies that once the mean H of the idiosyncratic shock Z is known, knowing the mean G of the factor Y does not help us to further update the distribution of idiosyncratic shocks Z. The last constraint (14) says that Y and Z are independent. It allows us to interpret Y as common factors and Z as idiosyncratic shocks. Under these assumptions, we can write

$$\begin{pmatrix} G \\ Y \\ H \\ Z \end{pmatrix} \sim \mathscr{N}\left(\begin{pmatrix} \mu_G \\ \mu_G \\ \mu_H \\ \mu_H \end{pmatrix}, \begin{pmatrix} \Sigma_G & \Sigma_G & \Sigma_{GH} & \Sigma_{GH} \\ \Sigma_G & \Sigma_Y & \Sigma_{GH} & 0 \\ \Sigma_{HG} & \Sigma_{HG} & \Sigma_H & \Sigma_H \\ \Sigma_{HG} & 0 & \Sigma_H & \Sigma_Z \end{pmatrix}\right). \tag{15}$$

Then,

$$\begin{pmatrix} Y \\ Z \end{pmatrix} \Bigg| \begin{pmatrix} G \\ H \end{pmatrix} \sim \mathscr{N}\left(\begin{pmatrix} G \\ H \end{pmatrix}, \begin{pmatrix} \Sigma_Y - \Sigma_G & -\Sigma_{GH} \\ -\Sigma_{HG} & \Sigma_Z - \Sigma_H \end{pmatrix}\right). \tag{16}$$

Let $\beta \in \boldsymbol{R}^{L \times N}$ and define the gross returns of N risky assets and the ambiguities involved in them by

$$X = \beta^\top Y + Z, \tag{17}$$

$$M = \beta^\top G + H. \tag{18}$$

Then β is the matrix of factor exposures, whose (ℓ, n) element is the exposure, or loading, of asset n to factor ℓ. Moreover,

$$\begin{pmatrix} M \\ X \\ G \\ Y \end{pmatrix} \sim \mathscr{N}\left(\begin{pmatrix} \mu_M \\ \mu_M \\ \mu_G \\ \mu_G \end{pmatrix}, \begin{pmatrix} \Sigma_M & \Sigma_M & \Sigma_{MG} & \Sigma_{MG} \\ \Sigma_M & \Sigma_X & \Sigma_{MG} & \Sigma_{XY} \\ \Sigma_{GM} & \Sigma_{GM} & \Sigma_G & \Sigma_G \\ \Sigma_{GM} & \Sigma_{YX} & \Sigma_G & \Sigma_Y \end{pmatrix}\right) \tag{19}$$

where

$$
\begin{aligned}
\mu_M &= \beta^\top \mu_G + \mu_H, && (20)\\
\Sigma_M &= \beta^\top \Sigma_G \beta + \beta^\top \Sigma_{GH} + \Sigma_{HG}\beta + \Sigma_H, && (21)\\
\Sigma_X &= \beta^\top \Sigma_Y \beta + \Sigma_Z, && (22)\\
\Sigma_{MG} &= \beta^\top \Sigma_G + \Sigma_{HG}, && (23)\\
\Sigma_{XY} &= \beta^\top \Sigma_Y. && (24)
\end{aligned}
$$

This implies that (M, X) is nothing but the model of risky assets that we have been considering so far, as long as $\Sigma_X \in \mathscr{S}^N_{++}$. This latter condition is satisfied if $\Sigma_Y \in \mathscr{S}^L_{++}$ and $\operatorname{Ker}\beta \cap \operatorname{Ker}\Sigma_Z = \{0\}$. Then,

$$
\left(\begin{array}{c} X \\ Y \end{array}\right) \middle| \left(\begin{array}{c} M \\ G \end{array}\right) \sim \mathscr{N}\left(\left(\begin{array}{c} M \\ G \end{array}\right), \left(\begin{array}{cc} \Sigma_X - \Sigma_M & -\Sigma_{MG} \\ -\Sigma_{GM} & \Sigma_Y - \Sigma_G \end{array}\right)\right).
$$

In particular, the conditional distribution of X given M is not changed by further conditioning X on G, and the conditional distribution of Y given G is not changed by further conditioning Y on M.

The converse also holds. Let (M, X) be the model of N risky assets that we have been considering so far, and (G, Y) be L factors involving ambiguity. We assume that M, X, G, and Y satisfies (19). Assume also that $\Sigma_Y \in \mathscr{S}^L_{++}$. Define $\beta = \Sigma_Y^{-1}\Sigma_{YX} \in \boldsymbol{R}^{L\times N}$. Then define H and Z to satisfy (17) and (18). Then G, Y, H, and Z satisfy (15) and (20) to (24). Moreover, since $\Sigma_X \in \mathscr{S}^N_{++}$ and $\Sigma_Y \in \mathscr{S}^L_{++}$, $\operatorname{Ker}\beta \cap \operatorname{Ker}\Sigma_Z = \{0\}$.

To summarize, we have shown that there are two equivalent ways to define a factor model with ambiguity. The first one is to specify any (G, Y, H, Z) satisfying $\Sigma_Y \in \mathscr{S}^L_{++}$ and (15) and any $\beta \in \boldsymbol{R}^{L\times N}$ satisfying $\operatorname{Ker}\beta \cap \operatorname{Ker}\Sigma_Z = \{0\}$. Then $\Sigma_X \in \mathscr{S}^N_{++}$. The other is to specify any (M, X, G, Y) satisfying $\Sigma_X \in \mathscr{S}^N_{++}$, $\Sigma_Y \in \mathscr{S}^L_{++}$, and (19), and define $\beta = \Sigma_Y^{-1}\Sigma_{YX} \in \boldsymbol{R}^{L\times N}$. Then $\operatorname{Ker}\beta \cap \operatorname{Ker}\Sigma_Z = \{0\}$. The two are related via (17) and (18).

In a factor model (G, Y, H, Z, β), if $\operatorname{rank}\beta = L$ and X is defined by (17), then $\operatorname{rank}\Sigma_{YX} = L$ by (24). Conversely, in a factor model (M, X, G, Y), if $\operatorname{rank}\Sigma_{YX} = L$ and $\beta = \Sigma_Y^{-1}\Sigma_{YX}$, then $\operatorname{rank}\beta = L$. In other words, these assumptions are equivalent to each other. Moreover, either of these two assumptions involves no loss of generality. It can indeed be easily shown that if either (and, hence, both) of the two ranks is equal to K, where $K < L$, then some linear combination of the L coordinates of Y that results in a K-dimensional random vector admits an equivalent factor model satisfying these assumptions.

5.2 Factor portfolios

Consider a factor model defined by a (G, Y, H, Z) satisfying $\Sigma_Y \in \mathscr{S}^L_{++}$ and (15), and a $\beta \in \boldsymbol{R}^{L \times N}$ satisfying $\operatorname{Ker}\beta \cap \operatorname{Ker}\Sigma_Z = \{0\}$. Assume in addition that $\operatorname{rank}\beta = L$ and $\operatorname{Ker}\beta + \operatorname{Ker}\Sigma_Z = \boldsymbol{R}^N$. The first assumption is without loss of generality, as we showed at the end of the previous subsection. The second one means that any portfolio can be decomposed into two portfolios, one without any common shock and the other without any idiosyncratic shock. Given the first assumption, we can state the second assumption in a number of equivalent ways, and one of them is that the linear mapping defined by β maps $\operatorname{Ker}\Sigma_Z$ onto $\boldsymbol{R}^L$.[5] Then the linear mapping defined by $\Sigma_Y \beta$ maps $\operatorname{Ker}\Sigma_Z$ onto $\boldsymbol{R}^L$. This can be interpreted as saying that each of the L factors is a return of a portfolio of traded assets, which we call factor portfolios. Although we do not need to assume the existence of the factor portfolios to derive the arbitrage pricing relation (Ross (1976)), the factor portfolios are often assumed to exist to simplify the theoretical deviation.[6] The assumption is also common in empirical studies, since the estimation and interpretation of factor risk premia are less complicated.[7]

Let us now characterize the $(\operatorname{Ker}\Sigma_M)$-based decomposition of Σ_X in the case of factor portfolios. Once this is done, the Risk-Ambiguity Decomposition Theorem (Theorem 2) allows us to characterize the optimal portfolio. We start with establishing a lemma, which is necessary to accommodate the case where the ambiguous mean vector G of factors Y and the ambiguous mean vector H of idiosyncratic shocks Z are correlated (although Y and Z are uncorrelated by assumption). For $\underline{\Sigma} \in \mathscr{S}^L$ and $\overline{\Sigma} \in \mathscr{S}^L$, we write $\underline{\Sigma} \leq \overline{\Sigma}$ whenever $\overline{\Sigma} - \underline{\Sigma} \in \mathscr{S}^L_+$.

[5] Since $\operatorname{Ker}\beta \cap \operatorname{Ker}\Sigma_Z = \{0\}$, this mapping is one-to-one on $\operatorname{Ker}\Sigma_Z$. Since, in addition, $\operatorname{Ker}\beta + \operatorname{Ker}\Sigma_Z = \boldsymbol{R}^N$, $\dim(\operatorname{Ker}\Sigma_Z) = \operatorname{rank}\beta$. Since $\operatorname{rank}\beta = L$, $\operatorname{Ker}\Sigma_Z$ is of dimension L and mapped onto $\boldsymbol{R}^L$.

[6] Huang and Litzenberger (1988, Chapter 4) contained such a derivation.

[7] Campbell, Lo, and MacKinlay (1997, Section 6.2) explored this case.

Lemma 3 *If* $\operatorname{rank}\beta = L$ *and* $\boldsymbol{R}^N = \operatorname{Ker}\beta + \operatorname{Ker}\Sigma_Z$*, then there is a* $\Gamma \in \mathscr{S}_+^L$ *such that* $\beta^\top \Sigma_{GH} + \Sigma_{HG}\beta = \beta^\top \Gamma \beta$ *and* $\Sigma_G + \Gamma \leq \Sigma_Y$*.*

With this lemma, we can characterize the two matrices Σ_{A} and Σ_{R} of the Risk-Ambiguity Decomposition Theorem (Theorem 2) in the case of factor portfolios.

Proposition 7 *Assume that* $\operatorname{rank}\beta = L$ *and* $\boldsymbol{R}^N = \operatorname{Ker}\beta + \operatorname{Ker}\Sigma_Z$*. Let* Γ *be the one defined in Lemma 3. Let* $\left(\Sigma_Y^{\mathrm{A}}, \Sigma_Y^{\mathrm{R}}\right)$ *be the* $(\operatorname{Ker}(\Sigma_G + \Gamma))$*-based decomposition of* Σ_Y *and* $\left(\Sigma_Z^{\mathrm{A}}, \Sigma_Z^{\mathrm{R}}\right)$ *be the* $(\operatorname{Ker}\Sigma_H)$*-based decomposition of* Σ_Z*. Write* $(\Sigma_{\mathrm{A}}, \Sigma_{\mathrm{R}}) = \left(\beta^\top \Sigma_Y^{\mathrm{A}} \beta + \Sigma_Z^{\mathrm{A}}, \beta^\top \Sigma_Y^{\mathrm{R}} \beta + \Sigma_Z^{\mathrm{R}}\right)$*. Then* $(\Sigma_{\mathrm{A}}, \Sigma_{\mathrm{R}})$ *is the* $(\operatorname{Ker}\Sigma_M)$*-based decomposition of* Σ_X*.*

Proposition 7 shows that the $(\operatorname{Ker}\Sigma_M)$-based decomposition of Σ_X can be obtained by applying the Risk-Ambiguity Decomposition Theorem (Theorem 2) twice. The covariance matrix Σ_Y of factors Y is decomposed into ambiguity-related Σ_Y^{A} and risk-related Σ_Y^{R}. The covariance matrix Σ_Z of idiosyncratic shocks Z is decomposed into ambiguity-related Σ_Z^{A} and risk-related Σ_Z^{R}. Thus the ambiguity-related covariance Σ_{A} consists of the systematic part Σ_Y^{A} and the idiosyncratic part Σ_Z^{A}. The risk-related covariance Σ_{R} consists of the systematic part Σ_Y^{R} and the idiosyncratic part Σ_Z^{R}. For both Σ_{A} and Σ_{R}, the factor-related part is also affected by the factor exposure matrix β.

5.3 Ambiguous alphas

Recall that a factor model is regarded as valid when the expected excess return of each risky asset is a linear combination of the expected excess return of factor portfolios, or, equivalently, when the intercept, known also as the alpha of the risky asset, of the linear regression of the excess return of the risky asset onto the excess returns of the factor portfolios is equal to zero. Given this, the investor's belief in the the validity of the factor model can be formulated as a distribution of the values of the alphas, where the investor believing dogmatically in the factor model has the prior distribution concentrated on zeros. This has been done by Pástor (2000), Pástor and Stambaugh (2000), and Wang (2005). In this subsection, we further specialize the model of factor portfolios in the previous subsection to accommodate ambiguous alphas as well as ambiguous factor returns. To do so, consider the following jointly normally distributed random vectors:

$$\begin{pmatrix} G \\ Y \\ F \\ U \end{pmatrix} \sim \mathscr{N}\left(\begin{pmatrix} \mu_G \\ \mu_G \\ \mu_F \\ \mu_F \end{pmatrix}, \begin{pmatrix} \Sigma_G & \Sigma_G & \Sigma_{GF} & \Sigma_{GF} \\ \Sigma_G & \Sigma_Y & \Sigma_{GF} & 0 \\ \Sigma_{FG} & \Sigma_{FG} & \Sigma_F & \Sigma_F \\ \Sigma_{FG} & 0 & \Sigma_F & \Sigma_U \end{pmatrix}\right), \tag{25}$$

where G and Y are L-variate and F and U are N-variate. In the following, we interpret F as alpha and U as the mean of alpha. Assume that

$$\mu_F = \left(\begin{array}{c} 0 \\ \hdashline \hat{\mu}_F \end{array}\right), \qquad \Sigma_U = \left(\begin{array}{c:c} 0 & 0 \\ \hdashline 0 & \hat{\Sigma}_U \end{array}\right) \tag{26}$$

and $\hat{\Sigma}_U \in \mathscr{S}_{++}^{N-L}$. Since $0 \le \Sigma_F \le \Sigma_U$, the matrix Σ_F also has the same form as in (26). Let $\beta \in \boldsymbol{R}^{L\times N}$ and assume that

$$\beta = \left(\begin{array}{c:c} I_L & \hat{\beta} \end{array}\right), \tag{27}$$

where $\hat{\beta} \in \boldsymbol{R}^{L\times(N-L)}$. This specification of β involves no loss of generality, given that $\operatorname{rank}\beta = L$. Denote by $\mathbf{1}_L$ the L-dimensional vector of which the coordinates are all equal to one. To avoid confusion, we write $\mathbf{1}_N$ in place of $\mathbf{1}$, which has been defined as the N-dimensional vector of which the coordinates are all equal to one. Then we define the risky asset returns X by

$$X - R\mathbf{1}_N = \beta^\top(Y - R\mathbf{1}_L) + U. \tag{28}$$

The definition (28) indeed falls into the case of factor portfolios of the previous subsection. Indeed, let $H = F + R(\mathbf{1}_N - \beta^\top\mathbf{1}_L)$ and $Z = U + R(\mathbf{1}_N - \beta^\top\mathbf{1}_L)$. Then $\Sigma_{YZ} = 0$ and $X = \beta^\top Y + Z$. Moreover, $\operatorname{rank}\beta = L$, $\operatorname{Ker}\beta \cap \operatorname{Ker}\Sigma_Z = \{0\}$, and $\operatorname{Ker}\beta + \operatorname{Ker}\Sigma_Z = \boldsymbol{R}^N$.

The mean vector μ_M and the covariance matrix σ_X are then given by

$$\mu_M = \left(\begin{array}{c} \mu_G \\ \hdashline \hat{\beta}^\top \mu_G \end{array}\right) + \left(\begin{array}{c} 0 \\ \hdashline \hat{\mu}_F + R\mathbf{1}_{N-L} - R\hat{\beta}^\top \mathbf{1}_L \end{array}\right), \tag{29}$$

$$\Sigma_X = \left(\begin{array}{c:c} \Sigma_Y & \Sigma_Y \hat{\beta} \\ \hdashline \hat{\beta}^\top \Sigma_Y & \hat{\beta}^\top \Sigma_Y \hat{\beta} \end{array}\right) + \left(\begin{array}{c:c} 0 & 0 \\ \hdashline 0 & \hat{\Sigma}_U \end{array}\right). \tag{30}$$

By (27), the L factors are, in fact, the first L risky assets and $\hat{\beta}$ represents the factor loadings of the last $N-L$ risky assets. By (26), the returns of the first L assets are not subject to the idiosyncratic shocks represented by U. Then

$$E\left[X - R\mathbf{1}_N \,|\, G, F\right] = \beta^\top (G - R\mathbf{1}_L) + F. \tag{31}$$

Thus, conditional on the expected return G of the L factors and the alphas F of the N risky assets, the expected excess returns of the N risky assets are equal to those of the factors, multiplied by the factor loadings β, plus the alphas F. Thus the covariance matrix Σ_F of F represents the degree of the investors (dis-)belief in the alphas being zero, or the factor model being valid. Since the returns Y of the factors may also be ambiguous, this factor model is more general than those of Pástor (2000), Pástor and Stambaugh (2000), and Wang (2005). The following proposition characterizes the optimal portfolios in two extreme cases of model uncertainty.

Proposition 8 *In the case of factor portfolios with model uncertainty, if $\Sigma_G = 0$ and $\hat{\Sigma}_U \in \mathscr{S}_{++}^{N-L}$, then define*

$$v_{\mathrm{R}} = \left(\begin{array}{c} \Sigma_Y^{-1}(\mu_G - R\mathbf{1}_L) \\ \hdashline 0 \end{array}\right),$$

$$v_{\mathrm{A}} = \left(\begin{array}{c} -\hat{\beta}\hat{\Sigma}_U^{-1}\hat{\mu}_F \\ \hdashline \hat{\Sigma}_U^{-1}\hat{\mu}_F, \end{array}\right)$$

where $\mathbf{1}_L$ and $\mathbf{1}_{N-L}$ be the vectors of N 1's and $(N-L)$ 1's. If $\Sigma_G \in \mathscr{S}_{++}^{L}$ and $\Sigma_F = 0$, then *define*

$$v_{\mathrm{R}} = \begin{pmatrix} -\hat{\beta}\hat{\Sigma}_U^{-1}\hat{\mu}_F \\ \hdashline \hat{\Sigma}_U^{-1}\hat{\mu}_F, \end{pmatrix},$$

$$v_{\mathrm{A}} = \begin{pmatrix} \Sigma_Y^{-1}(\mu_G - R\mathbf{1}_L) \\ \hdashline 0 \end{pmatrix}.$$

In either case, $\zeta(0) = v_{\mathrm{R}} + v_{\mathrm{A}}$ *and* $\zeta(\eta) \to v_{\mathrm{R}}$ *as* $\eta \to \infty$*.*

In the first one of this proposition, the factor returns are unambiguous but the validity of the factor model is ambiguous. In the second case, all the factor returns are ambiguous but the validity of the factor model are unambiguous. There are good reasons to assume that the factors are ambiguous. One of them is that the factors are difficult to identify, and one shortcut to capture this difficulty is to assume that the factors are ambiguous.[8] Another reason is that the expected returns of factors are time-varying and, thus, hard to identify. Yet another reason is that since the factors are much fewer than assets, and hence each one of them has a widespread impact on asset returns, the investor may well be more averse to the randomness of factors (which is measured by the coefficient γ) than to the randomness of individual asset returns (which is measured by the coefficient θ).

Note that the two portfolios v_{R} and v_{A} are swapped between the two extreme cases. In the first extreme case, where there is no ambiguity in the factors, there is no demand for the $N - L$ non-factor assets in the ambiguity-free portfolio v_{R}. In the second extreme case, where there is no ambiguity in the alphas, there is no demand for the $N - L$ non-factor assets in the ambiguity-related portfolio v_{A}.

[8] Campbell, Lo, and MacKinlay (1997) presented three approaches to identify factors. An alternative, probably more appropriate, way to capture the difficulty of identifying factors is to assume that the factor loadings β are ambiguous, and hence random variables for which the prior distributions need to be postulated.

6 Bayesian portfolio choice problem

6.1 Bayesian portfolio choice with ambiguity aversion

In the model of this paper, the parameters of the distributions of asset returns are assumed to be known but ambiguous. In reality, those parameters are unknown. It is necessary to estimate them to hold an optimal portfolio dictated by the model, such as a in (5). Merely plugging the estimated values in the formula of the optimal portfolio, however, tends to lead to suboptimal portfolios, because doing so ignores estimation errors. A potentially attractive approach to the optimal portfolio choice problem, initiated by Zellner and Chetty (1965), is the Bayesian approach, in which we postulate that the expected rates of returns of risky assets are random variables that follow some joint distribution, called the prior distribution, and obtain the predictive distribution of future asset returns based on past asset prices. The Bayesian optimal portfolio rule is derived from maximizing the expected utility with respect to the predictive distribution.

The setting of this paper is quite similar to that of the Bayesian approach. Indeed, since $M \sim \mathscr{N}(\mu_M, \Sigma_M)$ and $X|M \sim \mathscr{N}(M, \Sigma_X - \Sigma_M)$, we could think of M as representing the unknown means of risky asset returns, $\mathscr{N}(\mu_M, \Sigma_M)$ as the prior distribution of the means, and $\mathscr{N}(M, \Sigma_X - \Sigma_M)$ as the conditional distribution of asset returns when the true parameter value is M. The Bayesian approach can then be stated as follows. Suppose that T past returns, denoted by $X^1, \dots, X^T$, have been observed. Suppose also that these past returns and the return X, to be resolved after the investor chooses a portfolio, are independently and identically distributed conditional on M. Denote the sample mean $(1/T)\sum_{t=1}^{T} X^t$ by $\bar{X}$, which is a sufficient statistic for M. Then its covariance matrix, $\Sigma_{\bar{X}}$, is equal to $\Sigma_M + T^{-1}(\Sigma_X - \Sigma_M)$.

Now consider the following maximization problem:

$$\begin{aligned} \max_{(a,b)\in \boldsymbol{R}^N \times \boldsymbol{R}} \quad & E\left[u_\gamma\left(u_\theta^{-1}\left(E\left[u_\theta(a^\top X + b) \mid M\right]\right)\right) \mid \bar{X}\right] \\ \text{subject to} \quad & \mathbf{1}^\top a + b \le W. \end{aligned} \tag{32}$$

This is nothing but the optimal portfolio choice problem for the ambiguity-averse investor having the utility function $U_{\gamma,\theta}$ and knowing $\bar{X}$. The optimal portfolio in the Bayesian approach is a special case of this problem in which $\theta = \gamma$. It can thus be written as

$$\begin{aligned} \max_{(a,b)\in \boldsymbol{R}^N\times \boldsymbol{R}} \quad & E\left[u_\theta(a^\top X + b) \,|\, \bar{X}\right] \\ \text{subject to} \quad & \mathbf{1}^\top a + b \leq W. \end{aligned} \tag{33}$$

Proposition 9 *The solution to the problem* (32) *is given by*

$$a = \left(\theta(\Sigma_X - \Sigma_M) + \gamma\left(\Sigma_M - \Sigma_M\Sigma_{\bar{X}}^{-1}\Sigma_M\right)\right)^{-1}\left(\mu_M + \Sigma_M\Sigma_{\bar{X}}^{-1}\left(\bar{X} - \mu_M\right) - R\mathbf{1}\right). \tag{34}$$

In particular, if $\theta = \gamma$, *then*

$$a = \frac{1}{\theta}\left(\Sigma_X - \Sigma_M\Sigma_{\bar{X}}^{-1}\Sigma_M\right)^{-1}\left(\mu_M + \Sigma_M\Sigma_{\bar{X}}^{-1}\left(\bar{X} - \mu_M\right) - R\mathbf{1}\right). \tag{35}$$

This proposition shows that in the Bayesian approach, to find the optimal portfolio for an ambiguity-neutral investor, it is sufficient to know the predictive distribution $\mathscr{N}\left(\mu_M + \Sigma_M\Sigma_{\bar{X}}^{-1}(\bar{X} - \mu_M), \Sigma_X - \Sigma_M\Sigma_{\bar{X}}^{-1}\Sigma_M\right)$; and to find the optimal portfolio of an ambiguity-averse investor it is necessary to know how its covariance matrix $\Sigma_X - \Sigma_M\Sigma_{\bar{X}}^{-1}\Sigma_M$ can be decomposed into two parts. Indeed,

$$\Sigma_X - \Sigma_M\Sigma_{\bar{X}}^{-1}\Sigma_M = (\Sigma_X - \Sigma_M) + \left(\Sigma_M - \Sigma_M\Sigma_{\bar{X}}^{-1}\Sigma_M\right) \tag{36}$$

and the first term on the right hand side coincides with the covariance matrix of the purely risky asset returns (that is, the conditional covariance matrix given M), to which the aversion coefficient θ is applied, and the second term coincides with the conditional covariance matrix of parameter uncertainty given $\bar{X}$, to which the aversion coefficient γ is applied.

Garlappi, Uppal, and Wang (2007), Wang (2005), and Lutgens and Schotman (2010) considered ambiguity-averse investors with utility functions of Gilboa and Schmeidler (1989). Garlappi, Uppal and Wang (2007, Section 2.2) assumed that the set of possible values of unknown expected mean returns (the set of probability distributions over which the expected utility level is minimized) coincides with the confidence interval constructed from past asset returns. They also pointed out that an ambiguity-averse investor needs to know a sort of decomposition (36) to form the optimal portfolio, and we obtain an easier-to-grasp expression (36) thanks to the assumptions in this paper on asset returns and utility functions. In a factor model similar to that introduced in Section 5, Wang (2005) derived the optimal portfolio for investors who have the utility functions of Gilboa and Schmeidler (1989) of which the ambiguity is with regards to the validity of the factor model. Lutgens and Schotman (2010) assumed that an ambiguity-averse investor's utility function coincides with the minimum of the mean-variance utility functions, each based on a mean vector and a covariance matrix provided by an expert. They showed that the optimal portfolio coincides with the optimal portfolio of an ambiguity-neutral investor having the mean-variance utility function based on an weighted average of the mean vectors and the covariance matrices provided by the experts. They also established the mutual fund theorem, in the sense that the optimal portfolios in risky assets for such ambiguity-averse investors are scalar multiples of one another. This should be contrasted with our generalized mutual fund theorem (Theorem 1), and the difference can be attributed to the fact that Lutgens and Schotman (2010) assumed that all investors receive recommendations from the same group of experts and they all set their objective functions as the minimum of the mean-variance utility functions based on the experts' recommendations.

6.2 Various types of Bayesian priors

Although we mentioned some affinity between the Bayesian approach and the approach of this paper in the previous subsection, a fundamental difference lies in the way in which the distribution of $M \sim \mathscr{N}(\mu_M, \Sigma_M)$ is used and interpreted. The approach of this paper follows KMM's axiomatization, in which $\mathscr{N}(\mu_M, \Sigma_M)$ is built in the preference relation defined over a set of acts in the sense of Savage (1954, Section 2.5). Then the optimal portfolio is, quite legitimately, defined as the portfolio that is the most preferred one with respect to the preference relation. On the other hand, the prior distribution in the Bayesian approach is somewhat arbitrarily chosen based on the observer's prior knowledge regarding the source of randomness under consideration or on the ease with which the predictive distribution is derived from the prior distribution. This indicates the possibility for a host of alternative prior distributions.

The most commonly used prior distribution in the Bayesian approach is the diffuse prior, which embodies the idea that there is no prior knowledge on the process generating risky asset returns. For example, Zellner and Chetty (1965), Barry (1974), and Klein and Bawa (1976, Section 3) assumed that both the mean vector and covariance matrix of asset returns are unknown and the prior distribution of the mean vector is diffuse on $\boldsymbol{R}^N$. To see how our analysis will be modified in the context of the diffuse prior, assume that the distribution of M is diffuse (uniform) on $\boldsymbol{R}^N$ and that $X|M \sim \mathscr{N}(M, \Sigma_X)$,[9] then $M|\bar{X} \sim \mathscr{N}\left(\bar{X}, T^{-1}\Sigma_X\right)$, $X|(M, \bar{X}) \sim \mathscr{N}(M, \Sigma_X)$, and $X|\bar{X} \sim \mathscr{N}\left(\bar{X}, (1+T^{-1})\Sigma_X\right)$. Then the Bayesian optimal portfolio, or the solution to the maximization problem (33), coincides with

$$\frac{1}{(1+T^{-1})\theta}\Sigma_X^{-1}(\bar{X} - R\mathbf{1}).$$

This is the same as the solution to the maximization problem (3) in the case of an ambiguity-neutral investor (that is, $\gamma = \theta$) when the estimator $\bar{X}$ of the unknown mean M is plugged into the formula (5), except that the entire portfolio (for the risky assets) is scaled down by factor $(1+T^{-1})^{-1}$ due to the parameter uncertainty. This point was made by Barry (1974, Section 2), Klein and Bawa (1976, Section 3), and Avramov and Zhou (2010, Section 2.1).

The ambiguity-averse investor with the utility function $U_{\gamma,\theta}$, knowing $\bar{X}$, would optimally choose

$$\frac{1}{\theta + T^{-1}\gamma}\Sigma_X^{-1}\left(\bar{X} - R\mathbf{1}\right).$$

Again, this is a scaled-down version of the the solution to the maximization problem (3) in the case of an ambiguity-neutral investor, but scaled-down factor is now equal to $\left(1 + T^{-1} + T^{-1}\eta\right)^{-1}$, where $\eta = \gamma/\theta - 1$. Thus, if the investor is ambiguity-averse, then $\eta > 0$ and the factor is even lower than the factor of the Bayesian optimal portfolio. Thus, although the scaled-down factor is different, incorporating ambiguity does not change the proportions of wealth invested into the risky assets. This result is in sharp contrast with our generalized mutual fund theorem (Theorem 1), as the latter shows that an increase in η, in general, does not scale down the holding of a single mutual fund but changes the composition of multiple mutual

[9]Since M follows the uniform distribution on $\boldsymbol{R}^N$, the mean vector and the covariance matrix, which were denoted by μ_M and Σ_M, do not exist. We thus let $X|M \sim \mathscr{N}(M, \Sigma_X)$, rather than $X|M \sim \mathscr{N}(M, \Sigma_X - \Sigma_M)$.

funds. The difference arises from the specifications of prior distributions, in that while the prior is diffuse in the above-mentioned contributions, our distribution of M, $\mathscr{N}(\mu_M, \Sigma_M)$, is informative.

Unlike Zellner and Chetty (1965), Barry (1974), and Klein and Bawa (1976, Section 3), Frost and Savarino (1986) introduced, in their section III, informative priors such that both the mean vector and covariance matrix of asset returns are unknown and the prior (conditional) distribution of expected returns follow a normal distribution given a covariance matrix, and the prior (marginal) distribution of the covariance matrices is a Wishart distribution of which the mean has the form (7).[10]

Given the intrinsic arbitrariness of the choice of prior distributions, one could employ the so-called empirical Bayesian approach: While one posits a single prior distribution of the unknown parameter in the traditional Bayesian approach, one posits a parametric family of prior distributions of the unknown parameter in the empirical Bayesian approach. In the traditional Bayesian approach, one often uses a Bayesian estimator, which minimizes the mean squared error of estimating the unknown parameter. A Bayesian estimator, however, depends typically on the prior distribution being used and is not implementable in the empirical Bayesian approach. To circumvent this problem, one first constructs an estimator of the parameter specifying the prior distribution based on observed data and then plug it into a Bayesian estimator. The statistic thus constructed is called an empirical Bayesian estimator.

Although there is variety of empirical Bayesian estimators, many of them can be given the Bayes-Stein shrinkage interpretation. For example, Frost and Savarino (1986) applied, in their section IV, the empirical Bayesian estimators of the parameters defining the prior distributions that they introduced in their section III. Jorion (1986) assumed that the prior distributions of expected asset returns constitute a parametric family of normal distributions in which, in our notation, μ_M is always a scalar multiple of $\mathbf{1}$ (that is, the first assumption of Proposition 3 is met by every member of the parametric family) and Σ_M is always a positive multiple of Σ_X (that is, the assumption of Proposition 5 is met by every member

[10] We have, in contrast, assumed that the covariance matrix is known and the matrix of the form (7) is used for the covariance matrix of the normally distributed unknown mean vectors.

of the parametric family). He then used the James-Stein estimator, which can be regarded as an empirical Bayesian estimator, to construct a portfolio. Ledoit and Wolf (2003) also propose a variant of the James-Stein estimator to estimated the covariance matrix of stock returns. They derived the optimal level of the so-called shrinkage factor, relative to the loss function based on the Frobenius norm on the set of symmetric matrices, towards the covariance matrix in a single-factor model, paying special attention to the fact that the sample estimate of the covariance matrix and the estimator for the parameter specifying the prior distribution can be correlated. The portfolio they propose has a similar structure to our optimal portfolio (5) in the sense that covariance matrix used in the portfolio choice is a linear combination of two matrices. However the portfolio (5) is different from theirs because it combines two known covariance matrices with the weight solely determined by risk and ambiguity aversion coefficients. As mentioned in subsection 5.2, Pástor (2000) and Pástor and Stambaugh (2000) introduced priors that reflect an investor's belief in the validity of a factor model. Wang (2005) showed that the shrinkage factor of his variant of the James-Stein estimator is equal to 1/2, which means roughly that the factor model is valid with probability 1/2.

In the Bayesian approach, no single probability distribution, such as $\mathscr{N}(\mu_M, \Sigma_M)$ in our specification of smooth-ambiguity utility functions, can be used to assess which prior distribution is more desirable than another in the portfolio choice problem. In some cases, however, alternative estimators can be compared analytically without knowing true parameters. For example, Kan and Zhou (2007) showed that the Bayesian portfolio under the diffuse prior dominates that obtained simply by plugging in the sample mean. They also proposed (in Section IV) a portfolio that invests in three funds, the risk-free asset, the sample tangency portfolio, and the sample global minimum variance portfolio, and showed that it is similar to the portfolio based on the James-Stein estimator in Jorion (1986), but differs in ratios in which the three funds are invested in.

When no such comparison is possible, simulations of asset returns of which the means, variances, and covariances mimic those of historical stock returns are used. To name just a few example, Frost and Savarino (1986) used the average returns in certainty equivalents of CARA utility functions in simulations of asset returns. Jorion (1986) used the percentage loss in utility levels attained by CARA utility functions in simulations and showed that the portfolio he constructed dominates the portfolio obtained by simply plugging estimated expected returns into the formula of the optimal portfolio (5) and the portfolio obtained from the diffuse prior as in Zellner and Chetty (1965). Kan and Zhou (2007, Section V) also used CARA utility functions to show in simulations that the three-fund portfolio they proposed dominates the portfolio in Jorion (1986). In contrast, Garlappi, Uppal, and Wang (2007, Section 3) used the historical data on international stock market indices to construct confidence intervals of expected index returns for a 121st month from the preceding 120-month observations, for 259 rounds. They showed that the optimal portfolio for the ambiguity-averse investor (having CARA coefficient equal to one) they constructed attains a higher Sharpe ratio than the plug-in portfolio and the portfolio proposed by Jorion (1986).

Unfortunately, as we mentioned in Subsection 4.3, DeMiguel, Garlappi, and Uppal (2009) showed that these Bayesian optimal portfolios do not consistently perform better than the $1/N$ portfolio. They used, in addition to simulations of single-factor models of various numbers of assets, eight data sets of historical returns, such as those containing sector portfolios of S&P 500 portfolio, equity indices of eight countries, and Fama and French's HML and SMB portfolios (which will be mentioned again in Section 5). They calculated the Sharpe ratios and the average returns in certainty equivalents for a CARA utility function (with CARA coefficient equal to one) of fourteen portfolio selection rules that have been proposed in the literature, including the plug-in portfolio and the portfolios proposed by Jorion (1986) and Kan and Zhou (2007), and concluded that none of them consistently outperforms the $1/N$ portfolio in any data set in terms of either criterion.

7 Examples based on the U.S. equity data

In this section, to see if the ambiguity aversion in our model has any quantitatively significant impact on the optimal portfolio in the real-world financial markets, we use the monthly data in the U.S. equity market for the period of August 1926 to December 2013, obtained from Ken French's website. In the first subsection, we conduct the numerical exercise for the so called FF6 portfolios. In the second subsection, we do so in a factor model using the principal component analysis.

7.1 Optimal Portfolio using the FF6 portfolios

The FF6 portfolios are formed according to the values of market equity (abbreviated as ME, which is either Big or Small) and the ratio of book to market equity (abbreviated as B/M, which is either High, Neutral, or Low). The ME is the market cap at the end of June. Firms with negative book equity are not included in any portfolio.

Table 1 reports the sample mean, the sample standard deviation, and the Sharpe ratio of the FF6 portfolios, in addition to the risk-free rate and the market returns. Within each group of a common B/M, the Small ME portfolios (SH, SN, and SL) have higher average returns than the Big ME portfolios (BH, BN, and BL). Within each group of a common ME, a higher B/M leads to a higher average returns. Table 2 shows the sample covariance matrix of the FF6 portfolios. The sample variances of FF6 portfolios are not too dissimilar, except that those of BL and BN are smaller than the others'. In the examples of this subsection, we let the mean vector μ_M be the the sample mean in Table 1 for the six FF portfolios, the risk-free rate R be the average risk-free rate in Table 1, and the unconditional covariance matrix Σ_X be the sample covariance matrix in Table 2.

The following tables report the value of $\zeta(\eta)$, which coincides with the optimal portfolio for the investor with the coefficient θ of constant absolute risk aversion equal to one. Although there is an agreement among researchers that risk aversion is an important determinant of the optimal portfolio selection, there is no unanimous agreement on how large it actually is. For example, Beetsma and Schotman (2001) conclude, based on data from a Dutch television game show, that it is about 0.12. If this is the case, then $\zeta(\eta)$ in the following examples should be multiplied by 1/0.12, according to (5).

The ambiguity-neutral investor's optimal portfolio $\zeta(0) = \Sigma_X^{-1}(\mu_M - R\mathbf{1})$ is shown in Table 3, which is a mean-variance-efficient portfolio. As is often pointed out, such a portfolio is very sensitive to sample estimates and tends to involve large long and short positions. Our example is no exception: the portfolio $\zeta(0)$ sells short the SL portfolio, which has low Sharpe ratio, and holds long the SN portfolio, which has high Sharpe ratio. The SH and BL portfolios have high Sharpe ratios and are also held long. Although BN and BH portfolios have reasonably high Sharpe ratios, they are sold short, presumably due to the covariance structure. Although the investor with the coefficient $\theta = 1$ of constant absolute risk aversion allocates only 2.66% of his wealth into the risky assets, the investor with $\theta = 0.12$ allocates $2.66/0.12 = 22.17\%$ into the risky assets.

Here we consider four examples of the matrix Σ_M, which represents ambiguity in the expected return of the FF6 portfolios. In the first example, we let $\Sigma_M = (1/60)\Sigma_X$. The optimal portfolio $\zeta(\eta)$ is reported in Table 3. The matrix Σ_M roughly measures the standard error of the sample mean estimate with sixty-month observations. Since $Q = (1/60)I_6$, where I_6 is the 6×6 identity matrix, Theorem 1 implies that $\zeta(\eta)$ is a positive multiple of $\zeta(0)$ for every η. Thus the optimal portfolio weight of ambiguity-averse decision makers ($\eta > 0$) are indistinguishable from an ambiguity-neutral investor's optimal portfolio. However, the weight of wealth held in the risk-free asset is different and the wealth invested into each asset is also different. The investor with $\eta = 0$ allocates 97.34% of the wealth into the risk-free asset, although the investor with $\eta = 1000$ allocates 99.85%.

In the second example, we assume that the matrix Σ_M has the form of (7) of which the diagonal elements (variances) are strictly positive and the off-diagonal elements (covariances) are zero. The value of the diagonal elements cannot be arbitrarily chosen, because the conditional covariance matrix $\Sigma_X - \Sigma_M$ must be positive semidefinite. Indeed, $\Sigma_X - \Sigma_M$ is positive definite if the diagonal elements are equal to 0.5, which is, roughly, 1/60 of the smallest diagonal elements of Σ_X in Table 2, but not if they are large, say, 5. Thus the value of each diagonal element of the matrix Σ_M in this example is smaller than the value of the corresponding diagonal element of the matrix Σ_X. However this does not mean that the impact of Σ_M on the optimal portfolios is negligible.

Table 4 reports the optimal portfolio $\zeta(\eta)$ in the second example, with the diagonal elements of Σ_M all equal to 0.5. The investors with larger η invest less in the risky assets. For example, when $\eta = 1000$, only 0.67% of the total wealth is invested into the risky assets. There is no short sale in any assets. Proposition 2 shows that as $\eta \to \infty$, the proportion of wealth invested in each risky asset in the wealth invested in the N risky assets converges to $\Sigma_M^{-1}(\mu_M - R\mathbf{1}_6)$. Since $\Sigma_M^{-1} = 2I_6$, using the values in Table 1, we obtain $\Sigma_M^{-1}(\mu_M - R\mathbf{1}_6) = (1.41, 2.01, 2.41, 1.24, 1.38, 1.81)^\top$. We can observe that $\zeta(1000)$ in Table

4 is almost a scalar multiple of $\Sigma_M^{-1}(\mu_M - R\mathbf{1}_6)$. Indeed, once we normalize these two vectors so that the elements add up to one, then we see that the difference in the proportion of wealth invested in each FF6 portfolio is no larger than 3%.

In our third example, we assume that the diagonal elements of Σ_M are all equal to 0.5 and the off-diagonal elements are all equal to 0.1. Since the off-diagonal elements of Σ_M are positive, the mean returns of the risky assets tend to move together in the same direction. Table 5 reports the optimal portfolio $\zeta(\eta)$ in this case. As in the previous case, the portfolio $\zeta(1000)$ is nearly a scalar multiple of $\Sigma_M^{-1}(\mu_M - R\mathbf{1}_6) = (0.48, 1.23, 1.73, 0.27, 0.45, 0.98)^\top$. Once we normalize these two vectors so that the elements add up to one, then we see that the difference in the proportion of wealth invested in each FF6 portfolio is less than 2%.

In our fourth example, we assume that all the diagonal elements of Σ_M are equal to 0.5 and all off-diagonal elements are equal to -0.1. The mean returns of each pair of two risky assets tend to move in the opposite direction. Furthermore, since $-(6-1)^{-1} \cdot 0.5 = -0.1$, Proposition 4 is applicable: the portfolio $\zeta(\eta)$ converges to a scalar multiple of $\mathbf{1}$ as $\eta \to \infty$. Moreover, as shown in the proof of Proposition 4, $\operatorname{Ker} \Sigma_M$ coincides with the set of all scalar multiples of $\mathbf{1}$. Theorem 2, therefore, implies that there are a unique $(\operatorname{Ker} \Sigma_M)$-based decomposition $(\Sigma_{\mathrm{A}}, \Sigma_{\mathrm{R}})$ of Σ_X and a unique $(v_{\mathrm{R}}, v_{\mathrm{A}}) \in \operatorname{Ker}\, \Sigma_{\mathrm{A}} \times \operatorname{Ker}\, \Sigma_{\mathrm{R}}$ such that $\zeta(0) = v_{\mathrm{A}} + v_{\mathrm{R}}$ with $\zeta(\eta) \to v_{\mathrm{R}}$ as $\eta \to \infty$.

Table 6 shows the optimal portfolio $\zeta(0)$ and its decomposition v_{R} and v_{A} in our fourth example, which are obtained by following the proof method of Lemma 2 and Theorem 2. The ambiguity-related portfolio v_{A} allocates less than one percent in the risky asset but, once normalized so that their components add up to one, the weights among the risky assets can be seen to involve large long/short positions. On the other hand, the unambiguous portfolio v_{R} is nothing but the $1/N$ portfolio. Thus highly ambiguous averse investor holds the portfolio similar to the $1/N$ portfolio in this example, not as a rule of thumb but as a solution to the utility maximization problem.

7.2 Factor model based on the principal component analysis

In this subsection, we conduct a numerical exercise using a factor model with ambiguous alphas in subsection 5.3. The model is derived from the same data of the FF6 portfolios as in subsection 7.1 but based on a different numerical specification of the covariance matrix Σ_Z. Our purpose is to calculate the optimal portfolios for two specifications of ambiguity-related matrices Σ_G and Σ_F using Proposition 8.

Recall that a factor model of subsection 5.3 is given by (28), which is

$$X - R\mathbf{1}_N = \beta^\top (Y - R\mathbf{1}_L) + U.$$

To specify a factor model, we need to determine the number N of risky assets, the number L of factors, the mean vector μ_G and the covariance matrix Σ_Y of factor returns, the mean vector μ_F and the covariance matrix Σ_U of idiosyncratic shocks, and the factor loading matrix $\beta = \left(\begin{array}{c:c} I_L & \hat{\beta} \end{array} \right) \in \boldsymbol{R}^{L\times N}$ with $\hat{\beta} \in \boldsymbol{R}^{L\times(N-L)}$. We shall do so in six steps, spelt out below. The underlying ambiguity, on the other hand, is given by

$$M - R\mathbf{1}_N = \beta^\top (G - R\mathbf{1}_L) + F,$$

and we will specify the covariance matrices Σ_G and Σ_F just before deriving the optimal portfolios.

Step 1 Let $N = 9$ and $L = 3$.

Step 2 Find the three largest eigenvalues and the corresponding eigenvectors v_1, v_2, v_3 of the sample covariance matrix of Table 2. By multiplying scalars if necessary, we can assume without loss of generality that the N coordinates of v_k sum to one for each $k = 1, 2, 3$. This normalization allows us to regard each v_k as a portfolio of the FF6 portfolios, with its coordinates representing the wealth shares allocated to them. We thus call v_k the PCk portfolio (the k-th principal component of the sample covariance matrix), or, simply, PCk, for each $k = 1, 2, 3$.

Step 3 Using the return data of the FF6 portfolios, generate the return data of PC1, PC2, and PC3.

Step 4 Let μ_G and Σ_Y be the sample mean vector and the sample covariance matrix of the return data of PC1, PC2, and PC3 that are generated in step 7.2.

Step 5 Define $\hat{\beta} \in R^{3\times 6}$ so that its transpose, $\hat{\beta}^\top$, coincides with the regression coefficients of the return data of the FF6 portfolios on the return data of PC1, PC2, and PC3 generated in step 7.2. Let $\beta = \left(I_3 \mid \hat{\beta} \right) \in \boldsymbol{R}^{3\times 9}$. Let $\hat{\mu}_F$ be the estimated regression intercepts and $\tilde{\Sigma}_U$ be the estimated covariance matrix of the regression residuals.

Step 6 Denote by $\hat{\Sigma}_U$ the 6×6 matrix that can be obtained from $\tilde{\Sigma}_U$ by replacing all its off-diagonal elements by zeros, and let

$$\mu_F = \begin{pmatrix} 0 \\ \hline \hat{\mu}_F \end{pmatrix} \in \boldsymbol{R}^9, \qquad \Sigma_U = \left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & \hat{\Sigma}_U \end{array}\right) \in \mathscr{S}^9_+.$$

We have thereby determined μ_G, Σ_Y, μ_F, Σ_U, and β. In the factor model we have just constructed, there are three factors and nine assets. The three factors are the three principal components of the sample covariance matrix of the FF6 portfolios. They constitute the first three assets. The other six assets are similar but not identical to the FF6 portfolios, because the off-diagonal elements of the covariance matrix of the residuals in the regression of the FF6 have all been replaced by zeros. Thanks to this replacement, the nine assets are not redundant.

Let us now review some quantitative properties of this factor model.

Table 7 reports that the wealth shares in the factors PC1, PC2, and PC3, and the value weight (VW) portfolio. Among these four, PC1 is the closest to the $1/N$ portfolio. PC1 holds more Small ME portfolios and High B/E portfolios than VW. In other words, PC1 is the portfolio tilting towards small and value stocks.

PC2 contains large long/short positions in the FF6 portfolios. In particular, it holds SL and BH by more than 300% of the net wealth invested in this factor. The Small portfolios are sold short and the Big portfolios are held long. The High B/M portfolios are held more than the Low B/M portfolios. Thus PC2 is the portfolio tilting towards big and value portfolios.

PC3 also holds long/short positions by more than 100% of the net wealth invested in this factor. The differences in the holdings between the Small and Big portfolios is smaller in PC3 than in PC2. PC3 holds more of the Low B/M portfolios than PC2. In particular, it holds the BL portfolio by 200% of the net wealth invested in the factor. Thus PC3 is the portfolio tilting towards big and growth portfolios.

Table 8 reports that the sample mean vector and the sample covariance matrix of the returns of PC1, PC2, and PC3. PC1 has a larger monthly average return of 1.17% than the market return shown in Table 1, but also a larger variance of 45.27. PC2 also has a higher average return of 1.09% but also a huge variance of 444.88, due to its large long/short position. PC3 has a negative average return due to its big and growth property. By the construction using the principal component analysis, the estimated covariance between each factor portfolio returns is close to zero.

Table 9 shows the expected alphas $\hat{\mu}_F$ of the FF6 returns and their factor loadings $\hat{\beta}$. The expected alphas of SN, SH, and BL portfolios are positive but those of SL, BN, and BH are negative. The loadings of PC1 in BL and BN are 0.71 and 0.80, reflecting the fact that the weights on these two are much smaller (12.19% and 13.75%) than the value weighted portfolio (50.84% and 31.19%).

Table 10 shows the estimated covariance matrix $\tilde{\Sigma}_U$ of the residuals. The inverse of the matrix cannot be found numerically. The covariance matrix matrix $\hat{\Sigma}_U$ that we use in this subsection can be obtained by replacing all its off-diagonal elements by zeros. Its inverse matrix can be found numerically.

In the rest of this subsection, we consider the two extreme cases in subsection 5.3. The first extreme case is where the factor returns are unambiguous but the validity of the factor model is ambiguous. That is, $\Sigma_G = 0$. In the second extreme case, all the factor returns are ambiguous but the validity of the factor model are unambiguous. For this case, we assume that $\Sigma_G = \Sigma_Y$ and $\Sigma_F = 0$.

Table 11 reports the optimal portfolio $\zeta(0)$ and its decomposition v_{A} and v_{R} for the first case. The unambiguous (purely risky) portfolio v_{R} consists only of factor portfolios PC1, PC2, and PC3. PC1 is held long, PC2 is held only slightly, and PC3 is sold short. On the other hand, the ambiguity-related portfolio v_{A} essentially consists of FF6 portfolios. The portfolio weight for FF6 part depends largely on the value of alphas in Table 9. In fact, as shown in Table 9, SL, BN, and BH that have negative alpha are sold short. Others that have positive alpha are held long. PC1, PC2, and PC3 are also held in the ambiguity-related portfolio v_{A} to cancel out the purely risky parts of the returns of the FF6 portfolios, as indicated by Theorem 2.

Table 12 reports the optimal portfolio $\zeta(0)$ and its decomposition v_{R} and v_{A} in the second extreme case, where the three factors are ambiguous but the idiosyncratic shocks are unambiguous. As shown in Proposition 8, the optimal portfolio $\zeta(0)$ is the same as in the first case, but the two portfolio v_{R} and v_{A} are now swapped.

For the risk-neutral investor, the portfolios $\zeta(0)$ are identical in the two extreme cases. However, as the ambiguity parameter η becomes large, the optimal portfolios $\zeta(\eta)$ diverge from each other. In the first case, the alphas are ambiguous and the optimal portfolio is chosen to capture the (unambiguous) factor returns. The highly ambiguity averse investor holds almost none of each individual portfolio of FF6, as predicted by Proposition 8. In the second case, the factor returns are ambiguous, and the optimal portfolio is chosen to capture returns in (unambiguous) alphas. Our numerical analysis shows that the highly ambiguity averse investor hold large long/short positions in the FF6 portfolios.

To summarize this section, we have shown that the optimal portfolio based on the actual stock market data depends critically on ambiguity aversion. Moreover, the way we specify the ambiguity-related covariance matrix Σ_M has a significant impact on the optimal portfolio when the investor exhibits ambiguity aversion. Of particular interest is the factor model in which either the factor returns or the assets' alphas, but not both, are ambiguous. The ambiguity-neutral investor would hold the same portfolio, but the ambiguity-averse investor would hold rather different portfolios, depending on which of the two are ambiguous.

8 Conclusion

In this paper, we have studied the nature of the optimal portfolio for an investor who is not only risk-averse but also ambiguity-averse. Our focus has been on the validity of the mutual fund theorem and the asymptotic behavior of the optimal portfolio as the investor becomes extremely ambiguity-averse. We have introduced a factor model accommodating ambiguity aversion and compared the ambiguity-averse investor's portfolio choice problem with the Bayesian portfolio choice problem.

There are at least three possible directions of future research. The first one is to establish the generalized mutual fund theorem in a more general setting where the KMM utility functions need not be of the CARA type and the asset returns need not be normally distributed. The second one is to identify optimal portfolios when the covariance matrix of asset returns is also ambiguous. If this can be done, the connection with the Bayesian portfolio choice problem can be made clearer, as there is a large body of literature on the Bayesian portfolio choice problem that postulates a prior distribution on covariance matrices. The third one is to explore the implications on the equilibrium asset prices of the heterogeneity in investors's ambiguity aversion.

A Lemmas and Proofs

Proof of Lemma 1 By the properties of the moment generating function,

$$\begin{aligned}
& E\left[u_\theta\left(a^\top X + bR\right)|M\right] \\
=& -\exp(-\theta bR)E\left[\exp\left((-\theta a)^\top X\right)|M\right] \\
=& -\exp(-\theta bR)\exp\left((-\theta a)^\top M + \frac{1}{2}(\theta a)^\top \Sigma_{X|M}(\theta a)\right) \\
=& -\exp\left(-\theta\left(a^\top M + Rb - \frac{\theta}{2}a^\top \Sigma_{X|M} a\right)\right).
\end{aligned}$$

Then it follows from (2) that

$$\varphi_{\gamma,\theta}\left(E\left[u_\theta\left(a^\top X + bR\right)|M\right]\right) = -\exp\left(-\gamma\left(a^\top M + Rb - \frac{\theta}{2}a^\top \Sigma_{X|M} a\right)\right).$$

Thus, again by the properties of the moment generating function,

$$
\begin{aligned}
& U_{\gamma,\theta}(a^\top X + bR) \\
=& E\left[-\exp\left(-\gamma\left(a^\top M + Rb - \frac{\theta}{2}a^\top \Sigma_{X|M} a\right)\right)\right] \\
=& -\exp\left(-\gamma\left(Rb - \frac{\theta}{2}a^\top \Sigma_{X|M} a\right)\right) E\left[\exp\left((-\gamma a)^\top M\right)\right] \\
=& -\exp\left(-\gamma\left(Rb - \frac{\theta}{2}a^\top \Sigma_{X|M} a\right)\right) \exp\left(\mu_M^\top(-\gamma a) + \frac{1}{2}(-\gamma a)^\top \Sigma_M (-\gamma a)\right) \\
=& -\exp\left(-\gamma V_{\gamma,\theta}(a,b)\right).
\end{aligned}
$$

///

The following lemma is not covered by Bosch (1987) but can be proved by modifying the proof of his Theorem 5.

Lemma 4 *There is a basis of $\boldsymbol{R}^N$ that consists of eigenvectors of Q, and all the eigenvalues of Q belong to the closed unit interval* $[0,1]$.

Proof of Lemma 4 Since $\Sigma_X^{-1} \in \mathscr{S}_{++}^N$, there are an orthonormal matrix H (that is, H^{-1} exists and coincides with $H^\top$) and a diagonal matrix Γ, of which all the diagonal elements are strictly positive, such that $\Sigma_X^{-1} = H\Gamma H^{-1}$. Denote by $\Gamma^{1/2}$ the diagonal matrix of which the diagonal elements are the square roots of those of Γ, and write $A = H\Gamma^{1/2}H^{-1}$. Since H is orthonormal, A is symmetric and positive definite. Hence $A\Sigma_M A$ is symmetric and positive semidefinite. Therefore, there are an orthonormal matrix $\hat{H}$ and a diagonal matrix $\hat{\Gamma}$, of which all the diagonal elements are nonnegative, such that $A\Sigma_M A = \hat{H}\hat{\Gamma}\hat{H}^{-1}$. Since $A^2 = \Sigma_X^{-1}$,

$$
Q = A(A\Sigma_M A)A^{-1} = (A\hat{H})\hat{\Gamma}(\hat{H}^{-1}A^{-1}) = (A\hat{H})\hat{\Gamma}(A\hat{H})^{-1}.
$$

This shows that there is a basis of $\boldsymbol{R}^N$ that consists of eigenvectors of Q, and all the corresponding eigenvalues of are non-negative.

It remains to show that all the eigenvalues are less than one. To do so, let v be an eigenvalue of Q and λ be the eigenvalue that corresponds to v. Then $Qv = \lambda v$ and hence

$\Sigma_M v = \lambda \Sigma_X v$. Thus $v^\top \Sigma_M v = \lambda v^\top \Sigma_X v$, that is,

$$\lambda = \frac{v^\top \Sigma_M v}{v^\top \Sigma_M v + v^\top \Sigma_{X|M} v}.$$

Since $v^\top \Sigma_{X|M} v \geq 0$, $0 \leq \lambda \leq 1$. ///

Proof of Theorem 1 Let Λ be the set of all eigenvalues of Q. It follows from Lemma 4 that $\Lambda \subset [0, 1]$. For each $\lambda \in \Lambda$, denote by V_λ the eigenspace that correspond to λ. It also follows from Lemma 4 that V_λ is a linear subspace of $\boldsymbol{R}^N$, $(V_\lambda)_{\lambda \in \Lambda}$ is linearly independent (that is, if $v_\lambda \in V_\lambda$ for every $\lambda \in \Lambda$ and $\sum_{\lambda \in \Lambda} v_\lambda = 0$, then $v_\lambda = 0$ for every $\lambda \in \Lambda$), and $\sum_{\lambda \in \Lambda} V_\lambda = \boldsymbol{R}^N$.

Then, for each $\lambda \in \Lambda$, there is a $v_\lambda \in V_\lambda$ such that $\zeta(0) = \sum_{\lambda \in \Lambda} v_\lambda$. Since $\zeta(0) \neq 0$, there is a $\lambda \in \Lambda$ such that $v_\lambda \neq 0$. Let $\{\lambda_1, \lambda_2, \dots, \lambda_K\}$ be the set of all such λ's. We can assume that $\lambda_1 < \lambda_2 < \cdots < \lambda_K$. For each k, write $v_k = v_{\lambda_k}$, then $\zeta(0) = \sum_{k=1}^K v_k$.

Since $(I + \eta Q) v_k = (1 + \eta \lambda_k) v_k$, $(I + \eta Q)^{-1} v_k = (1 + \eta \lambda_k)^{-1} v_k$. Thus,

$$\zeta(\eta) = (I + \eta Q)^{-1} \zeta(0) = \sum_{k=1}^K (I + \eta Q)^{-1} v_k = \sum_{k=1}^K \frac{1}{1 + \lambda_k \eta} v_k.$$

///

Proof of Lemma 2 First, we prove this lemma when $\Sigma \in \mathscr{S}^N_{++}$. Then we prove it for the general case.

Assume that $\Sigma \in \mathscr{S}^N_{++}$. Let $S_1 = S$, and S_2 be the Σ-orthogonal complement of S_1. For each $i = 1, 2$, let W_i be the $N \times N$ matrix that represents the Σ-orthogonal projection onto S_{3-i}, that is, $W_i x = x$ for every $x \in S_{3-i}$ and $W_i x = 0$ for every $x \in S_i$. Then $W_1 + W_2 = I_N$ and $W_1^\top \Sigma W_2 = 0$. For each i, define $\Sigma_i = W_i^\top \Sigma W_i$. Then $(\Sigma_1, \Sigma_2) \in \mathscr{S}^N_+ \times \mathscr{S}^N_+$. Since $\Sigma \in \mathscr{S}^N_{++}$, $\operatorname{Ker} \Sigma_i = \operatorname{Ker} W_i = S_i$. Moreover,

$$\begin{aligned}\Sigma &= (W_1 + W_2)^\top \Sigma (W_1 + W_2) \\ &= W_1^\top \Sigma W_1 + W_1^\top \Sigma W_2 + W_2^\top \Sigma W_1 + W_2^\top \Sigma W_2 = \Sigma_1 + \Sigma_2.\end{aligned}$$

The proof of the existence of an S-based decomposition is thus completed.

To prove the uniqueness, it suffices to show that $\Sigma_1 = W_1^\top \Sigma W_1$ whenever (Σ_1, Σ_2) is an S-based decomposition of Σ. Note, then, that for each i, $\Sigma_i W_{3-i} = 0$ because $\operatorname{Col} W_{3-i} = S_i = \operatorname{Ker} \Sigma_i$. Thus

$$\begin{aligned} W_1^\top \Sigma W_1 &= W_1^\top (\Sigma_1 + \Sigma_2) W_1 = W_1^\top \Sigma_1 W_1 + W_1^\top \Sigma_2 W_1 \\ &= W_1^\top \Sigma_1 W_1 = (I_N - W_2)^\top \Sigma_1 (I_N - W_2) \\ &= \Sigma_1 - W_2^\top \Sigma_1 - \Sigma_1 W_2 + W_2^\top \Sigma_1 W_2 = \Sigma_1. \end{aligned}$$

As for the general case where $\Sigma \notin \mathscr{S}^N_{++}$, write $K = \operatorname{rank} \Sigma$ and let $V \in \boldsymbol{R}^{N \times K}$. Assume that the column vectors of V constitutes an orthonormal basis of $\operatorname{Col} \Sigma$. Then $V^\top \Sigma V \in \mathscr{S}^K_{++}$. Define $T = \{w \in \boldsymbol{R}^K \mid Vw \in S\}$. By this lemma for the case where Σ has a full rank, there is a unique T-based decomposition (Γ_1, Γ_2) of $V^\top \Sigma V$. For each $i = 1, 2$, define $\Sigma_i = V \Gamma_i V^\top$. We shall prove that (Σ_1, Σ_2) is an S-based decomposition of Σ. Since $\Gamma_i \in \mathscr{S}^K_+$, $\Sigma_i \in \mathscr{S}^N_+$. Moreover,

$$\Sigma_1 + \Sigma_2 = V \left(V^\top \Sigma V \right) V^\top = \left(V V^\top \right) \Sigma \left(V V^\top \right).$$

It can be easily shown that $VV^\top$ belongs to $\mathscr{S}^N_+$ and represents the orthogonal projection onto $\operatorname{Col} V$. Since $\operatorname{Col} V = \operatorname{Col} \Sigma$, $\left(V V^\top \right) \Sigma \left(V V^\top \right) = \Sigma$. Thus $\Sigma_1 + \Sigma_2 = \Sigma$. Since $\operatorname{rank} \Sigma_i = \operatorname{rank} \Gamma_i$ and $\operatorname{rank} \Gamma_1 + \operatorname{rank} \Gamma_2 = K$, $\operatorname{rank} \Sigma_1 + \operatorname{rank} \Sigma_2 = \operatorname{rank} \Sigma$. It remains to prove that $S = \operatorname{Ker} \Sigma_1$. Since $V^\top V = I_K$, $V^\top \Sigma_1 V = \Gamma_1$. Thus $\operatorname{Ker} \Gamma_1 = \{w \in \boldsymbol{R}^K \mid Vw \in \operatorname{Ker} \Sigma_1\}$. Since $\operatorname{Ker} \Gamma_1 = T$, $\{w \in \boldsymbol{R}^K \mid Vw \in S\} = \{w \in \boldsymbol{R}^K \mid Vw \in \operatorname{Ker} \Sigma_1\}$. Since $\operatorname{rank} V = K$, $S \cap \operatorname{Col} V = \operatorname{Ker} \Sigma_1 \cap \operatorname{Col} V$. Thus $S \cap \operatorname{Col} \Sigma = \operatorname{Ker} \Sigma_1 \cap \operatorname{Col} \Sigma$. It can be easily shown that since $S \supseteq \operatorname{Ker} \Sigma$, $S = (S \cap \operatorname{Col} \Sigma) + \operatorname{Ker} \Sigma$. Since $\Sigma - \Sigma_1 = \Sigma_2 \in \mathscr{S}^N_+$, $\operatorname{Ker} \Sigma_1 \supseteq \operatorname{Ker} \Sigma$. It can thus be analogously shown that $\operatorname{Ker} \Sigma_1 = (\operatorname{Ker} \Sigma_1 \cap \operatorname{Col} \Sigma) + \operatorname{Ker} \Sigma$. Hence $S = \operatorname{Ker} \Sigma_1$.

Let V and T be as in the previous paragraph. To show the uniqueness in the general case, it suffices to prove that for every S-based decomposition (Σ_1, Σ_2) of Σ, $\left(V^\top \Sigma_1 V, V^\top \Sigma_2 V \right)$ coincides the unique T-based decomposition of $V^\top \Sigma V$. Indeed, if this is the case, then let (Γ_1, Γ_2) be the unique T-based decomposition of $V^\top \Sigma V$. Then, for every S-based decomposition (Σ_1, Σ_2) of Σ, $\Gamma_i = V^\top \Sigma_i V$ for each $i = 1, 2$. Since $VV^\top$ represents the orthogonal onto $\operatorname{Col} \Sigma$ and $\operatorname{Col} \Sigma_i \subseteq \operatorname{Col} \Sigma$, $V \Gamma_i V^\top = \Sigma_i$. Since this is true for every S-based decomposition (Σ_1, Σ_2) of Σ, such a decomposition is, in fact, unique. Let's prove that for every S-based decomposition (Σ_1, Σ_2) of Σ, $\left(V^\top \Sigma_1 V, V^\top \Sigma_2 V\right)$ coincides with the unique T-based decomposition of $V^\top \Sigma V$. By construction, $\left(V^\top \Sigma_1 V, V^\top \Sigma_2 V\right) \in \mathscr{S}_+^K \times \mathscr{S}_+^K$ and $V^\top \Sigma_1 V + V^\top \Sigma_2 V = V^\top \Sigma V$. Since $\operatorname{Col} \Sigma_i \subseteq \operatorname{Col} \Sigma = \operatorname{Col} V$, $\operatorname{rank} V^\top \Sigma_i V = \operatorname{rank} \Sigma_i$ and $\operatorname{rank} V^\top \Sigma V = \operatorname{rank} \Sigma$. Hence $\operatorname{rank} V^\top \Sigma_1 V + \operatorname{rank} V^\top \Sigma_2 V = \operatorname{rank} \Sigma_1 + \operatorname{rank} \Sigma_2 = \operatorname{rank} \Sigma$. For every $w \in \boldsymbol{R}^K$, $w \in \operatorname{Ker} V^\top \Sigma_1 V$ if and only if $Vw \in \operatorname{Ker} \Sigma_1$. Since $S = \operatorname{Ker} \Sigma_1$, this holds if and only if $Vw \in S$. Thus $\operatorname{Ker} V^\top \Sigma_1 V = T$. The proof is thus completed. ///

Lemma 5 *Let $(\underline{\Sigma}, \overline{\Sigma}) \in \mathscr{S}_+^N \times \mathscr{S}_+^N$ and assume that $\overline{\Sigma} - \underline{\Sigma} \in \mathscr{S}_+^N$. Let $(\Sigma_1, \Sigma_2) \in \mathscr{S}_+^N \times \mathscr{S}_+^N$ be the $(\operatorname{Ker} \underline{\Sigma})$-based decomposition of $\overline{\Sigma}$. Then, for every $v \in \boldsymbol{R}^N$ and every $\lambda \neq 0$, $\lambda \overline{\Sigma} v = \underline{\Sigma} v$ if and only if $v \in \operatorname{Ker} \Sigma_2$ and $\lambda \Sigma_1 v = \underline{\Sigma} v$.*

Proof of Lemma 5 If $v \in \operatorname{Ker} \Sigma_2$ and $\lambda \Sigma_1 v = \underline{\Sigma} v$, then $\lambda \overline{\Sigma} = \lambda(\Sigma_1 + \Sigma_2) v = \lambda \Sigma_1 v = \underline{\Sigma} v$. Suppose, conversely, that $\lambda \overline{\Sigma} = \underline{\Sigma} v$. Then there is a $(v_1, v_2) \in \operatorname{Ker} \Sigma_1 \times \operatorname{Ker} \Sigma_2$ such that $v = v_1 + v_2$. Then $\overline{\Sigma} v = (\Sigma_1 + \Sigma_2)(v_1 + v_2) = \Sigma_1 v_2 + \Sigma_2 v_1$. Thus, $\underline{\Sigma} v - \lambda \Sigma_1 v_2 = \lambda \Sigma_2 v_1$. Since $\operatorname{Col} \underline{\Sigma} = \operatorname{Col} \Sigma_1$, $\underline{\Sigma} v - \lambda \Sigma_1 v_2 \in \operatorname{Col} \Sigma_1$. Since $\lambda \neq 0$, $\Sigma_2 v_1 \in \operatorname{Col} \Sigma_1 \cap \operatorname{Col} \Sigma_2 = \{0\}$. Thus, $v_1 \in \operatorname{Ker} \Sigma_2$ and hence $v_1 = 0$. Thus $v = v_2 \in \operatorname{Ker} \Sigma_2$. Hence $\overline{\Sigma} v = (\Sigma_1 + \Sigma_2) v = \Sigma_1 v$. Since $\lambda \overline{\Sigma} = \underline{\Sigma} v$, $\lambda \Sigma_1 v = \underline{\Sigma} v$. ///

Proof of Theorem 2 The existence of $(\Sigma_{\mathrm{A}}, \Sigma_{\mathrm{R}})$ follows from Lemma 2 by letting $\Sigma = \Sigma_X$ and $S = \operatorname{Ker} \Sigma_M$. Since $\boldsymbol{R}^N$ is the direct sum of $\operatorname{Row} \Sigma_{\mathrm{A}}$ and $\operatorname{Row} \Sigma_{\mathrm{R}}$, there exists a unique $(w_{\mathrm{A}}, w_{\mathrm{R}}) \in \operatorname{Row} \Sigma_{\mathrm{A}} \times \operatorname{Row} \Sigma_{\mathrm{R}}$ such that $\mu_M - R\mathbf{1} = w_{\mathrm{A}} + w_{\mathrm{R}}$. Note, then, that the linear transformation defined by Σ_{A} maps $\operatorname{Ker} \Sigma_{\mathrm{R}}$ onto $\operatorname{Row} \Sigma_{\mathrm{A}}$. Indeed, the linear transformation is one-to-one when its domain is restricted on $\operatorname{Ker} \Sigma_{\mathrm{R}}$, because $\operatorname{Ker} \Sigma_{\mathrm{A}} \cap \operatorname{Ker} \Sigma_{\mathrm{R}} = \{0\}$. Since $\dim \operatorname{Ker} \Sigma_{\mathrm{R}} = N - \dim \operatorname{Row} \Sigma_{\mathrm{R}} = \dim \operatorname{Row} \Sigma_{\mathrm{A}}$, the linear transformation maps $\operatorname{Ker} \Sigma_{\mathrm{R}}$ onto $\operatorname{Row} \Sigma_{\mathrm{A}}$. Thus, there exists a unique $v_{\mathrm{A}} \in \operatorname{Ker} \Sigma_{\mathrm{R}}$ such that $\Sigma_{\mathrm{A}} v_{\mathrm{A}} = w_{\mathrm{A}}$. Similarly, there exists a unique $v_{\mathrm{R}} \in \operatorname{Ker} \Sigma_{\mathrm{A}}$ such that $\Sigma_{\mathrm{R}} v_{\mathrm{R}} = w_{\mathrm{R}}$. Therefore,

$$\Sigma_X(v_{\mathrm{R}} + v_{\mathrm{A}}) = (\Sigma_{\mathrm{A}} + \Sigma_{\mathrm{R}})(v_{\mathrm{R}} + v_{\mathrm{A}}) = \Sigma_{\mathrm{A}} v_{\mathrm{A}} + \Sigma_{\mathrm{R}} v_{\mathrm{R}} = w_{\mathrm{A}} + w_{\mathrm{R}} = \mu_M - R\mathbf{1}$$

and $\zeta(0) = v_{\mathrm{R}} + v_{\mathrm{A}}$. Thanks to Theorem 1, to prove that $\zeta(\eta) \to v_{\mathrm{R}}$ as $\eta \to \infty$, it suffices to prove that $v_{\mathrm{R}} \in \operatorname{Ker} Q$ and v_{A} is a linear combination of the eigenvectors of Q that correspond to the strictly positive eigenvalues. Since $\operatorname{Ker}\Sigma_{\mathrm{A}} = \operatorname{Ker} Q$, $v_{\mathrm{R}} \in \operatorname{Ker} Q$. Since $v_{\mathrm{A}} \in \operatorname{Ker}\Sigma_{\mathrm{R}}$, it suffices to show that there is a basis of $\operatorname{Ker}\Sigma_{\mathrm{R}}$ that consists of the eigenvectors of Q that corresponds to the strictly positive eigenvalues. By Lemma 5, all eigenvectors of Q that correspond to strictly positive eigenvalues belong to $\operatorname{Ker}\Sigma_{\mathrm{R}}$. Moreover, since $\dim \operatorname{Ker}\Sigma_{\mathrm{R}} = N - \operatorname{Ker}\Sigma_{\mathrm{A}} = N - \operatorname{Ker} Q$, there is a basis of $\operatorname{Ker}\Sigma_{\mathrm{R}}$ that consists of these eigenvectors. ///

Proof of Proposition 2 Since $\Sigma_M \in \mathscr{S}^N_{++}$, all of its eigenevalues are strictly positive. In particular, $\lambda_1 > 0$. Hence, by Theorem 1,

$$\begin{aligned}\frac{1}{\mathbf{1}^\top \zeta(\eta)}\zeta(\eta) &= \frac{1}{\mathbf{1}^\top\left(\sum_{k=1}^K \frac{1}{1+\lambda_k\eta} v_k\right)} \sum_{k=1}^K \frac{1}{1+\lambda_k\eta} v_k = \frac{1}{\mathbf{1}^\top\left(\sum_{k=1}^K \frac{1+\lambda_1\eta}{1+\lambda_k\eta} v_k\right)} \sum_{k=1}^K \frac{1+\lambda_1\eta}{1+\lambda_k\eta} v_k \\ &\to \frac{1}{\mathbf{1}^\top\left(\sum_{k=1}^K \frac{\lambda_1}{\lambda_k} v_k\right)} \sum_{n=1}^K \frac{\lambda_1}{\lambda_k} v_k = \frac{1}{\mathbf{1}^\top\left(\sum_{k=1}^K \lambda_k^{-1} v_k\right)} \sum_{k=1}^K \lambda_k^{-1} v_k\end{aligned}$$

as $\eta \to \infty$. Since $Q\left(\sum_{k=1}^K \lambda_k^{-1} v_k\right) = \left(\sum_{k=1}^K \lambda_k^{-1} Q v_k\right) = \sum_{k=1}^K v_k = \zeta(0)$,

$$\sum_{k=1}^K \lambda_k^{-1} v_k = Q^{-1}\zeta(0) = \Sigma_M^{-1}\left(\Sigma_X\right)\left(\Sigma_X\right)^{-1}\left(\mu_M - R\mathbf{1}\right) = \Sigma_M^{-1}\left(\mu_M - R\mathbf{1}\right)$$

Thus the proof is completed. ///

Proof of Proposition 3 Since Σ_M is invertible, $\Sigma_M^{-1}\mathbf{1} = (\sigma^2 + (N-1)\kappa)^{-1}\mathbf{1}$. Hence

$$\Sigma_M^{-1}(\mu_M - R\mathbf{1}) = (\delta - R)(\sigma^2 + (N-1)\kappa)^{-1}\mathbf{1}.$$

Since $\delta - R \neq 0$,

$$\frac{1}{\mathbf{1}^\top \Sigma_M^{-1}(\mu_M - R\mathbf{1})} \Sigma_M^{-1}(\mu_M - R\mathbf{1}) = \frac{1}{N}\mathbf{1},$$

and the proof is completed by Proposition 2. ///

Proof of Corollary 1 Since $\Sigma_X - \Sigma_M = (1-\lambda)\Sigma_M \in \mathscr{S}^N_+$, $\lambda \leq 1$. Thus $1+\lambda\eta > 0$ for every $\eta > -1$. Hence, if $\zeta_n(\eta) > 0$ for some $\eta > -1$, then the n-th coordinate of v is also strictly positive. Thus $\zeta_n(\eta) > 0$ for every $\eta > -1$. ///

Proof of Proposition 6 By Theorem 1, it suffices to prove that λ is, if any, the only strictly positive eigenvalue of Q. So let $v \in \boldsymbol{R}^N \setminus \{0\}$ and $\kappa > 0$ and suppose that $Qv = \kappa v$. Then, by the only-if part of Lemma 5, $v \in \operatorname{Ker}(\Sigma_X - \lambda^{-1}\Sigma_M)$ and $\kappa(\lambda^{-1}\Sigma_M)v = \Sigma_M v$. Thus $\Sigma_X v = \lambda^{-1}\Sigma_M v$ and $(\kappa\lambda^{-1})\Sigma_M v = \Sigma_M v$. Hence $\lambda v = Qv$ and $\kappa = \lambda$ because $\Sigma_M v \neq 0$. Thus λ is, if any, the only strictly positive eigenvalue of Q. ///

Proof of Corollary 2 If $\lambda > 1$, then $\Sigma_X - \lambda^{-1}\Sigma_M = (1-\lambda^{-1})\Sigma_X + \lambda^{-1}(\Sigma_X - \Sigma_M) \in \mathscr{S}^N_{++}$ because $(1-\lambda^{-1})\Sigma_X \in \mathscr{S}^N_{++}$ and $\lambda^{-1}(\Sigma_X - \Sigma_M) \in \mathscr{S}^N_+$. Since $v_{\mathrm{A}} \in \operatorname{Ker}(\Sigma_X - \lambda^{-1}\Sigma_M)$ under the assumption of Proposition 6, $v_{\mathrm{A}} = 0$. On the other hand, since $v_{\mathrm{R}} \in \operatorname{Ker}\Sigma_M = \boldsymbol{R}^L \times \{0\}$, the n-th coordinate of v_{R} is equal to zero for every $n > L$. Thus this corollary has been trivially established for the case of $\lambda > 1$.

As for the case of $\lambda \leq 1$, note that $1+\lambda\eta > 0$ for every $\eta > -1$. Since the n-th coordinate of v_{R} is equal to zero for every $n > L$, if $\zeta_n(\eta) > 0$ for some $\eta > -1$, the n-th coordinate of v_{A} is strictly positive. Hence, for every $n > L$, $\zeta_n(\eta) > 0$ for every $\eta > -1$ and converges strictly decreasingly to 0 as $\eta \to \infty$. ///

Proof of Proposition 9 Since $\bar{X}|M \sim \mathscr{N}(M, (1/T)(\Sigma_X - \Sigma_M))$, the law of total variance implies that

$$\begin{pmatrix} M \\ X \\ \bar{X} \end{pmatrix} \sim \mathscr{N}\left(\begin{pmatrix} \mu_M \\ \mu_M \\ \mu_M \end{pmatrix}, \begin{pmatrix} \Sigma_M & \Sigma_M & \Sigma_M \\ \Sigma_M & \Sigma_X & \Sigma_M \\ \Sigma_M & \Sigma_M & \Sigma_{\bar{X}} \end{pmatrix}\right),$$

where $\Sigma_{\bar{X}} = \Sigma_M + T^{-1}(\Sigma_X - \Sigma_M)$. Thus

$$\begin{pmatrix} M \\ X \end{pmatrix}\Bigg| \bar{X} \sim \mathscr{N}\left(\begin{pmatrix} \mu_M + \Sigma_M\Sigma_{\bar{X}}^{-1}(\bar{X}-\mu_M) \\ \mu_M + \Sigma_M\Sigma_{\bar{X}}^{-1}(\bar{X}-\mu_M) \end{pmatrix}, \begin{pmatrix} \Sigma_M - \Sigma_M\Sigma_{\bar{X}}^{-1}\Sigma_M & \Sigma_M - \Sigma_M\Sigma_{\bar{X}}^{-1}\Sigma_M \\ \Sigma_M - \Sigma_M\Sigma_{\bar{X}}^{-1}\Sigma_M & \Sigma_X - \Sigma_M\Sigma_{\bar{X}}^{-1}\Sigma_M \end{pmatrix}\right). \tag{37}$$

Hence,

$$X \left| \begin{pmatrix} M \\ \bar{X} \end{pmatrix} \right. \sim \mathscr{N}(M, \Sigma_X - \Sigma_M). \tag{38}$$

The predictive distribution of return X given the sample estimate $\bar{X}$ thus coincides with $\mathscr{N}\left(\mu_M + \Sigma_M \Sigma_{\bar{X}}^{-1}(\bar{X} - \mu_M), \Sigma_X - \Sigma_M \Sigma_{\bar{X}}^{-1} \Sigma_M\right)$. We can then apply (37), (38), and (5) to establish (34). By putting $\theta = \gamma$, we obtain if (35). ///

Lemma 6 *For every $A \in \mathscr{S}^N$ with* $\operatorname{Row} A \subseteq \operatorname{Row} \beta$*, there is a $C \in \mathscr{S}^L$ such that $A = \beta^\top C \beta$.*

Proof of Lemma 6 Since A is symmetric, there are an orthonormal $N \times N$ matrix V and a diagonal $N \times N$ matrix D such that $A = V^\top D V$. Denote the diagonal elements of D by $d_1, d_2, \ldots, d_N$. Denote by D_1 the diagonal $N \times N$ matrix of which the n-diagonal elements is equal to $(\max\{d_n, 0\})^{1/2}$, and denote by D_2 the diagonal $N \times N$ matrix of which the n-th diagonal elements is equal to $(\max\{-d_n, 0\})^{1/2}$. For each $k = 1, 2$, write $A_k = V^\top D_k V \in \mathscr{S}_+^N$, then $\operatorname{Row} A_k \subseteq \operatorname{Row} A \subseteq \operatorname{Row} \beta$. Thus there is a $B_k \in \boldsymbol{R}^{N \times L}$ such that $A_k = B_k \beta$. Write $C_k = B_k^\top B_k \in \mathscr{S}_+^L$ and $C = C_1 - C_2$, then $C \in \mathscr{S}^L$ and

$$\begin{aligned} A &= V^\top D V = V^\top \left(D_1^2 - D_2^2\right) V = V^\top D_1^2 V - V^\top D_2^2 V \\ &= A_1^\top A_1 - A_2^\top A_2 = \beta^\top C_1 \beta - \beta^\top C_2 \beta = \beta^\top C \beta. \end{aligned}$$

///

Proof of Lemma 3 For every $v \in \boldsymbol{R}^N$, if $\beta v = 0$, then $v^\top \left(\beta^\top \Sigma_{GH} + \Sigma_{HG} \beta\right) v = 0$. Thus $\operatorname{Ker} \beta \subseteq \operatorname{Ker} \left(\beta^\top \Sigma_{GH} + \Sigma_{HG} \beta\right)$, or, equivalently, $\operatorname{Row} \beta \supseteq \operatorname{Row} \left(\beta^\top \Sigma_{GH} + \Sigma_{HG} \beta\right)$. Lemma 6 implies that there is a $\Gamma \in \mathscr{S}^L$ such that $\beta^\top \Sigma_{GH} + \Sigma_{HG} \beta = \beta^\top \Gamma \beta$.

For every such Γ, we shall now prove that $0 \leq \Sigma_G + \Gamma \leq \Sigma_Y$. As we showed in Footnote 5, for every $w \in \boldsymbol{R}^L$, there is a $v \in \operatorname{Ker} \Sigma_Z$ such that $w = \beta v$. Thus, it suffices to prove that $0 \leq (\beta v)^\top (\Sigma_G + \Gamma)(\beta v) \leq (\beta v)^\top \Sigma_Y (\beta v)$ for every $v \in \operatorname{Ker} \Sigma_Z$. Indeed, for every

$v \in \operatorname{Ker} \Sigma_Z$,

$$
\begin{aligned}
v^\top \Sigma_X v &= v^\top \left(\beta^\top \Sigma_Y \beta + \Sigma_Z \right) v = v^\top \beta^\top \Sigma_Y \beta v, \\
v^\top \Sigma_M v &= v^\top \left(\beta^\top \left(\Sigma_G + \Gamma \right) \beta + \Sigma_H \right) v = v^\top \beta^\top \left(\Sigma_G + \Gamma \right) \beta v,
\end{aligned}
$$

because $\operatorname{Ker} \Sigma_H \supseteq \operatorname{Ker} \Sigma_Z$. Since $0 \leq v^\top \Sigma_M v \leq v^\top \Sigma_X v$ (which is, in turn, because $0 \leq \Sigma_M \leq \Sigma_X$), the proof is completed. ///

Proof of Proposition 7 By construction, $(\Sigma_{\mathrm{A}}, \Sigma_{\mathrm{R}}) \in \mathscr{S}_+^N \times \mathscr{S}_+^N$ and $\Sigma_{\mathrm{A}} + \Sigma_{\mathrm{R}} = \Sigma_X$. Note that

$$
\begin{aligned}
\operatorname{Ker} \Sigma_M &= \operatorname{Ker} \beta^\top (\Sigma_G + \Gamma) \beta \cap \operatorname{Ker} \Sigma_H, \\
\operatorname{Ker} \Sigma_{\mathrm{A}} &= \operatorname{Ker} \beta^\top \Sigma_Y^{\mathrm{A}} \beta \cap \operatorname{Ker} \Sigma_Z^{\mathrm{A}}.
\end{aligned}
$$

Since $\operatorname{Ker} \beta^\top = \{0\}$,

$$
\begin{aligned}
\operatorname{Ker} \beta^\top (\Sigma_G + \Gamma) \beta &= \{ v \in \boldsymbol{R}^N \mid \beta v \in \operatorname{Ker} (\Sigma_G + \Gamma) \}, \\
\operatorname{Ker} \beta^\top \Sigma_Y^{\mathrm{A}} \beta &= \{ v \in \boldsymbol{R}^N \mid \beta v \in \operatorname{Ker} \Sigma_Y^{\mathrm{A}} \}
\end{aligned}
$$

Since $\operatorname{Ker} (\Sigma_G + \Gamma) = \operatorname{Ker} \Sigma_Y^{\mathrm{A}}$, $\operatorname{Ker} \beta^\top (\Sigma_G + \Gamma) \beta = \operatorname{Ker} \beta^\top \Sigma_Y^{\mathrm{A}} \beta$. By assumption, $\operatorname{Ker} \Sigma_H = \operatorname{Ker} \Sigma_Z^{\mathrm{A}}$. Hence $\operatorname{Ker} \Sigma_M = \operatorname{Ker} \Sigma_{\mathrm{A}}$.

It remains to prove that $\operatorname{Row} \Sigma_{\mathrm{A}} \cap \operatorname{Row} \Sigma_{\mathrm{R}} = \{0\}$. Let $v \in \operatorname{Row} \beta^\top \Sigma_Y^{\mathrm{A}} \beta \cap \operatorname{Row} \beta^\top \Sigma_Y^{\mathrm{R}} \beta$. Then there are a $w_1 \in \boldsymbol{R}^N$ and a $w_2 \in \boldsymbol{R}^N$ such that $w_1^\top \beta^\top \Sigma_Y^{\mathrm{A}} \beta = v = w_2^\top \beta^\top \Sigma_Y^{\mathrm{R}} \beta$. Since $\operatorname{rank} \beta = L$, $w_1^\top \beta^\top \Sigma_Y^{\mathrm{A}} = w_2^\top \beta^\top \Sigma_Y^{\mathrm{R}}$, and this vector belongs to $\operatorname{Row} \Sigma_Y^{\mathrm{A}} \cap \operatorname{Row} \Sigma_Y^{\mathrm{R}}$, which coincides with $\{0\}$ by assumption. Hence $v = 0$. Thus $\operatorname{Row} \beta^\top \Sigma_Y^{\mathrm{A}} \beta \cap \operatorname{Row} \beta^\top \Sigma_Y^{\mathrm{R}} \beta = \{0\}$. By assumption, $\operatorname{Row} \Sigma_Z^{\mathrm{A}} \cap \operatorname{Row} \Sigma_Z^{\mathrm{R}} = \{0\}$. Since $\operatorname{Row} \beta^\top \Sigma_Y^i \beta \subseteq \operatorname{Row} \beta$ and $\operatorname{Row} \Sigma_Z^i \subseteq \operatorname{Row} \Sigma_Z$ for each $i = 1, 2$ and since $\operatorname{Row} \beta \cap \operatorname{Row} \Sigma_Z = \{0\}$ by assumption,

$$
\operatorname{Row} \Sigma_{\mathrm{A}} \cap \operatorname{Row} \Sigma_{\mathrm{R}} = \operatorname{Row} \left(\beta^\top \Sigma_Y^{\mathrm{A}} \beta + \Sigma_Z^{\mathrm{A}} \right) \cap \operatorname{Row} \left(\beta^\top \Sigma_Y^{\mathrm{R}} \beta + \Sigma_Z^{\mathrm{R}} \right) = \{0\}.
$$

///

Proof of Proposition 8 Since $\Sigma_G = 0$ or $\Sigma_H = \Sigma_F = 0$, $\Gamma = 0$ in Lemma 3 in either of the two extreme cases. As noted right after Definition 1, in the first case, $(\Sigma_Y^{\mathrm{A}}, \Sigma_Y^{\mathrm{R}}) = (0, \Sigma_Y)$ is the unique $(\operatorname{Ker}\Sigma_G)$-based decomposition of Σ_Y and $(\Sigma_Z^{\mathrm{A}}, \Sigma_Z^{\mathrm{R}}) = (\Sigma_Z, 0)$ is the unique $(\operatorname{Ker}\Sigma_H)$-based decomposition of Σ_Z. By Proposition 7, $(\Sigma_{\mathrm{A}}, \Sigma_{\mathrm{R}}) = (\Sigma_Z, \beta^\top \Sigma_Y \beta)$. Define

$$w_{\mathrm{A}} = \begin{pmatrix} 0 \\ \hdashline \hat{\mu}_F \end{pmatrix}, \qquad w_{\mathrm{R}} = \begin{pmatrix} \mu_G - R\mathbf{1}_L \\ \hdashline \hat{\beta}^\top(\mu_G - R\mathbf{1}_L) \end{pmatrix}.$$

Then $w_{\mathrm{A}} + w_{\mathrm{R}} = \mu_M - R\mathbf{1}_N$. It remains to prove that $\Sigma_{\mathrm{R}} v_{\mathrm{R}} = w_{\mathrm{R}}$, and $\Sigma_{\mathrm{A}} v_{\mathrm{A}} = w_{\mathrm{A}}$. Indeed,

$$\Sigma_{\mathrm{A}} v_{\mathrm{A}} = \left(\begin{array}{c:c} 0 & 0 \\ \hdashline 0 & \hat{\Sigma}_U \end{array}\right) \begin{pmatrix} -\hat{\beta}\hat{\Sigma}_U^{-1}\hat{\mu}_F \\ \hdashline \hat{\Sigma}_U^{-1}\hat{\mu}_F \end{pmatrix} = w_{\mathrm{A}},$$

$$\Sigma_{\mathrm{R}} v_{\mathrm{R}} = \left(\begin{array}{c:c} \Sigma_Y & \Sigma_Y\hat{\beta} \\ \hdashline \hat{\beta}^\top\Sigma_Y & \hat{\beta}^\top\Sigma_Y\hat{\beta} \end{array}\right) \begin{pmatrix} \Sigma_Y^{-1}(\mu_G - R\mathbf{1}_L) \\ \hdashline 0 \end{pmatrix} = w_{\mathrm{R}}.$$

Then the optimal portfolio $\zeta(0)$ coincides with the sum of v_{A} and v_{R}:

$$\zeta(0) = \begin{pmatrix} \Sigma_Y^{-1}(\mu_G - R\mathbf{1}_L) - \hat{\beta}\hat{\Sigma}_U^{-1}\hat{\mu}_F \\ \hdashline \hat{\Sigma}_U^{-1}\hat{\mu}_F \end{pmatrix} = w_{\mathrm{A}} + w_{\mathrm{R}}.$$

The other extreme case can analogously be proven. ///

References

[1] Avramov, D., and G. Zhou. 2010. Bayesian Portfolio Analysis. *Annual Reviews of Financial Economics* 1:25–47.

[2] Barry, C. B. 1974. Portfolio Analysis Under Uncertain Means, Variances, and Covariances. *The Journal of Finance* 29:515–522.

[3] Beetsma, R. M. W. J., and P. C. Schotman. 2001. Measuring Risk Attitudes in a Natural Experiment: Data from the Television Game Show Lingo. *The Economic Journal* 111:821–848.

[4] Bosch, A. J. 1987. Note on the Factorization of a Square Matrix into Two Hermitian or Symmetric Matrices. *SIAM Review* 29:463–468.

[5] Campbell, J. Y., A. Lo, and C. MacKinlay. 1997. *The Econometrics of Financial Markets*. Princeton University Press.

[6] DeMiguel, V., L. Garlappi, and R. Uppal. 2009. Optimal Versus Naive Diversification: How Inefficient is the $1/N$ Portfolio Strategy? *The Review of Financial Studies* 5:1915–1953.

[7] Frost, P. A., and J. E. Savarino. 1986. An Empirical Bayes Approach to Efficient Portfolio Selection. *Journal of Financial and Quantitative Analysis* 3:293–305.

[8] Garlappi, L., R. Uppal, and T. Wang. 2007. Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach. *Review of Financial Studies* 20:41–81.

[9] Gilboa, I., and D. Schmeidler. 1989. Maxmin Expected Utility with Non-Unique Prior. *Journal of Mathematical Economics* 18:141–153.

[10] Harvey, C. R., Y. Liu, and H. Zhu. 2014. ... and the Cross-Section of Expected Returns, Working Paper.

[11] Huang, C. F., and R. H. Litzenberger. 1988. *Foundations for Financial eEonomics*. Prentice-Hall, New Jersey.

[12] Jorion, P. 1986. Bayes-Stein Estimation for Portfolio Analysis. *Journal of Financial and Quantitative Analysis* 21:279–292.

[13] Ju, N., and J. Miao. 2012. Ambiguity, Learning, and Asset Returns. *Econometrica* 80:559–591.

[14] Kan, R., and G. Zhou. 2007. Optimal Portfolio Choice with Parameter Uncertainty. *Journal of Financial and Quantitative Analysis* 42:621–656.

[15] Kandel, S., and R. F. Stambaugh. 1996. On the Predictability of Stock Returns: An Asset-Allocation Perspective. *Journal of Finance* 51:385-424.

[16] Klein, R.W., and V.S. Bawa. 1976. The Effect of Estimation Risk on Optimal Portfolio Choice. *Journal of Financial Economics* 3: 215–231.

[17] Klibanoff, P., M. Marinacci, and S. Mukerji. 2005. A Smooth Model of Decision Making under Ambiguity. *Econometrica* 73:1849–1892.

[18] Ledoit, O., and M. Wolf. 2003. Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection. *Journal of Empirical Finance* 10:603–621.

[19] Lutgens, F., and P. C. Schotman. 2010. Robust Portfolio Optimisation with Multiple Experts. *Review of Finance* 14:343-383.

[20] Maccheroni, F., M. Marinacci, and D. Ruffino. 2013. Alpha as Ambiguity: Robust Mean-Variance Portfolio Analysis. *Econometrica* 81:1075–1113.

[21] Pástor, L. 2000. Portfolio Selection and Asset Pricing Models. *Journal of Finance* 55:179–223.

[22] Pástor, L., and R. F. Stambaugh. 2000. Comparing Asset Pricing Models: An Investment Perspective. *Journal of Financial Economics* 56:335–381.

[23] Ross, S. A. 1976. The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory* 13:341-360.

[24] Ruffino, D. 2014. A Robust Capital Asset Pricing Model. Finance and Economics Discussion Paper Series, Divisions of Research and Statistics and Monetary Affairs, Federal Reserve Board.

[25] Savage, L. J. 1954. *The Foundations of Statistics*, Wiley.

[26] Wakai, K. 2013. Equilibrium Alpha in Asset Pricing under Ambiguity Averse Economy. manuscript. Graduate School of Economics, Kyoto University.

[27] Wang, Z. 2005. A Shrinkage Approach to Model Uncertainty and Asset Allocation. *Review of Financial Studies* 18:673–705.

[28] Wilson, R. 1968. The Theory of Syndicates. *Econometrica* 36:119–132.

[29] Zellner, A., and V. K. Chetty. 1965. Prediction and Decision Problems in Regression Models from the Bayesian Point of View. *Journal of the American Statistical Association* 60:608–616.

Table 1: Average Monthly Returns, Standard Deviations, and Sharpe Ratio of FF 6 Portfolios: July 1926-December 2013.

	Mean	Stand Dev	Sharpe Ratio
R	0.29	0.25	
R_{m}	0.94	5.40	0.12
SL	0.99	7.69	0.09
SN	1.29	7.04	0.14
SH	1.49	8.24	0.15
BL	0.90	5.36	0.12
BN	0.98	5.75	0.12
BH	1.19	7.17	0.13

Table 1 reports the average, the standard deviation, and the Sharpe ratio of the risk-free rate (R), the market portfolio (R_m), and the FF6 portfolio returns. "S" means small ME and "B" means big ME. "L" means low B/M, "N" means neutral, and "H" means high B/M.

Table 2: Sample Covariance Matrix of the FF6 Portfolios: July 1926-December 2013.

	SL	SN	SH	BL	BN	BH
SL	59.07	51.57	57.07	35.18	36.44	44.99
SN	51.57	49.60	55.87	31.89	35.77	45.27
SH	57.07	55.87	68.04	34.73	41.35	54.40
BL	35.18	31.89	34.73	28.66	27.48	31.72
BN	36.44	35.77	41.35	27.48	33.00	38.54
BH	44.99	45.27	54.40	31.72	38.54	51.49

Table 2 reports the covariance matrix of the FF 6 portfolio returns.

Table 3: Optimal Portfolios $\zeta(\eta)$ in the First Example.

	$\zeta(0)$	$\zeta(1)$	$\zeta(10)$	$\zeta(100)$	$\zeta(1000)$
R	97.34	97.39	97.72	99.00	99.85
SL	-8.57	-8.43	-7.35	-3.21	-0.49
SN	8.77	8.62	7.51	3.29	0.50
SH	3.06	3.01	2.62	1.15	0.17
BL	4.40	4.33	3.77	1.65	0.25
BN	-2.38	-2.34	-2.04	-0.89	-0.13
BH	-2.62	-2.58	-2.25	-0.98	-0.15

Table 3 reports the optimal portfolio weight in percent of $\zeta(\eta)$ for $\eta = 0, 1, 10, 100, 1000$ for the first case. The total wealth is assumed to be $W = 1$. The diagonal element of the matrix Σ_M is $(1/60)\Sigma_X$. "R" means the fraction (%) of wealth invested into the risk-free asset. The remaining "SL", "SN", "SH", "BL", "BN", and "BH" are the ratio of wealth invested into FF6 risky assets.

Table 4: Optimal Portfolios $\zeta(\eta)$ in the Second Example.

	$\zeta(0)$	$\zeta(1)$	$\zeta(10)$	$\zeta(100)$	$\zeta(1000)$
R	97.34	97.58	98.06	98.38	99.33
SL	−8.57	−6.64	−2.33	−0.14	0.08
SN	8.77	6.3	1.97	0.48	0.14
SH	3.06	3.02	1.84	0.61	0.17
BL	4.40	3.33	0.93	0.21	0.08
BN	−2.38	−1.57	−0.19	0.19	0.09
BH	−2.62	−2.03	−0.29	0.28	0.12

Table 4 reports the optimal portfolio weight in percent of $\zeta(\eta)$ for $\eta = 0, 1, 10, 100, 1000$ for the first case. The total wealth is assumed to be $W = 1$. The diagonal element of the matrix Σ_M is $\sigma^2 = 0.5$ and the off-diagonal element is $\kappa = 0$. "R" means the fraction (%) of wealth invested into the risk-free asset. The remaining "SL", "SN", "SH", "BL", "BN", and "BH" are the ratio of wealth invested into FF6 risky assets.

Table 5: Optimal Portfolios $\zeta(\eta)$ in the Third Example.

	$\zeta(0)$	$\zeta(1)$	$\zeta(10)$	$\zeta(100)$	$\zeta(1000)$
R	97.34	97.56	98.09	98.64	99.6
SL	−8.57	−6.94	−2.70	−0.26	0.03
SN	8.77	6.66	2.28	0.49	0.10
SH	3.06	3.05	2.03	0.68	0.14
BL	4.40	3.49	1.07	0.12	0.02
BN	−2.38	−1.69	−0.32	0.09	0.03
BH	−2.62	−2.13	−0.44	0.24	0.08

Table 5 reports the optimal portfolio weight in percent of $\zeta(\eta)$ for $\eta = 0, 1, 10, 100, 1000$ for the first case. The total wealth is assumed to be $W = 1$. The diagonal element of the matrix Σ_M is $\sigma^2 = 0.5$ and the off-diagonal element is $\kappa = 0.1$. "R" means the fraction (%) of wealth invested into the risk-free asset. The remaining "SL", "SN", "SH", "BL", "BN", and "BH" are the ratio of wealth invested into FF6 risky assets.

Table 6: Decomposition of the Optimal Portfolio $\zeta(0)$ by Theorem 2 in the Fourth Example.

	$\zeta(0)$	v_{A}	v_{A}
R	97.34	99.35	97.99
SL	-8.57	-8.91	0.33
SN	8.77	8.43	0.33
SH	3.06	2.73	0.33
BL	4.40	4.07	0.33
BN	-2.38	-2.71	0.33
BH	-2.62	-2.96	0.33

Table 6 reports the optimal portfolio $\zeta(0)$ and its decomposition v_{A} and v_{R} by Theorem 2 for the fourth example. The total wealth is assumed to be $W = 1$. The diagonal element of the matrix Σ_M is $\sigma^2 = 0.5$ and the off-diagonal element is $\kappa = -0.1$. "R" means the fraction (%) of wealth invested into the risk-free asset. The remaining wealth is invested into FF6 risky assets with the portfolio weights (%) described by "SL", "SN", "SH", "BL", "BN", "BH".

Table 7: Value-weighted portfolio and the three factors

	VW	PC1	PC2	PC3
SL	2.46	18.59	-387.31	56.37
SN	2.95	17.67	-142.83	-34.24
SH	2.08	20.45	-21.60	-157.27
BL	50.84	12.19	43.37	199.42
BN	31.19	13.75	253.92	77.22
BH	10.48	17.34	354.44	-41.50

Table 7 reports both the value weighted portfolio ("VW") and the factor portfolio corresponding to the three largest principal components ("PC1", "PC2", "PC3"). The value weighted portfolio is the time-series average of capitalization weight (%) of the FF6 portfolios. From the return data of FF6 portfolio, the factor mimicking portfolio weight (%) is constructed using the principal component analysis.

Table 8: Sample Mean μ_G and Covariance Σ_Y of factor portfolio

	μ_G	Σ_Y		
PC1	1.17	45.27	−0.18	0.10
PC2	1.09	−0.18	444.88	−0.25
PC3	−0.18	0.10	−0.25	74.73

Table 8 reports the average monthly return (%) of the factor portfolio. The covariance matrix of the factors are also reported. The former is used for the parameter μ_G and the latter is for Σ_Y in the following examples.

Table 9: Alpha $\hat{\mu}_G$ and beta $\hat{\beta}$ of the FF6 portfolios

	$\hat{\mu}_Y$	$\hat{\beta}^\top$		
		PC1	PC2	PC3
SL	−0.13	1.08	−0.11	0.07
SN	0.11	1.03	−0.04	−0.04
SH	0.06	1.19	−0.01	−0.21
BL	0.10	0.71	0.01	0.26
BN	−0.03	0.80	0.07	0.10
BH	−0.09	1.01	0.10	−0.05

Table 9 reports the estimated value of intercept and coefficients that are found by regressing excess returns of FF6 portfolio on factor portfolio returns (PC1, PC2, and PC3). The former is used as an input parameter $\hat{\mu}_Y$ and the latter is used for $\hat{\beta}$.

Table 10: Sample Covariance Matrix $\tilde{\Sigma}_U$

	$\hat{\Sigma}_U$					
	SL	SN	SH	BL	BN	BH
SL	0.60	−0.41	−0.37	−0.45	0.04	0.49
SN	−0.41	0.91	−0.31	−0.05	0.11	−0.17
SH	−0.37	−0.31	0.75	0.53	−0.00	−0.55
BL	−0.45	−0.05	0.53	0.72	−0.55	−0.17
BN	0.04	0.11	−0.00	−0.55	1.10	−0.64
BH	0.49	−0.17	−0.55	−0.17	−0.64	0.91

Table 10 reports the sample covariance matrix of the regression residuals. The estimated value is used as an input parameter $\hat{\Sigma}_U$.

Table 11: Decomposition in the first case

	$\zeta(0)$	v_{R}	v_{A}
R	95.86	98.49	97.37
PC1	5.87	1.95	3.92
PC2	−0.68	0.18	−0.86
PC3	−0.57	−0.62	0.05
SL	−22.04	0	−22.04
SN	11.63	0	11.63
SH	8.61	0	8.61
BL	13.70	0	13.70
BN	−2.41	0	−2.41
BH	−9.97	0	−9.97

Table 11 reports the optimal portfolio $\zeta(0)$ and its decomposition v_{R} and v_{A}. The total wealth is assumed to be $W = 1$. "R" means the fraction (%) of wealth invested into the risk-free asset. The remaining wealth is invested into factor portfolio and FF6 risky assets with the portfolio weights (%) described by "PC1", "PC2", "PC3", "SL", "SN", "SH", "BL", "BN", "BH".

Table 12: Decomposition in the second case

	$\zeta(0)$	v_{R}	v_{A}
R	95.86	97.37	98.49
PC1	5.87	3.92	1.95
PC2	−0.68	−0.86	0.18
PC3	−0.57	0.05	−0.62
SL	−22.04	−22.04	0
SN	11.63	11.63	0
SH	8.61	8.61	0
BL	13.70	13.70	0
BN	−2.41	−2.41	0
BH	−9.97	−9.97	0

Table 12 reports the optimal portfolio $\zeta(0)$ and its decomposition v_{R} and v_{A}. The total wealth is assumed to be $W = 1$. "R" means the fraction (%) of wealth invested into the risk-free asset. The remaining wealth is invested into factor portfolio and FF6 risky assets with the portfolio weights (%) described by "PC1", "PC2", "PC3", "SL", "SN", "SH", "BL", "BN", "BH".