Beta Ambiguity and Security Return Characteristics

Zhe Geng and Tan Wang *

May 10, 2020

Abstract

We study the cross-sectional properties of asset returns in the presence of ambiguity in asset returns. In our model, the cross-sectional expected returns are described by three factors, capturing risk, mean ambiguity and variance-covariance ambiguity, respectively. The expected returns exhibit cross-sectional characteristics consistent with the empirical fact that the overall beta-return relation and IVOL-return relation are both negative, but the beta-return relation is negative and stronger among over-priced stocks while positive and weaker among under-priced stocks, and the IVOL-return relation is negative and stronger among over-priced stocks but positive and weaker among under-priced stocks.

*Zhe Geng and Tan Wang are with Shanghai Advanced Institute of Finance at Shanghai Jiao Tong University. zGENG.15@saif.sjtu.edu.cn and tanwang@saif.sjtu.edu.cn. We are grateful for the valuable comments from Jun Pan, Jiang Wang, Weidong Tian, the discussant at the 2018 CICF, Jianfeng Yu, the discussant at the 2019 WFA, Yu Yuan, and the conference participants at the 2018 CICF, the 2019 AFA PhD Poster Session, the 2019 WFA.

1 Introduction

Among the numerous empirical stylized facts, the beta anomaly and the idiosyncratic volatility anomaly are perhaps the simplest and yet challenging to understand. Beta anomaly refers to the pattern in the cross-section of stock returns that the security market line is too flat relative to the one predicted by the CAPM theory (Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973)). Idiosyncratic volatility anomaly refers to the negative relation between idiosyncratic volatility and subsequent stock returns (Ang et al. (2006)).

There is a large literature that aims at explaining the two anomalies. Recently, however, Stambaugh, Yu, and Yuan (2015) and Liu, Stambaugh, and Yuan (2018) offer additional evidence that provides a new perspective on the anomalies and raised issues with the existing explanations. Stambaugh, Yu, and Yuan (2015) find that when examined for the subsample of over-priced and under-priced stocks, the relations between mispricing and idiosyncratic volatility have opposite signs, which they argue is a challenge to the existing explanations of the idiosyncratic volatility anomaly. Similarly, Liu, Stambaugh, and Yuan (2018) show that while the security market line for over-priced stocks is flatter than that predicted by the standard CAPM theory, the security market line for under-priced stocks is not flatter if not steeper. They argue that the existing explanations of the beta anomaly would have difficulty in reconciling with this evidence. They argue further that the beta anomaly is in fact closely related to the idiosyncratic volatility anomaly as a result of the positive correlation between beta and idiosyncratic volatility.

In this paper, we provide an explanation of the beta and idiosyncratic volatility anomalies that is consistent with the findings of Stambaugh, Yu, and Yuan (2015) and Liu, Stambaugh, and Yuan (2018). The individuals in our model are rational. The financial markets in our model are frictionless. The key ingredient of our model is that investors do not have perfect knowledge of the probability distribution of stock returns. As argued by Merton (1980), the mean returns are notoriously difficult to estimated precisely. Variances of the returns are in general better estimated. However, the challenge with the covariance matrix is of a different kind. One is that when the number of stocks is large, the determinant of the covariance matrix is very small.[1] A small error in the estimate of the covariance matrix can lead to significant difference in the agent's portfolio choice. Second, because the determinant of the covariance matrix is very small, an approximation of the estimated covariance matrix is used in applications, which also introduces ambiguity in the covariances. As a consequence of the ambiguity in the means and/or the covariances, investors ask for a premium as the compensation for that ambiguity. That premium is the mis-pricing, relative to the CAPM. Under natural assumptions, the premium exhibits a pattern that is consistent with the beta and idiosyncratic volatility anomalies.

In our model, agents are homogeneous. They are fully aware that there is ambiguity about the probability law of stock returns and that the data can only provide an approximation to the true distribution. Due to their aversion to ambiguity, they adjust their portfolios computed according to a reference distribution to account for the ambiguity. The adjustment leads to equilibrium returns that deviate from those computed according to the reference distribution. We show that the deviation can be tracked by two factor portfolios, one for the ambiguity in the expected returns of the stocks and the other for the ambiguity in the covariances of the returns of the stocks. As such, the premia on those two factors (portfolios) are interpreted as the premia for the two sources of ambiguity. It should, however, be emphasized that those two factors are not factors in the traditional sense. They do not track any fundamental market or macro risk variables, such as market return. They capture instead the systematic ambiguity in the asset returns.

When agents are ambiguity averse, the two factors earn positive ambiguity premia. Variation in the loadings of stocks on these two factors lead to variation in the cross section of expected returns. Stocks that have higher loading on those factors earn higher premia, while stocks that have lower or negative loading on those factors earn lower or even negative premia. As ex ante, there is no obvious reason that the reference distribution is related to the level of ambiguity, beta of stocks calculated according to the reference distribution is unlikely to be related to the systematic ambiguity of the stocks. As a consequence, if the

[1]For example, with just 500 stocks, the determinant of the covariance typically exceeds the computer limit and is indistinguishable from zero.

stocks are double-sorted on mis-pricing and beta or idiosyncratic volatility, the alphas can and in fact are likely to exhibit the pattern as shown in Stambaugh, Yu, and Yuan (2015) and Liu, Stambaugh, and Yuan (2018).

Through simulation we show that our model produces qualitatively similar patterns of alphas as shown in the literature. The security market line is flatter than predicted by CAPM. The overall idiosyncratic volatility-return relation is negative. However, the beta-return relation is negative and stronger among over-priced stocks while positive and weaker among under-priced stocks, and the idiosyncratic volatility-return relation is negative and stronger among over-priced stocks, but positive and weaker among under-priced stocks.

Our paper is related to two branches of the literature. One branch is that on ambiguity and its implications for asset prices. To model ambiguity averse agents, we follow the multiple-prior approach of Gilboa and Schmeidler (1989), the dynamic version of which is proposed by Epstein and Schneider (2003). In the study of asset pricing implications of ambiguity, similar approach has been taken by Dow and Werlang (1992), Epstein and Wang (1994, 1995), Chen and Epstein (2002) and Epstein and Miao (2003), Kogan and Wang (2003), Easley and O'Hara (2009, 2010), among many others. An alternative approach to modeling ambiguity averse agents is introduced by Hansen and Sargent (2001) and Anderson, Hansen, and Sargent (2003). That approach is taken by Uppal and Wang (2003), Maenhout (2004, 2006), Liu, Pan, and Wang (2005), among others. The third, smooth ambiguity preference, approach to modeling ambiguity averse agents is introduced by Klibanoff, Marinacci, and Mukerji (2005). Klibanoff, Marinacci, and Mukerji (2009), Hayashi and Miao (2011) provide a dynamic axiomatization of the smooth ambiguity preference. Ju and Miao (2012) propose a generalized recursive smooth ambiguity model which permits a three-way separation among risk aversion, ambiguity aversion, and inter-temporal substitution in a consumption-based asset-pricing model. The innovation of our model is that it allows for ambiguity both in the mean and in the variance-covariance matrix, while most of the existing literature assumes away the ambiguity in the variance-covariance matrix. Epstein and Ji (2013) consider ambiguity in the volatility of one asset. Liu and Zeng (2017) study the effect of correlation ambiguity on portfolio under-diversification. The paper that is closely related to ours is Kogan and Wang (2003). One difference between that paper and ours is our introduction of ambiguity in variance-covariance matrix. The key difference is, however, in our focus on the role of ambiguity for understanding of the beta and the idiosyncratic volatility anomalies.

The second branch is the literature on the beta anomaly and the idiosyncratic volatility anomaly. That literature is large. Blitz, Falkenstein, and van Vliet (2014), Liu, Stambaugh, and Yuan (2018), and Hou and Loh (2016), however, provide excellent summaries of the literature. We will just provide brief summaries of those that are most relevant. One common argument is based on trading constraints. In their explanation of the flat security market line, Black (1972) assumes constraint on riskless borrowing. Frazzini and Pedersen (2014) assumes leverage constraint, Hong and Sraer (2016) and Liu, Stambaugh, and Yuan (2018)) assume short-sale constraint. Another common argument is that investors exhibit particular behavioral preferences. It can be due to the desire to benchmark their portfolios (Baker, Bradley, and Wurgler (2011) and Christoffersen and Simutin (2017)), or preferences for positive skewness (Barberis and Huang (2008), Boyer, Mitton, and Vorkink (2010)), lottery-like payoffs (Bali, Cakici, and Whitelaw (2011)). Other explanation includes those based on earnings surprises (Jiang, Xu, and Yao (2009), Wong (2011)), one-month return reversal (Fu (2009), Huang et al. (2010)), illiquidity (Bali and Cakici (2008)), unpriced information risk (Johnson (2004)), and a missing factor (Chen and Petkova (2012)). As argued by Stambaugh, Yu, and Yuan (2015) and Liu, Stambaugh, and Yuan (2018) all the existing explanations have difficulty in reconciling with their empirical findings.

The remainder of this paper is organized as follows. Section 2 describes our model. Section 3 presents its equilibrium asset pricing implications. Section 4 focuses on the role of ambiguity for understanding beta and idiosyncratic anomalies. Section 5 summarizes the results and concludes.

2 The Model

2.1 The Setting

Similar to that in Kogan and Wang (2003), we consider a frictionless representative agent economy where the agent has constant absolute risk aversion utility with risk aversion parameter $\gamma > 0$,

$$u(x) = -\frac{e^{-\gamma x}}{\gamma},$$

The agent is endowed with an initial wealth W_0, which is, without loss of generality, assumed to be equal to one. Consumption takes place at the end of the period. The agent trades $N+1$ assets, one riskless asset with riskless return r and N risky assets whose returns follow a joint normal distribution. The representative agent knows that the returns are jointly normally distributed. She is, however, ambiguous about the expected return vector μ and variance-covariance matrix Ω. It is this ambiguity that differentiates our setting from that of the CAPM theory. We turn now to the description of the ambiguity and the agent's aversion to it.

2.2 Ambiguity and Ambiguity Averse Preferences

Due to the ambiguity, the agent's preference can not be represented by the standard expected utility. It is instead represented by a max-min utility (Gilboa and Schmeidler (1989)).

$$\min_{Q \in \mathcal{P}} E^Q[u(W)], \tag{1}$$

where $\mathcal{P}$ is a set of probability priors.

For our study, the specification of $\mathcal{P}$ is important. It is a confidence region around a reference probability measure P. Specifically, let P be a reference probability measure, or a reference model, under which the returns of the risky assets follow a joint normal distribution with mean return vector μ and variance-covariance matrix Ω. The density function of the return distribution under P is given by

$$f(R) = (2\pi)^{-N/2} |\Omega|^{-1/2} e^{-\frac{1}{2}(R-\mu)^\top \Omega^{-1}(R-\mu)}.$$

The set of priors, $\mathcal{P}$, takes the form

$$\mathcal{P} = \left\{ Q : v_{J_k}^{\top} \Omega_{J_k}^{-1} v_{J_k} \leq 2\eta_{1,k}, \ \mathrm{tr}(\Omega_{J_k}^{-1} U_{J_k}) - \ln |I_{J_k} + \Omega_{J_k}^{-1} U_{J_k}| \leq 2\eta_{2,k}, \ k = 1, ..., K \ \right\} \qquad (2)$$

In the set, each Q is a probability measure under which the returns of the risky assets are jointly normally distributed with density function given by

$$f_Q(R) = (2\pi)^{-N/2} |\hat{\Omega}|^{-1/2} e^{-\frac{1}{2}(R-\hat{\mu})^{\top} \hat{\Omega}^{-1} (R-\hat{\mu})},$$

where $\hat{\mu}$ and $\hat{\Omega}$ are the mean return vector and variance-covariance matrix, respectively. As a Q in $\mathcal{P}$ is in general different from the reference probability measure P, $v = (\mu - \hat{\mu})$ and $U = (\hat{\Omega} - \Omega)$, give the difference between the mean return vectors and the variance-covariance matrices. In (2), J_k, $k = 1, \ldots, K$, are subsets of $\{1, 2, \ldots, N\}$ and v_{J_k} denotes the sub-vector consisting of those elements of $v = (\mu - \hat{\mu})$ that are in the subset J_k. All the other notations with subscript J_k have similar meaning. I is the identity matrix and $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. In the set $\mathcal{P}$, the probability measure for which $v = 0$ and $U = 0$ is the reference probability measure P. The $\eta_{i,k}$ and $\eta_{2,k}$, $k = 1, \ldots, K$, are parameters for setting the level of confidence of the confidence region.

The motivation of the specific form of $\mathcal{P}$ is the same as in Kogan and Wang (2003) and Uppal and Wang (2003), which will be briefly described shortly. It is essentially a confidence region defined by log likelihood ratio or relative entropy (Anderson, Hansen, and Sargent (2003) and Uppal and Wang (2003)). In Kogan and Wang (2003), as there is no ambiguity about the variance-covariance matrix, $\eta_{2,k} = 0$, for $k = 1, \ldots, K$. The $\mathcal{P}$ in (2) can accommodate ambiguity both in the mean and in variance-covariance matrix.

We now provide the detailed explanation of what set $\mathcal{P}$ captures. We do so with two elaborated examples.

2.2.1 A Single Source of Information

As the true probability law of asset returns is unknown, the parameters of the model, μ and Ω. have to be estimated based on the data available. Suppose that there is only a single data source and the result of the estimation is the reference model P. This is the case where $K = 1$ and $J_1 = \{1, \ldots, N\}$. Since the data set is typically limited, as argued by Merton (1980), the estimated reference model P is unlikely the true model and the ($p\%$) confidence region provides the information on the ambiguity of where the true model is. Let Q be a probability measure that is potentially the true model. As the representative agent knows that the returns follow a joint normal distribution, the return under Q has density given by

$$f_Q(R) = (2\pi)^{-N/2}|\hat{\Omega}|^{-1/2}e^{-\frac{1}{2}(R-\hat{\mu})^\top\hat{\Omega}^{-1}(R-\hat{\mu})},$$

Under this measure, the expected return vector is $\hat{\mu}$ and the variance-covariance matrix is $\hat{\Omega}$. One measure of confidence the econometrician can use is the log likelihood ratio, $E^Q[\ln \xi]$, where $\xi = dQ/dP$ is the density of Q with respect to P. In terms of the reference probability, the likelihood ratio is the relative entropy, $E[\xi \ln(\xi)]$, of Q with respect to P. As argued in Kogan and Wang (2003) and Uppal and Wang (2003), $E[\xi \ln(\xi)]$ is a good approximation of the empirical log-likelihood when the number of observations is large.

It is readily verified that

$$\frac{dQ}{dP} = \xi(R) = \frac{|\Omega|^{\frac{1}{2}}}{|\hat{\Omega}|^{\frac{1}{2}}}e^{-\frac{1}{2}(R-\hat{\mu})^\top\hat{\Omega}^{-1}(R-\hat{\mu})+\frac{1}{2}(R-\mu)^\top\Omega^{-1}(R-\mu)},$$

A bit of algebra shows

$$E[\xi \ln(\xi)] = \frac{1}{2}\left[tr(\Omega^{-1}U) - \ln|I + \Omega^{-1}U| + (\mu - \hat{\mu})^\top\Omega^{-1}(\mu - \hat{\mu})\right] \tag{3}$$

where $U = \hat{\Omega} - \Omega$.

Suppose that Q is the true model and it is in the confidence region specified in (2). Consider first the case where there is no ambiguity about the true variance-covariance matrix. Following (3) the relative entropy in this case, denoted by L_{mean}, is given by

$$L^Q_{\text{mean}} = \frac{1}{2}(\mu - \hat{\mu})^\top\Omega^{-1}(\mu - \hat{\mu})$$

Thus the relative entropy of Q, $L^Q_{\text{mean}} < \eta_{1,1}$.

Suppose next the case where there is no ambiguity about the true mean return vector. The relative entropy in this case, denoted L_{cov}, is given by

$$L^Q_{\text{cov}} = \frac{1}{2}\left[tr(\Omega^{-1}U) - \ln|I + \Omega^{-1}U|\right]$$

This in this case, the relative entropy of Q, $L^Q_{\text{cov}} < \eta_{2,1}$.

Given L^Q_{mean} and L^Q_{cov}, what (2) says is that for Q to be in $\mathcal{P}$, its mean likelihood, measured by L^Q_{mean}, must be less than $\eta_{1,1}$ and its variance-covariance likelihood, measured by L^Q_{cov}, must be less than $\eta_{2,1}$.

2.2.2 Multiple Sources of Information

More realistically, the investors can obtain multiple data sources on the returns and each data source pertains to only a subset of the risky assets. To model multiple sources of information, let $J_k = \{j_1, j_2, ..., j_{N_k}\}$, $k = 1, 2, \ldots, K$, be subsets of $\{1, 2, \ldots, N\}$, and $\cup_k J_k = \{1, 2, ..., N\}$. So overall the agent has some information about each asset. The distribution of asset returns for any source of information J_k is that for $R_{J_k} = (R_{j_1}, R_{j_2}, ..., R_{j_{N_k}})$. We assume the reference probability law implied by the various sources of information coincides with the marginal distributions of the reference model P (denoted as P_{J_k}). The density function of R_{J_k} under the true model Q is

$$f(R_{J_k}) = (2\pi)^{-1}|\hat{\Omega}_{J_k}|^{-1/2}e^{-\frac{1}{2}(R_{J_k}-\hat{\mu}_{J_k})^\top \hat{\Omega}^{-1}_{J_k}(R_{J_k}-\hat{\mu}_{J_k})},$$

which is the marginal distribution of Q (denoted as Q_{J_k}), where $\hat{\mu}_{J_k}$ and $\hat{\Omega}_{J_k}$ are the mean return vector and variance-covariance return matrix of R_{J_k}. Thus, the likelihood ratio of the marginal distribution Q_{J_k} with respect to P_{J_k} is

$$\xi(R_{J_k}) = \frac{|\Omega_{J_k}|^{\frac{1}{2}}}{|\hat{\Omega}_{J_k}|^{\frac{1}{2}}}e^{-\frac{1}{2}(R_{J_k}-\hat{\mu}_{J_k})^\top \hat{\Omega}_{J_k}(R_{J_k}-\hat{\mu}_{J_k})+\frac{1}{2}(R_{J_k}-\mu_{J_k})^\top \Omega_{J_k}(R_{J_k}-\mu_{J_k})},$$

For convenience, we use the same notation $\hat{\Omega}^{-1}_{J_k}$ $(\Omega^{-1}_{J_k})$ to denote the $N \times N$-matrix whose elements in the j_m-th row and j_n-th column, for j_m and j_n in J_k, is the same as the elements in the m-th row and n-th column of the matrix $\hat{\Omega}^{-1}_{J_k}$ $(\Omega^{-1}_{J_k})$, otherwise it is zero. Then the relative entropy is

$$E[\xi_{J_k}\ln(\xi_{J_k})] = \frac{1}{2}\left[tr(\Omega^{-1}_{J_k}U_{J_k}) - \ln|I + \Omega^{-1}_{J_k}U_{J_k}| + (\mu-\hat{\mu})^\top \Omega^{-1}_{J_k}(\mu-\hat{\mu})\right] \tag{4}$$

With expression (4), we see that for a Q to be in the set $\mathcal{P}$, its mean likelihood and variance-covariance likelihood based on information k must be less than $\eta_{1,k}$ and $\eta_{2,k}$, respectively, for all $k = 1, \dots, K$.

3 Portfolio Choice

Because of the presence of ambiguity, the representative agent's portfolio choices will be different from that when there is no ambiguity. The agent will not only consider the trade-off between risk and return, but also the trade-off between those with ambiguity. To understand how the agent trades off ambiguity, risk and return, it is useful to introduce a metric for ambiguity. In the next two subsections, we will introduce our metric for mean return and variance-covariance ambiguity, respectively.

3.1 Measure of Mean Ambiguity

Suppose first that there is no variance-covariance ambiguity. In this case, the relative entropy including mean ambiguity only becomes

$$E[\xi \ln(\xi)] = \frac{1}{2}(\mu - \hat{\mu})^\top \Omega^{-1}(\mu - \hat{\mu}),$$

Let θ denote the portfolio of the risky assets of the agent and $\theta^\top R$ the portfolio return. The metric we use to measure the ambiguity in the mean return of the portfolio is given as

$$\Delta_1(\theta) = \sup_{Q \in \mathcal{P}} \ \theta^\top(\mu - \hat{\mu}), \tag{5}$$

where

$$\mathcal{P}_1 = \{Q : \ E[\xi_{J_k} \ln(\xi_{J_k})] = (\mu - \hat{\mu})^\top \Omega_{J_k}^{-1}(\mu - \hat{\mu}) \leq 2\eta_{1,k}, \ k = 1, 2, ..., K\}.$$

By the metric, the difference between the expected return of the portfolio under the reference model P and the true expected return of the portfolio, $\theta^\top(\mu - \hat{\mu})$, falls into the interval $[-\Delta_1(\theta), \Delta_1(\theta)]$. Thus $\Delta_1(\theta)$ is the maximum possible error in using the reference model P to gauge the true expected return of the portfolio, given the confidence region described by $\mathcal{P}_1$. Clearly, the smaller the $\Delta_1(\theta)$, the less ambiguity there is about the expected return of the portfolio. As shown in Kogan and Wang (2003), diversifiable ambiguity has no impact on $v(\theta)$.[2]

Lemma 1 *Let θ be a portfolio of the risky assets. A solution to (5) exists. If the portfolio θ is such that $\theta_i \neq 0$ for all $i = 1, \ldots N$, then the solution $v(\theta)$ is unique and is given by,*

$$v(\theta) = \Omega_\mu(\theta)\theta, \tag{6}$$

where $\Omega_\mu(\theta)$

$$\Omega_\mu(\theta) = \left(\sum_{k=1}^{K} \lambda_{1,k}(\theta) \Omega_{J_k}^{-1} \right)^{-1}$$

and $\lambda_{1,k}$, $k = 1, \ldots, K$, are Lagrangian multipliers for the K constraints in the definition of $\mathcal{P}_1$. Moreover, Ω_μ is positive definite.

Obviously, $\Delta_1(\theta) = \theta^\top v(\theta)$ depends on the set $\mathcal{P}_1$ and the portfolio θ. The Lagrangian multipliers $\lambda_{1,k}(\theta)$, $k = 1, \ldots, K$, measure how much each source of information contributes to the ambiguity of the portfolio. If $\lambda_{1,k}(\theta) = 0$, for example, the kth source of information does not help to reduce the ambiguity for the portfolio θ.

3.2 The Measure of Variance-Covariance Ambiguity

Now suppose that there is no ambiguity in the mean return vector. In this case,

$$E[\xi \ln \xi] = \frac{1}{2} \left[tr(\Omega^{-1}U) - \ln |I + \Omega^{-1}U| \right]$$

We define the measure of the ambiguity in variance-covariance by

$$\Delta_2(\theta) = \sup_{Q_2 \in \mathcal{P}_2} \theta^\top U \theta, \tag{7}$$

where $U = (\hat{\Omega} - \Omega)$ and

$$\mathcal{P}_2 = \left\{ Q : \frac{1}{2} \left[tr(\Omega_{J_k}^{-1} U_{J_k}) - \ln |I_{J_k} + \Omega_{J_k}^{-1} U_{J_k}| \right] \leq \eta_{2,k}, \; k = 1, 2, ..., K \right\}.$$

[2] For easy reference, the definition of diversifiability is provided in the appendix.

If $\hat{\Omega}$ is the true variance-covariance matrix, then the true variance of the portfolio return is $\theta^\top \hat{\Omega} \theta$. However, under the reference model P, the variance is $\theta^\top \Omega \theta$. Thus, by using the reference model, given the confidence region described by $\mathcal{P}_2$, the maximum error in the variance of the return of the portfolio is given by $\Delta_2(\theta)$.

Lemma 2 *If the portfolio θ is such that $\theta_i \neq 0$ for all $i = 1, \ldots N$, then the solution of (7) exits and is unique.*

3.3 Portfolio Choice of the Agent

Having defined the preference of the investor and the measure of ambiguity, we now turn to the portfolio choice problem of the agent. Using the utility function from (1), the representative agent's utility maximization problem is

$$\sup_{\theta} \min_{Q \in \mathcal{P}} \quad E^Q[-e^{-\gamma W}/\gamma],$$

where the set $\mathcal{P}$ is as given in (2), subject to the agent's wealth constraint $W = W_0[\theta(R - r\mathbf{1}) + 1 + r]$. where $\mathbf{1}$ is the N-vector $(1, 1, \ldots, 1)^\top$.

Proposition 3 *The agent's utility maximization problem has a solution θ given by,*

$$\theta = \gamma^{-1}(\Omega + U(\theta))^{-1}(\mu - r\mathbf{1} - v(\theta)), \tag{8}$$

where $v(\theta)$ and $U(\theta)$ are the solutions of (5) and (7), respectively, given the portfolio θ.

The solution (8) is fairly intuitive. When there is no ambiguity, that is, $v(\theta) = 0$ and $U(\theta) = 0$, (8) reduces to the standard mean-variance optimal portfolio. When there is only ambiguity in the expected returns, (8) reduces to that given in Kogan and Wang (2003). More generally, (8) says that in the presence of ambiguity, the agent behaves as if the true expected return vector of the assets is given by μ under the reference model P adjusted by $v(\theta)$ and the variance-covariance matrix is Ω adjusted by $U(\theta)$. The expected portfolio return is then $\theta^\top \mu - \Delta_1(\theta)$ and the variance of the portfolio return is $\theta^\top \Omega \theta + \Delta_2(\theta)$. That is, the agent behaves as if the expected portfolio return is that under the reference model adjusted downward by $\Delta_1(\theta)$, which is the ambiguity in the mean, and the variance is that under the reference model adjusted upward by $\Delta_2(\theta)$, which is the ambiguity in the variance.

4 Equilibrium Expected Returns

To derive the equilibrium, let θ_m denote the market portfolio of risky assets.[3] In equilibrium, the representative agent holds the market portfolio. It then follows from Proposition 3 that the expected return on the individual stocks and on the market must satisfy

$$\begin{aligned} \mu - r1 &= \gamma\Omega\theta_m + \gamma U(\theta_m)\theta_m + v(\theta_m) \qquad (9) \\ \mu_m - r &= \gamma\theta_m^\top\Omega\theta_m + \gamma\theta_m^\top U(\theta_m)\theta_m + \Delta_1(\theta_m). \qquad (10) \end{aligned}$$

Thus we have the following theorem.

Theorem 4 *The equilibrium vector of expected excess returns is given by*

$$\mu - r1 = \lambda\beta + \lambda_\mu\beta_\mu + \lambda_\Omega\beta_\Omega, \qquad (11)$$

where

$$\begin{aligned} \beta &= \frac{\Omega\theta_m}{\theta_m^\top\Omega\theta_m}, \quad \lambda = \gamma\theta_m^\top\Omega\theta_m = \gamma\sigma_m^2 \\ \beta_\mu &= \frac{\Omega_\mu(\theta_m)\theta_m}{\theta_m^\top\Omega_\mu(\theta_m)\theta_m}, \quad \lambda_\mu = \Delta_1(\theta_m) = \theta_m^\top\Omega_\mu(\theta_m)\theta_m \\ \beta_\Omega &= \frac{U(\theta_m)\theta_m}{\theta_m^\top U(\theta_m)\theta_m}, \quad \lambda_\Omega = \gamma\Delta_2(\theta_m) = \gamma\theta_m^\top U(\theta_m)\theta_m. \end{aligned}$$

where $\Omega_\mu(\theta_m)$ and $U(\theta)$ are solutions of (5) and (7), respectively.

Theorem 4 describes equilibrium asset expected returns in the cross-section. It has rich implications. In particular, equation (11) is what the analysis of beta anomaly and idiosyncratic volatility anomaly in Section 6 will be based on. The three terms on the right hand

[3]For easy comparison with the standard notion of β, we require that the market portfolio weights add up to one. In the following, θ that does not have m as part of its subscript does not have to meet the requirement that its components sum to one.

side of equation (11) have the natural interpretation that $\lambda\beta$ is the risk premium, $\lambda_\Omega\beta_\Omega$ is the variance-covariance ambiguity premium, and $\lambda_\mu\beta_\mu$ is the mean ambiguity premium. Clearly, when there is no ambiguity, the second and third terms on the right hand side of (11) are equal to zero and (11) reduces to the standard CAPM. The β is then the standard CAPM beta. Just as the interpretation for the risk premium where λ is the price of risk and β is the systematic risk, λ_μ and λ_Ω are the prices of ambiguity in the expected return and variance-covariance matrix, and β_μ and β_Ω are the systematic ambiguities in the expected return and variance-covariance matrix, respectively, which we explain now.

While the risk premium is well understood from the standard CAPM theory, what exactly are those ambiguity premia and how are they related to the ambiguity introduced earlier in (5) and (7)? To understand the relation, consider first the case where there is only mean ambiguity. Let θ_μ be the portfolio defined by $\theta_\mu = \Omega^{-1}\Omega_\mu(\theta_m)\theta_m$. The return of the portfolio is $R_\mu = \theta_\mu^\top R$. By Lemma 1, $\nu(\theta_m) = \Omega_\mu(\theta_m)\theta_m = \Omega\theta_\mu$. Next let θ be an arbitrary portfolio. As shown in Kogan and Wang (2003), the total ambiguity of the portfolio θ is $\theta^\top\nu(\theta)$ and its systematic ambiguity is $\theta^\top\nu(\theta_m)$. Using the portfolio θ_μ, the systematic mean ambiguity of the portfolio θ is $\theta^\top\Omega\theta_\mu$, which is the covariance between the return of the portfolio θ and that of θ_μ. According to Theorem 4, the mean ambiguity beta of the portfolio θ is

$$\beta_\mu(\theta) = \frac{\theta^\top\Omega_\mu(\theta_m)\theta_m}{\theta_m^\top\Omega_\mu(\theta_m)\theta_m} = \frac{\theta^\top\Omega\theta_\mu}{\theta_m^\top\Omega_\mu(\theta_m)\theta_m} = \frac{\mathrm{cov}(R_\theta, R_\mu)}{\theta_m^\top\Omega_\mu(\theta_m)\theta_m}$$

Therefore, the mean ambiguity beta of the portfolio θ, $\beta_\mu(\theta)$, is zero if and only if the systematic mean ambiguity of the portfolio θ is zero. In other words, a portfolio earns mean ambiguity premium if and only if its systematic mean ambiguity is non-zero, and that systematic ambiguity is captured by the covariance between the return of the portfolio θ and that of θ_μ. Because of the relationship between the systematic mean ambiguity of θ and the covariance between R_θ and R_μ, θ_μ is a factor portfolio for the ambiguity of the expected returns. Any asset or portfolio that has non-zero loading on the factor will earn a (mean) ambiguity premium.

Similarly, the total variance-covariance ambiguity of the portfolio θ is $\theta^\top U(\theta)\theta$ and its systematic ambiguity is $\theta^\top U(\theta_m)\theta_m$.[4] Let θ_Ω be the portfolio defined by $\theta_\Omega = \Omega^{-1}U(\theta_m)\theta_m$ and $R_\Omega = \theta_\Omega^\top R$ be its return. According to Theorem 4, the variance-covariance ambiguity beta of the portfolio θ is

$$\beta_\Omega(\theta) = \frac{\theta^\top U(\theta_m)\theta_m}{\theta_m^\top U(\theta_m)\theta_m} = \frac{\theta^\top \Omega\theta_\Omega}{\theta_m^\top U(\theta_m)\theta_m} = \frac{\mathrm{cov}(R_\theta, R_\Omega)}{\theta_m^\top U(\theta_m)\theta_m}$$

That is, the variance-covariance ambiguity beta of the portfolio θ, $\beta_\Omega(\theta)$, is zero if and only if the systematic variance-covariance ambiguity of the portfolio θ is zero. The portfolio earns variance-covariance ambiguity premium if and only if its systematic variance-covariance ambiguity, captured by $\mathrm{cov}(R_\theta, R_\Omega)$, is non-zero. The portfolio θ_Ω is a factor portfolio for the variance-covariance ambiguity.

As a simple example to illustrate the contrast between the standard CAPM and Theorem 4, lets consider a market neutral strategy. When there is no ambiguity, a zero-beta portfolio θ that neutralizes the standard market risk ($\theta^\top\beta = 0$) delivers the market neutral returns. When there is ambiguity, however, the return on that portfolio may no longer be market neutral. The three factor structure described in Theorem 4 suggests that a portfolio θ that also neutralizes ambiguity, that is, the portfolio such that $\theta^\top\beta = 0$, $\theta^\top\beta_\mu$ and $\theta^\top\beta_\Omega = 0$, is more likely to be market neutral.

Theorem 4 provides a three-factor structure for the expected returns of the asset. While we call R_μ and R_Ω factors, it should be emphasized that they are not factors in the traditional sense. That is, they do not track some fundamental market or macro risk variables. They arise not because of the presence of these risk variables. Rather it is because the agent lacks precise knowledge of the true distribution of the asset returns and that ambiguity cannot be diversified away. However, that is not to say there will not be economic variables that are highly correlated with R_μ and R_Ω. In that case, it must be that those economic variables track the systematic ambiguity of asset returns.

Before turning to the empirical implications of Theorem 4, we comment on whether the predictions of Theorem 4 are empirically distinguishable from those of the CAPM theory.

[4] In general, U may not be positive definite. However, as shown later, in the case of multiple non-overlapping sources of information, U is positive definite.

We provide two examples to elaborate on that.

One Source of Information

When there is only one source of information ($K = 1$), it can be shown that

$$\mu - r\mathbf{1} \;=\; (\gamma + \gamma\delta_1 + \delta_2)\Omega\theta_m$$

where $\delta_1 > 0$ and $\delta_2 > 0$ are two positive numbers. Thus it is as if the representative agent lives in a world with risk only and she has a higher level of risk aversion. The standard CAPM holds. This is reminiscient of the result in Anderson, Hansen, and Sargent (2003). This example shows that the presence of ambiguity does not necessarily leads to violation of CAPM. In this case, the standard zero-beta portfolio will neutralize with the confidence determined by $\mathcal{P}$, the uncertainty from both risk and ambiguity.

Multiple Non-overlapping Sources of Information

Another interesting case is one of non-overlapping sources of information. Suppose that there are K sources of information and they are non-overlapping in the sense that each source of information is about a subset of the N assets and the subsets do not overlap. In this case we can divide N assets into K non-overlapping groups and solve (5) and (7) to get explicit expressions for $\Delta_1(\theta)$ and $\Delta_2(\theta)$.

Lemma 5 *Let θ be a portfolio and θ_{J_k} be the sub-vector of θ containing components for assets in group k for $k = 1, \ldots, K$. If the K sources of information are non-overlapping, then the solutions of (5) and (7) are given by,*

$$v(\theta) = \begin{bmatrix} \frac{\sqrt{2\eta_{1,1}}}{\sigma_{J_1}}\Omega_{J_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\sqrt{2\eta_{1,K}}}{\sigma_{J_K}}\Omega_{J_K} \end{bmatrix} \begin{bmatrix} \theta_{J_1} \\ \vdots \\ \theta_{J_K} \end{bmatrix} \tag{12}$$

and

$$U(\theta) \quad = \quad \begin{bmatrix} \frac{f(\eta_{2,1})}{\sigma^2_{J_1}}\Omega_{J_1}\theta_{J_1}\theta_{J_1}^\top\Omega_{J_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{f(\eta_{2,K})}{\sigma^2_{J_K}}\Omega_{J_K}\theta_{J_K}\theta_{J_K}^\top\Omega_{J_K} \end{bmatrix} \tag{13}$$

where $\sigma^2_{J_k}(\theta_m) = \theta_{J_k}^\top \Omega_{J_k}\theta_{J_k}$, $f(\eta)$ *is implicitly defined by*

$$2\eta = f(\eta) - \ln\left(1 + f(\eta)\right),$$

and $\lambda_{1,k}$, $\lambda_{2,k}(\theta)$ *are given by*

$$\lambda_{1,k} = \sqrt{\frac{\sigma^2_{J_k}}{2\eta_{1,k}}}, \quad \lambda_{2,k}(\theta) = 2\theta_{J_k}^\top\Omega_{J_k}\theta_{J_k}\left(1 + \frac{1}{f(\eta_{2,k})}\right). \tag{14}$$

Moreover, U *is positive definite.*

Given the explicit solutions, it follows from Theorem 4 that, for the asset j in group k, the mean ambiguity beta is, for $j \in J_k$,

$$\beta_{\mu,j} = \frac{1}{\Delta_1(\theta_m)}v_j(\theta_m) = \frac{\sqrt{2\eta_{1,k}}\sigma_{J_k}}{\sum_{k=1}^K \sqrt{2\eta_{1,k}}\sigma_{J_k}}\beta_{J_k,j},$$

where $\beta_{J_k,j} = \mathrm{cov}(r_j, \theta_{J_k}^\top R_{J_k})/\sigma^2(\theta_{J_k})$. Interestingly, the mean ambiguity beta of the market portfolio is the risk beta of portfolio θ_{J_k} scaled down by a weight, with the weight being determined by the ambiguity.

For the variance-covariance ambiguity beta,

$$\beta_{\Omega,j} = \frac{[U\theta_m]_j}{\theta_m^\top U\theta_m} = \frac{\theta_{J_k}^\top U_{J_k}(\theta_m)\theta_{J_k}}{\Delta_2(\theta_m)}\frac{[U_{J_k}(\theta_m)\theta_{J_k}]_j}{\theta_{J_k}^\top U_{J_k}(\theta_m)\theta_{J_k}} = \frac{f(\eta_{2,k})\sigma^2_{J_k}}{\sum_{k=1}^K f(\eta_{2,k})\sigma^2_{J_k}}\beta_{J_k,j}.$$

Putting things together, we have the following corollary,

Corollary 6 *If the* K *sources of information are non-overlapping, then the expected return on the individual asset* j *in the group* k *is given by*

$$\mu_j - r = \gamma\sigma_m^2\beta_j + \left(\sqrt{2\eta_{1,k}}\sigma_{J_k} + \gamma f(\eta_{2,k})\sigma^2_{J_k}\right)\beta_{J_k,j}, \tag{15}$$

Corollary 6 shows that when there are more than one sources of information on the probability distribution of the returns, the equilibrium expected returns in our model differs from those in the CAPM theory.

5 Equilibrium Asset Prices

In this section, to prepare for the analysis in Section 6, we rewrite the equilibrium returns in Theorem 4 in terms of exogenous dividends and the equilibrium asset prices. Suppose that the vector of exogenous dividends D follows a joint normal distribution. The reference distribution is one with mean vector d and variance-covariance matrix Σ. Let P denote the equilibrium price vector. Let θ_m denote the market portfolio in terms of portfolio weights and $\bar{\theta}_m$ denotes the market portfolio in terms of shares. Then

$$R_j = \frac{D_j}{P_j} - 1, \quad \mu_j = \frac{d_j}{P_j} - 1, \quad \Omega = \mathrm{diag}(1/P)\,\Sigma\,\mathrm{diag}(1/P),$$
$$R_m = \frac{\bar{\theta}_m^\top D}{\bar{\theta}_m^\top P} - 1, \quad \theta_m = \mathrm{diag}(P)\bar{\theta}_m, \quad (\Omega\theta_m) = \mathrm{diag}(1/P)\Sigma\bar{\theta}_m.$$

where $\mathrm{diag}(x)$ is the diagonal matrix whose diagonal elements are given by the elements of vector x. Note that $\theta_m^\top \mathbf{1}$ is not necessarily equal to one as the riskless rate r is exogenously given.

When there is no ambiguity, the equilibrium price vector is given by,

$$P = \frac{1}{1+r}(d - \gamma\Sigma\bar{\theta}_m),$$

and the beta is given by

$$\beta = \frac{1}{\bar{\theta}_m^\top \Sigma \bar{\theta}_m}\mathrm{diag}(1/P)\Sigma\bar{\theta}_m,$$

The expected excess return of individual asset and market portfolio are respectively,

$$\mu - r\mathbf{1} = \gamma\mathrm{diag}(1/P)\Sigma\bar{\theta}_m, \quad \mu_m - r = \gamma\bar{\theta}_m^\top\Sigma\bar{\theta}_m,$$

The CAPM holds,

$$\mu_j - r = \frac{1}{P_j}\frac{(\Sigma\bar{\theta}_m)_j}{\bar{\theta}_m^\top\Sigma\bar{\theta}_m}(\mu_m - r).$$

When there is mean ambiguity and variance-covariance ambiguity under independent source of information, Corollary 6 in Section 4 shows that the equilibrium price for the asset

j in group k is

$$P_j = \frac{1}{1+r}\left(d_j - \gamma(\Sigma\bar{\theta}_m)_j - \left(\sqrt{\frac{2\eta_{1,k}}{\bar{\theta}_{J_k}^\top \Sigma_{J_k}\bar{\theta}_{J_k}}} + \gamma f(\eta_{2,k})\right)(\Sigma_{J_k}\bar{\theta}_{J_k})_j\right). \tag{16}$$

It is worth noting that if the means and variances, d and Σ, of dividends, as well as $\eta_{i,k}$, are all scaled up by a factor C, then the prices P are scaled up by C as well. This linear homogeneity is useful in the simulation analysis later.

Now we turn to the understanding of the beta and idiosyncratic volatility anomalies.

6 Understanding Anomalies

As discussed in the introduction, the literature has provided several possible explanations of the beta and idiosyncratic volatility anomalies. In this section, we show that the theory developed in the preceding sections can provide an alternative explanation of the beta and IVOL anomalies. While a serious empirical evaluation is beyond the scope of this paper, the simulation exercise provided highlights the economic mechanism that underlies the explanation.

6.1 Over-Pricing and Under-Pricing

An analysis of anomaly typically starts with the mis-pricing of assets according to a benchmark asset pricing theory.[5] The setting of our model is that of CAPM, except that the representative agent has max-min utility instead of the expected utility. Thus the benchmark theory for over-pricing and under-pricing is CAPM. That is,

$$\mu_j - r = \alpha_j + (\mu_m - r)\beta_j,$$

[5] It should be noted that there is not a universal benchmark theory. The benchmark theory used typically depends on the particular empirical anomaly being evaluated and the particular study. The benchmark we provide is based on the mean-variance framework we used to develop our theory.

and a non-zero α_j implies mis-pricing. Asset j is under-priced if $\alpha_j > 0$. It is over-priced if $\alpha_j < 0$. It then follows from Theorem 4 that

$$\alpha_j = [\lambda - (\mu_m - r)]\beta_j + \lambda_\mu \beta_{\mu,j} + \lambda_\Omega \beta_{\Omega,j}.$$

Since $\mu_m - r = \lambda + \lambda_\mu + \lambda_\Omega$, rewriting yields

$$\alpha_j = \left(\lambda_\mu \left[\frac{\beta_{\mu,j}}{\beta_j} - 1 \right] + \lambda_\Omega \left[\frac{\beta_{\Omega,j}}{\beta_j} - 1 \right] \right) \beta_j. \tag{17}$$

What this shows is that mispricing is driven by the interaction of the risk and ambiguity of the asset. Equation (17) is the basis on which we provide our analysis of the beta and idiosyncratic volatility anomalies. We will argue below in more detail how the interaction of risk and ambiguity of assets give rise to the beta and idiosyncratic volatility anomalies.

6.2 Beta Anomaly

In the classical CAPM of Sharpe (1964) and Lintner (1965) theory, stocks with higher betas should earn higher premia than stocks with lower betas. However, the empirical evidence shows that high-beta stocks earn too little compared to low-beta stocks (Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973)). As noted in the introduction, there are several explanations in the literature. Here in this section, based on the theory developed earlier, we provide an alternative explanation of the beta anomaly.

We will present first the results from a simulation. We focus on the special case of (17) where there are non-overlapping sources of information about the mean of the liquidating dividends. We assume there is no ambiguity about the variance-covariance matrix and simulate the model as follows.[6]

[6]In the simulation, the ambiguity is assumed to be about the distributions of dividends, as opposed to the distributions of asset returns. The set $\mathcal{P}$ in (2) can be readily converted into a set in terms of the probability distributions of dividends. In the case of mean ambiguity only, as

$$(\mu - \hat{\mu})\Omega^{-1}(\mu - \hat{\mu}) = (d - \hat{d})\text{Diag}(1/P)\left[\text{Diag}(1/P)\Sigma\text{Diag}(1/P)\right]^{-1}\text{Diag}(1/P)(d - \hat{d}) = (d - \hat{d})\Sigma^{-1}(d - \hat{d})$$

the set of probability measures is the same as $\mathcal{P}$ in (2).

1. Set the number of stocks n to be 1000. We make 1000 draws from the normal distribution $N(200, 5)$ as the mean vector d of the 1000 liquidating dividends.[7] We use US stocks monthly price and return data to estimate the monthly variance-covariance matrix of the liquidating dividends Σ as follows. We randomly choose 1000 stocks (we require that each stock should have over 20 years' monthly data) and calculate the correlation matrix. We then draw 1000 times from $N(0.45, 0.08)$ and take the square of the absolute values of the 1000 draws as the elements of the diagonal of Σ.[8] The supply of each asset equals to 1. The risk aversion coefficient is 2. The risk-free rate is set to be $r = 3\%$, annualized.

2. Assume that there are non-overlapping sources of information about mean ambiguity of the liquidating dividends. We divide the 1000 stocks into two groups of 500 each. Draw 600 times from the joint dividends distribution $N(d, \Sigma)$ and take those samples as realized dividends for the assets (dividend data for 50 years). Calculate the equilibrium return based on the simulated dividends, $r_{j,t} = D_{j,t}/P_j - 1$. The mean ambiguity confidence level of the first group and the second group are $\eta_1 = 200$ and $\eta_2 = 250$ respectively.[9]

3. The econometrician uses those realized returns to run regressions to estimate CAPM beta and to calculate the variance of the residuals as Ivol for each asset. Calculate the

[7] The particular choice of the mean of this distribution is not very important, as the price is proportional to the mean dividend when d, Σ and η_1 and η_2 are properly scaled. The standard deviation is to ensure the mean returns have some variation. The small ratio of 5/200 is only to ensure that the prices are all positive.

[8] This is a convenient way of generating the a 1000×1000 variance-covariance matrix of dividends whose correlation matrix mimics that of the 1000 stocks chosen. The absolute values of the parameters of the distribution, 0.45 and 0.08, are not important as, again, price is proportional to the mean dividend when d, Σ and η_1 and η_2 are properly scaled. So the ratio $0.08/\sqrt{0.45}$ is smaller than that for the 1000 stocks used for calculating the correlation matrix and is chosen to ensure positive prices. We have experimented with a variety of the ratios, the qualitative results are robust to the variation.

[9] If the mean of the reference model P is estimated using sample average, $\hat{\mu}$, then for the true model $Q \sim N(\mu, \Omega)$, $(\hat{\mu} - \mu)[\Omega/600]^{-1}(\hat{\mu} - \mu) = 600(\hat{\mu} - \mu)\Omega^{-1}(\hat{\mu} - \mu)$ follows χ^2 distribution with 500 degrees of freedom. The 5% critical value for that distribution is over 500. Since the parameters η_1 and η_2 represent both the presence of ambiguity and the agent's ambiguity aversion. By choosing η_1 and η_2 that is significantly below the 5% critical value, we are assuming that the representative agent is not very ambiguity averse. We have experimented with the levels of η_1 and η_2. The qualitative results are robust.

average of excess returns of each asset as the true return and the average of market excess returns as the true market excess return. Then define the alpha as the difference between the true return and the product term of CAPM betas multiplying average market excess return. We also use alphas to proxy for mispricing.

4. We double-sort the stocks by mis-pricing (α) and beta into 5 quintiles each and obtain 5×5 cells. For each cell, we compute the average of the αs of the stocks in that cell. We also compute the t-statistics of the average. We restrict to stocks with positive betas only, which is on average about 90% of the stocks, for comparison with Liu, Stambaugh, and Yuan (2018).

The result of the simulation is reported in the Table 1. Panel A reports the averages of the αs and Panel B reports the t-statistics of the averages. In the middle of Panel A are the 5×5 cells of double-sort. The last row reports the average of αs of all stocks sorted by beta. The second last row are the differences in α between the most over-priced and the most under-priced stocks. The last column of Panel A shows the differences, $H - L$, between the average αs of the stocks with the highest beta and that with the lowest beta.

The reported result is consistent with the existing literature. First, the last row of Panel A shows that there is a negative relation between α and beta, which is the beta anomaly reported in Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973), among others. Next, differentiating between over-priced and under-priced stocks, the first row of Panel A shows that among over-priced stocks, there is a negative relationship between α and beta, while the fifth row shows that among under-priced stocks, there is a positive relationship between α and beta. That is, if over-priced and under-priced stocks are differentiated, there is beta anomaly in the over-priced stocks and there is no beta anomaly in the under-priced stocks. If anything, the relation between αs and beta for the under-priced stocks is more likely to be positive, opposite to the sign in the beta anomaly. The middle rows, which are for stocks that are not obviously mispriced, α and beta exhibit a flat relation. Third, Panel B of Table 1 shows that the negative relations between α and beta for all stocks and for over-priced stocks, measured by H-L in the last column in Panel A, are statistically significant,

Table 1: Alphas for Portfolios Sorted on Beta and Mispricing

The table reports the alpha for portfolios formed by an independent 5 × 5 sort on Beta and Mispricing.

Mispricing	Beta Quintile					
Quintile	Lowest	2	3	4	Highest	H-L
			A. Alpha (%)			
Over-priced	-0.66	-0.63	-0.59	-0.69	-0.80	-0.15
2	-0.24	-0.24	-0.26	-0.25	-0.24	-0.01
3	0.02	-0.01	0.01	0.01	0.03	0.01
4	0.27	0.25	0.25	0.24	0.26	-0.01
Under-priced	0.66	0.65	0.67	0.71	0.72	0.07
Over-Under	-1.32	-1.28	-1.26	-1.40	-1.53	
All stocks	0.10	0.05	0.04	-0.07	-0.13	-0.24
			B. T statistics			
Over-priced	-14.99	-16.63	-15.55	-17.44	-27.39	-2.59
2	-15.54	-17.74	-20.74	-19.17	-15.80	-0.33
3	1.97	-0.65	1.22	1.17	1.87	0.65
4	21.32	24.28	22.70	19.04	13.05	-0.40
Under-priced	20.30	16.24	15.29	14.41	19.77	1.35
Over-Under	-24.17	-22.26	-20.36	-21.91	-32.73	
All stocks	3.29	1.80	1.35	-2.14	-2.85	-4.21

while the positive relation between α and beta for under-priced stocks is not statistically significant at the usual levels of confidence. Overall, the pattern of α reported in Table 1 is qualitatively similar to that reported in Liu, Stambaugh, and Yuan (2018).

Table 1 reports the results of one run of the simulation. To check the qualitative robustness of the results, we repeat the exercise 5000 times. The H-L cell in the over-price quintile is significant (at 5%) 89.5% of the 5000 simulation exercises. The corresponding numbers for H-L cells for the under-pricing quintile and all stocks are 14.6% and 100.0%, respectively.

The basic intuition of the reported simulation results can be explained as follows. When there is ambiguity in the mean only, (17) reduces to

$$\alpha_j = \lambda_\mu \left[\frac{\beta_{\mu,j}}{\beta_j} - 1\right] \beta_j. \tag{18}$$

In terms of averages,

$$\bar{\alpha}_{k,l} = \lambda_\mu \left(\bar{x}_{k,l}\bar{\beta}_{k,l} + \mathrm{cov}_{k,l}(x, \beta)\right), \qquad x_j = \beta_{\mu,j}/\beta_j - 1$$

where $\bar{\alpha}_{k,l}$ and $\bar{\beta}_{k,l}$ are the averages in each of the 5×5 mispricing by beta cells, indexed by (k, l). The covariance between $\beta_{\mu,j}/\beta_j$ and β_j is given by $\sum_j (x_j - \bar{x})(\beta_j - \bar{\beta})/(N_{ij} - 1)$ where $N_{k,l}$ is the number of stocks in the cell. Equation (18) suggests that for assets with positive betas, which is the case for most assets in the real world data and for most assets in our simulation,[10] there is over-pricing if and only if $\beta_{\mu,j} < \beta_j$. If $\beta_{\mu,j}/\beta_j$ and β_j are not highly correlated, then, for over-priced (under-priced) stocks with positive betas, α_j averaged for each beta quintile is decreasing (increasing) in β_j and hence H-L is negative (positive) for over-priced (under-priced) stocks. Thus, double sorting by mis-pricing and beta is likely to lead to what is seen in Panel A of Table 1.

The last row of Panel A of Table 1 is the beta anomaly. As risk beta increases from left to right, the mis-pricing decreases, leading to a flatter security market line. To understand it, note that according to Theorem 4,

$$\mu_j - r = \lambda\beta_j + \lambda_\mu\beta_{\mu,j}.$$

[10]See Table 1 of Liu, Stambaugh, and Yuan (2018). In our simulations, about 90% stocks have positive betas.

It is readily shown that standard OLS gives the following expression for the slope coefficient of β,

$$\hat{\lambda} \;=\; \mu_m - r + \lambda_\mu \left\{ \frac{Cov(\beta, \beta_\mu)}{Var\beta} - 1 \right\},$$

If $\mathrm{Var}(\beta_{\mu,j}) \leq \mathrm{Var}(\beta_j)$ and $\rho(\beta, \beta_\mu) < 1$, then $\frac{Cov(\beta,\beta_\mu)}{Var\beta} < 1$ and hence the slope is smaller than $\mu_m - r$. That is, the security market line is flatter than what the standard CAPM predicts.

The condition that $\rho(\beta, \beta_\mu) < 1$ seems sensible. It holds as long as β and β_μ are not perfectly correlated. For the condition that $\mathrm{Var}(\beta_{\mu,j}) \leq \mathrm{Var}(\beta_j)$, consider the regression of $\beta_{\mu,j}$ on β_j,

$$\beta_{\mu,j} = a + b\beta_j + \epsilon$$

Clearly, $\mathrm{Var}(\beta_{\mu,j}) = b^2\mathrm{Var}(\beta_j) + \mathrm{Var}(\epsilon)$. In our repeated simulation exercises, it is consistently found that $a > 0$ and $0 < b < 1$. Thus as long as $\mathrm{Var}(\epsilon)$ is not too large, $\mathrm{Var}(\beta_{\mu,j}) \leq \mathrm{Var}(\beta_j)$.

For a more intuitive understanding, let θ_S be a portfolio of a subset S of the N risky assets and $\sum_{j\in S} \theta_{S,j} = 1$. Specifically, let $\theta_{S,j} = 1/S$. The systematic risk of the portfolio is $\theta_S^\top \beta_S$ and the systematic ambiguity is $\theta_S^\top \beta_{\mu,S}$. Suppose that S is a portfolio of risky assets such that $\theta_S^\top \beta_S \approx 0$. Then,

$$\theta_S^\top \beta_{\mu,S} = a + b\theta_S^\top \beta_S + \theta_S^\top \epsilon_S \approx a + \theta_S^\top \epsilon_S$$

Since ϵ_j is conditionally uncorrelated with β_j and $E[\epsilon_j|\beta] = 0$, $\theta_S^\top \epsilon_S = \frac{1}{S}\sum_{j\in S} \epsilon_j \approx E[\epsilon_j] = E[E[\epsilon_j|\beta]] = 0$. Therefore $a > 0$ is equivalent to $\theta_S^\top \beta_{\mu,S} > 0$. Since $\theta_S^\top \beta_{\mu,S}$ measures the systematic ambiguity of the portfolio S, $a > 0$ simply means that the ambiguity averse investors ask for a positive premium for bearing the systematic ambiguity even though the systematic risk of the portfolio is approximately zero. Finally, as $\beta_m = \beta_{\mu,m} = 1$, $a > 0$ implies that the regression coefficient $b < 1$.

In summary, when the betas of the stocks are small (large), the ambiguity of the stocks may not be small (large) relatively. In regression terms, the regression of $\beta_{\mu,j}$ on β_j has positive intercept and coefficient of less than one. Because of that, small (large) beta stocks tend to be under-priced (over-priced). Therefore for small (large) beta stock, under-priced (over-priced) stocks tend to dominate the over-priced (under-priced) stocks. Consequently, as beta increases, the increase of α for under-priced stocks is not as steep as the decline of α for over-priced stocks. This asymmetry produces two effects. One is that H-L for under-priced stocks is less likely to be statistically significant. Second, H-L for all stocks is likely to be significant, which is shown in the last row of Panel A.

In Table 2, we report the results from 5000 repeated simulations. Panel A reports the averages of β_j. They are quite consistent across different mispricing quintiles. Panel B is the averages of the covariances between $\beta_{\mu,j}/\beta_j$ and β_j in each of the 5×5 mispricing and beta cells. They are all fairly small, ranging from -4.5% to 4.8%. Finally Panel C is the averages of $\beta_{\mu,j}/\beta_j - 1$. Their absolute values generally show a declining pattern across beta quintiles and there is asymmetry in the magnitude of the decline. In regression terms, when $\beta_{\mu,j}$ is regressed on β_j, the intercept term is positive and the regression coefficient is less than one. Indeed, in the 5000 simulations, the average intercept is 0.0658 and the average regression coefficient is 0.936.

The explanations of the beta anomaly provided in the literature are mostly based on short-selling or borrowing constraints. One argument is that when short-selling constraint is binding, investors behave as if they are holding the market portfolio and a zero-beta portfolio (Black (1972), Frazzini and Pedersen (2014)). The expected return on the zero-beta portfolio is higher than that of the riskless rate. Thus it appears that the security market line is flatter than the one predicted by the CAPM theory. Another argument is that heterogeneous expectations and short-sale constraints tend to lead to over-pricing of high beta stocks. Thus the security market line is flatter or even downward sloping in time of higher disagreement (Hong and Sraer (2016)). The third and more recent argument is that the beta anomaly maybe the consequence of the idiosyncratic volatility anomaly (Liu, Stambaugh, and Yuan (2018)).

As argued in Liu, Stambaugh, and Yuan (2018), while most of the explanations provided in the literature are consistent with the negative relation between α and beta as seen in the

Table 2: The Averages and Covariances of β and β_μ/β

The table reports the averages and covariances of β_j and $\beta_{\mu,j}$ across the 5 × 5 mispricing and beta quintiles. Panel A reports $\bar{\beta}$. Panel B reports cov(x, β). Panel C reports $\bar{x}$. Panel D reports the average number of stocks.

Mispricing	Beta Quintile				
Quintile	Lowest	2	3	4	Highest
			Panel A. β		
Over-priced	0.396	0.824	1.277	1.855	3.198
2	0.390	0.809	1.263	1.819	2.950
3	0.398	0.809	1.263	1.820	2.941
4	0.390	0.820	1.257	1.807	2.831
Under-priced	0.401	0.807	1.261	1.837	2.999
			Panel B. cov$(\beta_\mu/\beta, \beta)$		
Over-priced	0.048	0.010	0.004	0.004	0.020
2	0.025	0.004	0.002	0.002	0.009
3	0	0	0	0	0
4	-0.023	-0.005	-0.002	-0.002	-0.010
Under-priced	-0.045	-0.008	-0.006	-0.004	-0.014
			Panel C. $\beta_\mu/\beta - 1$		
Over-priced	-1.186	-0.515	-0.349	-0.255	-0.190
2	-0.533	-0.235	-0.145	-0.103	-0.066
3	0.010	0.006	0.002	0	-0.001
4	0.538	0.224	0.146	0.108	0.068
Under-priced	1.099	0.543	0.346	0.243	0.177
			Panel D. Number of Stocks		
Over-priced	25	25	31	41	63
2	33	33	33	37	33
3	38	39	37	29	22
4	34	30	35	34	22
Under-priced	37	40	31	27	26

last row of Panel A of Table 1, they are silent on the pattern of relations between α and beta when examine for over-priced and under-priced stocks separately, as shown in Panel A. Liu, Stambaugh, and Yuan (2018) provide their own explanation. Their argument is based on limits to arbitrage. Over-priced stocks are more difficult to arbitrage because of the higher cost in short sale, therefore the mispricing is stronger. Under-priced stocks on the other hand are easier to arbitrage. The positive relation between α and beta is weaker. The well-known beta anomaly is the net result of relative stronger effect of the negative relation between α and beta for over-priced stocks over that of the under-priced stocks. What we show is that even in the absence of short-sale constraints and limits to arbitrage, beta anomaly can still arise due to ambiguity.

6.3 Idiosyncratic Volatility (IVOL) Anomaly

Idiosyncratic volatility anomaly is a puzzling empirical pattern that was first documented by Ang et al. (2006). Stocks with higher idiosyncratic volatility have subsequent lower returns. It is puzzling because traditional theories predict either no relation between idiosyncratic volatility and expected returns (CAPM theory) or a positive relation due to market incompleteness and frictions (Merton (1987), Hirshleifer (1988)). As referred to in the introduction, a number of explanations have been provided in the literature. In this section, we provide a new angle for understanding the IVOL anomaly.

We first describe the simulation result. We use the same simulation data as in the preceding section, but double sort the data by idiosyncratic volatility instead of beta. Specifically, we independently assign stocks to Mispricing (α) Quintiles and IVOL Quintiles and obtain 5×5 intersecting cells.

The result is reported in the Table 3. Panel A reports the averages of the αs and Panel B reports the t-statistics of the averages. In the middle of Panel A are the 5×5 cells of double-sort. The last row reports the average of αs of all stocks sorted by IVOL. The second last row are the differences in α between the most over-priced and the most under-priced stocks. The last column of Panel A shows the differences, $H - L$, between the average αs of

Table 3: Alphas for Portfolios Sorted on IVOL and Mispricing

The table reports the alpha for portfolios formed by an independent 5 × 5 sort on IVOL and Mispricing.

Mispricing	IVOL Quintile					
Quintile	Lowest	2	3	4	Highest	H-L
			A. Alpha (%)			
Over-priced	-0.62	-0.60	-0.60	-0.62	-0.91	-0.29
2	-0.22	-0.25	-0.27	-0.21	-0.26	-0.04
3	0.00	0.01	0.00	0.00	0.01	0.01
4	0.26	0.26	0.28	0.26	0.28	0.01
Under-priced	0.60	0.60	0.61	0.67	0.79	0.18
Over-Under	-1.22	-1.20	-1.21	-1.30	-1.69	
All stocks	0.03	0.02	0.02	0.02	-0.09	-0.12
			B. T statistics			
Over-priced	-12.76	-21.44	-20.34	-17.27	-22.08	-3.46
2	-18.87	-19.26	-21.06	-17.59	-15.30	-1.88
3	0.43	1.11	-0.37	0.04	0.90	0.54
4	24.07	24.51	21.63	17.44	15.80	0.62
Under-priced	19.11	18.88	21.45	19.67	15.85	2.48
Over-Under	-22.09	-28.42	-29.46	-26.08	-26.50	
All stocks	1.19	0.64	0.59	0.48	-1.79	-2.13

the stocks with the highest IVOL and that with the lowest IVOL.

In Table 3, first we see that in the last row of Panel A, there is a negative relation between IVOL and return among over-priced stocks, which is the idiosyncratic volatility anomaly first reported in Ang et al. (2006). The rest of the result in Panel A has a similar pattern as in Panel of Table 1. When differentiated between over-priced and under-priced stocks, the first row of Panel A shows that among over-priced stocks, there is a negative relationship between α and IVOL, while the fifth row shows that among under-priced stocks, there is a positive relationship between α and IVOL. The middle rows, which are for stocks that are not obviously mispriced, α and IVOL exhibit a flat relation. Third, Panel B of Table 3 shows that the negative or positive relations between α and IVOL for all stocks, for over-priced stocks, or for under-priced stocks, measured by H-L in the last column in Panel A, are statistically significant. Overall, the result reported in Table 3 is qualitatively similar to that reported in Stambaugh, Yu, and Yuan (2015).

To provide the explanation of the result, we note that as in the case of beta anomaly, the mispricing is given by

$$\alpha_j = \lambda_\mu \left[\frac{\beta_{\mu,j}}{\beta_j} - 1\right] \beta_j = \lambda_\mu \left[\beta_{\mu,j} - \beta_j\right]. \tag{19}$$

Note next that

$$\beta_{\mu,j} = \frac{(\Omega_\mu \theta_m)_j}{\theta_m^\top \Omega_\mu \theta_m}, \quad \beta_j = \frac{(\Omega\theta)_j}{\theta^\top \Omega \theta}$$

Thus

$$\beta_{\mu,j} - \beta_j = \frac{\rho_{j,\Omega_\mu}\sigma_j}{\sqrt{\theta_m^\top \Omega_\mu \theta_m}} - \frac{\rho_{j,\Omega}\sigma_j}{\sqrt{\theta^\top \Omega \theta}} = \left[\frac{\rho_{j,\Omega_\mu}}{\sqrt{\theta_m^\top \Omega_\mu \theta_m}} - \frac{\rho_{j,\Omega}}{\sqrt{\theta^\top \Omega \theta}}\right]\sigma_j$$

$$U\theta_m = \frac{U\theta_m}{\sigma_j}\sigma_j$$

where ρ_{j,Ω_μ} is the correlation coefficient between the market portfolio and asset j when Ω_μ is taken as the variance-covariance matrix, and $\rho_{j,\Omega}$ is the correlation coefficient between the market portfolio and asset j when Ω is taken as the variance-covariance matrix. Clearly, ceteris paribus, the mispricing range is increasing in stock's total volatility (σ_j). However, ex ante, there is no reason that $\frac{\rho_{j,\Omega_\mu}}{\sqrt{\theta_m^\top \Omega_\mu \theta_m}}$ is alway greater than $\frac{\rho_{j,\Omega}}{\sqrt{\theta^\top \Omega \theta}}$, or vice versa. In fact for over-priced stocks, α_j is negative and $\frac{\rho_{j,\Omega_\mu}}{\sqrt{\theta_m^\top \Omega_\mu \theta_m}} < \frac{\rho_{j,\Omega}}{\sqrt{\theta^\top \Omega \theta}}$. Moreover, there is no ex ante reason that σ_j is strongly correlated with $\frac{\rho_{j,\Omega_\mu}}{\sqrt{\theta_m^\top \Omega_\mu \theta_m}} - \frac{\rho_{j,\Omega}}{\sqrt{\theta^\top \Omega \theta}}$. Then when over-priced stocked are divided into σ_j quintiles, there is likely a negative relation between mispricing and σ_j. As there is a strong correlation (over 95%) between total volatility and idiosyncratic volatility, that negative relation implies a negative relation between mispricing and IVOL, which explains the first row of Panel. The row for the under-priced stock in Panel can be explained by a similar argument.

Table 3 reports the results of one run of the simulation. Again, to check the qualitative robustness of the results, in the repeated 5000 exercises for beta anomaly, we also checked for the idiosyncratic volatility anomaly. The H-L cell in the over-price quintile is significant (at 5%) 99.6% of the 5000 simulation exercises. The corresponding numbers for H-L cells for the under-pricing quintile and all stocks are 99.9% and 47.8%, respectively. Interestingly, in contrast to the case of beta anomaly, the H-L cell for all stocks are significant for just under 50% of the 5000 simulations.

6.4 Beta Anomaly and Idiosyncratic Volatility Anomaly

So far, we have looked at the beta anomaly and idiosyncratic volatility separately. Since the same set simulation data exhibit both of these two anomalies as in the real world, one cannot help wonder if there is a deeper connection between the two. In our model, the connection is that they are both caused by ambiguity. However, the mechanisms are not exactly the same and hence there are some differences.

Now obviously, if there is a strong positive correlation between idiosyncratic volatility and beta, then beta anomaly and idiosyncratic volatility anomaly might be highly related. The total volatility can be decomposed as

$$\sigma_i^2 = \beta_i^2 \sigma_m^2 + \sigma_{i,\epsilon}^2.$$

Because of diversification, the total volatility of the market portfolio is typically much smaller than the volatility of individual stock's. Thus total volatility (σ_i) is highly correlated with idiosyncratic volatility ($\sigma_{i,\epsilon}$). This is consistent with empirical findings. Empirically, the correlations between total volatility and idiosyncratic volatility in the G7 countries are all over 95% (Ang et al. (2009)). On the other hand, the beta of an individual stock is

$$\beta_i = \rho_{i,m} \frac{\sigma_i}{\sigma_m}.$$

As there is no obvious reason that $\rho_{i,m}$ is highly correlated with σ_i, other things being equal, high total volatility should imply high risk beta. This argument suggests there is a positive relation between total volatility and beta, which is also what is true both in our simulation data set and in the real world data as well. Liu, Stambaugh, and Yuan (2018) reports a correlation coefficient of 0.33. In our 5000 repeated simulation exercises, the average is 0.83.

Liu, Stambaugh, and Yuan (2018) makes an interesting argument about the potential connection between the beta anomaly and the idiosyncratic volatility anomaly. They argue that the high correlation between idiosyncratic volatility and beta then implies that it is very likely that one sees both anomalies at the same time. They argue further that the beta anomaly is likely due to idiosyncratic volatility anomaly, as the latter is likely caused by limits to arbitrage (Stambaugh, Yu, and Yuan (2015)) while there is no such obvious reason for beta anomaly. Indeed, in the 5000 repeated simulations, for over-priced stocks, beta anomaly appears in 90% of the simulations and idiosyncratic volatility anomaly appears 99.6%. Interestingly, however, for under-priced stocks and for all stocks the two anomalies are not as highly correlated as the high correlation between idiosyncratic volatility and beta would suggest. In the 5000 repeated simulations, for under-priced stocks (all stocks), idiosyncratic volatility anomaly appears 99.88% (47.8%) while beta anomaly appears 14.84% (100%). Such asymmetry between under vs over-priced stocks is also evidenced in Stambaugh, Yu, and Yuan (2015) and Liu, Stambaugh, and Yuan (2018).

7 Conclusion

We develop a model that is useful for understanding the cross-sectional characteristics of asset returns. The model is otherwise standard. The additional ingredient is that the agent is ambiguous about the probability distribution of the returns of the assets and she is ambiguity averse. The ambiguity can be about the mean as well as the variance-covariance matrix of the returns. The equilibrium cross-sectional expected returns can be described by a three-factor model, capturing risk, mean ambiguity and variance-covariance ambiguity respectively. Expected returns include a mean ambiguity premium, a variance-covariance ambiguity premium, as well as the standard risk premium.

The cross-sectional asset returns in our model can exhibit a number of patterns that are silent in standard models. The most important is that mispricing, relative to standard models, occur not because the omitted important economic variables, but simply because the presence of ambiguity. In a simulation study, we examine two salient cross-sectional regularities of asset returns: the beta anomaly and the idiosyncratic volatility anomaly. We show that overall the alpha in our model decreases with beta. However, when sorted by mis-pricing, alpha of over-priced assets decreases with beta, while alpha of under-priced assets increases with beta. The alphas' exhibit similar characteristics when sorted by total or idiosyncratic volatility. Alpha of over-priced assets decreases with total or idiosyncratic volatility, while alpha of under-priced assets increases with total or idiosyncratic volatility. Overall alpha decreases with beta total or idiosyncratic volatility. As argued by Liu, Stambaugh, and Yuan (2018), reconciling these cross-sectional characteristics of asset returns is important for understanding the beta anomaly and the idiosyncratic volatility anomaly. They argue that limits to arbitrage and the resulting inefficiency may give rise the return characteristics. The study in this paper provides an alternative perspective.

A Appendix

A.1 Relative Entropy

Suppose that $R \sim N(\mu, \Omega)$ under P and $R \sim N(\hat{\mu}, \hat{\Omega})$ under Q. Then

$$\begin{aligned}
E[\xi \ln(\xi)] &= E^Q[\ln \xi] = \frac{1}{2} E^Q \left[\ln(\frac{|\Omega|}{|\hat{\Omega}|}) - (R-\hat{\mu})^\top \hat{\Omega}^{-1}(R-\hat{\mu}) + (R-\mu)^\top \Omega^{-1}(R-\mu) \right] \\
&= \frac{1}{2}\ln(\frac{|\Omega|}{|\hat{\Omega}|}) + \frac{1}{2} E^Q[-tr(\hat{\Omega}^{-1}(R-\hat{\mu})(R-\hat{\mu})^\top) + tr(\Omega^{-1}(R-\mu)(R-\mu)^\top] \\
&= \frac{1}{2}[\ln(\frac{|\Omega|}{|\hat{\Omega}|}) - N + tr(\Omega^{-1}\hat{\Omega}) + (\mu-\hat{\mu})^\top \Omega^{-1}(\mu-\hat{\mu})] \\
&= \frac{1}{2}[tr(\Omega^{-1}(\hat{\Omega}-\Omega)) - \ln|\Omega^{-1}\hat{\Omega}| + (\mu-\hat{\mu})^\top \Omega^{-1}(\mu-\hat{\mu})],
\end{aligned}$$

as is to be shown. ∎

A.2 Proof of Lemma 1

The proof for the first statement of the lemma is the same as that in Kogan and Wang (2003). The proof of the second statement is straightforward. ∎

A.3 Proof of Lemma 2

Uniqueness: Note that the objective function is a linear function of $\hat{\Omega}$. In order to prove the uniqueness of the solution, we first prove the convexity of the constraints function. For any $k \in K$, denote

$$\begin{aligned}
g(\hat{\Omega}_{J_k}) &= \frac{1}{2}[\ln(\frac{|\Omega_{J_k}|}{|\hat{\Omega}_{J_k}|}) - N_{J_k} + tr(\Omega_{J_k}^{-1}\hat{\Omega}_{J_k})] - \phi^2 \eta_{2,k}, \\
&= \frac{1}{2}[-\ln(|\hat{\Omega}_{J_k}|) + tr(\Omega_{J_k}^{-1}\hat{\Omega}_{J_k})] + \frac{1}{2}[\ln(|\Omega_{J_k}|) - N_{J_k}] - \eta_{2,k}, \\
&= \frac{1}{2}[-\ln(|\hat{\Omega}_{J_k}|) + tr(\Omega_{J_k}^{-1}\hat{\Omega}_{J_k})] + C_k,
\end{aligned}$$

where $C_k = \frac{1}{2}[\ln(|\Omega_{J_k}|) - N_{J_k}] - \eta_{2,k}$ is a constant. We need to show, for any $\hat{\Omega}^1_{J_k}$, $\hat{\Omega}^2_{J_k}$ and $a \in (0,1)$,

$$g(a\hat{\Omega}^1_{J_k} + (1-a)\hat{\Omega}^2_{J_k}) \le a g(\hat{\Omega}^1_{J_k}) + (1-a) g(\hat{\Omega}^2_{J_k}),$$

That is, we need to show,

$$-\ln(|a\hat{\Omega}^1_{J_k} + (1-a)\hat{\Omega}^2_{J_k}|) + tr[\Omega^{-1}_{J_k}(a\hat{\Omega}^1_{J_k} + (1-a)\hat{\Omega}^2_{J_k})] \leq$$
$$a[-\ln(|\hat{\Omega}^1_{J_k}|) + tr(\Omega^{-1}_{J_k}\hat{\Omega}^1_{J_k})] + (1-a)[-\ln(|\hat{\Omega}^2_{J_k}|) + tr(\Omega^{-1}_{J_k}a\hat{\Omega}^2_{J_k})],$$

which is,

$$\ln(|a\hat{\Omega}^1_{J_k} + (1-a)\hat{\Omega}^2_{J_k}|) \geq a\ln(|\hat{\Omega}^1_{J_k}|) + (1-a)\ln(|\hat{\Omega}^2_{J_k}|).$$

From the Minkowski Inequality, if A and B are positive semidefinite Hermite Matrices,

$$|A+B| \geq |A| + |B|,$$

Therefore,

$$\ln(|a\hat{\Omega}^1_{J_k} + (1-a)\hat{\Omega}^2_{J_k}|) \geq \ln(a|\hat{\Omega}^1_{J_k}| + (1-a)|\hat{\Omega}^2_{J_k}|) \geq a\ln(|\hat{\Omega}^1_{J_k}|) + (1-a)\ln(|\hat{\Omega}^2_{J_k}|),$$

as desired.

Suppose to the contrary that there exist two distinct solutions, $\hat{\Omega}^1$ and $\hat{\Omega}^2$. For any $a \in (0,1)$, denote $\hat{\Omega}^a = a\hat{\Omega}^1 + (1-a)\hat{\Omega}^2$ and let $\hat{\Omega}^h_{J_k}$, $h = (1,2,a)$ denote the corresponding solution for J_k. The convexity of all the constraints functions implies that

$$g(\hat{\Omega}^a_{J_k}) = \frac{1}{2}[\ln(\frac{|\Omega_{J_k}|}{|\hat{\Omega}^a_{J_k}|}) - N_{J_k} + tr(\Omega^{-1}_{J_k}\hat{\Omega}^a_{J_k})] - \phi^2\eta_{2,k} \leq ag(\hat{\Omega}^1_{J_k}) + (1-a)g(\hat{\Omega}^2_{J_k}) \leq 0,$$

for $k = 1, 2, \ldots, K$.

Let $k \in \{1, \ldots, K\}$. Suppose that k is such that

$$\frac{1}{2}[\ln(\frac{|\Omega_{J_k}|}{|\hat{\Omega}^a_{J_k}|}) - N_{J_k} + tr(\Omega^{-1}_{J_k}\hat{\Omega}^a_{J_k})] - \eta_{2,k} = 0, \quad \text{for } a = 0, 1, \bar{a}.$$

where $\bar{a} \in (0,1)$. Then by strict convexity, we must have $\hat{\Omega}^1_{J_k} = \hat{\Omega}^2_{J_k}$. Denote by A the set of all such k. If

$$J_A = \cup_{k\in A} J_k = \{1, 2, ..., N\},$$

then $\hat{\Omega}^1 = \hat{\Omega}^2$ and we have a contradiction. If $J_A \neq \{1, 2, ..., N\}$, without loss of generality, assume that $1 \notin J_A$ and $J_A = \{2, ..., N\}$. Then all $\hat{\Omega}$ of the following form

$$\hat{\Omega} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{12}^1 & \vdots & \sigma_{2N}^1 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2}^1 & \cdots & \sigma_{NN}^1 \end{bmatrix} \tag{20}$$

satisfy

$$\frac{1}{2}[\ln(\frac{|\Omega_{J_k}|}{|\hat{\Omega}_{J_k}|}) - N_{J_k} + tr(\Omega_{J_k}^{-1}\hat{\Omega}_{J_k})] - \eta_{2,k} = 0, \quad \text{for } k \in A.$$

where $\sigma_{1i} \in \mathcal{R},\ i = 1, 2, ..., N$ is the variance (covariance). Since $\hat{\Omega}^1$ and $\hat{\Omega}^2$ are solutions, we have, for each $k \notin A$,

$$\frac{1}{2}[\ln(\frac{|\Omega_{J_k}|}{|\hat{\Omega}_{J_k}^a|}) - N_{J_k} + tr(\Omega_{J_k}^{-1}\hat{\Omega}_{J_k}^a)] - \eta_{2,k} < 0, \quad \text{for } \ k \notin A.$$

for either $a = 1$ or 2. By strict convexity shown earlier, the strict inequality holds for all $a \in (0, 1)$, in particular for $a = 1/2$. Note that for $a = \frac{1}{2}$,

$$\hat{\Omega}^a = \begin{bmatrix} \frac{\sigma_{11}^1+\sigma_{11}^2}{2} & \frac{\sigma_{12}^1+\sigma_{12}^2}{2} & \cdots & \frac{\sigma_{1N}^1+\sigma_{1N}^2}{2} \\ \frac{\sigma_{21}^1+\sigma_{21}^2}{2} & \frac{\sigma_{22}^1+\sigma_{22}^2}{2} & \vdots & \frac{\sigma_{2N}^1+\sigma_{2N}^2}{2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{N1}^1+\sigma_{N1}^2}{2} & \frac{\sigma_{N2}^1+\sigma_{N2}^2}{2} & \cdots & \frac{\sigma_{NN}^1+\sigma_{NN}^2}{2} \end{bmatrix} = \begin{bmatrix} \frac{\sigma_{11}^1+\sigma_{11}^2}{2} & \frac{\sigma_{12}^1+\sigma_{12}^2}{2} & \cdots & \frac{\sigma_{1N}^1+\sigma_{1N}^2}{2} \\ \frac{\sigma_{21}^1+\sigma_{21}^2}{2} & \sigma_{22}^1 & \vdots & \sigma_{2N}^1 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{N1}^1+\sigma_{N1}^2}{2} & \sigma_{N2}^1 & \cdots & \sigma_{NN}^1 \end{bmatrix}$$

because $J_A = \{2, ..., N\}$. By continuity, for small $\epsilon > 0$, all the $\hat{\Omega}$ in (20) with $\sigma_{11} \in (\frac{\sigma_{11}^1+\sigma_{11}^2}{2} - \epsilon, \frac{\sigma_{11}^1+\sigma_{11}^2}{2} + \epsilon)$ and $\sigma_{1i} = \frac{\sigma_{1i}^1+\sigma_{1i}^2}{2}$, for $i = 2, ..., N$, satisfy

$$\frac{1}{2}[\ln(\frac{|\Omega_{J_k}|}{|\hat{\Omega}_{J_k}|}) - N_{J_k} + tr(\Omega_{J_k}^{-1}\hat{\Omega}_{J_k})] - \eta_{2,k} < 0, \quad \text{for } \ k \notin A.$$

Combining with the case $k \in A$, we have,

$$\frac{1}{2}[\ln(\frac{|\Omega_{J_k}|}{|\hat{\Omega}_{J_k}|}) - N_{J_k} + tr(\Omega_{J_k}^{-1}\hat{\Omega}_{J_k})] - \eta_{2,k} \leq 0, \quad \text{for } \ k = 1, 2, ..., K.$$

That is, all so defined $\hat{\Omega}$ are in the choice set. As the objective function is a linear function of $\hat{\Omega}$, so we have,

$$\theta^\top(\hat{\Omega} - \Omega)\theta = \theta^\top(\hat{\Omega}^{1/2} - \Omega)\theta + \theta_1^2(\sigma_{11} - (\sigma_{11}^1 + \sigma_{11}^2)/2)$$

Clearly, we can choose specific σ_{11} to achieve higher value of $\theta^\top(\hat{\Omega} - \Omega)\theta$, which is a contradiction!

Next, write down the Lagrangian function as follows,

$$\mathcal{L} = \theta^\top U\theta - \sum_{k=1}^{K} \lambda_{2,k} \left\{ \frac{1}{2} \left[tr(\Omega_{J_k}^{-1} U_{J_k}) - \ln |I_{J_k} + \Omega_{J_k}^{-1} U_{J_k}| \right] - \eta_{2,k} \right\},$$

Note that $\partial tr(\Omega^{-1}U)/\partial u_{ij} = tr(\Omega^{-1}U_{ij})$ where U_{ij} is the matrix which has zero everywhere except in the ith row and jth column where it is equal to 1. $\partial \ln |I + \Omega^{-1}U|/\partial u_{ij} = (\Omega + U)_{ij}^{-1}$. $tr(A^\top B) = \sum_i \sum_j A_{ij}B_{ij}$. FOC is

$$\frac{\partial \mathcal{L}}{\partial U} = \theta\theta^\top \circ S - \sum_{k=1}^{K} \frac{\lambda_{2,k}}{2} \left[-(\Omega_{J_k} + U_{J_k})^{-1} + \Omega_{J_k}^{-1} \right] = 0,$$

where S is a sign matrix whose elements take 1 if there is variance-covariance ambiguity information about the corresponding elements in Ω and takes 0 otherwise. $\theta\theta^\top \circ S$ is the entry-wise product between two matrices, which produces another matrix where each element ij is the product of elements ij of the original two matrices. So

$$\sum_{k=1}^{K} \lambda_{2,k}(\theta)(\Omega_{J_k} + U_{J_k})^{-1} = \sum_{k=1}^{K} \lambda_{2,k}(\theta)\Omega_{J_k}^{-1} - 2\theta\theta^\top \circ S.$$

Note, similar with Lemma 1, the proof above is also based on the assumption that there are multiple sources of information (K) and the information can cover all the assets in the market. So $\sum_{k=1}^{K} \lambda_{2,k}(\theta)\Omega_{J_k}^{-1}$ should be a full-rank matrix. If there is no variance-covariance ambiguity about some elements in the original Ω, the above equations becomes $0 = 0$ in the corresponding elements, which means those equations are redundant. ∎

A.4 Proof of Proposition 3

The agents utility maximization problem is

$$\sup_{\theta} \min_{Q\in\mathcal{P}} E^{Q}\left[-\frac{1}{\gamma}e^{-\gamma\left[\theta^{\top}(R-r\mathbf{1})+(1+r)\right]}\right] = \sup_{\theta}\left[-\frac{1}{\gamma}e^{-\gamma[\theta^{\top}(\mu-r\mathbf{1})+1+r-\Delta_1(\theta)]+\frac{1}{2}\gamma^2[\theta^{\top}\Omega\theta+\Delta_2(\theta)]}\right],$$

The FOC for θ is given by

$$\mu - r\mathbf{1} - \Delta_1'(\theta) - \gamma\Omega\theta - \frac{1}{2}\gamma\Delta_2'(\theta) = 0,$$

So the optimal portfolio choice follows

$$\mu - r\mathbf{1} = \Delta_1'(\theta) + \gamma\Omega\theta + \frac{1}{2}\gamma\Delta_2'(\theta).$$

By envelope theorem, we have, $\Delta_1'(\theta) = v(\theta)$ and $\Delta_2'(\theta) = 2U(\theta)\theta$. Thus

$$\mu - r\mathbf{1} = v(\theta) + \gamma\left[\Omega + U(\theta)\right]\theta,$$

as is to be shown. ∎

A.5 Proof of Lemma 5

(13) follows readily from Lemma 1. For variance-covariance ambiguity, let

$$\mathcal{L} = \theta^{\top}U\theta - \sum_{k=1}^{K}\lambda_{2,k}\left(\left[tr(\Omega_{J_k}^{-1}U_{J_k}) - \ln|I_{J_k} + \Omega_{J_k}^{-1}U_{J_k}|\right] - 2\eta_{2,k}\right)$$

It follows that the FOC is,

$$\sum_{k=1}^{K}\lambda_{2,k}(\theta)[\Omega_{J_k} + U_{J_k}^{*}(\theta)]^{-1} = \sum_{k=1}^{K}\lambda_{2,k}(\theta)\Omega_{J_k}^{-1} - 2\theta\theta^{\top}.$$

As J_k, $k = 1, \ldots, K$, are non-overlapping,

$$U_{J_k}^{*}(\theta) \;=\; \left(\Omega_{J_k}^{-1} - \frac{2}{\lambda_{2,k}(\theta)}\theta_{J_k}\theta_{J_k}^{\top}\right)^{-1} - \Omega_{J_k} = \frac{2}{\lambda_{2,k}(\theta) - 2\theta_{J_k}^{\top}\Omega_{J_k}\theta_{J_k}}\Omega_{J_k}\theta_{J_k}\theta_{J_k}^{\top}\Omega_{J_k}.$$

Plugging into the constraint, we can solve for $\lambda_2(\theta)$,

$$\begin{aligned} 2\eta_{2,k} &= -\ln|I + \Omega_{J_k}^{-1}U_{J_k}^*(\theta)| + tr(\Omega_{J_k}^{-1}U_{J_k}^*(\theta)) \\ &= -\ln(|I + \frac{2}{\lambda_{2,k}(\theta) - 2\theta_{J_k}^\top \Omega_{J_k}\theta_{J_k}}\theta_{J_k}\theta_{J_k}^\top \Omega_{J_k}|) + \mathrm{tr}\left(\frac{2}{\lambda_{2,k}(\theta) - 2\theta_{J_k}^\top \Omega_{J_k}\theta_{J_k}}\theta_{J_k}\theta_{J_k}^\top \Omega_{J_k}\right) \\ &= -\ln\left(1 + \frac{2\theta_{J_k}^\top \Omega_{J_k}\theta_{J_k}}{\lambda_{2,k}(\theta) - 2\theta_{J_k}^\top \Omega_{J_k}\theta_{J_k}}\right) + \frac{2\theta_{J_k}^\top \Omega_{J_k}\theta_{J_k}}{\lambda_{2,k}(\theta) - 2\theta_{J_k}^\top \Omega_{J_k}\theta_{J_k}}. \end{aligned}$$

as is to be shown. It is readily seen that U is positive definite. ∎

References

Anderson, E. W., L. P. Hansen, and T. J. Sargent. 2003. A quartet of semigroups for model specification, robustness, prices of risk, and model detection. *Journal of the European Economic Association* 1:68–123.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang. 2006. The cross-section of volatility and expected returns. *The Journal of Finance* 61:259–99.

———. 2009. High idiosyncratic volatility and low returns: International and further u.s. evidence. *Journal of Financial Economics* 11:1–23.

Baker, M., B. Bradley, and J. Wurgler. 2011. Benchmarks as limits to arbitrage: understanding the low-volatility anomaly. *Financial Analysts Journal* 67:4054–.

Bali, T. G., and N. Cakici. 2008. Idiosyncratic volatility and the cross section of expected returns. *Journal of Financial and Quantitative Analysis* 43:2958–.

Bali, T. G., N. Cakici, and R. F. Whitelaw. 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics* 99:427–46.

Barberis, N., and M. Huang. 2008. Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review* 98:2066–100.

Black, F. 1972. Capital market equilibrium with restricted borrowing. *The Journal of Business* 45:444–55.

Black, F., M. C. Jensen, and M. S. Scholes. 1972. The capital asset pricing model: some empirical tests. *Studies in the Theory of Capital Markets* .

Blitz, D., E. Falkenstein, and P. van Vliet. 2014. Explanations for the volatility effect: An overview based on the capm assumptions. *Journal of Portfolio Management* 40:61–76.

Boyer, B., T. Mitton, and K. Vorkink. 2010. Expected idiosyncratic skewness. *Review of Financial Studies* 23:169–202.

Chen, Z., and L. Epstein. 2002. Ambiguity, risk, and asset returns in continuous time. *Econometrica* 70:1403–43.

Chen, Z., and R. Petkova. 2012. Does idiosyncratic volatility proxy for risk exposure? *Review of Financial Studies* 25:2745–87.

Christoffersen, S. E. K., and M. Simutin. 2017. On the demand for high-beta stocks: Evidence from mutual funds. *Review of Financial Studies* 30:2596–620.

Dow, J., and S. R. d. C. Werlang. 1992. Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica* 60:197–204.

Easley, D., and M. O'Hara. 2009. Ambiguity and non-participation: The role of regulation. *Review of Financial Studies* 22:1817–44.

———. 2010. Microstructure and ambiguity. *Journal of Finance* 65:1827–46.

Epstein, L. G., and S. Ji. 2013. 'ambiguous volatility and asset pricing in continuous time. *Review of Financial Studies* 26:1740 – 1786.

Epstein, L. G., and J. Miao. 2003. A two-person dynamic equilibrium under ambiguity. *Journal of Economic Dynamics and Control* 27:1253 – 1288.

Epstein, L. G., and M. Schneider. 2003. Recursive multiple-priors. *Journal of Economic Theory* 113:1 – 31. ISSN 0022-0531.

Epstein, L. G., and T. Wang. 1994. Intertemporal asset pricing under knightian uncertainty. *Econometrica* 62:283–322.

———. 1995. Uncertainty, risk-neutral measures and security price booms and crashes. *Journal of Economic Theory* 67:40 – 82.

Fama, E., and J. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81:607–36.

Frazzini, A., and L. H. Pedersen. 2014. Betting against beta. *Journal of Financial Economics* 111:1–25.

Fu, F. 2009. Idiosyncratic risk and the cross-section of expected stock returns. *Journal of Financial Economics* 91:24–37.

Gilboa, I., and D. Schmeidler. 1989. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* 18:141 – 153.

Hansen, L. P., and T. J. Sargent. 2001. Robust control and model uncertainty. *The American Economic Review* 91:60–6.

Hayashi, T., and J. Miao. 2011. Recursive smooth ambiguity preferences,. *Theoretical Economics* 6:423–75.

Hirshleifer, D. 1988. Residual risk, trading costs, and commodity futures risk premia. *Review of Financial Studies* 1:173–93.

Hong, H., and D. Sraer. 2016. Speculative betas. *Journal of Finance* 71:2095–144.

Hou, K., and R. Loh. 2016. Have we solved the idiosyncratic volatility puzzle? *The Journal of Financial Economics* 121:167–94.

Huang, W., Q. Liu, S. G. Rhee, and L. Zhang. 2010. Return reversals, idiosyncratic risk, and expected returns. *Review of Financial Studies* 23:147–68.

Jiang, G. J., D. Xu, and T. Yao. 2009. The information content of idiosyncratic volatility. *Journal of Financial and Quantitative Analysis* 44:1–28.

Johnson, T. 2004. Forecast dispersion and the cross section of expected returns. *Journal of Finance* 59:1957–78.

Ju, N., and J. Miao. 2012. Ambiguity, learning, and asset returns. *Econometrica* 80:559–91.

Klibanoff, P., M. Marinacci, and S. Mukerji. 2005. A smooth model of decision making under ambiguity. *Econometrica* 73:1849–92.

———. 2009. Recursive smooth ambiguity preferences,. *Journal of Economic Theory* 144:930–76.

Kogan, L., and T. Wang. 2003. A simple theory of asset pricing under model uncertainty. Working paper.

Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics* 47:13–37.

Liu, J., J. Pan, and T. Wang. 2005. An equilibrium model of rare-event premia and its implication for option smirks. *Review of Financial Studies* 18:131–64.

Liu, J., R. F. Stambaugh, and Y. Yuan. 2018. Absolving beta of volatilitys effects. *Journal of Financial Economics* 128:1 – 15. ISSN 0304-405X.

Liu, J., and X. Zeng. 2017. Correlation ambiguity. *working paper* .

Maenhout, P. 2004. Robust portfolio rules and asset pricing. *Review of Financial Studies* 17:951–83.

———. 2006. Robust portfolio rules and detection-error probabilities for a mean-reverting risk premium. *Journal of Economic Theory* 128:136–63.

Merton, R. C. 1980. On estimating the expected return on the market: an exploratory investigation. *Journal of Financial Economics* 8:323–61.

———. 1987. A simple model of capital market equilibrium with incomplete information. *Journal of Finance* 42:483–510.

Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19:425–42.

Stambaugh, R. F., J. Yu, and Y. Yuan. 2015. Arbitrage asymmetry and the idiosyncratic volatility puzzle. *Journal of Finance* 70:1903–48.

Uppal, R., and T. Wang. 2003. Model misspecification and underdiversification. *Journal of Finance* 58:2465–86.

Wong, P. 2011. Earnings shocks and the idiosyncratic volatility discount in the cross-section of expected returns. *working paper* .