Submitted to *Operations Research*
manuscript (Please, provide the manuscript number!)

Beta and Coskewness Pricing: Perspective from Probability Weighting

Yun Shi
School of Statistics and Academy of Statistics and Interdisciplinary Sciences, East China Normal University, Shanghai 200062, China, yshi@fem.ecnu.edu.cn, ORCID https://orcid.org/0000-0002-4746-2312

Xiangyu Cui
School of Statistics and Management, Shanghai Institute of International Finance and Economics, Shanghai University of Finance and Economics, Shanghai, China. cui.xiangyu@mail.shufe.edu.cn, ORCID https://orcid.org/0000-0003-1749-4513

Xun Yu Zhou
Department of Industrial Engineering and Operations Research, and The Data Science Institute, Columbia University, New York, NY 10027, USA. xz2574@columbia.edu, ORCID https://orcid.org/0000-0001-9908-5697

The security market line is often flat or downward-sloping. We hypothesize that probability weighting plays a role and that one ought to differentiate between periods in which agents overweight extreme events and those in which they underweight them. Overweighting inflates the probability of extremely bad events and demands greater compensation for beta risk, whereas underweighting does the opposite. Unconditional on probability weighting, these two effects offset each other, resulting in a flat or slightly negative return-beta relationship. Similarly, overweighting the tails enhances the negative relationship between return and coskewness, while underweighting reduces it. We derive a three-moment conditional CAPM model for a market with rank-dependent utility agents to make these predictions, and we support our theory through an extensive empirical study.

Key words: behavioral finance, probability weighting, rank dependent utility, asset pricing, beta anomaly

Subject classifications: finance: asset pricing, portfolio selection

Area of review: Financial engineering

1. Introduction

The classical CAPM theory of Sharpe (1964), Lintner (1965), and Mossin (1966) asserts that expected returns increase with beta, leading to an upward-sloping security market line. However, empirical studies have shown that the return–beta slope is flat or even downward-sloping; see, e.g., Black et al. (1972), Fama and French (1992), and Baker et al. (2011). Various explanations have been offered for why beta is not or is negatively priced, ranging from misspecification/miscalculation of risk (Jagannathan and Wang 1996; Bali et al. 2017b), inefficiency of market proxies (Roll and Ross 1994), market frictions (Black et al. 1972; Frazzini and Pedersen 2014), aggregate disagreement (Hong and Sraer 2016), lottery-type demand (Bali et al. 2017a), to information gap (Andrei et al. 2020).

We approach the beta anomaly from a different perspective, one that involves probability weighting. The central task of asset pricing is to characterize how expected returns are related to risk and to investors' *perceptions* of risk. Probability weighting affects the perception of risk, especially in the two tails of the market returns; therefore, it is poised to play a role in beta pricing. Meanwhile, a positive/negative skewness measures the risk of large positive/negative realizations and can be viewed as a part of tail risk. As a result, probability weighting is also relevant to skewness/coskewness pricing because substantially right-/left-skewed events, such as winning a lottery (Barberis and Huang 2008; Bordalo et al. 2012; Green and Hwang 2012) or encountering a catastrophe (Kelly and Jiang 2014; Bollerslev et al. 2015; Kozhan et al. 2013) are even more attractive/undesirable to an investor under probability weighting.

In this paper, we provide a theory for beta and coskewness pricing using rank-dependent utilities (Quiggin 1982; Schmeidler 1989; Abdellaoui 2002; Quiggin 2012). The key difference between the rank-dependent utility theory (RDUT) and the classical expected utility theory (EUT) is that in the former the utility is weighted by a probability weighting function assigned to ranked outcomes.[1] So RDUT nests EUT but also captures risk attitude toward probabilities, especially those represented in the two tails of the return distributions. We first derive an equilibrium asset pricing formula for a static, representative RDUT economy, and we then examine the signs and magnitudes of the risk premia in covariance and coskewness. Our results show that the pricing kernel in this economy is the product of the marginal utility and the derivative of probability weighting, which implies that the *shape* of the weighting function matters for pricing.

The dominant view in the behavioral finance literature is that individuals overweight probabilities of both extremely good and extremely bad events, rendering an inverse S-shaped probability weighting function (Tversky and Kahneman 1992; Wu and Gonzalez 1996; Prelec 1998; Hsu et al. 2009; Tanaka et al. 2010). Some authors, however, have documented empirical evidence that individuals sometimes *underweight* tail events (Hertwig et al. 2004; Humphrey and Verschoor 2004; Harrison et al. 2009; Henrich et al. 2010), leading to S-shaped weighting functions. Accordingly, Barberis (2013) comments that the coexistence of overweighting and underweighting poses a challenge in the study of extreme events. Employing the generalized method of moments (GMM), we use the S&P 500 index option data to infer the probability weighting functions, updated monthly. Among the total 244 months tested in our study, about 58.6% of the implied weighting functions are inverse S-shaped and 41.4% are S-shaped. This observation suggests that the periods of overweighting and underweighting over a sufficiently long time period (over 20 years in our tests) are almost 6:4. Polkovnichenko and Zhao (2013) have similar observation by taking a nonparametric method.

Our main theoretical result is a closed-form expression of a static three-moment CAPM, characterizing the risk premia of covariance and coskewness of any given asset with respect to the market. These premia depend on the utility function, as well as on the probability weighting function *evaluated* at the probability value of having a positive market return. During an overweighing month (in which the weighting function is inverse S-shaped), the agent overweights both tails. Overweighting the left tail makes the RDUT agent more risk averse toward bad states (related to the convex part of the weighting function) than her EUT counterpart, whereas overweighting the right tail makes her more risk taking toward good states (the concave part). One may be tempted to think that these two effects offset each other. However, we find empirically that the average probabilities of having a *positive* market return are sufficiently large that they fall in the *convex* domain of Prelec's function, a baseline inverse S-shaped weighting function (see Figure 1). This yields that only the convex part of the weighting function is *relevant*. Intuitively, in a typical stock market, the chance of observing an extreme negative realization is much smaller than that of observing the same size positive realization. Thus, the impact of extremely bad states is much stronger than that of extremely good states, which causes the agent to pay more attention to the former. This *one-sided* effect results in an *enhanced* positive risk premium with respect to the covariance with the market; see the second term inside the brackets of (7) for the enhancing effect driven by probability weighting. Symmetrically, during an underweighting month in which the agent underweights both tails of the market return distribution, she is less risk averse toward bad states and less risk seeking toward good ones. Once again, the left tail effect dominates and only the former behavior is relevant because the probability of having a positive market return now falls in the *concave* domain of the S-shaped weighting function. In this case, the agent demands less compensation for the covariance with the market compared to an EUT maximizer, and she may even be willing to pay for the risk if and when she sufficiently underweights the bad states. Finally, for a sufficiently long sample period covering comparable runs of overweighting and underweighting months, the elevated beta during the former periods and the reduced beta during the latter periods may cancel each other, leading to an overall flat or possibly downward-sloping security market line (see Figure 7(a)).

Kraus and Litzenberger (1976) and Harvey and Siddique (2000) show that a typical EUT agent is willing to pay for the coskewness, implying a negative premium in coskewness. Probability weighting is likely to strengthen this skewness-inclined preference during an overweighting period. This can be explained as follows. When the asset has positive coskewness with the market, the asset return has a fat right tail with respect to the market portfolio. The hope of a larger gain relative to the market due to overwheighting further increases the attraction of the risky asset and requires even less compensation when compared to the case of no probability weighting. Symmetrically, when the coskewness is negative, the fear of a larger loss, which is also amplified by the overweighting, reduces the attraction of the risky asset and hence demands greater compensation. Combined, overweighting the tails inflates the negative relationship between expected return and coskewness, generating a significantly negative premium of the coskewness (see Figure 7(b)). When the agent underweights both tails of the market return, the probability weighting introduces a skewness-averse preference. Underweighting the tails deflates the negative relationship between expected return and coskewness, resulting in an insignificant (or even potentially positive) premium.

The characteristically different risk premia between the overweighting and underweighting periods suggest that we *need* to separate the whole sample into over- and under-weighting regimes and analyze the corresponding beta and coskewness pricing *conditionally and separately*. Indeed, the GMM analysis shows the clear time-variation in both risk aversion and probability weighting. This motivates us to formulate a *conditional* three-moment CAPM and put its predictions to empirical tests. Following the empirical asset pricing literature, we first perform portfolio analysis to study the impact of probability weighting on the mean-covariance and mean-coskewness relationships, respectively. We use all of the common stocks listed on the NYSE, AMEX and NASDAQ for the period January 1991 – April 2016. The one-way sort portfolio analysis shows that the covariance spread portfolio has a significantly positive return during overweighting months, a significantly negative return during underweighting months, and a slightly positive (but insignificant) average return for the entire period. The coskewness spread portfolio, on the other hand, has a significantly negative return during overweighting months, an insignificantly negative return during underweighting months, and a notably negative average return for the entire period. A Fama–MacBeth cross-sectional regression analysis (Fama and MacBeth 1973) further confirms these results.

The paper proceeds as follows. In Section 2 we deduce the static three-moment CAPM and study the impact of probability weighting on asset pricing theoretically. In Section 3 we derive the time-varying risk aversion and probability weighting using GMM, which motivates the formulation of the conditional CAPM. We test the theoretical predictions by an extensive empirical analysis of the conditional return and covariance/coskewness relationship in Section 4. We conclude in Section 5. All the proofs and a robustness analysis are presented in the online appendix.

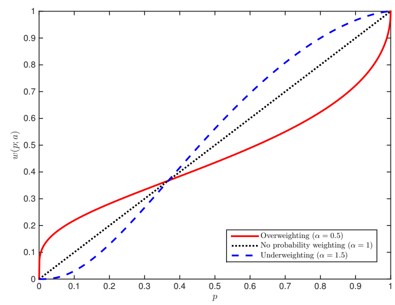
2. The Static CAPM

2.1. Portfolio Selection under Rank-Dependent Utility

Consider a one-period-two-date market where there are n risky assets and one risk-free asset. The set of possible states of nature at date 1 is Ω and the set of events at date 1 is a σ-algebra $\mathcal{F}$

of subsets of Ω. Any $\mathcal{F}$-measurable contingent claim ξ is priced as $\mathbb{E}[\tilde{m}\xi]$, where $\tilde{m}$ is the pricing kernel, which is an $\mathcal{F}$-measurable random variable such that $\mathbb{P}(\tilde{m} > 0) = 1$ and $\mathbb{E}[\tilde{m}] < \infty$.

Figure 1 Prelec's probability weighting functions with different α's



The representative agent in the market has an RDUT preference given by

$$U(\tilde{X}) = \int u(x)w'(1 - F_{\tilde{X}}(x))dF_{\tilde{X}}(x),$$

where $u(\cdot)$ is a utility function, $w(\cdot)$ is a probability weighting function, $w'(\cdot)$ is the first-order derivative of $w(\cdot)$, and $F_{\tilde{X}}(\cdot)$ is the cumulative distribution function (CDF) of the random payoff $\tilde{X}$. Specific parametric classes of probability weighting functions proposed in the literature include those by Tversky and Kahneman (1992), $w(p) = \frac{p^\alpha}{(p^\alpha + (1-p)^\alpha)^{1/\alpha}}$, and by Prelec (1998), $w(p) = \exp(-(-\log(p))^\alpha)$, where, in both cases, $0 < \alpha < 1$ corresponds to inverse S-shaped weighting functions (i.e., first concave and then convex; hence overweighting both tails) and $\alpha > 1$ corresponds to S-shaped one (i.e., first convex and then concave; hence underweighting both tails). In this paper, we will use Prelec's weighting functions, which are plotted in Figure 1. The smaller the $\alpha(< 1)$, the higher degree of overweighting, and the larger the $\alpha(> 1)$, the higher degree of underweighting.

Here we *assume* the existence of a representative investor – and hence the existence of a (well-diversified) market portfolio – in the RDUT market. Such an existence is *deduced* in Xia and Zhou (2016) under the assumption that all the investors have the same probability weighting function but may have heterogeneous utility functions. In a follow-up paper Jin et al. (2019), RDUT markets with heterogeneous probability weighting are studied, but the existence of a representative agent is not proved, which remains an outstanding open question to our best knowledge. To get around this difficulty, in this paper we take a "revealed preference" perspective, adopted also by Post and Levy (2005), based on the empirical observation that many investors are known to invest in broadly diversified passive security portfolios such as market index ETFs. The static three-moment CAPM established in this section lays out what we can do with *theoretical* deductions. It in turn motivates a structural model – a conditional CAPM – formulated in Section 3, which is further put to empirical tests in Section 4.

We make the following assumptions on the market.

ASSUMPTION 1. (i) The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ admits no atom.

(ii) u is strictly increasing, strictly concave, continuously differentiable on $(0,\infty)$, and satisfies the Inada condition: $u'(0+) = \infty,\ u'(\infty) = 0$. Moreover, without loss of generality, $u(\infty) > 0$. The asymptotic elasticity of u is strictly less than one: $\lim_{x\to\infty} \frac{xu'(x)}{u(x)} < 1$.

(iii) w is strictly increasing and continuously differentiable on $[0,1]$ and satisfies $w(0) = 0$, $w(1) = 1$.

(iv) The total return of the market portfolio $\widetilde{R}_M$ satisfies $1 - F_{\widetilde{R}_M}(1) > e^{-1}$.

Assumption 1(i) means that there is no state of nature ω such that $\mathbb{P}(\{\omega\}) > 0$. Being non-atomic is a standard assumption for *continuous* probability space to avoid non-essential technicalities. Assumptions 1(ii) and (iii) are standard and mild on the preference functions u and w. Finally, (iv) is a condition on the market return, yielding that $1 - F_{\widetilde{R}_M}(1)$ lies in the convex (concave) domain of $w(\cdot)$ when $\alpha < 1$ ($\alpha > 1$). In Section 3 of the online appendix, we show empirically that this condition is met by the S&P 500.

Denote by $\widetilde{R}_P$ the total return of a portfolio P. Our portfolio selection problem is formulated as follows,

$$\begin{aligned} \max_{\widetilde{R}_P \text{ is } \mathcal{F}\text{-measurable}} \quad & \int u(x) w'(1 - F_{\widetilde{R}_P}(x)) dF_{\widetilde{R}_P}(x), \\ \text{s.t.} \quad & \mathbb{E}[\tilde{m}\widetilde{R}_P] = 1, \\ & \mathbb{P}(\widetilde{R}_P \geq 0) = 1. \end{aligned} \tag{1}$$

This problem is a special instance of the one solved by Xia and Zhou (2016).[2] Applying Theorem 3.3 of Xia and Zhou (2016), we can derive the optimal portfolio $\widetilde{R}_P^*$ for (1).

2.2. The Three-Moment CAPM

Recall $\widetilde{R}_M$ is the total return of the market portfolio, and denote by $\tilde{r}_M$ and r_f the return rates of the market portfolio and the risk-free asset, respectively. By the market clear condition $\widetilde{R}_M = \widetilde{R}_P^*$, we can derive the equilibrium pricing kernel as well as the CAPM under RDUT as follows.

THEOREM 1. *When the market is in equilibrium with $\widetilde{R}_M$ having a continuous CDF $F_{\widetilde{R}_M}$, the pricing kernel is given by*

$$\tilde{m} = m(\widetilde{R}_M) \equiv \frac{w'(1-F_{\widetilde{R}_M}(\widetilde{R}_M))u'(\widetilde{R}_M)}{(1+r_f)\mathbb{E}\left[w'(1-F_{\widetilde{R}_M}(\widetilde{R}_M))u'(\widetilde{R}_M)\right]}. \tag{2}$$

Moreover, if $\widetilde{R}_M$ has a continuously differentiable density function $f_{\widetilde{R}_M}$, then, for any asset i in the market, we have the three-moment CAPM:

$$\mathbb{E}[\tilde{r}_i] = r_f + A\mathrm{Cov}(\tilde{r}_i, \tilde{r}_M) + \frac{1}{2}B\mathrm{Cov}(\tilde{r}_i, \tilde{r}_M^2) + o(\tilde{r}_M^2), \tag{3}$$

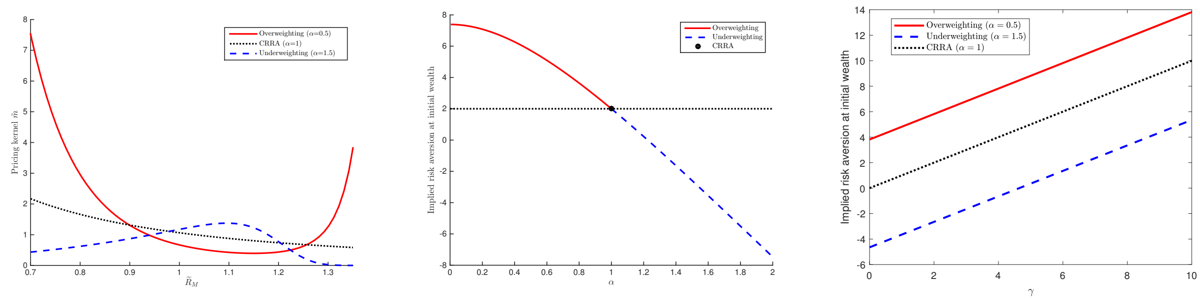
where $\tilde{r}_i$ is the return rate of the asset i, and the risk premia A and B are given as

$$\begin{aligned} A = &-\mathbb{E}^{-1}[w'(1-F_{\widetilde{R}_M}(\widetilde{R}_M))u'(\widetilde{R}_M)] \\ &\times \left[w'(1-F_{\widetilde{R}_M}(1))u''(1) - w''(1-F_{\widetilde{R}_M}(1))u'(1)f_{\widetilde{R}_M}(1)\right], \end{aligned} \tag{4}$$

$$\begin{aligned} B = &-\mathbb{E}^{-1}[w'(1-F_{\widetilde{R}_M}(\widetilde{R}_M))u'(\widetilde{R}_M)] \\ &\times \Big[w'(1-F_{\widetilde{R}_M}(1))u'''(1) + w'''(1-F_{\widetilde{R}_M}(1))u'(1)f^2_{\widetilde{R}_M}(1) \\ &- w''(1-F_{\widetilde{R}_M}(1))[2u''(1)f_{\widetilde{R}_M}(1) + u'(1)f'_{\widetilde{R}_M}(1)]\Big]. \end{aligned} \tag{5}$$

A notable feature of this result is that the equilibrium pricing kernel $\tilde{m}$ (indicated by expression (2)) is not only influenced by the utility function u, but also by the probability weighting function w *evaluated at* $1 - F_{\widetilde{R}_M}(\widetilde{R}_M)$ and, *consequently*, by the market return distribution. Figure 2(a) demonstrates how all three work together in determining pricing kernel. There, we consider Prelec's probability weighting function and the CRRA utility function $u(x) = \frac{1}{1-\gamma}x^{1-\gamma}$ with $\gamma = 2$, and take $\tilde{r}_M$ to be skew-normal with mean 7.6%, standard deviation 15.8% and skewness -0.339.[3] This specification can be easily verified to satisfy the condition $1 - F_{\widetilde{R}_M}(1) > e^{-1}$.

Figure 2 Pricing kernels and aggregate risk aversions



(a) Theoretical pricing kernel (b) Aggregate risk aversion vs α (c) Aggregate risk aversion vs γ

Figures 2(a) plots the pricing kernel against the market return. It shows that the RDUT agent overweighting both tails pays higher prices in the extreme states, resulting in a U-shaped pricing kernel.[4] The case of underweighting both tails has the opposite characteristics, with a bell-shaped pricing kernel. Such U-shaped or bell-shaped patterns driven by probability weighting will be reexamined empirically in Section 3.1.

Figures 2(b) depicts the *aggregate* relative risk aversion, $-m'(\widetilde{R}_M)\cdot\widetilde{R}_M/m(\widetilde{R}_M)$, at $\widetilde{R}_M = 1$, as a function of α. The classical CRRA case (the black dot at $\alpha = 1$) has a flat relative risk aversion, as expected. The solid red and dashed blue lines show how probability weighting changes risk attitudes by influencing investors' perceptions of tail risk. As shown in Figures 2(b), the aggregate risk aversion is a decreasing function of α. When $\alpha < 1$, overweighting both tails enhances the aggregate risk aversion, and hence demands a higher risk compensation. When $\alpha > 1$, underweighting both tails has the opposite effect. Indeed, the aggregate risk aversion is negative when one sufficiently underweights, yielding a risk-seeking behavior. Figures 2(c) shows the aggregate risk aversion as a function of γ. As expected, all the functions are increasing irrespective of the value of the probability weighting parameter.

2.3. Risk Premia, Risk Aversion and Probability Weighting

In the three-moment CAPM model, the terms A and B are the risk premia of the covariance and the coskewness, respectively. The signs of the two parameters determine the risk preference with respect to covariance and coskewness. As A and B depend on the utility function, the weighting function, and the distribution of the market return in a complex way, it is difficult to analytically study their signs *in general*. However, if we choose specific preference functions, say, Prelec's weighting function and CRRA utility function

$$u(x) = \frac{x^{1-\gamma}}{1-\gamma}, \qquad w(x) = \exp(-(-\log(p))^{\alpha}), \tag{6}$$

then analytical characterizations of the signs becomes possible.

Let us examine first the sign of A. From its expression on the right hand side of (4), it follows that the first term (including the factor in front of the brackets) is positive, representing the part of the risk aversion arising from the outcome utility function u. The second term is more complicated, representing the part of the risk aversion arising from the probability weighting, whose sign depends on whether the quantity $1 - F_{\widetilde{R}_M}(1)$ lies in the convex or concave domain of the weighting function w or equivalently whether the tails are overweighted or underweighted.

PROPOSITION 1. *Assume that the market is in equilibrium with $\widetilde{R}_M$ having a density function $f_{\widetilde{R}_M}$. Let $\alpha = \alpha(\gamma)$ be the unique solution of the algebraic equation*

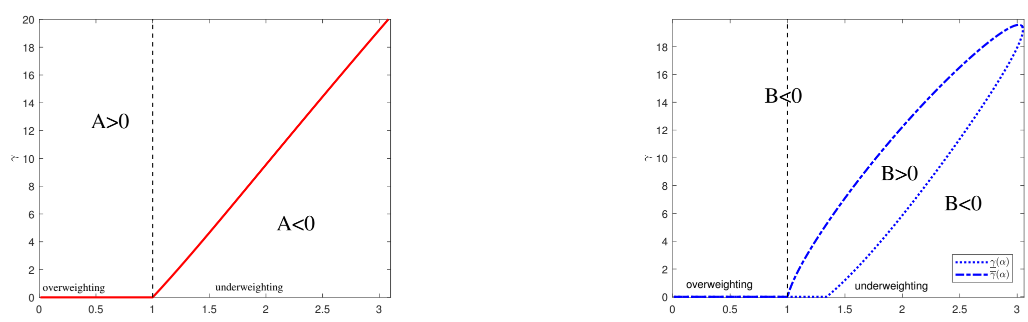
$$\gamma + (\alpha y_0^{\alpha} - \alpha + 1 - y_0) e^{y_0} y_0^{-1} f_{\widetilde{R}_M}(1) = 0$$

where $y_0 = -\log(1 - F_{\widetilde{R}_M}(1))$. Then $\alpha(\gamma) \geq 1$, $\forall \gamma \geq 0$, and

(i) *when $0 < \alpha < 1$ (overweighting), we have $A > 0$;*

(ii) *when $\alpha > 1$ (underweighting), we have $A > 0$ $(A < 0)$ if and only if $\alpha < \alpha(\gamma)$ $(\alpha > \alpha(\gamma))$.*

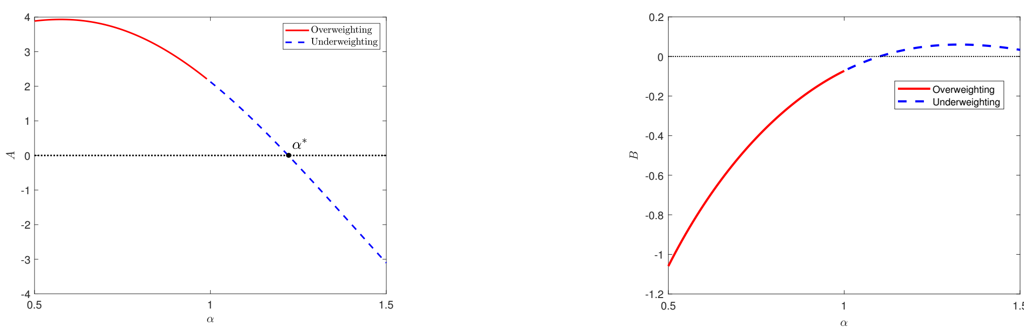
Figure 3 Theoretical signs of risk premia A and B



(a) Theoretical sign of risk premium A (b) Theoretical sign of risk premium B

We illustrate Proposition 1 by Figure 3(a) where the (α, γ)-plane is separated into two regions according to the signs of A. The red line in Figure 3(a) corresponds to the threshold value $\alpha(\gamma)$. Let us consider first the case $0 < \alpha < 1$ (w is inverse-S-shaped), in which the agent overweights both extremely good and extremely bad states. Overweighting the left tail makes the RDUT agent more risk averse toward bad states than her CRRA counterpart, and overweighting the right tail makes her less risk averse toward good states. While these two effects seem to offset each other, the condition $1 - F_{\widetilde{R}_M}(1) > e^{-1}$, which has been shown to be empirically satisfied by the S&P 500, dictates that only the former effect is *relevant*.[5] This results in a positive risk premium A; see the solid red line in Figure 4(a).

Figure 4 Theoretical relationship between risk premia and probability weighting parameter α



(a) Risk premium A against α (b) Risk premium B against α

Notes: The probability weighting function is that of Prelec with parameter α and the utility is CRRA with relative risk aversion $\gamma = 2$. The market return follows a skew-normal distribution with mean 7.6%, standard deviation 15.8%, and skewness -0.339.

There is more to it. Write

$$A = \frac{w'(1 - F_{\widetilde{R}_M}(1))u'(1)}{\mathbb{E}[w'(1 - F_{\widetilde{R}_M}(\widetilde{R}_M))u'(\widetilde{R}_M)]} \left[-\frac{u''(1)}{u'(1)} + \frac{w''(1 - F_{\widetilde{R}_M}(1))}{w'(1 - F_{\widetilde{R}_M}(1))} f_{\widetilde{R}_M}(1) \right]. \tag{7}$$

The first term inside the brackets, $-\frac{u''(1)}{u'(1)} > 0$, measures the risk aversion arising from the outcome utility function u in the classical EUT framework. The second term, $\frac{w''(1-F_{\widetilde{R}_M}(1))}{w'(1-F_{\widetilde{R}_M}(1))} f_{\widetilde{R}_M}(1)$, which arises from the probability weighting *independent* of the outcome utility function, is also positive when $0 < \alpha < 1$. Thus, the risk premium A is not only positive in this case, but indeed *elevated* by this second term compared with when there is no probability weighting. Such an increase of the level of beta risk premium is also consistent with Figure 4(a).

When $\alpha > 1$, on the other hand, w is S-shaped and the agent underweights the two tails of the market return distribution. This entails less risk aversion toward bad states and less risk seeking toward good states. However, the condition $1 - F_{\widetilde{R}_M}(1) > e^{-1}$ implies that $1 - F_{\widetilde{R}_M}(1)$ is now in the *concave* domain of w and hence only the former behavior is relevant. In this case, the agent demands less compensation for the beta risk compared with a classical utility maximizer and, indeed, may even become willing to pay for the risk (i.e., $A < 0$) when she *sufficiently* underweights bad states (i.e., when α is sufficiently large). This case is illustrated by the dashed blue line in Figure 4(a). Moreover, Figure 4(a) reveals that the risk premium A decreases in α, a theoretical prediction that will be confirmed empirically in Section 4.4.

It is interesting to note that Quiggin (2012) defines the term $-\frac{w''(p)}{w'(p)}$ as *probabilistic risk aversion*, analogous to the Arrow–Pratt absolute risk aversion measure. However, unlike the latter which always takes positive values for a concave outcome utility function, the probabilistic risk aversion can take both positive and negative values depending on where it is evaluated for an S-shaped or inverse-S-shaped weighting function. The formula (7) shows that when the probabilistic risk aversion is positive (negative) at the point of market break-even then overweighting (underweighting) of the left tail dominates.

Next we investigate the sign of B. First, when there is no probability weighting, we have $B = -\mathbb{E}^{-1}[u'(\widetilde{R}_M)]u'''(1) < 0$ assuming that $u''' > 0$, which is satisfied by most commonly used utility functions.[6] This implies that a typical EUT agent is willing to pay for the coskewness; see also Harvey and Siddique (2000). In the presence of probability weighting, the first term of B (see (5)) is negative if, again, $u''' > 0$. The second and third terms are more complicated, in that their signs depend on the utility function, the weighting function, and the market return distribution, all intertwined. However, as in the case of A, whether $1 - F_{\widetilde{R}_M}(1)$ lies in the convex or concave domain of w is critical for the sign of B. Here, we provide the necessary and sufficient conditions for B to be positive or negative.

Proposition 2. *Assume that the market is in equilibrium with $\widetilde{R}_M$ having a continuously differentiable density function $f_{\widetilde{R}_M}$. Define*

$$\underline{\gamma}(\alpha) = \begin{cases} \max(0.5[(-1 - 2g_1(\alpha) f_{\widetilde{R}_M}(1)) - \sqrt{\Delta}], 0), & \text{if } \Delta \geq 0, \\ 0, & \text{if } \Delta < 0, \end{cases}$$

$$\overline{\gamma}(\alpha)=\begin{cases}\max(0.5[(-1-2g_1(\alpha)f_{\widetilde{R}_M}(1))+\sqrt{\Delta}],0), & \text{if } \Delta\geq 0,\\ 0, & \text{if } \Delta<0,\end{cases}$$

where $\Delta=(1+2g_1(\alpha)f_{\widetilde{R}_M}(1))^2-4[-g_1(\alpha)f'_{\widetilde{R}_M}(1)+g_2(\alpha)f^2_{\widetilde{R}_M}(1)]$, $g_1(\alpha)=\frac{\alpha y_0^\alpha-\alpha+1-y_0}{e^{-y_0}y_0}$, $g_2(\alpha)=\frac{(\alpha y_0^\alpha-\alpha+1-y_0)^2+(1-\alpha)+(1-\alpha-y_0)(\alpha y_0^\alpha-y_0)}{e^{-2y_0}y_0^2}$, *and* $y_0=-\log(1-F_{\widetilde{R}_M}(1))$. *Then* $B<0$ *(*$B>0$*) if and only if* $\gamma\in[0,\underline{\gamma}(\alpha))\cup(\overline{\gamma}(\alpha),+\infty)$ *(*$\gamma\in(\underline{\gamma}(\alpha),\overline{\gamma}(\alpha))$*).*

The "if and only if conditions" in Proposition 2 separate the (α,γ)-plane into two regions where the risk premium B has different sign; see Figure 3(b) for visualization. The blue dot line ($\underline{\gamma}(\alpha)$) and blue dot-dash line ($\overline{\gamma}(\alpha)$) separates the sign of B. Note that $\underline{\gamma}(\alpha)=\overline{\gamma}(\alpha)=0$ when $0<\alpha<1$ (overweighting), in which case $B<0$. For $\alpha>1$ (underweighting), $\underline{\gamma}(\alpha)$ and $\overline{\gamma}(\alpha)$ form an elliptic region within which $B>0$; otherwise, $B<0$.

To interpret Proposition 2, we plot the risk premium B against the probability weighting parameter α while fixing the CRRA parameter $\gamma=2$ in Figure 4(b). We have noted earlier that in the absence of probability weighting, $B<0$ for "most commonly used utility functions," suggesting an inherent skewness-inclined preference in the classical EUT framework (as also shown by the negative B for $\alpha=1$ in Figure 4(b)). Now, when the agent overweights both tails of the market return (the case of $0<\alpha<1$), probability weighting is likely to *enhance* the skewness-inclined preference when overweighting occurs; refer to the red solid line in Figure 4(b). This enhancement can be explained intuitively as follows. When the asset has positive coskewness with the market, the asset return has a fat right tail with respect to the market portfolio. The hope of a larger gain relative to the market, which is strengthened by overweighting, further increases the attraction of the risky asset and requires even less compensation when compared to the case of no probability weighting. Symmetrically, when the coskewness is negative, the fear of a larger loss, which is also amplified by the overweighting, reduces the attraction of the risky asset and hence demands greater compensation. Combined, overweighting the tails *inflates* the negative relationship between expected return and coskewness, generating a more significantly negative B.

A similar (and opposite) argument yields that underweighting tails (i.e., $\alpha>1$) *deflates* the negative relationship between expected return and coskewness, which may even lead to an overall *positive* premium B. Whether B is eventually positive or negative depends on which one dominates:

the skewness-inclination inherent in u or the skewness-aversion in w. Proposition 2 *characterizes* when one of these is the case. Moreover, this dominance is not monotone as shown by Figure 3(b): for each fixed γ, as the probability weighting evolves from overweighting to underweighting, the skewness premium changes from negative to positive, and then to negative again. This theoretical predictions will be supported empirically in Section 4.4.

We conclude this section by noting that Propositions 1 and 2 are presented with specific, if commonly used, parametric forms of preference functions – the CRRA utility and Prelec's weighting. The results can be extended to more general functions involving more complex algebraic equations and less transparent assumptions and conclusions.

3. The Conditional CAPM

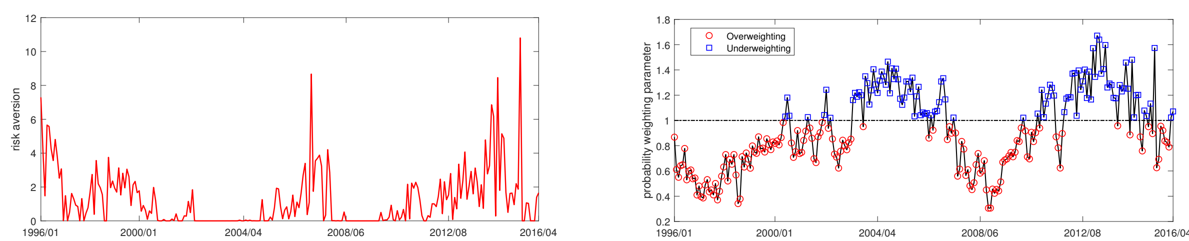
The implications of the theoretical static CAPM derived in the previous section depend crucially on whether agents overweights or underweights the tails. In this section we discuss the time-variation of preference parameters to motivate the formulation of *conditional* CAPM. In Subsection 3.1, we derive the implied preference parameters, including those of the risk aversion and the probability weighting, from option data, using GMM. The result demonstrates a strong time-varying structure including the changes between overweighting and underweighting. Then, in Subsection 3.2, we formulate the conditional CAPM.

3.1. Implied Preference Parameters via GMM

Option data have been used to estimate empirical pricing kernels in the literature (see, e.g., Rosenberg and Engle 2002). Recall that the pricing kernel under RDUT is the product of the marginal utility and the derivative of probability weighting. Hence, by specifying the utility function and the probability weighting function as in the simple one-parameter forms in (6), we can obtain the implied preference parameters, γ and α, from option data by GMM. A detailed GMM procedure is described in Section 4 of the online appendix.

The time series of the estimated risk aversion and probability weighting parameters are shown in Figure 5(a) and Figure 5(b), respectively. We have a couple of important observations. First, around 33% of the implied risk aversion is zero or close to zero, which means *only* the probability weighting plays a role during the corresponding months. Second, among the total 244 months, 143 months are identified as overweighting the tails ($\alpha < 1$) and 101 months underweighting the tails ($\alpha > 1$). So, both α and γ are time-varying and, in particular, Figure 5(b) confirms that investors do switch between overweighting and underweighting. The time variation of α was first documented by Polkovnichenko and Zhao (2013) under a fixed risk aversion parameter. Our result shows that this is prevalent regardless of whether one considers fixed or varying risk aversion. Moreover, Figure A1(a) in the online appendix indicates that the probability weighting index derived here by GMM and the one by Polkovnichenko and Zhao (2013) have a correlation coefficient 0.9646; hence the time variation of probability weighting is a robust property of the data that cannot be accounted for by the variation of the outcome utility function.

Figure 5 Time-varying risk aversion and probability weighting parameters



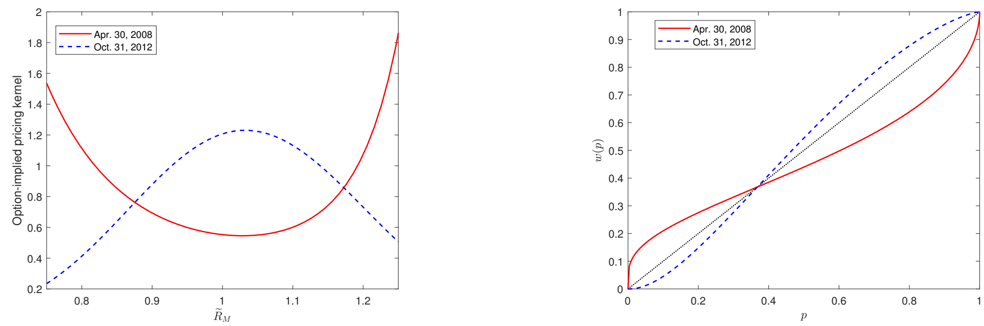
(a) risk aversion parameter γ_t (b) probability weighting parameter α_t

Figure 6 depicts the pricing kernels and the corresponding probability weighting functions in an overweighting period and an underweighting one. On Apr. 30, 2008 – during an overweighting period – the pricing kernel has a U-shape and the probability weighting function has an inverse S-shape. During an underweighting period on Oct. 31, 2012, the pricing kernel has a bell-shape and the weighting function has an S-shape. These empirical observations match our theoretical predictions demonstrated in Figure 2(a).

Note that for Prelec's weighting function, $0 < \alpha < 1$ corresponds to inverse S-shaped (overweighting) and $\alpha > 1$ to S-shaped (underweighting), with $\alpha = 1$ dividing the two. The smaller (greater) α, the greater the degree of overweighting (underweighting). This suggests that we can use the following index, PWI, to quantify the level of probability weighting:

$$PWI = \frac{1}{\alpha}. \tag{8}$$

Figure 6 Two typical option-implied pricing kernels and probability weighting functions



(a) Two typical pricing kernels (b) Two typical probability weighting functions

3.2. The Conditional CAPM

The previous subsections show that the preference parameters change over time, and hence so do the beta and coskewness risk premia. To capture such a dynamic structure, we assume that the three-moment CAPM holds in a conditional sense.

In a dynamic economy one needs not only to manage the types of risk that present in a static economy but also to anticipate and hedge against the possibility of adverse changes in both investment opportunities and risk preferences in the future. Complete characterizations of these hedging needs call for a full theoretical equilibrium model for a *dynamic* RDUT economy, which remains a significant and challenging open question, especially because of the time-inconsistency inherent with the probability weighting. However, following Merton (1980) and Jagannathan and Wang (1996), we assume that the hedging motives are not sufficiently important, and hence the conditional three-moment CAPM will hold in the following way:

$$\mathbb{E}_t(\tilde{r}_{i,t+1}) = r_{f,t+1} + A_t \mathrm{Cov}_t(\tilde{r}_{i,t+1}, \tilde{r}_{M,t+1}) + \frac{1}{2} B_t \mathrm{Cov}_t(\tilde{r}_{i,t+1}, \tilde{r}^2_{M,t+1}). \tag{9}$$

In the above, the subscript t indicates the relevant time period; so, for instance, $\tilde{r}_{i,t+1}$ denotes the return rate of asset i in period $t+1$, $\mathbb{E}_t$ the expectation conditional on the information up to the end of period t, Cov_t the corresponding conditional covariance, and A_t and B_t the resulting risk premia that are influenced by preferences (including outcome utility and probability weighting) as shown by (4) and (5).

It should be noted that our conditional CAPM model takes variation of risk premia as given which is then connected to variation of preferences. The latter is likely to be a manifestation of some more fundamental forces that the model in itself does not identify. The identification and study of those forces once again requires a dynamic model and remains an outstanding open question.

4. Empirical Analysis

The empirical analysis in this section tests our preceding theory regarding how probability weighting impacts the risk premia on covariance and coskewness. Subsections 4.1 contains the traditional one-way sort portfolio analysis to study the effect of probability weighting on the mean-covariance and mean-coskewness relationships, respectively. In Subsection 4.2 and 4.3, we run Fama-MacBeth regression to investigate the impact of probability weighting on cross-sectional asset pricing with and without control variables. In Subsection 4.4, we further test the structural dependence of the risk premia on the preference parameters.

4.1. One-way Sort Portfolio Analysis

We now conduct one-way sort portfolio empirical analysis to test the mean-covariance and mean-coskewness relationships during different probability weighting periods. In our study, we use all the common stocks (share codes 10 and 11) listed on the NYSE, AMEX, and NASDAQ to perform cross-sectional analysis. Data on prices, returns, and shares outstanding are obtained from CRSP monthly files, which range from January 1991 to May 2016.

At the end of each month t from January 1996 to April 2016, we use the immediate prior 60-month sample covariance and sample coskewness to estimate the pre-ranking covariance and coskewness in month t,

$$\text{covar}_{i,t} = \frac{1}{59}\sum_{j=0}^{59}(\tilde{r}_{i,t-j} - \bar{r}_i)(\tilde{r}_{M,t-j} - \bar{r}_M), \quad \text{coskew}_{i,t} = \frac{1}{59}\sum_{j=0}^{59}(\tilde{r}_{i,t-j} - \bar{r}_i)(\tilde{r}^2_{M,t-j} - \overline{r^2_M}), \tag{10}$$

where $\tilde{r}_{i,t-j}$ is the monthly return rate of stock i in month $t-j$, $\tilde{r}_{M,t-j}$ is the monthly return rate of the value-weighted market portfolio in month $t-j$, and $\bar{r}_i$, $\bar{r}_M$ and $\overline{r^2_M}$ are the sample means of $\tilde{r}_{i,t}$, $\tilde{r}_{M,t}$, and $\tilde{r}^2_{M,t}$, respectively.

At each t, we classify stocks into 10 covariance-sorted groups based on their pre-ranking covariance. Group 1 (10) contains stocks with the lowest (highest) covariance. For each group, we build a value-weighted portfolio.[7] Then, we construct a spread portfolio (10-1) that longs Portfolio 10 and shorts Portfolio 1. We then calculate the next month $(t+1)$'s post-ranking portfolio returns for these portfolios and repeat the procedure for the next month. Similarly, we construct the coskewness-sorted portfolios.

Table 1 reports the average next month's returns of covariance-sorted portfolios and coskewness-sorted portfolios for the entire period (January 1996 to April 2016), overweighting periods and underweighting periods, respectively, where the overweighting and underweighting periods are determined by the index PWI defined in (8). In order to balance the two sample sizes, in what follows we use the median of PWI, instead of 1, to separate the overweighting and underweighting months. We observe that the spread covariance-sorted portfolio (10-1) has a significantly positive monthly return during the overweighting periods and a significantly negative monthly return during the underweighting periods. Meanwhile, the spread coskewness-sorted portfolio (10-1) has a significantly negative monthly return during the overweighting periods and an insignificantly negative monthly return during the underweighting periods. Moreover, the negative return of this spread portfolio doubles its size from an average of -0.79% monthly over the entire period to an average of -1.62% monthly during the overweighting periods.

Table 1 One-way sort results based on PWI

Panel A: average returns of covariance-sorted value-weighted portfolios

	1	2	3	4	5	6	7	8	9	10	(10-1)	t-statistic
All	0.89%	1.07%	1.11%	1.09%	1.19%	1.12%	1.06%	1.12%	1.04%	0.99%	0.10%	(0.18)
Over	0.73%	0.96%	1.02%	1.03%	1.26%	1.38%	1.42%	1.84%	2.15%	2.68%	**1.95%****	(2.09)
Under	1.06%	1.19%	1.20%	1.15%	1.12%	0.84%	0.71%	0.40%	-0.08%	-0.69%	**-1.75%*****	(-2.85)

Panel B: average returns of coskewness-sorted value-weighted portfolios

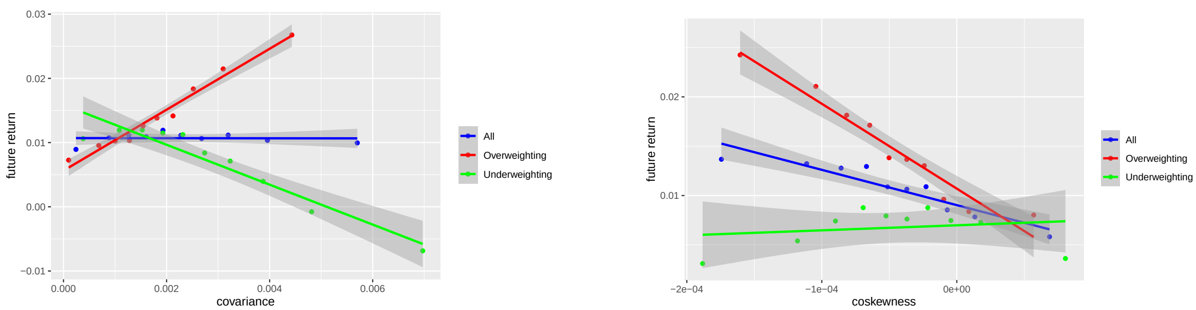
	1	2	3	4	5	6	7	8	9	10	(10-1)	t-statistic
All	1.37%	1.32%	1.28%	1.29%	1.09%	1.07%	1.09%	0.85%	0.78%	0.58%	**-0.79%****	(-2.36)
Over	2.42%	2.11%	1.81%	1.71%	1.38%	1.37%	1.30%	0.96%	0.84%	0.80%	**-1.62%*****	(-2.71)
Under	0.31%	0.54%	0.74%	0.88%	0.79%	0.76%	0.88%	0.75%	0.73%	0.36%	0.05%	(0.18)

Notes: The overweighting periods and underweighting periods are separated by the median of PWI. All periods (All), overweighting periods (Over), and underweighting periods (Under) include 244 months, 122 months, and 122 months, respectively. The spread portfolio (10-1) is constructed by longing the 10th portfolio and shorting the 1st portfolio. The t-statistics of the returns of the spread portfolios are reported. ***, ** , and * denote significance at the 1%, 5%, and 10% levels, respectively.

Figure 7(a) and Figure 7(b) plot the regression lines of the average next month's returns of value-weighted portfolios on the average covariance and on the average coskewness, respectively.

The return–covariance line has a steep upward (downward) slope during the overweighting (underweighting) periods. The return–coskewness line has a steep downward slope during the overweighting periods and a nearly flat slope during the underweighting periods. These findings are consistent with our theoretical predictions in Section 2.3. Finally, for the entire sample period, the return–covariance line is almost flat, reconciling with the beta anomaly, and the return–coskewness line is downward-sloping, conforming to the skewness-inclined preference that is well documented in the literature.

Figure 7 The relationship between average expected return and average covariance (coskewness)



(a) Covariance-sorted portfolios (b) Coskewness-sorted portfolios

4.2. Fama-MacBeth Cross-Sectional Regression

We conduct a Fama-MacBeth regression to investigate the impact of probability weighting on cross-sectional asset pricing. Following Bali et al. (2016), we first compute the covariance and coskewness of stock i for each month t from January 1996 to April 2016, by means of a 60-months-window ordinary least squares (OLS) regression:

$$\tilde{r}_{i,\tau} = k_{i,t} + \text{covariance}_{i,t}\tilde{r}_{M,\tau} + \text{coskewness}_{i,t}\tilde{r}^2_{M,\tau} + \epsilon_{i,\tau}, \quad \tau = t-59, \cdots, t, \tag{11}$$

where $k_{i,t}$ is the intercept and $\epsilon_{i,\tau}$ is the residual.[8] In contrast to computing the covariance and coskewness from the samples directly, as in Subsection 4.1, the above regression rules out interactions between covariance and coskewness. Next, for each t, we consider the following cross-sectional regression:

$$\tilde{r}_{i,t+1} = k_t + A_t\text{covariance}_{i,t} + \frac{1}{2}B_t\text{coskewness}_{i,t} + \varepsilon_{i,t+1}, \quad i = 1, \cdots, n. \tag{12}$$

The significance of the risk premia A and B is determined by a t-test on the time series $\{A_t\}$ and $\{B_t\}$.

Table 2 Fama-MacBeth regression results

Panel A: Fama-MacBeth results without control variables

	All periods	Overweighting periods	Underweighting periods
A	1.51	**11.17*** **	**-8.24*** **
	(0.65)	(2.72)	(-3.47)
B	**-0.88*** **	**-1.74*** **	0.01
	(-2.87)	(-2.98)	(0.06)
Adj. R^2	0.02	0.02	0.02
n	2695	2799	2590

Panel B: Fama-MacBeth results with control variables

	All periods			Overweighting periods			Underweighting periods		
A	1.24	2.30	1.73	**8.36*** **	**9.40*** **	**7.76*** **	**-5.87*** **	**-4.79*** **	**-4.31** **
	(0.65)	(1.20)	(1.03)	(2.58)	(2.88)	(2.81)	(-3.16)	(-2.67)	(-2.52)
B	**-0.38*** **	**-0.33*** **	**-0.24** **	**-1.37*** **	**-1.25*** **	**-0.88*** **	-0.14	-0.07	-0.06
	(-2.94)	(-2.63)	(-2.47)	(-2.87)	(-2.67)	(-2.61)	(-0.83)	(-0.44)	(-0.37)
$IVol$	0.01	-0.06	-0.05	0.10	0.01	0.01	**-0.08* **	**-0.13*** **	**-0.12*** **
	(0.15)	(-1.34)	(-1.35)	(0.99)	(0.13)	(0.19)	(-1.66)	(-2.91)	(-2.82)
$ISkew$	-0.74	**-1.04* **	-0.76	**-2.57** **	**-2.67** **	**-1.94** **	0.10	0.58	0.43
	(-1.19)	(-1.68)	(-1.42)	(-2.42)	(-2.53)	(-2.18)	(1.62)	(0.98)	(0.75)
$Size$		**-1.66*** **	**-1.43*** **		**-2.06** **	**-1.45* **		**-1.25*** **	**-1.42*** **
		(-3.27)	(-3.23)		(-2.27)	(-1.86)		(-2.76)	(-3.29)
BM		**1.68** **	**1.93*** **		1.02	1.37		**2.35*** **	**2.51*** **
		(2.35)	(2.71)		(0.88)	(1.16)		(2.82)	(3.10)
Mom			0.01			-0.03			**0.04** **
			(0.44)			(-0.72)			(2.49)
Rev			**-0.34*** **			**-0.53**			**-0.15**
			(-6.38)			(-5.96)			(-2.81)
Adj. R^2	0.03	0.04	0.05	0.03	0.04	0.06	0.02	0.03	0.04
n	2693	2691	2689	2796	2794	2792	2587	2585	2583

Notes: "Underweighting periods" and "Overweighting periods" are separated by the respective median for PWI. The average values of the coefficients, which are multiplied by 1000, are reported above and the corresponding t-statistics are reported below in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. The row labelled "Adj. R^2" reports the average adjusted R-square of each regression. The row labelled "n" reports the average number of stocks in each cross-sectional regression.

Panel A in Table 2 reports the Fama-MacBeth regression results, in which the market overweighting and underweighting periods are divided by the GMM-derived probability weighting index PWI. If we do not separate the time periods based on probability weighting, the overall risk premium of covariance, A, is not significant and the risk premium of coskewness, B, is modestly negative. These are consistent with the empirical findings in the literature. When we *do* separate the whole period into two regimes (overweighting and underweighting), however, A becomes significantly positive during overweighting periods and significantly negative during underweighting periods, and B almost doubles its overall value during overweighting periods yet becomes insignificant during underweighting periods. Again, these results reaffirm our theoretical and prior one-way and two-way sort portfolios analysis findings.

Finally, we note that during overweighting periods, the mean-covariance tradeoff and mean-coskewness tradeoff are also economically significant. Indeed, the standard deviation of covariance is 0.864 and that of coskewness is 10.591. Thus, during an overweighting period, a one-standard-deviation increase in covariance is associated with a roughly 0.97% (the std of covariance times A, i.e., $0.864 \times 11.17 \times 10^{-3}$) increase in expected monthly return, and a one-standard-deviation decrease in coskewness is associated with a roughly 0.92% (the std of coskewness times B, i.e., $0.5 \times (-10.591) \times (-1.74) \times 10^{-3}$) increase in expected monthly return. This finding strengthens the notion that probability weighting has important implications in pricing and investment.

4.3. With Common Control Variables

In this subsection, we conduct a Fama-MacBeth regression to test the robustness of our main findings after controlling six other common variables. These control variables are Size ($Size$), Book-to-market ratio (BM), Momentum (Mom), Short-term reversal (Rev), Idiosyncratic volatility ($IVol$), and Idiosyncratic skewness ($ISkew$). Among them, the first four are well-studied factors long known to be relevant in asset pricing, and the last two are relatively new in the literature. Liu et al. (2018) argue that the beta anomaly arises from beta's positive correlation with idiosyncratic volatility ($IVol$). Barberis and Huang (2008) find that CPT investors are willing to pay more for lottery-like stocks due to probability weighting. An et al. (2020) observe that lottery-like stocks significantly underperform their non-lottery-like counterparts, especially among stocks where investors have lost money.

Following Bali et al. (2016), these control variables are computed as follows

$$
\begin{aligned}
&Size_{i,t} = \log\left(\frac{PRC_{i,t} \times SHROUT_{i,t}}{1000}\right), && BM_{i,t} = \log\left(\frac{1000 \times BE_{i,y}}{PRC_{i,y} \times SHROUT_{i,y}}\right),\\
&Mom_{i,t} = 100\left[\prod_{j=1}^{11}(\tilde{r}_{i,t-j}+1)-1\right], && Rev_{i,t} = 100 \times \tilde{r}_{i,t},\\
&IVol_{i,t} = 100 \times \sqrt{\frac{1}{56}\sum_{j=0}^{59}\varepsilon_{i,t-j}^2} \times \sqrt{12}, && ISkew_{i,t} = \frac{\frac{1}{60}\sum_{j=0}^{59}\varepsilon_{i,t-j}^3}{\left(\frac{1}{60}\sum_{j=0}^{59}\varepsilon_{i,t-j}^2\right)^{3/2}},
\end{aligned}
$$

where y is the end of a fiscal year, t is the end of a month between June of year $y+1$, and May of year $y+2$, $PRC_{i,t}$ is the price of stock i at t, $SHROUT_{i,t}$ is the number of shares outstanding of stock i at t, $BE_{i,y}$ is the book value of common equity of stock i at y, $PRC_{i,y}$ is the price of stock

i at y, $SHROUT_{i,y}$ is the number of shares outstanding of stock i at y, $\tilde{r}_{i,t-j}$ is the return rate of stock i during month $t-j$, and $\varepsilon_{i,t-j}$, $j=0,\cdots,59$, are the residuals of the following regression:

$$\tilde{r}_{i,\tau} = k_{i,t} + \delta^1_{i,t} MKT_\tau + \delta^2_{i,t} SMB_\tau + \delta^3_{i,t} HML_\tau + \varepsilon_{i,\tau}, \quad \tau = t-59,\cdots,t,$$

where MKT_τ is the return of the market factor during month τ and SMB_τ and HML_τ are the returns of the size and value factor mimicking portfolios, respectively, during month τ.[9]

The Fama-MacBeth regression results with control variables are reported in Panel B of Table 2. They show that even with the commonly used variables controlled in a cross-sectional analysis, our main empirical finding that the risk premium of covariance (coskewness) is significantly positive (negative) during overweighting periods and significantly negative (insignificant) during underweighting periods still stands.

It is interesting to note that, in Panel B of Table 2, IVol exhibits a significantly negative relationship with expected return only during underweighting periods. In other words, the mispricing (the negative risk premia for beta and idiosyncratic volatility) only appears when investors underweight the tail risk, which echoes the finding in Liu et al. (2018). On the other hand, similar to coskewness, ISkew shows a significantly negative relation with expected return only during overweighting periods. This implies that overweighting both tails leads to RDUT investors developing a strong taste for both coskewness and idiosyncratic skewness.

In addition to controlling the common variables presented here, we also conduct several robustness checks, including the effect of time-varying risk aversion, the effect of stochastic volatility, the result from the UK market, the result based on both ITM and OTM options and the role of standard state variables; see Section 5 of the online appendix for details.

4.4. The Dependence of Risk Premia on PWI and Risk Aversion

Our previous empirical analysis studies risk premia A and B conditional on a binary variable: whether the market is overweighting or underweighting tails. In this subsection, we investigate the time-varying structural relationship between the risk premia and the preference parameters directly using GMM.

In the conditional CAPM model (9), we further assume that the risk premia A_t and B_t depend on γ_t, PWI_t and $\gamma_t PWI_t$ as follows

$$A_t(\vartheta) = a_1\gamma_t + a_2 PWI_t + a_3\gamma_t PWI_t, \quad B_t(\vartheta) = b_1\gamma_t + b_2 PWI_t + b_3\gamma_t PWI_t, \tag{13}$$

where $\vartheta = (a_1, a_2, a_3, b_1, b_2, b_3)$. Here, we add the product term $\gamma_t PWI_t$ to reflect the nonlinearity in the theoretical expressions of A_t and B_t in γ_t and PWI_t; see Theorem 1. Now, the conditional expected return of a risky asset i satisfies

$$\mathbb{E}_t(\tilde{r}_{i,t+1}) = r_{f,t+1} + A_t(\vartheta)\mathrm{Cov}_t(\tilde{r}_{i,t+1}, \tilde{r}_{M,t+1}) + \frac{1}{2}B_t(\vartheta)\mathrm{Cov}_t(\tilde{r}_{i,t+1}, \tilde{r}^2_{M,t+1}),$$

which further yields

$$\begin{aligned} f_i(\vartheta) &\stackrel{\Delta}{=} \mathbb{E}(\mathbb{E}_t(\tilde{r}_{i,t+1})) = \mathbb{E}[r_{f,t+1} + A_t(\vartheta)\mathrm{Cov}_t(\tilde{r}_{i,t+1}, \tilde{r}_{M,t+1}) + \frac{1}{2}B_t(\vartheta)\mathrm{Cov}_t(\tilde{r}_{i,t+1}, \tilde{r}^2_{M,t+1})] \\ &= \mathbb{E}[r_{f,t+1}] + \mathbb{E}[A_t(\vartheta)]\mathbb{E}[\mathrm{Cov}_t(\tilde{r}_{i,t+1}, \tilde{r}_{M,t+1})] + \frac{1}{2}\mathbb{E}[B_t(\vartheta)]\mathbb{E}[\mathrm{Cov}_t(\tilde{r}_{i,t+1}, \tilde{r}^2_{M,t+1})] \\ &\quad + \mathrm{Cov}(A_t(\vartheta), \mathrm{Cov}_t(\tilde{r}_{i,t+1}, \tilde{r}_{M,t+1})) + \frac{1}{2}\mathrm{Cov}(B_t(\vartheta), \mathrm{Cov}_t(\tilde{r}_{i,t+1}, \tilde{r}^2_{M,t+1})). \end{aligned} \tag{14}$$

We have the time series γ_t, PWI_t, $r_{f,t+1}$, $\mathrm{Cov}_t(\tilde{r}_{i,t+1}, \tilde{r}_{M,t+1})$ and $\mathrm{Cov}_t(\tilde{r}_{i,t+1}, \tilde{r}^2_{M,t+1})$ for 244 months from Jan. 1996 to Apr. 2016. Each of the expectations on the right hand side of (14) is computed as the corresponding average over these 244 samples.

Define the following set of moment conditions

$$\tilde{g}(\vartheta) = \left[f_1(\vartheta) - \frac{1}{244}\sum_{t=1}^{244}\tilde{r}^{real}_{1,t}, \cdots, f_n(\vartheta) - \frac{1}{244}\sum_{t=1}^{244}\tilde{r}^{real}_{n,t} \right],$$

where $\tilde{r}^{real}_{i,t}$ is the observed return of risky asset i in month t and n is the number of risky assets. The estimated parameters $\hat{\vartheta}$ are then obtained by solving

$$\hat{\vartheta} = \arg\min_{\vartheta \in \mathbb{R}^6} \quad \tilde{g}(\vartheta) W \tilde{g}(\vartheta)^T,$$

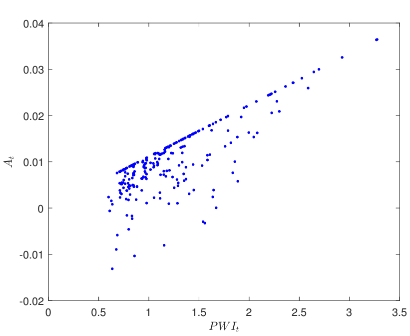
where, again, W is chosen to be the identity matrix. According to Hansen (1982), the estimated parameter vector $\hat{\vartheta}$ follows $\hat{\vartheta} \sim N(\vartheta, V/244)$, where $V = [GWG^T]^{-1}GWSWG^T[GWG^T]^{-1}$, S is the covariance matrix of the n risky assets, and $G = \frac{\partial \tilde{g}(\vartheta)}{\partial \vartheta}$ is an $6 \times n$ matrix.

We use 1412 stocks, including 890 stocks whose monthly return data are available for the whole period and 522 stocks with numbers of missing monthly return data less than 36, to estimate the parameters. Table 3 reports the results. The estimated coefficients, which are multiplied by 1000, are reported above and the corresponding t-statistics are reported below in parentheses. We also draw the corresponding scatter diagrams of the time-varying risk premia A_t and B_t against PWI_t respectively in Figure 8. Table 3 and Figure 8 show clearly that the probability weighting index PWI_t has a statistically significant positive influence on the risk premium of covariance A, and a statistically significant negative influence on the risk premium of coskewness B, consistent with our theoretical prediction as illustrated in Figure 4. On the other hand, although the signs of influence of γ_t on A and B are the same as those of PWI_t, the levels of influence are statistically insignificant.

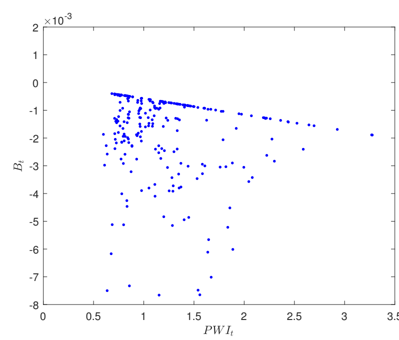
Table 3 The dependence of risk premia on preference parameters

Preference parameters	Coefficients of A_t	Coefficients of B_t
γ_t	-0.65	-0.29
	(-0.21)	(0.72)
PWI_t	**11.12*****	**-0.58*****
	(5.74)	(-2.58)
$\gamma_t * PWI_t$	-1.97	0.46
	(-0.69)	(1.45)

Figure 8 The relationship between risk premia and preference parameters



(a) A_t against PWI_t



(b) B_t against PWI_t

5. Conclusions

The Nobel-prize-winning CAPM theory takes an asset's beta - the correlation with the market - as the risk measure and predicts that high-beta assets earn higher average returns than low-beta ones. However, extensive empirical studies do not support this prediction - a puzzle named the "beta anomaly". An underlying yet unstated assumption of the classical CAPM is that individuals are able to evaluate probabilities objectively. However, behavioral economics study finds that people tend to "distort" or "weight" probabilities of rare events, and this probability weighting plays a crucial role in dictating risk preference and decision making. For example, buying insurance or lottery tickets reflect exaggeration of small probabilities of very "bad" or "good" events. Now, if an investor inflates the small probability of significant downside of an asset then she will demand more return compensation than the CAPM predicts. Likewise, if she deflates that small probability then she will be happy with a smaller average return. Our thesis is then that probability weighting can potentially explain the beta anomaly, at least partly.

The purpose of this paper is to test this thesis, both theoretically and empirically. We show that probability weighting is indeed a key driver in beta and coskewness pricing. Whether overweighting or underweighting is in force changes dynamically over time, but the total length of overweighting periods is comparable to that of underweighting periods. If we analyze a sufficiently long sample period, the aggregate effects of overweighting and underweighting may cancel out, potentially deceiving us into overlooking the significance of probability weighting. It is, therefore, important to *separate out* the overweighting and underweighting periods and to study them individually and respectively. In doing so, as demonstrated in this paper, we are able to understand the significant role that probability weighting plays in pricing and in building profitable portfolios.

We carry out the theoretical analysis using a static RDUT model and extend it to a conditional CAPM to account for the time-variation in both risk aversion and probability weighting. We then confirm the theoretical predictions with extensive empirical tests. We believe that our model offers a partial explanation of the beta anomaly and hints a viable way to potentially resolve other puzzles in asset pricing.

Finally, the results of this paper may motivate new research problems. One of them is why individuals sometimes overweight tail events and sometimes underweight them? The conditional CAPM approach in this paper has helped identify variation in preferences and connect it to expected return, but the source of this variation remains hidden and a challenge to understand. Is there a model built upon psychological foundation that can explain the switch between overweighting and underweighting? Another outstanding problem is a full dynamic CAPM taking into consideration the inherent time-inconsistency arising from probability weighting. The answers to these questions are beyond the scope of this paper, which we hope to see in future research.

Endnotes

1. Another prominent theory that extends EUT to include probability weighting is the cumulative prospect theory (CPT; Tversky and Kahneman 1992). Based on CPT, Barberis et al. (2016) construct a TK factor and find its predicting power derives mainly from probability weighting. Hence, we focus on RDUT in this paper.

2. In Problem (2.2) of Xia and Zhou (2016), setting $I = 1$, $u_{0i}(\cdot) = 0$, $u_{1i}(\cdot) = u(\cdot)$, $w_i(\cdot) = w(\cdot)$, $\beta_i = 1$, $\mathbb{E}[\tilde{\rho}\tilde{e}_{1i}] = 1$, $e_{0i} = 0$, and $\tilde{c}_{1i} = \widetilde{R}_P$, we recover our Problem (1).

3. The distribution of $\tilde{r}_M$ is estimated based on the annual historical data of the S&P 500 from January 1946 to January 2009, which is reported in De Giorgi and Legg (2012).

4. U-shaped pricing kernels have been presented in literature under various settings. For example, Shefrin (2008) finds that the pricing kernel has a U-shape when agents have heterogeneous risk tolerance. Baele et al. (2019) observe that pricing kernels under the cumulative prospect theory (CPT) are also U-shaped.

5. This observation is straightforward from the mathematical expression of A, (4), which depends on w only through its value at $1 - F_{\widetilde{R}_M}(1)$, a point in the *convex* domain of w.

6. Brockett and Golden (1987) refer to the class of increasing utility functions with derivatives that alternate in sign as the class that contains "all commonly used utility functions." In particular, $u''' > 0$ implies that the utility function has nonincreasing absolute risk aversion, which is one of the essential properties (also termed risk *prudence*) of a risk-averse individual; see Harvey and Siddique (2000).

7. There are two different approaches to form portfolios: equal-weighted or value-weighted. Our main results, including the one-way and two-way portfolio analysis, remain unchanged for the equal-weighted portfolios.

8. Besides Bali et al. (2016)'s approach, Harvey and Siddique (2000) propose to compute covariance$_{i,t}$ and coskewness$_{i,t}$ by $\tilde{r}_{i,\tau} = k_{i,t} + \text{covariance}_{i,t}\tilde{r}_{M,\tau} + \epsilon_{i,\tau}, \tau = t - 59, \cdots, t$, and $\text{coskewness}_{i,t} = \frac{\frac{1}{59}\sum_{j=0}^{59}(\epsilon_{i,t-j})(\tilde{r}^2_{M,t-j} - \overline{r^2_M})}{\sqrt{\frac{1}{60}\sum_{j=0}^{59}\epsilon^2_{i,t-j}} \cdot \frac{1}{59}\sum_{j=0}^{59}(\tilde{r}_{M,t-j} - \bar{r}_M)^2}$, where $\bar{r}_M$ and $\overline{r^2_M}$ are the sample means of $\tilde{r}_{M,\tau}$ and $\tilde{r}^2_{M,\tau}$, respectively. We obtain the same results based on this method (omitted here due to space limit).

9. The $IVol_{i,t}$ and $ISkew_{i,t}$ are estimated based on monthly data in a rolling window fashion. We have also estimated the $IVol_{i,t}$ and $ISkew_{i,t}$ based on daily data in each month t, and the main results still hold.

Acknowledgments

The authors are grateful to Jianjun Gao, Duan Li, Wei Li, Jinming Xie, Jing Yao, Shushang Zhu, the Area Editor, an anonymous Associate Editor, and three anonymous reviewers for their detailed and constructive comments, which have led to a much improved version of the paper. The authors also thank seminar and conference participants at 2018 INFORMS International Conference, 8th Annual Conference of Financial Engineering and Financial Risk Management Branch of China Operation Research Society, 2018 Annual Conference of Society for Management Science and Engineering of China, Workshop on Finance, Insurance, Probability and Statistics (FIPS 2019), East China Normal University, Shanghai Jiaotong University for their comments. The authors make equal contributions to this paper. This work was supported by National Natural Science Foundation of China [Grants 71971083, 71931004, 71732003, 72171138]; Program for Innovative Research Team of Shanghai University of Finance and Economics [Grant 2020110930]; Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science-Ministry of Education; A start-up grant and the Nie Center for Intelligent Asset Management at Columbia University.

References

Abdellaoui M (2002) A genuine rank-dependent generalization of the von neumann-morgenstern expected utility theorem. *Econometrica* 70(2):717–736.

An L, Wang H, Wang J, Yu J (2020) Lottery-related anomalies: The role of reference-dependent preferences. *Management Science* 66(1):473–501.

Andrei D, Cujean J, Wilson MI (2020) The lost capital asset pricing model. *Available at SSRN 2922598* .

Baele L, Driessen J, Ebert S, Londono JM, Spalt OG (2019) Cumulative prospect theory, option returns, and the variance premium. *The Review of Financial Studies* 32(9):3667–3723.

Baker M, Bradley B, Wurgler J (2011) Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. *Financial Analysts Journal* 67(1):40–54.

Bali TG, Brown SJ, Murray S, Tang Y (2017a) A lottery-demand-based explanation of the beta anomaly. *Journal of Financial and Quantitative Analysis* 52(6):2369–2397.

Bali TG, Engle RF, Murray S (2016) *Empirical asset pricing: The cross section of stock returns* (John Wiley & Sons).

Bali TG, Engle RF, Tang Y (2017b) Dynamic conditional beta is alive and well in the cross-section of daily stock returns. *Management Science* 63(11):3760–3779.

Barberis N (2013) The psychology of tail events: Progress and challenges. *American Economic Review* 103(3):611–16.

Barberis N, Huang M (2008) Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review* 98(5):2066–2100.

Barberis N, Mukherjee A, Wang B (2016) Prospect theory and stock returns: An empirical test. *The Review of Financial Studies* 29(11):3068–3107.

Black F, Jensen MC, Scholes M (1972) *The capital asset pricing model: Some empirical tests*, 9–124 (Praeger Publishers Inc.).

Bollerslev T, Todorov V, Xu L (2015) Tail risk premia and return predictability. *Journal of Financial Economics* 118(1):113–134.

Bordalo P, Gennaioli N, Shleifer A (2012) Salience theory of choice under risk. *The Quarterly Journal of Economics* 127(3):1243–1285.

Brockett PL, Golden LL (1987) A class of utility functions containing all the common utility functions. *Management Science* 33(8):955–964.

De Giorgi EG, Legg S (2012) Dynamic portfolio choice and asset pricing with narrow framing and probability weighting. *Journal of Economic Dynamics and Control* 36(7):951–972.

Fama EF, French KR (1992) The cross-section of expected stock returns. *The Journal of Finance* 47(2):427–465.

Fama EF, MacBeth JD (1973) Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81(3):607–636.

Frazzini A, Pedersen LH (2014) Betting against beta. *Journal of Financial Economics* 111(1):1–25.

Green TC, Hwang BH (2012) Initial public offerings as lotteries: Skewness preference and first-day returns. *Management Science* 58(2):432–444.

Hansen LP (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 1029–1054.

Harrison GW, Humphrey SJ, Verschoor A (2009) Choice under uncertainty: Evidence from ethiopia, india and uganda. *The Economic Journal* 120(543):80–104.

Harvey CR, Siddique A (2000) Conditional skewness in asset pricing tests. *The Journal of Finance* 55(3):1263–1295.

Henrich J, Heine SJ, Norenzayan A (2010) The weirdest people in the world? *Behavioral and Brain Sciences* 33(2-3):61–83.

Hertwig R, Barron G, Weber EU, Erev I (2004) Decisions from experience and the effect of rare events in risky choice. *Psychological Science* 15(8):534–539.

Hong H, Sraer DA (2016) Speculative betas. *The Journal of Finance* 71(5):2095–2144.

Hsu M, Krajbich I, Zhao C, Camerer CF (2009) Neural response to reward anticipation under risk is nonlinear in probabilities. *Journal of Neuroscience* 29(7):2231–2237.

Humphrey SJ, Verschoor A (2004) The probability weighting function: Experimental evidence from uganda, india and ethiopia. *Economics Letters* 84(3):419–425.

Jagannathan R, Wang Z (1996) The conditional capm and the cross-section of expected returns. *The Journal of Finance* 51(1):3–53.

Jin H, Xia J, Zhou XY (2019) Arrow–debreu equilibria for rank-dependent utilities with heterogeneous probability weighting. *Mathematical Finance* 29(3):898–927.

Kelly B, Jiang H (2014) Tail risk and asset prices. *The Review of Financial Studies* 27(10):2841–2871.

Kozhan R, Neuberger A, Schneider P (2013) The skew risk premium in the equity index market. *The Review of Financial Studies* 26(9):2174–2203.

Kraus A, Litzenberger RH (1976) Skewness preference and the valuation of risk assets. *The Journal of Finance* 31(4):1085–1100.

Lintner J (1965) Security prices, risk, and maximal gains from diversification. *The Journal of Finance* 20(4):587–615.

Liu J, Stambaugh RF, Yuan Y (2018) Absolving beta of volatility's effects. *Journal of Financial Economics* 128(1):1–15.

Merton RC (1980) On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8(4):323–361.

Mossin J (1966) Equilibrium in a capital asset market. *Econometrica* 34(4):768–783.

Polkovnichenko V, Zhao F (2013) Probability weighting functions implied in options prices. *Journal of Financial Economics* 107(3):580–609.

Post T, Levy H (2005) Does risk seeking drive stock prices? a stochastic dominance analysis of aggregate investor preferences and beliefs. *Review of Financial Studies* 18:925–953.

Prelec D (1998) The probability weighting function. *Econometrica* 66(3):497–527.

Quiggin J (1982) A theory of anticipated utility. *Journal of Economic Behavior & Organization* 3(4):323–343.

Quiggin J (2012) *Generalized expected utility theory: The rank-dependent model* (Springer Science & Business Media).

Roll R, Ross SA (1994) On the cross-sectional relation between expected returns and betas. *The Journal of Finance* 49(1):101–121.

Rosenberg JV, Engle RF (2002) Empirical pricing kernels. *Journal of Financial Economics* 64(3):341–372.

Schmeidler D (1989) Subjective probability and expected utility without additivity. *Econometrica* 57(3):571–587.

Sharpe WF (1964) Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance* 19(3):425–442.

Shefrin H (2008) *A Behavioral Approach to Asset Pricing* (New York: Elsevier.).

Tanaka T, Camerer CF, Nguyen Q (2010) Risk and time preferences: Linking experimental and household survey data from vietnam. *American Economic Review* 100(1):557–71.

Tversky A, Kahneman D (1992) Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5(4):297–323.

Wu G, Gonzalez R (1996) Curvature of the probability weighting function. *Management Science* 42(12):1676–1690.

Xia J, Zhou XY (2016) Arrow–debreu equilibria for rank-dependent utilities. *Mathematical Finance* 26(3):558–588.

"Beta and Coskewness Pricing: Perspective from Probability Weighting"

Online Appendix

This document contains all the proofs and other supplement materials in the paper "Beta and Coskewness Pricing: Perspective from Probability Weighting".

1 Proofs

1.1 Proof of Theorem 1

Applying Theorem 3.3 of Xia and Zhou (2016), we can solve the optimal portfolio selection problem and obtain the total return of the optimal portfolio as follows

$$\widetilde{R}_P^* = (u')^{-1}\left(\lambda^* \hat{N}'\left(1 - w(F_{\tilde{m}}(\tilde{m}))\right)\right),$$

where the Lagrange multiplier λ^* is determined by $\mathbb{E}\left[\tilde{m}(u')^{-1}\left(\lambda^* \hat{N}'\left(1 - w(F_{\tilde{m}}(\tilde{m}))\right)\right)\right] = 1$, $N(q) = -\int_{\bar{w}^{-1}(q)}^{1-} Q_{\tilde{m}}^-(1-p)dp,\ \ q \in [0,1]$, $\bar{w}(p) = 1 - w(1-p)$, $\hat{N}$ is the concave envelope of N, $\hat{N}'$ the right derivative of $\hat{N}$ and $Q_{\tilde{m}}^-(p) \triangleq \sup\{x \in \mathbb{R} \mid F_{\tilde{m}}(x) < p\},\ p \in (0,1]$, is the lower quantile function of $\tilde{m}$.

When the market is in equilibrium, the total return of the optimal portfolio of the representative investor equals the total return of the market portfolio, i.e., $\widetilde{R}_P^* = \widetilde{R}_M$. Applying Theorem 5.2 of Xia and Zhou (2016), we obtain

$$\tilde{m} = (\lambda^*)^{-1} w'(1 - F_{\widetilde{R}_M}(\widetilde{R}_M))u'(\widetilde{R}_M).$$

However, the pricing formula of the risk-free asset yields

$$1 = \mathbb{E}[\tilde{m}(1 + r_f)] = \mathbb{E}[\tilde{m}](1 + r_f),$$

which in turn implies

$$\lambda^* = (1 + r_f)\mathbb{E}\left[w'(1 - F_{\widetilde{R}_M}(\widetilde{R}_M))u'(\widetilde{R}_M)\right]. \tag{A1}$$

This establishes the equation (2) in the paper.

Next we rewrite

$$\tilde{m} = (\lambda^*)^{-1} w'(1 - F_{\widetilde{R}_M}(1 + \tilde{r}_M))u'(1 + \tilde{r}_M).$$

Applying the Taylor expansion up to the second order of $\tilde{r}_M$, we can express the pricing kernel as follows:[1]

$$\begin{aligned}\tilde{m} =&(\lambda^*)^{-1}w'(1-F_{\widetilde{R}_M}(1))u'(1)\\&+(\lambda^*)^{-1}\left[w'(1-F_{\widetilde{R}_M}(1))u''(1)-w''(1-F_{\widetilde{R}_M}(1))u'(1)f_{\widetilde{R}_M}(1)\right]\tilde{r}_M\\&+\frac{1}{2}(\lambda^*)^{-1}\Big\{w'(1-F_{\widetilde{R}_M}(1))u'''(1)-w''(1-F_{\widetilde{R}_M}(1))[2u''(1)f_{\widetilde{R}_M}(1)+u'(1)f'_{\widetilde{R}_M}(1)]\\&+w'''(1-F_{\widetilde{R}_M}(1))u'(1)f^2_{\widetilde{R}_M}(1)\Big\}\tilde{r}_M^2+o(\tilde{r}_M^2).\end{aligned} \tag{A2}$$

Applying the pricing formula to the asset i, we have

$$\mathbb{E}[\tilde{m}(1+\tilde{r}_i)]=1.$$

Hence,

$$\mathrm{Cov}(\tilde{m},(1+\tilde{r}_i))+\mathbb{E}[1+\tilde{r}_i]\mathbb{E}[\tilde{m}]=1,$$

or

$$\mathbb{E}[\tilde{r}_i]=\frac{1-\mathrm{Cov}(\tilde{m},\tilde{r}_i)}{\mathbb{E}[\tilde{m}]}-1=r_f+A\mathrm{Cov}(\tilde{r}_i,\tilde{r}_M)+\frac{1}{2}B\mathrm{Cov}(\tilde{r}_i,\tilde{r}_M^2)+o(\tilde{r}_M^2),$$

thanks to (A1) and (A2).

1.2 Proof of Proposition 1

i) When $\alpha<1$, w is inverse S-shaped. Hence, $1-F_{\widetilde{R}_M}(1)$ lies within the convex domain of w under the assumptions of the proposition. Therefore, the two terms inside the brackets of the equation (5) in the paper are both negative. This implies $A>0$.[2]

ii) When $\alpha>1$, w is S-shaped, and $1-F_{\widetilde{R}_M}(1)$ lies within the concave domain of w. With $w(p)=\exp(-(-\log(p))^\alpha)$ and $u(x)=\frac{1}{1-\gamma}x^{1-\gamma}$, we have

$$\begin{aligned}A=&-\frac{w'(1-F_{\widetilde{R}_M}(1))}{\mathbb{E}[w'(1-F_{\widetilde{R}_M}(\widetilde{R}_M))u'(\widetilde{R}_M)]}\left[u''(1)-\frac{w''(1-F_{\widetilde{R}_M}(1))}{w'(1-F_{\widetilde{R}_M}(1))}u'(1)f_{\widetilde{R}_M}(1)\right]\\=&\frac{w'(1-F_{\widetilde{R}_M}(1))}{\mathbb{E}[w'(1-F_{\widetilde{R}_M}(\widetilde{R}_M))u'(\widetilde{R}_M)]}\left[\gamma+g_1(\alpha)f_{\widetilde{R}_M}(1)\right],\end{aligned}$$

[1] Here, we Taylor expand $F_{\widetilde{R}_M}(1+\tilde{r}_M)$ around 1, then expand $w'(1-F_{\widetilde{R}_M}(1+\tilde{r}_M))$ around $1-F_{\widetilde{R}_M}(1)$, and expand $u'(1+\tilde{r}_M)$ around 1.

[2] This part of the proof does not depend on the specific forms of w and u; it only requires $1-F_{\widetilde{R}_M}(1)$ to be in the convex domain of w and u to be a concave function.

where $g_1(\alpha) = (\alpha y_0^\alpha - \alpha + 1 - y_0)e^{y_0}y_0^{-1}$, $y_0 = -\log(1 - F_{\widetilde{R}_M}(1))$. As $w'(1 - F_{\widetilde{R}_M}(\widetilde{R}_M)) > 0$ and $u'(\widetilde{R}_M) > 0$, $A > 0$ (< 0) if and only if

$$\gamma + g_1(\alpha)f_{\widetilde{R}_M}(1) > 0 \quad (< 0).$$

Next, as $1 > 1 - F_{\widetilde{R}_M}(1) > e^{-1}$, we have $1 > y_0 > 0$, which implies that

$$\frac{dg_1(\alpha)}{\alpha} = (y_0^\alpha - 1 + \alpha y_0^\alpha \log(y_0))e^{y_0}y_0^{-1} < 0, \quad \text{for} \quad \alpha \geq 1.$$

Since $g_1(\alpha)$ is an decreasing function of α, $\gamma + g_1(1)f_{\widetilde{R}_M}(1) = \gamma \geq 0$ and $\lim_{\alpha\to+\infty} \gamma + g_1(1)f_{\widetilde{R}_M}(1) = -\infty$, there exists a unique $\alpha(\gamma) \geq 1$ such that $\gamma + g_1(\alpha(\gamma))f_{\widetilde{R}_M}(1) = 0$. Furthermore, $\gamma + g_1(\alpha)f_{\widetilde{R}_M}(1) > 0$ (< 0) if and only if $\alpha < \alpha(\gamma)$ $(\alpha > \alpha(\gamma))$.

1.3 Proof of Proposition 2

Under the assumption, we have

$$\begin{aligned} B = &- \frac{w'(1 - F_{\widetilde{R}_M}(1))u'(1)}{\mathbb{E}[w'(1 - F_{\widetilde{R}_M}(\widetilde{R}_M))u'(\widetilde{R}_M)]}\Bigg[\gamma(\gamma+1) - \frac{\alpha y_0^\alpha - \alpha + 1 - y_0}{e^{-y_0}y_0}(-2\gamma f_{\widetilde{R}_M}(1) + f'_{\widetilde{R}_M}(1)) \\ &+ \frac{(\alpha y_0^\alpha - \alpha + 1 - y_0)^2 + (1-\alpha) + (1 - \alpha - y_0)(\alpha y_0^\alpha - y_0)}{e^{-2y_0}y_0^2} f^2_{\widetilde{R}_M}(1)\Bigg] \\ = &- \frac{w'(1 - F_{\widetilde{R}_M}(1))u'(1)}{\mathbb{E}[w'(1 - F_{\widetilde{R}_M}(\widetilde{R}_M))u'(\widetilde{R}_M)]}[\gamma^2 + (1 + 2g_1(\alpha)f_{\widetilde{R}_M}(1))\gamma - g_1(\alpha)f'_{\widetilde{R}_M}(1) + g_2(\alpha)f^2_{\widetilde{R}_M}(1)], \end{aligned}$$

where $y_0 = -\log(1 - F_{\widetilde{R}_M}(1))$.

As $w'(1 - F_{\widetilde{R}_M}(\widetilde{R}_M)) > 0$ and $u'(\widetilde{R}_M) > 0$, $B > 0$ (< 0) if and only if

$$\gamma^2 + (1 + 2g_1(\alpha)f_{\widetilde{R}_M}(1))\gamma - g_1(\alpha)f'_{\widetilde{R}_M}(1) + g_2(\alpha)f^2_{\widetilde{R}_M}(1) < 0 \quad (> 0).$$

The desired conclusion then follows from the standard results regarding signs of a quadratic function.

2 Estimation and Simulation of EGARCH Model

In this subsection, we explain some details about the estimation and simulation of an EGARCH model used for various purposes in this online appendix. At the end of each month (denoted as time 0), we collect the immediately prior 1250 daily S&P 500 index return data. Then, we compute the maximum likelihood estimations of the parameters $(\mu, \nu, \beta, \zeta, \varkappa)$ in the following EGARCH(1,1) model for daily return:

$$\begin{aligned} \log(S_{M,t}) &= \log(S_{M,t-1}) + \mu + \sqrt{h_t} z_t, \\ \log(h_t) &= \nu + \beta \log(h_{t-1}) + \zeta(|z_{t-1}| - E[|z_{t-1}|]) + \varkappa z_{t-1}, \end{aligned} \tag{A3}$$

where z_t follows a standard Gaussian distribution.

For a given time horizon $[0, T_j]$, we already have the presample variance h_0, presample innovation z_0 and presample index price $S_{M,0}$. To generate a sample path of $\{S_{M,1}, \cdots, S_{M,T_j}\}$, we compute the conditional variance h_1 based on the second equation in (A3), generate z_1 from a standard Gaussian distribution, and then compute the realization of $S_{M,1}$ according to the first equation in (A3). Repeating the same procedure, we can simulate the whole path $\{S_{M,1}, \cdots, S_{M,T_j}\}$, which in turn gives the value of $\widetilde{R}_{M,T_j} = \log(S_{M,T_j}) - \log(S_{M,0})$. We apply this method to simulate 100,000 paths of $\{S_{M,1}, \cdots, S_{M,T_1}\}$ and $\{S_{M,1}, \cdots, S_{M,T_2}\}$, yielding 100,000 samples of S_{M,T_1}, $\widetilde{R}_{M,T_1}$, S_{M,T_2} and $\widetilde{R}_{M,T_2}$.

3 Verifying the Condition $1 - F_{\widetilde{R}_M}(1) > e^{-1}$

Notice that $1 - F_{\widetilde{R}_M}(1) = \mathbb{P}(\widetilde{R}_M > 1)$, i.e., the probability that the market has a positive return rate. So this quantity reflects the trend of the overall economy/market. Let us examine this condition for the S&P 500. Following De Giorgi and Legg (2012), we assume that the S&P 500 annual return rate, $\tilde{r}_M$, follows either a normal distribution, a skew-normal distribution, or a log-normal distribution. Based on the annual historical data of the S&P 500 from January 1946 to January 2009, the annual return rate, $\tilde{r}_M$, has a mean of 7.6% and a standard deviation of 15.8%. Moreover, in the case of skew-normal, its skewness is estimated to be -0.339. Based on these values, assuming $\tilde{r}_M$ is respectively normal, skew-normal, and log-normal, we can estimate the corresponding value of $1 - F_{\widetilde{R}_M}(1)$ to be respectively 0.6847, 0.6980, and 0.6543, all greater than e^{-1}.

Next, we consider the S&P 500 *monthly* return, as our main empirical analysis is conducted at the monthly frequency. Following Polkovnichenko and Zhao (2013), we use the daily historical data of the S&P 500 from January 2, 1991 to April 29, 2016 and estimate an EGARCH model for the S&P 500 with 1250 daily index return data up to the end of each month from January 1996 to April 2016. Then, we simulate 100,000 samples based on the EGARCH model following the same approach to be described in Section 2 of the online appendix and compute the nonparametric probability density function for the monthly return of the S&P 500. We find that $1 - F_{\widetilde{R}_M}(1) > e^{-1}$ holds for *all* the months tested.

4 GMM for Implied Preference Parameters

In this section, we provide the detailed procedure of GMM to derive the implied preference parameters α and γ used in Subsection 3.1 of the paper. We use the S&P 500 index option data and the S&P 500 index data to derive the option-implied preference parameters for any given period of time (monthly in this section). These indices are then used to gauge γ and α for the period concerned. The option data are obtained from OptionMetrics, with a time window from January 4, 1996 to April 29, 2016. The index data are obtained from the Centre for Research in Security Prices (CRSP), with a time window from January 2, 1991 to April 29, 2016. We derive the preference parameters by matching the theoretical implied volatilities with the observed implied volatilities based on out-of-the-money (OTM) options only. Following a convention in literature, we use OTM options with nearest maturity T_1 (less than a month) and second nearest maturity T_2 (larger than a month) to compute one-month implied volatilities. The general reasons for using OTM options only include better liquidity and price reliability compared with in-the-money (ITM) options (Christoffersen et al., 2013; CBOE, 2021), and the specific reason in our setting is that probability weighting is driven by tail behaviors of returns that are more relevant to OTM options.[3]

With the parametric preference functions (6), it follows from Theorem 1 in the paper that the pricing kernel $\tilde{m}_j$ is, now denoted as a function of the preference parameter $\theta = (\gamma, \alpha)$

$$\tilde{m}_j(\theta) = \frac{\alpha e^{-(-\log(1-F_{\widetilde{R}_{M,T_j}}(\widetilde{R}_{M,T_j})))^{\alpha}} (-\log(1 - F_{\widetilde{R}_{M,T_j}}(\widetilde{R}_{M,T_j}))^{\alpha-1} (\widetilde{R}_{M,T_j})^{-\gamma}}{(1 + r_{f,T_j})(1 - F_{\widetilde{R}_{M,T_j}}(\widetilde{R}_{M,T_j}))\mathbb{E}\left[w'(1 - F_{\widetilde{R}_{M,T_j}}(\widetilde{R}_{M,T_j}))u'(\widetilde{R}_{M,T_j})\right]},$$

where $\widetilde{R}_{M,T_j}$ and r_{f,T_j} are respectively the total return of the S&P 500 index and the risk-free return with time horizon T_j, $j = 1, 2$. The OTM options have theoretical prices

$$C(\theta, K^C_{i,j}, T_j) = \mathbb{E}[\tilde{m}_j(\theta)(S_{M,T_j} - K^C_{i,j})_+], \quad P(\theta, K^P_{\ell,j}, T_j) = \mathbb{E}[\tilde{m}_j(\theta)(K^P_{\ell,j} - S_{M,T_j})_+], \quad j = 1, 2,$$

where S_{M,T_j} is the S&P 500 index value at time T_j, $K^C_{i,j}$ is the strike price for call i, $i = 1, \cdots, n_j$, $K^P_{\ell,j}$ is the strike price for put ℓ, $\ell = 1, \cdots, m_j$, n_j and m_j are the numbers of OTM calls and puts with maturity T_j, respectively. Notice that n_j and m_j may vary for different months. Based on the Black–Scholes option pricing formula, we can further obtain the theoretical implied volatilities, denoted by $IV_C(\theta, K^C_{i,j}, T_j)$ and $IV_P(\theta, K^P_{\ell,j}, T_j)$. Then, we simulate 100,000 samples of $\widetilde{R}_{M,T_j}$ and S_{M,T_j} based on the EGARCH model. Some details of the simulation procedure are provided in Section 2 of the online appendix. The theoretical values for implied volatilities are then approximated by summing over the 100,000 samples.

At the end of each month, we compute the differences between the theoretical and real implied volatilities of OTM options:

$$\begin{aligned} g(\theta) = [&IV_C(\theta, K^C_{1,1}, T_1) - IV^{real}_C(\theta, K^C_{1,1}, T_1), \cdots, IV_C(\theta, K^C_{n_2,2}, T_2) - IV^{real}_C(\theta, K^C_{n_2,2}, T_2), \\ &IV_P(\theta, K^P_{1,1}, T_1) - IV^{real}_P(\theta, K^P_{1,1}, T_1), \cdots, IV_P(\theta, K^P_{m_2,2}, T_2) - IV^{real}_P(\theta, K^P_{m_2,2}, T_2)], \end{aligned}$$

[3] The authors are very grateful to an anonymous referee for suggesting using OTM only. A comparison of results between OTM only and the full cross section (including ITM) is presented in Section 5.7 of the online appendix.

where $IV_C^{real}(\theta, K_{i,j}^C, T_j)$ and $IV_P^{real}(\theta, K_{\ell,j}^P, T_j)$ are the real implied volatilities of OTM calls and puts, respectively. In total, there are $(n_1 + n_2 + m_1 + m_2)$ moment conditions. Finally, we estimate θ by solving

$$\hat{\theta} = \arg \min_{\gamma \geq 0, \alpha \geq 0} g(\theta) W g(\theta)^T,$$

where ^{T} denotes the transpose of a vector or a matrix, W is the weighting matrix for the moment conditions, chosen in our study to be the identity matrix yielding an equal weight to all the components of g. In view of the forward-looking nature of option price, the estimated preference parameters at the end of each month can be used to predict the representative agent's preference in the next month.

With the estimated preference parameters $(\hat{\alpha}, \hat{\gamma})$ obtained by GMM as well as the EGARCH-estimated market return distribution $F_{\widetilde{R}_M}$, we can get the theoretical values of A and B from the equations (4)–(5) in the paper.

As discussed in Subsection 3.1 of the main body of the paper, there are 244 months in our sample, containing 143 overweighting months and 101 underweighting ones. For the 143 overweighting months, the majority of A (135 out of 143) are positive and those of B are all negative. For the 101 underweighting periods, A is negative during 58 periods. Moreover, the mean and standard error of A are respectively -0.6795 and 0.1704 conditional on underweighting, implying that the negativity of A is statistically significant at 1% level. On the other hand, B is positive during 69 months out of the 101 underweighting months. The corresponding mean and standard error are 1.2634 and 1.0234 respectively. Hence the mean of B is not statistically significant due to the large standard error.

So the theoretical predictions based on the estimated market and preference data are that A's are mostly positive under overweighting and mostly negative under underweighting, while B's are all negative under overweighting but insignificant under underweighting. These predictions is confirmed empirically in Section 4 of the paper.

5 Robustness Analysis

5.1 Two-way Sort Portfolio Analysis

In this section, we conduct two-way sort portfolio analysis to separate the effects of covariance and coskewness. For each month t from January 1996 to April 2016, we classify all the stocks into five coskewness-sorted groups based on their pre-ranking coskewness. Then, in each coskewness-sorted group, we further group the stocks into five covariance-sorted groups based on their pre-ranking covariance. Thus, there are now a total of 25 groups. For each group, we construct the value-weighted portfolio, calculate next month $(t+1)$'s post-ranking portfolio returns, and repeat the procedure from the next month onwards. As the stocks have similar values of coskewness within the same coskewness-sorted group, we can thereby control the influence of the coskewness and compare the performance of the five covariance-sorted portfolios. Similarly, we also analyze the 25 covariance-coskewness-sorted portfolios based first on their pre-ranking covariance and then on their pre-ranking coskewness.

Table A1 reports the average next month's returns of value-weighted portfolios for the entire period, overweighting periods and underweighting periods, respectively, where the overweighting periods and underweighting periods are determined by PWI. Panel A shows that with the coskewness controlled, the spread portfolios (5-1) still have significantly positive monthly returns during the overweighting periods and significantly negative monthly returns during the underweighting periods. Panel B, on the other hand, yields that with covariance controlled the spread portfolios (5-1) have significantly negative monthly returns during the overweighting periods and insignificantly monthly returns during the underweighting periods.

Our portfolio analysis suggests that one can significantly improve the performance of the long-short covariance-sorted and coskewness-sorted strategies by making use of the additional information from the PWI proposed in this paper. It confirms that probability weighting is an important factor for pricing securities and building profitable portfolios.

5.2 The Effect of Time-Varying Risk Aversion

From our empirical analysis in the paper, it follows that the risk premia depend critically on whether the market is overweighting or underweighting tails, or in general on the time-variation of probability weighting. However, risk preferences consist of two parts in our model, risk aversion and probability weighting, which are both time-varying as shown empirically in Subsection 3.1 of the main body of the paper. As a result, the impacts of the time-variation of the two may be entangled in our previous analysis. In this subsection, we disentangle them by controlling the risk aversion.

To do this, we assume that the risk aversion is constant over time and time-variation is attributed all to the probability weighting. Following the procedure of Polkovnichenko and Zhao (2013), we derive the implied probability weighting function, denoted as PWI^{PZ}, also from option data. Different from the GMM method, Polkovnichenko and Zhao (2013)'s method makes no assumption on the functional form of the probability weighting; instead it assumes a power utility function with a fixed risk aversion ($\gamma = 2$ in our analysis). Figure A1(a) puts together PWI^{PZ} and PWI for easy comparison. The former has a much larger variation than the latter. This is expected, because all the time-variation of risk preferences is attributed to probability weighting under Polkovnichenko and Zhao (2013)'s framework. On the other hand, Figure A1(a) also shows that the two indices tend to move in the same direction. When we separate overweighting and underweighting periods by the two indices respectively, there are only 12 months (among 244 months in total) that are labeled differently.

Panel A in Table A2 reports the Fama-MacBeth cross-sectional regression results under PWI^{PZ}, which is characteristically similar to our main results presented in Panel A of Table 2 in the paper. In Panel A of Table A2, A is significantly positive during overweighting periods and significantly negative during underweighting periods, and B almost doubles its overall value during overweighting periods yet becomes insignificant during underweighting periods. Comparing Panel A of Table 2 in the paper with Panel A of Table A2 here, the results of the

Table A1: Two-way sort results based on PWI

Panel A: average returns of coskewness-covariance-sorted value-weighted portfolios

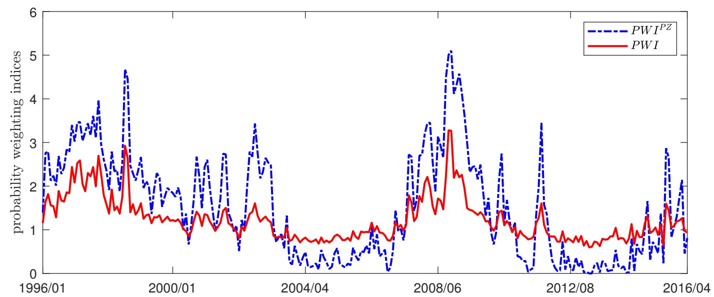
Periods	coskewness	covariance						
		1	2	3	4	5	(5-1)	t-statistic of (5-1)
All	1	1.17%	1.25%	1.33%	1.28%	0.79%	-0.38%	(-0.99)
	2	1.06%	1.20%	1.24%	1.01%	0.89%	-0.17%	(-0.39)
	3	1.08%	1.20%	1.15%	1.11%	1.09%	0.01%	(0.03)
	4	0.90%	0.95%	1.06%	1.13%	1.09%	0.19%	(0.44)
	5	0.71%	1.11%	0.93%	0.95%	1.20%	0.49%	(0.82)
Over	1	1.28%	1.45%	1.70%	2.16%	1.88%	0.60%	(0.93)
	2	0.97%	1.09%	1.48%	1.41%	2.18%	**1.22%***	(1.72)
	3	1.01%	1.13%	1.18%	1.41%	2.30%	**1.28%****	(1.97)
	4	0.69%	0.83%	1.04%	1.36%	2.36%	**1.67%****	(2.37)
	5	0.30%	1.20%	1.25%	2.00%	2.93%	**2.63%*****	(2.64)
Under	1	1.07%	1.06%	0.97%	0.41%	-0.29%	**-1.36%*****	(-3.19)
	2	1.15%	1.32%	0.99%	0.61%	-0.40%	**-1.55%*****	(-3.37)
	3	1.15%	1.26%	1.13%	0.81%	-0.11%	**-1.26%*****	(-2.78)
	4	1.11%	1.08%	1.07%	0.90%	-0.18%	**-1.29%*****	(-2.66)
	5	1.12%	1.01%	0.62%	-0.10%	-0.53%	**-1.65%*****	(-2.78)

Panel B: average returns of covariance-coskewness-sorted value-weighted portfolios

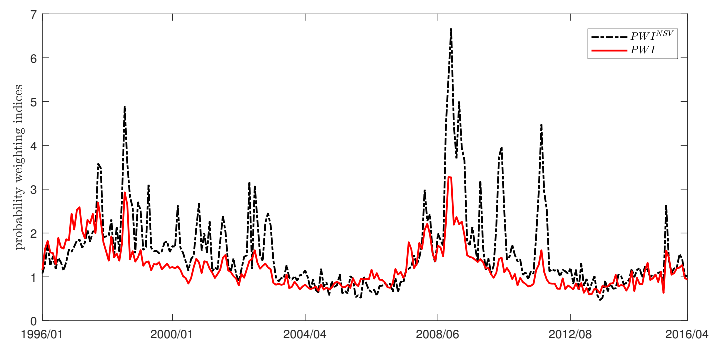
Periods	covariance	coskewness						
		1	2	3	4	5	(5-1)	t-statistic of (5-1)
All	1	1.47%	1.20%	1.04%	0.99%	0.69%	**-0.78%****	(-2.33)
	2	1.41%	1.15%	1.21%	0.98%	0.70%	**-0.71%*****	(-3.03)
	3	1.36%	1.21%	1.12%	1.06%	0.93%	**-0.44%****	(-1.94)
	4	1.46%	1.14%	1.08%	1.00%	0.85%	**-0.61%****	(-2.50)
	5	0.83%	0.88%	1.20%	1.07%	0.88%	0.06%	(0.22)
Over	1	1.75%	1.27%	0.90%	0.80%	0.27%	**-1.47%****	(-2.40)
	2	1.67%	1.10%	1.18%	0.80%	0.56%	**-1.10%*****	(-2.73)
	3	1.85%	1.44%	1.13%	1.13%	0.89%	**-0.96%*****	(-2.59)
	4	2.47%	1.88%	1.51%	1.27%	1.13%	**-1.34%*****	(-3.16)
	5	2.22%	2.20%	2.47%	2.51%	2.22%	0.00%	(0.01)
Under	1	1.19%	1.14%	1.18%	1.19%	1.11%	-0.08%	(-0.32)
	2	1.14%	1.20%	1.23%	1.16%	0.83%	-0.31%	(-1.35)
	3	0.88%	0.98%	1.10%	0.99%	0.97%	0.09%	(0.37)
	4	0.46%	0.40%	0.66%	0.74%	0.57%	0.11%	(0.48)
	5	-0.57%	-0.43%	-0.07%	-0.37%	-0.45%	0.11%	(0.45)

Notes: The overweighting periods (Over) and underweighting periods (Under) are separated by the median of PWI. The spread portfolio (5-1) is constructed by longing the 5th portfolio and shorting the 1st portfolio. The t-statistics of the returns of the spread portfolios are reported. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

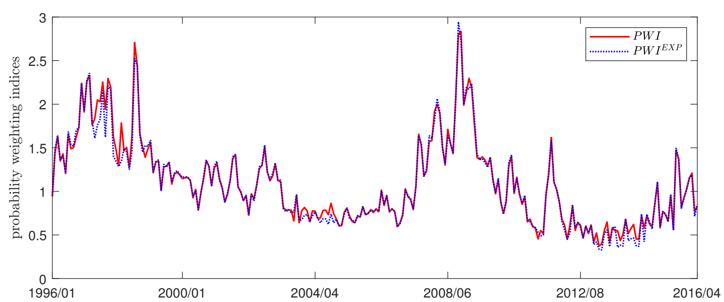
Figure A1: Various probability weighting indices



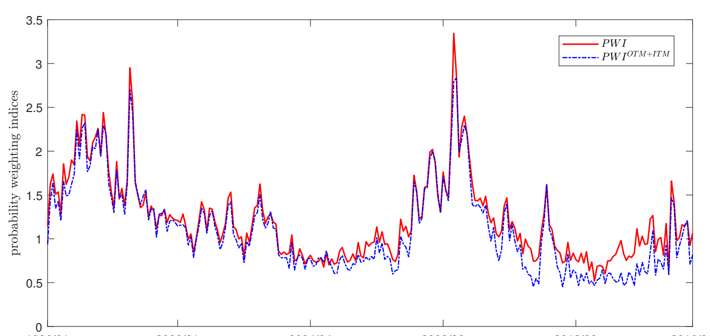
(a) PWI vs PWI^{PZ} (probability weighting with fixed risk aversion)



(b) PWI vs PWI^{NSV} (probability weighting excluding stochastic volatility)



(c) PWI vs PWI^{EXP} (probability weighting under exponential utility)



(d) PWI vs $PWI^{OTM+ITM}$ (probability weighting with both OTM and ITM) former with time-varying risk aversion are stronger with larger average absolute values of A and B.

5.3 The Effect of Stochastic Volatility

The main thrust of this paper is to use probability weighting to explain empirical patterns including the U-shape/bell-shape pricing kernels. In literature, a more conventional explanation for the option pricing kernel puzzle is stochastic volatility (see Christoffersen et al., 2013). In this subsection, we aim to disentangle the effects of probability weighting and stochastic volatility.

Following Christoffersen et al. (2013), we characterize the S&P 500 index by Heston and Nandi (2000)'s GARCH model,

$$\begin{aligned} \log(S_{M,t}) &= \log(S_{M,t-1}) + r_f + \left(\mu - \tfrac{1}{2}\right) h_t + \sqrt{h_t} z_t, \\ h_t &= \nu + \beta h_{t-1} + \zeta \left(z_{t-1} - \tau\sqrt{h_{t-1}}\right)^2, \end{aligned} \tag{A4}$$

where z_t has a standard normal distribution, and parameters $(\mu, \nu, \beta, \zeta, \tau)$ in each month are estimated based on 1250 daily S&P 500 index return data up to the end of that month. We further assume the following pricing kernel with stochastic volatility,

$$\hat{m}_t^{sv} = \left(\frac{S_{M,t}}{S_{M,0}}\right)^{\phi} \exp\left(\delta t + \eta \sum_{s=1}^{t} h_s + \xi(h_{t+1} - h_1)\right),$$

where the parameters δ and η govern the time preference, and ϕ and ξ capture the respective aversions to equity and variance risk. According to Christoffersen et al. (2013), the pricing kernel $\hat{m}_t^{sv}$ is consistent with the GARCH dynamics (A4) with the following mapping parameters: $\delta = -(\phi+1)r_f - \xi\nu + \frac{1}{2}\ln(1-2\xi\zeta)$, $\phi = -(\mu - \frac{1}{2} + \tau)(1-2\xi\zeta) + \tau - \frac{1}{2}$, $\eta = -(\mu - \frac{1}{2})\phi - \xi\zeta\tau^2 + (1-\beta)\xi - \frac{(\phi - 2\xi\zeta\tau)^2}{2(1-2\xi\zeta)}$. Thus, ξ is the only free parameter in determining $\hat{m}_t^{sv}$, which we now denote by $\hat{m}_t^{sv} = \hat{m}_t^{sv}(\xi)$.

To our best knowledge, there are no existing theoretical results for portfolio selection under RDUT preference and stochastic volatility. Thus, we opt to disentangle the effects of probability weighting and stochastic volatility empirically. To this end, we hypothesize that the pricing formulae in the current model with both stochastic volatility and probability weighting are

$$\begin{aligned} C^{sv+pw}(\varphi, K, T_j) &= \mathbb{E}[\hat{m}_{T_j}^{sv}(\xi)(S_{M,T_j} - K)_+ w'(1 - F_{\widetilde{R}_{M,T_j}}(\widetilde{R}_{M,T_j}))], \\ P^{sv+pw}(\varphi, K, T_j) &= \mathbb{E}[\hat{m}_{T_j}^{sv}(\xi)(K - S_{M,T_j})_+ w'(1 - F_{\widetilde{R}_{M,T_j}}(\widetilde{R}_{M,T_j}))], \end{aligned} \tag{A5}$$

where C^{sv+pw} and P^{sv+pw} are the theoretical option prices and $\varphi = (\xi, \alpha)$ with α being the parameter in the probability weighting function. Note that (A5) provides a flexible model covering the two ingredients. Taking the call option formula for example, the price under only stochastic volatility, C^{sv}, is obtained from (A5) with $\alpha = 1$ or $w'(\cdot) = 1$, and the price under only probability weighting, C^{pw}, follows from (A5) with $\hat{m}^{sv} = u'(\cdot)$, as proposed in our paper.

Similar to Subsection 3.1 of the main body of the paper, we use GMM to estimate the parameter value $\hat{\varphi}$ at the end of each month by matching the theoretical implied volatilities with the real implied volatilities of OTM options (using the nearest maturity T_1 and the second nearest maturity T_2):

$$\hat{\varphi} = \arg\min_{\alpha \geq 0} \ \bar{g}(\varphi)\bar{g}(\varphi)^T,$$

where $\bar{g}(\varphi) = [IV_C^{sv+pw}(\varphi, K_{1,1}^C, T_1) - IV_C^{real}(K_{1,1}^C, T_1), \cdots, IV_C^{sv+pw}(\varphi, K_{n_2,2}^C, T_2) - IV_C^{real}(K_{n_2,2}^C, T_2), IV_P^{sv+pw}(\varphi, K_{1,1}^P, T_1) - IV_P^{real}(K_{1,1}^P, T_1), \cdots, IV_P^{sv+pw}(\varphi, K_{m_2,2}^P, T_2) - IV_P^{real}(K_{m_2,2}^P, T_2)]$. Figure A1(b) reports the newly derived probability weighting index $PWI^{NSV} = 1/\hat{\alpha}$ (the black dash dot line) and the original GMM-derived probability weighting index PWI in Section 3.1 of the main body of the paper (the red solid line). These two indices again move together, having a correlation coefficient 0.7550. Panel B of Table A2 reports the Fama-MacBeth cross-sectional regression results under the new probability weighting index, PWI^{NSV}. Comparing Panel B of Table A2 here with Panel A of Table 2 in the paper, we can see that the new probability weighting index PWI^{NSV} still has the ability to distinguish the signs of the risk premia, if with slightly less effectiveness for the risk premium A and more effectiveness for the risk premium B. In sum, probability weighting still plays a significant role in pricing even taking into consideration the dependence of the pricing kernel on stochastic volatility.

5.4 The Effect of Data Frequency

We have used monthly data in our empirical analysis in the paper. In this section, we change the data frequency from monthly to weekly to check the robustness of our results.[4] As the one-month maturity options are considered to be the most liquid options (see Bakshi et al., 2010, Polkovnichenko and Zhao, 2013, Baele et al., 2019), we still use the one-month maturity option prices to estimate the risk aversion and probability weighting parameters at the end of each week. Then, following the same GMM estimation procedure described in Section 3.1, we derive the preference parameters (α_t and γ_t) at the closing of each Friday.

Furthermore, we use a 52-week window to estimate the covariance and coskewness for each stock i at week t:

$$\tilde{r}_{i,\tau} = k_{i,t} + \text{covariance}_{i,t}\tilde{r}_{M,\tau} + \text{coskewness}_{i,t}\tilde{r}_{M,\tau}^2 + \epsilon_{i,\tau}, \quad \tau = t - 52, \cdots, t.$$

Note that because we use a smaller window to estimate the covariance and coskewness, we have more stocks with available values in a given month, and thus have more stocks in the cross-sectional regression; see the row of n in Panel C of Table A2.

[4]We choose not to test the daily data because they are known to be too noisy compared with weekly or monthly data. Moreover, the switch between overweighting and underweighting tails occurs very slowly; see Figure 5(b) in the paper based on monthly estimated probability weighting indices. It would then be not necessary to use daily estimated probability weighting parameters to capture the much less frequent overweighting/underweighting changes.

Panel C of Table A2 reports the Fama-MacBeth cross-sectional regression results with weekly data. Compared with the monthly data in Panel A of Table 2 in the paper, probability weighting still has the ability to distinguish the signs of the risk premia, if with less statistical significance and smaller values of the coefficients. Such slightly weak (but still statistically significant) results are expected, since weekly data are known to be much noisier than monthly data.

5.5 Results under Exponential Utility

Hitherto we have used power CRRA as the outcome utility function. Meyer and Meyer (2005) argue that an accepted property of a utility function is that its elasticity is zero or (slightly) positive. CRRA has zero elasticity and the exponential utility has a positive elasticity. So, in this subsection, we take the latter as an alternative to analyze the robustness of our empirical results.

Specifically, with the exponential utility we re-estimate the probability weighting parameter by the same GMM procedure in Section 3.1. For easy comparison we put the derived probability weighting indices under both power and exponential utilities together in Figure A1(c). Clearly, the two are almost identical with a correlation close to 0.9945. Moreover, these two probability weighting indices result in exactly the same separation of overweighting and underweighting periods. Therefore, the Fama-MacBeth results under the exponential utility is the same as Panel A in Table 2 of the paper. In sum, changing CRRA to exponential does not change our main results.

5.6 Results on the UK Data

We have carried out empirical analysis on the US market in the paper. In this section, we test on non-US data, specifically the UK data.

We use FTSE 100 index option data and FTSE 100 index data to derive the monthly option-implied probability weighting index. The monthly option data are obtained from Bloomberg with a time window from January 2000 to December 2015. The daily index data are from Bloomberg with a time window from January 2, 1995 to December 31, 2015. Following the same GMM estimation procedure described in Section 3.1, we derive the preference parameters (α_t and γ_t) at the end of each month, and separate the whole period into overweighting months and underweighting months by PWI^{UK}.

We use the FTSE 100 constituents to perform the Fama-MacBeth regression analysis.[5] The corresponding price data are obtained from Bloomberg ranging from January 1995 to December 2015. Panel D in Table A2 reports the Fama-MacBeth regression results, which are largely consistent with those on the US data. Specifically, during the overweighting periods the risk premium of covariance is significantly positive and that of coskewness significantly

[5] We use only the FTSE 100 constituents because during the studied period, all the stocks listed in the European markets were traded in London Stock Exchange and it is hard to clearly define the UK stocks from the dataset.

Table A2: Various Fama-MacBeth regression results

Panel A: Results under PWI^{PZ} with fixed risk aversion

	All periods	Overweighting periods	Underweighting periods
A	1.51	**8.89****	**-5.87*****
	(0.65)	(2.20)	(-2.74)
B	**-0.88*****	**-1.64*****	-0.09
	(-2.87)	(-2.89)	(-0.53)
Adj. R^2	0.02	0.02	0.02
n	2695	2819	2571

Panel B: Results under PWI^{NSV} excluding stochastic volatility effect

	All periods	Overweighting periods	Underweighting periods
A	1.51	**9.15****	**-6.22*****
	(0.65)	(2.09)	(-3.11)
B	**-0.88*****	**-1.78*****	0.05
	(-2.87)	(-2.96)	(0.25)
Adj. R^2	0.02	0.02	0.02
n	2695	2801	2589

Panel C: Results under PWI^{week} with weekly data

	All periods	Overweighting periods	Underweighting periods
A	-0.46	**1.82***	**-2.69*****
	(-0.75)	(1.76)	(-4.26)
B	**-0.16****	**-0.37*****	0.02
	(-2.23)	(-2.70)	(0.91)
Adj. R^2	0.02	0.02	0.02
n	4726	5248	4217

Panel D: Results under PWI^{UK} with UK data

	All periods	Overweighting periods	Underweighting periods
A	3.91	**15.85****	-5.18
	(1.02)	(2.02)	(-1.62)
B	**-1.78*****	**-3.91*****	-0.28
	(-2.77)	(-2.98)	(-0.48)
Adj. R^2	0.06	0.07	0.06
n	82	80	83

Panel E: Results under $PWI^{OTM+ITM}$ with both OTM and ITM options

	All periods	Overweighting periods	Underweighting periods
A	1.51	**9.62****	**-6.55*****
	(0.65)	(2.27)	(-2.87)
B	**-0.88*****	**-1.71*****	-0.04
	(-2.87)	(-2.85)	(-0.21)
Adj. R^2	0.02	0.02	0.02
n	2695	2819	2568

Notes: "Underweighting periods" and "Overweighting periods" are separated by the median for PWI^{PZ}, PWI^{NSV}, PWI^{week}, PWI^{UK}, $PWI^{OTM+ITM}$. The average values of the coefficients, which are multiplied by 1000, are reported above and the corresponding t-statistics are reported below in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. The row labelled "Adj. R^2" reports the average adjusted R-square of each regression. The row labelled "n" reports the average number of stocks in each cross-sectional regression.

negative. During the underweighting periods, the UK risk premium of covariance is still negative but not significant; thus the result is slightly weaker than the US counterpart.

5.7 Results Based on Both ITM and OTM Options

In Section 3.1 of the paper, we derive the probability weighting index based on OTM options only, for the reasons of better liquidity, price reliability, and the relevance of probability weighting to the tail behavior of pricing kernel. In this subsection, we investigate how the result will be affected by including both ITM and OTM options. The corresponding probability weighting index derived is denoted by $PWI^{OTM+ITM}$, which contains the information from the entire cross section.

The real traded S&P 500 index options at the end of each month usually do not have exactly one month maturity; hence we obtain the interpolated prices of one-month maturity options from OptionMetrics.[6] Following Baele et al. (2019), we consider 13 index calls with deltas ranging from 0.2 to 0.8 (from OTM to ITM), and 13 index puts with deltas ranging from -0.8 to -0.2 (from ITM to OTM). We further denote the strike prices of these calls and puts as $K_1^C, \cdots, K_{13}^C$ and $K_1^P, \cdots, K_{13}^P$, respectively.

Given pricing kernel $\tilde{m}$, the theoretical prices of call and put options with strike price K and one-month maturity are

$$C(\theta, K) = \mathbb{E}[\tilde{m}(\theta)(S_M - K)_+], \quad P(\theta, K) = \mathbb{E}[\tilde{m}(\theta)(K - S_M)_+], \tag{A6}$$

where S_M is the S&P 500 index value in one month, $\tilde{m}(\theta)$ is the pricing kernel with $\theta = (\gamma, \alpha)$ being the preference parameters.

At the end of each month, we estimate an EGARCH model for the S&P 500 index with 1250-daily-window data. Then, based on the EGARCH model, we simulate 100,000 samples. The theoretical options' prices in (A6) are then approximated by summing over the 100,000 samples. Then, we compute

$$\begin{aligned} g(\theta) = [&C(\theta, K_1^C) - C^{real}(K_1^C), \cdots, C(\theta, K_{13}^C) - C^{real}(K_{13}^C), \\ &P(\theta, K_1^P) - P^{real}(K_1^P), \cdots, P(\theta, K_{13}^P) - P^{real}(K_{13}^P)], \end{aligned}$$

where C^{real} and P^{real} are respectively the observed prices of the calls and puts from OptionMetrics. In total, there are 26 moment conditions. Finally, we estimate θ by solving

$$\hat{\theta} = \arg \min_{\gamma \geq 0, \alpha \geq 0} g(\theta) g(\theta)^T.$$

The time series of $PWI^{OTM+ITM}$ (the blue dashed-dot line) and PWI (OTM only, the red solid line) are plotted together in Figure A1(d). It is evident that $PWI^{OTM+ITM}$ is highly

[6] OptionMetrics constructs synthetic one-month implied volatilities (IVs) by using kernel regression approximations to observed IVs for all traded options, and provides the standardized prices by inverting standardized IVs at fixed delta and maturity values to prices based on Black–Scholes formula.

correlated with PWI (with correlation coefficient 0.9783). On the other hand, $PWI^{OTM+ITM}$ takes slightly lower values than PWI. This is expected, as OTM options contain information about the tail behavior of the pricing kernel leading to higher values of PWI. As a result, $PWI^{OTM+ITM}$ tends to yield more underweighting months. Indeed, during the total of 244 months, there are 121 overweighting months and 123 underweighting months by $PWI^{OTM+ITM}$, resulting in a ratio between overweighting and underweighting months around 5:5. By contrast, the corresponding ratio is around 6:4 by PWI.

We also run the Fama-MecBeth cross-sectional regression based on $PWI^{OTM+ITM}$ and report the results in Panel E of Table A2, which show that $PWI^{OTM+ITM}$ still has the ability to distinguish the signs of the risk premia, if with less effectiveness.

5.8 An Alternative Model for Testing Pricing Kernel

One of the referees suggested using an alternative model to test whether the pricing kernel (2) in the main body of the paper really explains the cross-sectional differences between stocks and market portfolios. The suggested model is

$$\frac{\mathbb{E}[\tilde{R}_i - R_F]}{\mathbb{E}[\tilde{R}_M - R_F]} = beta(\tilde{R}_i, \tilde{m}) := \frac{\mathrm{Cov}(\tilde{R}_i - R_F, \tilde{m})}{\mathrm{Cov}(\tilde{R}_M - R_F, \tilde{m})}, \tag{A7}$$

which follows from the general fundamental equation $\mathbb{E}[(\tilde{R}_i - R_F)\tilde{m}] = 0$ regardless whether probability weighting is considered or not. Recall that our three-moment CAPM model is also derived from the same fundamental equation but with the specific form of the pricing kernel under probability weighting along with its Taylor expansion up to the second order:

$$\mathbb{E}[\tilde{r}_i] = r_f + A\mathrm{Cov}(\tilde{r}_i, \tilde{r}_M) + \frac{1}{2}B\mathrm{Cov}(\tilde{r}_i, \tilde{r}_M^2) + o(\tilde{r}_M^2). \tag{A8}$$

Compared with (A8), we believe the equation (A7) has some drawbacks when used for empirical testing. Indeed, $\tilde{r}_i$ and $\tilde{r}_M$ are directly observable, enabling the accurate computation of $\mathrm{Cov}(\tilde{r}_i, \tilde{r}_M)$ and $\mathrm{Cov}(\tilde{r}_i, \tilde{r}_M^2)$ in (A8). By contrast, the pricing kernel $\tilde{m}$ in (A7) is not observable and can only be estimated (e.g. from option data); hence the terms $\mathrm{Cov}(\tilde{R}_i - R_F, \tilde{m})$ and $\mathrm{Cov}(\tilde{R}_M - R_F, \tilde{m})$ may suffer from larger, compounded estimation errors. Moreover, $\mathrm{Cov}(\tilde{r}_i, \tilde{r}_M)$ and $\mathrm{Cov}(\tilde{r}_i, \tilde{r}_M^2)$ are risk measures and the coefficients (A, B) are the corresponding risk premia, which all have clear, well-established economic meanings. However, the quantity $beta(\tilde{R}_i, \tilde{m})$ in (A7) as the ratio of two covariances is foreign to us and its economic interpretation is not clear.

Nonetheless, we go ahead to conduct empirical testing by carrying out the following Fama–MacBeth cross-sectional regression:

$$\frac{R_{i,t+1} - R_{F,t+1}}{R_{M,t+1} - R_{F,t+1}} = C_t \cdot \frac{\mathrm{Cov}_t(R_{i,t} - R_{F,t}, m_t)}{\mathrm{Cov}_t(R_{M,t} - R_{F,t}, m_t)} + \epsilon_{i,t+1}, \tag{A9}$$

where m_t is estimated at the end of month t using the method presented in the paper (with probability weighting), and $\mathrm{Cov}_t(R_{i,t} - R_{F,t}, m_t)$ and $\mathrm{Cov}_t(R_{M,t} - R_{F,t}, m_t)$ are estimated with a 60-month rolling window. Then we do a t-test for the estimated coefficients $\{C_t\}$. The following Table A3 reports the results:

Table A3: Empirical results for equation (A9)

parameter	mean	t-statistic	p-value
C_t	4.211	7.999	5.23×10^{-14}

The results indicate that the coefficient C_t in (A9) is positive and significant; hence the pricing kernel (2) derived in the main body of the paper at least reconciles with the cross-sectional differences between stocks and market portfolios.

5.9 The Effect of Other State Variables

The same referee asked whether time varying probability weighting is driven out by standard state and conditioning variables such as option implied volatility, option-implied skew, market liquidity and credit yield spread. We collect the monthly VIX index as the option implied volatility from the CBOE website, compute the monthly option-implied skew from the option data at OptionMetrics using the approach of Bakshi et al. (2003), compute the monthly Amihud illiquidity measure of S&P 500 index based on the index data from Yahoo Finance website using the method of Amihud (2002), and obtain the monthly credit yield spread between Baa corporate and US Treasure bonds from the website https://www.longtermtrends.net/bond-yield-credit-spreads/. The time series of these variables are denoted as VIX_t, OIS_t, $ILLIQ_t$ and CYS_t, respectively.

We then conduct linear regressions for the estimated PWI in month $t+1$ with respect to the above state variables in month t. The following Table A4 reports the results. The adjusted R-square values indicate that these four state variables together can explain only about half of the PWI's variation. We stress that identifying more fundamental state variables driving probability weighting is an outstanding open problem and deserves additional serious research effort.

Table A4: Linear regression results for PWI_{t+1} onto various state variables in month t

	M1	M2	M3	M4	M5
Intercept	**0.32*****	**1.45*****	**0.79*****	**0.98*****	**0.31***
	(4.36)	(8.89)	(20.04)	(9.42)	(1.78)
VIX	**0.04*****				**0.04*****
	(11.32)				(8.17)
OIS		**0.35****			-0.11
		(2.17)			-(0.93)
$ILLIQ$			**5.46*****		**3.07*****
			(10.58)		(4.82)
CYS				0.05	**-0.12*****
				(1.20)	-(2.63)
Adj. R^2	0.34	0.02	0.31	0.00	0.51

Notes: The regression coefficients are reported above and the corresponding t-statistics are reported below in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. The row labelled "Adj. R^2" reports the adjusted R-square values of the regressions.

References

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56.

Baele, L., Driessen, J., Ebert, S., Londono, J. M., and Spalt, O. G. (2019). Cumulative prospect theory, option returns, and the variance premium. *The Review of Financial Studies*, 32(9):3667–3723.

Bakshi, G., Kapadia, N., and Madan, D. (2003). Stock return characteristics, skew laws, and the differential pricing of individual equity options. *The Review of Financial Studies*, 16(1):101–143.

Bakshi, G., Madan, D., and Panayotov, G. (2010). Returns of claims on the upside and the viability of u-shaped pricing kernels. *Journal of Financial Economics*, 97(1):130–154.

CBOE (2021). White paper cboe volatility index. *Available at: https://cdn.cboe.com/resources/vix/vixwhite.pdf. Last accessed: June 25, 2022.*

Christoffersen, P., Heston, S., and Jacobs, K. (2013). Capturing option anomalies with a variance-dependent pricing kernel. *The Review of Financial Studies*, 26(8):1963–2006.

De Giorgi, E. G. and Legg, S. (2012). Dynamic portfolio choice and asset pricing with narrow framing and probability weighting. *Journal of Economic Dynamics and Control*, 36(7):951–972.

Heston, S. L. and Nandi, S. (2000). A closed-form garch option valuation model. *The Review of Financial Studies*, 13(3):585–625.

Meyer, D. J. and Meyer, J. (2005). Relative risk aversion: What do we know? *Journal of Risk and Uncertainty*, 31(3):243–265.

Polkovnichenko, V. and Zhao, F. (2013). Probability weighting functions implied in options prices. *Journal of Financial Economics*, 107(3):580–609.

Xia, J. and Zhou, X. Y. (2016). Arrow–debreu equilibria for rank-dependent utilities. *Mathematical Finance*, 26(3):558–588.