Prospect Theory and Stock Market Anomalies

Nicholas Barberis, Lawrence Jin, and Baolian Wang

February 2019*

Abstract

Prospect theory was developed more than 30 years ago, but its predictions for basic aspects of asset prices such as the cross-section of average returns are still not well understood. In this paper, we build a new model of asset prices in which investors evaluate risk in part according to prospect theory, and show how the model can be used to make quantitative predictions about average stock returns. We then examine the model's ability to explain 22 prominent stock market anomalies. We find that the model is helpful for thinking about a majority of the anomalies we consider, including the momentum, volatility, distress, and profitability anomalies. It performs most poorly on the value anomaly. We present suggestive evidence that this is due to investors having incorrect beliefs about the distribution of value stock returns.

JEL classification:

Keywords:

*This is preliminary draft intended only for submission to the AFA – please do not forward it, distribute it, or make it public. The authors are grateful to seminar participants at Boston College for helpful feedback, and to Jonathan Ingersoll for extensive discussions.

1 Introduction

Research in behavioral finance tries to make sense of financial data using psychologically-grounded assumptions about investor behavior – for example, about investor attitudes to risk. The central idea in psychology about risk attitudes is prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). While this theory was developed over 30 years ago, we still do not have a full understanding of its implications for basic aspects of asset prices, such as the cross-section of average returns. For example, if investors evaluate risk according to Expected Utility, then, under some additional assumptions, average returns are described by the CAPM. But what is the structure of average returns if investors instead evaluate risk according to prospect theory?

In this paper, we try to answer this question. We build a new model of asset prices that incorporates prospect theory, as well as a related concept known as narrow framing, into investor preferences. We show how the model can be used to make quantitative predictions about the cross-section of average stock returns. In our main application, we take 22 prominent stock market anomalies and examine whether our model can help explain them. We find that the model is able to shed light on a majority of these anomalies.

Prospect theory posits that people derive utility from gains and losses measured relative to a reference point, and that they process these gains and losses using a utility function that is kinked at its origin, so that they are more sensitive to losses than to gains; concave over gains, so that they are risk averse over moderate-probability gains; and convex over losses, so that they are risk-seeking over moderate-probability losses. Moreover, they weight outcomes not by objective probabilities but by transformed probabilities that overweight the tails of the distribution they are thinking about. Prospect theory is often implemented in conjunction with narrow framing, a phenomenon observed in experimental studies whereby, when an individual is thinking about taking on a new risk, he evaluates it to some extent *in isolation*, separately from his other risks. In the stock market context, this means that, when an investor is wondering how much money to allocate to a particular stock, he thinks, at least in part, about the potential gains and losses in his holdings of the stock itself.

Intuition suggests that, in an economy with prospect theory investors who engage in narrow framing, the price of an asset will depend in part on three asset characteristics: the *volatility* of the asset's returns; the *skewness* of the asset's returns; and the average investor's

prior gain or loss in the stock, a quantity known as the stock's "capital gain overhang" (Grinblatt and Han, 2005). All else equal, investors require a *higher* average return on more volatile assets: since these investors evaluate each asset to some extent in isolation, and since they are loss averse, they find assets with volatile returns unappealing. All else equal, investors require a *lower* average return on assets with more positively-skewed returns: since these investors focus on a stock's own distribution of potential gains and losses, and since they overweight the tails of this distribution, they find assets with positively-skewed returns attractive. Finally, the convexity of the utility function over losses means that, all else equal, investors require a *lower* average return on stocks that they have prior losses in.

To study the implications of prospect theory for the cross-section of average returns, we start by building a model of asset prices that incorporates this concept. In our model, investor preferences have two components. The first is the traditional mean-variance preference specification – a component that, taken alone, leads to the CAPM. The second component incorporates prospect theory and narrow framing. Since, in a framework of this kind, an asset's volatility, skewness, and capital gain overhang are likely to be important determinants of the asset's average return, we make sure that our model can take account of all three characteristics.

While the structure of our model is simple, solving for equilibrium prices presents a challenge. In our economy, all investors are identical. In an Expected Utility framework, this would imply that, in equilibrium, all investors hold identical portfolios. Strikingly, such an equilibrium does not exist once we introduce prospect theory preferences. To break this logjam, we propose another type of equilibrium – one in which investors hold different portfolios that correspond to non-unique optima of their objective function. We find that this equilibrium does exist, and it is the one we study.

In our main application, we examine whether the model can explain 22 prominent stock market anomalies. To see whether our model can explain a particular anomaly – the size anomaly, say – we need to compute what it predicts for the average return of the typical small-cap stock. As explained above, this average return will depend on the volatility, skewness, and capital gain overhang of the typical small-cap stock. We estimate these quantities from historical U.S. data, plug them into our model, and record the model's prediction for the average return of a typical small-cap stock. We repeat this process for the typical stock in each of the ten market capitalization deciles. The results reveal how much, if any, of the size anomaly our model can explain. We proceed in the same way for all 22 anomalies.

Our historical estimates of the volatility, skewness, and capital gain overhang of the typical stock in each anomaly decile are interesting in their own right. We find that the three characteristics are strongly correlated across anomaly portfolios: if the typical stock in decile 1 has more positively skewed returns than the typical stock in anomaly decile 10, then it almost always also has more volatile returns and trades at a more negative capital gain. For example, in the case of the size anomaly, the typical small-cap stock not only has more positively-skewed returns than the typical large-cap stock, but also has more volatile returns and a more negative capital gain overhang.

This last observation points to the necessity of the quantitative approach we take in this paper. Consider once again the size anomaly. Empirically, the returns of the typical small-cap stock are much more volatile than those of the typical large-cap stock. All else equal, this would lead prospect theory investors who engage in narrow framing to charge a *higher* average return on small-cap stocks than on large-cap stocks, thereby helping to explain the size anomaly. However, the typical small-cap stock also has more positively-skewed returns, and a more negative capital gain overhang, than the typical large-cap stock. All else equal, these two factors would lead prospect theory investors to charge a *lower* average return on small-cap stocks, thereby hampering the model's ability to explain the size anomaly. It is only through a quantitative, model-based approach like the one we present here that we can find out what the overall prediction of prospect theory for the size anomaly is, once *all* of these factors are properly combined.

We find that our model is helpful for thinking about 13 of the 22 anomalies we consider, in the sense that it predicts a higher average return for the extreme anomaly decile portfolio that actually has a higher average return, empirically. These are the momentum, failure probability, idiosyncratic volatility, profitability, expected idiosyncratic skewness, return on assets, maximum daily return, O-score, external finance, composite equity issuance, net stock issuance, post-earnings announcement drift, and difference of opinion anomalies. Our model explains these anomalies in the same way. For all of these 13 anomalies, the typical stock in the extreme decile with the lower average return is more highly skewed, more volatile, and has a lower gain overhang than the typical stock in the other extreme decile. The greater skewness and lower gain overhang of the former stock leads investors to charge a lower average return on it, while its higher volatility leads investors to charge a higher average return on it. Quantitatively, the first effect dominates. As a consequence, our model's prediction about the anomaly is in line with the empirical facts.

While our model is helpful for thinking about many anomalies, there are some anomalies where it performs poorly – most notably, the size and value anomalies. For example, value stocks are more highly skewed and have a more negative capital gain than growth stocks. All else equal, this leads prospect theory investors to charge a lower average return on value stocks. However, value stocks are also more volatile, which, all else equal, leads investors to charge a higher average return on them. Quantitatively, the first effect dominates. Overall, the model predicts a lower average return on value stocks, contrary to the empirical facts.

In making its predictions, our model assumes that investors have rational beliefs: when pricing a stock, they are aware of the volatility and skewness of the stock's future return distribution, and of its capital gain overhang. One possible explanation for our model's poor performance on some anomalies is that investors have incorrect beliefs about the statistical properties of stocks' return distributions. To evaluate this possibility, we use data on stock options to extract investors' risk-neutral beliefs about the variance and skewness of individual stocks. Since these are risk-neutral beliefs, they constitute only suggestive evidence. Nonetheless, they suggest that investors have sensible beliefs about most anomaly returns. Interestingly, however, they also suggest that investors have incorrect beliefs about the returns on value stocks. Specifically, investors appear to think that the returns on these stocks have similar skewness to the returns on growth stocks, even though, empirically, value stock returns are more highly skewed than growth stock returns. These incorrect beliefs may help explain our model's counterfactual prediction about the value premium.

A number of other papers have linked prospect theory to stock market anomalies, and we discuss these in detail in the main body of the paper. Overall, we advance this line of research on three dimensions – in terms of theory, in terms of empirics, and in terms of scope. First, we present a new model of the cross-section that incorporates all the elements of prospect theory. Second, in order to get a quantitative sense of prospect theory's predictions about an anomaly, we use empirical measures of return volatility, return skewness, and gain overhang as inputs to the model. And third, while previous papers on prospect theory and market anomalies have each focused on one particular anomaly, we widen the scope of this research by looking at 22 different anomalies.

In Section 2, we review prospect theory and narrow framing. In Section 3, we present a model that incorporates these concepts and discuss the structure of the equilibrium. In Section 4, we list the 22 anomalies that are the focus of our study and compute the empirical characteristics of the stocks in the various anomaly deciles that serve as inputs to our model

– specifically, the volatility, skewness, and capital gain overhang of the typical stock in each anomaly decile. In Section 5, we present our main results. In Section 6, we discuss some additional issues raised by our analysis, while Section 7 concludes.

2 Prospect theory and narrow framing

Our goal is to study asset prices in an economy where investors have prospect theory preferences and engage in narrow framing. In this section, we briefly review these two psychological concepts. Readers already familiar with them may prefer to go directly to Section 3.

2.1 Prospect theory

The original version of prospect theory is described in Kahneman and Tversky (1979). While this paper contains all of the theory's essential insights, the specific model it presents has some limitations: it can be applied only to gambles with at most two nonzero outcomes, and it predicts that people will sometimes choose dominated gambles. Tversky and Kahneman (1992) propose a modified version of the theory known as cumulative prospect theory that resolves these problems. This is the version that is typically used in economic analysis and is the version we adopt in this paper.[1]

To see how cumulative prospect theory works, consider the gamble

$$(x_{-m}, p_{-m}; \ldots; x_{-1}, p_{-1}; x_0, p_0; x_1, p_1; \ldots; x_n, p_n), \tag{1}$$

which should be read as "gain or lose x_{-m} with probability p_{-m}, x_{-m+1} with probability p_{-m+1}, and so on," where $x_i < x_j$ for $i < j$ and where $x_0 = 0$, so that x_{-m} through x_{-1} are losses and x_1 through x_n are gains, and where $\sum_{i=-m}^{n} p_i = 1$. For example, a 50:50 bet to win \$110 or lose \$100 would be written as $(-\$100, \frac{1}{2}; \$110, \frac{1}{2})$. In the expected utility framework, an individual with utility function $U(\cdot)$ evaluates the gamble in (1) by computing

$$\sum_{i=-m}^{n} p_i U(W + x_i), \tag{2}$$

[1]While our analysis is based on cumulative prospect theory, we often abbreviate this to "prospect theory."

where W is his current wealth. A cumulative prospect theory individual, by contrast, assigns the gamble the value

$$\sum_{i=-m}^{n} \pi_i v(x_i), \tag{3}$$

where

$$\pi_i = \begin{cases} w(p_i + \ldots + p_n) - w(p_{i+1} + \ldots + p_n) & \text{for } 0 \leq i \leq n \\ w(p_{-m} + \ldots + p_i) - w(p_{-m} + \ldots + p_{i-1}) & \text{for } -m \leq i < 0 \end{cases}, \tag{4}$$

and where $v(\cdot)$ and $w(\cdot)$ are known as the value function and probability weighting function, respectively.[2] Tversky and Kahneman (1992) propose the functional forms

$$v(x) = \begin{cases} x^\alpha & \text{for } x \geq 0 \\ -\lambda(-x)^\alpha & \text{for } x < 0 \end{cases} \tag{5}$$

and

$$w(P) = \frac{P^\delta}{(P^\delta + (1-P)^\delta)^{1/\delta}}, \tag{6}$$

where α, $\delta \in (0, 1)$ and $\lambda > 1$. The left panel in Figure 1 plots the value function in (5) for $\alpha = 0.5$ and $\lambda = 2.5$. The right panel in the figure plots the weighting function $w(P)$ in (6) for $\delta = 0.4$ (the dashed line), for $\delta = 0.65$ (the solid line), and for $\delta = 1$, which corresponds to no probability weighting (the dotted line). Note that $v(0) = 0$, $w(0) = 0$, and $w(1) = 1$.

There are four important differences between (2) and (3). First, the carriers of value in prospect theory are gains and losses, not final wealth levels: the argument of $v(\cdot)$ in (3) is x_i, not $W + x_i$. Second, while $U(\cdot)$ is typically differentiable everywhere, the value function $v(\cdot)$ is kinked at the origin, as shown in Figure 1, so that the individual is more sensitive to losses – even small losses – than to gains of the same magnitude. This element of prospect theory is known as loss aversion and is designed to capture the widespread aversion to bets such as

$$(-\$100, \frac{1}{2}; \$110, \frac{1}{2}). \tag{7}$$

The severity of the kink is determined by the parameter λ; a higher value of λ implies a greater relative sensitivity to losses. Tversky and Kahneman (1992) estimate $\lambda = 2.25$ for their median subject.

Third, while $U(\cdot)$ is typically concave everywhere, $v(\cdot)$ is concave only over gains; over losses, it is convex. This pattern, which can be seen in Figure 1, captures the experimental finding that people tend to be risk averse over moderate-probability gains – they prefer a

[2]When $i = n$ or $i = -m$, equation (4) reduces to $\pi_n = w(p_n)$ and $\pi_{-m} = w(p_{-m})$, respectively.

certain gain of \$500 to $(\$1000, 0.5)$ – but risk-seeking over moderate-probability losses, in that they prefer $(-\$1000, 0.5)$ to a certain loss of \$500.[3] The degree of concavity over gains and convexity over losses are both governed by the parameter α; a lower value of α means greater concavity over gains and greater convexity over losses. Using experimental data, Tversky and Kahneman (1992) estimate $\alpha = 0.88$ for their median subject.

Finally, under cumulative prospect theory, the individual does not use objective probabilities when evaluating a gamble, but rather, transformed probabilities obtained from objective probabilities via the weighting function $w(\cdot)$. Equation (4) shows that, to obtain the probability weight π_i for an outcome $x_i \geq 0$, we take the total probability of all outcomes equal to or better than x_i, namely $p_i + \ldots + p_n$, the total probability of all outcomes strictly better than x_i, namely $p_{i+1} + \ldots + p_n$, apply the weighting function $w(\cdot)$ to each, and compute the difference. To obtain the probability weight for an outcome $x_i < 0$, we take the total probability of all outcomes equal to or worse than x_i, the total probability of all outcomes strictly worse than x_i, apply the weighting function $w(\cdot)$ to each, and compute the difference.

The main consequence of the probability weighting in (4) and (6) is that the individual overweights the *tails* of any distribution he faces. In equations (3)-(4), the most extreme outcomes, x_{-m} and x_n, are assigned the probability weights $w(p_{-m})$ and $w(p_n)$, respectively. For the functional form in (6) and for $\delta \in (0, 1)$, $w(P) > P$ for low, positive P; the right panel of Figure 1 illustrates this for $\delta = 0.4$ and $\delta = 0.65$. If p_{-m} and p_n are small, then, we have $w(p_{-m}) > p_{-m}$ and $w(p_n) > p_n$, so that the most extreme outcomes – the outcomes in the tails – are overweighted.

The overweighting of tails in (4) and (6) is designed to capture the simultaneous demand many people have for both lotteries and insurance. For example, people typically prefer $(\$5000, 0.001)$ to a certain \$5, but also prefer a certain loss of \$5 to $(-\$5000, 0.001)$. By overweighting the tail probability of 0.001 sufficiently, cumulative prospect theory can capture both of these choices. The degree to which the individual overweights tails is governed by the parameter δ; a lower value of this parameter implies more overweighting of tails. Tversky and Kahneman (1992) estimate $\delta = 0.65$ for their median subject.

[3] We abbreviate $(x, p; 0, q)$ as (x, p).

2.2 Narrow framing

Traditional models, in which utility functions are defined over wealth or consumption, make a clear prediction as to how an individual evaluates a new gamble he is offered: he merges the new gamble with other risks he is already facing to determine its effect on the distribution of his future wealth or consumption, and then checks if the new distribution is an improvement.

Research on decision-making under risk has uncovered many instances in which people do not appear to evaluate gambles in this way: instead of merging a new gamble with other risks they are already facing and checking if the combination is attractive, they often evaluate the new gamble in isolation, separately from their other risks. This is known as "narrow framing". More formally, narrow framing means that, when an individual is deciding whether to accept a gamble, he uses a utility function that depends directly on the outcome of the gamble, not just indirectly via the gamble's contribution to his total wealth.

Tversky and Kahneman (1981) present early laboratory evidence of narrow framing. More recently, Barberis, Huang, and Thaler (2006) argue that the commonly-observed rejection of the 110/100 gamble in (7) is evidence not only of loss aversion, but of narrow framing as well. To see why, note that most of the people who are offered this gamble are typically already facing *other* kinds of risk, such as labor income risk, housing risk, or financial market risk. In the absence of narrow framing, they must therefore evaluate the 110/100 gamble by mixing it with these other risks and then checking if the combination is attractive. It turns out that the combination *is* almost always attractive: since the 110/100 gamble is independent of other risks, it offers useful diversification benefits, which, even if loss averse, people can enjoy. The rejection of the 110/100 gamble therefore suggests that people are *not* fully merging the gamble with their other risks, but that, to some extent, they are evaluating it in isolation; in other words, that they are framing it narrowly.[4]

[4]For more discussion and evidence of narrow framing, see Kahneman and Tversky (1983), Tversky and Kahneman (1986), Redelmeier and Tversky (1992), Kahneman and Lovallo (1993), and Read, Lowenstein, and Rabin (2000).

3 Model and Equilibrium Structure

In this section, we present a model of asset prices in which investors evaluate risks in the way described by prospect theory and also engage in narrow framing. In an economy with such investors, three asset characteristics are particularly important for the pricing of an asset: the volatility of the asset's return distribution; the skewness of the asset's return distribution; and the average paper gain or loss in investors' holdings of the asset. Because these characteristics affect the price of the asset, it is important to build a model that takes account of all three.

Before we present our model, we explain in more detail *why* these three characteristics are important for pricing. Prospect theory investors who engage in narrow framing evaluate a stock by thinking about the potential gains and losses in their holdings of the stock, and then computing the prospect theory value of this distribution of potential gains and losses. Since they are loss averse, they dislike stocks with more volatile return distributions; all else equal, they require a higher average return on such stocks. Moreover, since, according to probability weighting, they overweight the tails of the distribution they are thinking about, they like stocks with positively-skewed return distributions; all else equal, they require a lower average return for such stocks. Finally, if a stock is trading at a gain for the typical investor, this investor finds himself to the right of the kink in the utility function in Figure 1. Since he is risk averse at this point, he demands a high average return to hold the stock. If, on the other hand, the stock has recently decreased in value, giving the typical investor a paper loss in the stock, then this investor finds himself in the convex region of the utility function, where he is risk-seeking. As a result, he requires a lower average return for holding the stock.

We now present a model of asset prices in which investors exhibit prospect theory and narrow framing, and which takes into account the three asset characteristics discussed above. Specifically, in Section 3.1, we present the model, and in Section 3.2, we discuss the structure of the equilibrium, a structure that differs from that of Expected Utility models.

3.1 Model setup

We consider a model with three dates, $t = -1$, 0, and 1. There is a risk-free asset with gross per-period return R_f. There are also N risky assets. The gross per-period return of risky asset i is $\widetilde{R}_i$, and the return vector $\widetilde{R} = (\widetilde{R}_1, \ldots, \widetilde{R}_N)$ has a multivariate distribution with cumulative distribution function $P(\widetilde{R})$; we specify $P(\cdot)$ below. The expected return of asset i is $\overline{R}_i$ and the covariance matrix of returns is $\Sigma = \{\sigma_{ij}\}$.

There is a representative investor in the economy. His wealth at time t is W_t, and the fraction of his time 0 wealth allocated to risky asset i is Θ_i, so that

$$\widetilde{W}_1 = W_0(\Theta'\widetilde{R} + (1 - 1'\Theta)R_f). \tag{8}$$

A crucial step is to specify the investor's preferences. At date 0, the investor solves:

$$\begin{aligned} & \max_{\Theta_1,\ldots,\Theta_N} \mathrm{E}(\widetilde{W}_1) - \frac{\gamma}{2}\mathrm{Var}(\widetilde{W}_1) + b_0 \sum_{i=1}^{N} V(\widetilde{G}_i) \\ = & \max_{\Theta_1,\ldots,\Theta_N} W_0(\Theta'\overline{R} + (1 - 1'\Theta)R_f) - \frac{\gamma}{2} W_0^2 \Theta'\Sigma\Theta + b_0 \sum_{i=1}^{N} V(\widetilde{G}_i), \end{aligned} \tag{9}$$

where

$$\widetilde{G}_i = W_0\Theta_i(\widetilde{R}_i - R_f) + W_{-1}\Theta_{i,-1}g_i. \tag{10}$$

The first two terms in (9) are the traditional mean-variance specification of preferences; γ measures aversion to portfolio-level risk. The third term in (9) is new, and captures prospect theory and narrow framing. This third term is the sum of N components, where the i'th component, $V(\widetilde{G}_i)$, corresponds to asset i. Specifically, $\widetilde{G}_i$ is the gain or loss on asset i, and $V(\widetilde{G}_i)$ is the cumulative prospect theory value of this gain or loss. The parameter b_0 controls the importance of the prospect theory term relative to the mean-variance terms.

The gain or loss on asset i, $\widetilde{G}_i$, is defined in (10). It is the sum of two terms. The first term, $W_0\Theta_i(\widetilde{R}_i - R_f)$, is the gain or loss between time 0 and time 1: specifically, it is the value of the investor's holdings of asset i at time 0 multiplied by the return on the asset in excess of the risk-free rate. For example, if the investor's holdings of asset 1 are worth \$100 at time 0, and the net asset return and risk-free rate between time 0 and time 1 are 20% and 2% respectively, then this first term is equal to $\$120 - \$102 = \$18$. We use the risk-free rate as a benchmark in part for tractability, but also because it may be psychologically plausible: the investor may think of the outcome of his investment in asset i as a gain only if this outcome is better than what he would have earned by investing in the risk-free asset.

The second term in (10), $W_{-1}\Theta_{i,-1}g_i$, is the gain or loss the investor experienced in his holdings of asset i *prior to* arriving at time 0: W_{-1} is the investor's wealth at some earlier time $t = -1$, $\Theta_{i,-1}$ is the fraction of wealth allocated to asset i at that earlier time, and g_i is the net return on asset i since that time. We think of g_i as the capital gain overhang on asset i: if the investor experienced a gain of 30% on asset i since purchase, then $g_i = 0.3$, while if he experienced a loss of 30%, then $g_i = -0.3$.

Equation (10) indicates that, at time 0, the investor merges the potential future gain or loss on asset i between time 0 and time 1 with his prior gain or loss on the asset, and computes the prospect theory value of this *overall* gain or loss. This assumption – an assumption about how the investor does his "mental accounting" – is consistent with a wide range of prior evidence that, when an individual has an ongoing investment in an asset, he integrates the potential future gain or loss in the asset with his past gain or loss (Andrikogiannopoulou and Papakonstantinou, 2018; Frydman, Hartzmark, and Solomon, 2018).

The second term on the right-hand side of (10) should be thought of as exogenous: for any asset i, we will empirically estimate the typical investor's prior gain or loss g_i in that asset. In addition, as we explain in Section 4, we will set $\Theta_{i,-1}$ to a neutral value, namely asset i's weight in the market portfolio. As such, $\Theta_{i,-1}$ is not a control variable that the investor chooses; the only control variable is Θ_i, the investor's allocation to asset i at time 0, and it features only in the first term in (10). Finally, we approximate W_{-1} by setting $W_{-1} = W_0$.[5]

We noted above that $V(\widetilde{G}_i)$ is the cumulative prospect theory value of the gain or loss $\widetilde{G}_i$. We can write $V(\widetilde{G}_i)$ in full as

$$\begin{aligned} &-\lambda W_0^{\alpha} \int_{-\infty}^{R_f - \Theta_{i,-1}g_i/\Theta_i} (\Theta_i(R_f - R_i) - \Theta_{i,-1}g_i)^{\alpha} dw(P(R_i)) \\ &-W_0^{\alpha} \int_{R_f - \Theta_{i,-1}g_i/\Theta_i}^{\infty} (\Theta_i(R_i - R_f) + \Theta_{i,-1}g_i)^{\alpha} dw(1 - P(R_i)). \end{aligned} \tag{11}$$

The expression in (11) uses a standard implementation of cumulative prospect theory for gambles with continuous distributions. The top row corresponds to losses, and is therefore multiplied by loss aversion λ. The bottom row corresponds to gains.

To complete the description of the investor's problem, we need to specify the probability

[5]We have also considered more accurate approximations, such as $W_{-1} = W_0/1.04$, where 4% is a measure of the historical average return on investor wealth. However, these alternative specifications have only a minor impact on our results; we therefore stick with the simpler approximation $W_{-1} = W_0$.

distribution $P(\widetilde{R})$ for asset returns. Since skewness plays an important role in our analysis, we need a distribution that can incorporate it in a flexible way. An attractive distribution for this purpose is the "generalized hyperbolic (GH) skewed t" distribution, a distribution that is increasingly seen as an appealing way of modeling skewness in asset returns. The probability density function for the multivariate form of this distribution is

$$p(R) = \frac{2^{1-\frac{\nu+N}{2}}}{\Gamma(\frac{\nu}{2})(\pi\nu)^{\frac{N}{2}}|S|^{\frac{1}{2}}}.\frac{K_{\frac{\nu+N}{2}}(\sqrt{(\nu+(R-\mu)'S^{-1}(R-\mu))\zeta'S^{-1}\zeta})\exp((R-\mu)'S^{-1}\zeta)}{\left(\sqrt{(\nu+(R-\mu)'S^{-1}(R-\mu))\zeta'S^{-1}\zeta}\right)^{-\frac{\nu+N}{2}}(1+(R-\mu)'S^{-1}(R-\mu)/\nu)^{\frac{\nu+N}{2}}}, \quad \text{for } \zeta \neq 0 \tag{12}$$

$$p(R) = \frac{\Gamma(\frac{\nu+N}{2})}{\Gamma(\frac{\nu}{2})(\pi\nu)^{\frac{N}{2}}|S|^{\frac{1}{2}}}.(1+(R-\mu)'S^{-1}(R-\mu)/\nu)^{-\frac{\nu+N}{2}}, \text{ for } \zeta = 0, \tag{13}$$

where $\Gamma(\cdot)$ is the Gamma function and K_l is the modified Bessel function of the second kind with order l.

The above distribution has four parameters: μ, S, ζ, and ν. Here, $\mu = (\mu_1, \ldots, \mu_N)$, the vector of location parameters, helps to determine the mean of the distribution; $S = \{S_{ij}\}$, the dispersion matrix, controls the dispersion in returns; $\zeta = (\zeta_1, \ldots, \zeta_N)$, the vector of asymmetry parameters, governs the skewness of returns; and ν, a degree of freedom scalar, affects the relationship between skewness and volatility. The mean of the distribution is given by

$$\mu + \zeta\frac{\nu}{\nu-2}. \tag{14}$$

When we apply the model in Sections 4 and 5, we will set three of the four parameters – S, ζ, and ν – to match the empirical volatility and skewness of asset returns, and will then search for values of $\mu = (\mu_1, \ldots, \mu_N)$ so that the market for each asset clears. The assets' expected returns – the quantities we want to determine – are then given by (14).

In our model, the expected return of asset i is significantly affected by the $V(\widetilde{G}_i)$ term in (9). This term, in turn, depends on the *marginal* distribution of asset i's return. This is a one-dimensional GH skewed t distribution, whose density function is

$$p(R_i) = \frac{2^{1-\frac{\nu+1}{2}}}{\Gamma(\frac{\nu}{2})(\pi\nu S_{ii})^{\frac{1}{2}}}.\frac{K_{\frac{\nu+1}{2}}(\sqrt{(\nu+(R_i-\mu_i)^2/S_{ii})\zeta_i^2/S_{ii}})\exp((R_i-\mu_i)\zeta_i/S_{ii})}{\left(\sqrt{(\nu+(R_i-\mu_i)^2/S_{ii})\zeta_i^2/S_{ii}}\right)^{-\frac{\nu+1}{2}}(1+(R_i-\mu_i)^2\nu^{-1}/S_{ii})^{\frac{\nu+1}{2}}}, \quad \text{for } \zeta \neq 0 \tag{15}$$

$$p(R_i) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})(\pi\nu S_{ii})^{\frac{1}{2}}}.(1+(R_i-\mu_i)^2\nu^{-1}/S_{ii})^{-\frac{\nu+1}{2}}, \text{ for } \zeta = 0. \tag{16}$$

The mean, variance, and skewness of this distribution are given by

$$E(\widetilde{R}_i) = \overline{R}_i = \mu_i + \frac{\nu}{\nu-2}\zeta_i \tag{17}$$

$$\text{Var}(\widetilde{R}_i) = \frac{\nu}{v-2}S_{ii} + \frac{2\nu^2}{(\nu-2)^2(\nu-4)}\zeta_i^2 \tag{18}$$

$$\text{Skew}(\widetilde{R}_i) = \frac{2\zeta_i\sqrt{\nu(\nu-4)}}{\sqrt{S_{ii}}(2\nu\zeta_i^2/S_{ii}+(\nu-2)(\nu-4))^{\frac{3}{2}}}\left[3(\nu-2)+\frac{8\nu\zeta_i^2}{\sigma_i^2(\nu-6)}\right]. \tag{19}$$

We think of the preference specification in (9) as having a "foundation" of mean-variance preferences, one that we then supplement with a prospect theory term. Why do we take mean-variance preferences to be the foundation, rather than some other Expected Utility specification? Our main application in this paper is to see if prospect theory can explain stock market anomalies, in other words, empirical deviations from the CAPM. As such, we want the *non*-prospect theory part of our preference specification to be one that generates CAPM pricing: this will allow us to clearly identify the deviations from the CAPM that prospect theory generates. The simplest specification that generates a CAPM is the mean-variance one.

3.2 Equilibrium structure

In this section, we discuss the structure of the equilibrium in our economy. Remarkably, the equilibrium structure that is almost always used when analyzing Expected Utility models no longer applies in a model like the one in (9). This is a fundamental challenge that arises when trying to understand prospect theory's implications for the cross-section, and one of our contributions is to propose a way of surmounting it – specifically, to propose a new equilibrium structure. Below, we lay out three possible types of equilibrium and explain why we study the one that we do.

Full rationality with homogeneous holdings. Our model features a representative agent. We think of this representative agent as standing in for a large number of identical investors – investors who all have the same wealth at time 0, the same preference parameters, and the same prior gain or loss in each risky asset. Given that the investors are identical, it is natural to think that, in equilibrium, they would all have exactly the same portfolio holdings $\{\Theta_i\}_{i=1}^N$, in other words, that they would each hold the market supply of each risky asset. Formally, such an equilibrium would consist of a location vector $\mu = (\mu_1, \ldots, \mu_N)$ such that,

for this μ, the objective function in (9) has a unique global maximum $\Theta^* = (\Theta_1^*, \ldots, \Theta_N^*)$ with $\Theta_i^* = \Theta_{M,i}$ for all i, where $\Theta_{M,i}$ is the market value of asset i divided by the total market value of all traded assets. This equilibrium structure, which we refer to as a "full-rationality equilibrium with homogeneous holdings," is the one used in Expected Utility models with a representative agent.

Remarkably, however, for the wide range of parameter values we have examined, this type of equilibrium does not exist for the model in (9). In Section 4, after parameterizing the model, we illustrate this non-existence with an example. For now, we explain it in general terms. Suppose that, for some value of μ_1, the location parameter for asset 1, the objective function in (9) is maximized at $\Theta_1 = \Theta_1^*$, where Θ_1^* exceeds the market supply $\Theta_{M,1}$ of the asset. This suggests that, in order to clear the market, we simply need to lower the value of μ_1. However, it turns out that as we do so, the value of Θ_1 for which the objective function in (9) attains its maximum jumps *discontinuously*, from a Θ_1^* that exceeds $\Theta_{M,1}$ to a Θ_1^* that lies below it. As such, there is *no* value of μ_1 for which the objective function is maximized at a Θ_1^* that equals $\Theta_{M,1}$. An equilibrium where investors have identical holdings for all assets therefore does not exist.

Full rationality with heterogeneous holdings. Given that the homogeneous-holdings equilibrium does not exist, what kind of equilibrium can we consider instead? One alternative equilibrium structure involves multiple global maxima. In other words, it may be that there exists a location vector $\mu = (\mu_1, \ldots, \mu_N)$ such that, for this μ, the objective function in (9) has multiple global maxima, and that by allocating the appropriate number of investors to each optimum, we can clear the market in each asset.

The difficulty with this equilibrium structure is that it is computationally infeasible to determine whether or not it exists. To see why, suppose that we consider 100 candidate values for each element of the location vector μ. This means that we need to try 100^N different location vectors μ. Moreover, since we are thinking of the risky assets as individual stocks, N is a large number, on the order of 1000. For each of the 100^N location vectors, we need to solve the problem in (9) and to determine if there are multiple global optima, something that, because of numerical error, is very hard to do: if two values of $\Theta = (\Theta_1, \ldots, \Theta_N)$ give similar values for the objective function, are they both global optima, or is one slightly better than the other, so that there is actually a unique optimum? Even if we manage to get this far, we then need to determine whether, by allocating investors to the various optima, we can clear the market. Overall, this procedure is computationally infeasible even for $N = 2$

risky assets, let alone for $N = 1000$, the more realistic value we use below.

Bounded rationality with heterogeneous holdings. To find a way around the difficulties described above, we introduce a mild bounded-rationality assumption, one that makes it feasible to find a heterogeneous-holdings equilibrium. Specifically, we assume that, when trying to determine the allocation Θ_i to asset i that maximizes the objective function in (9), an investor assumes that his holdings of the *other* $N-1$ risky assets equal the market supply of those assets – in other words, that $\Theta_j = \Theta_{M,j}$ for all $j \neq i$. This will not be exactly true – as we will see, investors' actual portfolios are less diversified than the market portfolio – but it will be approximately true: the return on the investor's actual holdings of the remaining $N-1$ assets is highly correlated with the return on the market portfolio.

We then define a bounded-rationality equilibrium with heterogeneous holdings as consisting of a location vector μ such that, for this μ, and under the bounded rationality assumption, the investor's solution to the problem in (9) has multiple global maxima, and by allocating each investor to one of the maxima, we can clear the market. More precisely, for each i in turn, we take the objective function in (9), view it as a function of Θ_i, and then – this is where the bounded-rationality assumption comes in – set $\Theta_j = \Theta_{M,j}$ for all $j \neq i$. Up to a linear transformation, the resulting function can be written:

$$\begin{aligned}
&\Theta_i(\mu_i + \frac{\nu\zeta_i}{\nu-2} - R_f) - \frac{\widehat{\gamma}}{2}(\Theta_i^2\sigma_i^2 + 2\Theta_i\sum_{j\neq i}\Theta_{M,j}\sigma_{ij}) \\
&-\lambda\widehat{b}_0\int_{-\infty}^{R_f-\Theta_{i,-1}g_i/\Theta_i}(\Theta_i(R_f-R_i)-\Theta_{i,-1}g_i)^\alpha dw(P(R_i)) \\
&-\widehat{b}_0\int_{R_f-\Theta_{i,-1}g_i/\Theta_i}^{\infty}(\Theta_i(R_i-R_f)+\Theta_{i,-1}g_i)^\alpha dw(1-P(R_i)),
\end{aligned} \tag{20}$$

where

$$\widehat{\gamma} = \gamma W_0 \tag{21}$$

$$\widehat{b}_0 = b_0 W_0^{\alpha-1}. \tag{22}$$

A bounded-rationality equilibrium with heterogeneous holdings consists of a location vector $(\mu_1, \ldots, \mu_N)$ such that, for each i, the function in (20) has either a unique global maximum at $\Theta_i = \Theta_{M,i}$, or has multiple global maxima, one of which lies below $\Theta_{M,i}$, and one of which lies above it.[6]

[6]The reader may wonder whether, under the bounded-rationality assumption, there is an equilibrium with *homogeneous* holdings. The conditions that must be satisfied for such an equilibrium to exist are the same as those that must be satisfied for a *full*-rationality equilibrium with homogeneous holdings to exist. Since the latter equilibrium does not exist, nor can the former.

In (20), Θ_i is the fraction of the investor's *total* wealth invested in asset i. When implementing the model, it is easier to work instead with the fraction of the investor's *risky asset* portfolio that is invested in asset i, a quantity we denote by θ_i. From now on, then, we think of the investor choosing θ_i rather than Θ_i. In a similar way, we define $\theta_{i,-1}$ to be the fraction of the investor's risky asset portfolio allocated to asset i at time -1, and $\theta_{M,i}$ to be the market value of asset i relative to the market value of all risky assets. In the Appendix, we show that, when reformulated with θ_i as the choice variable, the investor's decision problem has exactly the same form as in (20), subject only to a rescaling of $\widehat{\gamma}$ and $\widehat{b}_0$. We also show that, once the problem is reformulated in terms of risky portfolio shares, we can simplify the variance term in the first row of (20) by introducing asset i's beta, denoted β_i.

For clarity, we now restate the the definition of equilibrium in terms of θ_i. A bounded-rationality equilibrium consists of a location vector $\mu = (\mu_1, \ldots, \mu_N)$, such that, for each i, the function

$$\begin{aligned}
&\theta_i(\mu_i + \frac{\nu\zeta_i}{\nu-2} - R_f) - \frac{\widehat{\gamma}}{2}(\theta_i^2\sigma_i^2 + 2\theta_i(\beta_i\sigma_M^2 - \theta_{M,i}\sigma_i^2)) \\
&-\lambda\widehat{b}_0 \int_{-\infty}^{R_f - \theta_{i,-1}g_i/\theta_i} (\theta_i(R_f - R_i) - \theta_{i,-1}g_i)^\alpha dw(P(R_i)) \\
&-\widehat{b}_0 \int_{R_f - \theta_{i,-1}g_i/\theta_i}^{\infty} (\theta_i(R_i - R_f) + \theta_{i,-1}g_i)^\alpha dw(1 - P(R_i)),
\end{aligned} \tag{23}$$

where

$$\widehat{\gamma} = \gamma W_0 \theta_R \tag{24}$$

$$\widehat{b}_0 = b_0 W_0^{\alpha-1} \theta_R^{\alpha-1}, \tag{25}$$

and where θ_R is the fraction of the investor's overall wealth allocated to risky assets, has either a unique global maximum at $\theta_i = \theta_{M,i}$ or has multiple global maxima that straddle $\theta_{M,i}$.[7]

We find that this equilibrium exists for a wide range of parameter values, and it is the one we focus on. We note two things about it. First, in this equilibrium, investors need not have heterogeneous holdings for *all* the risky assets. Indeed, when we implement the equilibrium,

[7] In the type of model we are working with, two of the following three quantities – the investor's risk aversion γ, the fraction of overall wealth allocated to risky assets θ_R, and the equity premium – determine the third. When we implement the model, we will set $\widehat{\gamma}$ and $\widehat{b}_0$ to match the historical equity premium. Since $\widehat{\gamma}$ and $\widehat{b}_0$ are functions of γ and θ_R, we are effectively setting γ and θ_R and letting these determine the equity premium.

we find that, for many risky assets, investors have identical holdings. However, for at least one risky asset, they have different holdings, and this is what makes it a heterogeneous-holdings equilibrium. Second, for any asset i where investors do have heterogeneous holdings, the function in (23) has just two global maxima, θ_i^* and θ_i^{**}. These maxima straddle the market supply of $\theta_{M,i}$, so that $\theta_i^* < \theta_{M,i} < \theta_i^{**}$, and this allows us to clear the market in the asset by assigning some investors to the θ_i^* allocation and the rest to the θ_i^{**} allocation. In many cases, although not all, $\theta_i^* = 0$.

The bounded-rationality assumption significantly simplifies the investor's optimization problem: for each asset i, the search for the optimal allocation θ_i becomes a one-dimensional problem. Moreover, precisely because it is a one-dimensional problem, it is easy to determine whether the function in (23) has multiple global maxima or a unique global maximum. Later in the paper, we explain in full how we determine whether investors have identical or heterogeneous holdings in an asset, and how we then compute the asset's expected return.

4 Anomalies and Model Parameter Values

The goal of our paper is to lay out a way of generating quantitative predictions about the cross-section of average returns in an economy where investors evaluate risk according to prospect theory. To illustrate the broad applicability of our approach, we use the model of Section 3 to see if prospect theory can help us make sense of 22 prominent stock market anomalies. In setting the number of anomalies to 22, we are trying to be as comprehensive as possible, subject to the computational constraints we face. Table 1 lists the anomalies, along with the abbreviations we use in subsequent tables to refer to them. As noted in the Introduction, the 22 anomalies are intended to include those that, to date, have received the most attention from researchers. To construct the set of anomalies, we start with the 11 anomalies studied by Stambaugh, Yu, and Yuan (2012) and then add 11 more anomalies by choosing, from among the 97 anomalies studied by McLean and Pontiff (2016), those that appear to us to have received the most attention. The list is not based on any prior beliefs we have about whether prospect theory is helpful for explaining an anomaly – again, it is intended to be nothing more than a representative set of anomalies.

To see if our model can explain a particular anomaly, we proceed as follows. We consider an economy with $N = 1000$ stocks; each anomaly decile therefore contains 100 stocks. For

example, in the case of the size anomaly, assets 1 to 100 belong to size decile 1, assets 101 to 200 to size decile 2, and so on. We treat all stocks in a given decile as identical: we assign them the same characteristics – specifically, the empirical characteristics of the typical stock in that anomaly decile. We then examine what the model predicts about average stock returns, and, in particular, whether it can capture the empirical fact that the average return on small-cap stocks is higher than the average return on large-cap stocks. Note that, since all the assets in a given anomaly decile are identical, they have the same average return. For each anomaly, then, we need only compute the average return of ten stocks, consisting of one randomly-chosen stock from each decile. We say that our model can help explain the size anomaly if the average return it predicts for the randomly-chosen stock in decile 1 is higher than for the randomly-chosen stock in decile 10.[8]

What are the empirical inputs that we need in order to compute the average return of an asset in our model? In an economy with prospect theory investors who engage in narrow framing, an asset's average return will depend on the asset's return volatility, return skewness, and capital gain overhang. Since the investor's objective function in (9) also includes the traditional mean-variance terms, the asset's average return will also depend on the asset's beta. Therefore, to see what our model predicts about the average return of a stock in some anomaly decile, we need to know, for the typical stock in that decile, its return volatility, its return skewness, its gain overhang, and its beta. We estimate these crucial inputs from historical data. To explain how we do so, we focus on the example of the size anomaly; the process is the same for all the anomalies we consider.

Each month from July 1963 to December 2014, we rank all stocks listed on the NYSE, Amex, or Nasdaq on their market capitalization and then group them into deciles. (For each of the other anomalies, we instead rank stocks on the relevant anomaly characteristic – for example, on the past 12-month return in the case of the momentum anomaly.) Decile 1 corresponds to stocks with the lowest market capitalizations, and decile 10 to stocks with the highest market capitalizations. Suppose that, in some particular month, each decile contains 100 stocks. For each decile in turn, we compute the returns, over the next year, of the 100 stocks in that decile. We then compute the standard deviation of these 100 returns, and also their skewness. Also, for each of the 100 stocks, we compute its capital gain overhang – the

[8]Why do we not simply consider an economy with $N = 10$ assets, where each asset represents the typical stock in one of the anomaly deciles? The reason is that the expected return our model predicts for an asset depends on the asset's market weight. We therefore want to capture the fact that each stock makes up only a small fraction of the overall market.

percentage gain or loss in the stock for the average investor in the stock – and then average these 100 numbers. For each of the 100 stocks, we then compute its beta and average these 100 numbers. Finally, for each of the 100 stocks, we compute the return of the stock over the next month, and then average these 100 returns. At the end of this process, we have five quantities in hand for each anomaly decile in that month: the volatility, skewness, gain overhang, beta, and average return for the typical stock in that decile.

We repeat the above exercise for each month in our sample. This gives us, for each market-capitalization decile, a time series for each of the five quantities: return volatility, return skewness, capital gain overhang, beta, and average return. In the final step, we compute the mean of each time series. For each market capitalization decile, this leaves us with five numbers corresponding to the typical stock in that decile: the standard deviation of its returns; the skewness of its returns; its capital gain overhang; its beta; and its average return. We feed the first four of these numbers into our model to see if it can match the fifth number, the average return, which is the one we are trying to understand.

Before presenting the empirical characteristics of the 22 anomalies, we clarify the computation of one key variable: the capital gain overhang for a stock. There are two approaches to computing this quantity, one due to Grinblatt and Han (2005) and the other to Frazzini (2006). Grinblatt and Han (2005) use data on a stock's past trading volume to estimate how long each of the investors in the stock has been holding it; this then allows them to compute a measure of the average investor's gain or loss in the stock. Frazzini (2006) uses data on mutual fund holdings to directly compute the average gain or loss in a stock across the mutual fund investors in that stock. Frazzini's data source, the Thomson-Reuters database, also includes information on the holdings of other institutional investors. Another variant of the Frazzini measure of capital gain overhang uses this additional data to compute the average gain or loss in a stock across *all* institutional investors in the stock.

In the empirical analysis described above, we computed all three of these measures of capital gain overhang – the Grinblatt and Han (2005) measure and the two Frazzini (2006) measures. The three methods lead to similar quantitative estimates of capital gain overhang, and the asset pricing results we present in Section 5 are not particularly sensitive to which of the three measures we use. We therefore pick one, the Grinblatt and Han (2005) measure, and stick with it throughout.

Tables 2 and 3 present the results of the above empirical exercise. The first column of

Table 2 lists the 22 anomalies we are interested in. The second and third columns of the table list, for each anomaly, the average return of the typical stock in decile 1 and decile 10, respectively. By definition of what an anomaly is, these average returns differ in a way that is not captured by beta. The fourth and fifth columns of the table report the skewness of the return of the typical stock in deciles 1 and 10, respectively, computed as described above. The second and third columns of Table 3 list the standard deviation of the return of the typical stock in deciles 1 and 10, respectively. Finally, the fourth and fifth columns of Table 3 report the capital gain overhang of the typical stock in deciles 1 and 10, respectively.

We make three observations about the results in Tables 2 and 3. First, the tables show that, for most of the anomalies, the typical stock in decile 1 differs significantly from the typical stock in decile 10 in its return skewness, standard deviation of returns, and gain overhang. Consider, for example, the size anomaly: the typical stock in decile 1 has a standard deviation of 76%, while the typical stock in decile 10 has a standard deviation of just 25%. Similarly, while the typical small-cap stock has a skewness of 4.3, the typical large-cap stock has skewness of 0.7. And while the typical small-cap stock has a negative gain overhang of -15%, the typical large-cap stock has a positive gain overhang of 17%.

The fact that, for most anomalies, deciles 1 and 10 differ substantially in their return skewness, standard deviation of returns, and gain overhang means that our model is likely to make strong predictions about the average returns of the stocks in these deciles – predictions that can be compared to the data. Put differently, if the stocks in the two deciles had similar levels of standard deviation, skewness, and gain overhang, then our model would predict similar average returns for the two deciles, and would therefore not be helpful in explaining the anomalies.

The second observation is that the three characteristics we are focused on – standard deviation, skewness, and gain overhang – are strongly correlated across anomaly deciles: for 21 of the 22 anomalies, if the typical stock in decile 1 has a higher return skewness than the typical stock in decile 10, then it also has a higher standard deviation; the only exception is for post-earnings announcement drift. Furthermore, for 21 of the 22 anomalies – the only exception is the net operating assets anomaly – if the typical stock in decile 1 has a higher return skewness than the typical stock in decile 10, then it also has a more negative gain overhang.

Figure 2 illustrates these relationships more clearly. Consider the top-left graph in the

figure. Each asterisk in the graph corresponds to an anomaly decile; since there are 22 anomalies, this makes for a total of 220 asterisks. The horizontal and vertical axes in the graph measure the standard deviation and skewness, respectively, of the typical stock in an anomaly decile. The graph clearly shows the positive correlation between these two quantities. In a similar way, the other two graphs show the negative correlation, across anomaly deciles, between skewness and gain overhang, and between volatility and gain overhang.

The empirical relationships in Figure 2 point to the necessity of the quantitative approach we are taking in this paper. Suppose that one of the extreme decile portfolios – decile 1, say – has a higher return skewness, higher standard deviation, and lower capital gain overhang than the typical stock in the other extreme decile, decile 10. Note again that 21 of the 22 anomalies follow this pattern. Notice that, in this case, it is not possible to tell, without a quantitative model, whether prospect theory will be able to explain the anomaly. The reason is that there are counteracting forces. Decile 1 stocks have more volatile returns than decile 10 stocks. Since prospect theory investors are loss averse, this will lead them, all else equal, to require a higher average return on decile 1 stocks than on decile 10 stocks. However, decile 1 stocks also have more skewed returns than decile 10 stocks. Since prospect theory investors exhibit probability weighting, this will lead them, all else equal, to charge a lower average return on decile 1 stocks. Finally, decile 1 stocks trade at a loss relative to purchase price, while decile 10 stocks trade at a gain. Due to the diminishing sensitivity component of prospect theory, this will lead prospect theory investors, all else equal, to require a lower average return on decile 1 stocks. Since two of these forces go in one direction, and the other goes in the other direction, we need a quantitative model to determine whether prospect theory can explain the anomaly – a model of asset prices under prospect theory that takes the empirical values of return volatility, return skewness, and capital gain overhang and uses them to predict an average return.

So far, we have noted that, for almost all of the anomalies, one of the extreme decile portfolios has higher skewness, higher standard deviation, and a lower gain overhang than the other extreme decile. Our third observation about the results in Tables 2 and 3 is that the extreme decile with the higher skewness and volatility and lower gain overhang usually also has a lower average return. This is the case for 13 of the 22 anomalies. For just five of the anomalies does the extreme decile with the lower average return have substantially lower skewness. And for the remaining four anomalies, the skewness of the two extreme deciles is similar. We return to this observation in Section 5.

The empirical results in Tables 2 and 3 are incorporated into the model through the values we assign the model parameters. We now explain how we set these parameter values.

4.1 Model parameters

To see whether our model can capture a particular anomaly, we proceed as follows. We consider an economy with $N = 1000$ assets, and assign 100 of these stocks to each anomaly decile: assets 1 to 100 belong to anomaly decile 1, assets 101 to 200 to anomaly decile 2, and so on. For any given decile, we take all the stocks in that decile to be identical: they have the same standard deviation, skewness, and capital gain overhang, namely the empirical standard deviation, skewness, and capital gain overhang of the typical stock in that anomaly decile, computed as described in the previous section. We set the parameters S, ζ, and ν of the GH skewed t distribution to capture these empirical values. We then search for a location vector μ so that the conditions for equilibrium described in Section 3.2 around equation (23) are satisfied. Our model's prediction for the assets' average returns is then given by (14). Note that all stocks in a given decile will have the same μ_i and hence the same average return.

We now explain in more detail how we parameterize the model. While the model features several parameters, all of the important parameters are disciplined by either field data or experimental data. The asset-level parameters are R_f, the gross risk-free rate; N, the number of risky assets; S, the dispersion matrix for asset returns; ζ, the vector of asymmetry parameters for asset returns; ν, the degree of freedom parameter; g_i, asset i's capital gain overhang; and θ_M, the market supply of the risky assets. The investor-level parameters are γ, portfolio risk aversion; b_0, the relative importance of the prospect theory term in investor preferences; $(\alpha, \delta, \lambda)$, the prospect theory preference parameters; and θ_{-1}, the investor's prior allocation to the risky assets.

We start with the asset-level parameters, and in particular, the parameters of the GH skewed t distribution. We set $\nu = 7.5$; this allows an asset's skewness to increase substantially with its volatility. We then set the diagonal elements of the dispersion matrix, $\{S_{ii}\}_{i=1}^{N}$, and the elements of the asymmetry vector $\zeta = (\zeta_1, \ldots, \zeta_N)$. To do this, recall from equations (18) and (19) that, for the GH skewed t distribution,

$$std(\widetilde{R}_i) \quad = \quad \left[\frac{\nu}{\upsilon - 2} S_{ii} + \frac{2\nu^2}{(\nu-2)^2(\nu-4)}\zeta_i^2\right]^{0.5} \tag{26}$$

$$skew(\widetilde{R}_i) = \frac{2\zeta_i\sqrt{\nu(\nu-4)}}{\sqrt{S_{ii}}(2\nu\zeta_i^2/S_{ii}+(\nu-2)(\nu-4))^{\frac{3}{2}}}\left[3(\nu-2)+\frac{8\nu\zeta_i^2}{S_{ii}(\nu-6)}\right]. \quad (27)$$

To set S_{ii} and ζ_i for a stock i that belongs to a particular anomaly decile, we take the empirical standard deviation and skewness of the typical stock in that anomaly decile, and plug them into the left-hand side of equations (26) and (27). These two equations then allow us to solve for the two unknowns, S_{ii} and ζ_i. For example, in the case of the size anomaly, assets 1 to 100 belong to the lowest market capitalization decile. From Tables 2 and 3, we see that the empirical standard deviation and skewness for the typical stock in this decile are 0.76 and 4.3, respectively. Accordingly, for this anomaly, to set the values of S_{ii} and ζ_i for $i \in \{1,\ldots,100\}$, we solve

$$0.76 = \left(\frac{7.5}{7.5-2}S_{ii}+\frac{2(7.5)^2}{(7.5-2)^2(7.5-4)}\zeta_i^2\right)^{0.5} \quad (28)$$

$$4.3 = \frac{2\zeta_i\sqrt{7.5(7.5-4)}}{\sqrt{S_{ii}}(2(7.5)\zeta_i^2/S_{ii}+(7.5-2)(7.5-4))^{\frac{3}{2}}}\left[3(7.5-2)+\frac{8(7.5)\zeta_i^2}{S_{ii}(7.5-6)}\right]. \quad (29)$$

We do not need to set the off-diagonal elements of S. Under our bounded-rationality assumption, all the information we need about asset i's covariance with other assets is contained in its beta. For an asset i that belongs to some anomaly decile, we set its beta equal to the empirical beta of the typical stock in that decile, computed as described in the previous section.

In terms of asset-level parameters, this leaves only g_i, the capital gain overhang; θ_M, the market supply of the risky asset; and the risk-free rate R_f. For an asset i in some anomaly decile, we set g_i to the empirical capital gain overhang of the typical stock in that anomaly decile, computed as described earlier in this section and displayed in Tables 2 and 3 for the extreme deciles of each anomaly. For θ_M, we start with neutral values of $\theta_{M,i} = 1/N$ for all i, but also consider empirical values. Finally, we set the risk-free rate R_f to 1.

We now turn to the investor-level parameters. We set $\widehat{\gamma}$, the scaled portfolio-level risk aversion in (24), and $\widehat{b}_0$, the scaled weight on the prospect theory term in (25), to generate an aggregate equity premium of around 6%. There is some flexibility here: we can achieve a high equity premium by increasing either $\widehat{\gamma}$ or $\widehat{b}_0$. Our quantitative predictions for average returns are similar in either case. We therefore set both $\widehat{\gamma}$ and $\widehat{b}_0$ to 0.6. We set $\theta_{i,-1}$, the investor's allocation to asset i at time -1, equal to $\theta_{M,i}$, the weight of asset i in the market portfolio, which, as noted above, equals $1/N$.

The three prospect theory preference parameters $(\alpha, \delta, \lambda)$ play an important role in our analysis, so we set them carefully. One well-known set of estimates comes from Tversky and Kahneman (1992), who report $(\alpha, \delta, \lambda) = (0.88, 0.65, 2.25)$ for their median subject. However, in the wake of Tversky and Kahneman's work, several other studies have also estimated these parameters, and it seems reasonable to set α, δ, and λ to values that are consistent with the average estimates across all these studies. Booij, van Praag, and van de Kuilen (2010) tabulate the estimates of α, δ, and λ across many studies. Tversky and Kahneman's (1992) estimate of $\delta = 0.65$ is similar to the average estimate across the full range of studies; we therefore adopt this value in our analysis. However, their estimates of $\alpha = 0.88$ and $\lambda = 2.25$ are high relative to the findings of other studies. In our analysis, we use $\alpha = 0.7$ and $\lambda = 1.5$, which are closer to the median estimate across other studies.

By setting the asset-level parameters – specifically, S, ζ, and g_i – in the way described above, we are assuming that investors have rational beliefs: they are aware of the volatility and skewness of a given stock's future return distribution, and of its gain overhang. In our view, this is a plausible benchmark case to study. An investor can obtain a good estimate of a stock's future return volatility simply by looking at its past return volatility, or at the past return volatility of stocks with similar characteristics. In addition, a stock's past return is a good proxy for its gain overhang. It may be more difficult for an investor to estimate a stock's future return skewness – a stock's past return skewness is a poor predictor of its future return skewness. However, an investor can obtain a better estimate of future skewness by looking at other stock characteristics that predict skewness – past volatility, for example; by looking at the cross-sectional skewness of the past returns of similar stocks; or by using economic intuition: for example, since the prices of distressed stocks have typically been beaten down, it is natural to conclude that these stocks have limited downside but a large upside.

While the case of rational beliefs strikes us as a good benchmark, it is possible that, for some anomalies, investors do not have sensible beliefs. In Section 6, we use evidence on investor beliefs derived from option prices to suggest that, while investors often do have sensible beliefs, they occasionally do not, and that this may help explain some of the model's failures.

4.2 Illustration of the equilibrium structure

The investors in our model are identical in their preferences and in the past gains and losses they have experienced in each of the N stocks. Nonetheless, as noted in Section 3.2, they do not hold identical portfolios in equilibrium: an equilibrium with homogeneous holdings does not exist. Instead, the equilibrium we study involves heterogeneous holdings.

Now that we have parameterized the model, we can illustrate these heterogeneous holdings. In the Appendix, we provide a full description of how we determine whether investors have identical or heterogeneous holdings for an asset, and how we then compute its expected return. We illustrate the equilibrium for the case of the momentum anomaly. We note that it is only for *some* of the risky assets that the investors have heterogenous holdings. For other assets, they have identical, homogeneous holdings; in other words, they each hold the supply of the risky asset, $\theta_{M,i} = 1/N$. This is the case for all stocks in momentum decile 1, namely stocks 1 to 100. Specifically, for stock 1, there is a value of μ_1, the location parameter for asset 1, such that the function in (23), for $i = 1$, has a unique global maximum at $\theta_1 = \theta_{M,1}$. Figure 3 plots the function in (23) for this value of μ_1, namely $\mu_1 = 0.0337$. The global maximum at $\theta_1^* = \theta_{M,1} = 1/N = 0.001$ is clearly visible.

For the stocks in momentum decile 10, namely stocks 901 to 1000, investors do not have identical holdings. For example, for stock 901, there is no value of μ_{901} such that the function in (23), for $i = 901$, has a global maximum at $\theta_{901} = \theta_{M,901} = 0.001$. We can see this in Figure 4. The dashed line plots the function in (23) for $\mu_{901} = 0.585$. For this value of μ_{901}, the function has a unique global maximum at $\theta_{901} = 0.167 > \theta_{M,901}$. Since demand for asset 901 exceeds supply at this value of μ_{901}, it appears that, to clear the market, we need to lower the value of μ_{901}. However, as we do so, the value of θ_{901} at which the function attains its maximum jumps discontinuously downward: the dashed-dot line, which plots the objective function for a slightly lower value of μ_{901}, namely 0.576, shows that the unique maximum is now at $\theta_{901} = 0 < \theta_{M,901}$.

The figure also shows, however, that there is a heterogeneous-holdings equilibrium. The solid line plots the function for an intermediate value of μ_{901}, namely 0.581. For this value of μ_{901}, there are *two* global optima, one at $\theta_{901}^* = 0 < \theta_{M,901}$ and one at $\theta_{901}^{**} = 0.129 > \theta_{M,901}$. Since these optima straddle the market supply of $\theta_{M,901} = 0.001$, we can clear the market by allocating most investors to the first optimum, $\theta_{901}^* = 0$, and the rest to the other optimum.

Figures 3 and 4 illustrate a pattern that holds almost without exception across all the anomalies we consider: investors have identical holdings in stocks that are trading at a loss – stocks for which $g_i < 0$ – but heterogeneous holdings in stocks that are trading at a gain, for which $g_i > 0$. Why is this the case? Note first that the expression in (23) is a function of θ_i, θ_i^2, and θ_i^α, where the θ_i and θ_i^2 terms come from the mean-variance component of preferences, while the θ_i^α term comes from the prospect theory component. Since $\alpha < 1$, it is the prospect theory term that drives the behavior of (23) for small positive values of θ_i.

Now consider a stock trading at a gain. If investors are to have identical holdings in this stock, the expression in (23) must attain a global maximum for a very small value of θ_i, namely $\theta_{M,i}$. However, for this stock, the investor is in the concave gain region of the value function where he is very risk averse. As a consequence, as we increase θ_i, starting from $\theta_i = 0$, the objective function initially falls sharply in value. This, in turn, means that it is unlikely that the expression in (23) can attain a maximum for a small positive θ_i. Of course, we can try to overcome this effect by raising the stock's expected return. However, the expected return needed to overcome it is so high that, for this expected return, the investor would want to take a *large* positive position in the stock, one that is much higher than the market supply $\theta_{M,i}$. As a result, in equilibrium, investors cannot have identical holdings in this stock; rather, they have heterogeneous holdings.

Conversely, for a stock trading at a loss, it is much more likely that the equilibrium involves identical holdings. The convexity of the value function in the region of losses means that, as we increase θ_i starting from $\theta_i = 0$, the objective function in (23) initially increases sharply. This makes it more likely that the function can attain a unique maximum for a low value of θ_i.

5 Results

We now present our main results – our findings as to whether our model can help explain the 22 anomalies in Table 1.

To determine whether our model can "explain" an anomaly, we focus on anomaly alphas. For any given anomaly, we compute the empirical alphas for the ten anomaly deciles, namely $\alpha^d(1), \ldots, \alpha^d(10)$, where the "$d$" superscript denotes "data." We then compute the alphas

predicted by our model for each of the ten deciles, namely $\alpha^m(1), \ldots, \alpha^m(10)$, where "$m$" stands for "model." Since, within each decile, all stocks are identical and, in particular, have the same expected return and alpha, we can compute the alpha of decile l as the alpha of any stock in that decile – for example, as the alpha of stock $100l$:

$$\alpha^m(l) = \overline{R}_{100l} - (R_f + \beta_{100l}(\overline{R}_M - R_f)),$$

where $\overline{R}_M = \theta'_M \overline{R}$.

We say that our model can help explain an anomaly if

$$sign(\alpha^d(10) - \alpha^d(1)) = sign(\alpha^m(10) - \alpha^m(1)) \text{ and } |\alpha^m(10) - \alpha^m(1)| > 0.015. \quad (30)$$

The first condition in (30) is that the model correctly predicts the sign of the difference between $\alpha^d(10)$ and $\alpha^d(1)$, in other words, predicts that $\alpha(10) > \alpha(1)$ if this is empirically the case and that $\alpha(10) < \alpha(1)$ if that is empirically the case. The second condition in (30) is that the model makes a "strong" prediction, in other words, predicts a substantial difference between the two alphas. Similarly, we say that the model fails to explain an anomaly if

$$sign(\alpha^d(10) - \alpha^d(1)) = -sign(\alpha^m(1) - \alpha^m(1)) \text{ and } |\alpha^m(10) - \alpha^m(1)| > 0.015, \quad (31)$$

in other words, if the model makes a strong prediction but this prediction is incorrect, for example predicting that $\alpha(10) > \alpha(1)$ when the opposite is true in the data. Finally, we say that the model does not make a strong prediction about an anomaly if

$$|\alpha(10) - \alpha(1)| < 0.015. \quad (32)$$

We find that our model is helpful for thinking about a majority of the anomalies we consider; we review these anomalies in Section 5.1. In Section 5.2, we discuss the anomalies where our model performs poorly. And in Section 5.3, we note the anomalies where our model does not make a strong prediction.

5.1 Anomalies where prospect theory performs well

Our model is helpful for thinking about 13 of the 22 anomalies: for these anomalies, the conditions in (30) are satisfied. Specifically, these are the momentum, failure probability, idiosyncratic volatility, profitability, idiosyncratic skewness, return on assets, MAX, O-Score,

external finance, composite equity issuance, net stock issuance, post-earnings announcement drift, and difference of opinion anomalies. We present our model's predictions about these anomalies in Figures 5 through 11. Figures 5 to 9 showcase five anomalies that the model makes particularly strong predictions about. Figures 10 and 11 present the results for eight other anomalies where the model also appears to be helpful.

Figure 5 shows the results for the momentum anomaly. The horizontal axis corresponds to the ten decile portfolios, 1 through 10, while the vertical axis measures their alphas. The graph has two lines on it. The dashed line plots the empirical alphas of each decile; the fact that it is upward sloping indicates that, historically, and controlling for beta, stocks with high past 12-month returns earned a higher average return than stocks with low past 12-month returns. The solid line plots the alphas predicted by our model for the ten anomaly deciles.

Figures 6 through 9 present results for the failure probability, idiosyncratic volatility, profitability, and idiosyncratic skewness anomalies. These graphs have the same structure as Figure 5: the dashed lines plot the empirical alphas, while the solid lines plot the model-predicted alphas. Figures 10 and 11 present results for eight more anomalies, again using the same format.

Figures 5 through 11 show that, for all 13 anomalies in these graphs, our model predicts a lower alpha for the extreme decile that, empirically, has a lower alpha. The intuition for why our model "works" for these anomalies is the following. For these anomalies, the extreme decile with the lower empirical alpha contains stocks with more volatile returns, more skewed returns, and a more negative capital gain overhang. On the one hand, the higher volatility of these stocks leads the investors in our economy to charge a higher average return on them. On the other hand, their higher skewness and more negative gain overhang leads investors to charge a lower average return on them. The latter force dominates, so that the model predicts a low average return on these stocks, consistent with the data.

The results in Figures 5 to 11 highlight the necessity of our quantitative approach: given that some stock characteristics help prospect theory explain these anomalies, while other characteristics hinder its ability to do so, it is only by using a quantitative model that we can see whether, once all the characteristics are accounted for, prospect theory can explain the anomalies.

The overarching contribution of our paper is to show how prospect theory can be used to generate quantitative predictions about the cross-section of average returns in a way that is broadly applicable. Our analysis goes beyond previous efforts on several dimensions – in the way we consider all elements of prospect theory, including probability weighting; in our use of empirical values of stocks' volatility, skewness, and gain overhang as inputs to the model; and in the wide scope of our application, where we consider 22 different anomalies.

Our analysis also leads to more specific insights. Prior studies have linked the diminishing-sensitivity component of prospect theory to the some of the anomalies we consider – specifically, to momentum and post-earnings announcement drift (Grinblatt and Han, 2005; Frazzini, 2006; Li and Yan, 2013). For example, in the case of momentum, the idea is that stocks in momentum decile 10 have capital gains which bring investors into the concave, risk averse region of the value function, leading them to charge a higher average return on these stocks. Our analysis confirms this idea, but also shows that it is incomplete in an important way. Stocks in momentum decile 10 have not only a capital gain, but are also less volatile and less skewed than stocks in momentum decile 1 – characteristics that, due to loss aversion and probability weighting, *also* affect the average return that prospect theory predicts for momentum deciles. Our analysis shows that, once we take all three characteristics – volatility, skewness, and gain overhang – into account, prospect theory can indeed explain the momentum anomaly, but it is only through a quantitative framework like the one we develop in this paper that this conclusion can be drawn.

Similarly, prior work has linked the probability-weighting component of prospect theory to some of the anomalies we consider – specifically, to the idiosyncratic volatility, skewness, and failure probability anomalies (Boyer, Mitton, and Vorkink, 2010). The idea is that, for these anomalies, stocks in decile 10 have positively-skewed returns, which, due to probability weighting, leads investors to charge a lower average return on them. Our analysis confirms that this mechanism helps to explain these anomalies, but also indicates that the argument is incomplete. Highly skewed stocks also tend to have more volatile returns and to trade at a loss – characteristics that, in a prospect theory framework, will also affect their average returns. We show that once all three features are taken into account, prospect theory can explain the above anomalies – but, again, it is only through our quantitative approach that this can be confirmed.

To underscore the importance of taking all three of an asset's volatility, skewness, and capital gain into account when computing prospect theory's prediction for its average return,

we take one particular anomaly – the failure probability anomaly – and isolate the contribution that each characteristic makes to the overall expected return. We present the results in Figure 12. For ease of interpretation, we work with average returns, rather than alphas, in this figure.

The solid line in Figure 12 plots the average return predicted by our model for each failure probability decile when we take into account all three stock characteristics: volatility, skewness, and gain overhang. By contrast, the dashed line shows the average returns predicted by our model when we take into account each stock's volatility and skewness, but not its gain overhang. To generate this line, we use the same parameter values as for the solid line, with one exception: we set the capital gain g_i of stock i equal to zero, for all i. Finally, the dash-dot line shows the average return predicted by the model when we take into account only a stock's volatility, but not its skewness or gain overhang. To produce this line, we again set the gain overhang g_i for stock i equal to zero for all i, but also set the asymmetry parameter ζ_i equal to zero for all i; this fixes the skewness of all assets at zero. To set S_{ii} – the diagonal element of the dispersion matrix that corresponds to asset i – we solve equation (26) while imposing $\zeta_i = 0$. All other parameters take the same values as for the solid line.

The dash-dot line in Figure 12 shows that, when we account only for volatility, the model predicts, counterfactually, a much higher average return for stocks with high failure probability: these stocks have more volatile returns; since investors are loss averse, they require a higher average return on them. The dashed line shows that accounting for skewness lowers the predicted average return for all deciles: the typical stock in each decile is positively-skewed, a characteristic that, due to probability weighting, prospect theory investors find appealing. This effect is much more pronounced for high failure probability stocks because these stocks have particularly skewed returns. Finally, the solid line shows that further accounting for the capital gain has little impact on the predicted average return for stocks in deciles 1 to 6: for these deciles, the capital gain of the typical stock is close to zero. However, it has a major impact on the predicted average return for stocks in deciles 7 to 10: these stocks trade at a large loss, which, due to diminishing sensitivity, leads investors to charge a much lower average return on them.

Overall, Figure 12 shows that all three of an asset's volatility, skewness, and gain overhang have a significant effect on the average return that prospect theory predicts for that asset, and that it is therefore important to take all of them into account.

Finally, we emphasize that Figures 5 to 11 draw a connection between prospect theory and some anomalies that, to our knowledge, has not previously been noted. For example, they show that prospect theory is helpful for thinking about the profitability, return on assets, external finance, composite equity issuance, net stock issuance, and difference of opinion anomalies anomalies – again, anomalies that prospect theory has not previously been linked to.

5.2 Anomalies where prospect theory performs poorly

For seven of the 22 anomalies, the model performs poorly, in that, as laid out in the conditions in (31), it predicts a substantial difference between the alphas for deciles 1 and 10, but of the wrong sign. These are the size, value, long-term reversal, short-term reversal, accrual, asset growth, and investment anomalies. We discuss two of these – the size and value anomalies – in more detail, and the others more briefly.

Figure 13 presents the results for the size anomaly. As in the earlier graphs, the dashed line plots the empirical alphas, while the solid line plots the alphas predicted by the model. The graph shows that our model has mixed success in explaining this anomaly. On the one hand, the model captures, in part, the decline in the empirical alphas as we move from decile 2 to decile 10. On the other hand, it completely fails to explain the positive historical alpha for decile 1, which contains the stocks with the lowest market capitalizations. These stocks are more volatile, which, all else equal, leads investors to charge a higher average return on them. However, they also have positively-skewed returns and trade at a loss, which leads investors to charge a lower average return on them. Our analysis shows that this second effect overwhelms the first one.

Figure 14 presents the results for the value anomaly. The model fails to explain this anomaly. The reason is that value stocks are more positively-skewed than growth stocks, and trade at a larger loss, both of which lead investors to charge a lower average return on them. It is true that value stocks are also more volatile than growth stocks, which, all else equal, leads investors to charge a higher average return on value stocks. However, the figure shows that the second effect is overwhelmed by the first one.

Figure 15 and the top-left panel in Figure 16 present results for five other anomalies that our model performs poorly on – the long-term reversal, short-term reversal, accrual, asset

growth, and investment anomalies.

5.3 Anomalies where prospect theory does not make a strong prediction.

For two other anomalies – the net operating assets and organizational capital anomalies – prospect theory does not make a strong prediction, in the sense that, as specified in (32), it predicts alphas for the two extreme decile portfolios that differ by less than 1.5% in absolute magnitude. The lower panel of Figure 16 presents results for these anomalies.

6 Discussion

In this section, we discuss some other issues raised by our analysis.

6.1 Investor beliefs

While our model is helpful for thinking about a majority of the anomalies in Table 1, Figures 13 to 16 show that, for some of the anomalies, our model performs poorly. Why is this? The simplest answer is that, for some anomalies, other forces play a more important role than prospect theory. For example, for the value and long-term reversal anomalies, extrapolative beliefs about cash flows or returns may be more important than any aspect of preferences, while for the short-term reversal anomaly, liquidity-driven price pressure may be the primary driver.

There is a second answer, however, one that applies within our prospect theory framework. In generating Figures 5 to 16, we assumed that investors have accurate beliefs about stocks' volatility, skewness, and gain overhang. However, it may be that, for certain types of stocks, investors' beliefs about these quantities are incorrect.

To investigate this, we obtain data on stock option prices from the Ivy Options database, and use these to extract estimates of risk-neutral volatility and risk-neutral skewness; see the Appendix for a description of our methodology. While these are estimates of risk-neutral

quantities, they may contain some information about beliefs, and have previously been used for this purpose in helpful ways (Birru and Wang, 2016).

Table 4 presents, for each anomaly, the risk-neutral variance and risk-neutral skewness of the typical stock in each of the two extreme anomaly deciles. By comparing these results to those in Tables 2 and 3, we see that, for most types of stocks, investors appear to have sensible beliefs: the risk-neutral volatility and skewness are higher for stocks whose actual return volatility and skewness are higher. There is one notable exception however: the risk-neutral skewness of value stocks is similar to that of growth stocks, even though the *actual* skewness of value stock returns is much higher than that of growth stock returns.

This last finding may help explain why our model performs poorly on the value anomaly. To generate Figure 14, we assumed that investors *know* that value stocks have more positively-skewed returns than growth stocks; this, in turn, leads the model to predict a lower average return on value stocks. However, if investors think that it is actually growth stocks that have the more positively-skewed returns, this will lead our model to predict a lower return on growth stocks, reducing the gap between the empirical and model-predicted alphas.

6.2 Sensitivity to the values of the prospect theory parameters

As discussed in Section 4, for our main results, we use the prospect theory parameter values

$$(\alpha, \delta, \lambda) = (0.7, 0.65, 1.5). \tag{33}$$

Figure 17 shows, in the context of the failure probability anomaly, how our results are affected when we adjust these values. For ease of interpretation, we work with average returns rather than alphas in this figure.

The solid line in Figure 17 shows the average returns predicted by our model for the benchmark parameter values in (33). The dotted line shows the results when we keep α and δ fixed at the values in (33), but increase λ from 1.5 to 2. The dash-dot line shows the results if we keep α and λ fixed at the values in (33), but lower δ from 0.65 to 0.6. Finally, the dashed line shows the results if we keep δ and λ fixed at the values in (33), but lower α from 0.7 to 0.65.

The results are very consistent with the intuition laid out earlier in the paper. Comparing the dotted and solid lines, we see that raising investor loss aversion leads to an increase in the

required returns on all stocks, but especially for the most volatile stocks – those in decile 10. Comparing the dash-dot and solid lines, we see that strengthening the degree of probability weighting leads to lower returns for all stocks, but particularly for the most skewed stocks, those in decile 10. Finally, comparing the dashed and solid lines, we see that strengthening the degree of diminishing sensitivity raises the average return of stocks in decile 1 – stocks with gains – but lowers the average return of stocks in decile 10, which have capital losses.

7 Conclusion

Prospect theory was developed more than 30 years ago, but its predictions for basic aspects of asset prices such as the cross-section of average returns are still not well understood. In this paper, we build a new model of asset prices in which investors evaluate risk in part according to prospect theory, and show how the model can be used to make quantitative predictions about average stock returns. We then examine the model's ability to explain 22 prominent stock market anomalies. We find that the model is helpful for thinking about a majority of the anomalies we consider, including the momentum, volatility, distress, and profitability anomalies. It performs most poorly on the value anomaly. We present suggestive evidence that this is due to investors having incorrect beliefs about the distribution of value stock returns.

8 Appendix

8.1 Reformulating the decision problem in terms of risky portfolio shares

Let θ_R be the fraction of the investor's portfolio that is allocated to risky assets, and θ_i the fraction of his risky asset portfolio allocated to asset i, so that $\Theta_i = \theta_R \theta_i$. If we substitute this into (9), set $\theta_j = \theta_{M,j}$ for all $j \neq i$, and focus only on terms in θ_i, then, up to a linear transformation, we obtain

$$\theta_i(\mu_i + \frac{\nu\gamma_i}{\nu - 2} - R_f) - \frac{\widehat{\gamma}}{2}(\theta_i^2\sigma_i^2 + 2\sum_{j\neq i}\theta_i\theta_{M,j}\sigma_{ij})$$

$$-\lambda\widehat{b}_0\int_{-\infty}^{R_f-\theta_{i,-1}g_i/\theta_i}(\theta_i(R_f-R_i)-\theta_{i,-1}g_i)^\alpha dw(P(R_i))$$
$$-\widehat{b}_0\int_{R_f-\theta_{i,-1}g_i/\theta_i}^{\infty}(\theta_i(R_i-R_f)+\theta_{i,-1}g_i)^\alpha dw(1-P(R_i)), \tag{34}$$

where

$$\widehat{\gamma} = \gamma W_0\theta_R \tag{35}$$
$$\widehat{b}_0 = b_0 W_0^{\alpha-1}\theta_R^{\alpha-1}. \tag{36}$$

Note that the expression in (34), viewed as a function of θ_i, is identical to the expression in (20), viewed as a function of Θ_i, subject only to a rescaling of $\widehat{\gamma}$ and $\widehat{b}_0$.

Note also that, now that we are working with shares of the risky asset portfolio, we can simplify the variance term in the first row of (34). Specifically, the quantity

$$\sum_{j\neq i}\theta_{M,j}\sigma_{ij}$$

can be rewritten as

$$-\theta_{M,i}\sigma_i^2+\sum_j\theta_{M,j}\sigma_{i,j}=-\theta_{M,i}\sigma_i^2+cov(\widetilde{R}_i,\widetilde{R}_M)=-\theta_{M,i}\sigma_i^2+\beta_i\sigma_M^2,$$

where $\widetilde{R}_M$ is the return on the market portfolio of all risky assets, β_i is the beta of asset i, and σ_M^2 is the standard deviation of the market portfolio. Substituting this into (34) leads to the expression in (23).

8.2 Computing the equilibrium with heterogeneous holdings

The equilibrium structure we consider – a bounded-rationality equilibrium with heterogeneous holdings – consists of a location vector $(\mu_1,\ldots,\mu_N)$ so that, for each i, the expression in (23) has either a unique global maximum at $\theta_i^*=\theta_{M,i}$ or else has two global maxima at θ_i^* and θ_i^{**} with $\theta_i^*<\theta_{M,i}<\theta_i^{**}$.

We now describe a way of computing this heterogeneous-holdings equilibrium. For a given risky asset $i\in\{1,\ldots,N\}$, we first check whether investors have homogeneous holdings in that asset, in other words, whether they all hold the same per-capita market supply $\theta_{M,i}$. To do this, we take the first derivative of (23), substitute in $\theta_i=\theta_{M,i}$, and set the resulting

expression to 0. This gives:

$$
\begin{aligned}
0 = &(\mu_i + \frac{\nu\zeta_i}{\nu-2} - R_f) - \widehat{\gamma}\beta_i\sigma_M^2 \\
&-\alpha\lambda\widehat{b}_0 \int_{-\infty}^{R_f-\theta_{i,-1}g_i/\theta_i} (\theta_{M,i}(R_f - R_i) - \theta_{i,-1}g_i)^{\alpha-1}(R_f - R_i)dw(P(R_i)) \\
&-\alpha\widehat{b}_0 \int_{R_f-\theta_{i,-1}g_i/\theta_i}^{\infty} (\theta_{M,i}(R_i - R_f) + \theta_{i,-1}g_i)^{\alpha-1}(R_i - R_f)dw(1-P(R_i)). \qquad (37)
\end{aligned}
$$

We then check whether, for the μ_i that solves (37), the function in (23) has a unique maximum at $\theta_i = \theta_{M,i}$, as opposed to only a local maximum or a local minimum. If $\theta_i = \theta_{M,i}$ is indeed a global maximum, then all investors have identical holdings of risky asset i, each holding the per-capita supply of the asset, namely $\theta_{M,i}$. If the function in (23) does not have a global maximum at this θ_i, then the equilibrium does not involve homogeneous holdings and we must search for a heterogeneous-holdings equilibrium.

The simplest way to find the heterogeneous-holdings equilibrium is to search for a value of μ_i so that the maximum value of the function in (23) in the range $[0, \theta_{M,i})$ is equal to the maximum value of the function in the range $(\theta_{M,i}, \infty)$. If we find such a μ_i, then there is an equilibrium where investors have heterogeneous holdings, with some of them allocated to $\theta_i = \theta_i^*$, the point at which the function in (23) is maximized in $[0, \theta_{M,i})$, and others allocated to $\theta_i = \theta_i^{**}$, the point at which the objective function is maximized in $(\theta_{M,i}, \infty)$. We find that the value of μ_i for which this holds is typically close to the value of μ_i that solves the first-order condition (37). Therefore, if we find that the value of μ_i that solves (37) does not correspond to a global maximum at $\theta_i = \theta_{M,i}$, we simply search in the neighborhood of that μ_i for a heterogeneous-holdings equilibrium.

8.3 Inferring beliefs from option prices

In Section 6.1, we propose using option prices to back out estimates of the risk-neutral volatility and skewness investors perceive. This, in turn, can help us understand whether investors have sensible beliefs about stocks' actual volatility and skewness. In this section, we describe our methodology, which follows that of Birru and Wang (2016).

Bakshi, Kapadia, and Madan (2003) show that a stock's risk-neutral variance and skewness can be computed as

$$
V_{i,t}(t,\tau) \;=\; \int_{S(t)}^{\infty} \frac{2(1-\ln(K/S(t)))}{K^2} C(t,\tau;K)dK
$$

$$
\begin{aligned}
& + \int_0^{S(t)} \frac{2(1+\ln(K/S(t)))}{K^2} P(t,\tau;K)dK \\
W_{i,t}(t,\tau) &= \int_{S(t)}^{\infty} \frac{6\ln(K/S(t)) - 3(\ln(K/S(t)))^2}{K^2} C(t,\tau;K)dK \\
& - \int_0^{S(t)} \frac{6\ln(K/S(t)) + 3(\ln(K/S(t)))^2}{K^2} P(t,\tau;K)dK \\
X_{i,t}(t,\tau) &= \int_{S(t)}^{\infty} \frac{12\ln(K/S(t)) - 4(\ln(K/S(t)))^2}{K^2} C(t,\tau;K)dK \\
& - \int_0^{S(t)} \frac{12\ln(K/S(t)) + 4(\ln(K/S(t)))^2}{K^2} P(t,\tau;K)dK \\
\mu_{i,t}(t,\tau) &= e^{r\tau} - 1 - e^{r\tau} V_{i,t}(t,\tau)/2 - e^{r\tau} W_{i,t}(t,\tau)/6 - e^{r\tau} X_{i,t}(t,\tau)/24.
\end{aligned}
$$

In principle, the above equations should be implemented using a continuum of options, but since there are only a finite number of options traded on a given stock, we use a discrete approximation instead. Dennis and Mayhew (2012) examine some potential biases that could arise from the discrete approximation, and conclude that these are very small, so long as the calculation makes use of at least two out-of-the-money (OTM) put options and two OTM call options. As a result, we require at least two OTM puts and two OTM calls. We compute a 30-day risk-neutral variance and skew by interpolating the skewness of the option with expiration closest to, but less than, 30 days and the option with expiration closest to, but greater than, 30 days. If there is no option with maturity longer (shorter) than 30 days, we choose the longest (shortest) available maturity. We have also computed results using 60-day options and one-year options; these lead to similar conclusions.

Our data on options comes from the Ivy DB Option Metrics database, which provides option prices, volume, and open interest from January 1996 to December 2012. We include options on all securities classified as common stock. To minimize the impact of data errors, we remove options missing best bid or offer prices, as well as those with bid prices less than or equal to \$0.05. We also remove options that violate arbitrage bounds, and options with zero open interest. Unless stated explicitly, we also exclude options with special settlement arrangements, and options for which the underlying stock price is under \$10. The midpoint of the best bid and best offer is taken as the option price.

9 References

An, L., Wang, H., Wang, J., and J. Yu (2017), "Lottery-related Anomalies: The Role of Reference-dependent Preferences," Working paper.

Andrikogiannopoulou, A. and F. Papakonstantinou (2018), " History-dependent Risk Preferences: Evidence from Individual Choices and Implications for the Disposition Effect," Working paper.

Bakshi, G., Kapadia, N., and D. Madan (2003), "Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options," *Review of Financial Studies* 16, 101-143.

Barberis, N., and M. Huang (2001), "Mental Accounting, Loss Aversion, and Individual Stock Returns," *Journal of Finance* 56, 1247-1292.

Barberis, N., and M. Huang (2008), "Stocks as Lotteries: The Implications of Probability Weighting for Security Prices," *American Economic Review* 98, 2066-2100.

Barberis, N., Huang, M., and R. Thaler (2006), "Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing," *American Economic Review* 96, 1069-1090.

Birru, J. and B. Wang (2016), "Nominal Price Illusion," *Journal of Financial Economics* 119, 578-598.

Booij, A., van Praag, B., and G. van de Kuilen (2010), "A Parametric Analysis of Prospect Theory's Functionals for the General Population," *Theory and Decision* 68, 115-148.

Boyer, B., Mitton, T., and K. Vorkink (2010), "Expected Idiosyncratic Skewness," *Review of Financial Studies* 23, 169-202.

Frazzini, A. (2006), "The Disposition Effect and Underreaction to News," *Journal of Finance* 61, 2017-2046.

Grinblatt, M. and B. Han (2005), "Prospect Theory, Mental Accounting, and Momentum," *Journal of Financial Economics* 78, 311-339.

Frydman, C., Hartzmark, S. and D. Solomon (2018), "Rolling Mental Accounts," *Review of Financial Studies* 31, 362-397.

Kahneman, D. and A. Tversky (1979), "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica* 47, 263-291.

Li, Y. and L. Yang (2013), "Prospect Theory, the Disposition Effect, and Asset Prices," *Journal of Financial Economics* 107, 715-739.

Tversky, A. and D. Kahneman (1981), "The Framing of Decisions and the Psychology of Choice," *Science* 211, 453-458.

Tversky, A. and D. Kahneman (1992), "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty* 5, 297-323.

Wang, H., Yan, J., and J. Yu (2017), "Reference-dependent Preferences and the Risk-Return Trade-off," *Journal of Financial Economics* 123, 395-414.

Table 1. List of anomalies.

Anomaly	Abbreviation
Idiosyncratic volatility	VOL
Market capitalization	SIZE
Value	VAL
Expected idiosyncratic skewness	EISKEW
Momentum	MOM
Failure probability	FPROB
O-Score	OSC
Net stock issuance	NSI
Composite equity issuance	CEI
Accrual	ACC
Net operating assets	NOA
Gross profitability	PROF
Asset growth	AG
Return on assets	ROA
Investment	INV
Maximum daily return	MAX
Organizational capital	ORGCP
Long-term reversal	LTREV
External finance	XFIN
Short-term reversal	STREV
Difference of opinion	DOO
Post-earnings announcement drift	PEAD

Table 2. Empirical properties of anomaly deciles.

Anomaly	Average return Decile 1	Average return Decile 10	Skewness Decile 1	Skewness Decile 10
VOL	11.9	-3.2	2.46	3.79
SIZE	14.0	10.63	4.27	0.69
VAL	10.7	17.7	1.85	2.66
EISKEW	12.4	8.0	1.33	3.54
MOM	-2.4	20.6	3.84	2.46
FPROB	17.5	0.1	2.31	3.9
OSC	13.9	3.8	2.56	3.5
NSI	16.0	6.8	2.71	3.2
CEI	14.0	6.8	2.46	2.68
ACC	15.4	7.4	3.2	3.0
NOA	14.8	6.9	3.13	2.95
PROF	8.8	14.4	3.49	2.68
AG	14.8	7.0	3.1	3.07
ROA	4.1	13.1	3.22	2.22
INV	15.9	8.0	3.6	3.31
MAX	10.9	1.2	2.76	3.61
ORGCP	10.9	15.8	2.76	3.38
LTREV	16.7	11.0	3.23	1.77
XFIN	13.0	4.2	3.17	3.44
STREV	14.3	7.7	3.63	3.03
DOO	15.5	10.0	1.36	1.74
PEAD	9.25	16.8	2.49	2.3

Table 3. Empirical properties of anomaly deciles, continued.

Anomaly	Standard deviation Decile 1	Standard deviation Decile 10	Gain overhang Decile 1	Gain overhang Decile 10
VOL	36.8	94.3	10.5	-31.6
SIZE	76.0	24.6	-15.4	17.0
VAL	54.1	71.0	12.1	-24.1
EISKEW	33.3	90.3	13.7	-24.8
MOM	91.4	63.1	-44.6	30.8
FPROB	51.5	110.4	28.2	-46.1
OSC	51.6	84.4	6.1	-20.6
NSI	54.2	70.8	5.7	-8.4
CEI	42.1	66.5	7.4	-11.2
ACC	85.0	72.0	-18.7	-5.4
NOA	66.4	65.3	-2.6	-4.9
PROF	73.0	62.6	-12.0	1.1
AG	82.5	70.7	-12.0	1.1
ROA	85.3	55.3	-29.9	15.1
INV	81.8	72.2	-14.0	-5.4
MAX	46.3	90.4	3.2	-25.1
ORGCP	63.3	77.5	-3.8	-7.9
LTREV	88.9	47.7	-32.1	17.1
XFIN	58.7	83.4	-0.5	-15.6
STREV	83.1	73.3	-22.2	-1.9
DOO	40.2	57.9	16.6	-7.8
PEAD	58.2	58.3	-9.4	7.5

Table 4. Risk-neutral beliefs from option prices.

Anomaly	RN-variance Decile 1	RN-variance Decile 10	RN-skewness Decile 1	RN-skewness Decile 10
VOL	0.03	0.17	-0.53	-0.21
SIZE	0.23	0.02	-0.12	-0.60
VAL	0.06	0.13	-0.36	-0.36
EISKEW	0.04	0.17	-0.45	-0.18
MOM	0.17	0.07	-0.21	-0.32
FPROB	0.05	0.22	-0.35	-0.24
OSC	0.05	0.18	-0.36	-0.18
NSI	0.05	0.10	-0.41	-0.28
CEI	0.04	0.10	-0.43	-0.28
ACC	0.12	0.09	-0.27	-0.26
NOA	0.10	0.07	-0.28	-0.34
PROF	0.13	0.06	-0.25	-0.34
AG	0.12	0.09	-0.27	-0.29
ROA	0.17	0.05	-0.19	-0.37
INV	0.09	0.08	-0.30	-0.32
MAX	0.03	0.14	-0.46	-0.29
ORGCP	0.08	0.08	-0.34	-0.31
LTREV	0.17	0.06	-0.25	-0.35
XFIN	0.05	0.11	-0.39	-0.25
STREV	0.12	0.09	-0.12	-0.4
DOO	0.04	0.11	-0.44	-0.29
PEAD	0.08	0.06	-0.33	-0.36

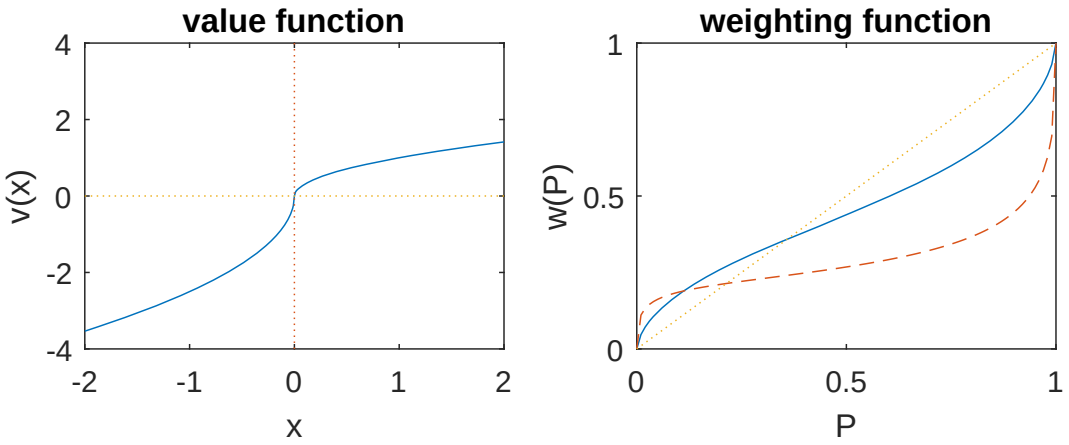


Figure 1. The left panel plots the value function proposed by Tversky and Kahneman (1992) as part of their cumulative prospect theory, namely, $v(x) = x^{\alpha}$ for $x \geq 0$ and $v(x) = -\lambda(-x)^{\alpha}$ for $x < 0$, for $\alpha = 0.5$ and $\lambda = 2.5$. The right panel plots the probability weighting function proposed by Tversky and Kahneman (1992), namely, $w(P) = P^{\delta}/(P^{\delta} + (1-P)^{\delta})^{1/\delta}$, for three different values of δ. The dashed line corresponds to $\delta = 0.4$, the solid line to $\delta = 0.65$, and the dotted line to $\delta = 1$.

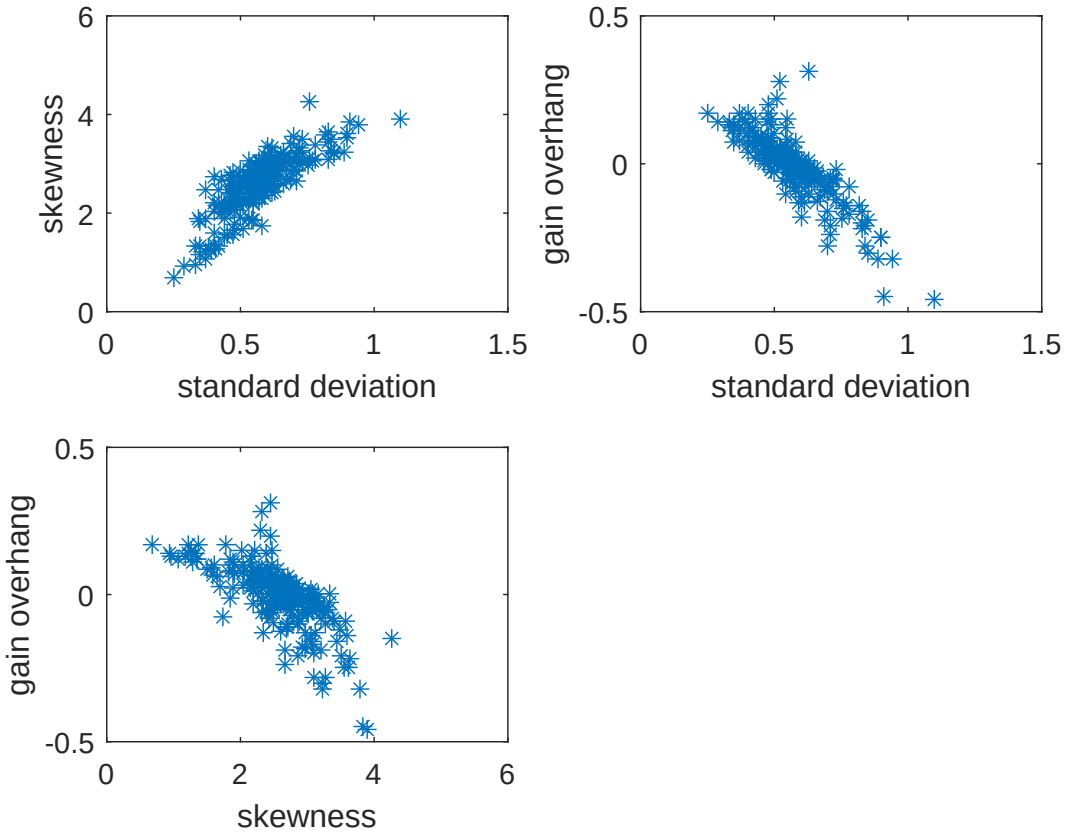


Figure 2. The three graphs show the relationship between volatility, skewness, and capital gain overhang for 220 stocks, where each stock is a typical stock in one of the ten deciles of one of 22 anomalies. Each asterisk corresponds to one of the 220 stocks.

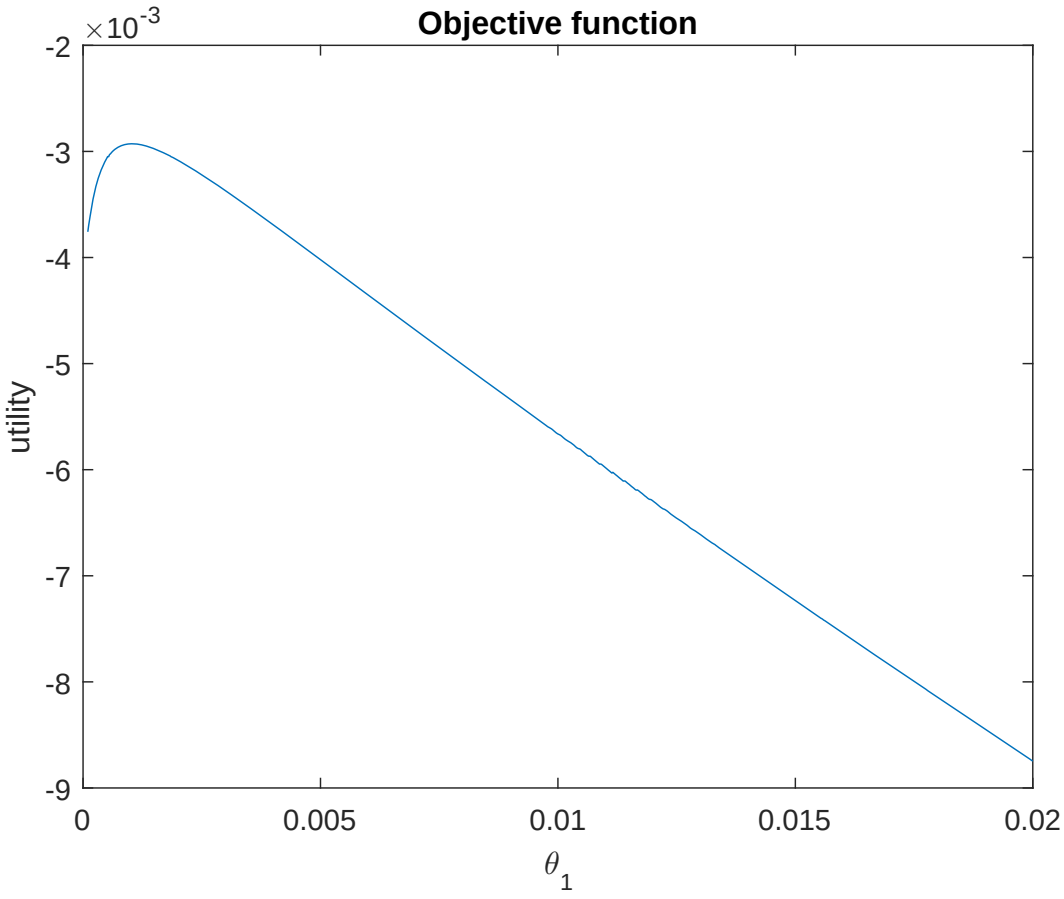


Figure 3. The graph shows that all investors have identical holdings of each stock in momentum decile 1. The solid line plots the value of an investor's objective function in equilibrium, as a function of θ_1, the fraction of the investor's risky asset portfolio allocated to stock 1. This function has a unique global maximum at the point where θ_1 equals the weight of stock 1 in the market portfolio, namely 0.001.

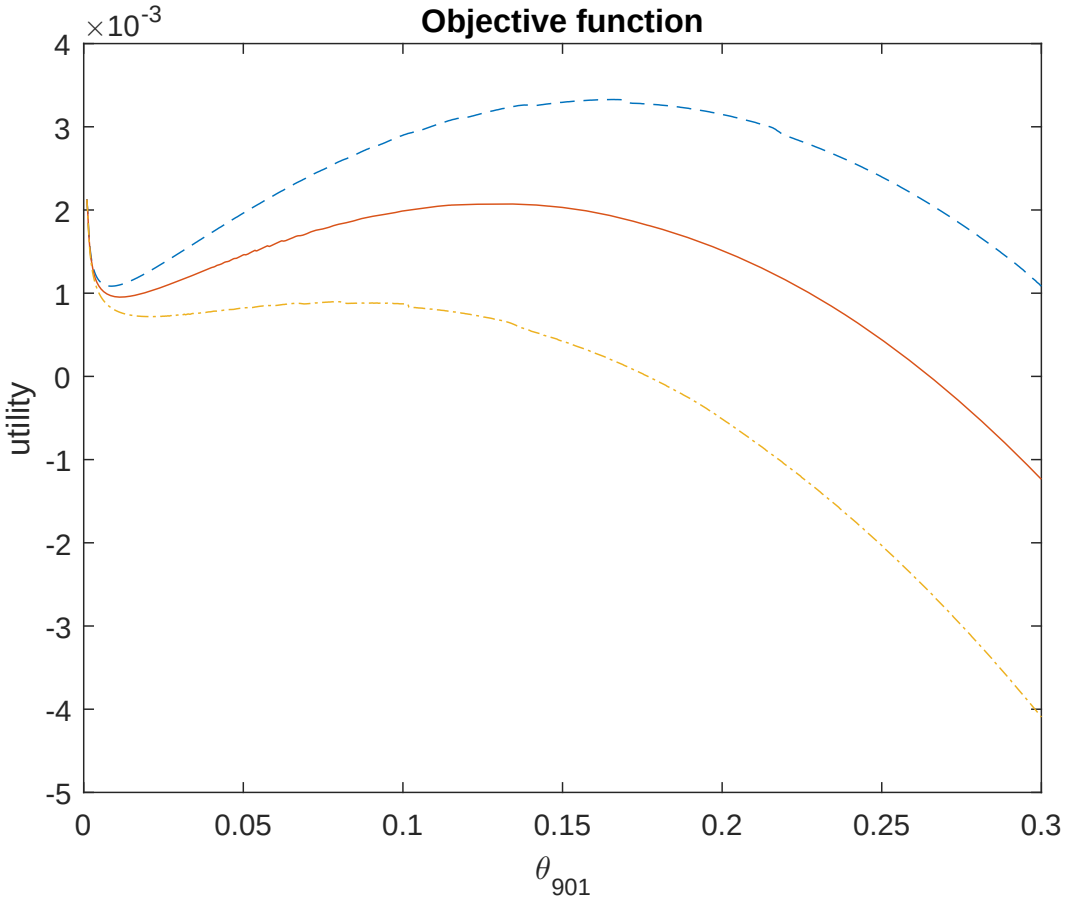


Figure 4. The graph shows that investors have heterogeneous holdings of each stock in momentum decile 10. The solid line plots the value of an investor's objective function in equilibrium, as a function of θ_{901}, the fraction of the investor's risky asset portfolio allocated to stock 901. This function has two global maxima that straddle the weight of stock 901 in the market portfolio, namely 0.001. The dashed line plots the objective function for a higher expected return on the stock, while the dash-dot line plots the objective function for a lower expected return.

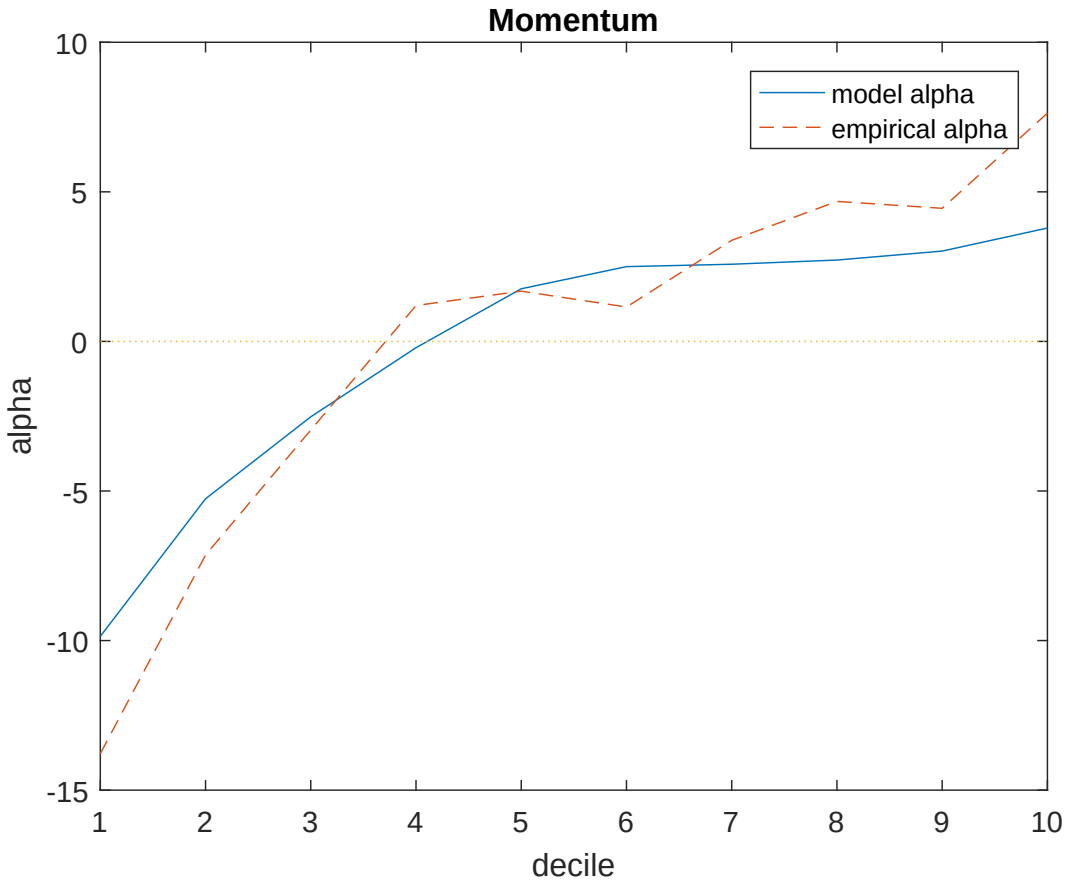


Figure 5. The dashed line plots the historical alpha of each momentum anomaly decile. The solid line plots the alphas predicted by a model where investors evaluate risk according to prospect theory and engage in narrow framing.

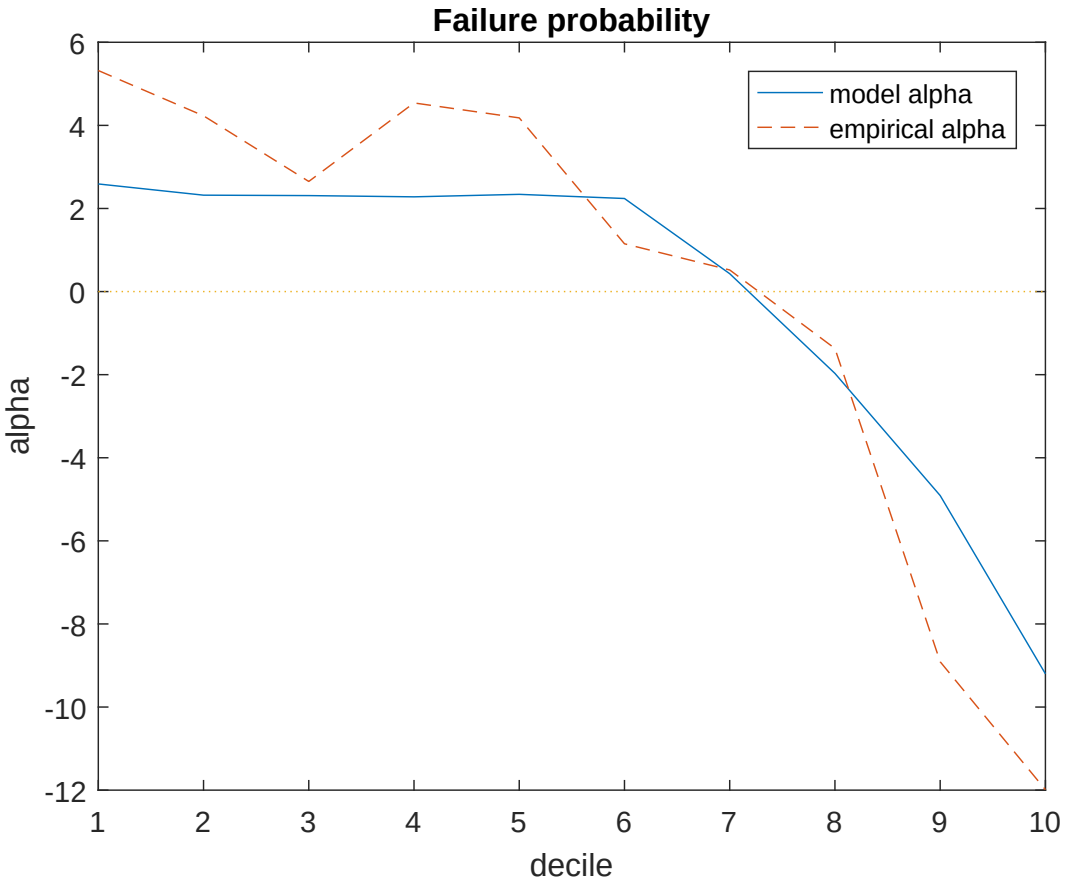


Figure 6. The dashed line plots the historical alpha of each failure probability anomaly decile. The solid line plots the alphas predicted by a model where investors evaluate risk according to prospect theory and engage in narrow framing.

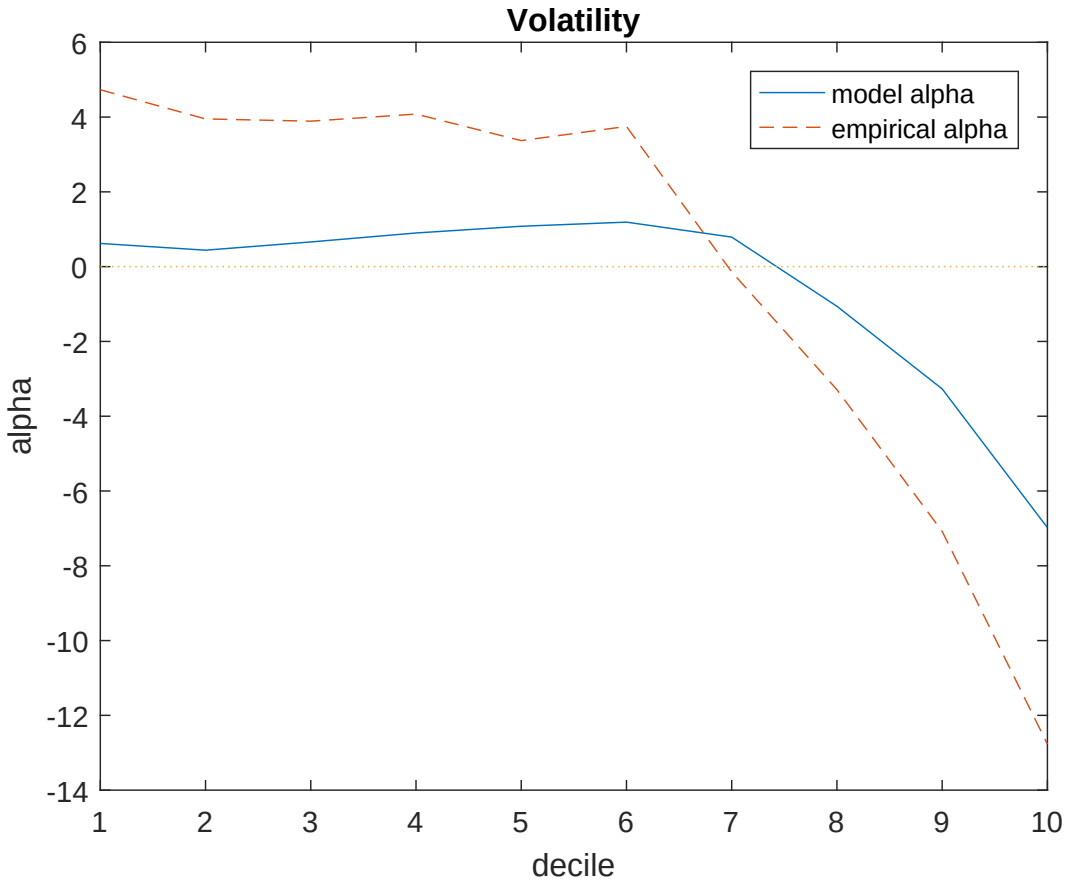


Figure 7. The dashed line plots the historical alpha of each idiosyncratic volatility anomaly decile. The solid line plots the alphas predicted by a model where investors evaluate risk according to prospect theory and engage in narrow framing.

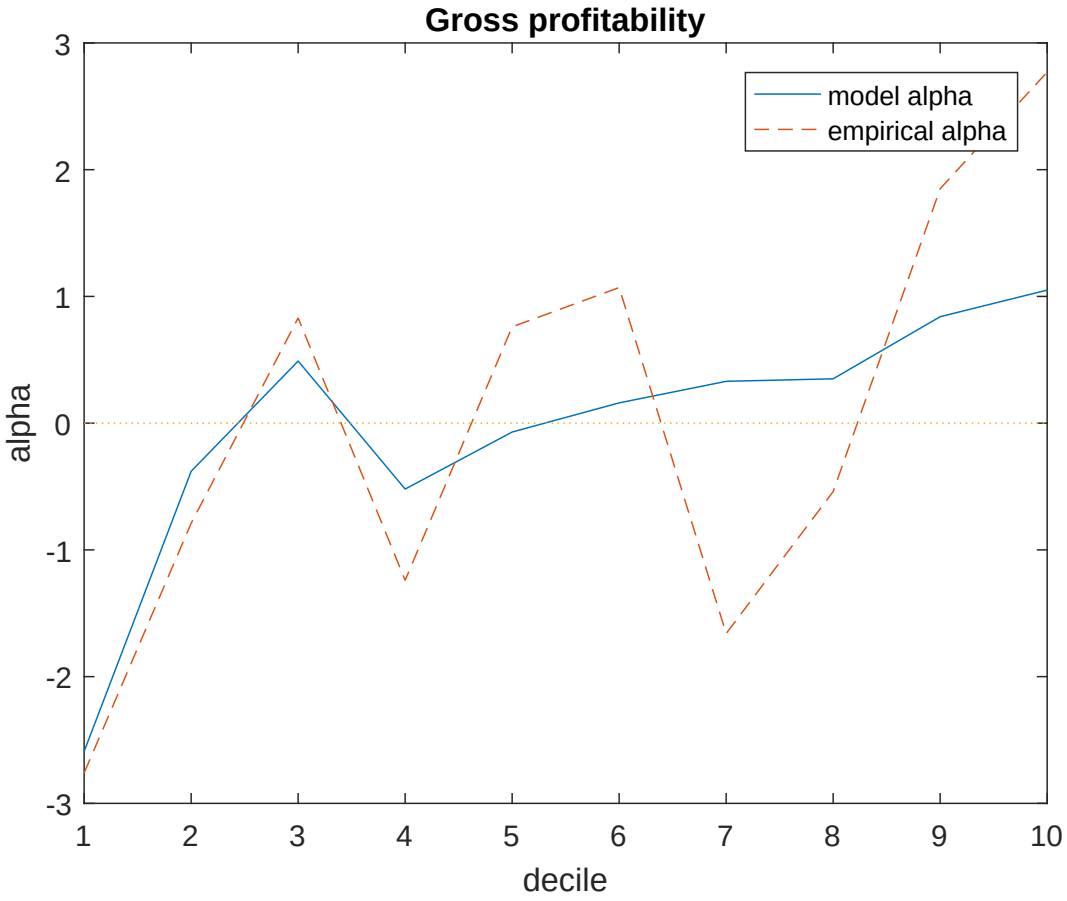


Figure 8. The dashed line plots the historical alpha of each gross profitability anomaly decile. The solid line plots the alphas predicted by a model where investors evaluate risk according to prospect theory and engage in narrow framing.

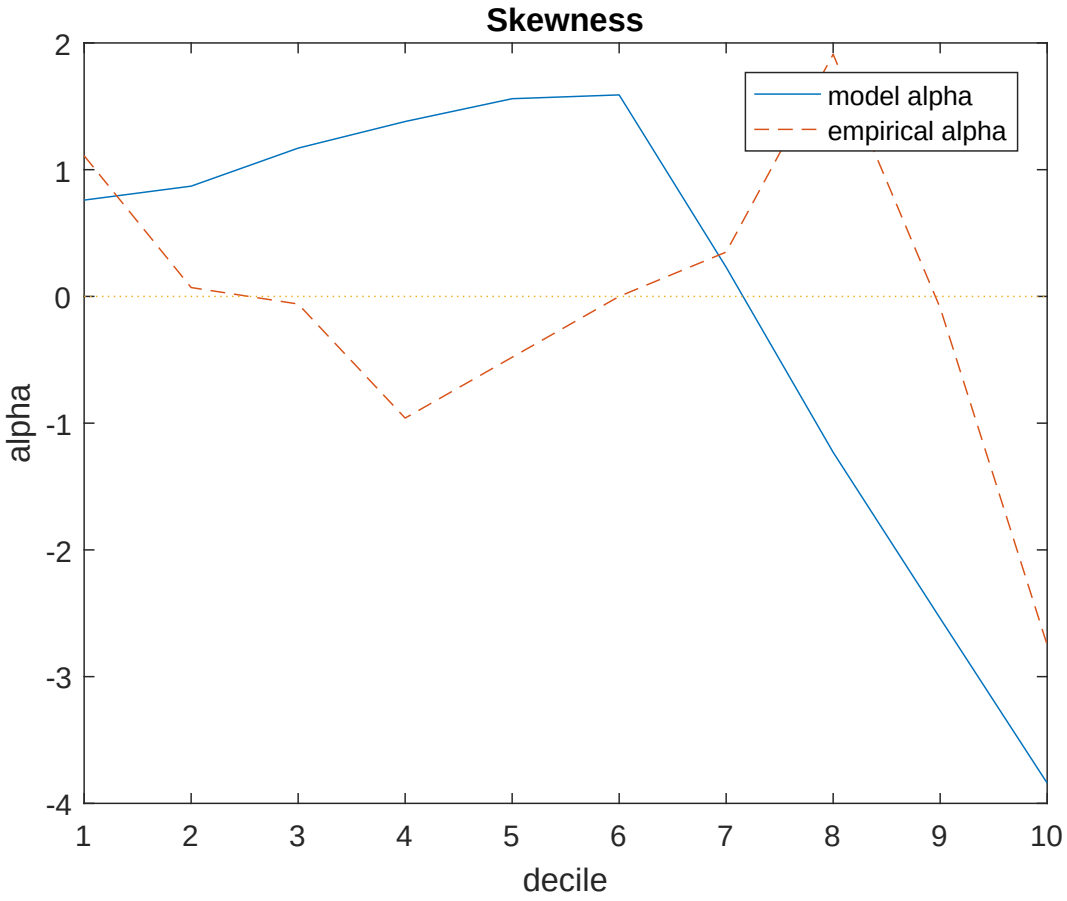


Figure 9. The dashed line plots the historical alpha of each idiosyncratic skewness anomaly decile. The solid line plots the alphas predicted by a model where investors evaluate risk according to prospect theory and engage in narrow framing.

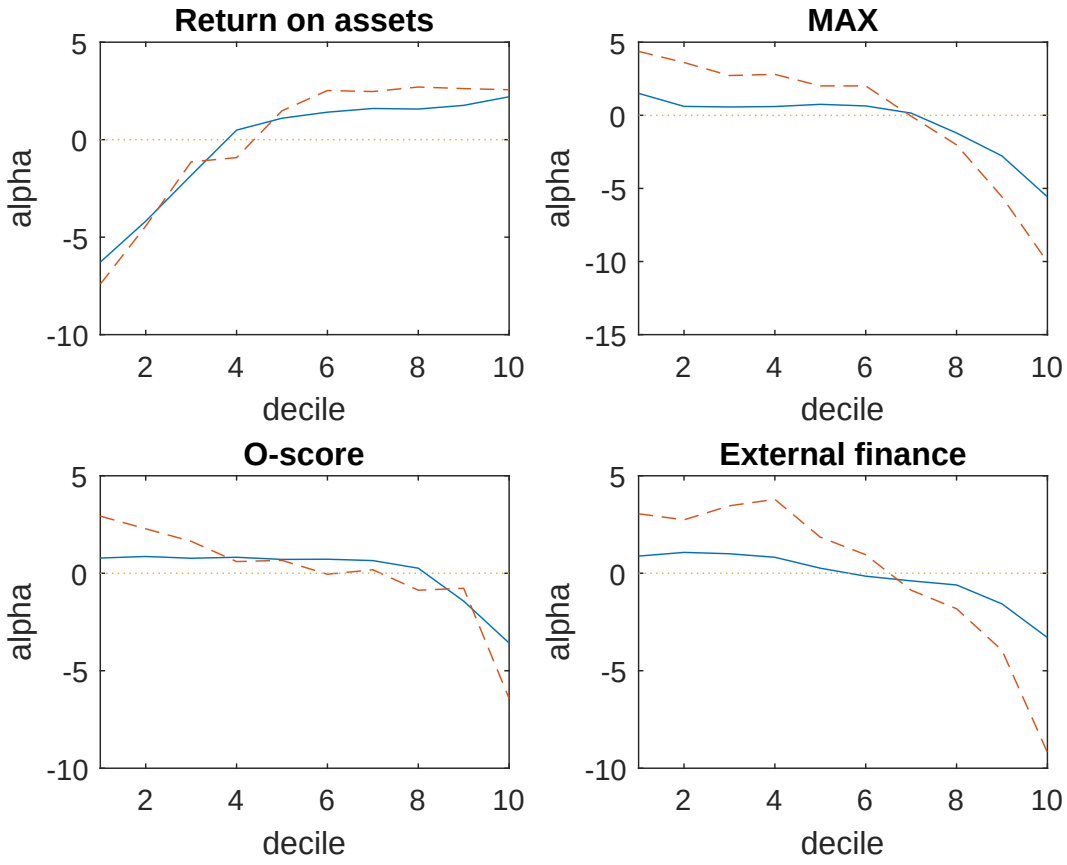


Figure 10. In each graph, the dashed lines plot the historical alpha of each anomaly decile for the return on assets, maximum daily return, O-score, and external finance anomalies. The solid lines plot the alphas predicted by a model where investors evaluate risk according to prospect theory and engage in narrow framing.

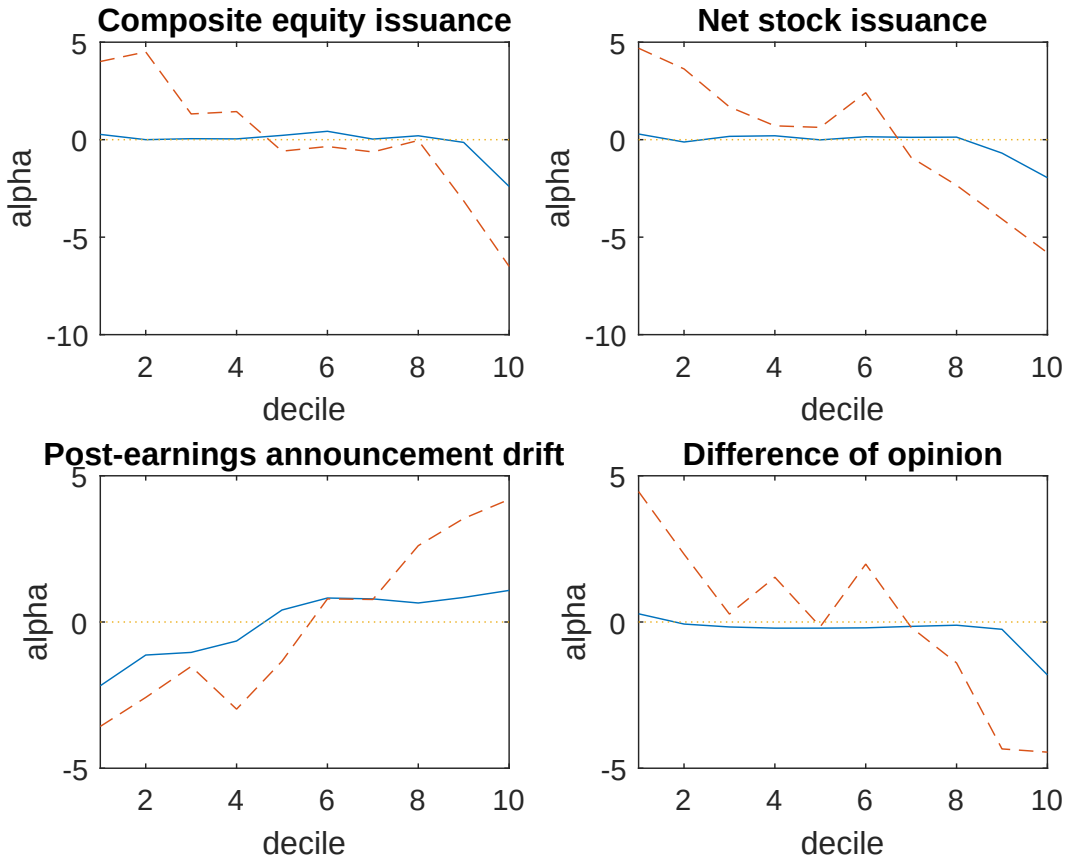


Figure 11. In each graph, the dashed lines plot the historical alpha of each anomaly decile for the composite equity issuance, net stock issuance, post-earnings announcement drift, and difference of opinion anomalies. The solid lines plot the alphas predicted by a model where investors evaluate risk according to prospect theory and engage in narrow framing.

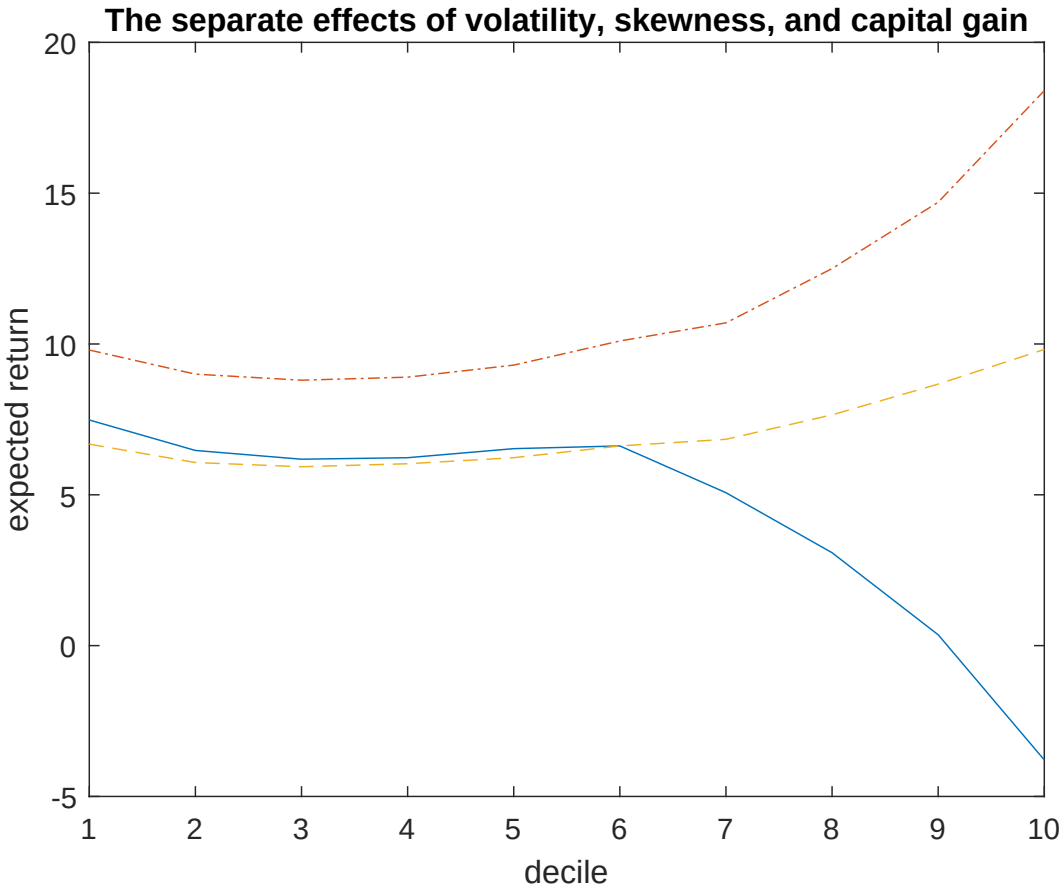


Figure 12. The solid line plots the expected return of each failure probability anomaly decile in a model where investors evaluate risk according to prospect theory and engage in narrow framing. The dashed line shows the expected returns when we take account of stocks' volatility and skewness, but not of their gain overhang. The dash-dot line shows the expected returns when we take account of stocks' volatility, but not of their skewness or gain overhang.

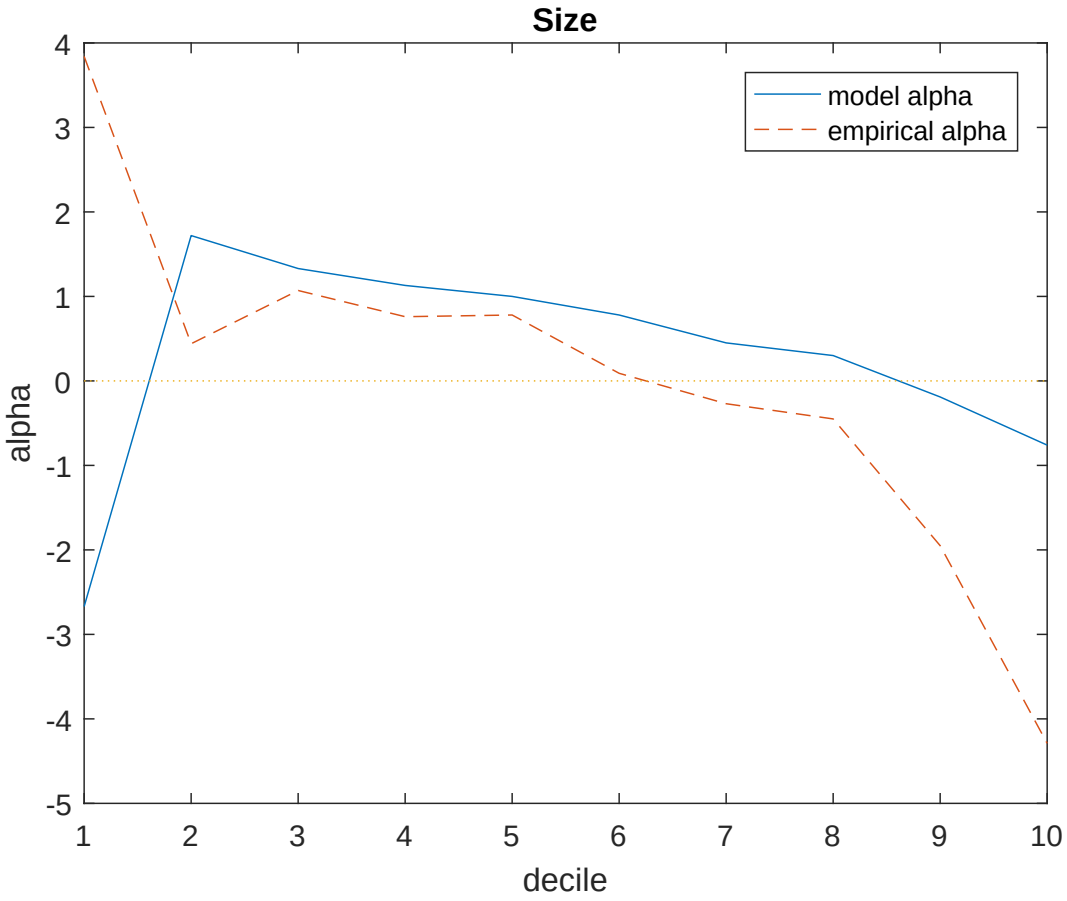


Figure 13. The dashed line plots the historical alpha of each size anomaly decile. The solid line plots the alphas predicted by a model where investors evaluate risk according to prospect theory and engage in narrow framing.

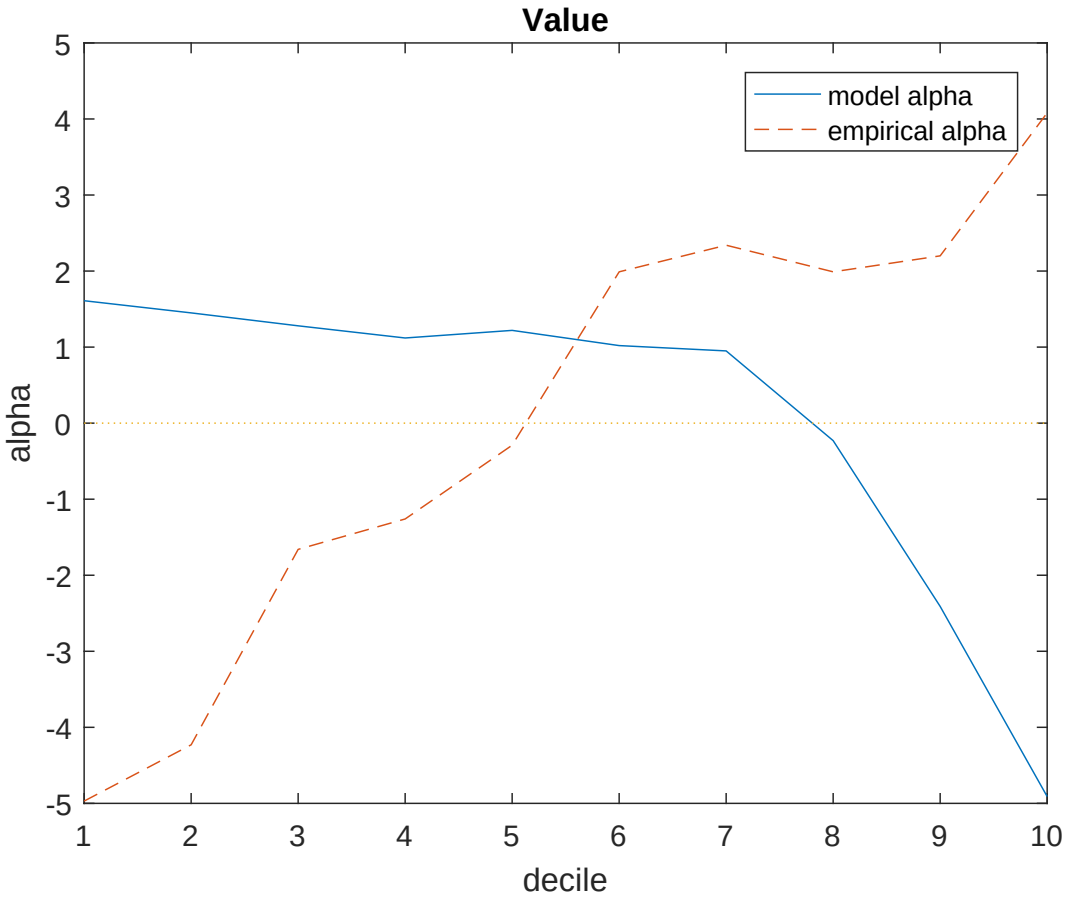


Figure 14. The dashed line plots the historical alpha of each value anomaly decile. The solid line plots the alphas predicted by a model where investors evaluate risk according to prospect theory and engage in narrow framing.

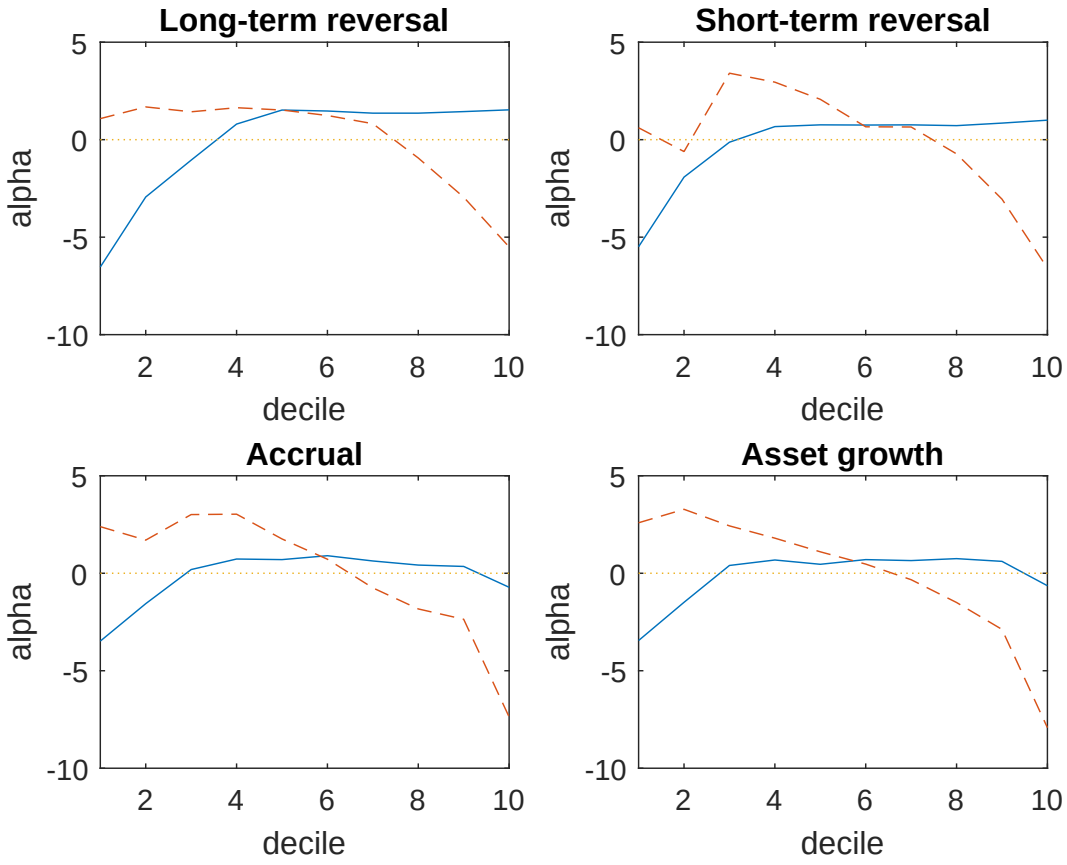


Figure 15. In each graph, the dashed line plots the historical alpha of each anomaly decile for the long-term reversal, short-term reversal, accrual, and asset growth anomalies. The solid lines plots the alphas predicted by a model where investors evaluate risk according to prospect theory and engage in narrow framing.

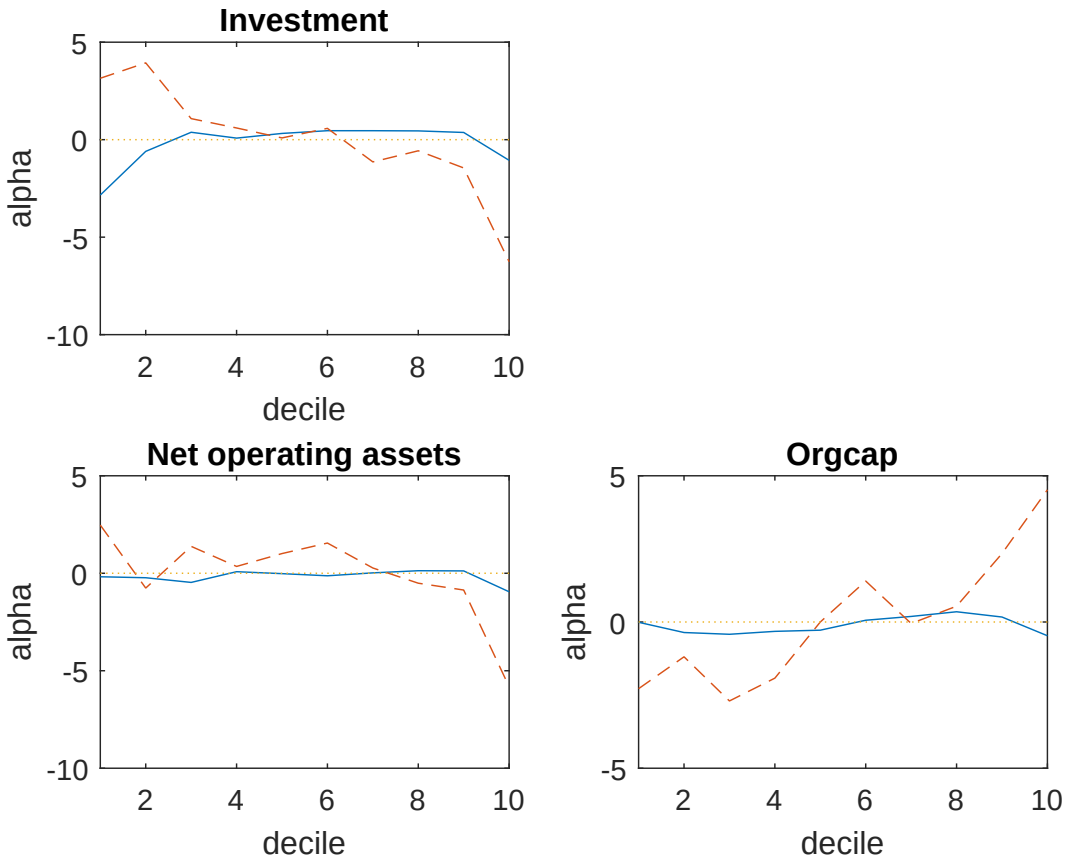


Figure 16. In each graph, the dashed line plots the historical alpha of each anomaly decile for the investment, net operating assets, and organizational capital anomalies. The solid lines plot the alphas predicted by a model where investors evaluate risk according to prospect theory and engage in narrow framing.

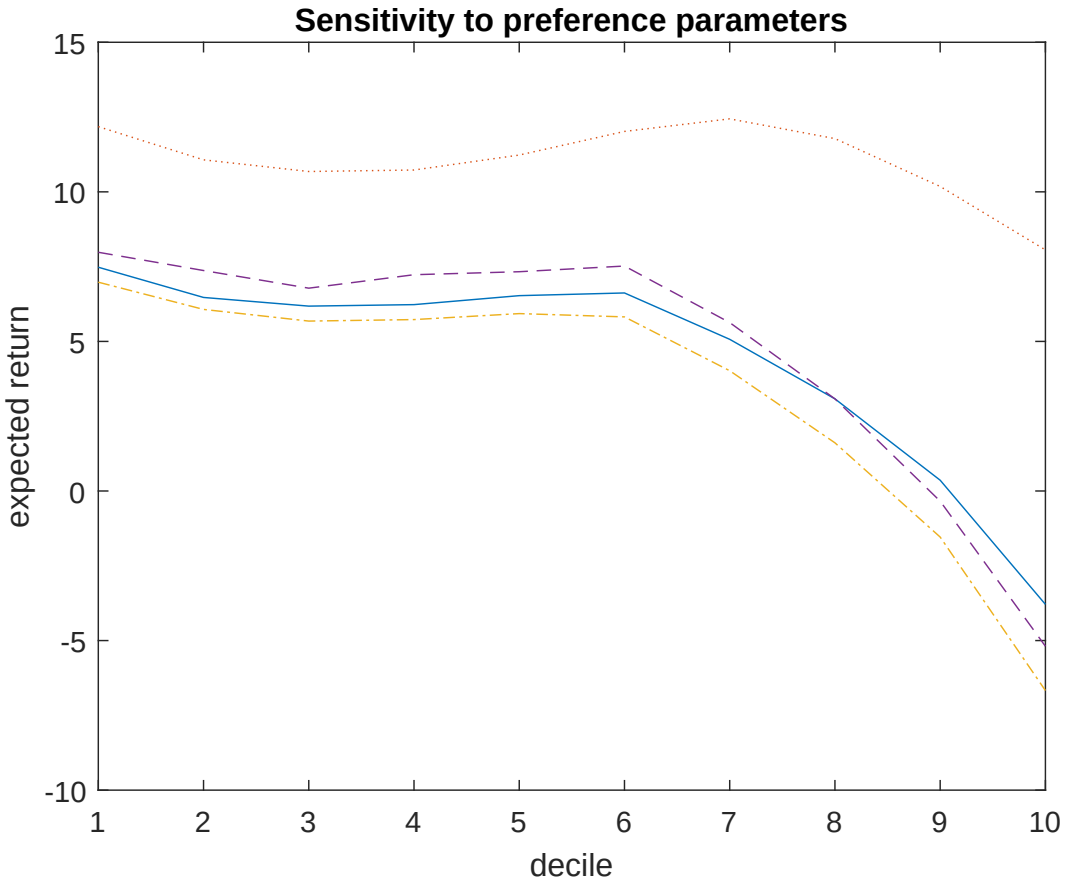


Figure 17. The solid line plots the expected return of each failure probability anomaly decile in a model where investors evaluate risk according to prospect theory and engage in narrow framing. The dotted line shows the expected returns when the preference parameter λ is raised to 2; the dash-dot line shows the expected returns when the preference parameter δ is lowered to 0.6; and the dashed line shows the expected returns when the preference parameter α is lowered to 0.65.