Prospect Theory and Stock Market Anomalies

Nicholas Barberis, Lawrence Jin, and Baolian Wang

August 2019*

Abstract

We present a new model of asset prices in which investors evaluate risk according to prospect theory and examine its ability to explain 22 prominent stock market anomalies. The model incorporates all the elements of prospect theory, takes account of investors' prior gains and losses, and makes quantitative predictions about an asset's average return based on empirical estimates of its beta, volatility, skewness, and capital gain overhang. 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:

*The authors are grateful to seminar participants at Boston College and the University of Maryland for helpful feedback, and to Jonathan Ingersoll for extensive discussions.

1 Introduction

Prospect theory, due to Kahneman and Tversky (1979, 1992), is a highly influential theory of decision-making under risk. In a parsimonious way, it summarizes a wide range of experimental evidence on attitudes to risk. As such, it has the potential to shed light on asset prices and investor behavior. However, despite years of effort, we still do not understand its implications for basic aspects of asset prices, such as the cross-section of average returns. Under mean-variance preferences, average returns are described by the CAPM. But what determines average returns when investors instead evaluate risk according to prospect theory? What does prospect theory predict about the relative average returns of small-cap stocks and large-cap stocks, or of value stocks and growth stocks? Full answers to these basic questions are still not available.

In this paper, we answer these questions. 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 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, where the utility function is kinked at its origin, so that people are more sensitive to losses than to gains ("loss aversion"), and also concave over gains and convex over losses, so that they are risk averse over moderate-probability gains and risk-seeking over moderate-probability losses ("diminishing sensitivity"). In addition, they weight outcomes not by objective probabilities but by transformed probabilities that overweight the tails of the distribution they are thinking about ("probability weighting"). 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 thinking about how much money to allocate to a particular stock, he focuses, at least in part, on the potential gains and losses in his holdings of the stock itself.

Intuition and prior research suggest 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 asset, a quantity known as the asset'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 an asset'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 utility function's concavity over gains and convexity over losses mean that, all else equal, investors require a *higher* average return on assets where they have larger prior gains.[1]

The above intuitions indicate that, to understand prospect theory's implications for asset prices, we need a model that incorporates all the elements of prospect theory *and* accounts for investors' prior gains and losses in each risky asset. No existing model fulfills both conditions; we therefore build a new one that does. In our model, investor preferences have two components. The first is the traditional mean-variance preference specification; taken alone, it leads to the CAPM. The second embeds prospect theory and narrow framing.

While our model has a simple structure, solving for equilibrium prices presents a challenge. In the model, 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 construct an alternative equilibrium, one in which investors hold different portfolios that correspond to non-unique optima of their objective function. We then show how the model can be used to generate quantitative predictions about the expected return on any risky asset.

In our main application, we examine whether the model can explain 22 prominent stock market anomalies. To see if our model can explain a particular anomaly – the size anomaly, say – we compute what it predicts for the average return of the typical small-cap stock. As explained above, this average return will depend on the return volatility, return 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

[1]The three intuitions described here are outlined in Barberis and Huang (2001); Barberis and Huang (2008); and Grinblatt and Han (2005) and Li and Yang (2013), respectively.

three characteristics are strongly correlated across anomaly deciles: if the typical stock in decile 1 for some anomaly has more positively skewed returns than the typical stock in decile 10 for that anomaly, 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 each 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 sensible 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 stock returns. Since these are risk-neutral beliefs, they constitute only suggestive evidence. Nonetheless, they indicate that investors have sensible beliefs about the return distributions of many types of stocks. Interestingly, however, they also suggest that investors have incorrect beliefs about the returns on value stocks: 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 average return on value stocks.

We noted above that, to fully understand the implications of prospect theory for asset prices, we need a model that incorporates all the elements of prospect theory *and* takes account of investors' prior gain or loss in each asset. Most of the earlier models incorporate only a subset of the elements of prospect theory: only loss aversion (Barberis and Huang, 2001), only loss aversion and diminishing sensitivity (Li and Yang, 2013), or only loss aversion and probability weighting (Baele et al., 2018). Meanwhile, the two prior models of the cross-section that do incorporate all the elements of prospect theory, Barberis and Huang (2008) and Barberis, Mukherjee, and Wang (2016), are both one-period models; as such, they cannot account for investors' prior gains and losses. A new model is needed, and we develop one in this paper.

More generally, we advance research on prospect theory applications in finance 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, one that overcomes the limitations of prior approaches. Second, in order to derive quantitative predictions about average returns, 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 at most a small set of anomalies, we widen the scope of this research by looking at 22 different anomalies. To our knowledge, our paper marks the first time a "behavioral" model

of either beliefs or preferences has been used to make quantitative predictions about such a wide range of 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 introduce the 22 anomalies that are the focus of our study and compute the empirical characteristics that serve as inputs to our model – specifically, the return volatility, return skewness, and capital gain overhang of the typical stock in each anomaly decile. In Section 5, we present our model's predictions about stock market anomalies. In Section 6, we discuss some additional issues raised by our analysis, and 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 review these concepts. Readers already familiar with prospect theory may prefer to go directly to Section 2.2 on narrow framing.

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.[2]

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

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

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

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

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.[3] 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.

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

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

finding that people tend to be risk averse over moderate-probability gains – they prefer a 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.[4] 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.

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.

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

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

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". 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.

3 Model and Equilibrium Structure

In the Introduction, we noted that, in an economy with prospect theory investors who engage in narrow framing, three asset characteristics are particularly important for the pricing of an asset: the volatility of the asset's returns; the skewness of the asset's returns; and the average paper gain or loss in investors' holdings of the asset. We now 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 gains and losses. Since they are loss averse, they dislike stocks with 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 on such stocks. Finally, if a stock is trading at a gain for the typical investor, this investor finds himself in the concave region 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 typical investor has a paper loss in the stock, then he finds himself in the convex region to the left of the kink, where he is risk-seeking. As a result, he requires a low average return for holding the stock.

The above intuitions make it clear that, to fully understand prospect theory's implications for asset prices, we need a model that incorporates all the elements of prospect theory *and* accounts for investors' prior gains and losses in each risky asset. As noted in the Introduction, no existing model fulfills both conditions. We now present a new model that does.

3.1 Model setup

We consider a model with three dates, $t = -1$, 0, and 1; our focus is on investor decision-making at time 0. 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 cumulative distribution function $P(\widetilde{R})$ that we specify below. The vector of expected returns on the risky assets is $\overline{R} = (\overline{R}_1, \ldots, \overline{R}_N)$ and the covariance matrix of returns is $\Sigma = \{\sigma_{ij}\}$.

The economy contains a large number of investors who are identical in their preferences; in their wealth at time -1, W_{-1}; and in their wealth at time 0, W_0. The fraction of time 0 wealth that an investor allocates to risky asset i is Θ_i, so that wealth at time 1 is

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

where $\Theta = (\Theta_1, \ldots, \Theta_N)$. To determine Θ, at date 0, each investor solves:

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

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 preferences; γ measures aversion to portfolio risk. The third term in (9) is new, and captures prospect theory and narrow framing. It 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 potential 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 potential *future* gain or loss on asset i 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 return on asset 1 and on the risk-free asset between time 0 and time 1 are 20% and 2% respectively, then the realized value of this first term will be \$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. Here, W_{-1} is the investor's wealth at time -1, $\Theta_{i,-1}$ is the fraction of wealth allocated to asset i at time -1, and g_i is the capital gain on asset i between time -1 and time 0: if the investor experienced a capital gain of 30% on asset i between $t = -1$ and $t = 0$, then $g_i = 0.3$, while if he experienced a capital loss of 30%, then $g_i = -0.3$.[5]

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 between time -1 and time 0, and computes the prospect theory value of this *overall* gain or loss. This assumption is consistent with 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 (Thaler and Johnson, 1990; Imas, 2016; Andrikogiannopoulou and Papakonstantinou, 2019).

The second term on the right-hand side of (10) is identical across investors. Each investor in the model has the same prior gain or loss g_i in asset i, one that we will empirically estimate as the average capital gain or loss across all holders of the asset. In addition, for each investor, we will set $\Theta_{i,-1}$ to a neutral value, namely asset i's weight in the market portfolio. As such, the $W_{-1}\Theta_{i,-1}g_i$ term can be thought of as exogenous: $\Theta_{i,-1}$ 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, which appears in the first term in (10). Finally, we use the approximation $W_{-1} \approx W_0$.[6]

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 $V(\widetilde{G}_i)$ as

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

where $P(R_i)$ is the marginal distribution of asset i's returns and where the full expressions for $dw(P(R_i))$ and $dw(1 - P(R_i))$ are given in Appendix A. The expression in (11) uses a

[5]There are small differences in the way the past and the future gains and losses are defined: for simplicity, the past gain or loss does not account for dividends or correct for the risk-free rate. Adjusting for these has a very minor impact on our results.

[6]More accurate approximations, such as $W_{-1} \approx W_0/1.04$, where 4% is a measure of the historical average return on investor wealth, have a very minor impact on our quantitative predictions. We therefore stick with the simpler approximation $W_{-1} \approx W_0$.

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 investors' decision problem, we need to specify the probability distribution $P(\widetilde{R})$ for asset returns. Since skewness plays an important role in our analysis, we need a distribution that can capture it as accurately as possible. The "generalized hyperbolic (GH) skewed t" distribution is increasingly seen as a superior way of modeling skewness and fat tails in asset returns, and we adopt it here. The probability density function for the multivariate form of this distribution is

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

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

where $\Gamma(\cdot)$ is the Gamma function and K_l is the modified Bessel function of the second kind with order l.[7]

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 heaviness of the tails of the distribution. The mean of the distribution is given by

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

When we implement 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). As seen in (11), this term depends on the *marginal* distribution of asset i's return.

[7]See Aas and Haff (2006), Hu and Kercheval (2010), Birge and Chavez-Bedoya (2016), and Kwak and Prince (2018) for more discussion of the GH skewed t distribution. This distribution is appealing not only for its ability to capture skewness and fat tails in asset returns, but also because it is closed under conditioning, regularization, and linear transformation.

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

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

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

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

The objective function in (9) combines a traditional component, namely mean-variance preferences, with a non-traditional one that incorporates prospect theory. Why do we take the traditional term to be mean-variance preferences, 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 traditional part of our preference specification to be one that delivers CAPM pricing; this will allow us to cleanly identify the deviations from the CAPM that prospect theory generates. The simplest preferences that lead to the CAPM are mean-variance preferences.[8]

3.2 Equilibrium structure

In this section, we discuss the form of the equilibrium in our economy. The equilibrium structure that is typically used to analyze Expected Utility models does not apply for the model in (9). This is a roadblock to understanding prospect theory's implications for the cross-section, and one of our contributions is to surmount it by way of a new equilibrium

[8]This explains why we do not include a preference for portfolio skewness in the Expected Utility component of the objective function. If we did, it would be unclear whether deviations from the CAPM predicted by the model are due to the Expected Utility term – specifically, to coskewness – or to the prospect theory term. Other papers have examined whether coskewness can shed light on stock market anomalies (Harvey and Siddique, 2000); here, we examine whether prospect theory can do so.

structure. Below, we describe three types of equilibrium and explain why we study the one that we do.

Full rationality with homogeneous holdings. At time 0, the investors in our economy are identical in their preferences, their wealth, and their prior gain or loss in each risky asset. It is therefore natural to think that, in equilibrium, at time 0, they would all choose 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 label a "full-rationality equilibrium with homogeneous holdings," is the one used in Expected Utility models with identical investors.

Remarkably, however, for the wide range of parameter values we have examined, this type of equilibrium does not exist for the model in Section 3.1. 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 asset's market supply $\Theta_{M,1}$. This suggests that, to clear the market, we simply need to lower the value of μ_1, as this will lower the asset's expected return. However, it turns out that as we do so, the value of Θ_1 for which the objective function 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.[9]

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 maximum, we can clear the market in each asset.

The difficulty with this equilibrium structure is that it is computationally infeasible to

[9]Why does the value of Θ_1^* at which (9) is maximized jump discontinuously as we lower μ_1? When $b_0 = 0$, the expression in (9), viewed as a function of Θ_1, depends only on Θ_1 and Θ_1^2. It therefore has a single local maximum that is also its global maximum. When $b_0 > 0$, (9) becomes a function of Θ_1, Θ_1^2, *and* Θ_1^α. As such, it can have multiple local maxima. As we lower μ_1, the global maximum switches from one local maximum to another; this, in turn, makes Θ_1^* a discontinuous function of μ_1.

determine if it exists. To see why, suppose that we consider 100 candidate values for each element of the location vector μ, which, in turn, leads to 100^N possible location vectors μ. 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 N-dimensional optimization problem in (9) and determine if there are multiple global maxima. We then need to see whether, by allocating investors to the various maxima, we can clear the market. This procedure is challenging even for $N = 2$ risky assets; for $N = 1000$, the more realistic value we use below, it is completely infeasible.

Bounded rationality with heterogeneous holdings. To overcome 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 – investors' actual portfolios will be less diversified than the market portfolio – but, as we explain below, this discrepancy has a negligible impact on our results.

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 solution to the problem in (9) involves 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), in the range $\Theta_i \in [0, \infty)$, 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, thereby allowing us to clear the market in asset i by allocating some investors to the lower optimum and others to the upper optimum. The restriction $\Theta_i \in [0, \infty)$ indicates that, for simplicity, we are imposing a short-sale constraint.

The bounded-rationality assumption greatly simplifies the investors' optimization problem: by turning the multivariate function in (9) into the univariate function in (20), it converts the search for the optimal allocation Θ_i to asset i into a one-dimensional problem, one where investors trade off a larger allocation to asset i and lower allocation to the risk-free asset against the reverse. Moreover, because the problem is now one-dimensional, it is easy to determine whether the function in (20) has multiple global maxima or a unique global maximum.

We find that a bounded-rationality equilibrium with heterogeneous holdings 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, 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, we find that, for any asset i where investors have heterogeneous holdings, the function in (20) has just two global maxima in the range $\Theta_i \in [0, \infty)$, Θ_i^* and Θ_i^{**}. These maxima straddle the market supply $\Theta_{M,i}$, 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. We also find that Θ_i^* is usually only slightly lower than $\Theta_{M,i}$, while Θ_i^{**} is substantially higher than $\Theta_{M,i}$. As such, to clear the market, we assign the vast majority of investors to the Θ_i^* allocation and the remaining few to the Θ_i^{**} allocation.

To understand the portfolios that the investors in our economy hold, suppose that there are $N = 1000$ risky assets, and that for 500 of them, all investors have identical holdings – for these assets, the function in (20) has a unique global maximum – while for the remaining 500, the function in (20) has two global maxima, so that investors have heterogeneous holdings; this equal split between assets with homogeneous holdings and assets with heterogeneous holdings approximates what we find when we implement the equilibrium. All investors then hold the first 500 assets in proportion to their market weights. For the vast majority of the remaining 500 assets, the fraction of his portfolio that a given investor allocates to each one is slightly lower than its market weight – this is the Θ_i^* optimum – but for a small handful of these assets, he holds a large position, given by the Θ_i^{**} optimum. Overall, then, each investor combines a diversified portfolio of many assets with a small number of concentrated holdings – a portfolio structure that mirrors that of many real-world investors.

An investor's assumption, when solving for his optimal allocation to asset i, that his holdings of the remaining assets equal their market weights, is not exactly correct: by the nature of the heterogeneous-holdings equilibrium, he may have an undiversified position in a small number of these other risky assets. However, this discrepancy has a negligible impact on our model's predictions. The reason is that the investor's assumption is *almost* correct: in the 1000-asset example above, the typical investor holds either the market supply or an amount slightly below market supply for approximately 996 of the 1000 assets. We confirm that, if, when solving for his allocation to asset i, the investor instead makes the *correct* assumption that he has a few undiversified holdings, the quantitative predictions for expected returns are very similar to those of the simpler bounded-rationality equilibrium defined through equations (20)-(22).[10]

To make the model easier to implement, we rescale it. Specifically, let $\Theta_{M,R}$ be the market value of all risky assets relative to the market value of all assets, and define

$$\begin{aligned} \theta_i &= \Theta_i/\Theta_{M,R} \\ \theta_{M,i} &= \Theta_{M,i}/\Theta_{M,R} \\ \theta_{i,-1} &= \Theta_{i,-1}/\Theta_{M,R}. \end{aligned} \tag{23}$$

From now on, we think of investors as choosing θ_i rather than Θ_i. In Appendix B, 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 $\widehat{b}_0$. The rescaled problem is simpler to implement because it is easier to compute an empirical counterpart for $\theta_{M,i}$ than for $\Theta_{M,i}$, and because – see Appendix B for details – the rescaling allows us to simplify the variance term in the first row of (20) by introducing asset i's beta, denoted β_i.[11]

For completeness, we restate the definition of equilibrium in terms of θ_i. There is nothing conceptually new in the equations below; they are simply a rescaled version of (20)-(22). A

[10]We have also studied the following iterative procedure. When solving for his allocation Θ_i to asset i, the investor starts with an assumption A_1 about his remaining holdings – specifically, that they equal market weights. He then uses the resulting optimal portfolio P_1 as a new assumption A_2 about the structure of his portfolio: he again solves for his allocation Θ_i, this time under assumption A_2 about his remaining holdings. He then takes the new optimal portfolio P_2 as a new assumption about his remaining holdings, and so on. This iterative procedure converges to a "self-consistent" heterogeneous-holdings equilibrium, one where the assumption the investor makes about his remaining holdings when solving for his allocation to a particular asset is consistent with the portfolio he actually ends up choosing to hold. The expected asset returns in this self-consistent equilibrium are quantitatively very similar to those in the simpler bounded-rationality equilibrium in equations (20)-(22).

[11]Due to heterogeneous holdings, for almost *no* individual investor will $\sum_i \theta_i$ equal 1. However, $\sum_i \theta_{M,i} = 1$.

bounded-rationality equilibrium consists of a location vector $\mu = (\mu_1, \ldots, \mu_N)$ such that, for each i, and over the range $\theta_i \in [0, \infty)$, the function

$$\begin{aligned}
&\theta_i(\mu_i + \frac{\nu\zeta_i}{\nu - 2} - R_f) - \frac{\widehat{\gamma}}{2}(\theta_i^2\sigma_i^2 + 2\theta_i(\beta_i\sigma_M^2 - \theta_{M,i}\sigma_i^2)) \\
&-\lambda\widehat{b}_0 \int_{-\infty}^{R_f - \theta_{i,-1}g_i/\theta_i} (\theta_i(R_f - R_i) - \theta_{i,-1}g_i)^\alpha dw(P(R_i)) \\
&-\widehat{b}_0 \int_{R_f - \theta_{i,-1}g_i/\theta_i}^{\infty} (\theta_i(R_i - R_f) + \theta_{i,-1}g_i)^\alpha dw(1 - P(R_i)),
\end{aligned} \tag{24}$$

where

$$\widehat{\gamma} = \gamma W_0 \Theta_{M,R} \tag{25}$$

$$\widehat{b}_0 = b_0 W_0^{\alpha-1} \Theta_{M,R}^{\alpha-1}, \tag{26}$$

has either a unique global maximum at $\theta_i = \theta_{M,i}$ or multiple global maxima that straddle $\theta_{M,i}$.

In Appendix C, we explain in full the procedure we use to determine whether investors have identical or heterogeneous holdings in an asset, and to then compute the asset's expected return. Due to the numerical integration in (24), this computation takes a few minutes. This is fast enough for the application we consider in this paper. In an Online Appendix, we derive a faster approximation that is accurate for assets with very low market supply $\theta_{M,i}$ and that may be useful in other applications.

4 Anomalies and Model Parameter Values

In Section 3, we presented a model that generates quantitative predictions about the cross-section of average returns when investors evaluate risk according to prospect theory. We now use our framework to answer a basic but long-standing question: Can prospect theory shed light on stock market anomalies? To be as comprehensive as possible, subject to the computational constraints we face, we consider 22 prominent anomalies. They are listed in Table 1, along with the abbreviations we use in subsequent tables to refer to them; Appendix D reports the predictor variable associated with each anomaly. 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 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 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, stocks 1 to 100 belong to size decile 1, stocks 101 to 200 to size decile 2, and so on. All stocks in a given decile are identical: they have the same characteristics, namely, the empirical characteristics of the typical stock in that anomaly decile. For each decile in turn, we choose one stock at random and compute our model's prediction for its expected return. Since all stocks in a given decile are identical, this immediately tells us the expected return of *all* the stocks in that decile. Our model can help explain the size anomaly if the expected return it predicts for the randomly-chosen stock in decile 1 is higher than the expected return it predicts for the randomly-chosen stock in decile 10.[12]

What are the empirical inputs we need to compute the expected return of a stock in our model? Equation (24) shows that, to determine μ_i, and hence stock i's expected return, we need to know σ_i, ζ_i, g_i, and β_i. In words, to compute our model's prediction for the expected 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 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 book-to-market ratio in the case of the value 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 record the returns, over the next year, of the 100 stocks in that decile. We then calculate 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 percentage capital gain or loss in the stock for the average investor in the stock – and then

[12]Why do we not simply consider an economy with $N = 10$ risky 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 weight in the market portfolio. We therefore need to capture the fact that, in reality, any given stock makes up only a small fraction of the overall market. Setting $N = 1000$ allows us to do this. Since all 100 stocks in each decile are identical, the computational burden in the case of $N = 1000$ is the same as for $N = 10$.

average these 100 numbers. Finally, for each of the 100 stocks, we compute its beta [HOW?] and then average these 100 numbers. At the end of this process, we have four quantities in hand for each anomaly decile in that month: the volatility, skewness, gain overhang, and beta 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 four quantities: return volatility, return skewness, capital gain overhang, and beta. In the final step, we compute the mean of each time series. For each market capitalization decile, this leaves us with four numbers pertaining to the typical stock in that decile: the standard deviation of its returns; its return skewness; its capital gain overhang; and its beta. We feed these four numbers into our model to see what it predicts for the expected return of the typical stock in that anomaly decile.

In our calculations, we compute the volatility and skewness of *annual* stock returns. Why is this? In our model, we focus on decision-making at time 0, which lies somewhere between time -1, when the investor purchases a stock, and time 1, when he disposes of it. Over our sample period, stock market turnover is approximately 50% per year, which implies a holding period for an individual stock of about two years. If the interval between time -1 and time 1 is two years, it is natural to take the interval between time 0 and time 1 to be one half of this, namely one year.

Our approach to computing beta and gain overhang differs slightly from our approach to computing volatility and skewness. To compute the beta and gain overhang of the typical stock in some anomaly decile, we take the betas and capital gains of the 100 stocks in that decile, and average them. To compute the volatility and skewness of the typical stock in the decile, we instead use a "forward-looking" approach: we compute these quantities as as the *cross-sectional* volatility and skewness of the subsequent returns of the 100 stocks in the decile. We adopt this approach because it focuses on what a rational investor is interested in: the likely volatility and skewness of a stock going forward, rather than the stock's past volatility and skewness.[13]

Before presenting the empirical characteristics of the 220 anomaly deciles, we clarify

[13]We have also considered an alternative forward-looking approach. For each of the 100 stocks in a given anomaly decile in a given month, we compute the volatility and skewness of the stock's daily returns over the next year and then average these quantities across the 100 stocks. We can then use the volatility and skewness of daily returns to make inferences about the volatility and skewness of annual returns. However, this last step presents some difficulties, in that it relies on additional assumptions about the autocorrelation of stock returns. We prefer our "cross-sectional" approach as a simpler and more reliable way of estimating volatility and skewness at the annual horizon that is our focus.

the definition of one key variable: the capital gain overhang. 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 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 calculate the average investor's capital gain or loss in the stock. Frazzini (2006) uses data on mutual fund holdings to compute the average gain or loss in a stock across the mutual fund investors in that stock. We computed both measures of gain overhang, and find that they lead to similar quantitative estimates and similar model predictions. We therefore pick one – specifically, the Grinblatt and Han (2005) measure, because it is simpler and accounts for both individual and institutional investors – and stick with it throughout.

Table 2 presents the results of the above empirical exercise. The first column of the table lists the 22 anomalies we are interested in. The second and third columns report, for each anomaly, the value-weighted return of decile 1 stocks and decile 10 stocks, respectively, averaged across the 618 months of our sample and annualized. 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 standard deviation of the annual return of the typical stock in deciles 1 and 10, respectively, computed as described above. The sixth and seventh columns list the skewness of the annual return of the typical stock in deciles 1 and 10, respectively. Finally, the eighth and ninth columns report the capital gain overhang of the typical stock in deciles 1 and 10, respectively.

We make two observations about the results in Table 2. First, for most of the anomalies, the typical stock in decile 1 differs substantially from the typical stock in decile 10 in its return volatility, return skewness, and gain overhang – in other words, in the three characteristics that, aside from beta, determine expected returns in our model. Consider, for example, the size anomaly: the typical stock in decile 1 has an annual return standard deviation of 76%, while the typical stock in decile 10 has an annual return standard deviation of just 25%. Similarly, while the typical small-cap stock has an annual return skewness of 4.3, the typical large-cap stock has an annual return skewness of just 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 second, more striking, 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, and vice-versa; the only exception is for post-earnings announcement drift (PEAD). Furthermore, for

21 of the 22 anomalies – the only exception is the net operating assets (NOA) 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, and vice-versa.

Figure 2 illustrates these relationships. 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 standard deviation and gain overhang, and between skewness 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, for one of the extreme decile portfolios – decile 1, say – the typical stock in that decile has a higher return skewness, higher return volatility, and lower capital gain overhang than the typical stock in the other extreme decile, decile 10; again, 20 of the 22 anomalies follow this pattern. It is then impossible to tell, without a quantitative model, whether prospect theory can 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, while decile 10 stocks trade at a gain. Due to diminishing sensitivity, 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 with prospect theory investors that takes the empirical values of return volatility, return skewness, and capital gain overhang and uses them to predict an expected return.

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

4.1 Parameter values

To see if our model can capture a particular anomaly, we proceed as follows. We consider an economy with $N = 1000$ stocks, and assign 100 of these stocks to each anomaly decile: stocks 1 to 100 belong to anomaly decile 1, stocks 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, capital gain overhang, and beta, namely the empirical standard deviation, skewness, capital gain overhang, and beta of the typical stock in that anomaly decile, computed as described above. We set the parameters S, ζ, and g 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 (24) are satisfied. Our model's prediction for assets' expected returns is then given by (14). Note that all stocks in a given decile will have the same μ_i and hence the same expected return.

We now explain in more detail how we parameterize the model. While the model features several parameters, all of them 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 stocks; S, the dispersion matrix for stock returns; ζ, the vector of asymmetry parameters for stock returns; ν, the degree of freedom parameter; g, the vector of capital gains; σ_M, the standard deviation of stock market returns; and θ_M, the vector of market weights for the N stocks. The investor-level parameters are γ, portfolio risk aversion; b_0, the importance of the prospect theory term in investor preferences; $(\alpha, \delta, \lambda)$, the prospect theory preference parameters; and θ_{-1}, the vector of investors' prior allocations to the N stocks.

We start with the asset-level parameters, and in particular, the parameters of the GH skewed t distribution. We set $\nu = 7.5$, which represents a substantial degree of fat-tailedness in stock returns. 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,

$$\text{Std}(\widetilde{R}_i) = \left[\frac{\nu}{v-2} S_{ii} + \frac{2\nu^2}{(\nu-2)^2(\nu-4)}\zeta_i^2\right]^{0.5} \tag{27}$$

$$\text{Skew}(\widetilde{R}_i) = \frac{2\zeta_i\sqrt{\nu(\nu-4)}}{\sqrt{S_{ii}}(2\nu\zeta_i^2/S_{ii} + (\nu-2)(\nu-4))^{\frac{3}{2}}}\left[3(\nu-2) + \frac{8\nu\zeta_i^2}{S_{ii}(\nu-6)}\right]. \tag{28}$$

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 (27) and (28). These equations then allow us to solve for the two unknowns, S_{ii} and ζ_i. For example, in the case of the size anomaly,

stocks 1 to 100 belong to the lowest market capitalization decile. From Table 2, we see that the empirical standard deviation and skewness for the typical stock in this decile are 0.76 and 4.27, respectively. Accordingly, for this anomaly, to set the values of S_{ii} and ζ_i for $i \in \{1, \ldots, 100\}$, we solve

$$0.76 = \left(\frac{7.5}{7.5-2} S_{ii} + \frac{2(7.5)^2}{(7.5-2)^2(7.5-4)} \zeta_i^2 \right)^{0.5} \tag{29}$$

$$4.27 = \frac{2\zeta_i \sqrt{7.5(7.5-4)}}{\sqrt{S_{ii}}(2(7.5)\zeta_i^2/S_{ii} + (7.5-2)(7.5-4))^{\frac{3}{2}}} \left[3(7.5-2) + \frac{8(7.5)\zeta_i^2}{S_{ii}(7.5-6)} \right]. \tag{30}$$

We do not need to set the off-diagonal elements of S. As a consequence of our bounded-rationality assumption, all the information investors need about asset i's covariance with other assets is contained in its beta. For a given stock i that belongs to some anomaly decile, we set its beta equal to the empirical beta of the typical stock in that decile.

In terms of asset-level parameters, this leaves g, the vector of capital gains; σ_M, the standard deviation of annual stock market returns; θ_M, the vector of market weights; and the gross risk-free rate R_f. For a stock 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 Table 2 for the two extreme deciles of each anomaly. We set σ_M to the empirically reasonable level of 0.25, and the gross risk-free rate R_f to 1.

We set θ_M, the vector of market weights, to match empirical market weights. Take, for example, the volatility anomaly. In each month of our sample, we compute the fraction of the total market value of all stocks in our sample in that month that is made up by the stocks in each volatility decile. We then compute the time-series averages of these proportions. We find that, on average, volatility decile 1 makes up 29.6% of total stock market value. Since, in our model, there are 100 identical stocks in decile 1, we set $\theta_{M,i} = 0.296/100$ for all stocks in decile 1, in other words, for $i = 1, \ldots, 100$. We proceed similarly for the other deciles.

We now turn to the investor-level parameters. We set $\widehat{\gamma}$, the scaled portfolio risk aversion in (25), and $\widehat{b}_0$, the scaled weight on the prospect theory term in (26), to generate an aggregate equity premium of around 6%. There are many pairs $(\widehat{\gamma}, \widehat{b}_0)$ that produce an equity premium of 6%. How do we choose one? As we increase b_0, we not only increase the equity premium, but also the size of the predicted deviations from the CAPM, as well as investors' degree of under-diversification. To put an approximate upper bound on the level of mispricing that prospect theory investors can generate, we therefore choose, from the among the $(\widehat{\gamma}, \widehat{b}_0)$ pairs that generate a 6% equity premium, the one with the highest level of $\widehat{b}_0$ that still produces plausible levels of under-diversification. This is the pair $(\widehat{\gamma}, \widehat{b}_0) = (0.6, 0.6)$. We set $\theta_{i,-1}$,

investors' allocation to stock i at time -1, to a neutral value, namely $\theta_{M,i}$, the weight of stock i in the market portfolio of risky assets, which, as noted above, is based on empirical values.

Finally, we set the preference parameters α, δ, and λ. A well-known set of values for these parameters comes from Tversky and Kahneman (1992), who estimate $(\alpha, \delta, \lambda) = (0.88, 0.65, 2.25)$ for the median participant in their experiment. However, these estimates are almost 30 years old and are based on a small number of participants. Given that the values we assign to these parameters will play an important role in our results, it seems prudent to base these values on a wide range of studies, not just one.

Tversky and Kahneman's (1992) results have led to the widespread view that the degree of loss aversion λ is approximately 2. However, recent studies indicate that the true level of loss aversion in the population is significantly lower. In a meta-analysis of experimental estimates of loss aversion, Walasek, Mullett, and Stewart (2018) find that the median estimate of λ is just 1.31. Chapman et al. (2018) further argue that these experimental estimates may be based on non-representative samples of participants. Using a large, representative sample, they obtain estimates of λ that are lower still – as low as $\lambda = 0.98$ for their median participant, albeit somewhat higher for those with greater cognitive ability. To reflect these findings, in a conservative way, we set $\lambda = 1.5$.

Booij, van Praag, and van de Kuilen (2010) compile a list of experimental estimates of α and δ. The median estimate of δ is close to Tversky and Kahneman's (1992) estimate; we therefore maintain $\delta = 0.65$. Experimental estimates of α span a fairly wide range; most lie between 0.5 and 0.95. We set α near the midpoint of this range, at 0.7.

In summary, then, our preference parameter values are

$$(\alpha, \delta, \lambda) = (0.7, 0.65, 1.5). \tag{31}$$

While these values reflect our reading of the experimental literature to date, there is still substantial uncertainty about them. Given this, one interpretation of our results in Section 5 is that they indicate the levels of loss aversion, diminishing sensitivity, and probability weighting that are most consistent with the evidence on stock market anomalies.[14]

By setting the stock-level parameters – specifically, S, ζ, and g – in the way described

[14]The values of α, δ, and λ in (31) are also in line with prior research on financial applications of prospect theory. Barberis and Xiong (2009) find that prospect theory is more consistent with investor trading behavior for values of α and λ that are lower than those estimated by Tversky and Kahneman (1992). Meanwhile, even the lower level of loss aversion in (31) is strong enough to generate a high equity premium and non-participation in the stock market – two prominent applications of loss aversion in finance.

above, we are assuming that investors have sensible beliefs: they are aware of the return volatility and return skewness of the typical stock in each anomaly decile, and of its gain overhang. This is a natural case to study. However, it is possible that, for certain types of stocks – for value stocks, say, or for small-cap stocks – investors do not have accurate beliefs. In Section 6, we use evidence on investor beliefs derived from option prices to suggest that, while investors often do have accurate beliefs, they occasionally do not, and that this may explain some of the model's failures.

4.2 Illustration of the equilibrium structure

At date 0, 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. Instead, the equilibrium we study involves heterogeneous holdings.

Now that we have parameterized the model, we can illustrate these heterogeneous holdings, using the momentum anomaly as our example. We note that it is only for *some* of the stocks that the investors have heterogenous holdings. For other stocks, they have identical holdings, each holding the market supply of each stock, $\theta_{M,i}$. This is the case for all stocks in momentum decile 1, namely stocks 1 to 100. For example, for stock 1, there is a value of μ_1, the location parameter for stock 1, such that the function in (24), for $i = 1$, has a unique global maximum at $\theta_1 = \theta_{M,1} = 0.000185$. Figure 3 plots the function in (24) for this value of μ_1, namely $\mu_1 = 0.0124$. The global maximum at $\theta_1^* = \theta_{M,1}$ 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 (24), for $i = 901$, has a global maximum at $\theta_{901} = \theta_{M,901} = 0.00075$. We can see this in Figure 4. The dashed line plots the function in (24) for $\mu_{901} = 0.5793$. For this value of μ_{901}, the function has a unique global maximum at $\theta_{901} = 0.14 > \theta_{M,901}$. 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 0.5743, 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 0.5768. For this value

of μ_{901}, there are *two* global optima, one at $\theta^*_{901} = 9.5\text{E-}5 < \theta_{M,901}$ and one at $\theta^{**}_{901} = 0.119 > \theta_{M,901}$. Since these optima straddle the market supply, we can clear the market by allocating most investors to the first optimum and the rest to the other optimum.

5 Application

In this section, we use the model developed in Section 3 and parameterized in Section 4 to answer a basic but long-standing question: Can the risk attitudes captured by prospect theory shed light on the prominent anomalies in Table 1?

To determine whether our model can help explain an anomaly, we focus on anomaly alphas. For any given anomaly, we compute the empirical annual alphas for the ten anomaly deciles over our 1963-2014 sample, and denote them as $\alpha^d(1), \ldots, \alpha^d(10)$, where the "$d$" superscript stands for "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 $\overline{R}_M = \sum_{i=1}^{N} \theta_{M,i}\overline{R}_i$.

We say that our model can help explain an anomaly if

$$\text{sign}(\alpha^d(10) - \alpha^d(1)) = \text{sign}(\alpha^m(10) - \alpha^m(1)) \text{ and } |\alpha^m(10) - \alpha^m(1)| > 0.015. \tag{32}$$

The first condition in (32) 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 (32) is that the model makes a "strong" prediction, in other words, predicts a substantial difference between the two alphas; while the 1.5% cutoff is somewhat arbitrary, it allows for a simple way of organizing our results. Similarly, we say that the model fails to explain an anomaly if

$$\text{sign}(\alpha^d(10) - \alpha^d(1)) = -\text{sign}(\alpha^m(1) - \alpha^m(1)) \text{ and } |\alpha^m(10) - \alpha^m(1)| > 0.015, \tag{33}$$

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

$$|\alpha(10) - \alpha(1)| < 0.015. \tag{34}$$

We find that our model is helpful for thinking about a strikingly large number – a majority, in fact – 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 13 anomalies, the conditions in (32) are satisfied. Specifically, these are the momentum, failure probability, idiosyncratic volatility, profitability, idiosyncratic skewness, return on equity, 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 is 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 in 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; indeed, for most of these 13 anomalies, our model is able to explain a large fraction of the spread in empirical alphas. The intuition for why our model helps to explain these anomalies is the following.

For these anomalies, the extreme decile with the lower empirical alpha – for example, decile 1 in the case of the momentum anomaly and decile 10 in the case of the volatility anomaly – 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 – one that accounts for all the relevant characteristics, including gain overhang, and for all the elements of prospect theory – that we can properly evaluate whether prospect theory can explain the anomalies.[15]

Figures 5 to 11 show that our model can explain not only the alphas for the extreme deciles, but also those for the intermediate deciles. In particular, for several of the anomalies, the dashed lines, which plot the empirical alphas, are concave: the alphas are similar for most deciles, but fall rapidly as we approach the extreme decile with the most skewed stocks; this pattern is particularly stark for the volatility and failure probability anomalies in Figures 6 and 7, but is present for a number of other anomalies as well. The solid lines in the graphs show, strikingly, that our model captures this concavity. This successful prediction is a direct consequence of the heterogeneous-holdings equilibrium structure, and shows that, while this structure is novel and unusual, it has useful predictive content.

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). 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 that this mechanism helps to explain these anomalies, but also shows that the argument is incomplete in important

[15]Our focus is on understanding unconditional anomaly returns. However, our framework may also be helpful for thinking about other facts about anomaly returns. For example, Wang, Yan, and Yu (2017) and An et al. (2019) show that the low returns on volatile stocks relative to less volatile stocks, and on skewed stocks relative to less skewed stocks, are observed primarily among stocks trading at a loss, a pattern that, the authors suggest, may be explained by prospect theory. Although beyond the scope of our analysis, the model we present in this paper may be helpful for studying and formalizing the authors' hypothesis.

ways. It is not just that stocks in momentum decile 10 trade at a gain; their returns are also less volatile and less skewed than the returns of 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 this mechanism, 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, also affect their average returns. We show that, when all three characteristics 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.

Figures 5 to 11 also draw a connection between prospect theory and a number of 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 equity, external finance, composite equity issuance, net stock issuance, and difference of opinion 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 (33), 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 12 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 fails to explain the positive historical alpha for decile 1, which contains the stocks with the lowest market capitalizations. These stocks are very volatile, which, all else equal, leads investors to charge a high average return on them. However, they also have positively-skewed returns and trade at a loss, which leads investors to charge a low average return on them. Our analysis shows that this second effect overwhelms the first one.

Figure 13 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; this leads investors to charge a lower average return on value stocks. 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 14 and the top-left panel in Figure 15 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. In Section 6, we discuss some possible reasons why the model performs poorly for some 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 that, as specified in (34), it predicts alphas for the two extreme decile portfolios that differ by less than 1.5% in absolute magnitude. The lower panel of Figure 15 presents the results for these anomalies.

6 Discussion

In Section 5, we used the framework of Sections 3 and 4 to see if prospect theory can help explain 22 prominent anomalies, and found that it is able to shed light on a majority of them. We now discuss some other issues raised by our analysis. In Section 6.1, we revisit the assumption that investors have sensible beliefs about the return volatility and return skewness of the typical stock in each anomaly decile. And in Section 6.2, we review the limits to arbitrage that allow investors with prospect theory preferences to have a substantial

impact on asset prices, and discuss how allowing for investor heterogeneity would impact our results.

6.1 Investor beliefs

Figures 12 to 15 show that, for some anomalies, our model performs poorly. Why is this? One answer is that, for some anomalies, the risk attitudes captured by prospect theory are not the primary driver of average returns. 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 most relevant factor.

There is another answer, however, one that applies within the prospect theory framework. In generating Figures 5 to 15, we assumed that investors have accurate beliefs about stocks' return volatility, return skewness, and gain overhang. However, it may be that, for certain types of stocks, investors' beliefs about these characteristics are incorrect.

To investigate this, we take data on the prices of stock options from the IvyDB OptionMetrics database, and use them to extract estimates of risk-neutral volatility and risk-neutral skewness; we describe our methodology in more detail in Appendix E. While these are estimates of risk-neutral quantities, they nonetheless reflect investors' beliefs about the physical return distribution, and prior studies have exploited this in useful ways (Birru and Wang, 2016).

Table 3 reports, for each anomaly, the 30-day risk-neutral variance and risk-neutral skewness of the typical stock in each of the two extreme anomaly deciles. One-year risk-neutral moments would be a better match for the volatility and skewness of annual returns that we report in Table 2. We focus on 30-day moments because short-maturity options are more heavily traded and therefore likely to lead to more reliable estimates. However, we have also computed 60-day and one-year moments and find that these lead to similar conclusions. We compute the risk-neutral moments using a similar procedure to that described in Section 4 for the physical moments. For a given month and anomaly decile, we compute the risk-neutral moments for each stock in the decile for which option data are available, and then average these results across the stocks in the decile. We then compute time-series means across all months in the sample, which runs from January 1996 to December 2012.

Comparing the results in Table 3 to those in Table 2, we see that, for most types of stocks, investors appear to have sensible beliefs: the risk-neutral volatility and skewness are almost

always higher for stocks whose actual return volatility and skewness are higher. There is, however, one notable exception: 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 explain why our model performs poorly on the value anomaly. To generate Figure 13, 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 the return skewness of value stocks and growth stocks is similar, this will increase the average return that our model predicts for value stocks, reducing the gap between the empirical and model-predicted alphas.[16]

6.2 Limits to arbitrage and investor heterogeneity

A theme of this paper is that investors who evaluate risk according to prospect theory leave a footprint on asset prices. Can traditional Expected Utility investors attenuate these price effects?

For several reasons, it is difficult for Expected Utility investors to do so. One reason is the form of the mispricing that prospect theory investors generate – specifically, *mild* underpricing but *severe* overpricing. This pattern characterizes the model's predictions for almost all of the 22 anomalies. For example, in Figure 7, which corresponds to the volatility anomaly, stocks in deciles 1 to 7 are slightly underpriced, while the highly-skewed stocks in deciles 9 and 10 are very overpriced. Moreover, the overpriced stocks are not only highly skewed, but also tend to be small-cap stocks. To significantly attenuate the mispricing caused by prospect theory investors, Expected Utility investors need to correct the overpricing; what underpricing there is, is already small in magnitude. But to correct the overpricing, they need to short a large number of highly-skewed, small-cap stocks, a strategy that entails high costs and fees.

Expected Utility investors also face other, more general limits to arbitrage. For most anomalies, the stocks in each extreme decile comove in their returns; for example, there is a common factor in the returns of value stocks, and also in the returns of growth stocks. As such, a strategy that buys the stocks in one extreme decile and shorts the stocks in the other can have very volatile returns, which limits arbitrage. The risk that the mispricing

[16] Bordalo, Gennaioli, and Shleifer (2013) suggest that investors find the potential upside of growth stocks, and the potential downside of value stocks, to be more salient. Such a mechanism may underlie investors' incorrect beliefs about the relative skewness of value and growth stock returns.

will worsen in the short run, leading to outflows or margin calls, is an additional deterrent to arbitrageurs (Shleifer and Vishny, 1997).

Finally, to exploit the mispricing caused by prospect theory investors, the Expected Utility investors would first need to detect it, and this can take many years. For example, the idiosyncratic volatility anomaly was present in the data for decades before it became widely known with the publication of Ang et al. (2006). If some part of the volatility anomaly was due to the actions of prospect theory investors, these actions went undetected for a very long time.

In our model, investors at time 0 are identical in their wealth, preferences, and past gain or loss in each risky asset. It is beyond the scope of our paper to formally explore investor heterogeneity; the case we consider is challenging enough. However, the reasons given above for why Expected Utility investors are unlikely to attenuate the mispricing caused by prospect theory investors also suggest that our predictions will not be strongly affected by taking heterogeneity among prospect theory investors into account.

For example, in our model, a risky asset is priced by the investors who are holding it. Suppose that we also take into account prospect theory investors who are not holding the asset. Such investors may attenuate the underpricing shown in Figures 5 to 15, but this underpricing is mild in the first place. The more severe mispricing is *over*-pricing – but, as noted above, to exploit this, prospect theory investors not holding the overpriced assets would need to short them in large quantities, and this is a costly strategy.

7 Conclusion

We present a new model of asset prices in which investors evaluate risk according to prospect theory and examine its ability to explain 22 prominent stock market anomalies. The model incorporates all the elements of prospect theory, takes account of investors' prior gains and losses, and makes quantitative predictions about an asset's average return based on empirical estimates of its beta, volatility, skewness, and capital gain overhang. 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.

Appendix A: The probability weighting terms

Here, we lay out in full the $dw(P(R_i))$ and $dw(1-P(R_i))$ terms which appear in (11), (20), and (24). We can write

$$dw(P(R_i)) = \frac{dw(P(R_i))}{dP(R_i)}\frac{dP(R_i)}{dR_i}dR_i.$$

By differentiating the probability weighting function in (6), and using P as shorthand for $P(R_i)$, we can write the right-hand side as

$$\frac{\delta P^{\delta-1}(P^\delta+(1-P)^\delta)-P^\delta(P^{\delta-1}-(1-P)^{\delta-1})}{(P^\delta+(1-P)^\delta)^{1+\frac{1}{\delta}}}p(R_i)dR_i,$$

where the probability density function $p(R_i)$ is given in (15) and (16). Similarly,

$$\begin{aligned} dw(1-P(R_i)) &= \frac{dw(1-P(R_i))}{dP(R_i)}\frac{dP(R_i)}{dR_i}dR_i \\ &= -\frac{\delta(1-P)^{\delta-1}(P^\delta+(1-P)^\delta)-(1-P)^\delta((1-P)^{\delta-1}-P^{\delta-1})}{(P^\delta+(1-P)^\delta)^{1+\frac{1}{\delta}}}p(R_i)dR_i. \end{aligned}$$

Appendix B: Rescaling the decision problem

Substituting the definitions in (23) into (20) and multiplying the resulting expression by the exogenous parameter $\Theta_{M,R}^{-1}$, we obtain

$$\begin{aligned} &\theta_i(\mu_i+\frac{\nu\zeta_i}{\nu-2}-R_f)-\frac{\widehat{\gamma}}{2}(\theta_i^2\sigma_i^2+2\sum_{j\neq i}\sigma_{ij}\theta_i\theta_{M,j}) \\ &-\lambda\widehat{b}_0\int_{-\infty}^{R_f-\theta_{i,-1}g_i/\theta_i}(\theta_i(R_f-R_i)-\theta_{i,-1}g_i)^\alpha dw(P(R_i)) \\ &-\widehat{b}_0\int_{R_f-\theta_{i,-1}g_i/\theta_i}^{\infty}(\theta_i(R_i-R_f)+\theta_{i,-1}g_i)^\alpha dw(1-P(R_i)), \end{aligned} \tag{35}$$

where

$$\widehat{\gamma}=\gamma W_0\Theta_{M,R}, \qquad \widehat{b}_0=b_0W_0^{\alpha-1}\Theta_{M,R}^{\alpha-1}.$$

It follows that, if Θ_i maximizes (20), then $\theta_i=\Theta_i/\Theta_{M,R}$ maximizes (35), and conversely that, if θ_i maximizes (35), then $\Theta_i=\theta_i\Theta_{M,R}$ maximizes (20). Maximizing (20) is therefore equivalent to maximizing (35).

The rescaling also allows us to simplify the variance term in the first row of (35). Specifically, the quantity

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

can be rewritten as

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

where $\widetilde{R}_M$ is the return on the market portfolio of all risky assets, σ_M is the standard deviation of this return, and β_i is the beta of asset i. Substituting this into (35) leads to the expression in (23).

Appendix C: Procedure for computing expected returns

The equilibrium structure we consider – a bounded-rationality equilibrium with heterogeneous holdings – consists of a location vector $(\mu_1, \ldots, \mu_N)$ such that, for each i, and in the range $\theta_i \in [0, \infty)$, the expression in (24) 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 explain how we compute this heterogeneous-holdings equilibrium. For a given risky asset i, we first check whether investors have identical 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 (24), substitute in $\theta_i = \theta_{M,i}$, and set the resulting expression to 0. This gives:[17]

$$\begin{aligned} 0 = {} & (\mu_i + \frac{\nu\zeta_i}{\nu - 2} - R_f) - \widehat{\gamma}\beta_i\sigma_M^2 \\ & -\alpha\lambda\widehat{b}_0 \int_{-\infty}^{R_f - \theta_{i,-1}g_i/\theta_i} (\theta_{M,i}(R_f - R_i) - \theta_{i,-1}g_i)^{\alpha-1}(R_f - R_i)dw(P(R_i)) \\ & -\alpha\widehat{b}_0 \int_{R_f - \theta_{i,-1}g_i/\theta_i}^{\infty} (\theta_{M,i}(R_i - R_f) + \theta_{i,-1}g_i)^{\alpha-1}(R_i - R_f)dw(1 - P(R_i)). \end{aligned} \tag{36}$$

We then check whether, for the μ_i that solves (36), the function in (24) has a global maximum at $\theta_i = \theta_{M,i}$, as opposed to only a local maximum or a local minimum. If $\theta_i = \theta_{M,i}$ indeed corresponds to 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 (24) does not have a global maximum at this θ_i, then, in equilibrium, investors do not have identical holdings of asset i. We must instead look for an equilibrium with heterogeneous holdings of this asset.

To find an equilibrium with heterogeneous holdings of asset i, we search for a value of μ_i such that the maximum value of the function in (24) in the range $[0, \theta_{M,i})$, attained at $\theta_i = \theta_i^*$, say, is equal to the maximum value of the function in the range $(\theta_{M,i}, \infty)$, attained at $\theta_i = \theta_i^{**}$. If we find such a μ_i, then there is an equilibrium where investors have heterogeneous holdings of the asset, with some investors allocated to $\theta_i = \theta_i^*$ and others allocated to $\theta_i = \theta_i^{**}$. The value of μ_i for which this holds is typically close to the value of μ_i that solves the first-order condition (36). Therefore, if we find that, for the value of μ_i that solves (36), the function in (24) does not have a global maximum at $\theta_i = \theta_{M,i}$, we search in the neighborhood of that

[17] As discussed in Section 4, we take $\theta_{i,-1} = \theta_{M,i}$, which further simplifies (36).

μ_i for an equilibrium with heterogeneous holdings of asset i.

Appendix D: Stock market anomalies

Here, we list the predictor variable associate with each of the 22 anomalies that we study.

VOL. Idiosyncratic volatility. Standard deviation of the residuals from a regression of daily stock returns on daily innovations of the three Fama-French factors over the past month. See Ang, Hodrick, Xing, and Zhang (2006).

SIZE. Market capitalization. The log of the product of price per share and number of shares outstanding, computed at the end of the previous month.

VAL. Book-to-market. The log of book value of equity scaled by market value of equity, computed following Fama and French (1992) and Fama and French (2008); firms with negative book value are excluded from the analysis.

EISKEW. Expected idiosyncratic skewness, computed as in Boyer, Mitton, and Vorkink (2010).

MOM. Momentum. Measured at time t as the stock's cumulative return from the start of month $t-12$ to the end of month $t-2$.

FPROB. Failure probability. Estimated using a dynamic logit model with both accounting and equity market variables as explanatory variables. See Campbell, Hilscher, and Szilagyi (2008).

OSC. O-Score. Uses accounting variables to estimate the probability of bankruptcy. See Ohlson (1980).

NSI. Net stock issuance. Growth rate of split-adjusted shares outstanding over the previous fiscal year. See Stambaugh, Yu, and Yuan (2012).

CEI. Composite equity issuance. Five-year change in number of shares outstanding, excluding changes due to dividends and splits. See Daniel and Titman (2006).

ACC. Accruals. See Sloan (1996).

NOA. Net operating assets. See Hirshleifer, Hou, Teoh, and Zhang (2004).

PROF. Gross profitability. Measured as revenue minus cost of goods sold at time t, divided by assets at time $t-1$. See Novy-Marx (2010).

AG. Asset growth. Yearly percentage change in total assets. See Cooper, Gulen, and Schill (2008).

ROE. Return on equity. Net income dividend by book equity. See Chen, Novy-Marx, and Zhang (2011).

INV. Investment to assets. The annual change in gross property, plant, and equipment plus the annual change in inventory, scaled by the lagged book value of assets. See Stambaugh, Yu, and Yuan (2012).

MAX. A stock's maximum one-day return in month $t-1$. See Bali, Cakici, and Whitelaw (2011).

ORGCP. Organizational capital. See Eisfeldt and Papanikolaou (2013).

LTREV. Long-term reversal. The stock's cumulative return from the start of month $t-60$ to the end of month $t-13$.

XFIN. External finance. Total net external financing scaled by total assets. See Bradshaw, Richardson, and Sloan (2006).

STREV. Short-term reversal. The stock's return in month $t-1$.

DOO. Dispersion of opinion. The standard deviation of earnings forecasts (unadjusted IBES file, item STDEV) divided by the absolute value of the consensus mean forecast (unadjusted file, item MEANEST). We use the forecasts for the current fiscal year. See Diether, Malloy, and Scherbina (2002).

PEAD. Post-earnings announcement drift. Measured as standardized unexpected earnings: the change in the quarterly earnings per share from its value four quarters before divided by the standard deviation of this change in quarterly earnings over the previous eight quarters. See Foster, Olsen, and Shevlin (1984).

Appendix E: Inferring beliefs from option prices

In Section 6.1, we use option prices to back out estimates of the risk-neutral volatility and skewness investors perceive for individual stocks. Below, we describe our methodology, which follows that of Birru and Wang (2016).

We compute risk-neutral volatility and risk-neutral skewness using equations (1)-(4) in Birru and Wang (2016). These equations were originally derived by Bakshi, Kapadia, and Madan (2003); see their Theorem 1. In principle, these equations should be implemented

using a continuum of options, but since only a finite number of options are traded on any given stock, we use a discrete approximation instead. We implement the calculation only for stocks with at least two out-of-the-money (OTM) puts and two OTM calls; Dennis and Mayhew (2012) argue that the biases that arise from the discrete approximation are small in this case. We compute a 30-day risk-neutral variance and skewness 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 use the option with the longest (shortest) available maturity. We have also computed 60-day and one-year risk-neutral moments; these lead to similar conclusions.

Our options data comes from the IvyDB OptionMetrics 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 with 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; options with zero open interest; options with special settlement arrangements; and options for which the underlying stock price is under \$10. We take the option price to be the midpoint of the best bid and best offer.

References

Aas, K. and I. Haff (2006), "The Generalized Hyperbolic Skew Student's t-Distribution," *Journal of Financial Econometrics* 4, 275-309.

An, L., Wang, H., Wang, J., and J. Yu (2019), "Lottery-related anomalies: The Role of Reference-dependent Preferences," forthcoming, *Management Science*.

Andrikogiannopoulou, A. and F. Papakonstantinou (2019), "History-dependent Risk Preferences: Evidence from Individual Choices and Implications for the Disposition Effect," Working paper, University of London.

Ang, A., Hodrick, R., Xing, Y., and X. Zhang (2006), "The Cross-section of Volatility and Expected Returns," *Journal of Finance* 61, 259-299.

Baele, L., Driessen, J., Ebert, S., Londono, J., and O. Spalt (2018), "Cumulative Prospect Theory, Option Returns, and the Variance Premium," *Review of Financial Studies* 32, 3667-3723.

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.

Bali, T., Cakici, N., and R. Whitelaw (2011), "Maxing Out: Stocks as Lotteries and the Cross-section of Expected Returns," *Journal of Financial Economics* 99, 427-446.

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. and X. Wiong (2009), "What Drives the Disposition Effect? An Analysis of a Long-standing Preference-based Explanation," *Journal of Finance* 64, 751-784.

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.

Barberis, N., Mukherjee, A., and B. Wang (2016), "Prospect Theory and Stock Returns: An Empirical Test," *Review of Financial Studies* 29, 3068-3107.

Birge, J. and L. Chavez-Bedoya (2016), "Portfolio Optimization Under a Generalized Hyperbolic Skewed t Distribution and Exponential Utility," *Quantitative Finance* 16, 1019-1036.

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.

Bordalo, P., Gennaioli, N., and A. Shleifer (2013), "Salience and Asset Prices," *American Economic Review Papers and Proceedings* 103, 623-628.

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

Chapman, J., Snowberg, E., Wang, S., and C. Camerer (2018), "Loss Attitudes in the U.S. Population: Evidence from Dynamically Optimized Sequential Experimentation," Working paper, California Institute of Technology.

Dennis, P. and S. Mayhew (2002), "Risk-Neutral Skewness: Evidence from Stock Options," *Journal of Financial and Quantitative Analysis* 37, 471-493.

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.

Harvey, C. and A. Siddique (2000), "Conditional Skewness in Asset Pricing Tests," *Journal of Finance* 55, 1263-1295.

Hu, W. and A. Kercheval (2010), "Portfolio Optimization for Student t and Skewed t Returns," *Quantitative Finance* 10, 91-105.

Imas, A. (2016), "The Realization Effect," *American Economic Review* 106, 2086-2109.

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

Kwak, M. and T. Pirvu (2018), "Cumulative Prospect Theory with Generalized Hyperbolic Skewed t Distribution," *SIAM Journal on Financial Mathematics* 9, 54-89.

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

McLean, D. and J. Pontiff (2016), "Does Academic Research Destroy Stock Return Predictability," *Journal of Finance* 71, 5-32.

Shleifer, A. and R. Vishny (1997), "The Limits of Arbitrage," *Journal of Finance* 52, 35-55.

Stambaugh, R., Yu, J., and Y. Yuan (2012), "The Short of It: Investor Sentiment and Anomalies," *Journal of Financial Economics* 104, 288-302.

Thaler, R. and E. Johnson (1990), "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice," *Management Science* 36, 643-660.

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 Repre-

sentation of Uncertainty," *Journal of Risk and Uncertainty* 5, 297-323.

Walasek, L., Mullett, T., and N. Stewart (2018), "A Meta-analysis of Loss Aversion in Risky Contexts," Working paper, University of Warwick.

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

Table 1. Stock market 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 equity	ROE
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. The first column lists 22 prominent anomalies. The remaining columns report, for each anomaly, the average annual return, standard deviation of annual returns, skewness of annual returns, and capital gain overhang of the typical stock in anomaly decile 1 and anomaly decile 10.

Anomaly	Average return Decile 1	Average return Decile 10	Standard deviation Decile 1	Standard deviation Decile 10	Skewness Decile 1	Skewness Decile 10	Gain overhang Decile 1	Gain overhang Decile 10
VOL	11.9	–3.2	36.8	94.3	2.46	3.79	10.5	–31.6
SIZE	14.0	10.6	76.0	24.6	4.27	0.69	–15.4	17.0
VAL	10.7	17.7	54.1	71.0	1.85	2.66	12.1	–24.1
EISKEW	12.4	8.0	33.3	90.3	1.33	3.54	13.7	–24.8
MOM	-2.4	20.6	91.4	63.1	3.84	2.46	–44.6	30.8
FPROB	17.5	0.1	51.5	110.4	2.31	3.9	28.2	–46.1
OSC	13.9	3.8	51.6	84.4	2.56	3.5	6.1	–20.6
NSI	16.0	6.8	54.2	70.8	2.71	3.2	5.7	–8.4
CEI	14.0	6.8	42.1	66.5	2.46	2.68	7.4	–11.2
ACC	15.4	7.4	85.0	72.0	3.2	3.0	–18.7	–5.4
NOA	14.8	6.9	66.4	65.3	3.13	2.95	–2.6	–4.9
PROF	8.8	14.4	73.0	62.6	3.49	2.68	–12.0	1.1
AG	14.8	7.0	82.5	70.7	3.1	3.07	–20.6	–5.6
ROE	4.1	13.1	85.3	55.3	3.22	2.22	–29.9	15.1
INV	15.9	8.0	81.8	72.2	3.6	3.31	–14.0	–5.4
MAX	10.9	1.2	46.3	90.4	2.76	3.61	3.2	–25.1
ORGCP	10.9	15.8	63.3	77.5	2.76	3.38	–3.8	–7.9
LTREV	16.7	11.0	88.9	47.7	3.23	1.77	–32.1	17.1
XFIN	13.0	4.2	58.7	83.4	3.17	3.44	–0.5	–15.6
STREV	14.3	7.7	83.1	73.3	3.63	3.03	–22.2	–1.9
DOO	15.5	10.0	40.2	57.9	1.36	1.74	16.6	–7.8
PEAD	9.25	16.8	58.2	58.3	2.49	2.3	–9.4	7.5

Table 3. The first column lists 22 anomalies. The remaining columns report, for each anomaly, the 30-day risk-neutral variance and risk-neutral skewness of the typical stock in anomaly decile 1 and anomaly decile 10. The risk-neutral measures are extracted from the prices of options on individual stocks over the 1996-2012 sample period.

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
ROE	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

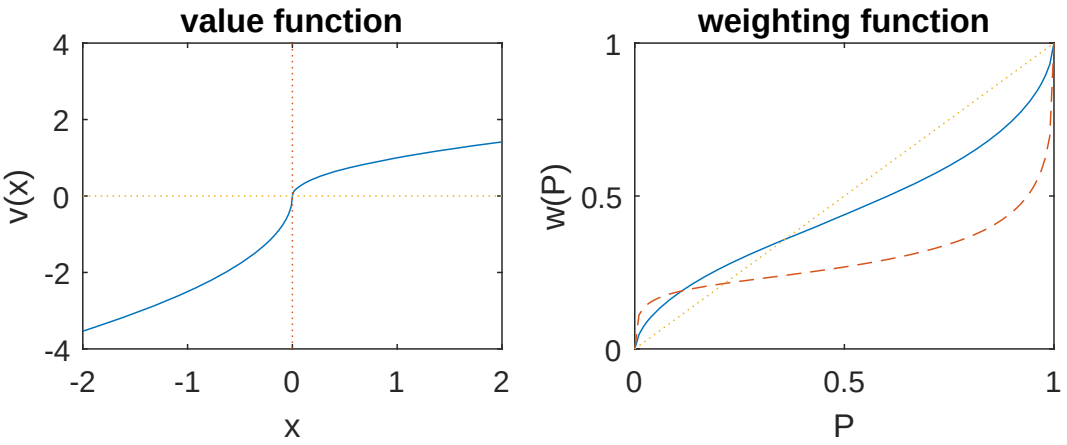


Figure 1. The left panel plots the value function proposed by Tversky and Kahneman (1992) as part of their cumulative prospect theory, namely, $v(x) = x^{\alpha}$ for $x \geq 0$ and $v(x) = -\lambda(-x)^{\alpha}$ for $x < 0$, for $\alpha = 0.5$ and $\lambda = 2.5$. The right panel plots the probability 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 $\delta = 1$.

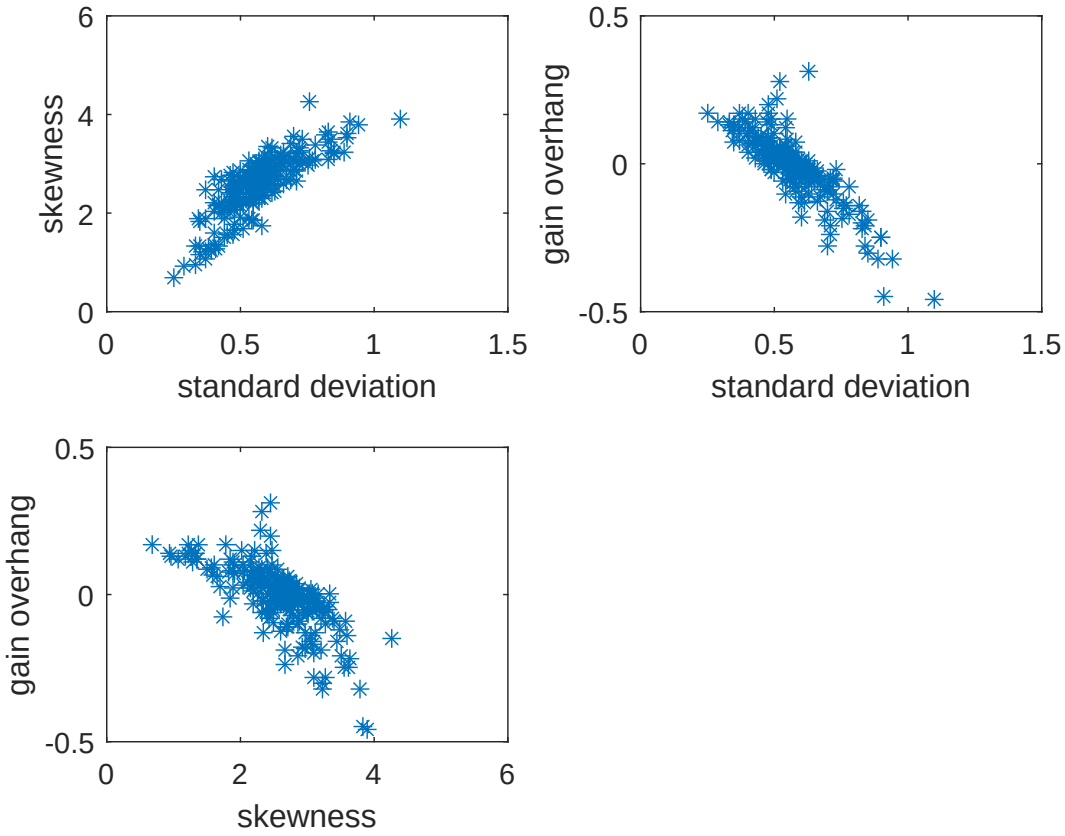


Figure 2. Each graph plots 220 asterisks, where each asterisk corresponds to one of 10 deciles for one of 22 anomalies. In the top-left graph, a given asterisk that corresponds to some anomaly decile marks the standard deviation of returns and the return skewness of the typical stock in that anomaly decile. In the top-right graph, each asterisk marks the standard deviation of returns and the gain overhang of the typical stock in some anomaly decile. In the bottom-left graph, each asterisk marks the return skewness and gain overhang of the typical stock in some anomaly decile.

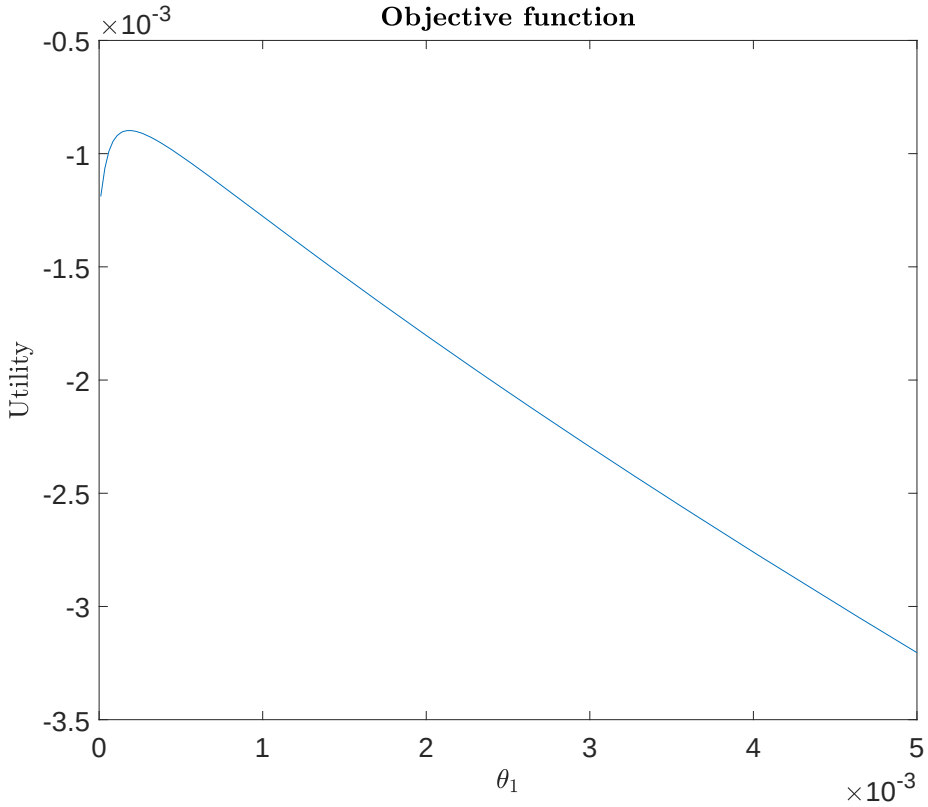


Figure 3. The graph shows that all investors have identical holdings of each stock in momentum decile 1. The solid line plots the value of an investor's objective function in equilibrium, as a function of θ_1, the (scaled) fraction of the investor's portfolio allocated to stock 1, which belongs to momentum decile 1. The function has a unique global maximum at the point where θ_1 equals the weight of stock 1 in the market portfolio, namely 0.000185.

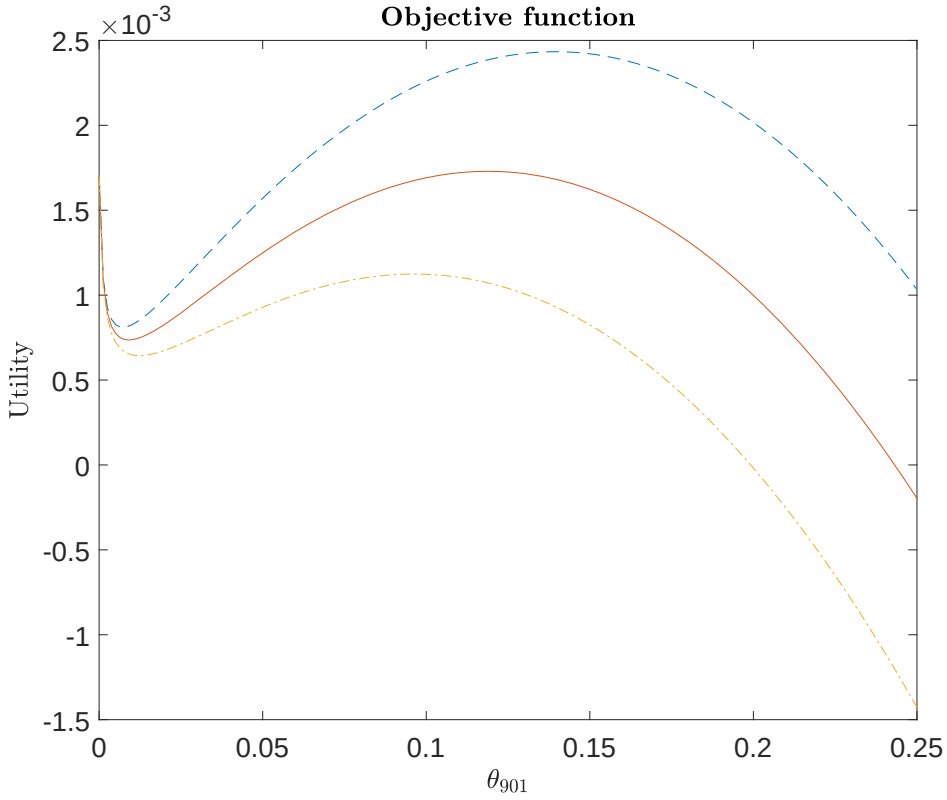


Figure 4. The graph shows that investors have heterogeneous holdings of each stock in momentum decile 10. The solid line plots the value of an investor's objective function in equilibrium, as a function of θ_{901}, the (scaled) fraction of the investor's portfolio allocated to stock 901, which belongs to momentum decile 10. The function has two global maxima which straddle the weight of stock 901 in the market portfolio, namely 0.00075. 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.

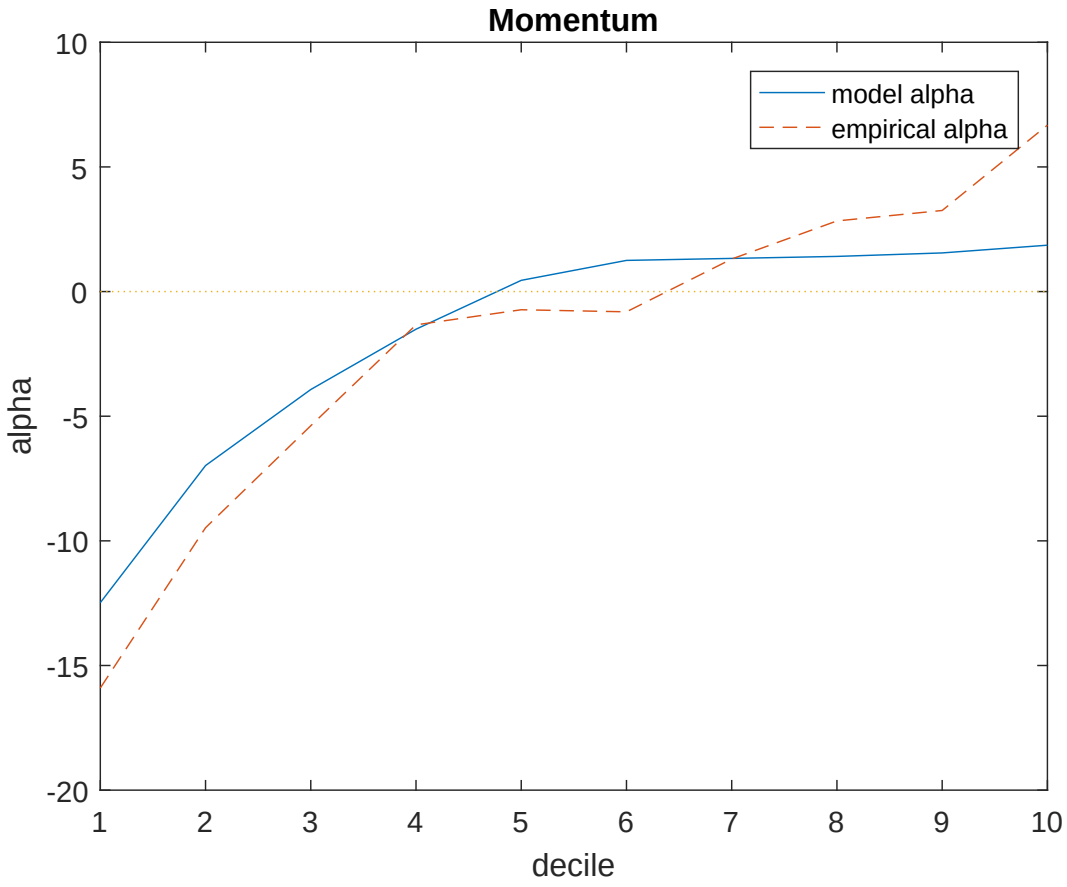


Figure 5. The dashed line plots the historical annual alpha of each momentum anomaly decile. The solid line plots the alphas predicted by a model where investors evaluate risk according to prospect theory.

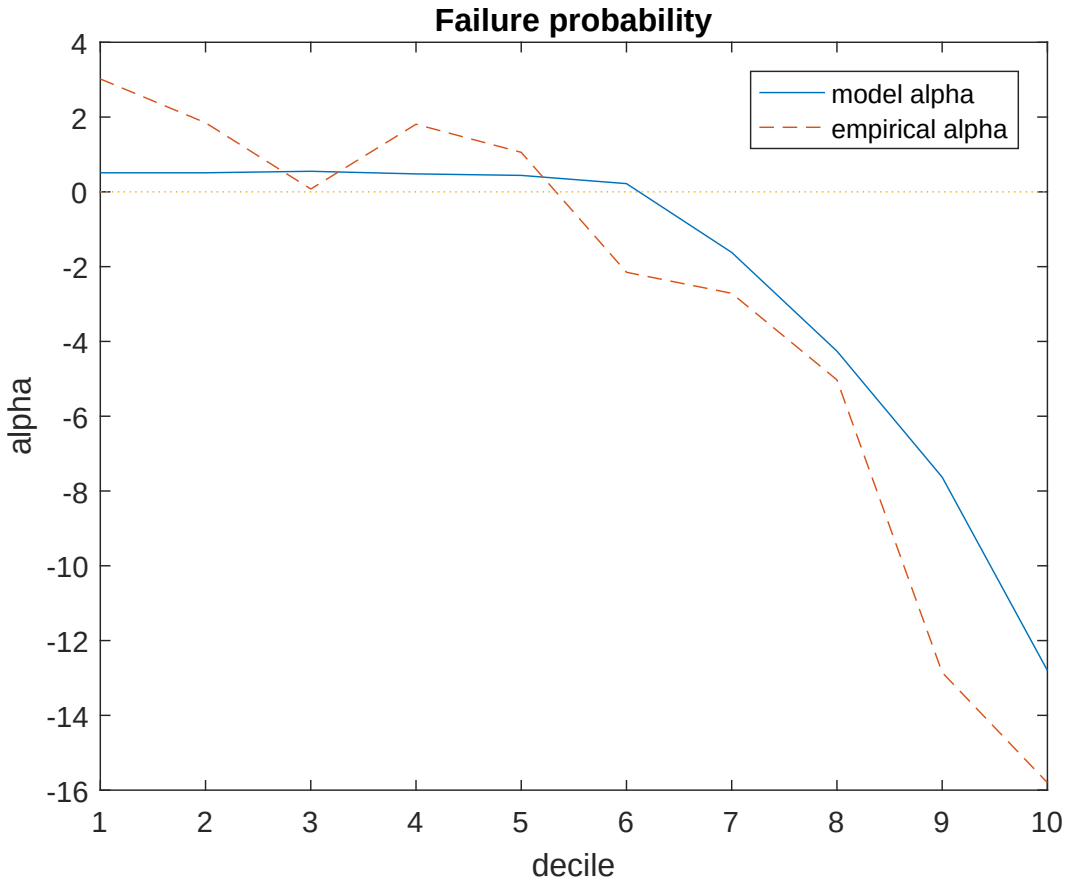


Figure 6. The dashed line plots the historical annual alpha of each failure probability anomaly decile. The solid line plots the alphas predicted by a model where investors evaluate risk according to prospect theory.

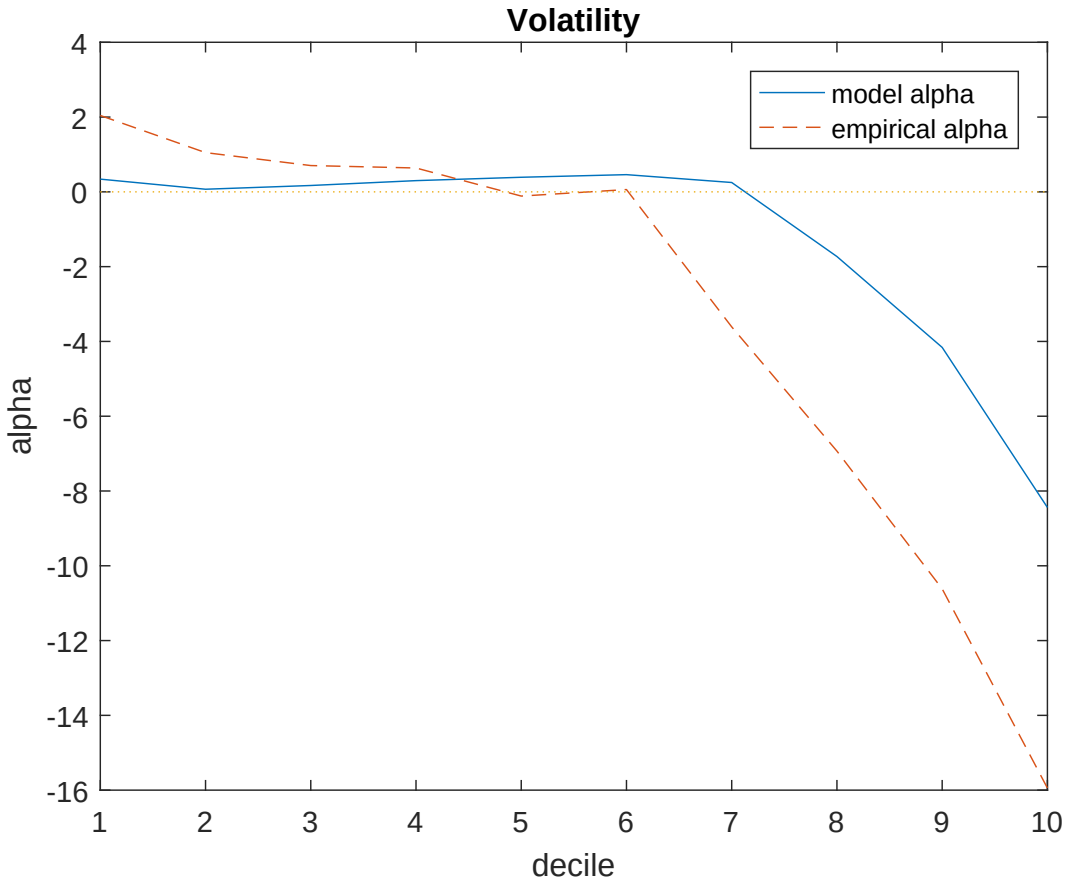


Figure 7. The dashed line plots the historical annual alpha of each idiosyncratic volatility anomaly decile. The solid line plots the alphas predicted by a model where investors evaluate risk according to prospect theory.

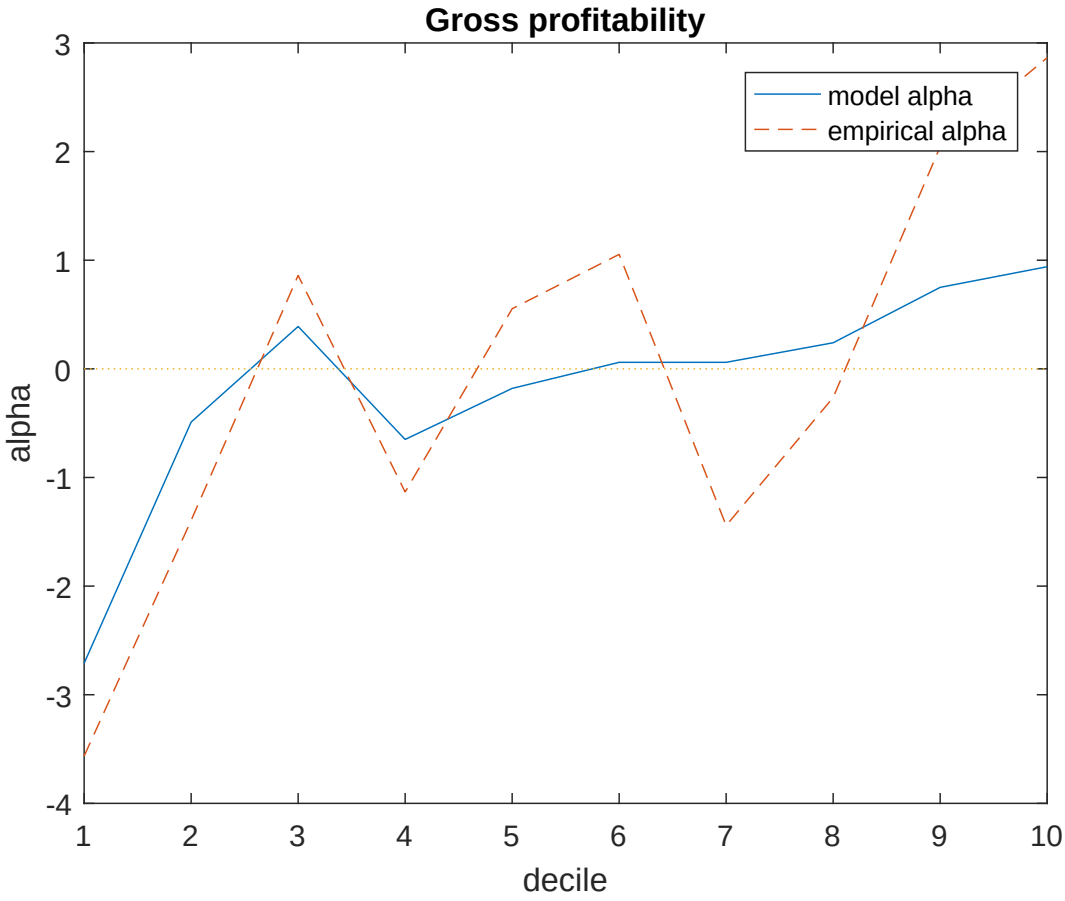


Figure 8. The dashed line plots the historical annual alpha of each gross profitability anomaly decile. The solid line plots the alphas predicted by a model where investors evaluate risk according to prospect theory.

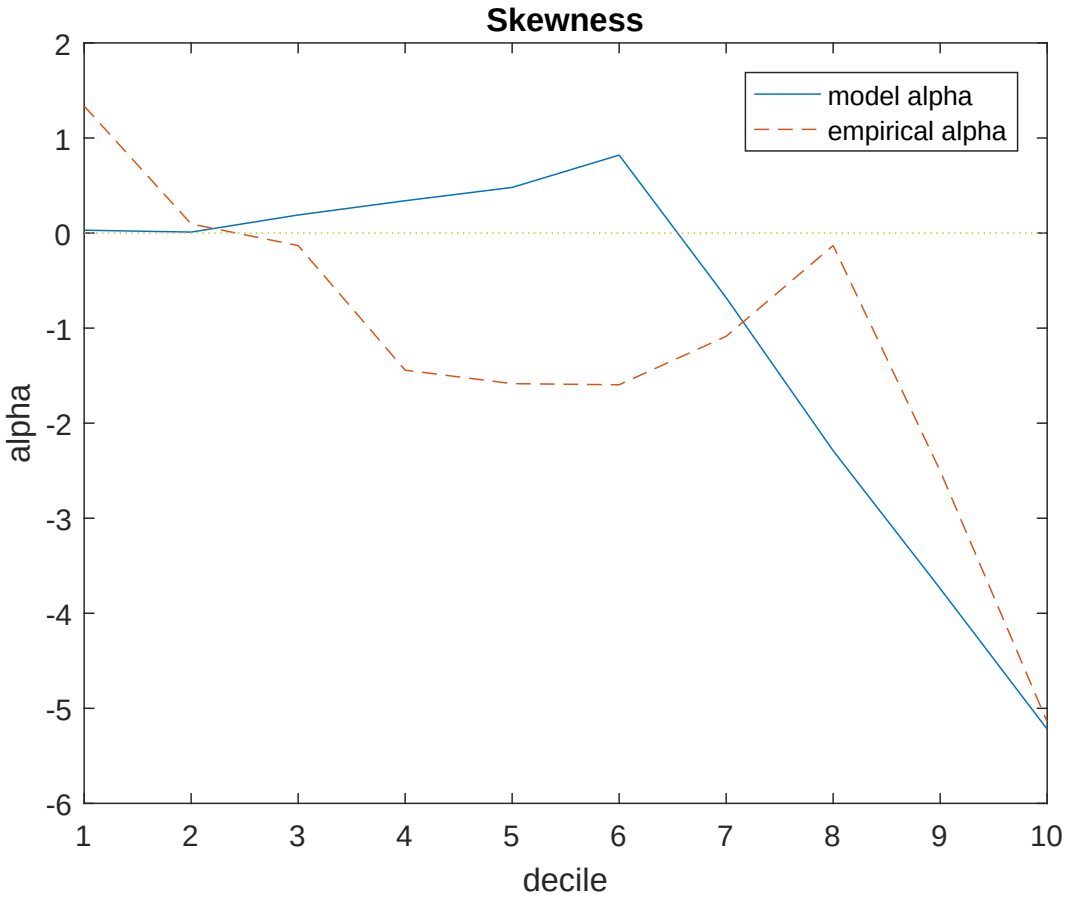


Figure 9. The dashed line plots the historical annual alpha of each idiosyncratic skewness anomaly decile. The solid line plots the alphas predicted by a model where investors evaluate risk according to prospect theory.

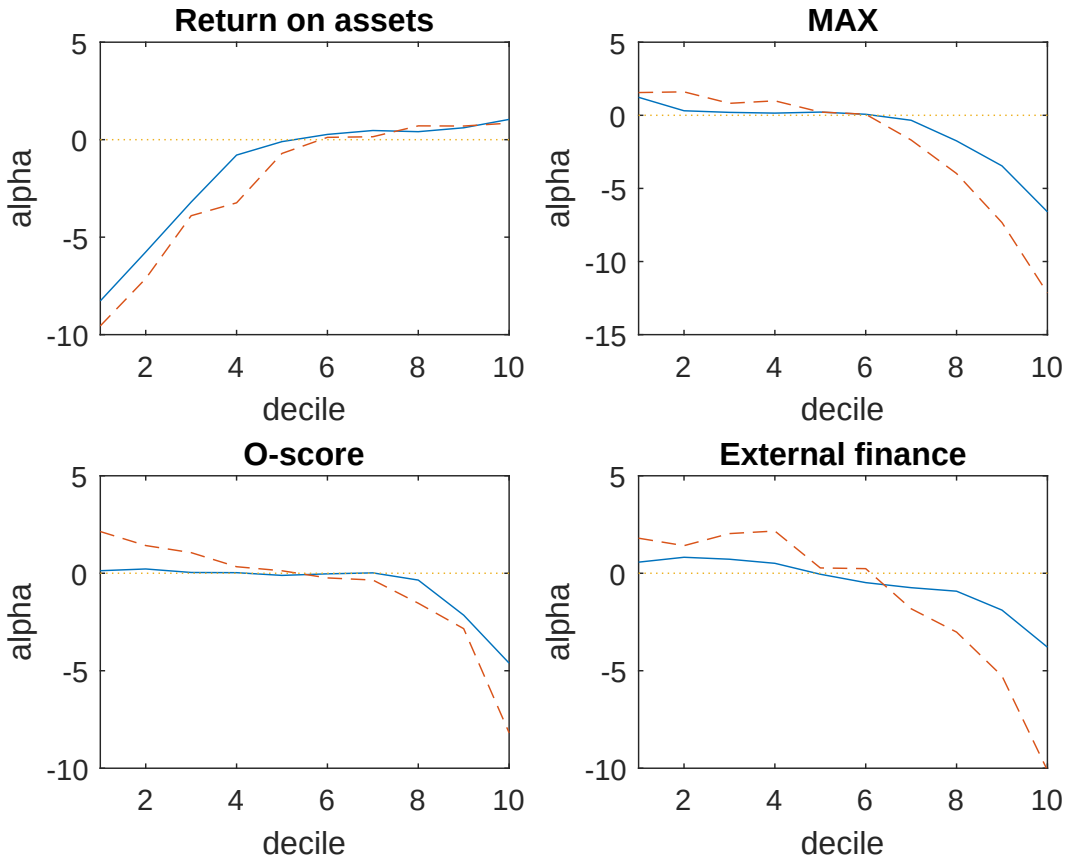


Figure 10. In each graph, the dashed lines plot the historical annual alpha of each anomaly decile for the return on equity, maximum daily return, O-score, and external finance anomalies. The solid lines plot the alphas predicted by a model where investors evaluate risk according to prospect theory.

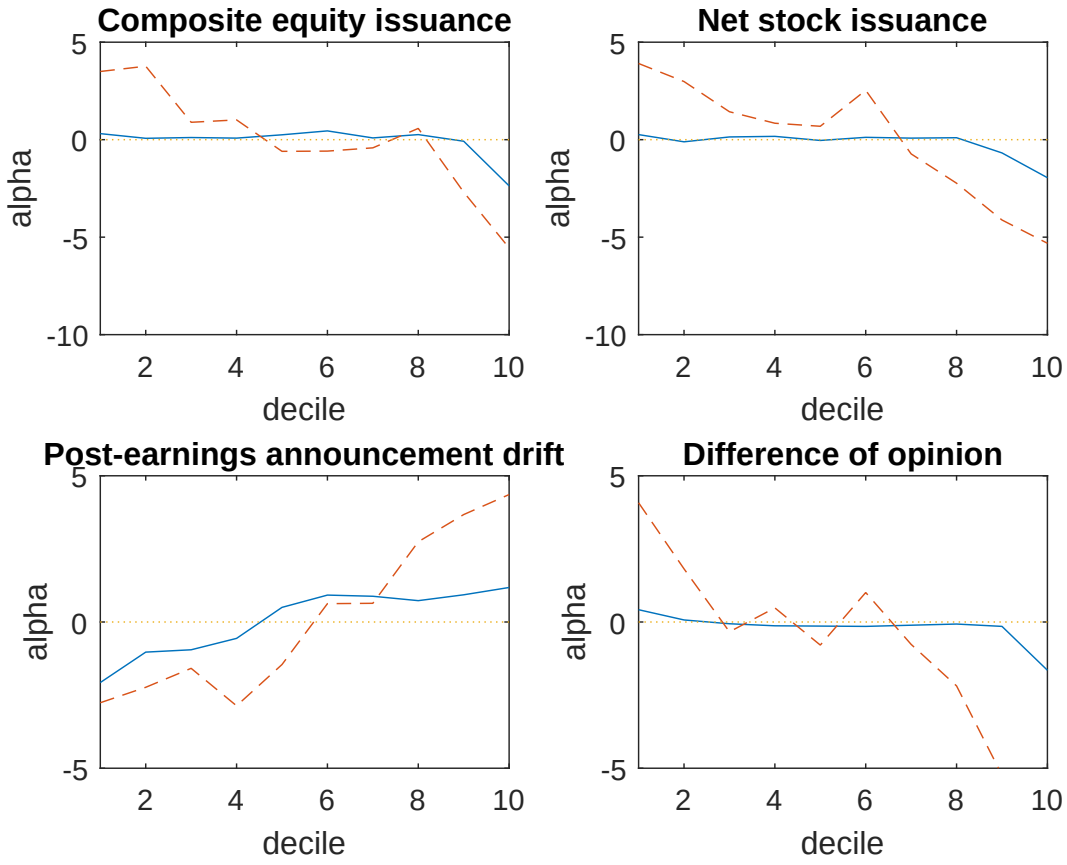


Figure 11. In each graph, the dashed lines plot the historical annual alpha of each anomaly decile for the composite equity issuance, net stock issuance, post-earnings announcement drift, and difference of opinion anomalies. The solid lines plot the alphas predicted by a model where investors evaluate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

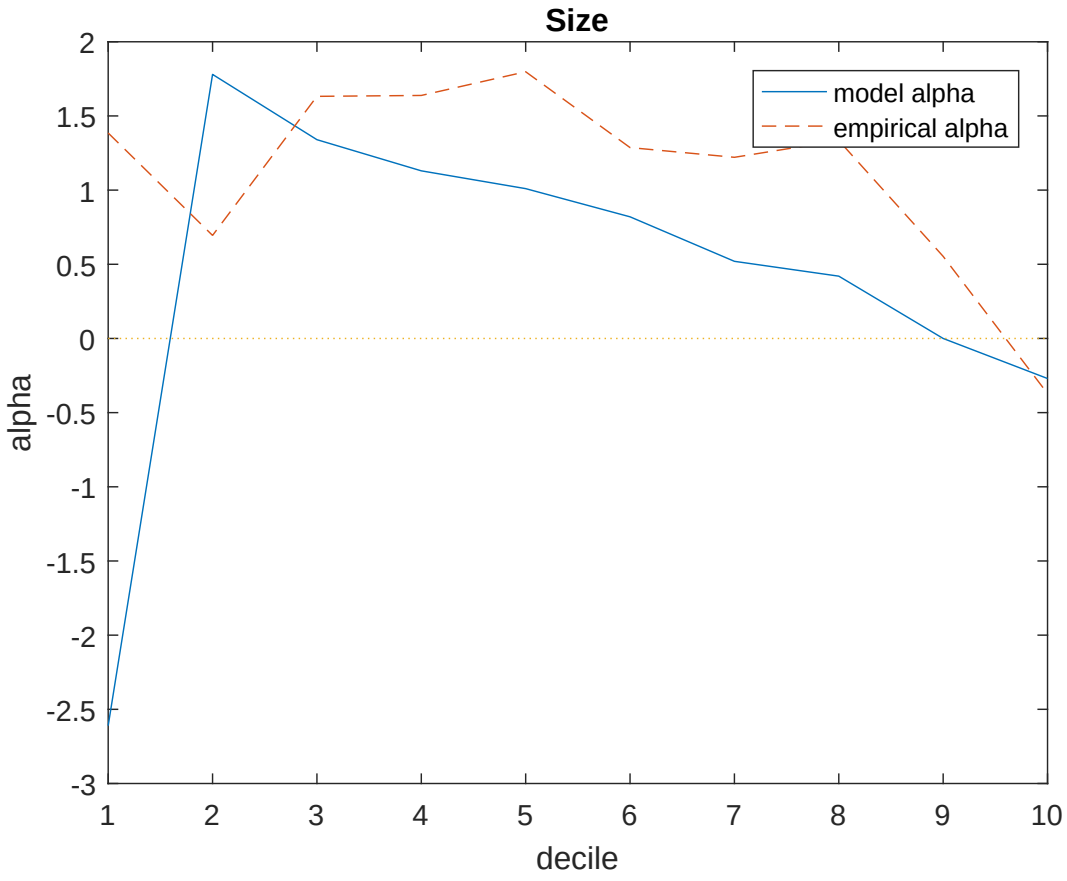


Figure 12. The dashed line plots the historical annual alpha of each size anomaly decile. The solid line plots the alphas predicted by a model where investors evaluate risk according to prospect theory.

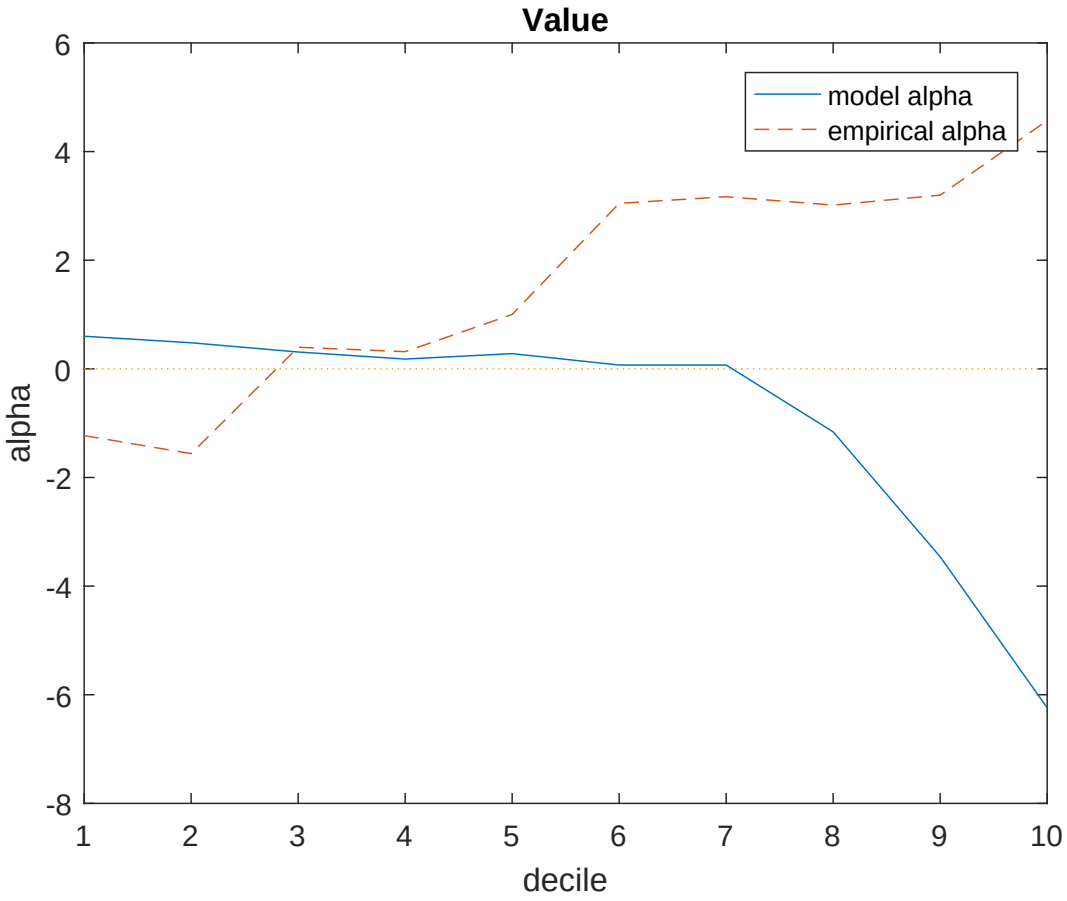


Figure 13. The dashed line plots the historical annual alpha of each value anomaly decile. The solid line plots the alphas predicted by a model where investors evaluate risk according to prospect theory.

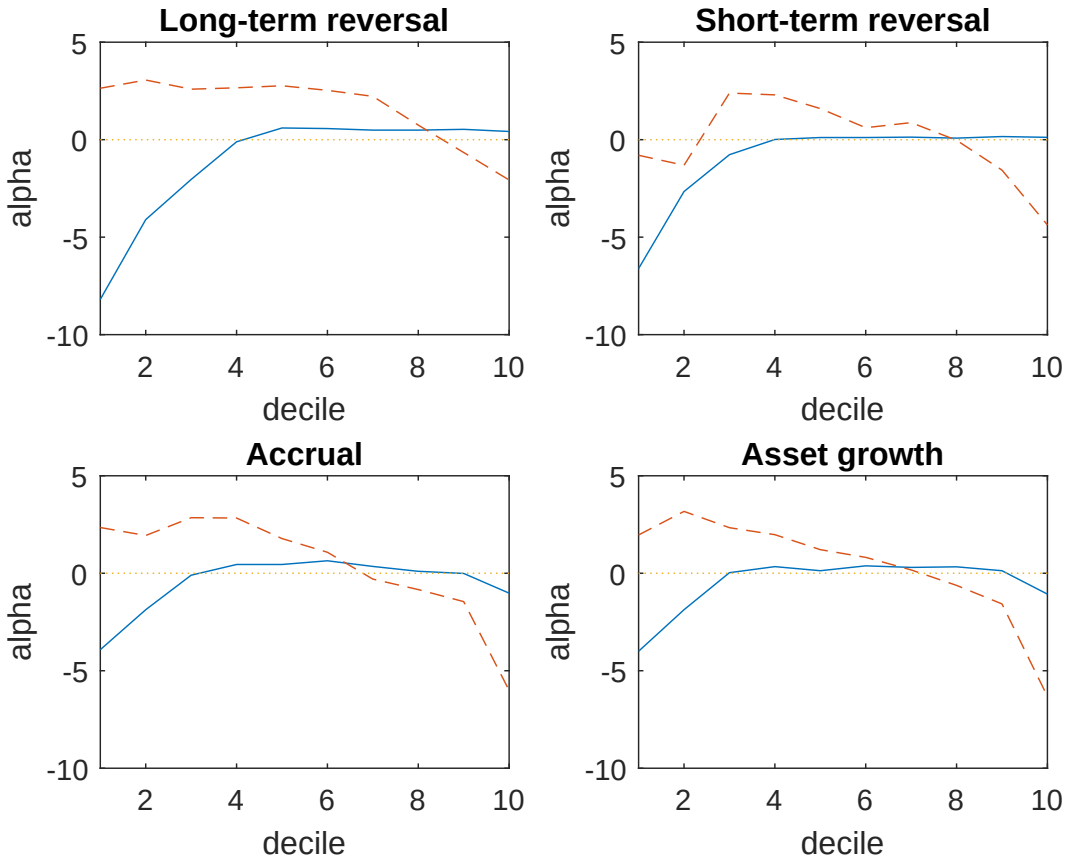


Figure 14. In each graph, the dashed line plots the historical annual alpha of each anomaly decile for the long-term reversal, short-term reversal, accrual, and asset growth anomalies. The solid lines plots the alphas predicted by a model where investors evaluate risk according to prospect theory.

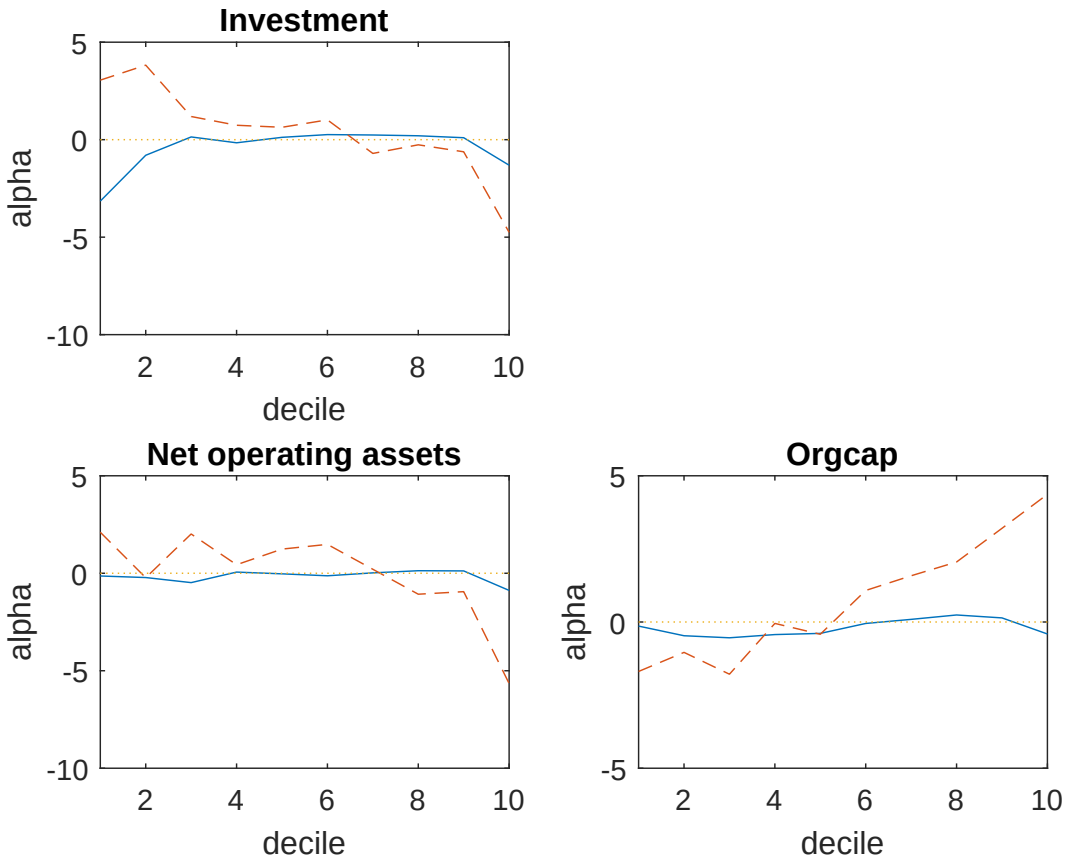


Figure 15. In each graph, the dashed line plots the historical annual alpha of each anomaly decile for the investment, net operating assets, and organizational capital anomalies. The solid lines plot the alphas predicted by a model where investors evaluate risk according to prospect theory.