

# Prospect Theory and Stock Market Anomalies

Nicholas Barberis, Lawrence Jin, and Baolian Wang

May 2020\*

## 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 volatility, skewness, and past capital gain. We find that the model is helpful for thinking about a majority of the 22 anomalies.

*JEL classification:* G11, G12

*Keywords:* prospect theory, loss aversion, probability weighting, cross-section of returns

---

\*The authors' affiliations are Yale University, Caltech, and the University of Florida, respectively; their e-mails are [nick.barberis@yale.edu](mailto:nick.barberis@yale.edu), [lawrence.jin@caltech.edu](mailto:lawrence.jin@caltech.edu), and [baolian.wang@warrington.ufl.edu](mailto:baolian.wang@warrington.ufl.edu). They are grateful to John Campbell, Jonathan Ingersoll, Eben Lazarus, Ian Martin, Dimitri Vayanos, Keith Vorkink, and seminar participants at Boston College, Caltech, the London Business School, the London School of Economics, Princeton University, the University of California at Berkeley, the University of Maryland, the University of Miami, the University of Notre Dame, Washington University, Yale University, the AFA, the Miami Behavioral Finance conference, and the NBER Behavioral Finance conference for very helpful comments.

# 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 captures 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 some 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 evaluate risk using a utility function that is defined over gains and losses; that has a kink at its origin, capturing a greater sensitivity to losses than to gains (“loss aversion”); and that is concave over gains and convex over losses, capturing risk aversion over moderate-probability gains and risk-seeking over moderate-probability losses (“diminishing sensitivity”). It also states that people 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.

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 prior gain or loss since purchase across investors holding 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.<sup>1</sup>

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 of asset prices fulfills both conditions; we therefore build a new one that does. In our model, investor preferences have two components. The first is traditional mean-variance preferences; taken alone, they lead 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. With this equilibrium structure in hand, we are able 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 empirical estimates of the volatility, skewness, and gain overhang of the typical stock in each anomaly decile are interesting in their own right. We find that the three characteristics are strongly correlated across anomaly deciles: if the typical stock in decile

---

<sup>1</sup>The three intuitions described here are outlined in Barberis and Huang (2001) and Li and Yang (2013); in Section III.G of Barberis and Huang (2008); and in Grinblatt and Han (2005) and Li and Yang (2013), respectively.

1 for some anomaly has more volatile returns than the typical stock in decile 10 for that anomaly, then it almost always also has more positively-skewed returns and a more negative gain overhang. For example, in the case of the size anomaly, the typical small-cap stock not only has more volatile returns than the typical large-cap stock, but also has more skewed returns and a more negative gain overhang.

This last observation points to the necessity of the quantitative approach we take in this paper. Consider 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 leads 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 gain overhang, than the typical large-cap stock. All else equal, these two factors 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 can help explain 13 of the 22 anomalies we consider, in the sense that it predicts a higher CAPM alpha for the extreme anomaly decile portfolio that actually has a higher alpha, empirically. These are the momentum, failure probability, idiosyncratic volatility, gross 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. The 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. To evaluate the model's performance more formally, we compare its average absolute pricing error across the 22 anomalies to that of several widely-used factor models. Our model achieves similar performance to the Carhart four-factor model – a striking result, given that the Carhart model was designed in full knowledge of several major anomalies, while ours was not.

For some anomalies, most notably the size and value anomalies, our model performs

poorly. For example, value stocks are more highly skewed and have a more negative gain overhang 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. The model therefore predicts a lower average return on value stocks, contrary to the empirical facts. One possible explanation for the model’s poor performance on some anomalies is that investors have incorrect beliefs about the return distributions of certain types of stocks – for example, they may under-estimate the skewness in value stock returns. We discuss evidence from option prices that offers some support for this hypothesis.

Our analysis builds on intuitions laid out in earlier papers. The idea that, due to loss aversion, more volatile assets should, all else equal, have a higher average return, is discussed by Barberis and Huang (2001) and Li and Yang (2013); the idea that, due to diminishing sensitivity, an asset’s gain overhang should be positively related to its average return, is developed by Grinblatt and Han (2005) and Li and Yang (2013); and the idea that, due to probability weighting, an asset’s return skewness should be negatively related to its average return, is studied by Barberis and Huang (2008) and Baele et al. (2019), among others.

Despite these advances, the basic questions listed in our opening paragraph remain unanswered. To answer these questions – to understand prospect theory’s predictions for the cross-section of returns – we need to combine the three above intuitions. And to do that, 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., 2019). Meanwhile, the prior models of the cross-section that do incorporate all the elements of prospect theory, those of Barberis and Huang (2008), Ingersoll (2014), and Barberis, Mukherjee, and Wang (2016), are all 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. The model is not an “easy extension” of prior models, but rather requires an entirely new equilibrium structure and solution method.

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 allows us to answer the questions laid out at the start of the paper in a way that was not previously possible. Second, 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 a very 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 the model. In Section 5, we present the model’s predictions about stock market anomalies. Section 6 discusses some other aspects of our analysis, while Section 7 concludes. Additional technical details can be found in Section 8 and in the Internet Appendix.

## 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 prospect theory and narrow framing in turn. Readers already familiar with these concepts may prefer to go directly to Section 3.

### 2.1 Prospect theory

The original version of prospect theory is described in Kahneman and Tversky (1979). This version of the theory has some limitations: it can be applied 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 overcomes these limitations. This is the version we adopt in this paper.<sup>2</sup>

To see how cumulative prospect theory works, consider the gamble

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

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

---

<sup>2</sup>While our analysis is based on cumulative prospect theory, we often abbreviate this as “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

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

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

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

where

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

and where  $v(\cdot)$  and  $w(\cdot)$  are known as the value function and probability weighting function, respectively.<sup>3</sup> Tversky and Kahneman (1992) propose the functional forms

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

and

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

where  $\alpha, \delta \in (0, 1)$  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 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

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

The severity of the kink is determined by the parameter  $\lambda$ ; a higher value of  $\lambda$  implies a greater sensitivity to losses.

---

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

Third, while  $U(\cdot)$  in (2) is typically concave everywhere,  $v(\cdot)$  in (3) is concave only over gains; over losses, it is convex. This pattern, which can be seen in Figure 1, captures the experimental finding that people tend to be risk averse over moderate-probability gains – they prefer a 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.<sup>4</sup> The degree of concavity over gains and convexity over losses are governed by the parameter  $\alpha$ ; a lower value of  $\alpha$  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  $w(\cdot)$ . 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 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  $\delta$ ; a lower value of  $\delta$  implies more overweighting of tails.<sup>5</sup>

## 2.2 Narrow framing

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

---

<sup>4</sup>We abbreviate  $(x, p; 0, q)$  as  $(x, p)$ .

<sup>5</sup>Prospect theory has impressive predictive power in experimental settings: Fudenberg et al. (2019) find that, at least for simple gambles, it attains almost the maximum possible level of predictive ability. One aspect of prospect theory – the “rank-dependent” formulation of probability weighting in (4) – has not found empirical support (Bernheim and Sprenger, 2019). Fortunately, our results do not rely on rank dependence: we have repeated our analysis with *non*-rank-dependent probability weighting and obtain very similar findings.

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 gamble in (7) is evidence not only of loss aversion, but of narrow framing as well.

Prospect theory and narrow framing are both widely seen as describing people’s intuitive mode of thinking. As such, models with prospect theory investors often also incorporate narrow framing. We follow this practice here.

### 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.

Prospect theory investors who engage in narrow framing evaluate a risky asset by thinking about the potential gains and losses in their holdings of the asset, and then computing the prospect theory value of this distribution of gains and losses. Since they are loss averse, they dislike assets with volatile returns; all else equal, they require a higher average return on these assets. Moreover, since, according to probability weighting, they overweight the tails of the distribution they are thinking about, they like assets with positively-skewed returns; all else equal, they require a lower average return on such assets. Finally, if an asset 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 value function in Figure 1. Since he is risk averse at this point, he demands a high average return to hold the asset. If, on the other hand, the typical investor has a paper loss in the asset, 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 asset.

The above intuitions make it clear 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 gain or loss in each risky asset. As noted above, no existing

model of asset prices fulfills both conditions. We now present a new model that does. Solving a model of this kind presents significant challenges; to keep the model tractable, we necessarily make some simplifying assumptions. Nonetheless, the model captures the three essential intuitions described above in a robust way.

### 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  $\tilde{R}_i$ , and the return vector  $\tilde{R} = (\tilde{R}_1, \dots, \tilde{R}_N)$  has a cumulative distribution function  $P(\tilde{R})$  that we specify below. The vector of expected returns on the risky assets is  $\bar{R} = (\bar{R}_1, \dots, \bar{R}_N)$  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  $\Theta_i$ , so that wealth at time  $1$  is

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

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

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

where

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

The first two terms in (9) are the traditional mean-variance preferences;  $\gamma$  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,  $V(\tilde{G}_i)$ , corresponds to asset  $i$ . Specifically,  $\tilde{G}_i$  is the potential gain or loss on asset  $i$ , and  $V(\tilde{G}_i)$  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  $i$ ,  $\tilde{G}_i$ , is defined in (10). It is the sum of two terms. The first

term,  $W_0\Theta_i(\tilde{R}_i - R_f)$ , 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 view the risk-free rate as a psychologically plausible benchmark: 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. Our framework can also accommodate other choices of benchmark.

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  $g_i = -0.3$ .<sup>6</sup>

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 mounting 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).

To keep the model tractable, we take the second term on the right-hand side of (10) to be 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 gain or loss, since purchase, 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  $\Theta_i$ , the investor's allocation to asset  $i$  at time 0, which appears in the first term in (10). In the Internet Appendix, we study the impact of allowing for heterogeneity across investors in either their gain or loss  $g_i$  or their initial allocation  $\Theta_{i,-1}$ ; we find that this impact is quantitatively small. Finally, we use the approximation  $W_{-1} \approx W_0$ .<sup>7</sup>

---

<sup>6</sup>There are small differences in how 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.

<sup>7</sup>More accurate approximations, such as  $W_{-1} \approx W_0/1.04$ , where 4% is a measure of the historical average

By defining utility over asset-level gains and losses, we are incorporating narrow framing. This framing assumption makes our model more tractable. In addition, as noted above, there are psychological reasons to couple prospect theory with narrow framing. Consistent with this, we find that the narrow framing assumption improves the model's predictive power. In the Internet Appendix, we describe an analysis we conducted of a model with broad framing, one where investors derive prospect theory utility from portfolio-level gains and losses. We find that the model with narrow framing is more successful at explaining the observed facts.

The quantity  $V(\tilde{G}_i)$  is the cumulative prospect theory value of the gain or loss  $\tilde{G}_i$ . For  $\Theta_i > 0$ , we can write  $V(\tilde{G}_i)$  as

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

where  $P(R_i)$  is the marginal CDF of asset  $i$ 's returns and where  $dw(P(R_i))$  and  $dw(1 - P(R_i))$  are given in full in Section 8.1. The expression in (11) uses a standard implementation of cumulative prospect theory for gambles with continuous distributions. The top row corresponds to losses, and is therefore multiplied by loss aversion  $\lambda$ . The bottom row corresponds to gains.

To complete the description of the decision problem, we need to specify the probability distribution  $P(\tilde{R})$  for asset returns. Since skewness plays an important role in our analysis, we need a distribution that can capture it as accurately as possible. One distribution that is increasingly seen as a superior way of modeling skewness and fat tails in asset returns is the “generalized hyperbolic (GH) skewed  $t$ ” distribution, and we adopt it here. The vector of asset returns  $\tilde{R} = (\tilde{R}_1, \dots, \tilde{R}_N)$  has an  $N$ -dimensional GH skewed  $t$  distribution; we present the density function for this distribution in Internet Appendix A. For our computations, we need only the marginal distribution of an asset's return. This is a one-dimensional GH skewed  $t$  distribution; for asset  $i$ , its density function is<sup>8</sup>

$$p(R_i) = \frac{2^{1-\frac{\nu+1}{2}}}{\Gamma(\frac{\nu}{2})(\pi\nu S_i)^{\frac{1}{2}}} \cdot \frac{K_{\frac{\nu+1}{2}}(\sqrt{(\nu + (R_i - \mu_i)^2/S_i)\zeta_i^2/S_i}) \exp((R_i - \mu_i)\zeta_i/S_i)}{\left(\sqrt{(\nu + (R_i - \mu_i)^2/S_i)\zeta_i^2/S_i}\right)^{-\frac{\nu+1}{2}} (1 + (R_i - \mu_i)^2\nu^{-1}/S_i)^{\frac{\nu+1}{2}}},$$

for  $\zeta \neq 0$

---

return on investor wealth, have a very minor impact on our quantitative predictions. We therefore stick with the simpler approximation  $W_{-1} \approx W_0$ .

<sup>8</sup>We use  $p(\cdot)$  and  $P(\cdot)$  to denote the probability density function and the cumulative distribution function, respectively.

$$p(R_i) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})(\pi\nu S_i)^{\frac{1}{2}}} \cdot (1 + (R_i - \mu_i)^2 \nu^{-1} / S_i)^{-\frac{\nu+1}{2}}, \text{ for } \zeta = 0, \quad (12)$$

where  $\Gamma(\cdot)$  is the Gamma function and  $K_l$  is the modified Bessel function of the second kind with order  $l$ .<sup>9</sup>

The above distribution has four parameters:  $\mu_i$ ,  $S_i$ ,  $\zeta_i$ , and  $\nu$ . Here,  $\mu_i$ , the location parameter, helps to determine the mean of the distribution;  $S_i$ , the dispersion parameter, controls the dispersion in returns;  $\zeta_i$ , the asymmetry parameter, governs the skewness of returns; and  $\nu$ , a degree of freedom scalar, affects the heaviness of the tails of the distribution. The mean, variance, and skewness of the distribution are

$$E(\tilde{R}_i) = \bar{R}_i = \mu_i + \frac{\nu}{\nu - 2} \zeta_i \quad (13)$$

$$\text{Var}(\tilde{R}_i) = \frac{\nu}{\nu - 2} S_i + \frac{2\nu^2}{(\nu - 2)^2(\nu - 4)} \zeta_i^2 \quad (14)$$

$$\text{Skew}(\tilde{R}_i) = \frac{2\zeta_i \sqrt{\nu(\nu - 4)}}{\sqrt{S_i} (2\nu\zeta_i^2 / S_i + (\nu - 2)(\nu - 4))^{\frac{3}{2}}} \left[ 3(\nu - 2) + \frac{8\nu\zeta_i^2}{S_i(\nu - 6)} \right]. \quad (15)$$

In Sections 4 and 5, using equations (14) and (15), we will set  $S_i$ ,  $\zeta_i$ , and  $\nu$  to match the empirical volatility and skewness of asset  $i$ 's returns. We will then search for a value of  $\mu_i$  so that the market for asset  $i$  clears. The asset's expected return – the quantity we want to determine – is then given by (13).<sup>10</sup>

The objective function in (9) combines a traditional component, namely mean-variance preferences, with a non-traditional one that incorporates prospect theory. As such, it is consistent with the approach advocated by Koszegi and Rabin (2006) among others, namely that models of gain-loss utility should retain a traditional utility term. One interpretation of this two-part utility function draws on the influential “two-system” framework in psychology (Kahneman, 2013). Under this view, the prospect theory term in (9) and the narrow framing it incorporates capture a person's “fast,” intuitive sense of how to invest, while the

---

<sup>9</sup>See Aas and Haff (2006) and Hu and Kercheval (2010) for more discussion of the GH skewed  $t$  distribution. This distribution has one “heavy” tail and one “semi-heavy” tail, making it particularly useful for capturing skewness and fat tails in asset returns. Simpler distributions such as the log-normal and skew-normal are not suitable for our purposes. The log-normal distribution has two parameters; setting these to match an asset's volatility and skewness also fixes the asset's mean, preventing it from being determined in equilibrium by market clearing. The skew-normal distribution cannot accommodate skewness levels higher than 0.995; this makes it a poor fit for our application.

<sup>10</sup>This procedure is analogous to that for the CAPM. In the case of the CAPM, we estimate a second moment – an asset's beta – from the data and use it to determine the asset's expected return. Here, we will estimate second and third moments – the asset's beta, volatility, and skewness – and use them all to predict the expected return.

mean-variance term corresponds to an investment strategy based on “slow,” effortful reasoning. The individual’s final course of action is a combination of these intuitive and reasoned judgments: he anchors on his initial intuitive impression, which then leaves an imprint on his final decision.

Why do we take the traditional term in (9) 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.<sup>11</sup>

### 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 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 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  $\mu = (\mu_1, \dots, \mu_N)$  such that, for this  $\mu$ , the objective function in (9) has a unique global maximum  $\Theta^* = (\Theta_1^*, \dots, \Theta_N^*)$  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 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 the Internet Appendix, we

---

<sup>11</sup>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; see Harvey and Siddique (2000). Here, we examine whether prospect theory can do so.

illustrate this non-existence with an example. Here, we explain it in general terms. Suppose that, for some value of  $\mu_1$ , the location parameter for asset 1, the objective function in (9) is maximized at  $\Theta_1 = \Theta_1^*$ , where  $\Theta_1^*$  exceeds the asset's market supply  $\Theta_{M,1}$ . This suggests that, to clear the market, we simply need to lower the value of  $\mu_1$ , as this will lower the asset's expected return. However, it turns out that, as we do so, the value of  $\Theta_1$  for which the objective function attains its maximum jumps *discontinuously* from a  $\Theta_1^*$  that exceeds  $\Theta_{M,1}$  to a  $\Theta_1^*$  that lies below it. As such, there is *no* value of  $\mu_1$  for which the objective function is maximized at a  $\Theta_1^*$  that equals  $\Theta_{M,1}$ . An equilibrium where investors have identical holdings for all assets therefore does not exist.<sup>12</sup>

**Full rationality with heterogeneous holdings.** Given that the homogeneous-holdings equilibrium does not exist, what kind of equilibrium can we consider instead? One alternative structure involves multiple global maxima. In other words, it may be that there exists a location vector  $\mu = (\mu_1, \dots, \mu_N)$  such that, for this  $\mu$ , 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 determine if it exists. To see why, suppose that we consider 100 candidate values for each element of the location vector  $\mu$ ; this implies  $100^N$  possible location vectors  $\mu$ . 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  $\Theta_i$  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

---

<sup>12</sup>Why does the value of  $\Theta_1^*$  at which (9) is maximized jump discontinuously as we lower  $\mu_1$ ? When  $b_0 = 0$ , the expression in (9), viewed as a function of  $\Theta_1$ , depends only on  $\Theta_1$  and  $\Theta_1^2$ . It therefore has a single local maximum that is also its global maximum. When  $b_0 > 0$ , (9) becomes a function of  $\Theta_1$ ,  $\Theta_1^2$ , and additional powers of  $\Theta_1$ , including  $\Theta_1^\alpha$ , where  $\alpha \in (0, 1)$ . As such, it can have multiple local maxima. As we lower  $\mu_1$ , there comes a point where the global maximum jumps from one local maximum to another; this makes  $\Theta_1^*$  a discontinuous function of  $\mu_1$ .

explain below, this discrepancy is likely to have a negligible impact on our results. We think of this bounded-rationality assumption not just as a computational technique for solving the problem in (9), but also as a psychological assumption as to how an individual might in reality go about solving this problem.

We then define a bounded-rationality equilibrium with heterogeneous holdings as consisting of a location vector  $\mu$  such that, for this  $\mu$ , 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 risky asset  $i$  in turn, we take the objective function in (9), view it as a function of  $\Theta_i$ , 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:

$$\begin{aligned} & \Theta_i(\mu_i + \frac{\nu\zeta_i}{\nu-2} - R_f) - \frac{\hat{\gamma}}{2}(\Theta_i^2\sigma_i^2 + 2\Theta_i \sum_{j \neq i} \Theta_{M,j}\sigma_{ij}) \\ & - \lambda \hat{b}_0 \int_{-\infty}^{R_f - \Theta_{i,-1}g_i/\Theta_i} (\Theta_i(R_f - R_i) - \Theta_{i,-1}g_i)^\alpha dw(P(R_i)) \\ & - \hat{b}_0 \int_{R_f - \Theta_{i,-1}g_i/\Theta_i}^{\infty} (\Theta_i(R_i - R_f) + \Theta_{i,-1}g_i)^\alpha dw(1 - P(R_i)), \end{aligned} \quad (16)$$

where

$$\hat{\gamma} = \gamma W_0 \quad (17)$$

$$\hat{b}_0 = b_0 W_0^{\alpha-1}. \quad (18)$$

A bounded-rationality equilibrium with heterogeneous holdings consists of a location vector  $(\mu_1, \dots, \mu_N)$  such that, for each  $i$ , the function in (16) 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 bounded-rationality assumption greatly simplifies the investors' optimization problem: by turning the multivariate function in (9) into the univariate function in (16), it converts the search for the optimal allocation  $\Theta_i$  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 opposite strategy. Moreover, because the problem is now one-dimensional, it is easy to determine whether the function in (16) 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. When we implement the equilibrium, we find that, for many risky assets, investors have identical holdings. Second, we find that, for any asset  $i$  where investors have heterogeneous holdings, the function in (16) has just two global maxima,  $\Theta_i^*$  and  $\Theta_i^{**}$ , both of which are positive. 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  $\Theta_i^*$  allocation and the rest to the  $\Theta_i^{**}$  allocation. We also find that  $\Theta_i^*$  is always much closer to  $\Theta_{M,i}$  than is  $\Theta_i^{**}$ . As such, to clear the market, we assign the vast majority of investors to the  $\Theta_i^*$  allocation and the remaining few to the  $\Theta_i^{**}$  allocation.

In Section 4, after parameterizing the model, we present a numerical example that illustrates investors' heterogeneous holdings. For now, we emphasize that there is a simple and realistic intuition for the two optima,  $\Theta^*$  and  $\Theta^{**}$ . The heterogeneous holdings typically arise for stocks with a positive gain overhang which, in part because of this, have a high expected return. The lower optimum  $\Theta^*$  reflects investors' desire to lock in their prior gains in these stocks by selling some of their holdings; the upper optimum  $\Theta^{**}$  reflects investors' desire to benefit from the high expected return by increasing their allocation to these stocks. In equilibrium, the two strategies are equally attractive.

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 (16) has a unique global maximum – while for the remaining 500, the function in (16) 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 lower than its market weight – this is the  $\Theta_i^*$  optimum – but for a small handful of these assets, he holds a large position, given by the  $\Theta_i^{**}$  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 is likely to have a negligible impact on the model's predictions: we find 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 (16)-(18).<sup>13</sup>

To make the model easier to implement, we rescale it. Specifically, let  $\Theta_{M,R} = \sum_{i=1}^N \Theta_{M,i}$  be the market value of all risky assets relative to the market value of all assets, and define

$$\begin{aligned}\theta_i &= \Theta_i / \Theta_{M,R} \\ \theta_{M,i} &= \Theta_{M,i} / \Theta_{M,R} \\ \theta_{i,-1} &= \Theta_{i,-1} / \Theta_{M,R}.\end{aligned}\tag{19}$$

From now on, we think of investors as choosing  $\theta_i$  rather than  $\Theta_i$ . In Section 8.2, we show that, when reformulated with  $\theta_i$  as the choice variable, the investor's decision problem has exactly the same form as in (16), subject only to a rescaling of  $\hat{\gamma}$  and  $\hat{b}_0$ . 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 Section 8.2 for details – the rescaling allows us to simplify the variance term in the first row of (16) by introducing asset  $i$ 's beta, denoted  $\beta_i$ .<sup>14</sup>

For completeness, we restate the definition of equilibrium in terms of  $\theta_i$ . The equations below are simply a rescaled version of (16)-(18). A bounded-rationality equilibrium consists of a location vector  $\mu = (\mu_1, \dots, \mu_N)$  such that, for each  $i$ , the function

$$\begin{aligned}\theta_i(\mu_i + \frac{\nu\zeta_i}{\nu-2} - R_f) - \frac{\hat{\gamma}}{2}(\theta_i^2\sigma_i^2 + 2\theta_i(\beta_i\sigma_M^2 - \theta_{M,i}\sigma_i^2)) \\ - \lambda\hat{b}_0 \int_{-\infty}^{R_f - \theta_{i,-1}g_i/\theta_i} (\theta_i(R_f - R_i) - \theta_{i,-1}g_i)^\alpha dw(P(R_i)) \\ - \hat{b}_0 \int_{R_f - \theta_{i,-1}g_i/\theta_i}^{\infty} (\theta_i(R_i - R_f) + \theta_{i,-1}g_i)^\alpha dw(1 - P(R_i)),\end{aligned}\tag{20}$$

---

<sup>13</sup>We have also studied the following iterative procedure. When solving for his allocation  $\Theta_i$  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  $\Theta_i$ , 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 (16)-(18).

<sup>14</sup>Since  $\Theta_{M,R} = \sum_{i=1}^N \Theta_{M,i}$ , it follows that  $\sum_{i=1}^N \theta_{M,i} = 1$ . However, due to investors' heterogeneous holdings,  $\sum_{i=1}^N \theta_i$  will almost never equal 1 for any given investor.

where

$$\hat{\gamma} = \gamma W_0 \Theta_{M,R} \quad (21)$$

$$\hat{b}_0 = b_0 W_0^{\alpha-1} \Theta_{M,R}^{\alpha-1}, \quad (22)$$

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

In Section 8.3, 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. Since it involves numerical integration, this calculation takes a few minutes of computing time; this is fast enough for the application we consider in this paper.

## 4 Anomalies and Model Parameter Values

The model of Section 3 generates quantitative predictions about the cross-section of average returns when investors evaluate risk according to prospect theory. We now use the model 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 anomalies. They are listed in Table 1, along with the abbreviations we use in subsequent tables to refer to them; Internet Appendix B 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 well-known 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. We number the stocks so that, in the case of the value anomaly, say, stocks 1 to 100 belong to decile 1, which contains stocks with low book-to-market ratios; stocks 101 to 200 belong to 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 value anomaly if the expected return it predicts for the randomly-chosen stock in decile 10 is significantly higher than the expected return it predicts for the randomly-chosen stock in decile 1.<sup>15</sup>

What are the empirical inputs we need to compute the expected return of a stock in our model? Equation (20) shows that, to determine  $\mu_i$ , and hence stock  $i$ 's expected return, we need to know  $\sigma_i$ ,  $\zeta_i$ ,  $g_i$ , and  $\beta_i$ . 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 value 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 book-to-market ratio and then group them into deciles. (For each of the other anomalies, we instead rank stocks on the relevant anomaly characteristic – for example, on their idiosyncratic volatility in the case of the volatility anomaly.) Decile 1 corresponds to stocks with the lowest book-to-market ratios, and decile 10 to stocks with the highest book-to-market ratios. Suppose that, in some particular month, each decile contains 100 stocks. Take decile 4, say. To compute the beta of the typical stock in decile 4 in this month, we calculate the betas of each of the 100 stocks in the decile and average them.<sup>16</sup> To compute the capital gain overhang of the typical stock in decile 4, we calculate the gain overhang for each stock in the decile – the percentage gain or loss since purchase for the average investor in the stock – and average these 100 numbers. To compute the volatility and skewness of the typical stock in decile 4 over the next year, we record the returns, over the next year, of the 100 stocks in the decile and compute the cross-sectional volatility and skewness of these 100 returns. We conduct this exercise for each decile in this month. At the end of this process, we have four quantities in hand for each anomaly decile in this 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 book-to-market decile, a time series for each of the four quantities: return volatility, return skewness, gain overhang, and beta. In the final step, we compute the mean of each time

---

<sup>15</sup>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  $N = 10$ .

<sup>16</sup>For a given stock in a given month, we compute its beta using daily returns over the following year.

series. For each book-to-market 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 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.<sup>17</sup>

In the calculations described above, 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 an 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 that an individual stock is held for about two years, on average (Chordia, Subrahmanyam, and Anshuman, 2001). 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.

As noted above, we compute the volatility and skewness of the typical stock in an anomaly decile as the cross-sectional volatility and skewness of the subsequent returns of the 100 stocks in the decile. This approach has a number of advantages. By measuring the likely volatility and skewness of a stock going forward, rather than the stock's past volatility and skewness, it focuses on what a rational, forward-looking investor is interested in. It also captures a natural way in which real-world investors may judge the future volatility or skewness of a stock: when trying to estimate a stock's skewness, say, they may take a set of stocks with similar characteristics and then check how often stocks in this set post an extreme right-tail return. The cross-sectional volatility and skewness are fairly stable from month to month, which means that investors can learn them from even a short sample of data. One limitation of this approach is that, due to common factors in the returns of stocks in the same decile, it may underestimate the volatility and skewness of these returns. However, if this bias is similar across deciles, its impact on the alphas we compute later will be minor.<sup>18</sup>

Before presenting the empirical characteristics of the 220 anomaly deciles, we clarify the definition of one key variable: the capital gain overhang. There are two approaches to

---

<sup>17</sup>For four of the anomalies – O-Score, post-earnings announcement drift, failure probability, and difference of opinion – data availability requires us to begin the computation after July 1963: in January 1972, December 1973, December 1974, and January 1980, respectively. For the size anomaly, we follow standard practice in using NYSE rather than CRSP breakpoints. For this one anomaly, then, decile 1 contains many more stocks than decile 10.

<sup>18</sup>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 is challenging, particularly in the case of skewness, in that it relies on additional assumptions about the autocorrelations of stock returns. Weighing the advantages and disadvantages of the time-series and cross-sectional approaches, we view the latter as superior.

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 – the Grinblatt and Han (2005) measure, because it is easier to compute and accounts for both individual and institutional investors – and stick with it throughout.<sup>19</sup>

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; these are computed month by month, 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 explained 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 standard deviation of 76%, while the typical stock in decile 10 has an annual 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 – 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

---

<sup>19</sup>Grinblatt and Han (2005) compute the gain overhang as  $(P - R)/P$ , where  $P$  is a stock's current price and  $R$  is the average purchase price. We compute it slightly differently, as  $(P - R)/R$ . This measure is a more precise match for the capital gain variable  $g$  in our model.

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 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.<sup>20</sup>

The empirical relationships in Figure 2 point to the necessity of our quantitative approach. 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 gain overhang than the typical stock in the other extreme decile, decile 10; of the 22 anomalies, 20 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 opposite direction, we need a quantitative model to determine whether prospect theory can explain the anomaly.

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.

---

<sup>20</sup>The correlations between volatility, skewness, and gain overhang shown in Figure 2 also hold at the individual stock level; see Table 1 in An et al. (2020). However, the relationships are significantly stronger at the anomaly decile level – and it is these decile-level relationships that matter when predicting decile-level average returns.

## 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 the decile to be identical: they have the same standard deviation, skewness, gain overhang, and beta, namely the empirical standard deviation, skewness, gain overhang, and beta of the typical stock in that anomaly decile, computed as described above. We set the parameters  $S_i$  and  $\zeta_i$  of the skewed  $t$  distribution, the capital gain  $g_i$ , and the beta  $\beta_i$  to capture these empirical values. We then search for a location parameter  $\mu_i$  so that the conditions for equilibrium described in Section 3.2 around equation (20) are satisfied. Our model's prediction for a stock's expected return is then given by (13). Note that all stocks in a given decile will have the same  $\mu_i$  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;  $\{S_i\}$ , the dispersion parameters for stock returns;  $\{\zeta_i\}$ , the asymmetry parameters for stock returns;  $\nu$ , the degree of freedom parameter;  $\{\beta_i\}$ , the stocks' betas;  $\{g_i\}$ , the stocks' capital gains;  $\sigma_M$ , the standard deviation of stock market returns; and  $\{\theta_{M,i}\}$ , the market weights for the  $N$  stocks. The investor-level parameters are  $\hat{\gamma}$ , portfolio risk aversion;  $\hat{b}_0$ , the importance of the prospect theory term in investor preferences;  $(\alpha, \delta, \lambda)$ , the prospect theory preference parameters; and  $\{\theta_{i,-1}\}$ , investors' prior allocations to the  $N$  stocks.<sup>21</sup>

We start with the asset-level parameters. We set  $\nu = 7.5$ , which represents a reasonable degree of fat-tailedness in stock returns; our results are not very sensitive to the value of  $\nu$ . We then set the dispersion parameters  $\{S_i\}$  and the asymmetry parameters  $\{\zeta_i\}$ . To do this, recall from equations (14) and (15) that, for the GH skewed  $t$  distribution,

$$\text{Std}(\tilde{R}_i) = \left[ \frac{\nu}{v-2} S_i + \frac{2\nu^2}{(\nu-2)^2(\nu-4)} \zeta_i^2 \right]^{0.5} \quad (23)$$

---

<sup>21</sup>An alternative to calibrating the model is to estimate it – specifically, to estimate the values of  $\alpha$ ,  $\delta$ ,  $\lambda$ ,  $\hat{b}_0$ , and  $\hat{\gamma}$ . However, without the use of approximations, estimation is not feasible: it requires generating expected returns for the 220 anomaly deciles for thousands of different sets of parameter values; because of the numerical integration involved, generating the 220 expected returns even for *one* set of parameter values takes a substantial amount of time. One appealing feature of our calibration approach is that it allows us to see how well the model performs when parameterized using data that are *independent* of the anomalies it is trying to explain.

$$\text{Skew}(\tilde{R}_i) = \frac{2\zeta_i \sqrt{\nu(\nu-4)}}{\sqrt{S_i}(2\nu\zeta_i^2/S_i + (\nu-2)(\nu-4))^{\frac{3}{2}}} \left[ 3(\nu-2) + \frac{8\nu\zeta_i^2}{S_i(\nu-6)} \right]. \quad (24)$$

To set  $S_i$  and  $\zeta_i$  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 (23) and (24). These equations then allow us to solve for the two unknowns,  $S_i$  and  $\zeta_i$ . 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  $S_i$  and  $\zeta_i$  for  $i \in \{1, \dots, 100\}$ , we solve

$$0.76 = \left( \frac{7.5}{7.5-2} S_i + \frac{2(7.5)^2}{(7.5-2)^2(7.5-4)} \zeta_i^2 \right)^{0.5} \quad (25)$$

$$4.27 = \frac{2\zeta_i \sqrt{7.5(7.5-4)}}{\sqrt{S_i}(2(7.5)\zeta_i^2/S_i + (7.5-2)(7.5-4))^{\frac{3}{2}}} \left[ 3(7.5-2) + \frac{8(7.5)\zeta_i^2}{S_i(7.5-6)} \right]. \quad (26)$$

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  $\{g_i\}$ , the stocks' capital gains;  $\sigma_M$ , the standard deviation of annual stock market returns;  $\{\theta_{M,i}\}$ , the stocks' 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 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  $\sigma_M$  to 0.25 and the gross risk-free rate  $R_f$  to 1.

We set  $\{\theta_{M,i}\}$ , the stocks' 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 anomaly decile. We then compute the time-series averages of these fractions. 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  $i = 1, \dots, 100$ . We proceed similarly for the other deciles.

We now turn to the investor-level parameters. We set  $\hat{\gamma}$ , the scaled portfolio risk aversion in (21), and  $\hat{b}_0$ , the scaled weight on the prospect theory term in (22), to generate an aggregate equity premium of 6%. There are many pairs  $(\hat{\gamma}, \hat{b}_0)$  that produce an equity premium of 6%.

How do we choose one? As we increase  $b_0$ , we increase not only 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 choose, from among the  $(\hat{\gamma}, \hat{b}_0)$  pairs that generate a 6% equity premium, the one with the highest level of  $\hat{b}_0$  that still produces plausible levels of under-diversification. This is the pair  $(\hat{\gamma}, \hat{b}_0) = (0.6, 0.6)$ .<sup>22</sup> 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. As noted earlier and discussed further in the Internet Appendix, our results are not very sensitive to this initial allocation.

Finally, we set the preference parameters  $\alpha$ ,  $\delta$ , and  $\lambda$ . 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 play a significant 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  $\lambda$  is approximately 2. However, recent studies suggest 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  $\lambda$  is just 1.31. Chapman et al. (2018) argue that even this estimate may be too high. Based on a large group of people who are more representative of the population than the participants in the typical experimental study, they obtain estimates of  $\lambda$  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  $\alpha$  and  $\delta$ . The median estimate of  $\delta$  is close to Tversky and Kahneman's (1992) estimate; we therefore maintain  $\delta = 0.65$ . Experimental estimates of  $\alpha$  span a fairly wide range; most lie between 0.5 and 0.95. We set  $\alpha$  near the midpoint of this range, at 0.7. Our preference

---

<sup>22</sup>For these parameter values, the upper optimum  $\theta^{**}$  for the anomaly deciles with heterogeneous holdings is almost always much below 0.2. If we take  $\Theta_{M,R} = 0.5$ , this implies that, for any stock in which an investor has an undiversified position, the stock constitutes significantly less than 10% = 0.5(0.2) of his financial wealth.

parameter values are therefore<sup>23</sup>

$$(\alpha, \delta, \lambda) = (0.7, 0.65, 1.5). \quad (27)$$

By setting the stock-level parameters – specifically,  $S_i$ ,  $\zeta_i$ , and  $g_i$  – in the way described 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. We discuss this possibility in Section 6.1.

Now that we have parameterized the model, we can illustrate the equilibrium structure using the momentum anomaly as an example. For stocks in momentum decile 1, namely stocks 1 to 100, investors have identical holdings, each holding the market supply of each stock,  $\theta_{M,i}$ . For example, for stock 1, there is a value of  $\mu_1$ , the location parameter for stock 1, such that the function in (20), for  $i = 1$ , has a unique global maximum at  $\theta_1 = \theta_{M,1} = 1.85 \times 10^{-4}$ . Figure 3 plots the function in (20) for this value of  $\mu_1$ , 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 have heterogeneous holdings. For example, for stock 901, there is a value of  $\mu_{901}$  such that the function in (20), for  $i = 901$ , has two global maxima that straddle the market supply. The solid line in Figure 4 plots the function in (20) for this value of  $\mu_{901}$ , namely  $\mu_{901} = 0.5678$ . It shows the two global maxima, one at  $\theta_{901}^* = 9.5 \times 10^{-5} < \theta_{M,901}$  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.

As noted earlier, there is a simple and realistic intuition for the two global maxima in Figure 4. The heterogeneous holdings typically arise for stocks that are trading at a gain and that, in part because of this, have a high expected return. The lower optimum  $\theta^*$  reflects investors' desire to lock in their prior gains in these stocks by selling some of their holdings; the upper optimum  $\theta^{**}$  reflects investors' desire to benefit from the high expected return by increasing their allocation to these stocks. In equilibrium, the two strategies are equally

---

<sup>23</sup>The values of  $\alpha$ ,  $\delta$ , and  $\lambda$  in (27) 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  $\alpha$  and  $\lambda$  that are lower than those estimated by Tversky and Kahneman (1992). Meanwhile, even the lower level of loss aversion in (27) is strong enough to generate a high equity premium and non-participation in the stock market – two prominent applications of loss aversion in finance.

attractive.<sup>24</sup>

In Internet Appendix C, we give more details about the shape of the solid lines in Figures 3 and 4. With reference to the dashed and dash-dot lines in Figure 4, we also explain why, for the stocks in momentum decile 10, there is no equilibrium with identical holdings.

## 5 Application

We now use our model to answer a 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 alphas for the ten anomaly deciles over our 1963-2015 sample – these are value-weighted CAPM alphas computed from a monthly regression and annualized – and denote them as  $\alpha^d(1), \dots, \alpha^d(10)$ , where the “ $d$ ” superscript stands for “data.” We then compute the alphas predicted by our model for each of the ten deciles, namely  $\alpha^m(1), \dots, \alpha^m(10)$ , where “ $m$ ” stands for “model.” Since, within each decile, all stocks are identical and hence 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  $100l$ :

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

where  $\bar{R}_M = \sum_{i=1}^N \theta_{M,i} \bar{R}_i$ .

We say that our model can help explain an anomaly if

$$\text{sign}(\alpha^d(10) - \alpha^d(1)) = \text{sign}(\alpha^m(10) - \alpha^m(1)) \text{ and } |\alpha^m(10) - \alpha^m(1)| > 0.015. \quad (28)$$

The first condition in (28) 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 (28) 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

---

<sup>24</sup>Heterogeneous holdings due to non-unique optima of the objective function also arise in other settings; see Brunnermeier, Gollier, and Parker (2007) and Barberis and Huang (2008). However, the forces driving the non-unique optima are different in each case.

anomaly if

$$\text{sign}(\alpha^d(10) - \alpha^d(1)) = -\text{sign}(\alpha^m(10) - \alpha^m(1)) \text{ and } |\alpha^m(10) - \alpha^m(1)| > 0.015, \quad (29)$$

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

$$|\alpha^m(10) - \alpha^m(1)| < 0.015. \quad (30)$$

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. Finally, in Section 5.4, we compare our model’s performance to that of some widely-used factor models.

## 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 (28) are satisfied. Specifically, these are the momentum, failure probability, idiosyncratic volatility, gross profitability, 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. We present our model’s predictions about these anomalies in Figures 5-8.

The top-left graph in 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 alpha of each decile; the fact that it is upward sloping indicates that, historically, and controlling for beta, stocks with high medium-term past returns earned a higher average return than stocks with low medium-term past returns. The solid line plots the alphas predicted by our model for the ten anomaly deciles.

Figures 5-8 also present results for 12 other anomalies that our model can help explain. The graphs all have the same structure: the dashed lines plot the empirical alphas, while the solid lines plot the model-predicted alphas. For all 13 anomalies in these graphs, our model predicts a lower alpha for the extreme decile that, empirically, has a lower alpha. Moreover,

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, the stocks' 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.

Figures 5-8 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 Figure 5, 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.

Prior studies have linked the diminishing-sensitivity component of prospect theory to 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 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, idiosyncratic skewness, and failure probability anomalies (Campbell, Hilscher, and Szilagyi, 2008; Boyer, Mitton, and Vorkink, 2010; Conrad, Kapadia, and Xing, 2014). 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.<sup>25</sup>

Wang, Yan, and Yu (2017) propose and document empirically that, in a prospect theory framework, greater volatility can lead to *lower* average returns for assets trading at a loss because investors are in the convex region of the value function. This effect turns out not to play a major role in our results. The reason is that, as shown in the top-right graph of Figure 2, for each of the 220 stocks we are pricing, the stock’s return volatility exceeds the absolute value of its gain overhang. As such, the investors in these stocks are still quite close to the kink in the value function, and they price the stocks accordingly: greater volatility interacts with the kink to generate a *higher* average return.<sup>26</sup>

Figures 5-8 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 gross profitability, return on assets, 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 (29), 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.

The top-left graph in Figure 9 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

---

<sup>25</sup>We find that all three characteristics – volatility, skewness, and gain overhang – play a significant role in determining the average return; it is not that one of them plays an outsize role.

<sup>26</sup>Similarly, An et al. (2020) propose and document empirically that, in a prospect theory framework, greater skewness lowers an asset’s average return primarily for assets trading at a loss. This effect is also not playing a major role in our results: we find that, for the 220 stocks we are pricing, a marginal increase in skewness leads to a lower average return regardless of whether the stock is trading at a gain or at a loss.

predicted by the model. The graph shows that the model has mixed success in explaining this anomaly. On the one hand, the model captures the small 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.

The top-right graph in Figure 9 presents the results for the value anomaly. The model fails to explain this anomaly. The reason is that value stocks have more positively-skewed returns 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.

Figures 9-10 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 these 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 (30), it predicts alphas for the two extreme decile portfolios that differ by less than 1.5% in absolute magnitude. Figure 11 presents the results for these anomalies.

### 5.4 Model performance

The results presented so far show that our model is helpful for thinking about a majority of the 22 anomalies. We now examine how the model performs relative to other pricing models – in particular, relative to the CAPM; the Fama-French three-factor model; the Carhart four-factor model; the more recently-proposed Fama-French five-factor model; and a six-factor model that augments the five-factor model with a momentum factor. Specifically, we

compute the models' average absolute pricing errors for the 22 long-short portfolios that, for each anomaly, go long the stocks in anomaly decile 1 and short the stocks in anomaly decile 10. For example, for the prospect theory model, the pricing error for a given long-short anomaly portfolio is

$$[\bar{R}^d(1) - \bar{R}^d(10)] - [\bar{R}^m(1) - \bar{R}^m(10)],$$

where  $\bar{R}(l)$  is the average return of the stocks in anomaly decile  $l$ , and where  $d$  and  $m$  refer to the data and the model, respectively. The pricing error for the three-factor model, say, is computed in the usual way as the intercept in a regression of the long-short portfolio return on the three Fama-French factors.

Table 3 reports the average absolute monthly pricing error across the 22 anomalies for the prospect theory model and for the five factor models. The table shows that our model performs better than the CAPM and three-factor model, similarly to the four-factor model, and less well than the five- and six-factor models.

It is striking that our model performs as well as the four-factor model, given that the latter was developed in full knowledge of the size, value, and momentum anomalies. By contrast, our model was developed independently of any information about the anomalies: prospect theory itself was designed with no prior knowledge of any anomalies, and we calibrated the model based on experimental evidence that is again independent of the anomalies. Despite being handicapped in this way, our model is able to match the four-factor model in its performance. It is not surprising that our model underperforms the five- and six-factor models, as these were developed in full knowledge of multiple anomalies.

## 6 Discussion

In Section 5, we saw that the prospect theory model can help explain a majority of the 22 anomalies in Table 1. 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. In the Internet Appendix, we discuss the impact of investor heterogeneity on our results and contrast our predictions with those of a model that assumes broad, rather than narrow, framing.

## 6.1 Investor beliefs

Figures 9 and 10 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. Indeed, for a number of the anomalies that our model performs poorly on – for example, size, value, and investment – plausible explanations are already available. This suggests that, for these anomalies, other factors are more important than prospect theory in determining average returns.

There is another answer, however, one that applies within the prospect theory framework. In generating Figures 5-11, we assumed that investors have accurate beliefs about stocks' return volatility, return skewness, and gain overhang. However, for certain types of stocks, investors' beliefs about these characteristics may be incorrect.

In Internet Appendix D and Table A1, we present some evidence on this issue. Specifically, we take data on the prices of stock options and use them to extract estimates of risk-neutral variance and risk-neutral skewness – for example, the risk-neutral skewness of the typical stock in decile 1 of the value anomaly or decile 10 of the failure probability anomaly. Since these are estimates of risk-neutral quantities, they can offer only preliminary evidence on investors' beliefs about the physical return distribution. However, prior studies suggest that a substantial part of the variation in risk-neutral moments is due to variation in beliefs about physical returns (Birru and Wang, 2016).

The results in Table A1 suggest that, for most types of stocks, investors have sensible beliefs: the risk-neutral variance and skewness are almost always higher for stocks whose actual return variance and skewness are higher. For example, the risk-neutral skewness of the typical stock in decile 10 of the volatility anomaly is higher than the risk-neutral skewness of the typical stock in decile 1 for this anomaly, suggesting that investors are aware of the higher physical skewness of volatile stocks. There is, however, one notable exception: the risk-neutral skewness of value stocks is the same as that of growth stocks, even though the *actual* return skewness of value stocks is much higher than that of growth stocks.

This last finding may explain why our model performs poorly on the value anomaly. To generate the value anomaly graph in Figure 9, 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 the model predicts for value stocks relative to growth stocks, reducing the gap between

the empirical and model-predicted alphas.<sup>27</sup>

## 6.2 Limits to arbitrage

In our model, investors who evaluate risk according to prospect theory leave an imprint 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 relates to the specific form of the mispricing that prospect theory investors generate, namely, *mild* underpricing but *severe* overpricing. This pattern can be seen in almost all of the 22 anomalies. For example, in the volatility anomaly graph in Figure 5, the stocks in deciles 1 to 7 are slightly underpriced – they are predicted to earn small positive alphas – while the stocks in deciles 9 and 10 are very overpriced. Importantly, the stocks in deciles 9 and 10 are not only very volatile, but also have highly-skewed returns and low market capitalizations. To 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 have 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 will worsen in the short run, leading to fund outflows or margin calls, is an additional deterrent to professional arbitrageurs (Shleifer and Vishny, 1997).

Finally, to exploit the mispricing caused by prospect theory investors, 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.

---

<sup>27</sup>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.

## 7 Conclusion

As a highly influential model of risk attitudes, prospect theory has the potential to shed light on investor behavior. However, despite years of effort, a basic question about the theory's implications for asset prices remains unanswered: What does it predict about the relative average returns of different types of assets? In this paper, we answer this question, and in particular show that prospect theory can help explain a majority of 22 prominent stock market anomalies.

Our paper joins a growing body of work that shows that prospect theory is useful for thinking about a wide range of facts in finance. It also shows that a framework that is deeply rooted in psychology can nonetheless make clear quantitative predictions. We are not aware of a previous effort to use a behavioral model to make quantitative predictions about a large number of stock market anomalies, but our analysis shows that this goal can be achieved. We expect to see much more finance research that is both psychological *and* quantitative in the years ahead.

## 8 Supplementary details

### 8.1 The probability weighting terms

Here, we provide the full expressions for the  $dw(P(R_i))$  and  $dw(1 - P(R_i))$  terms which appear in (11), (16), and (20). 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 above 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 (12). Similarly,

$$dw(1 - P(R_i)) = \frac{dw(1 - P(R_i))}{dP(R_i)} \frac{dP(R_i)}{dR_i} dR_i$$

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

## 8.2 Rescaling the decision problem

Substituting the definitions in (19) into (16) and multiplying the resulting expression by the exogenous parameter  $\Theta_{M,R}^{-1}$ , we obtain

$$\begin{aligned} & \theta_i(\mu_i + \frac{\nu\zeta_i}{\nu-2} - R_f) - \frac{\hat{\gamma}}{2}(\theta_i^2\sigma_i^2 + 2\sum_{j\neq i}\sigma_{ij}\theta_i\theta_{M,j}) \\ & - \lambda\hat{b}_0\int_{-\infty}^{R_f-\theta_{i,-1}g_i/\theta_i}(\theta_i(R_f - R_i) - \theta_{i,-1}g_i)^\alpha dw(P(R_i)) \\ & - \hat{b}_0\int_{R_f-\theta_{i,-1}g_i/\theta_i}^\infty(\theta_i(R_i - R_f) + \theta_{i,-1}g_i)^\alpha dw(1 - P(R_i)), \end{aligned} \quad (31)$$

where

$$\hat{\gamma} = \gamma W_0 \Theta_{M,R}, \quad \hat{b}_0 = b_0 W_0^{\alpha-1} \Theta_{M,R}^{\alpha-1}.$$

It follows that if  $\Theta_i$  maximizes (16), then  $\theta_i = \Theta_i/\Theta_{M,R}$  maximizes (31), and conversely that if  $\theta_i$  maximizes (31), then  $\Theta_i = \theta_i\Theta_{M,R}$  maximizes (16). Maximizing (16) is therefore equivalent to maximizing (31).

The rescaling also allows us to simplify the variance term in the first row of (31). Specifically, the quantity

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

can be rewritten as

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

where  $\tilde{R}_M$  is the return on the market portfolio of all risky assets,  $\sigma_M$  is the standard deviation of this return, and  $\beta_i$  is the beta of asset  $i$ . Substituting this into (31) leads to the expression in (20).

## 8.3 Procedure for computing expected returns

The equilibrium structure we consider – a bounded-rationality equilibrium with heterogeneous holdings – consists of a location vector  $(\mu_1, \dots, \mu_N)$  such that, for each  $i$ , the expression

in (20) has either a unique global maximum at  $\theta_i^* = \theta_{M,i}$  or else two global maxima at  $\theta_i^*$  and  $\theta_i^{**}$  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 (20), substitute in  $\theta_i = \theta_{M,i}$ , and set the resulting expression to 0. This gives<sup>28</sup>

$$0 = (\mu_i + \frac{\nu\zeta_i}{\nu - 2} - R_f) - \hat{\gamma}\beta_i\sigma_M^2 - \alpha\lambda\hat{b}_0 \int_{-\infty}^{R_f - \theta_{i,-1}g_i/\theta_i} (\theta_{M,i}(R_f - R_i) - \theta_{i,-1}g_i)^{\alpha-1}(R_f - R_i)dw(P(R_i)) - \alpha\hat{b}_0 \int_{R_f - \theta_{i,-1}g_i/\theta_i}^{\infty} (\theta_{M,i}(R_i - R_f) + \theta_{i,-1}g_i)^{\alpha-1}(R_i - R_f)dw(1 - P(R_i)). \quad (32)$$

We then check whether, for the  $\mu_i$  that solves (32), the function in (20) 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 (20) does not have a global maximum at this  $\theta_i$ , then, in equilibrium, investors do not have identical holdings of asset  $i$ . We must instead look for an equilibrium with heterogeneous holdings.

To find an equilibrium with heterogeneous holdings of asset  $i$ , we search for a value of  $\mu_i$  such that the maximum value of the function in (20) in the range  $(-\infty, \theta_{M,i})$ , 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  $\mu_i$ , 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  $\mu_i$  for which this holds is typically close to the value of  $\mu_i$  that solves the first-order condition (32). Therefore, if we find that, for the value of  $\mu_i$  that solves (32), the function in (20) does not have a global maximum at  $\theta_i = \theta_{M,i}$ , we search in the neighborhood of that  $\mu_i$  for an equilibrium with heterogeneous holdings.

---

<sup>28</sup>As discussed in Section 4, we take  $\theta_{i,-1} = \theta_{M,i}$ , which further simplifies (32).

## 9 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 (2020), “Lottery-related Anomalies: The Role of Reference-dependent Preferences,” *Management Science* 66, 473-501.

Andrikogiannopoulou, A. and F. Papakonstantinou (2019), “History-dependent Risk Preferences: Evidence from Individual Choices and Implications for the Disposition Effect,” forthcoming, *Review of Financial Studies*.

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., Driessens, J., Ebert, S., Londono, J., and O. Spalt (2019), “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 W. Xiong (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.

Bernheim, D. and C. Sprenger (2019), “Direct Tests of Cumulative Prospect Theory,” Working paper.

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.

Brunnermeier, M., Gollier, C., and J. Parker (2007), “Optimal Beliefs, Asset Prices, and the Preference for Skewed Returns,” *American Economic Review Papers and Proceedings* 97, 159-165.

Campbell, J.Y., Hilscher, J., and J. Szilagyi (2008), “In Search of Distress Risk,” *Journal of Finance* 63, 2899-2939.

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.

Chordia, T., Subrahmanyam, A., and V. R. Anshuman (2001), “Trading Activity and Expected Stock Returns,” *Journal of Financial Economics* 59, 3-22.

Conrad, J., Kapadia, N., and Y. Xing (2014), “Death and Jackpot: Why Do Individual Investors Hold Overpriced Stocks?,” *Journal of Financial Economics* 113, 455-475.

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.

Fudenberg, D., Kleinberg, J., Liang, A., and S. Mullainathan (2019), “Measuring the Completeness of Theories,” Working paper.

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.

Ingersoll, J. (2014), “Cumulative Prospect Theory, Aggregation, and Pricing,” *Critical Finance Review* 4, 1-55.

Kahneman, D. (2013), *Thinking, Fast and Slow*. Farrar, Straus, and Giroux.

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

Koszegi, B. and M. Rabin (2006), “A Model of Reference-dependent Preferences,” *Quarterly Journal of Economics* 121, 1133-1165.

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 Representation 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.

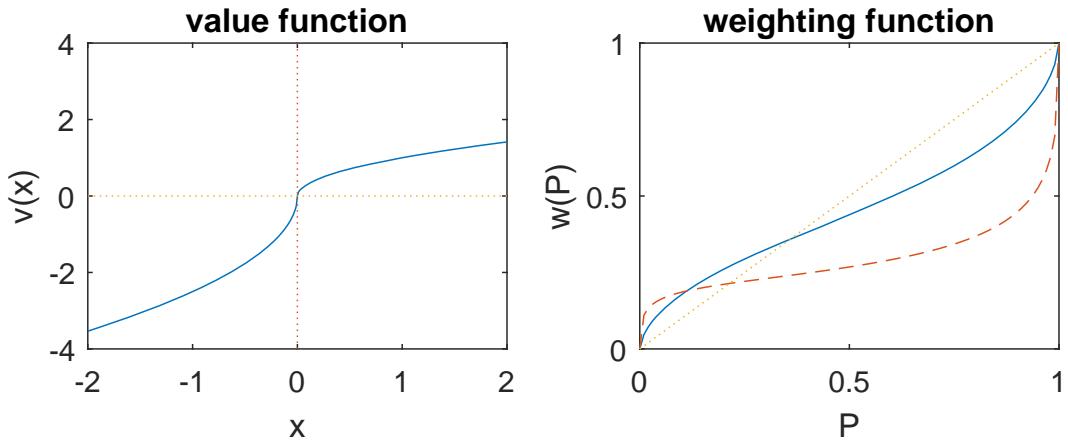
Anomaly	Abbreviation
Idiosyncratic volatility	IVOL
Market capitalization	SIZE
Value	VAL
Expected idiosyncratic skewness	EISKEW
Momentum	MOM
Failure probability	FPROB
O-Score	OSC
Net stock issuance	NSI
Composite equity issuance	CEI
Accrual	ACC
Net operating assets	NOA
Gross profitability	PROF
Asset growth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

**Table 2.** Empirical properties of anomaly deciles. The first column lists 22 anomalies; the acronyms are defined in Table 1. The remaining columns report, for each anomaly, the average annual return, standard deviation of annual returns, skewness of annual returns, and capital gain overhang (in percent) of the typical stock in anomaly decile 1 and anomaly decile 10.

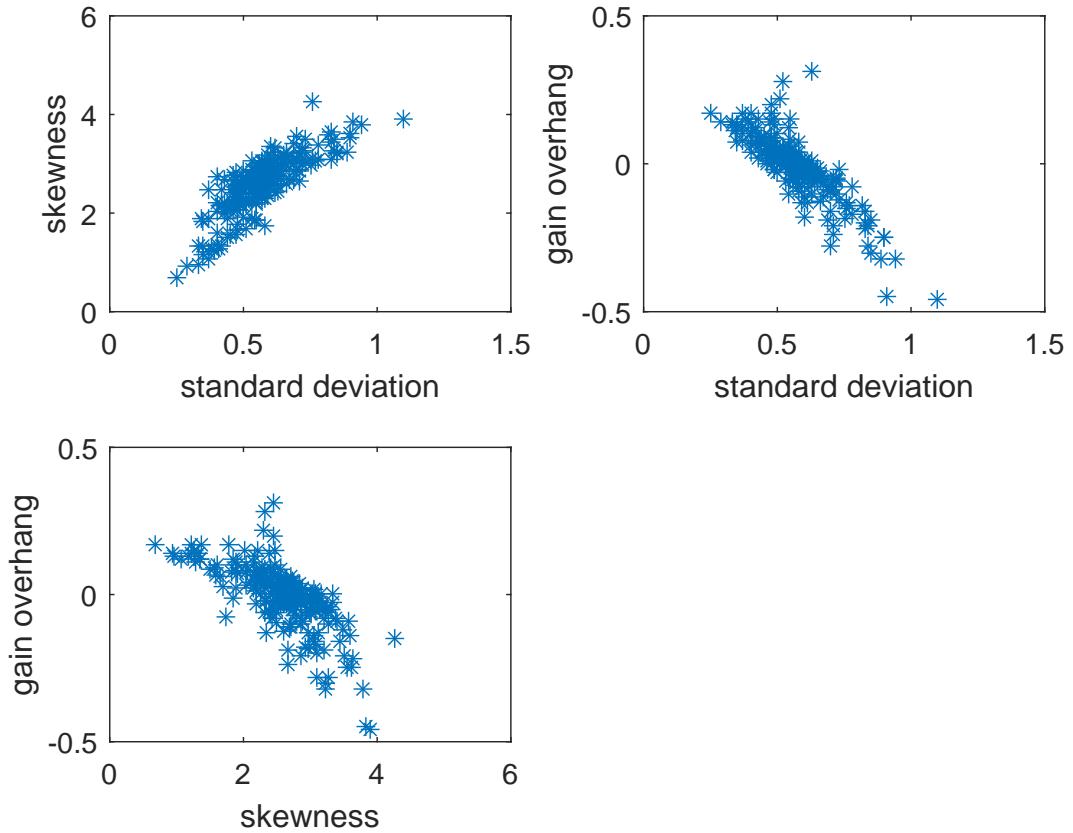
Anomaly	Average return	Average return	Standard deviation	Standard deviation	Skewness	Skewness	Gain overhang	Gain overhang
	Decile 1	Decile 10	Decile 1	Decile 10	Decile 1	Decile 10	Decile 1	Decile 10
IVOL	11.9	-3.2	36.8	94.3	2.46	3.79	10.5	-31.6
SIZE	14.0	10.6	76.0	24.6	4.27	0.69	-15.4	17.0
VAL	10.7	17.7	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
ROA	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
ORGCAP	10.9	15.8	63.3	77.5	2.76	3.38	-3.8	-7.9
LTRREV	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
DOP	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 table reports the average absolute monthly pricing error across the 22 anomalies for the prospect theory model of Section 3 and for five factor models: the CAPM; the Fama-French three-factor model; the Carhart four-factor model; the Fama-French five-factor model; and a six-factor model that augments the five-factor model with a momentum factor. For a given factor model, we compute the alphas of the 22 long-short portfolios that, for the 22 anomalies, go long the stocks in anomaly decile 1 and short the stocks in anomaly decile 10; we report the average absolute alpha. For the prospect theory model, we compute, for the 22 anomalies, the difference in the empirical and theoretically-predicted average return of the 22 long-short portfolios; we report the average absolute return difference.

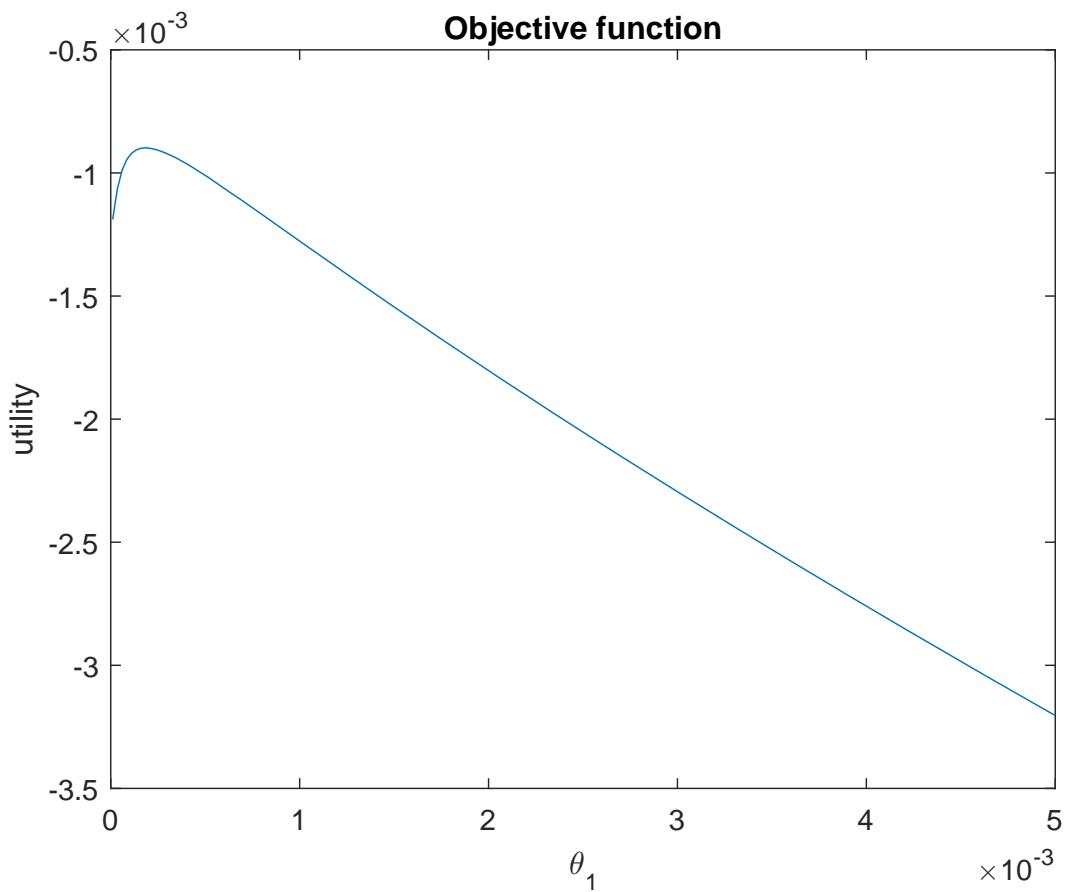
Model	Average absolute pricing error
Prospect theory	0.56
CAPM	0.81
Three-factor model	0.80
Four-factor model	0.56
Five-factor model	0.45
Six-factor model	0.32



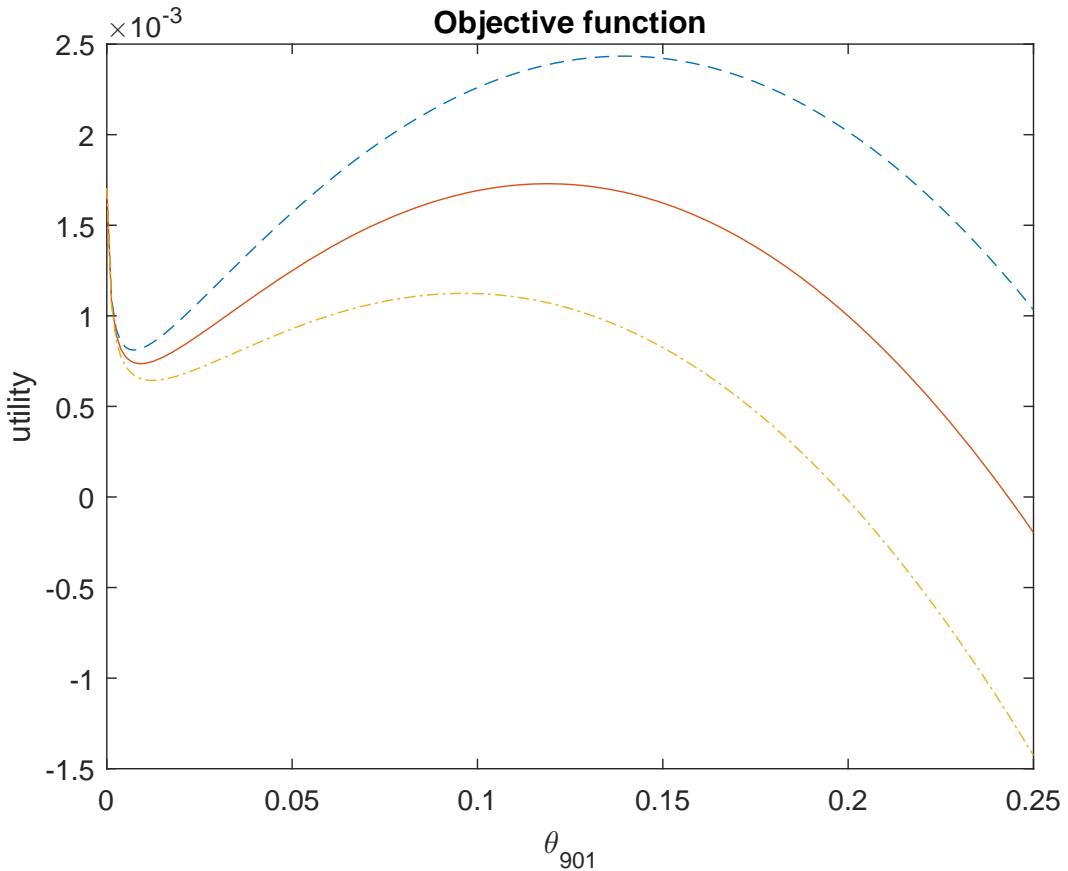
**Figure 1.** The left panel plots the value function proposed by Tversky and Kahneman (1992) as part of their cumulative prospect theory, namely,  $v(x) = x^\alpha$  for  $x \geq 0$  and  $v(x) = -\lambda(-x)^\alpha$  for  $x < 0$ , for  $\alpha = 0.5$  and  $\lambda = 2.5$ . The right panel plots the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



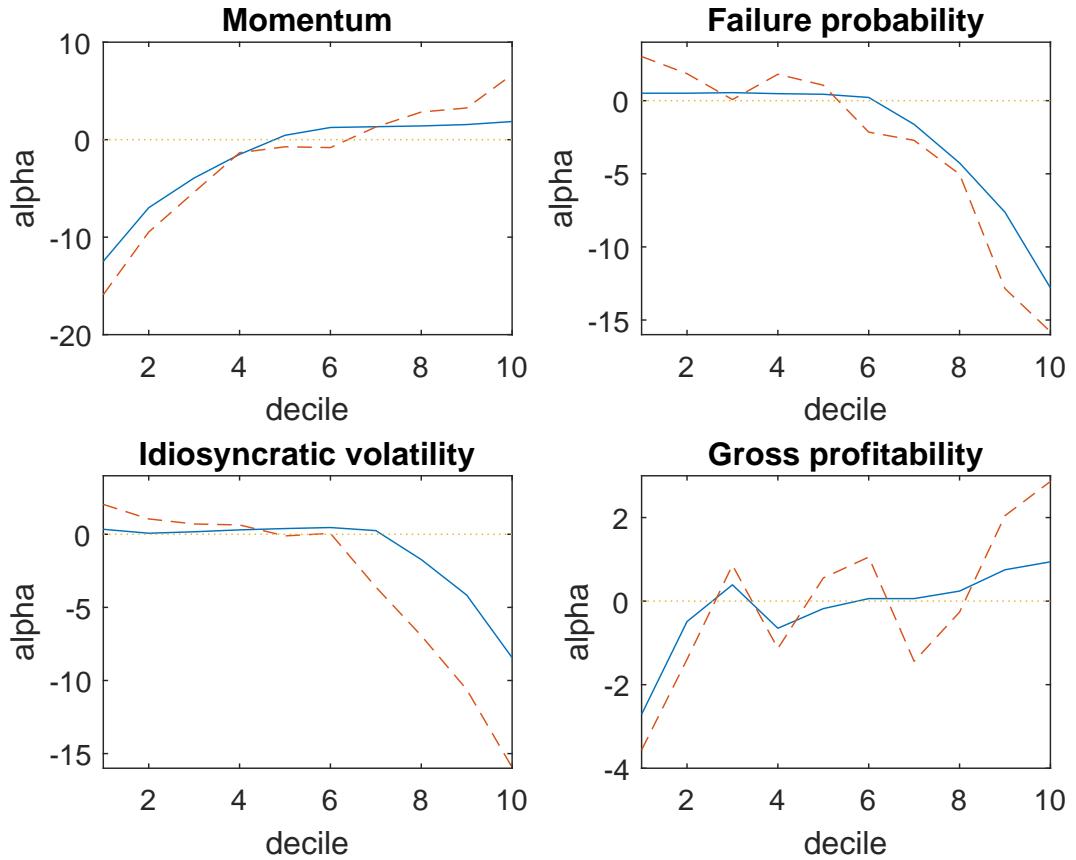
**Figure 2.** Each graph plots 220 asterisks, where each asterisk corresponds to one of 10 deciles for one of 22 anomalies. In the top-left graph, a given asterisk that corresponds to some anomaly decile marks the standard deviation of returns and the return skewness of the typical stock in that anomaly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



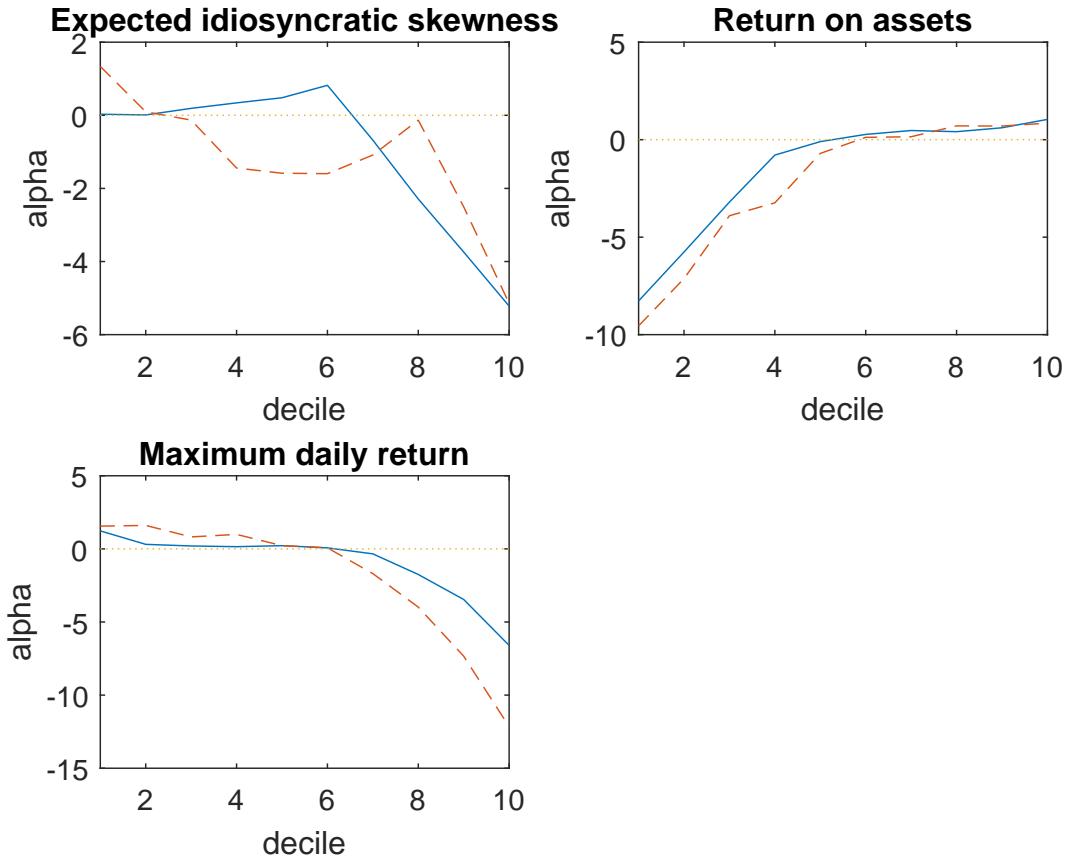
**Figure 3.** The graph shows that investors have identical holdings of each stock in momentum decile 1. The solid line plots the value of an investor's objective function in equilibrium, as a function of  $\theta_1$ , the (scaled) fraction of the investor's portfolio allocated to stock 1, which belongs to momentum decile 1. The function has a unique global maximum at the point where  $\theta_1$  equals the weight of stock 1 in the market portfolio, namely  $1.85 \times 10^{-4}$ .



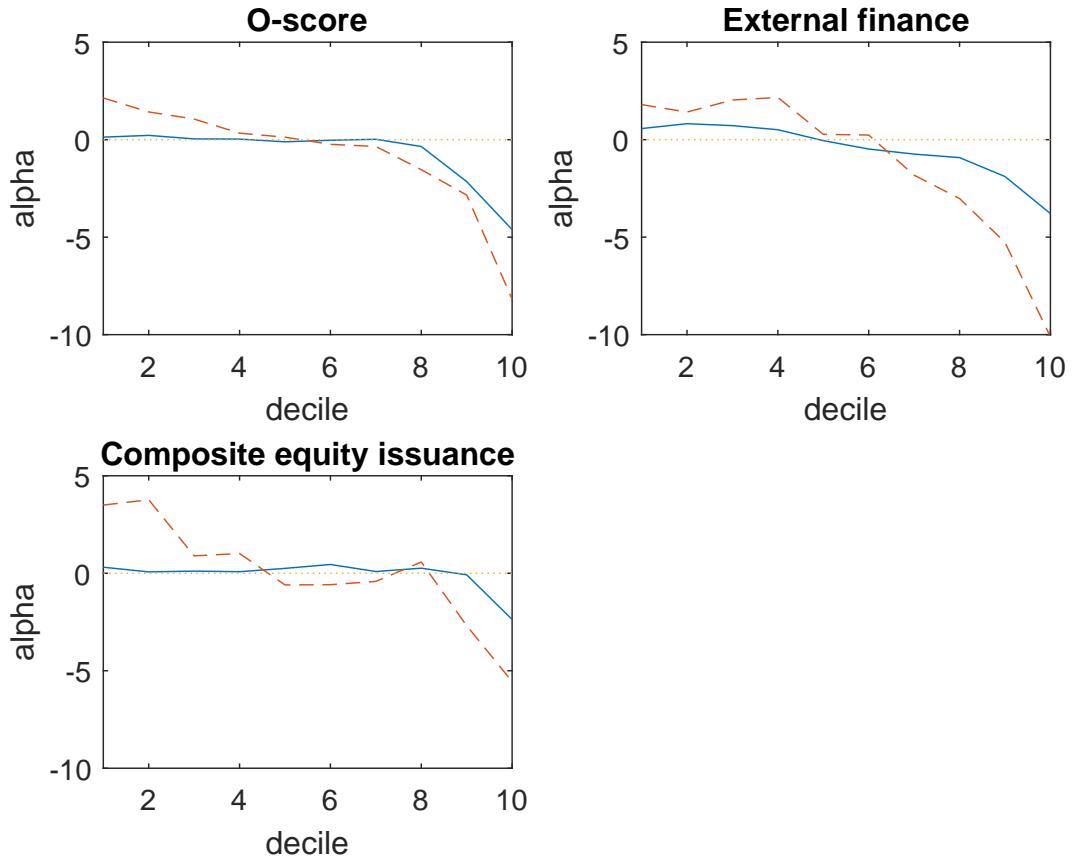
**Figure 4.** The graph shows that investors have heterogeneous holdings of each stock in momentum decile 10. The solid line plots the value of an investor's objective function in equilibrium, as a function of  $\theta_{901}$ , the (scaled) fraction of the investor's portfolio allocated to stock 901, which belongs to momentum decile 10. The function has two global maxima which straddle the weight of stock 901 in the market portfolio, namely  $7.5 \times 10^{-4}$ . 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.



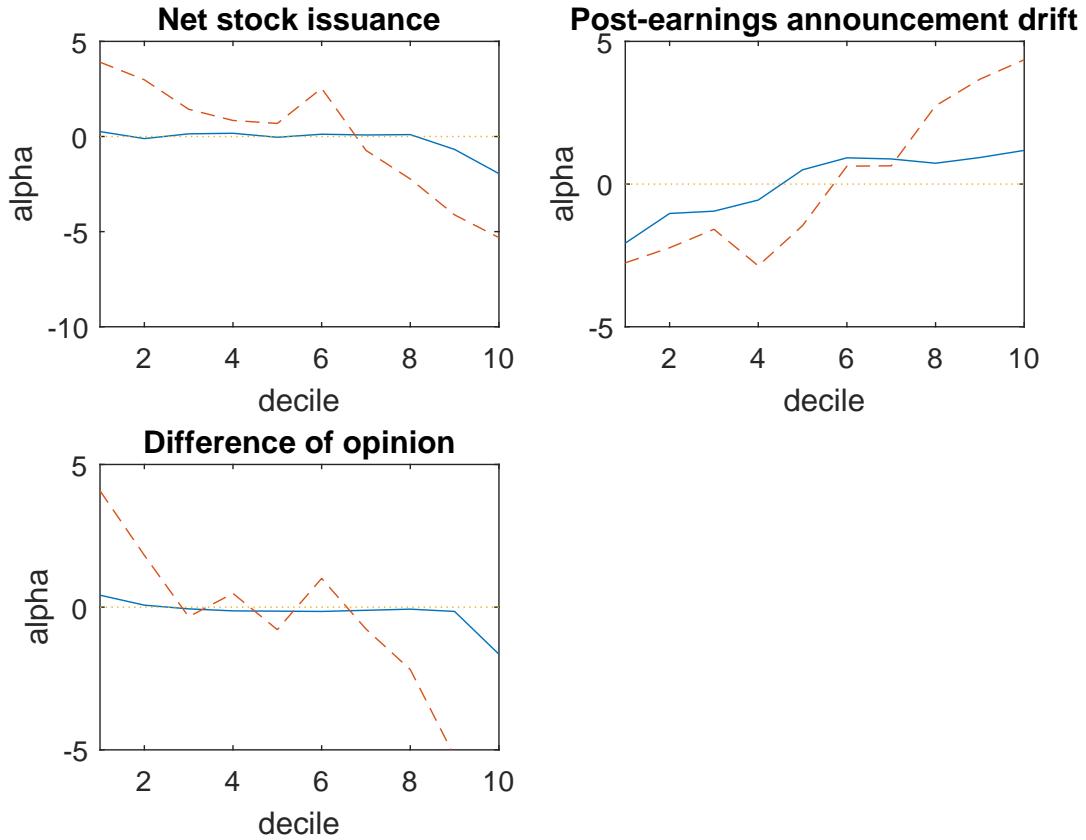
**Figure 5.** The dashed lines in the graphs plot the historical annual alpha of each anomaly decile for the momentum, failure probability, idiosyncratic volatility, and gross profitability anomalies. The solid lines plot the alphas predicted by a model in which investors evaluate risk according to prospect theory.



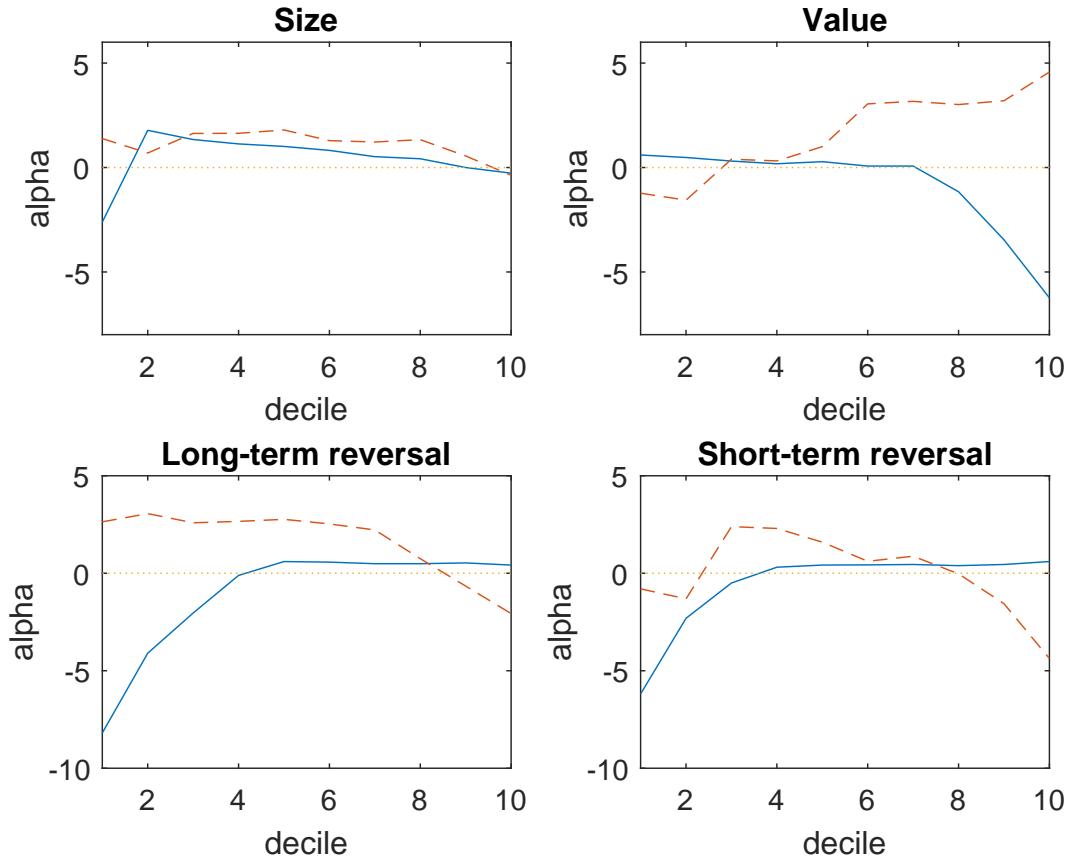
**Figure 6.** The dashed lines in the graphs plot the historical annual alpha of each anomaly decile for the expected idiosyncratic skewness, return on assets, and maximum daily return anomalies. The solid lines plot the alphas predicted by a model in which investors evaluate risk according to prospect theory.



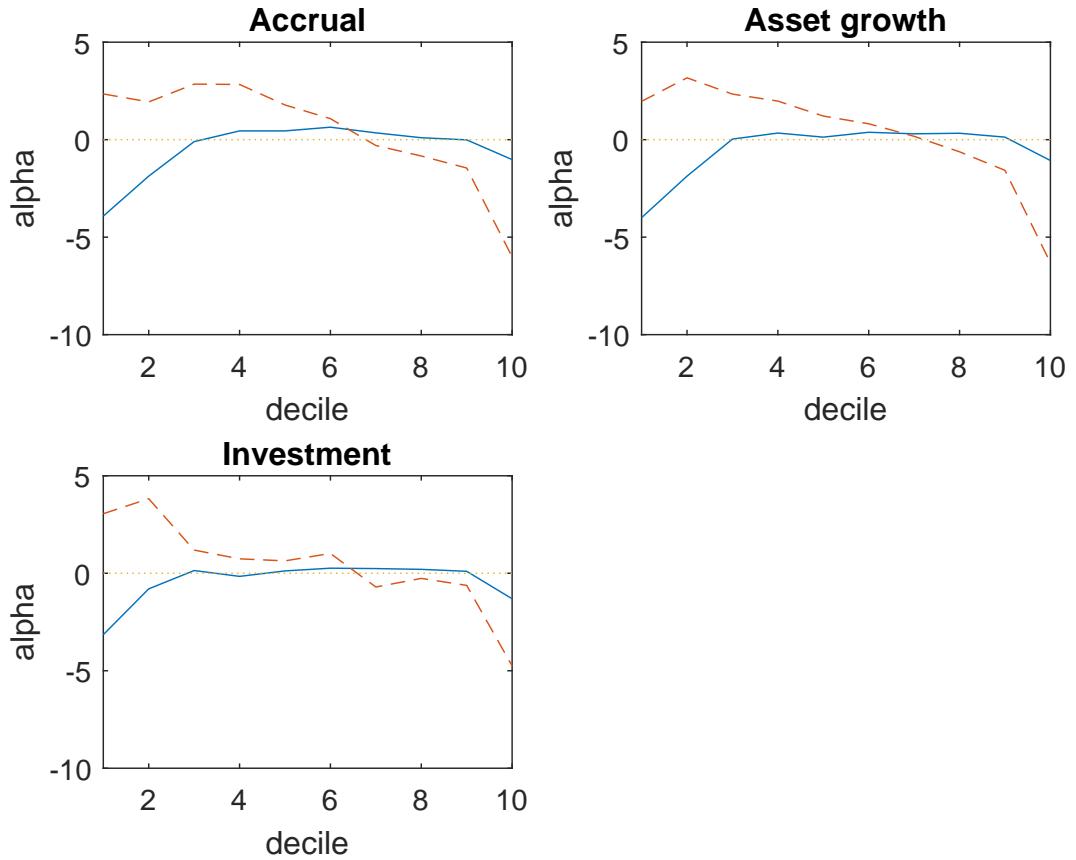
**Figure 7.** The dashed lines in the graphs plot the historical annual alpha of each anomaly decile for the O-score, external finance, and composite equity issuance anomalies. The solid lines plot the alphas predicted by a model in which investors evaluate risk according to prospect theory.



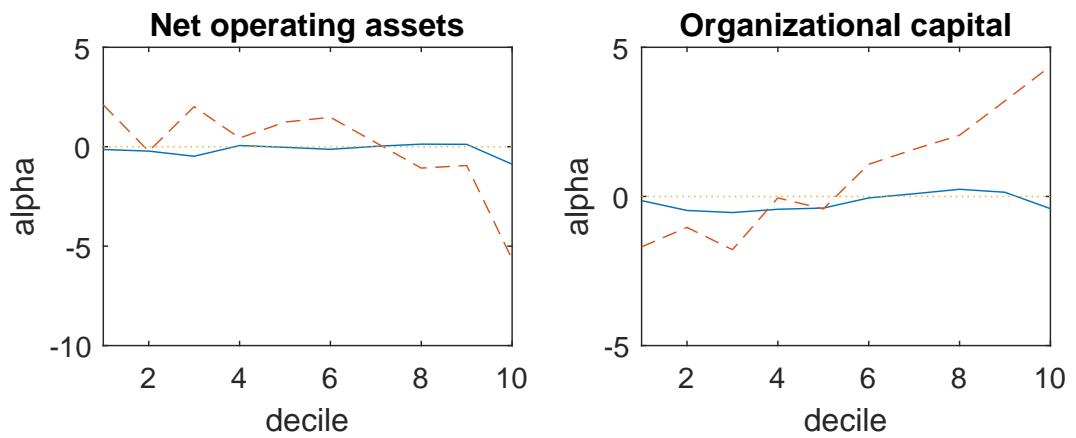
**Figure 8.** The dashed lines in the graphs plot the historical annual alpha of each anomaly decile for the net stock issuance, post-earnings announcement drift, and difference of opinion anomalies. The solid lines plot the alphas predicted by a model in which investors evaluate risk according to prospect theory.



**Figure 9.** The dashed lines in the graphs plot the historical annual alpha of each anomaly decile for the size, value, long-term reversal, and short-term reversal anomalies. The solid lines plot the alphas predicted by a model in which investors evaluate risk according to prospect theory.



**Figure 10.** The dashed lines in the graphs plot the historical annual alpha of each anomaly decile for the accrual, asset growth, and investment anomalies. The solid lines plot the alphas predicted by a model in which investors evaluate risk according to prospect theory.



**Figure 11.** The dashed lines in the graphs plot the historical annual alpha of each anomaly decile for the net operating assets and organizational capital anomalies. The solid lines plot the alphas predicted by a model in which investors evaluate risk according to prospect theory.

## INTERNET APPENDIX

### A. The Generalized Hyperbolic (GH) Skewed $t$ distribution

In our model, the vector of asset returns  $\tilde{R} = (\tilde{R}_1, \dots, \tilde{R}_N)$  follows an  $N$ -dimensional GH skewed  $t$  distribution. The density function for this distribution is

$$\begin{aligned} p(R) &= \frac{2^{1-\frac{\nu+N}{2}}}{\Gamma(\frac{\nu}{2})(\pi\nu)^{\frac{N}{2}}|S|^{\frac{1}{2}}} \cdot \frac{K_{\frac{\nu+N}{2}}(\sqrt{(\nu + (R - \mu)'S^{-1}(R - \mu))\zeta'S^{-1}\zeta}) \exp((R - \mu)'S^{-1}\zeta)}{\left(\sqrt{(\nu + (R - \mu)'S^{-1}(R - \mu))\zeta'S^{-1}\zeta}\right)^{-\frac{\nu+N}{2}} (1 + (R - \mu)'S^{-1}(R - \mu)/\nu)^{\frac{\nu+N}{2}}}, \\ &\text{for } \zeta \neq 0 \\ p(R) &= \frac{\Gamma(\frac{\nu+N}{2})}{\Gamma(\frac{\nu}{2})(\pi\nu)^{\frac{N}{2}}|S|^{\frac{1}{2}}} \cdot (1 + (R - \mu)'S^{-1}(R - \mu)/\nu)^{-\frac{\nu+N}{2}}, \text{ for } \zeta = 0, \end{aligned} \quad (33)$$

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

The above distribution has four parameters:  $\mu$ ,  $S$ ,  $\zeta$ , and  $\nu$ . Here,  $\mu = (\mu_1, \dots, \mu_N)$ , the vector of location parameters, helps to determine the mean of the distribution;  $S = \{S_{ij}\}$ , the dispersion matrix, controls the dispersion in returns;  $\zeta = (\zeta_1, \dots, \zeta_N)$ , the vector of asymmetry parameters, governs the skewness of returns; and  $\nu$ , a degree of freedom scalar, affects the heaviness of the tails of the distribution. The mean of the distribution is

$$\mu + \frac{\nu}{\nu - 2} \zeta. \quad (34)$$

For our computations, we need only the marginal distribution of each asset's return; this is the one-dimensional GH skewed  $t$  distribution in equation (12). In that equation,  $\mu_i$  and  $\zeta_i$  are the  $i$ 'th elements of the vectors  $\mu$  and  $\zeta$ , respectively, while  $S_i$  is the  $i$ 'th diagonal element of  $S$ .

### B. Stock market anomalies

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

IVOL. Idiosyncratic volatility. Standard deviation of the residuals from a firm-level regression of daily stock returns on the daily Fama-French three factors using data from the past month. See Ang et al. (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 (2013).

AG. Asset growth. Percentage change in total assets over the previous year. See Cooper, Gulen, and Schill (2008).

ROA. Return on assets. Income before extraordinary items divided by total assets. See Stambaugh, Yu, and Yuan (2012).

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).

ORGCAP. 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  $t - 1$ .

DOP. Difference 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).

## C. Additional details about the equilibrium structure

In this section, we present additional information about the equilibrium structure illustrated in Figures 3 and 4.

Figure 3 plots the objective function, in equilibrium, for a stock in momentum decile 1, stock 1, a stock for which investors have identical holdings. Why does it have this shape? In our model, investors usually have identical holdings in stocks that are trading at a loss and that, in part because of this, have a low, even negative, expected return. As an investor increases his allocation to stock 1, utility initially rises: he is in the convex region of the value function, where he is risk-seeking. As he further increases his allocation, utility falls: a larger fraction of time 1 outcomes bring the investor to the right of the kink in the value function, increasing his risk aversion; the stock's low expected return and the greater portfolio volatility caused by the larger allocation to the stock decrease utility further.

The solid line in Figure 4 plots the objective function, in equilibrium, for a stock in momentum decile 10, stock 901, a stock for which investors have heterogeneous holdings.

Why does it have this shape? The heterogeneous holdings typically arise for stocks that are trading at a gain and that, in part because of this, have a high expected return. As an investor increases his allocation to the stock, utility initially falls; this is because he is in the gain region of the prospect theory value function, where he is risk averse. As he further increases his allocation, utility starts to rise; this is due to the stock's high average return. Finally, as the investor increases his allocation even more, utility falls again: the large holdings of the stock create high portfolio volatility, which lowers utility.<sup>29</sup>

For any stock in momentum decile 10, there is no equilibrium in which investors have identical holdings in the stock. We can see this in Figure 4 for the case of stock 901. The dashed line plots the function in (20) for  $\mu_{901} = 0.5793$ . For this value of  $\mu_{901}$ , the function has a unique global maximum at  $\theta_{901} = 0.14 > \theta_{M,901} = 7.5 \times 10^{-4}$ . Since demand for stock 901 exceeds supply at this value of  $\mu_{901}$ , it appears that, to clear the market, we need to lower the value of  $\mu_{901}$ . However, as we do so, the value of  $\theta_{901}$  at which the function attains its maximum jumps discontinuously downward: the dash-dot line, which plots the objective function for a slightly lower value of  $\mu_{901}$ , namely 0.5743, shows that the unique maximum is now at  $\theta_{901} = 0 < \theta_{M,901}$ . As such, there is no value of  $\mu_{901}$  such that the function in (20), for  $i = 901$ , has a unique global maximum at  $\theta_{901} = \theta_{M,901}$ .

At time 0, for a stock in which investors have heterogeneous holdings, there is substantial volume as investors adjust their holdings of the stock from their time  $-1$  level,  $\theta_{M,i}$ , to either  $\theta_i^*$  or  $\theta_i^{**}$ . For a stock in which, at time 0, investors have identical holdings, there is no volume. A stock is more likely to have heterogeneous holdings if it has a positive gain overhang. Our model therefore predicts higher trading volume for stocks with a high positive gain overhang than for stocks with a low negative gain overhang. We find that, if, in historical U.S. data, we sort stocks by their gain overhang or, more simply, by their return over the previous year, then, consistent with this prediction, stocks in the decile with the highest gain overhang indeed have higher trading volume than stocks in the decile with the lowest gain overhang.

## D. Inferring beliefs from option prices

When generating our model's predictions for average returns, we assume that investors have sensible beliefs about the return volatility and return skewness of the typical stock in each anomaly decile. We now present some evidence on this assumption. Specifically, we

---

<sup>29</sup>As the investor increases his allocation to stock 901 from zero, the objective function actually rises slightly before falling. The initial rise is not visible to the naked eye in Figure 4, but it is the reason the lower optimum  $\theta^*$  is strictly greater than zero. For stock 901, the investor is in the gain region of the value function at time 0. While the value function is concave over gains, it exhibits second-order risk aversion. This, and the stock's high expected return, explain the objective function's initial rise.

take data on the prices of stock options and use them to extract estimates of risk-neutral variance and risk-neutral skewness.

Our options data come from the IvyDB OptionMetrics database, which provides option prices, volume, and open interest from January 1996 to December 2015. We use these data to compute, for each anomaly, the risk-neutral variance and risk-neutral skewness of the typical stock in each of the two extreme anomaly deciles. Specifically, 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 the results across the stocks in the decile. We then compute time-series means across all months in the sample.

Our methodology follows that of Birru and Wang (2016). We compute risk-neutral variance and skewness using equations (1)-(4) in their paper, which were originally derived by Bakshi, Kapadia, and Madan (2003); see their Theorem 1. These equations assume a continuum of options; we use a discrete approximation. We implement the calculation only for stocks with at least two out-of-the-money (OTM) puts and two OTM calls; Dennis and Mayhew (2002) argue that the biases that arise from the discrete approximation are small in this case.

We compute 30-day risk-neutral moments. One-year moments would be a better match for the volatility and skewness of annual returns reported in Table 2. We focus on 30-day moments because short-maturity options are more heavily traded and therefore likely to give us more reliable estimates. However, we have also computed 60-day and one-year moments and find that these lead to similar conclusions.

To compute the 30-day moments, we interpolate the moments 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.

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.

Table A1 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. Comparing these results to those in Table 2, we see that, for most types of stocks, investors have sensible

beliefs: the risk-neutral variance 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 the same as that of growth stocks, even though the *actual* return skewness of value stocks is much higher than that of growth stocks.

As noted in Section 6.1, this last finding may explain why our model performs poorly on the value anomaly. To generate the value anomaly graph in Figure 9, 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 the model predicts for value stocks relative to growth stocks, reducing the gap between the empirical and model-predicted alphas.

## E. Investor heterogeneity

To keep our model tractable, we assume that, along several dimensions, the investors in the economy are identical. For example, at time 0, the second term on the right-hand side of equation (10) is the same for all investors: they all have the same initial allocation  $\Theta_{-1}$  to each risky asset and experience the same capital gain  $g$ .

While it is not possible to study heterogeneity in a general way, our model does allow us to consider simple forms of heterogeneity. The outcome of this exercise is reassuring: the model predictions turn out to be quite robust to investor heterogeneity.

Take, for example, momentum decile 10. As shown in Table 2, the gain overhang of the typical stock in this decile is 30.8%. Accordingly, in our main analysis, we assume that, for this stock, all investors have the same prior gain of 30.8%, an assumption that leads the model to predict an expected return of 8.47% for the stock. Suppose instead that, for this stock, half of the investors have a gain of  $30.8 - 10 = 20.8\%$  while the other half have a gain of  $30.8 + 10 = 40.8\%$ . We find that there is then an equilibrium in which the objective function for the 40.8-gain investors has a single global optimum at  $\theta^* = 1.24 \times 10^{-4} < \theta_M = 7.47 \times 10^{-4}$ , while the objective function for the 20.8-gain investors has two global optima at  $\theta^* = 6.17 \times 10^{-5} < \theta_M$  and  $\theta^{**} = 0.109 > \theta_M$ . In equilibrium, the stock has an expected return of 8.35%. As such, heterogeneity affects the expected return, but in a minor way.

For momentum decile 1, where investors have identical holdings of each stock, the impact of heterogeneity is even smaller. As shown in Table 2, the gain overhang of the typical stock in this decile is  $-44.6\%$ . In our main analysis, we assume that all investors have the same prior gain of  $-44.6\%$  in the stock, and this leads to a predicted expected return of  $-6.66\%$ .

Suppose that we instead assume that half of the investors have a gain of  $-44.6 - 10 = -54.6\%$  while the other half have a gain of  $-44.6 + 10 = -34.6\%$ . In this case, there is an equilibrium in which the objective function of the  $-54.6$ -gain investors has a single global optimum at  $\theta^* = 2.25 \times 10^{-4} > \theta_M = 1.85 \times 10^{-4}$ , while that for the  $-34.6$ -gain investors has a single global optimum at  $\theta^* = 1.45 \times 10^{-4} < \theta_M$ . The predicted expected return is *the same* as in the equilibrium with identical investors, namely  $-6.66\%$ .

The fact that heterogeneity in  $g$  has a relatively small impact on our results immediately implies that heterogeneity in  $\Theta_{-1}$ , investors' time  $-1$  allocation to risky assets, will also have little impact: equation (10) shows that allowing for heterogeneity in  $\Theta_{-1}$  is mathematically equivalent to allowing for heterogeneity in  $g$ .

## F. Broad vs. narrow framing

The model of Section 3 assumes that investors engage in narrow framing: they derive prospect theory utility from *stock*-level gains and losses. How do the predictions of this model compare with those of a model with broad framing, in other words, one where investors derive prospect theory utility from *portfolio* gains and losses?

Models that apply prospect theory at the portfolio level are less tractable than those with narrow framing. However, under the same bounded-rationality assumption that we adopt in Section 3, we have been able to solve a model with broad framing, and have used it to generate predictions about the anomalies in Table 1. We find that the broad framing model is less successful than the narrow framing model at explaining the anomalies. Under narrow framing, a stock's gain overhang affects its average return in a way that can shed light on several anomalies. Under broad framing, however, investors' focus on portfolio gains and losses means that a stock's own gain overhang has much less impact on its average return; this, in turn, means that the broad framing model is less able to explain the observed facts. The greater success of the narrow framing model offers support for our conjecture, in Section 3, that narrow framing is a realistic psychological assumption.

**Table A1.** The first column lists 22 anomalies; the acronyms are defined in Table 1. 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 1996-2015 sample period.

Anomaly	RN-variance	RN-variance	RN-skewness	RN-skewness
	Decile 1	Decile 10	Decile 1	Decile 10
IVOL	0.03	0.17	-0.53	-0.21
SIZE	0.23	0.02	-0.12	-0.60
VAL	0.06	0.13	-0.36	-0.36
EISKEW	0.04	0.17	-0.45	-0.18
MOM	0.17	0.07	-0.21	-0.32
FPROB	0.05	0.22	-0.35	-0.24
OSC	0.05	0.18	-0.36	-0.18
NSI	0.05	0.10	-0.41	-0.28
CEI	0.04	0.10	-0.43	-0.28
ACC	0.12	0.09	-0.27	-0.26
NOA	0.10	0.07	-0.28	-0.34
PROF	0.13	0.06	-0.25	-0.34
AG	0.12	0.09	-0.27	-0.29
ROA	0.17	0.05	-0.19	-0.37
INV	0.09	0.08	-0.30	-0.32
MAX	0.03	0.14	-0.46	-0.29
ORGCAP	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	-0.40
DOP	0.04	0.11	-0.44	-0.29
PEAD	0.08	0.06	-0.33	-0.36