Prospect Theory and Stock Market Anomalies

NICHOLAS BARBERIS, LAWRENCE J. JIN, and BAOLIAN WANG*

ABSTRACT

We present a new model of asset prices in which investors evaluate risk according to prospect theory and examine its ability to explain 23 prominent stock market anomalies. The model incorporates all of the elements of prospect theory, accounts for investors' prior gains and losses, and makes quantitative predictions about an asset's average return based on empirical estimates of the asset's return volatility, return skewness, and past capital gain. We find that the model can help explain a majority of the 23 anomalies.

* Nicholas Barberis is at Yale University, Lawrence J. Jin is at the California Institute of Technology, and Baolian Wang is at the University of Florida. We are grateful to Stefan Nagel (the Editor), the Associate Editor, two anonymous referees, John Campbell, Allen Hu, Alex Imas, Jonathan Ingersoll, Eben Lazarus, Erik Loualiche, Ian Martin, Dimitri Vayanos, Keith Vorkink, and seminar participants at Baruch College, Boston College, Caltech, the London Business School, the London School of Economics, Michigan State University, 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, the NBER Behavioral Finance conference, and the SFS Cavalcade for very helpful comments. We have read the *Journal of Finance* disclosure policy and have no conflicts of interest to disclose.

Prospect theory, due to Kahneman and Tversky (1979) and Tversky and Kahneman (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 Capital Asset Pricing Model (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? 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 examine whether the model can account for 23 prominent stock market anomalies. We find that the model can help explain a majority of the 23 anomalies.

Prospect theory posits that people evaluate risk with 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"). The theory further posits 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 for which they have larger prior gains.[1]

The above intuitions indicate that, to understand prospect theory's implications for asset prices, we need a model that incorporates *all* of the elements of prospect theory *and* accounts for investors' prior gains and losses in each risky asset. No existing model of asset prices meets both conditions, and so we 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. Such an equilibrium does not exist, however, once we introduce prospect theory preferences. To break this logjam, we construct an alternative equilibrium in which investors hold different portfolios that correspond to nonunique maxima of their objective function. With this 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 23 prominent stock market anomalies. To see if the model can explain a particular anomaly – for instance, the size anomaly – we compute what it predicts for the average return of the typical small-cap stock. As noted above, this average return depends 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 the model, and record the model's prediction for the average return of the typical small-cap stock. We repeat this process for the typical stock in each of the 10 market capitalization deciles. The results reveal how much, if any, of the size anomaly the model can explain. We proceed in the same way for all 23 anomalies.

Our empirical estimates of the return volatility, return 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 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

[1]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 (2008a); and in Grinblatt and Han (2005) and Li and Yang (2013), respectively.

stock has not only more volatile returns than the typical large-cap stock, but also more skewed returns and a more negative gain overhang.

The 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 to charge a *higher* average return on small-cap stocks than on large-cap stocks, which helps 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, the latter two factors lead prospect theory investors to charge a *lower* average return on small-cap stocks, which reduces the model's ability to explain the size anomaly. Since one economic force goes in one direction and the other two go in the opposite direction, the only way to determine prospect theory's prediction for the size anomaly, or indeed for any anomaly, is to develop a quantitative model that combines all three forces. This has not been done before, and is what we do in this paper.

We find that our model can help explain 14 of the 23 anomalies we consider, in the sense that it predicts a substantially higher CAPM alpha for the extreme anomaly decile portfolio that actually has a higher alpha empirically. In particular, the model helps explain the momentum, failure probability, idiosyncratic volatility, gross profitability, expected idiosyncratic skewness, return on assets, capital gain overhang, maximum daily return, Z-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 the 14 anomalies, the typical stock in the extreme decile with the lower average return has more volatile returns, more positively skewed returns, and a lower gain overhang than the typical stock in the other extreme decile. The greater skewness and lower gain overhang of the former stock lead 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, the 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 23 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 with full knowledge of several major anomalies, while ours was not. We also find that our model can explain time-variation in anomaly alphas – variation that it attributes to changes in stocks' volatility, skewness, and gain overhang over time.

For some anomalies, most notably the size and value anomalies, our model performs poorly. For example, value stocks have more positively skewed returns and 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 also have more volatile returns than growth stocks, which, all else equal, leads investors to charge a higher average return on value stocks. Quantitatively, the first effect dominates. The model therefore predicts a lower average return on value stocks, contrary to the empirical facts. We are able to shed light on why the model sometimes performs poorly. For most of the anomalies that the model fails to explain, a large part of the anomaly return comes around earnings announcement dates. This suggests that these anomalies are driven not by the risk attitudes embedded in prospect theory, but rather by incorrect beliefs about firms' future outcomes – incorrect beliefs that are corrected by earnings announcements.

Our analysis builds on three intuitions laid out in earlier papers. The idea that, due to loss aversion, assets with more volatile returns 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 (2008a) and Baele et al. (2019), among others.

Despite these advances, the basic questions posed in our opening paragraph remain unanswered. The reason is that the three intuitions above have opposite effects on average returns. As such, to determine what prospect theory predicts for the size anomaly, the value anomaly, or any anomaly, we have to *combine* the three intuitions in a single model, something that has never been done. To do this, we need a model that incorporates all of the elements of prospect theory and also accounts for 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)). Prior models of the cross-section that do incorporate all of the elements of prospect theory, namely Barberis and Huang (2008a), Ingersoll (2014), and Barberis, Mukherjee, and Wang (2016), are all one-period models; as a result, they do not account for investors' prior gains and losses. To answer the questions we laid out at the start, a new model is needed, and we develop one in this paper. The effort pays off: our model can explain many more of the 23 anomalies than can models that omit elements of prospect theory or that ignore investors' prior gains and losses.

In summary, our paper accomplishes three things. First, by way of a new model of the cross-section, it answers the long-standing question: "What does prospect theory predict for stock market anomalies?" Second, by helping to explain a majority of 23 prominent anomalies, it offers a psychological account of multiple stock market puzzles. Finally, 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 a wide range of anomalies.

The paper is organized as follows. In Section I, we review prospect theory and narrow framing. In Section II, we present a model that incorporates these concepts and discuss the structure of the equilibrium. In Section III, we introduce the 23 anomalies that are the focus of our application and compute the empirical characteristics that serve as inputs to the model. In Section IV, we present the model's predictions for the stock market anomalies. Section V discusses additional results, while Section VI concludes. Further analysis can be found in the Internet Appendix.[2]

I. Prospect Theory and Narrow Framing

Our aim is to study asset prices in an economy in which investors have prospect theory preferences and engage in narrow framing. In this section, we review prospect theory and narrow framing in turn. Readers familiar with these concepts may prefer to go directly to Section II.

A. Prospect Theory

The original version of prospect theory is due to 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.[3]

To see how cumulative prospect theory works, consider the gamble

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

[2] The Internet Appendix is available in the online version of this article on the *Journal of Finance* website.

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

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 $i < j$, where $x_0 = 0$ so that x_{-m} through x_{-1} are losses and x_1 through x_n are gains, and where $\sum_{i=-m}^{n} p_i = 1$. For example, a 50:50 bet to win \$110 or lose \$100 is written as $(-\$100, \frac{1}{2}; \$110, \frac{1}{2})$. In the Expected Utility framework, an individual with utility function $U(\cdot)$ evaluates the gamble in (1) by computing

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

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

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

where

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

and where $v(\cdot)$ and $w(\cdot)$ are known as the value function and the probability weighting function, respectively.[4] Tversky and Kahneman (1992) propose the functional forms

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

and

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

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

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

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

The degree of loss aversion is determined by the parameter λ; a higher value of λ implies greater sensitivity to losses.

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 $(\$1000, \frac{1}{2})$ – but risk-seeking over moderate-probability losses, in that they prefer $(-\$1000, \frac{1}{2})$ to a certain loss of \$500.[5] The degree of concavity over gains and convexity over losses is governed by the parameter α; a lower value of α means greater concavity over gains and greater convexity over losses.

Finally, under cumulative prospect theory, the individual does not use objective probabilities when evaluating a gamble, but rather transformed probabilities obtained from objective probabilities via the weighting function $w(\cdot)$. 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) and (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 $\delta \in (0, 1)$, we have $w(P) > P$ for low, positive P; the right panel of Figure 1 illustrates this for $\delta = 0.4$ and $\delta = 0.65$. If p_{-m} and p_n are small, then, we have $w(p_{-m}) > p_{-m}$ and $w(p_n) > p_n$, so that the most extreme outcomes – the outcomes in the tails – are overweighted.

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

B. Narrow Framing

Traditional models, in which utility functions are defined over wealth or consumption,

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

[6] Prospect theory has impressive predictive power in experimental settings: Fudenberg et al. (2021) find that, at least for simple gambles, it attains nearly the maximum possible predictive power. However, one aspect of cumulative prospect theory – the "rank-dependent" formulation of probability weighting in (4) – has not found empirical support (Bernheim and Sprenger (2020)). Fortunately, our results do not rely on rank dependence: when we repeat our analysis with *non*-rank-dependent probability weighting, we obtain very similar findings. The intuition is that the main property of probability weighting that drives our results is the strong preference it induces for positively skewed returns. Such a preference can be generated by both rank-dependent and non-rank-dependent probability weighting.

make a clear prediction as to how an individual evaluates a new gamble he is offered: he merges the new gamble with other risks that he already faces to determine its effect on the distribution of his future wealth or consumption, and then checks if the new distribution is an improvement. However, 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 that they already face and checking if the combination is attractive, they often evaluate the new gamble to some extent in isolation, separately from their other risks. This is known as "narrow framing." Tversky and Kahneman (1981) present early laboratory evidence of this phenomenon. 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.

II. Model and Equilibrium Structure

In the introduction to the paper, we note 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, due 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, 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 of the elements of prospect theory *and* accounts for investors' prior gain or loss in each risky asset. No existing model of asset prices meets both conditions. Below, we present a new model that does. 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.

A. Model Setup

We consider an economy with three dates, $t = -1$, 0, and 1. Our focus is on investor decision-making at date 0. There is a risk-free asset with gross per-period return R_f. There are also N risky assets. The gross per-period return of risky asset i is $\widetilde{R}_i$, and the return vector $\widetilde{R} = (\widetilde{R}_1, \ldots, \widetilde{R}_N)'$ has cumulative distribution function $P(\widetilde{R})$, which we specify below. The vector of expected returns on the risky assets is $\overline{R} = (\overline{R}_1, \ldots, \overline{R}_N)'$ and the covariance matrix of returns is $\Sigma = \{\sigma_{ij}\}$.

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

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

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

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

where

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

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

The gain or loss on asset i, $\widetilde{G}_i$, is defined in (10). It is the sum of two terms. The first term, $W_0\Theta_i(\widetilde{R}_i - R_f)$, is the potential *future* gain or loss on asset i between time 0 and time 1. Specifically, it is the value of the investor's holdings of asset i at time 0 multiplied by the return on the asset in excess of the risk-free rate. For example, if the investor's holdings of asset 1 are worth \$100 at time 0, and if 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 thinks 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 for the benchmark return.

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* 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 (loss) of 30% on asset i between $t = -1$ and $t = 0$, then $g_i = 0.3$ (-0.3). Equation (10) indicates that, at time 0, the investor merges the potential future gain or loss on asset i with his prior gain or loss on the asset and computes the prospect theory value of this overall gain or loss.[7]

For simplicity, we take the second term on the right-hand side of (10) to be identical across investors. All investors in the model have the same prior gain or loss g_i in asset i, which we empirically estimate as the average gain or loss since purchase across all holders of the asset.[8] In addition, for all investors, we 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 an investor chooses; rather, the only control variable is Θ_i, the investor's allocation to asset i at time 0, which appears in the first term in (10). We have studied the effects of allowing for heterogeneity across investors in their gain or loss g_i and of endogenizing the initial allocation $\Theta_{i,-1}$. As we discuss in Section V, our conclusions are robust to these modifications. Finally, we use the approximation $W_{-1} \approx W_0$.[9]

We emphasize three aspects of the gain or loss in equation (10). First, we assume narrow,

[7]There are small differences in how past and future gains and losses are defined. For simplicity, the past gain or loss does not account for dividends or adjust for the risk-free rate. The impact on our results is minor.

[8]The model also accommodates other measures of the prior gain or loss on the asset – for example, the prior gain or loss over a fixed time interval such as the previous six months.

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

rather than broad, framing: equations (9) and (10) show that investors derive utility from asset-level gains and losses, not from portfolio gains and losses. As noted in Section I.B, there is both psychological and experimental support for this assumption.

Second, when making his decision at time 0, an investor does not segregate the prior gain or loss from the future gain or loss, deriving utility separately from each one; rather, he merges the prior and future gains and losses and derives utility from the integrated gain or loss. This is in line with mounting evidence from both experiments and betting markets 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 (2020)).

Finally, we assume that the investor derives utility from paper gains and losses rather than from realized gains and losses: if he sells shares of an asset at a gain at time 0, he does not derive utility at time 0 from the realized gain, but only at time 1 from the accumulated gain or loss on the asset.

To understand the impact of these assumptions, we have also analyzed a model with broad rather than narrow framing, a model with utility from segregated rather than integrated gains, and a model with utility from realized rather than paper gains. We discuss our findings in Section V.A. In brief, our narrow framing and integrated gain assumptions significantly improve the model's ability to explain stock market anomalies – a result that underscores the psychological realism of these assumptions. A model with utility from realized gains, while less tractable, delivers predictions similar to those of the model laid out above.

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

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

where $P(R_i)$ is the marginal cumulative distribution function of asset i's returns and where $dw(P(R_i))$ and $dw(1 - P(R_i))$ are written out in full in Section I of the Internet Appendix. 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 the degree of loss aversion λ. The bottom row corresponds to gains. We allow investors to sell short. In Section II of the Internet Appendix, we show how the expression in (11) is amended in the case of a negative allocation to asset i, $\Theta_i < 0$.

To complete the description of the decision problem, we need to specify the probability distribution $P(\widetilde{R})$ for asset returns. Since skewness plays an important role in our analysis, we need a distribution that can capture it as accurately as possible. 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 $\widetilde{R} = (\widetilde{R}_1, \ldots, \widetilde{R}_N)'$ has an N-dimensional GH skewed t distribution; we present the density function for this distribution in Section III of the Internet Appendix. 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[10]

$$
\begin{aligned}
p(R_i) &= \frac{2^{1-\frac{\nu+1}{2}}}{\Gamma(\frac{\nu}{2})(\pi\nu S_i)^{\frac{1}{2}}} \cdot \frac{K_{\frac{\nu+1}{2}}\left(\sqrt{(\nu + (R_i-\mu_i)^2/S_i)\zeta_i^2/S_i}\right)\exp((R_i-\mu_i)\zeta_i/S_i)}{\left(\sqrt{(\nu + (R_i-\mu_i)^2/S_i)\zeta_i^2/S_i}\right)^{-\frac{\nu+1}{2}}(1+(R_i-\mu_i)^2\nu^{-1}/S_i)^{\frac{\nu+1}{2}}}, \\
&\quad \text{for } \zeta_i \neq 0 \\
p(R_i) &= \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})(\pi\nu S_i)^{\frac{1}{2}}} \cdot (1+(R_i-\mu_i)^2\nu^{-1}/S_i)^{-\frac{\nu+1}{2}}, \text{ for } \zeta_i = 0, \qquad (12)
\end{aligned}
$$

where $\Gamma(\cdot)$ is the Gamma function and K_l is the modified Bessel function of the second kind with order l.[11]

The distribution in (12) has four parameters: μ_i, S_i, ζ_i, and ν. Here, μ_i, the location parameter, helps to determine the mean of the distribution; S_i, the dispersion parameter, controls the dispersion in returns; ζ_i, the asymmetry parameter, governs the skewness of returns; and ν, a degree of freedom scalar, affects the heaviness of the tails of the distribution. The mean, variance, and skewness of the distribution are

$$
\begin{aligned}
E(\widetilde{R}_i) = \overline{R}_i &= \mu_i + \frac{\nu}{\nu-2}\zeta_i \qquad (13) \\
\mathrm{Var}(\widetilde{R}_i) &= \frac{\nu}{\nu-2}S_i + \frac{2\nu^2}{(\nu-2)^2(\nu-4)}\zeta_i^2 \qquad (14) \\
\mathrm{Skew}(\widetilde{R}_i) &= \frac{2\zeta_i\sqrt{\nu(\nu-4)}}{\sqrt{S_i}(2\nu\zeta_i^2/S_i + (\nu-2)(\nu-4))^{\frac{3}{2}}}\left[3(\nu-2) + \frac{8\nu\zeta_i^2}{S_i(\nu-6)}\right], \qquad (15)
\end{aligned}
$$

[10] We use $p(\cdot)$ and $P(\cdot)$ to denote the probability density function and the cumulative distribution function, respectively.

[11] See Aas and Haff (2006), Hu and Kercheval (2010), Birge and Chavez-Bedoya (2016), and Kwak and Pirvu (2018) 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, which prevents it from being determined in equilibrium by market clearing. The skew-normal distribution cannot accommodate skewness levels higher than 0.995, which makes it a poor fit for our application.

respectively.

In Sections III and IV, using equations (14) and (15), we set S_i, ζ_i, and ν to match the empirical volatility and skewness of asset i's returns. We then search for a value of μ_i that clears the market for asset i. The asset's expected return – the quantity we want to determine – is then given by (13).[12]

The objective function in (9) combines a traditional component – 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) captures a person's "fast," intuitive sense of how to invest, while the mean-variance term corresponds to an investment strategy based on "slow," effortful reasoning. The individual's final course of action combines these intuitive and reasoned judgments.

Why do we take the traditional term in (9) to be mean-variance preferences, rather than some other Expected Utility specification? In our main application in this paper, we examine whether 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 deliver CAPM pricing, as 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.

B. Equilibrium Structure

We now 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 of asset returns. One of our contributions is to address this challenge 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

[12] 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 estimate second and third moments – the asset's beta, volatility, and skewness – and use them all to predict the expected return.

identical in their preferences, wealth, and prior gain or loss in each risky asset. It is then 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 $\mu = (\mu_1, \ldots, \mu_N)'$ such that, for this μ, the objective function in (9) has a unique global maximum $\Theta^* = (\Theta_1^*, \ldots, \Theta_N^*)'$ with $\Theta_i^* = \Theta_{M,i}$ for all i, where $\Theta_{M,i}$ is the market value of asset i divided by the total market value of all traded assets. This equilibrium structure is the one used in Expected Utility models with identical investors.

Remarkably, however, for the wide range of parameter values that we have examined, this type of equilibrium does not exist for the model in Section II.A. Later in the paper, we illustrate this nonexistence with an example. Here, we explain it in general terms. Suppose that, for some value of μ_1, the location parameter for asset 1, the objective function in (9) is maximized at $\Theta_1 = \Theta_1^*$, where Θ_1^* exceeds the asset's market supply $\Theta_{M,1}$. This suggests that, to clear the market, we simply need to lower the value of μ_1, as this will lower the asset's expected return. However, it turns out that as we do so, the value of Θ_1 at which the objective function attains its maximum can jump discontinuously from a Θ_1^* that exceeds $\Theta_{M,1}$ to a Θ_1^* that lies below it. As a result, there is *no* value of μ_1 for which the objective function is maximized at a Θ_1^* that equals $\Theta_{M,1}$. An equilibrium in which investors have identical holdings for all assets therefore does not exist.[13]

Full rationality with heterogeneous holdings. Given that the homogeneous-holdings equilibrium does not exist, what kind of equilibrium can we consider instead? One possibility involves multiple global maxima. In particular, it may be the case that there exists a location vector $\mu = (\mu_1, \ldots, \mu_N)'$ such that, for this μ, the objective function in (9) has multiple global maxima, and by allocating the appropriate fraction 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 μ; this implies 100^N possible location vectors μ. Since we are thinking of the risky assets as individual stocks, N is a large number, on the order of 1,000. 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

[13]Why does the value of Θ_1^* at which (9) is maximized jump discontinuously as we lower μ_1? When $b = 0$, the expression in (9), viewed as a function of Θ_1, depends only on Θ_1 and Θ_1^2. It therefore has a single local maximum that is also its global maximum. When $b > 0$, (9) becomes a function of Θ_1, Θ_1^2, and additional powers of Θ_1, including Θ_1^α, where $\alpha \in (0, 1)$. As such, it can have multiple local maxima. As we lower μ_1, there can come a point where the global maximum jumps from one local maximum to another. This makes Θ_1^* a discontinuous function of μ_1.

whether, by allocating investors to the various maxima, we can clear the market. This procedure is challenging even for $N = 2$ risky assets; for $N = 1,000$, the more realistic value we use below, it is not feasible.

*Bounded rationality with heterogeneous holdings***.** To overcome the difficulties described above, we introduce a mild bounded-rationality assumption, one that makes it feasible to find a heterogeneous-holdings equilibrium. Specifically, we assume that, when trying to determine the allocation Θ_i to asset i that maximizes the objective function in (9), an investor assumes that his holdings of the *other* $N-1$ risky assets equal the market supply of those assets – in other words, that $\Theta_j = \Theta_{M,j}$ for all $j \neq i$. This will not be exactly true – investors' actual portfolios will be less diversified than the market portfolio – but, as we explain below, this discrepancy 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 actually go about solving this problem.

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

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

where

$$\widehat{\gamma} = \gamma W_0 \tag{17}$$

$$\widehat{b} = bW_0^{\alpha-1}. \tag{18}$$

A bounded-rationality equilibrium with heterogeneous holdings consists of a location vector $(\mu_1, \ldots, \mu_N)'$ such that, for each i, the function in (16) has either a unique global maximum at $\Theta_i = \Theta_{M,i}$ or 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 Θ_i to asset i into a one-dimensional problem, one in which investors trade off a larger allocation to asset i and lower allocation to the risk-free asset against the opposite strategy. Since the problem is now one-dimensional, it is straightforward 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 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 for which investors have heterogeneous holdings, the function in (16) has just two global maxima, Θ_i^* and Θ_i^{**}, both of which are nonnegative. These maxima straddle the market supply $\Theta_{M,i}$, so that $\Theta_i^* < \Theta_{M,i} < \Theta_i^{**}$, which allows us to clear the market in the asset by assigning some investors to the Θ_i^* allocation and the rest to the Θ_i^{**} allocation. We also find that Θ_i^* is always much closer to $\Theta_{M,i}$ than is Θ_i^{**}. As such, to clear the market, we assign the vast majority of investors to the Θ_i^* allocation and the remaining few to the Θ_i^{**} allocation.

The two global maxima arise because investor preferences embed elements of risk-seeking – in particular, as described in Section I, risk-seeking over moderate-probability losses and over low-probability gains; if investors were instead uniformly risk-averse, the objective function would be concave and have only one global maximum. There is a simple intuition for the two optima, Θ^* and Θ^{**}. The heterogeneous holdings typically arise for assets with a positive gain overhang, which, in part because of this overhang, have a high expected return. The lower optimum Θ^* reflects investors' desire to lock in their prior gains in these assets by selling some of their holdings; the upper optimum Θ^{**} reflects investors' desire to benefit from the high expected return by increasing their allocation to these assets. In equilibrium, the two strategies are equally attractive. In Section III.B, we discuss the multiple optima in more detail by way of a numerical example.

To understand the portfolios that the investors in our economy hold, suppose that there are $N = 1,000$ 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. For any given investor, then, his allocations to the first 500 assets equal the market weights of these assets. For each of the remaining 500 assets, the investor's allocation to the asset will be lower than its market weight with very high probability – this is the Θ_i^* optimum – but, with very low probability, he will hold a large position in the asset given by the Θ_i^{**} optimum. Overall, then, the average investor combines a diversified portfolio of many assets with a small number of concentrated holdings. In Section III.A, we check that the level of underdiversification for this average investor is similar to that observed in actual household portfolios.[14]

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) to (18).

To make the model easier to implement, we rescale it. 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 θ_i rather than Θ_i. In Section IV of the Internet Appendix, we show that, when reformulated with θ_i as the choice variable, the investor's decision problem has exactly the same form as in (16), subject only to a rescaling of $\widehat{\gamma}$ and $\widehat{b}$. 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 the rescaling allows us to simplify

[14]If, for some asset i, Θ_i^* and Θ_i^{**} are such that, to clear the market, 99% of investors need to hold Θ_i^* and 1% need to hold Θ_i^{**}, then, for any given investor, he has a 99% chance of being assigned to Θ_i^* and a 1% chance of being assigned to Θ_i^{**}. We can think of this probabilistic assignment as being independent across assets for a given investor. As such, some investors will have no concentrated holdings: they will not be assigned to Θ_i^{**} for any asset. Other investors will have one concentrated holding, others two, and so on. Of course, this is just one way of assigning investors to the two maxima. The model therefore does not fully determine each investor's holdings. It simply says that, for any asset with heterogenous holdings, the vast majority of investors hold a small position in it while a small handful hold a large position in it.

the variance term in the first row of (16) by introducing asset i's beta, denoted β_i.[15]

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

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

where

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

$$\widehat{b} = bW_0^{\alpha-1}\Theta_{M,R}^{\alpha-1}, \quad (22)$$

has either a unique global maximum at $\theta_i = \theta_{M,i}$ or two global maxima that straddle $\theta_{M,i}$. In Section V of the Internet Appendix, 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 computation time for each asset; this is fast enough for the application we consider in this paper.

In Section II of the Internet Appendix, we show how the prospect theory terms in (20) are modified when an investor sells asset i short, so that $\theta_i < 0$. It turns out that, for our particular application in Sections III and IV below, no investor sells short in equilibrium: even for an asset i in which investors have heterogeneous holdings, the lower optimum θ_i^* is always greater than or equal to zero. We discuss this further in Section III.B.

III. Anomalies and Model Parameter Values

The model of Section II 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

[15]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 one for any given investor.

stock market anomalies? To be as comprehensive as possible, subject to the computational constraints we face, we consider 23 anomalies. We list these anomalies in Table I, along with the acronyms we use in subsequent tables to refer to them; in Section VI of the Internet Appendix, we define the predictor variable associated with each anomaly. The 23 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 10 anomalies drawn from Stambaugh, Yu, and Yuan (2012). We then add 12 anomalies from the 97 studied by McLean and Pontiff (2016) and one suggested by a referee. 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. In Section V.B, we check that our results are not specific to the 23 anomalies in Table I by repeating our analysis for another set of anomalies.

To see if our model can explain a particular anomaly, we proceed as follows. We consider an economy with $N = 1,000$ stocks; each anomaly decile therefore contains 100 stocks. We number the stocks so that, in the case of the value anomaly, for example, 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, 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 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.[16]

What are the empirical inputs we need to compute the expected return of a stock in our model? Equation (20) shows that, to determine μ_i, and hence stock i's expected return, we need to know σ_i, ζ_i, g_i, and β_i. In words, to compute the 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 use the value anomaly as an example; the process is the same for all of the anomalies we consider.

[16]Why do we not simply consider an economy with $N = 10$ risky assets, where each asset corresponds to one of the anomaly deciles? The reason is that the expected return that 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 $N = 1,000$ allows us to do this. Since all 100 stocks in each decile are identical, the computational burden in the case of $N = 1,000$ is the same as for $N = 10$.

Each month from July 1963 to December 2014, we rank all stocks listed on the NYSE, NASDAQ, and Amex 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. 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.[17] 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 the given 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 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.[18]

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. In the best-known data set of individual investor trading, portfolio turnover is on the order of 50% per year (Barber and Odean (2000)), which implies that an individual stock is held for roughly two years, on average.[19] 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

[17] For a given stock in a given month, we compute its beta using daily returns over the following year.

[18] For three of the anomalies – post-earnings announcement drift, failure probability, and difference of opinion – data availability requires that we begin the computation after July 1963: in 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.

[19] Barber and Odean (2000) report a median annual portfolio turnover of 32% and a mean turnover of 78%. Both numbers are informative for our purposes.

this, that is, one year.

We noted above that 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: for instance, when trying to estimate a stock's skewness, they may take a set of stocks with similar characteristics and check whether any stocks in this set have posted an extreme right-tail return. Finally, 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.[20]

Before presenting the empirical characteristics of the 230 anomaly deciles, we clarify the definition of a key variable: the capital gain overhang. There are two approaches to computing this variable, 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 the stock. We compute 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.[21]

Table II presents the results of the empirical exercise above. The first column of the table lists the 23 anomalies. 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 the definition of an anomaly, these average returns differ in a way that is not explained by beta.

[20]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; we 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, because it relies on additional assumptions about the autocorrelations of stock returns. After weighing the pros and cons of the time-series and cross-sectional approaches, we view the latter as superior.

[21]Grinblatt and Han (2005) compute the gain overhang as $(P_i - R_i)/P_i$, where P_i is stock i's current price and R_i is the average purchase price. We compute it slightly differently, as $(P_i - R_i)/R_i$; this is a more precise match for the capital gain variable g_i in our model.

The fourth and fifth columns 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 II. 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 standard deviation of the annual return of the typical stock in decile 1 is 76%, while for the typical stock in decile 10, the standard deviation is 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 22 of the 23 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 22 of the 23 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. Each asterisk corresponds to an anomaly decile; since there are 23 anomalies, this makes for a total of 230 asterisks. The horizontal and vertical axes 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 two other graphs show the negative correlation, across anomaly deciles, between the standard deviation and gain overhang, and between skewness and gain overhang.[22]

The empirical patterns in Figure 2 point to the necessity of our quantitative approach.

[22]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.

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 23 anomalies, 21 follow this pattern. Without a quantitative model, it is impossible to tell 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 II 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.

A. Parameter Values

To see if the model can capture a particular anomaly, we proceed as follows. We consider an economy with $N = 1,000$ 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 ζ_i of the skewed t distribution, the capital gain g_i, and the beta β_i to capture these empirical values. We then search for a location parameter μ_i such that the conditions for equilibrium described in Section II.B around equation (20) are satisfied. The 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 μ_i and hence the same expected return.

We now explain in more detail how we parameterize the model. While the model features several parameters, all of them are disciplined by either field data or experimental data. The asset-level parameters are R_f, the gross risk-free rate; N, the number of stocks; $\{S_i\}$, the dispersion parameters for stock returns; $\{\zeta_i\}$, the asymmetry parameters for stock returns; ν, the degree of freedom parameter; $\{\sigma_i\}$, the standard deviations of stocks' returns; $\{\beta_i\}$, the stocks' betas; $\{g_i\}$, the stocks' capital gains; σ_M, the standard deviation of stock market returns; and $\{\theta_{M,i}\}$, the stocks' market weights. The investor-level parameters are $\widehat{\gamma}$, portfolio risk aversion; $\widehat{b}$, 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.

We start with the asset-level parameters. For a given stock i that belongs to some decile, we set the standard deviation σ_i of its returns equal to the empirical standard deviation of the typical stock in that decile. 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 ν. 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,

$$\mathrm{Std}(\widetilde{R}_i) = \sigma_i = \left[\frac{\nu}{\nu-2}S_i + \frac{2\nu^2}{(\nu-2)^2(\nu-4)}\zeta_i^2\right]^{0.5} \tag{23}$$

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

To set S_i and ζ_i for a stock i that belongs to a particular anomaly decile, we take the empirical standard deviation and skewness of the typical stock in that anomaly decile and place them on the left-hand side of equations (23) and (24). These equations then allow us to solve for the two unknowns, S_i and ζ_i. For example, in the case of the size anomaly, stocks 1 to 100 belong to the lowest market capitalization decile. From Table II, 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 ζ_i for $i \in \{1, \ldots, 100\}$, we solve

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

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

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; σ_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 reported in Table II for the two extreme deciles of each anomaly. We set σ_M to 0.25 and the gross risk-free rate R_f to one.

We set $\{\theta_{M,i}\}$, the stocks' market weights, to match empirical market weights. Take, for example, the volatility anomaly. For each month of our sample, we compute the fraction of the total market value of all stocks in the 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 the model, there are 100 identical stocks in decile 1, we set $\theta_{M,i} = 0.296/100$ for all stocks in decile 1, in other words, for $i = 1, \ldots, 100$. We proceed similarly for the other deciles.

We now turn to the investor-level parameters. We set $\theta_{i,-1}$, 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; we discuss the robustness of our results to this assumption in Section V.B. We set $\widehat{\gamma}$, the scaled portfolio risk aversion in (21), and $\widehat{b}$, the scaled weight on the prospect theory term in (22), to generate an aggregate equity premium of 6%. There are many pairs $(\widehat{\gamma}, \widehat{b})$ that produce an equity premium of 6%. How do we choose one? As we increase $\widehat{b}$, we increase not only the equity premium, but also investors' degree of underdiversification. We therefore choose, from among the $(\widehat{\gamma}, \widehat{b})$ pairs that generate a 6% equity premium, a pair with a value of $\widehat{b}$ that is consistent with the degree of underdiversification in actual investor portfolios, as estimated by Calvet, Campbell, and Sodini (2007). This is the pair $(\widehat{\gamma}, \widehat{b}) = (0.6, 0.6)$.[23] For this value of $\widehat{b}$, investors put modest weight on the prospect theory term in (9): it makes up just 12% of the value of the overall objective function. In Section VII of the Internet Appendix, we also check that, for this value of $\widehat{b}$, investors have sensible attitudes to large and small gambles: their objective function satisfies the restrictions on these attitudes suggested by Barberis, Huang, and Thaler (2006) and Barberis and Huang (2008b), and also captures the experimental evidence in Kahneman and Tversky (1979) that motivates the elements of prospect theory.

Finally, we set the preference parameters α, δ, and λ. A well-known set of values for these parameters comes from Tversky and Kahneman (1992), who estimate $(\alpha, \delta, \lambda) = (0.88, 0.65, 2.25)$ for the median participant in their experiment. However, these estimates

[23]Calvet, Campbell, and Sodini (2007) find that the average Sharpe ratio of the household portfolios in their sample is 19% lower than the Sharpe ratio of a market benchmark. For $(\widehat{\gamma}, \widehat{b}) = (0.6, 0.6)$, the average Sharpe ratio of investors' portfolios in our economy is 12.5% lower than that of the market portfolio – a similar level of underdiversification.

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 them on a wide range of studies, not just one.

Tversky and Kahneman's (1992) results have led to the widespread view that the degree of loss aversion λ is approximately two. 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 λ 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 λ that are lower still – as low as $\lambda = 0.99$ for their median participant, albeit somewhat higher for those with greater cognitive ability. To reflect these findings in a conservative way, we set $\lambda = 1.5$.

Booij, van Praag, and van de Kuilen (2010) compile a list of experimental estimates of α and δ. The median estimate of δ is close to Tversky and Kahneman's (1992) estimate; we therefore maintain $\delta = 0.65$. Experimental estimates of α span a fairly wide range; most lie between 0.5 and 0.95. We set α near the midpoint of this range, at 0.7. Our preference parameter values are therefore[24]

$$(\alpha, \delta, \lambda) = (0.7, 0.65, 1.5). \tag{27}$$

B. Illustration of Equilibrium Structure

Now that we have parameterized the model, we can illustrate the equilibrium structure. We use the momentum anomaly as an example. For any stock in momentum decile 1 (i.e., any of stocks 1 to 100), investors have identical holdings; each investor holds the market supply of the stock, $\theta_{M,i}$. For example, for stock 1, there is a value of μ_1, the location parameter for stock 1, such that for $i = 1$, the function in (20) has a unique global maximum at $\theta_1 = \theta_{M,1} = 1.86 \times 10^{-4}$. Figure 3 plots the function in (20) for this value of μ_1, namely $\mu_1 = 0.0108$. The global maximum at $\theta_1^* = \theta_{M,1}$ is clearly visible.

Why does the solid line in Figure 3 have the shape that it does? In the model, investors

[24]The values of α, δ, and λ 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 α and λ that are lower than those estimated by Tversky and Kahneman (1992). In addition, even the lower level of loss aversion in (27) is enough to generate a high equity premium and nonparticipation in the stock market – two prominent applications of loss aversion in finance.

typically have identical holdings in stocks that are trading at a loss and that, in part because of this negative overhang, have a low, or even negative, expected return. As an investor increases his allocation to stock 1 from zero, utility initially rises: he is in the convex region of the prospect theory 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.

Figure 3 also shows that no investor will choose to short stock 1. As an investor decreases his allocation to the stock from zero, utility initially rises: shorting the stock has a high expected return, and, due to his prior loss, the investor is risk-seeking. However, as he further lowers his allocation, utility falls: shorting the stock exposes him to strong negative skewness, which, due to probability weighting, he finds aversive.

For any stock in momentum decile 10 (i.e., any of stocks 901 to 1,000), investors have heterogeneous holdings. For example, for stock 901, there is a value of μ_{901} such that for $i = 901$, the function in (20) has two global maxima that straddle the stock's market supply of $\theta_{M,901} = 7.26 \times 10^{-4}$. The solid line in Figure 4 plots the function in (20) for this value of μ_{901}, namely $\mu_{901} = 0.5853$. It shows the two global maxima, one at $\theta^*_{901} = 9.1 \times 10^{-5} < \theta_{M,901}$ and one at $\theta^{**}_{901} = 0.119 > \theta_{M,901}$. We can clear the market by allocating most investors to the first optimum and the rest to the second optimum.

To explain the shape of the solid line in Figure 4, it is helpful to break it into its two components: the mean-variance component, represented by the dashed line – this is the first row of equation (20) – and the prospect theory component, represented by the dash-dot line, which is the second and third rows of equation (20); the solid line is the sum of the dashed and dash-dot lines. Heterogeneous holdings typically arise for stocks that are trading at a gain and that, in part because of this positive overhang, have a high expected return. The mean-variance component has the usual quadratic form: as the investor increases his allocation from zero, it first rises due to the high expected return and then falls due to the increased volatility. By contrast, the prospect theory component initially falls sharply: the investor is in the gain region of the prospect theory value function, where he is risk averse. As he increases his allocation further, more of the stock's time 1 outcomes fall in the convex segment of the prospect theory value function to the left of the kink. This slows the rate of decline of the dash-dot line, which allows the mean-variance component to give the overall objective function a second global maximum at $\theta^{**}_{901} = 0.119$.[25]

[25] As the investor increases his allocation to stock 901 from zero, the prospect theory part of the objective

Figure 4 also shows that no investor will choose to short stock 901. Shorting the stock gives an investor a low expected return and exposes him to negative skewness. The value of the objective function therefore falls sharply as the investor lowers his allocation to the stock from zero.

The intuition above points to a simple interpretation of the two global maxima in Figure 4. The lower optimum θ^* reflects investors' desire to lock in their prior gains in the stock by selling some of their initial holdings, while the upper optimum θ^{**} reflects investors' desire to benefit from the stock's high expected return by increasing their allocation to it. In equilibrium, the two strategies are equally attractive. To generate the two maxima, it is crucial that investors exhibit an element of risk-seeking: if they were uniformly risk-averse, the objective function would be concave and have only one global maximum. In Figure 4, the risk-seeking stems primarily from the convex segment of the prospect theory value function; this leads the dash-dot line to fall more slowly as the investor increases his allocation, which allows the objective function – the solid line – to have a second global maximum.

The value of the objective function at the two global maxima in Figure 4 is, by definition, the same. One may worry that this represents a "knife-edge" situation that arises only for particular values of the preference parameters. This is not the case. For a wide range of preference parameter values, if investors do not have identical holdings in a stock, we can always find a heterogeneous-holdings structure like the one in Figure 4. In each case, the stock's expected return adjusts until the two local maxima have the same height. The intuition we give above for the shape of the solid line in Figure 4 helps explain why this structure arises as commonly as it does.

In Section VIII of the Internet Appendix, we discuss additional aspects of the equilibrium structure. We explain why, for stocks in momentum decile 10, there is no equilibrium with identical holdings; and why, for stocks in momentum decile 1, there is no equilibrium with heterogeneous holdings. We also contrast the heterogeneous holdings that arise in our setting with those that arise in other models, such as that of Barberis and Huang (2008a).

IV. Application

We now use our model to answer a long-standing question: Can the risk attitudes cap-

function – the dash-dot line – 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 θ^* 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 initial rise.

tured by prospect theory shed light on the prominent anomalies in Table I?

To determine whether the model can help explain an anomaly, we focus on anomaly alphas. For any given anomaly, we compute the empirical alphas for the 10 anomaly deciles over our 1963 to 2015 sample – these are value-weighted CAPM alphas computed with monthly data and annualized – and denote them by $\alpha^d(1), \ldots, \alpha^d(10)$, where the "$d$" superscript stands for "data." We then compute the alphas predicted by the model for each of the 10 deciles, namely $\alpha^m(1), \ldots, \alpha^m(10)$, where the "$m$" 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 $100l$:[26]

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

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

We say that the model can help explain an anomaly if

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

The first condition in (28) states that the model correctly predicts the sign of the difference between $\alpha^d(10)$ and $\alpha^d(1)$; in other words, it 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) states that the model makes a "strong" prediction, in other words, predicts a substantial difference between the two alphas; while the 1.5% cutoff is somewhat arbitrary, it allows for a simple way of organizing our results. Similarly, we say that the model fails to explain an anomaly if

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

in other words, if the model makes a strong prediction but the prediction is incorrect – for example, if it predicts 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

$$|\alpha^m(10) - \alpha^m(1)| < 0.015. \tag{30}$$

We find that the model is helpful for thinking about a strikingly large number – a majority,

[26]We use the notation α both for anomaly alphas and for the degree of diminishing sensitivity in prospect theory. The context will make clear which of the two we are referring to.

in fact – of the anomalies we consider; we review these anomalies in Section IV.A. In Section IV.B, we discuss the anomalies for which the model performs poorly. In Section IV.C, we note the anomalies for which the model does not make a strong prediction. Finally, in Section IV.D, we compare our model's performance to that of several widely used factor models.

A. Anomalies for which the Model Performs Well

We present the model's predictions in Table III. The horizontal lines in the table divide the 23 anomalies into three groups. The upper group corresponds to anomalies for which the model performs well; for these anomalies, the conditions in (28) are satisfied. The middle group corresponds to anomalies for which the model does not make a strong prediction; for these anomalies, the condition in (30) holds. Finally, the lower group corresponds to anomalies for which the model performs poorly; for these anomalies, conditions (29) are satisfied. For each anomaly, the table reports the model's predictions for the alphas of deciles 1 and 10, $\alpha^m(1)$ and $\alpha^m(10)$; the model-predicted alpha spread $\alpha^m(10) - \alpha^m(1)$ between the two deciles; and the empirical alpha spread, $\alpha^d(10) - \alpha^d(1)$.

The table shows that the model is helpful for thinking about 14 of the 23 anomalies – specifically, the momentum, gain overhang, failure probability, return on assets, idiosyncratic volatility, maximum daily return, idiosyncratic skewness, Z-Score, external finance, gross profitability, post-earnings announcement drift, composite equity issuance, net stock issuance, and difference of opinion anomalies.

In Figures 5 and 6, we plot results for eight of these 14 anomalies. The top-left graph in Figure 5 shows the results for the momentum anomaly. The horizontal axis corresponds to deciles 1 through 10. The dashed line plots the empirical alpha of each decile, while the solid line plots the alphas predicted by the model. The other graphs in Figures 5 and 6 have the same structure. Table III and Figures 5 and 6 show that, for many of the 14 anomalies, the model explains a large fraction of the spread in empirical alphas between anomaly deciles 1 and 10.[27]

The intuition for why the model helps to explain the 14 anomalies is as follows. 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

[27]For six of the 14 anomalies, the model explains a sufficiently large fraction of the empirical spread that the difference between the empirical and model-predicted spreads is not statistically significant. These are the gain overhang, idiosyncratic skewness, failure probability, Z-Score, gross profitability, and return on assets anomalies.

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 lead 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.[28]

Figures 5 and 6 show that the model can explain the alphas not only for the extreme deciles, but also for the intermediate deciles. 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 other anomalies as well. The solid lines in the graphs show, strikingly, that the 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 returns 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 such as the one we develop here 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 prob-

[28]The GH skewed t distribution has unbounded support and therefore assigns a small probability mass to gross returns below zero. Since this is more pronounced for stocks with volatile returns, it makes our results all the more striking: the investors in the model are willing to accept a low average return on the typical stock in decile 10 of the volatility anomaly even though this stock exposes them to unpleasant left-tail outcomes. Consistent with this, if we repeat the analysis with a GH skewed t distribution that is truncated at the point of zero gross return, we obtain an even larger predicted alpha spread for the volatility anomaly.

ability weighting, leads investors to require a lower average return on them. Our analysis confirms this mechanism, but also indicates that the argument is incomplete. The stocks in decile 10 for these anomalies are not only highly skewed; they also have volatile returns and 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.[29]

B. Anomalies for which the Model Performs Poorly

For seven of the 23 anomalies – the seven anomalies in the bottom group in Table III, namely, the size, value, long-term reversal, short-term reversal, accrual, asset growth, and investment 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. We discuss two of these anomalies – the size and value anomalies – in more detail, and the others more briefly.

Table III shows that the model incorrectly predicts a negative alpha for decile 1 of the size anomaly, which contains the stocks with the lowest market capitalization. These stocks tend to have very volatile returns, which, all else equal, leads the investors in our economy to charge a high average return on them. However, they also tend to have positively skewed returns and to trade at a loss, which leads investors to charge a low average return on them. Our quantitative analysis shows that the second effect dominates the first; as a result, the model makes an incorrect prediction.

Table III shows that the model also fails to explain the value anomaly. The reason is that value stocks typically 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 also tend to have more volatile returns than growth stocks, which, all else equal, leads investors to charge a higher average return on value stocks. However, the second effect is dominated by the first, and so the model makes the wrong prediction. Table III also reports results for five other anomalies that the model performs poorly on – the long-term

[29]Wang, Yan, and Yu (2017) propose that, in a prospect theory framework, higher return volatility can lead to a *lower* average return for an asset trading at a loss because investors are in the convex region of the value function. This mechanism does not play a role in our results. The reason is that, as shown in the top-right graph of Figure 2, for almost all of the 230 stocks we are pricing, the stock's return volatility exceeds the absolute value of its gain overhang, usually by a wide margin. As such, the investors in these stocks are still quite close to the kink in the value function, and they price the stocks accordingly: higher return volatility interacts with the kink to generate a *higher* average return.

reversal, short-term reversal, accrual, asset growth, and investment anomalies.

Why does the model perform poorly for some anomalies? One possibility is that, for these anomalies, any effect that prospect theory risk attitudes have on prices is swamped by other factors. In Section IX of the Internet Appendix, we present evidence in support of this view. We show that, for five of the seven anomalies that prospect theory does not explain – the value, investment, long-term reversal, accrual, and asset growth anomalies – a large fraction of their return comes around earnings announcement dates. This suggests that these anomalies are driven to a significant extent by investors' incorrect beliefs about firms' future prospects – incorrect beliefs that are corrected around earnings announcements – and not by prospect theory risk attitudes.[30] This, in turn, raises the possibility that many anomalies can be put into one of two categories: "preference-based" anomalies that are driven by risk attitudes of the kind captured by prospect theory, and "belief-based" anomalies that are driven by incorrect beliefs.[31]

Another approach to understanding the model's poor performance for some anomalies is also related to investor beliefs, albeit within the context of the model. When generating the results in Table III and Figures 5 and 6, we assume 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. For example, when generating the model's predictions for the value anomaly, we assume that investors know that value stocks have more positively skewed returns than growth stocks, which leads the model to predict a lower average return on value stocks. However, if investors think that growth stocks have the more positively skewed returns, 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. In Section X of the Internet Appendix, we examine how large the error in beliefs about skewness has to be for the model to predict a substantially positive value premium in line with the empirical facts.[32]

C. Anomalies for which the Model Does Not Make a Strong Prediction

[30]An implicit assumption in this argument – one that is widely adopted in prior research – is that earnings announcements serve to correct investors' biased forecasts. An alternative view is that these announcements may sometimes amplify investor biases.

[31]A sixth anomaly that the model performs poorly on, the short-term reversal anomaly, is widely thought to be driven by liquidity considerations. This may explain why the model does not capture it.

[32]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. This may lead investors to overestimate the skewness of growth stocks and underestimate the skewness of value stocks.

For two other anomalies – the middle group in Table III, namely, the net operating assets and organizational capital anomalies – the model 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.

D. Model Performance

The results presented so far show that our model is helpful for thinking about a majority of the 23 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 Fama-French five-factor model, and a six-factor model that augments the five-factor model with a momentum factor. We compute the models' average absolute pricing errors for the 23 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

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

where $\overline{R}(l)$ 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, for instance, 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 IV reports the average absolute monthly pricing error across the 23 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 has similarly good performance 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 independent of any information about the anomalies: prospect theory itself was designed with no 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.

V. Discussion

We now discuss some additional issues raised by the analysis in Section IV.

A. The Impact of Each Model Ingredient

Which of the three elements of prospect theory – loss aversion, diminishing sensitivity, or probability weighting – is most useful for explaining stock market anomalies? To address this question, for each of the 14 anomalies that the model can help explain, we decompose the model-predicted alpha spread between deciles 1 and 10 into three parts that correspond to the three elements of prospect theory.

We start by computing the model-predicted alpha spreads for the 14 anomalies when investors exhibit loss aversion, but not diminishing sensitivity or probability weighting, so that the preference parameter values are $(\alpha, \delta, \lambda) = (1, 1, 1.5)$. We then compute the predicted alpha spreads when investors exhibit loss aversion and probability weighting, but not diminishing sensitivity; this corresponds to $(\alpha, \delta, \lambda) = (1, 0.65, 1.5)$. With these in hand, we can decompose the full model's prediction for the alpha spread between deciles 1 and 10 for a given anomaly, $\alpha^m(10) - \alpha^m(1)$, as

$$\begin{aligned} \alpha^m(10) - \alpha^m(1) \quad = \quad & \alpha^{LA}(10) - \alpha^{LA}(1) \\ & +(\alpha^{LA,PW}(10) - \alpha^{LA,PW}(1)) - (\alpha^{LA}(10) - \alpha^{LA}(1)) \\ & +(\alpha^m(10) - \alpha^m(1)) - (\alpha^{LA,PW}(10) - \alpha^{LA,PW}(1)), \end{aligned} \tag{31}$$

where $\alpha^{LA}(j)$ and $\alpha^{LA,PW}(j)$ are the alpha predicted by the model for decile j when investors exhibit only loss aversion, and when they exhibit only loss aversion and probability weighting, respectively. The first row on the right-hand side of equation (31) is the component of the alpha spread generated by loss aversion alone; the second and third rows are the marginal contributions to the alpha spread of probability weighting and diminishing sensitivity, respectively.

For each anomaly, we compute the fraction of the spread generated by each of the three elements of prospect theory and then average these fractions across the 14 anomalies. The fractions of the spread that come from loss aversion, probability weighting, and diminishing sensitivity are -21%, 48%, and 73%, respectively; the negative fraction for loss aversion means that, taken alone, loss aversion typically predicts an alpha spread of the opposite sign from the empirical alpha spread.[33]

These numbers indicate the importance of diminishing sensitivity. However, they understate the role of probability weighting. In addition to the analysis above, we also computed the alpha spreads predicted by the model when investors exhibit only loss aversion and diminishing sensitivity. In this case, the model can explain just eight of the 23 anomalies; it is only when we also add probability weighting that the model can explain 14 anomalies. As such, probability weighting almost doubles the number of anomalies we can explain. The way to reconcile this result with the numbers that suggest an outsize role for diminishing sensitivity is to note that it is probability weighting that allows diminishing sensitivity to have a large effect. Since the typical stock in any anomaly decile has positively skewed returns, probability weighting lowers the expected return on all deciles. This means that, for a stock with a prior loss at time 0, many of its potential time 1 gains and losses are likely to remain in the convex region of the value function where investors are risk-seeking. This lowers the stock's expected return even further, allowing the model to explain the very low average returns of several anomaly deciles, such as decile 1 of the momentum anomaly and decile 10 of the idiosyncratic volatility anomaly.

Given that diminishing sensitivity and probability weighting play a major role in the model's successful predictions, it is useful to know how sensitive these predictions are to the values of the associated parameters, α and δ. To shed light on this, we pick a representative anomaly that the model can help explain – the idiosyncratic volatility anomaly – and compute the model's predicted alpha spread for this anomaly for different values of the diminishing-sensitivity parameter α, while keeping the other preference parameter values fixed at $\lambda = 1.5$ and $\delta = 0.65$. Similarly, we compute the predicted alpha spread for different values of the probability-weighting parameter δ, while maintaining $\alpha = 0.7$ and $\lambda = 1.5$. We present the results in Figure IA.1 in the Internet Appendix. The figure shows that, by the criterion in (28), the model can help explain the volatility anomaly for a sizable range of values of the preference parameters.

Our model assumes that, when making their allocation decision at time 0, investors engage in narrow framing: they derive utility from stock-level gains and losses. It also assumes that investors merge their prior gain or loss in a stock with their potential future gain or loss in the stock and derive utility from the integrated gain or loss. Both of these assumptions have experimental support. They also significantly improve the model's ability to explain the stock market anomalies.

To demonstrate this, we study a model in which, at time 0, investors consider only the

[33] The decomposition for all 14 anomalies is reported in Table IA.I in the Internet Appendix.

potential future gain or loss on a stock, without merging it with their prior gains; this corresponds to setting $g_i = 0$ for each stock i in equations (10) and (20). We find that, by the criterion in (28), this specification can explain only five of the 23 anomalies. Under the same bounded-rationality assumption that we adopt in Section II, we have also studied a model with broad, rather than narrow, framing. This model also explains fewer anomalies than the model with narrow framing.

What is the intuition for these results? A mechanism that helps our model explain a majority of the anomalies in Table I is the one studied by Grinblatt and Han (2005) and Li and Yang (2013), whereby a stock with a prior gain brings its investors into the concave region of the prospect theory value function where they are risk-averse and require a high average return. This mechanism relies on narrow framing and on investors taking their prior gains and losses into account. Without these assumptions, the mechanism loses its force and the model's explanatory power is weaker.

We have also studied a model in which investors derive utility from realized gains and losses rather than from paper gains and losses. While less tractable, this model delivers similar predictions to the model of Section II. The reason is that it features a mechanism similar to the one described in the previous paragraph. At time 0, investors are keen to sell an asset trading at a gain because they want to enjoy positive realization utility. To clear the market, the stock must therefore have a high expected return. Conversely, at time 0, investors are reluctant to sell a stock with a prior loss because they want to avoid negative realization utility. They are therefore willing to hold the market supply even if the stock has a low expected return.

B. Initial Holdings and An Alternative Set of Anomalies

We now examine the robustness of our results along two dimensions: to endogenizing investors' initial holdings, and to considering an alternative set of anomalies.[34]

When computing the expected return of a stock i, we take each investor's initial position in the stock at time -1, $\theta_{i,-1}$, to equal the stock's market supply $\theta_{M,i}$ estimated from the data. We can instead endogenize the initial holdings. Specifically, we take the objective function in (20) and set $g_i = 0$; this captures the fact that, when an investor initially buys stock i, his prior gain is zero. We look for a value of the location parameter μ_i that clears the market at time -1 and take investors' initial holdings of stock i to be the maxima of

[34]In Section XI of the Internet Appendix, we also study the robustness of our conclusions to allowing for heterogeneity across investors in their prior gains and losses on a stock.

the objective function for this μ_i. We then use these initial holdings as inputs to the time 0 pricing problem. While this analysis is less tractable, it leads to similar conclusions: using the criterion in (28), the model again helps to explain 14 of the 23 anomalies, namely the 14 anomalies in the top group in Table III; and it again produces an average absolute pricing error similar to that of the Carhart four-factor model. We provide more details, along with an example, in Section XII of the Internet Appendix.

We chose the 23 anomalies in Table I because they have received a lot of attention in prior research. However, we want to be sure that the conclusion we draw in Section IV – that prospect theory is helpful for thinking about stock market anomalies – is not particular to this set of anomalies. After completing the analysis in Section IV, we came across another set of 23 anomalies, one constructed by Novy-Marx and Velikov (2016) as part of a study of transaction costs, in other words, for reasons that have nothing to do with prospect theory. We use our model to generate predictions about this alternative set of 23 anomalies and report the results in Section XIII of the Internet Appendix. We find that, by the criterion in (28), the model can help explain 13 of the 23 Novy-Marx and Velikov (2016) anomalies. This is a very similar proportion to that reported in Section IV, where the model explained 14 of the 23 anomalies in Table I. This finding reinforces our conclusion that prospect theory is helpful for understanding stock market anomalies.

C. Comparison With Other Prospect Theory Models

Our model accounts for all of the elements of prospect theory – loss aversion, diminishing sensitivity, and probability weighting – as well as for investors' prior gains and losses. In this section, we compare our model to other prospect theory models. We find that our model can explain almost all of the facts captured by prior models, but that the converse is not true: the earlier models explain substantially fewer of the anomalies in Table I than our model does.

We classify the earlier prospect theory models into three groups and consider each group in turn.[35] Barberis, Huang, and Santos (2001) and Barberis and Huang (2001) incorporate only the loss aversion element of prospect theory and show that this predicts a high aggregate equity premium. Our model also delivers a high equity premium, and for similar reasons: due to loss aversion, investors find the volatility of stock returns unappealing and require a high average return as compensation. Barberis, Huang, and Santos (2001) and Barberis and Huang (2001) explain two facts that our model does not: high stock market volatility

[35]Table IA.II in the Internet Appendix lists several prospect theory models of investor behavior along with the assumptions that each model makes.

and the value premium in the cross-section. However, they do so only by introducing an assumption that lies outside of prospect theory, namely that investors' loss aversion varies over time. Our framework can accommodate such an assumption, and doing so may allow it to explain a wider set of phenomena, but we do not pursue this approach here.

Grinblatt and Han (2005), Barberis and Xiong (2009), and Li and Yang (2013) consider models that account for loss aversion and diminishing sensitivity, and also for investors' prior gains and losses. Together, these papers show that their assumptions can generate a disposition effect, momentum in the cross-section, and a positive relationship between volume and return. Our model also delivers these predictions: as Figure 5 shows, the model captures momentum; and consistent with the disposition effect and the positive volume-return correlation, the investors in our model sell stocks at time 0 only in the case of heterogeneous holdings, a case that arises primarily for stocks with prior gains.

Finally, Barberis and Huang (2008a), Ingersoll (2014), and Barberis, Mukherjee, and Wang (2016) consider models that account for all of the elements of prospect theory but not for investors' prior gains and losses. Together, these models predict that an asset's idiosyncratic skewness will be priced and that some investors will hold undiversified portfolios. Our model also makes these predictions.

While our model can explain most of the empirical facts captured by previous models, the converse is not true: the prior models explain far fewer of the 23 anomalies in Table I than our model does. To be clear, none of the previous papers makes explicit predictions, either qualitative or quantitative, about a wide range of anomalies. However, we can use our model to see what they would predict. A model that, like Barberis, Huang, and Santos (2001) or Barberis and Huang (2001), features loss aversion but not diminishing sensitivity or probability weighting – this corresponds to preference parameter values $(\alpha, \delta, \lambda) = (1, 1, 1.5)$ – explains only three of the 23 anomalies by the criterion in (28). A model that, like Grinblatt and Han (2005), Barberis and Xiong (2009), or Li and Yang (2013), features loss aversion and diminishing sensitivity but not probability weighting – this corresponds to parameter values $(\alpha, \delta, \lambda) = (0.7, 1, 1.5)$ – can explain just eight of the 23 anomalies. Finally, as noted in Section V.A, a model that, like Barberis and Huang (2008a), Ingersoll (2014), or Barberis, Mukherjee, and Wang (2016), captures all of the elements of prospect theory but does not account for prior gains and losses – this corresponds to setting $g_i = 0$ in equation (10) for all i – can explain only five of the 23 anomalies. It is only when we incorporate all of the elements of prospect theory *and* account for prior gains and losses that we can explain a full 14 of the 23 anomalies.

D. Time-Variation in Alphas and Interaction Effects

In Section IV, we use our model to explain anomaly alphas in the full sample from 1963 to 2015. Can the model also explain variation in these alphas over time? In Section XIV of the Internet Appendix, we show that it can. We divide the full sample into four equal subperiods and use the procedure of Section III to compute, separately for each subperiod, the return volatility, return skewness, and gain overhang of the typical stock in each anomaly decile. For each of the four subperiods, we use these inputs to generate model-predicted alphas for each anomaly decile, which we then compare to the empirical alphas in that subperiod. We find that the model is able to explain a significant fraction of the time-variation in the alphas – variation that it attributes to changes in stocks' volatility, skewness, and gain overhang from one subperiod to the next.

In Section XV of the Internet Appendix, we consider another aspect of time-variation in anomaly alphas. We show that the anomalies for which the prospect theory model performs well have better out-of-sample performance – better performance in the years after the publication of the journal articles that first documented them – than the anomalies for which the model performs poorly. One interpretation of this result is that the anomalies for which the model performs well have a stronger theoretical foundation, one rooted in prospect theory, which in turn explains their better out-of-sample performance.

The model can also help explain three interaction effects related to the anomalies in Table I. Using double sorts, Wang, Yan, and Yu (2017) show that for stocks with a large negative overhang, there is a negative relationship between idiosyncratic volatility and subsequent returns, but that for stocks with a large positive overhang, this relationship turns positive. Similarly, An et al. (2020) show that for stocks with a large negative overhang, there is a negative relationship between expected idiosyncratic skewness and subsequent returns, but that this relationship is positive for stocks with a large positive overhang. Finally, Frazzini (2006) shows that a strategy that buys stocks with positive earnings surprises performs particularly well for stocks with a large positive overhang, and that a strategy that buys stocks with negative earnings surprises performs particularly poorly for stocks with a large negative overhang.

All three sets of authors suggest, based on informal reasoning, that the risk attitudes captured by prospect theory may be driving their results. In Section XVI of the Internet Appendix, we provide formal support for this claim: we show that our model can help explain all three interaction patterns.

E. Prospect Theory and Real-World Investors

The experimental evidence that motivates the elements of prospect theory comes primarily from studies of ordinary individuals. This raises the following question: Given that, during our sample period, individual investors are a minority of the trader population, how large an effect on asset prices can we expect prospect theory risk attitudes to have? We make three observations about this.

First, even if it is primarily individual investors who have prospect theory preferences, these preferences can still have a significant impact on asset prices. This will be the case if, despite being a minority, individuals trade in a correlated way, according to prospect theory, and thereby exert demand pressure on stocks, and if institutional investors are not able to fully absorb this demand pressure. There is reason to think that this will be the case because of the type of mispricing that prospect theory generates, namely, overpricing of stocks with volatile and skewed returns. This pattern can be seen in all of the graphs in Figures 5 and 6. For example, Figure 5 shows that, in the case of the volatility anomaly, the most severe mispricing is the overpricing of the volatile, skewed stocks in deciles 9 and 10. To correct this mispricing, institutional investors would need to short a large number of highly volatile, highly skewed, small-cap stocks. This strategy is risky and costly, which will limit their ability to absorb individual investor demand. Individual investors can therefore have a substantial impact on asset prices.

Second, several studies suggest that at least some institutional investors are likely to have prospect theory preferences. Haigh and List (2005) and Abdellaoui, Bleichrodt, and Kammoun (2013) take some of the experiments that have been used to document the elements of prospect theory for the general population and repeat them for financial professionals – for 54 traders in the first study and 46 investment managers in the second. Haigh and List (2005) test specifically for loss aversion and find support for it; Abdellaoui, Bleichrodt, and Kammoun (2013) test for all of loss aversion, diminishing sensitivity, and probability weighting and find support for all three. In addition, Coval and Shumway (2005) find strong evidence of diminishing sensitivity in the behavior of futures traders, while Larson, List, and Metcalfe (2016) find evidence of loss aversion in the trading of 151 financial professionals. Moreover, option prices – assets that are traded by more sophisticated investors – exhibit price patterns consistent with probability weighting (Polkovnichenko and Zhao (2013), Boyer and Vorkink (2014), Baele et al. (2019)).

Other studies, even if less directly related, also bolster the view that some professional investors exhibit the elements of prospect theory in their decision-making. Kachelmeier and

Shehata (1992) and Post et al. (2008) find evidence of prospect theory even in situations with high financial stakes, while Pope and Schweitzer (2011) document loss aversion in the behavior of highly trained professional athletes in a competitive, high-stakes environment.

Finally, for the value of $\widehat{b}$ that we assume in our calculations – recall that $\widehat{b}$ is the effective weight on the prospect theory terms in (9) – these terms make up only 12% of the value of the overall objective function. As such, even if investors put only modest weight on prospect theory in their decision-making, the effect on asset prices can be substantial.

VI. 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 23 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.

REFERENCES

Aas, Kjersti, and Ingrid H. Haff, 2006, The generalized hyperbolic skew student's t-distribution, *Journal of Financial Econometrics* 4, 275-309.

Abdellaoui, Mohammed, Han Bleichrodt, and Hilda Kammoun, 2013, Do financial professionals behave according to prospect theory? An experimental study, *Theory and Decision* 74, 411-429.

An, Li, Huijun Wang, Jian Wang, and Jianfeng Yu, 2020, Lottery-related anomalies: The role of reference-dependent preferences, *Management Science* 66, 473-501.

Andrikogiannopoulou, Angie, and Filippos Papakonstantinou, 2020, History-dependent risk preferences: Evidence from individual choices and implications for the disposition effect, *Review of Financial Studies* 33, 3674-3718.

Baele, Lieven, Joost Driessen, Sebastian Ebert, Juan Londono, and Oliver Spalt, 2019, Cumulative prospect theory, option returns, and the variance premium, *Review of Financial Studies* 32, 3667-3723.

Barber, Brad, and Terrance Odean, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, *Journal of Finance* 55, 773-806.

Barberis, Nicholas, and Ming Huang, 2001, Mental accounting, loss aversion, and individual stock returns, *Journal of Finance* 56, 1247-1292.

Barberis, Nicholas, and Ming Huang, 2008a, Stocks as lotteries: The implications of probability weighting for security prices, *American Economic Review* 98, 2066-2100.

Barberis, Nicholas, and Ming Huang, 2008b, The loss aversion / narrow framing approach to the equity premium puzzle, in Rajnish Mehra, ed.: *Handbook of the Equity Risk Premium* (Elsevier).

Barberis, Nicholas, Ming Huang, and Tano Santos, 2001, Prospect theory and asset prices, *Quarterly Journal of Economics* 116, 1-53.

Barberis, Nicholas, Ming Huang, and Richard Thaler, 2006, Individual preferences, monetary gambles, and stock market participation: A case for narrow framing, *American Economic Review* 96, 1069-1090.

Barberis, Nicholas, Abhiroop Mukherjee, and Baolian Wang, 2016, Prospect theory and stock returns: An empirical test, *Review of Financial Studies* 29, 3068-3107.

Barberis, Nicholas, and Wei Xiong, 2009, What drives the disposition effect? An analysis of a long-standing preference-based explanation, *Journal of Finance* 64, 751-784.

Bernheim, Douglas, and Charles Sprenger, 2020, On the empirical validity of cumulative prospect theory: Experimental evidence of rank-independent probability weighting, *Econometrica* 88, 1363-1409.

Birge, John, and Luis Chavez-Bedoya, 2016, Portfolio optimization under a generalized hyperbolic skewed t distribution and exponential utility, *Quantitative Finance* 16, 1019-1036.

Booij, Adam, Bernard Van Praag, and Gijs Van de Kuilen, 2010, A parametric analysis of prospect theory's functionals for the general population, *Theory and Decision* 68, 115-148.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2013, Salience and asset prices, *American Economic Review Papers and Proceedings* 103, 623-628.

Boyer, Brian, Todd Mitton, and Keith Vorkink, 2010, Expected idiosyncratic skewness, *Review of Financial Studies* 23, 169-202.

Boyer, Brian, and Keith Vorkink, 2014, Stock options as lotteries, *Journal of Finance* 69, 1485-1527.

Calvet, Laurent, John Y. Campbell, and Paolo Sodini, 2007, Down or out: Assessing the welfare costs of household investment mistakes, *Journal of Political Economy* 115, 707-747.

Campbell, John Y., Jens Hilscher, and Jan Szilagyi, 2008, In search of distress risk, *Journal of Finance* 63, 2899-2939.

Chapman, Jonathan, Erik Snowberg, Stephanie Wang, and Colin Camerer, 2018, Loss attitudes in the U.S. population: Evidence from dynamically optimized sequential experimentation, Working paper, California Institute of Technology.

Conrad, Jennifer, Nishad Kapadia, and Yuhang Xing, 2014, Death and jackpot: Why do individual investors hold overpriced stocks? *Journal of Financial Economics* 113, 455-475.

Coval, Joshua, and Tyler Shumway, 2005, Do behavioral biases affect prices? *Journal of Finance* 60, 1-34.

Frazzini, Andrea, 2006, The disposition effect and underreaction to news, *Journal of Finance* 61, 2017-2046.

Fudenberg, Drew, Jon Kleinberg, Annie Liang, and Sendhil Mullainathan, 2021, Measuring the completeness of economic models, Working paper, Massachusetts Institute of Technology.

Grinblatt, Mark, and Bin Han, 2005, Prospect theory, mental accounting, and momentum, *Journal of Financial Economics* 78, 311-339.

Haigh, Michael, and John List, 2005, Do professional traders exhibit myopic loss aversion? An experimental analysis, *Journal of Finance* 60, 523-534.

Hu, Wenbo, and Alec Kercheval, 2010, Portfolio optimization for student t and skewed t returns, *Quantitative Finance* 10, 91-105.

Imas, Alex, 2016, The realization effect, *American Economic Review* 106, 2086-2109.

Ingersoll, Jonathan, 2014, Cumulative prospect theory, aggregation, and pricing, *Critical Finance Review* 4, 1-55.

Kachelmeier, Steven, and Mohamed Shehata, 1992, Examining risk preferences under high monetary incentives: Experimental evidence from the People's Republic of China, *American Economic Review* 82, 1120-1141.

Kahneman, Daniel, 2013, *Thinking, Fast and Slow* (Farrar, Straus, and Giroux, New York, NY).

Kahneman, Daniel, and Amos Tversky, 1979, Prospect theory: An analysis of decision under risk, *Econometrica* 47, 263-291.

Koszegi, Botond, and Matthew Rabin, 2006, A model of reference-dependent preferences, *Quarterly Journal of Economics* 121, 1133-1165.

Kwak, Minsuk, and Traian Pirvu, 2018, Cumulative prospect theory with generalized hyperbolic skewed t distribution, *SIAM Journal on Financial Mathematics* 9, 54-89.

Larson, Francis, John List, and Robert Metcalfe, 2016, Can myopic loss aversion explain the equity premium puzzle? Evidence from a natural field experiment with professional traders, NBER Working paper 22605.

Li, Yan, and Liyan Yang, 2013, Prospect theory, the disposition effect, and asset prices, *Journal of Financial Economics* 107, 715-739.

McLean, David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability? *Journal of Finance* 71, 5-32.

Novy-Marx, Robert, and Mihail Velikov, 2016, A taxonomy of anomalies and their trading costs, *Review of Financial Studies* 29, 104-147.

Polkovnichenko, Valery, and Feng Zhao, 2013, Probability weighting functions implied in option prices, *Journal of Financial Economics* 107, 580-609.

Pope, Devin, and Maurice Schweitzer, 2011, Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes, *American Economic Review* 101, 129-157.

Post, Thierry, Martijn van den Assem, Guido Baltussen, and Richard Thaler, 2008, Deal or no deal? Decision-making under risk in a large-payoff game show, *American Economic Review* 98, 38-71.

Stambaugh, Robert, Jianfeng Yu, and Yu Yuan, 2012, The short of it: Investor sentiment and anomalies, *Journal of Financial Economics* 104, 288-302.

Thaler, Richard, and Eric 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, Amos, and Daniel Kahneman, 1981, The framing of decisions and the psychology of choice, *Science* 211, 453-458.

Tversky, Amos, and Daniel Kahneman, 1992, Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and Uncertainty* 5, 297-323.

Walasek, Lukasz, Timothy Mullett, and Neil Stewart, 2018, A meta-analysis of loss aversion in risky contexts, Working paper, University of Warwick.

Wang, Huijun, Jinghua Yan, and Jianfeng Yu, 2017, Reference-dependent preferences and the risk-return tradeoff, *Journal of Financial Economics* 123, 395-414.

Table I
Stock Market Anomalies

The table lists 23 prominent anomalies that we try to explain using a model in which investors evaluate risk according to prospect theory. The second column lists the acronyms that we use in subsequent tables to refer to the anomalies. The predictor variable associated with each anomaly is defined in Section VI of the Internet Appendix.

Anomaly	Acronym
Idiosyncratic volatility	IVOL
Market capitalization	SIZE
Value	VAL
Expected idiosyncratic skewness	EISKEW
Momentum	MOM
Failure probability	FPROB
Z-Score	ZSC
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
Capital gain overhang	CGO

Table II
Empirical Properties of Anomaly Deciles

The first column lists 23 anomalies; the acronyms are defined in Table I. The remaining columns report, for each anomaly, the average annual return (in percent), standard deviation of annual returns (in percent), skewness of annual returns, and capital gain overhang (in percent) of the typical stock in anomaly decile 1 and anomaly decile 10. The sample period is 1963 to 2015 except for FPROB, DOP, and PEAD, for which, as described in the text, the sample period is shorter due to data availability.

Anomaly	Average Return Decile 1	Average Return Decile 10	Standard Deviation Decile 1	Standard Deviation Decile 10	Skewness Decile 1	Skewness Decile 10	Gain Overhang Decile 1	Gain Overhang Decile 10
IVOL	11.9	–3.2	36.8	94.3	2.47	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	–2.8	20.9	91.8	63.1	3.81	2.39	–45.0	31.0
FPROB	17.5	0.1	51.5	110.4	2.31	3.9	28.2	–46.2
ZSC	3.8	13.9	84.5	51.6	3.51	2.56	–20.6	6.1
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	10.2	16.3	54.5	77.7	2.17	3.21	–1.7	–8.6
LTREV	16.7	11.0	88.9	47.7	3.23	1.77	–32.1	17.1
XFIN	13.0	4.1	58.7	83.4	3.17	3.44	–0.5	–15.6
STREV	14.4	7.6	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	9.2	16.8	58.3	58.3	2.49	2.3	–9.4	7.5
CGO	5.8	15.5	93.9	47.8	3.58	2.17	–57.4	57.7

Table III
Prospect Theory's Predictions for Anomaly Alphas

The first column lists 23 anomalies; the acronyms are defined in Table I. The remaining columns report, for each anomaly, the model-predicted annual alpha in percent for anomaly deciles 1 and 10, the model-predicted alpha spread between these two deciles, and the empirical annual value-weighted alpha spread between the two deciles. The sample period is 1963 to 2015 except for FPROB, DOP, and PEAD, for which, as described in the text, the sample period is shorter due to data availability. The horizontal lines in the table divide the anomalies into three groups: the 14 anomalies for which the model performs well (top group), the two anomalies for which the model does not make a strong prediction (middle group), and the seven anomalies for which the model performs poorly (bottom group).

Anomaly	Model Alpha Decile 1	Model Alpha Decile 10	Model Alpha Spread	Empirical Alpha Spread
CGO	–15.79	1.53	17.32	11.95
MOM	–12.52	2.0	14.52	23.26
FPROB	0.51	–12.8	–13.3	–18.82
ROA	–8.27	1.04	9.31	10.39
IVOL	0.34	–8.44	–8.78	–17.99
MAX	1.23	–6.6	–7.84	–13.71
EISKEW	0.03	–5.22	–5.25	–6.48
ZSC	-4.1	0.34	4.44	10.32
XFIN	0.57	–3.79	–4.36	–11.91
PROF	–2.71	0.94	3.65	6.43
PEAD	–2.3	0.96	3.26	7.12
CEI	0.31	–2.37	–2.68	–9.06
NSI	0.26	–1.95	–2.21	–9.23
DOP	0.42	–1.65	–2.07	–9.46
NOA	–0.14	–0.88	–0.74	–7.76
ORGCAP	–0.14	–0.41	–0.27	6.06
INV	–3.15	–1.31	1.84	–7.8
SIZE	–2.61	–0.27	2.34	–1.76
ACC	–3.92	–1.02	2.89	–8.35
AG	–4.0	–1.07	2.93	–8.28
STREV	–6.19	0.6	6.79	–3.57
VAL	0.6	–6.24	–6.85	5.79
LTREV	–8.19	0.42	8.61	–4.7

Table IV
Pricing Errors

The table reports the average absolute monthly pricing error across the 23 anomalies in Table I for the prospect theory model of Section II.A 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 23 long-short portfolios that, for the 23 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 23 anomalies, the difference in the empirical and theoretically predicted average return of the 23 long-short portfolios; we report the average absolute return difference. The sample period is 1963 to 2015.

Model	Average Absolute Pricing Error
Prospect theory	0.57
CAPM	0.82
Three-factor model	0.83
Four-factor model	0.55
Five-factor model	0.47
Six-factor model	0.31

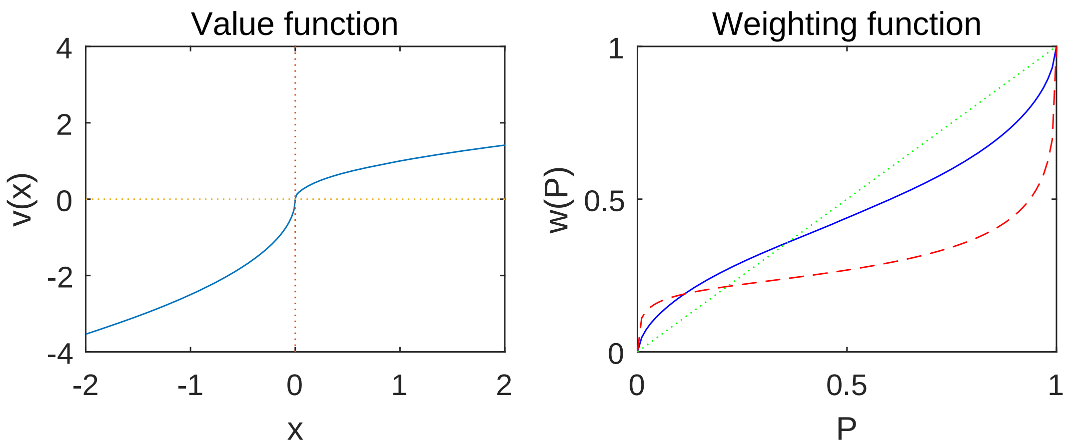

Figure 1. The prospect theory value function and probability weighting function. The left panel plots the value function proposed by Tversky and Kahneman (1992) as part of their cumulative prospect theory, namely, $v(x) = x^\alpha$ for $x \geq 0$ and $v(x) = -\lambda(-x)^\alpha$ for $x < 0$, for $\alpha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

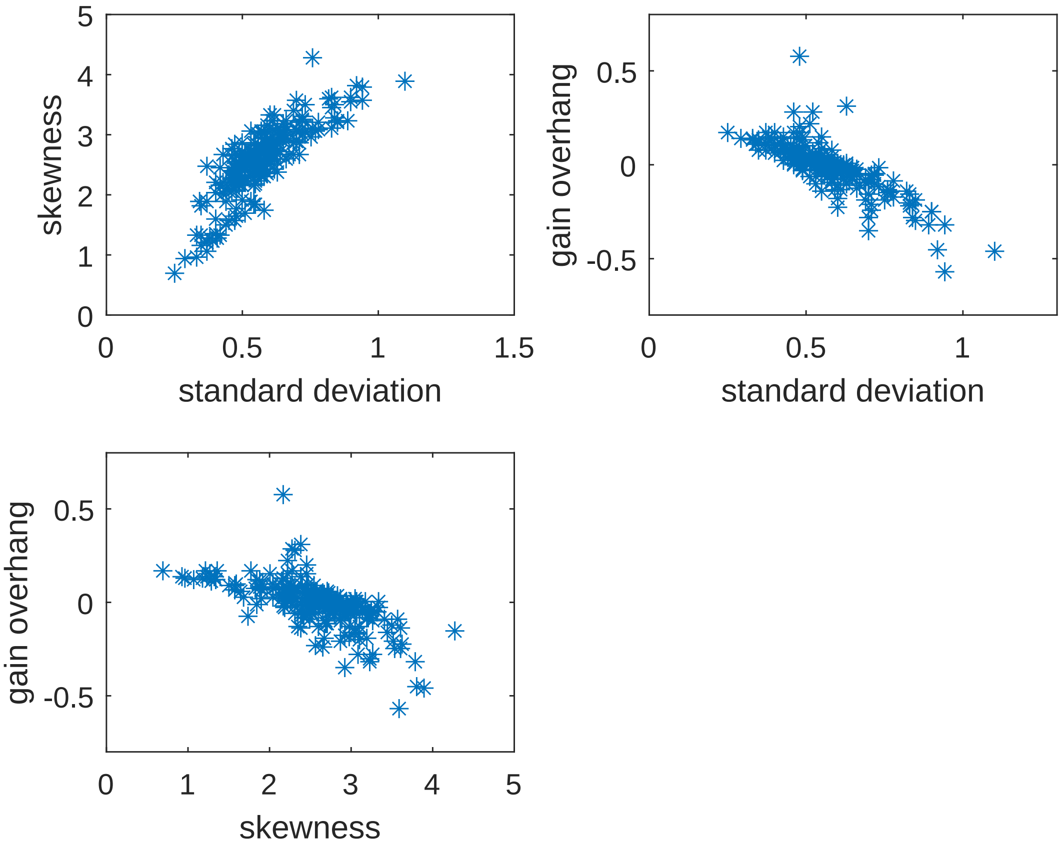

Figure 2. Empirical properties of anomaly deciles. Each graph plots 230 asterisks, where each asterisk corresponds to one of 10 deciles for one of 23 stock market anomalies. In the top-left graph, an asterisk that corresponds to some anomaly decile marks the standard deviation of the annual return and the skewness of the annual return for 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 the corresponding anomaly decile. In the bottom-left graph, each asterisk marks the return skewness and capital gain overhang of the typical stock in the corresponding anomaly decile.

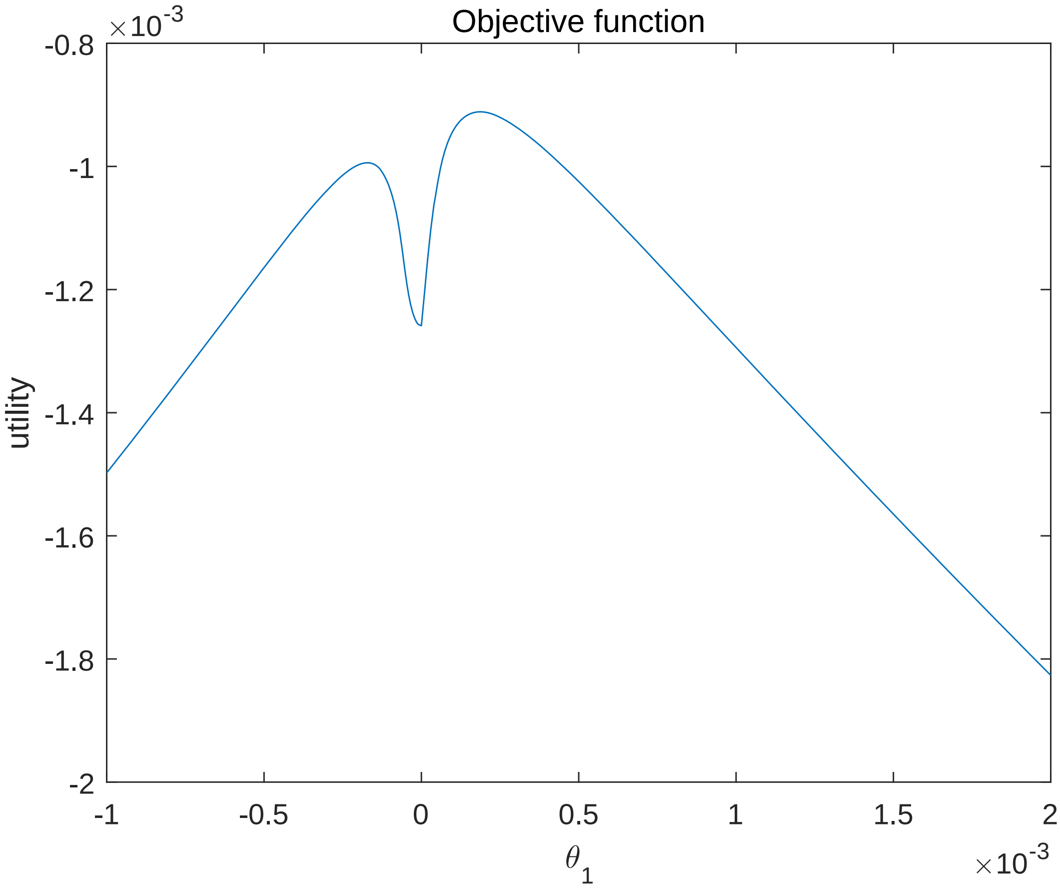

Figure 3. Investors' objective function for a stock in momentum decile 1. The solid line plots the value of an investor's objective function in equilibrium as a function of θ_1, the (scaled) fraction of the investor's portfolio allocated to stock 1, which belongs to momentum decile 1. The function has a unique global maximum at the point where θ_1 equals the weight of stock 1 in the market portfolio, namely 1.86×10^{-4}. The graph implies that, for any stock in momentum decile 1, investors have identical holdings of that stock.

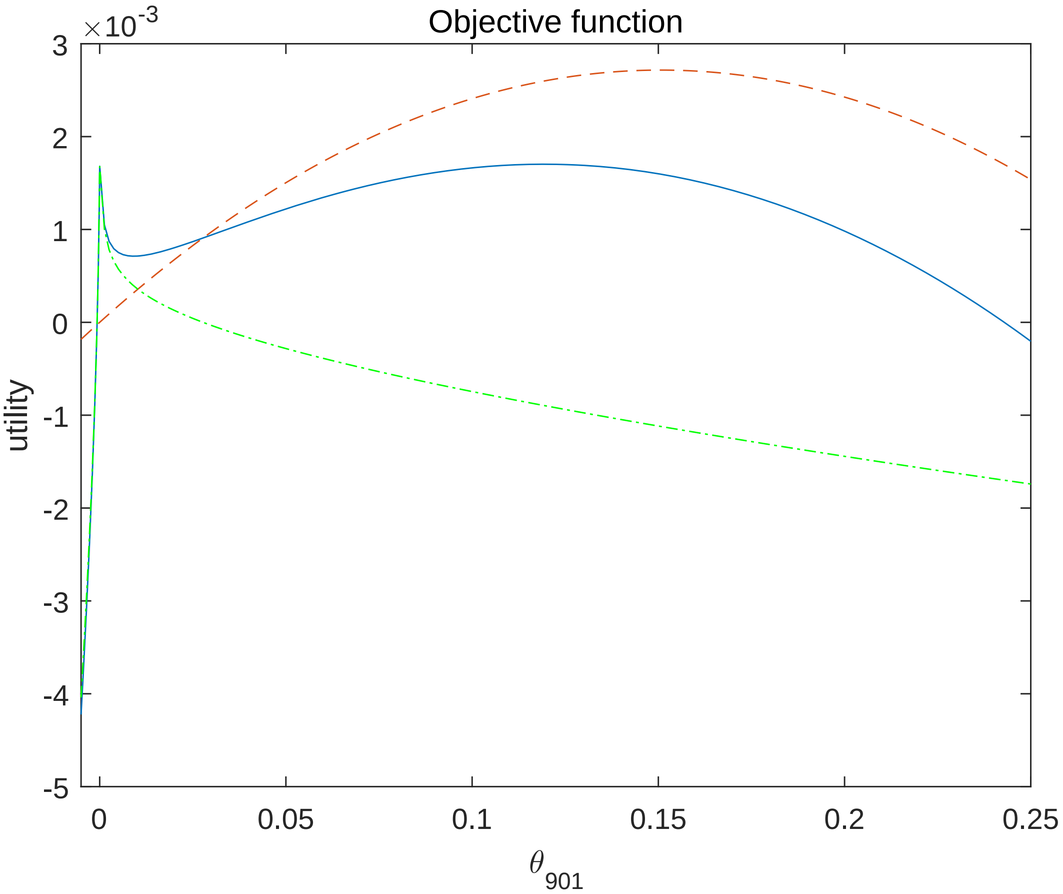

Figure 4. Investors' objective function for a stock in momentum decile 10. The solid line plots the value of an investor's objective function in equilibrium as a function of θ_{901}, the (scaled) fraction of the investor's portfolio allocated to stock 901, which belongs to momentum decile 10. The function has two global maxima at $\theta_{901} = 9.1 \times 10^{-5}$ and $\theta_{901} = 0.119$ which straddle the weight of stock 901 in the market portfolio, namely 7.26×10^{-4}. The dashed and dash-dot lines plot the mean-variance and prospect theory components of the objective function, respectively; the solid line is the sum of the dashed and dash-dot lines. The graph implies that, for any stock in momentum decile 10, investors have heterogeneous holdings of that stock.

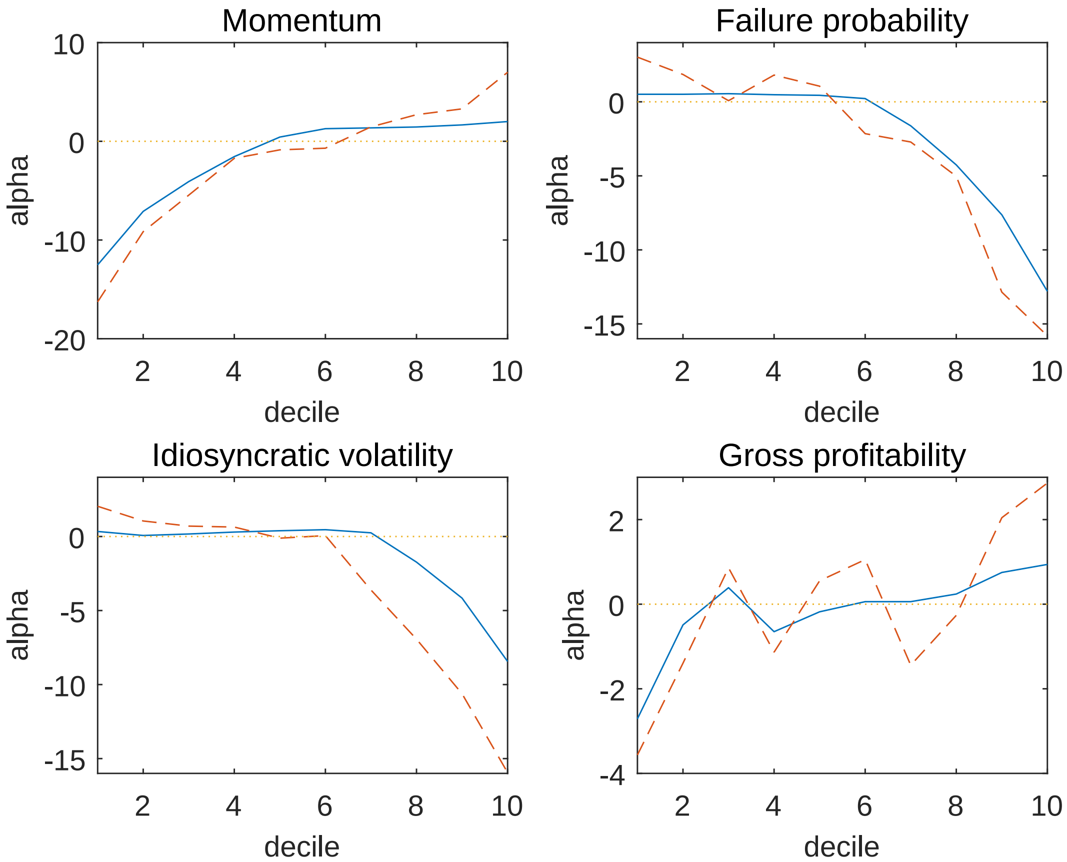

Figure 5. Model-predicted alphas for four prominent anomalies. The graphs present results for four of the anomalies for which the prospect theory model performs well: the momentum, failure probability, idiosyncratic volatility, and gross profitability anomalies. The dashed lines in the graphs plot the historical annual value-weighted alpha of each anomaly decile. The sample period is 1963 to 2015 except for the failure probability anomaly for which it is 1974 to 2015 due to data availability. The solid lines plot the alphas predicted by the model.

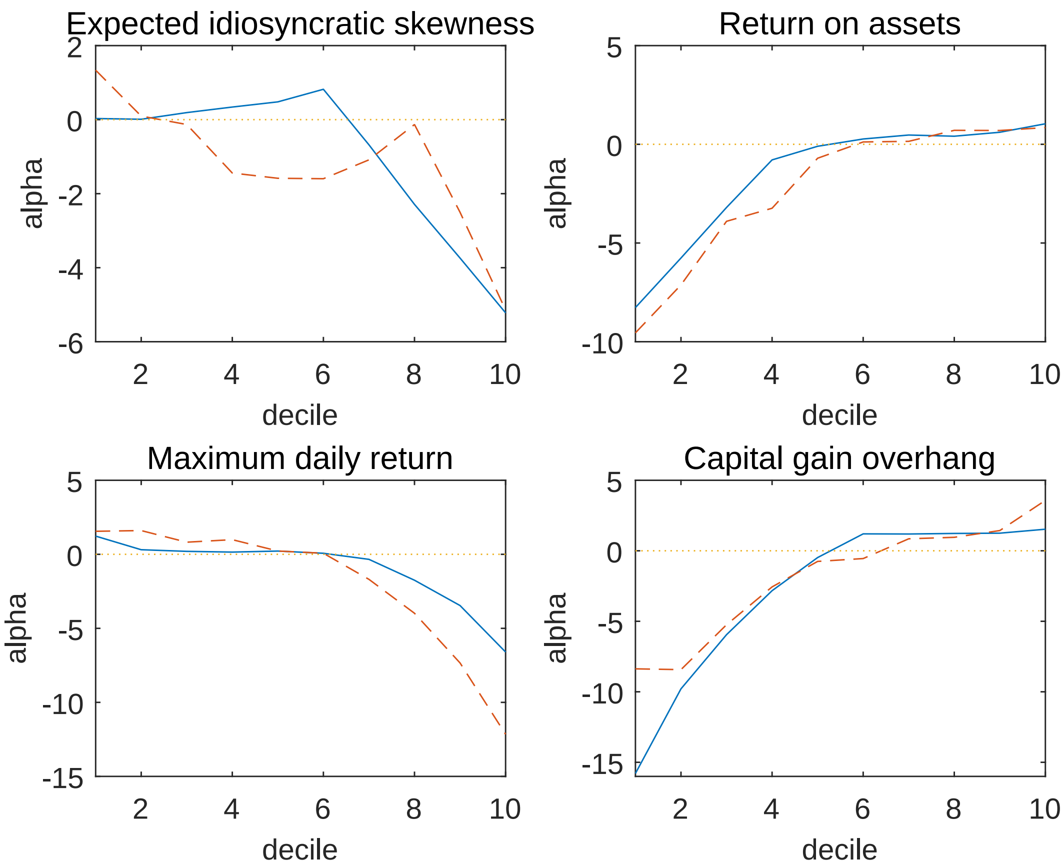

Figure 6. Model-predicted alphas for four prominent anomalies. The graphs present results for four of the anomalies for which the prospect theory model performs well: the expected idiosyncratic skewness, return on assets, maximum daily return, and capital gain overhang anomalies. The dashed lines in the graphs plot the historical annual value-weighted alpha of each anomaly decile. The sample period is 1963 to 2015. The solid lines plot the alphas predicted by the model.