NBER WORKING PAPER SERIES

PROSPECT THEORY AND STOCK MARKET ANOMALIES

Nicholas C. Barberis
Lawrence J. Jin
Baolian Wang

Working Paper 27155
http://www.nber.org/papers/w27155

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2020, Revised January 2021

The authors are grateful to John Campbell, Alex Imas, Jonathan Ingersoll, Eben Lazarus, Erik Loualiche, Ian Martin, Dimitri Vayanos, Keith Vorkink, and seminar participants at Boston College, Caltech, the London Business School, the London School of Economics, Princeton University, the University of California at Berkeley, the University of Maryland, the University of Miami, the University of Notre Dame, Washington University, Yale University, the AFA, the Miami Behavioral Finance conference, the NBER Behavioral Finance conference, and the SFS Cavalcade for very helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Prospect Theory and Stock Market Anomalies
Nicholas C. Barberis, Lawrence J. Jin, and Baolian Wang
NBER Working Paper No. 27155
May 2020, Revised January 2021
JEL No. G11,G12

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 the elements of prospect theory, takes account of investors’ prior gains and losses, and makes quantitative predictions about an asset’s average return based on empirical estimates of its volatility, skewness, and past capital gain. We nd that the model is helpful for thinking about a majority of the 23 anomalies.

Nicholas C. Barberis
Yale School of Management
135 Prospect Street
P O Box 208200
New Haven, CT 06520-8200
and NBER
nick.barberis@yale.edu

Lawrence J. Jin
California Institute of Technology
1200 E. California Blvd. MC 228-77
Pasadena, CA 91125
lawrence.jin@caltech.edu

Baolian Wang
Warrington College of Business
University of Florida
314 Stuzin Hall
Gainesville, FL 32611
Baolian.Wang@warrington.ufl.edu

1 Introduction

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

In this paper, we answer these questions. We build a new model of asset prices that incorporates prospect theory, as well as a related concept known as narrow framing, into investor preferences. We show how the model can be used to make quantitative predictions about the cross-section of average returns. In our main application, we take 23 prominent stock market anomalies and examine whether our model can help explain them. We find that the model is able to shed light on a majority of these anomalies.

Prospect theory posits that people evaluate risk using a utility function that is defined over gains and losses; that has a kink at its origin, capturing a greater sensitivity to losses than to gains ("loss aversion"); and that is concave over gains and convex over losses, capturing risk aversion over moderate-probability gains and risk-seeking over moderate-probability losses ("diminishing sensitivity"). It also states that people weight outcomes not by objective probabilities but by transformed probabilities that overweight the tails of the distribution they are thinking about ("probability weighting"). Prospect theory is often implemented in conjunction with narrow framing, a phenomenon observed in experimental studies whereby, when an individual is thinking about taking on a new risk, he evaluates it to some extent in isolation, separately from his other risks.

Intuition and prior research suggest that, in an economy with prospect theory investors who engage in narrow framing, the price of an asset will depend in part on three asset characteristics: the volatility of the asset's returns; the skewness of the asset's returns; and the average prior gain or loss since purchase across investors holding the asset, a quantity known as the asset's "capital gain overhang" (Grinblatt and Han, 2005). All else equal, investors require a *higher* average return on more volatile assets: since these investors evaluate each asset to some extent in isolation, and since they are loss averse, they find assets with volatile returns unappealing. All else equal, investors require a *lower* average return on assets with more positively-skewed returns: since these investors focus on an asset's own distribution of potential gains and losses, and since they overweight the tails of this distribution, they find assets with positively-skewed returns attractive. Finally, the utility function's concavity over gains and convexity over losses mean that, all else equal, investors require a *higher* average return on assets where they have larger prior gains.[1]

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

While our model has a simple structure, solving for equilibrium prices presents a challenge. In the model, all investors are identical. In an Expected Utility framework, this would imply that, in equilibrium, all investors hold identical portfolios. Strikingly, such an equilibrium does not exist once we introduce prospect theory preferences. To break this logjam, we construct an alternative equilibrium, one in which investors hold different portfolios that correspond to non-unique optima of their objective function. With this equilibrium structure in hand, we are able to generate quantitative predictions about the expected return on any risky asset.

In our main application, we examine whether the model can explain 23 prominent stock market anomalies. To see if our model can explain a particular anomaly – the size anomaly, say – we compute what it predicts for the average return of the typical small-cap stock. As explained above, this average return will depend on the return volatility, return skewness, and capital gain overhang of the typical small-cap stock. We estimate these quantities from historical U.S. data, plug them into our model, and record the model's prediction for the average return of a typical small-cap stock. We repeat this process for the typical stock in each of the ten market capitalization deciles. The results reveal how much, if any, of the size anomaly our model can explain. We proceed in the same way for all 23 anomalies.

Our empirical estimates of the volatility, skewness, and gain overhang of the typical stock in each anomaly decile are interesting in their own right. We find that the three characteristics are strongly correlated across anomaly deciles: if the typical stock in decile

[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.

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

This last observation points to the necessity of the quantitative approach we take in this paper. Consider again the size anomaly. Empirically, the returns of the typical small-cap stock are much more volatile than those of the typical large-cap stock. All else equal, this leads prospect theory investors to charge a *higher* average return on small-cap stocks than on large-cap stocks, thereby helping to explain the size anomaly. However, the typical small-cap stock also has more positively-skewed returns, and a more negative gain overhang, than the typical large-cap stock. All else equal, these two factors lead prospect theory investors to charge a *lower* average return on small-cap stocks, thereby hampering the model's ability to explain the size anomaly. Since one economic force goes in one direction and the other forces go in the opposite direction, the only way to determine prospect theory's prediction for the size anomaly is to develop a quantitative model that combines all three forces. This has not been done before; it 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. These are the momentum, failure probability, idiosyncratic volatility, gross profitability, expected idiosyncratic skewness, return on assets, capital gain overhang, maximum daily return, O-Score, external finance, composite equity issuance, net stock issuance, post-earnings announcement drift, and difference of opinion anomalies. The model explains these anomalies in the same way. For each of these 14 anomalies, the typical stock in the extreme decile with the lower average return is more positively skewed, more volatile, and has a lower gain overhang than the typical stock in the other extreme decile. The greater skewness and lower gain overhang of the former stock leads investors to charge a lower average return on it, while its higher volatility leads investors to charge a higher average return on it. Quantitatively, the first effect dominates. As a consequence, our model's prediction about the anomaly is in line with the empirical facts. To evaluate the model's performance more formally, we compare its average absolute pricing error across the 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 in 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 are more positively skewed and have a more negative gain overhang than growth stocks. All else equal, this leads prospect theory investors to charge a lower average return on value stocks. However, value stocks are also more volatile, which, all else equal, leads investors to charge a higher average return on them. Quantitatively, the first effect dominates. The model therefore predicts a lower average return on value stocks, contrary to the empirical facts. 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 intuitions laid out in earlier papers. The idea that, due to loss aversion, more volatile assets should, all else equal, have a higher average return, is discussed by Barberis and Huang (2001) and Li and Yang (2013); the idea that, due to diminishing sensitivity, an asset's gain overhang should be positively related to its average return, is developed by Grinblatt and Han (2005) and Li and Yang (2013); and the idea that, due to probability weighting, an asset's return skewness should be negatively related to its average return, is studied by Barberis and Huang (2008a) and Baele et al. (2019), among others.

Despite these advances, the basic questions listed in our opening paragraph remain unanswered. The reason is that the three above intuitions have opposite effects on average returns. As such, to determine what prospect theory predicts for the size anomaly, the value anomaly, or indeed any anomaly, we have to *combine* the three intuitions in a single model, something that has never been done before. To do this, we need a model that incorporates all the elements of prospect theory *and* takes account of investors' prior gain or loss in each asset. Most of the earlier models incorporate only a subset of the elements of prospect theory: only loss aversion (Barberis and Huang, 2001), only loss aversion and diminishing sensitivity (Li and Yang, 2013), or only loss aversion and probability weighting (Baele et al., 2019). Meanwhile, the prior models of the cross-section that do incorporate all the elements of prospect theory, those of Barberis and Huang (2008a), Ingersoll (2014), and Barberis, Mukherjee, and Wang (2016), are all one-period models; as such, they cannot 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 model is not an "easy extension" of prior models, but rather requires an entirely new equilibrium structure and solution method. The effort to develop these 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, this paper makes three contributions. 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-level puzzles. Finally, to our knowledge, the 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.

In Section 2, we review prospect theory and narrow framing. In Section 3, we present a model that incorporates these concepts and discuss the structure of the equilibrium. In Section 4, we introduce the 23 anomalies that are the focus of our study and compute the empirical characteristics that serve as inputs to the model. In Section 5, we present the model's predictions about stock market anomalies. Section 6 discusses some other aspects of our analysis, while Section 7 concludes. Additional details are in the Internet Appendix.

2 Prospect Theory and Narrow Framing

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

2.1 Prospect theory

The original version of prospect theory is described in Kahneman and Tversky (1979). This version of the theory has some limitations: it can be applied to gambles with at most two nonzero outcomes, and it predicts that people will sometimes choose dominated gambles. Tversky and Kahneman (1992) propose a modified version of the theory known as cumulative prospect theory that overcomes these limitations. This is the version we adopt in this paper.[2]

To see how cumulative prospect theory works, consider the gamble

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

which should be read as "gain or lose x_{-m} with probability p_{-m}, x_{-m+1} with probability p_{-m+1}, and so on," where $x_i < x_j$ for $i < j$ and where $x_0 = 0$, so that x_{-m} through x_{-1} are

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

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

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

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

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

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 $\delta = 0.4$ (the dashed line), for $\delta = 0.65$ (the solid line), and for $\delta = 1$, which corresponds to no probability weighting (the dotted line). Note that $v(0) = 0$, $w(0) = 0$, and $w(1) = 1$.

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

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

The severity of the kink is determined by the parameter λ; a higher value of λ implies a

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

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.[4] The degree of concavity over gains and convexity over losses are 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)-(4), the most extreme outcomes, x_{-m} and x_n, are assigned the weights $w(p_{-m})$ and $w(p_n)$, respectively. For the functional form in (6) and for $\delta \in (0,1)$, $w(P) > P$ for low, positive P; the right panel of Figure 1 illustrates this for $\delta = 0.4$ and $\delta = 0.65$. If p_{-m} and p_n are small, then, we have $w(p_{-m}) > p_{-m}$ and $w(p_n) > p_n$, so that the most extreme outcomes – the outcomes in the tails – are overweighted.

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

2.2 Narrow framing

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

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

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

his future wealth or consumption, and then checks if the new distribution is an improvement.

Research on decision-making under risk has uncovered many instances in which people do not appear to evaluate gambles in this way: instead of merging a new gamble with other risks they are already facing and checking if the combination is attractive, they often evaluate the new gamble in isolation, separately from their other risks. This is known as "narrow framing." Tversky and Kahneman (1981) present early laboratory evidence of narrow framing. More recently, Barberis, Huang, and Thaler (2006) argue that the commonly-observed rejection of the gamble in (7) is evidence not only of loss aversion, but of narrow framing as well.

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

3 Model and Equilibrium Structure

In the Introduction, we noted that, in an economy with prospect theory investors who engage in narrow framing, three asset characteristics are particularly important for the pricing of an asset: the volatility of the asset's returns; the skewness of the asset's returns; and the average paper gain or loss in investors' holdings of the asset. We now explain in more detail why these three characteristics are important.

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

The above intuitions make it clear that, to understand prospect theory's implications for asset prices, we need a model that incorporates all the elements of prospect theory *and* accounts for investors' prior gain or loss in each risky asset. No existing model of asset prices fulfills both conditions. We now present a new model that does. Solving a model of this kind presents significant challenges; to keep the model tractable, we necessarily make some simplifying assumptions. Nonetheless, the model captures the three essential intuitions described above in a robust way.

3.1 Model setup

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

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

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

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

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

where

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

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

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

The second term in (10), $W_{-1}\Theta_{i,-1}g_i$, is the gain or loss the investor experienced in his holdings of asset i *prior to* arriving at time 0. Here, W_{-1} is the investor's wealth at time -1, $\Theta_{i,-1}$ is the fraction of wealth allocated to asset i at time -1, and g_i is the capital gain on asset i between time -1 and time 0: if the investor experienced a capital gain of 30% on asset i between $t = -1$ and $t = 0$, then $g_i = 0.3$, while if he experienced a capital loss of 30%, then $g_i = -0.3$.[6] 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.

To keep the model tractable, we take the second term on the right-hand side of (10) to be identical across investors. Each investor in the model has the same prior gain or loss g_i in asset i, one that we will empirically estimate as the average gain or loss, since purchase, across all holders of the asset. In addition, for each investor, we will set $\Theta_{i,-1}$ to a neutral value, namely asset i's weight in the market portfolio. As such, the $W_{-1}\Theta_{i,-1}g_i$ term can be thought of as exogenous: $\Theta_{i,-1}$ is not a control variable that the investor chooses; the only control variable is Θ_i, the investor's allocation to asset i at time 0, which appears in the first term in (10). We have studied the impact 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 6, our conclusions are robust to these modifications. Finally, we use the approximation $W_{-1} \approx W_0$.[7]

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

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

We emphasize three aspects of the gain or loss in equation (10). First, we assume narrow, rather than broad, framing: equations (9) and (10) show that the investor derives utility from asset-level gains and losses, not from portfolio gains and losses. As noted in Section 2, there is both psychological and experimental support for this assumption.

Second, when making his decision at time 0, the 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 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.

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 6.1. In brief, we find that 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 CDF of asset i's returns and where $dw(P(R_i))$ and $dw(1-P(R_i))$ are given in full in Section A 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 loss aversion λ. The bottom row corresponds to gains. We allow investors to sell short. In Section B 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 C 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[8]

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

where $\Gamma(\cdot)$ is the Gamma function and K_l is the modified Bessel function of the second kind with order l.[9]

The above distribution 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

$$E(\widetilde{R}_i) = \overline{R}_i = \mu_i + \frac{\nu}{\nu - 2}\zeta_i \qquad (13)$$

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

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

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

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

In Sections 4 and 5, using equations (14) and (15), we will set S_i, ζ_i, and ν to match the empirical volatility and skewness of asset i's returns. We will then search for a value of μ_i so that the market for asset i clears. The asset's expected return – the quantity we want to determine – is then given by (13).[10]

The objective function in (9) combines a traditional component, namely mean-variance preferences, with a non-traditional one that incorporates prospect theory. As such, it is consistent with the approach advocated by Koszegi and Rabin (2006) among others, namely that models of gain-loss utility should retain a traditional utility term. One interpretation of this two-part utility function draws on the influential "two-system" framework in psychology (Kahneman, 2013). Under this view, the prospect theory term in (9) 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 is a combination of these intuitive and reasoned judgments: he anchors on his initial intuitive impression, which then leaves an imprint on his final decision.

Why do we take the traditional term in (9) to be mean-variance preferences, rather than some other Expected Utility specification? Our main application in this paper is to see if prospect theory can explain stock market anomalies, in other words, empirical deviations from the CAPM. As such, we want the traditional part of our preference specification to deliver CAPM pricing; this will allow us to cleanly identify the deviations from the CAPM that prospect theory generates. The simplest preferences that lead to the CAPM are mean-variance preferences.[11]

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

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

3.2 Equilibrium structure

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

Full rationality with homogeneous holdings. At time 0, the investors in our economy are identical in their preferences, their wealth, and their prior gain or loss in each risky asset. It is therefore natural to think that, in equilibrium, at time 0, they would all choose the same portfolio holdings $\{\Theta_i\}_{i=1}^{N}$, in other words, that they would each hold the market supply of each risky asset. Formally, such an equilibrium would consist of a location vector $\mu = (\mu_1, \ldots, \mu_N)'$ 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 we have examined, this type of equilibrium does not exist for the model in Section 3.1. Later in the paper, we illustrate this non-existence 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 for 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 such, there can be *no* value of μ_1 for which the objective function is maximized at a Θ_1^* that equals $\Theta_{M,1}$. An equilibrium where investors have identical holdings for all assets therefore does not exist.[12]

Full rationality with heterogeneous holdings. Given that the homogeneous-holdings equilibrium does not exist, what kind of equilibrium can we consider instead? One alternative structure involves multiple global maxima. In other words, it may be that there exists

[12]Why does the value of Θ_1^* at which (9) is maximized jump discontinuously as we lower μ_1? When $b_0 = 0$, the expression in (9), viewed as a function of Θ_1, depends only on Θ_1 and Θ_1^2. It therefore has a single local maximum that is also its global maximum. When $b_0 > 0$, (9) becomes a function of Θ_1, Θ_1^2, and 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.

a location vector $\mu = (\mu_1, \ldots, \mu_N)'$ such that, for this μ, the objective function in (9) has multiple global maxima, and that by allocating the appropriate number of investors to each maximum, we can clear the market in each asset.

The difficulty with this equilibrium structure is that it is computationally infeasible to 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 1000. For each of the 100^N location vectors, we need to solve the N-dimensional optimization problem in (9) and determine if there are multiple global maxima. We then need to see whether, by allocating investors to the various maxima, we can clear the market. This procedure is challenging even for $N = 2$ risky assets; for $N = 1000$, the more realistic value we use below, it is completely infeasible.

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

We 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 in turn, we take the objective function in (9), view it as a function of Θ_i, and then – this is where the bounded-rationality assumption comes in – set $\Theta_j = \Theta_{M,j}$ for all $j \neq i$. Up to a linear transformation, the resulting function can be written:

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

$$-\widehat{b}_0 \int_{R_f-\Theta_{i,-1}g_i/\Theta_i}^{\infty} (\Theta_i(R_i - R_f) + \Theta_{i,-1}g_i)^{\alpha} dw(1 - P(R_i)), \tag{16}$$

where

$$\widehat{\gamma} = \gamma W_0 \tag{17}$$
$$\widehat{b}_0 = b_0 W_0^{\alpha-1}. \tag{18}$$

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 has multiple global maxima, one of which lies below $\Theta_{M,i}$ and one of which lies above it, thereby allowing us to clear the market in asset i by allocating some investors to the lower optimum and others to the upper optimum.

The bounded-rationality assumption greatly simplifies the investors' optimization problem: by turning the multivariate function in (9) into the univariate function in (16), it converts the search for the optimal allocation Θ_i to asset i into a one-dimensional problem, one where investors trade off a larger allocation to asset i and lower allocation to the risk-free asset against the opposite strategy. Moreover, because the problem is now one-dimensional, it is easy to determine whether the function in (16) has multiple global maxima or a unique global maximum.[13]

We find that a bounded-rationality equilibrium with heterogeneous holdings exists for a wide range of parameter values, and it is the one we focus on. We note two things about it. First, in this equilibrium, investors need not have heterogeneous holdings for *all* the risky assets. When we implement the equilibrium, we find that, for many risky assets, investors have identical holdings. Second, we find that, for any asset i where investors have heterogeneous holdings, the function in (16) has just two global maxima, Θ_i^* and Θ_i^{**}, both of which are non-negative. These maxima straddle the market supply $\Theta_{M,i}$, so that $\Theta_i^* < \Theta_{M,i} < \Theta_i^{**}$, and this allows us to clear the market in the asset by assigning some investors to the Θ_i^* allocation and the rest to the Θ_i^{**} allocation. We also find that Θ_i^* is 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 – as described in Section 2, risk-seeking over moderate-probability losses and over low-probability gains; if investors were instead uniformly risk averse, the objective function

[13]If a symmetric equilibrium were to exist under full rationality, then this equilibrium would coincide with the bounded-rationality equilibrium.

would be concave and would have only one global maximum. We discuss this in more detail in Section 4.2 by way of a numerical example. For now, we note that there is a simple intuition for the two optima, Θ^* and Θ^{**}. The heterogeneous holdings typically arise for stocks with a positive gain overhang which, in part because of this, have a high expected return. The lower optimum Θ^* reflects investors' desire to lock in their prior gains in these stocks 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 stocks. In equilibrium, the two strategies are equally attractive.

To understand the portfolios that the investors in our economy hold, suppose that there are $N = 1000$ risky assets, and that for 500 of them, all investors have identical holdings – for these assets, the function in (16) has a unique global maximum – while for the remaining 500, the function in (16) has two global maxima, so that investors have heterogeneous holdings; this equal split between assets with homogeneous holdings and assets with heterogeneous holdings approximates what we find when we implement the equilibrium. All investors then hold the first 500 assets in proportion to their market weights. For the vast majority of the remaining 500 assets, the fraction of his portfolio that a given investor allocates to each one is lower than its market weight – this is the Θ_i^* optimum – but for a small handful of these assets, he holds a large position given by the Θ_i^{**} optimum. Overall, then, each investor combines a diversified portfolio of many assets with a small number of concentrated holdings – a portfolio structure that mirrors that of many real-world investors.

An investor's assumption, when solving for his optimal allocation to asset i, that his holdings of the remaining assets equal their market weights, is not exactly correct: by the nature of the heterogeneous-holdings equilibrium, he may have an undiversified position in a small number of these other risky assets. However, this discrepancy is likely to have a negligible impact on the model's predictions: we find that, if, when solving for his allocation to asset i, the investor instead makes the *correct* assumption that he has a few undiversified holdings, the quantitative predictions for expected returns are very similar to those of the simpler bounded-rationality equilibrium defined through equations (16)-(18).[14]

[14]We have also studied the following iterative procedure. When solving for his allocation Θ_i to asset i, the investor starts with an assumption A_1 about his remaining holdings – specifically, that they equal market weights. He then uses the resulting optimal portfolio P_1 as a new assumption A_2 about the structure of his portfolio: he again solves for his allocation Θ_i, this time under assumption A_2 about his remaining holdings. He then takes the new optimal portfolio P_2 as a new assumption about his remaining holdings, and so on. This iterative procedure converges to a "self-consistent" heterogeneous-holdings equilibrium, one where the assumption the investor makes about his remaining holdings when solving for his allocation to a particular asset is consistent with the portfolio he actually ends up choosing to hold. The expected asset returns in this self-consistent equilibrium are quantitatively very similar to those in the simpler bounded-rationality equilibrium in equations (16)-(18).

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

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

From now on, we think of investors as choosing θ_i rather than Θ_i. In Section D 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 $\widehat{b}_0$. The rescaled problem is simpler to implement because it is easier to compute an empirical counterpart for $\theta_{M,i}$ than for $\Theta_{M,i}$ and because the rescaling allows us to simplify 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)-(18). A bounded-rationality equilibrium consists of a location vector $\mu = (\mu_1, \ldots, \mu_N)'$ such that, for each i, the function

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

where

$$\widehat{\gamma} = \gamma W_0 \Theta_{M,R} \tag{21}$$

$$\widehat{b}_0 = b_0 W_0^{\alpha-1} \Theta_{M,R}^{\alpha-1}, \tag{22}$$

has either a unique global maximum at $\theta_i = \theta_{M,i}$ or multiple global maxima that straddle $\theta_{M,i}$. In Section B 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, in equilibrium, no investor sells short: even for a stock i where investors have heterogeneous holdings, the lower optimum θ_i^* is always greater than or equal to zero. We return to this in Section 4.2.

In Section E of the Internet Appendix, we explain in full the procedure we use to de-

[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 1 for any given investor.

termine whether investors have identical or heterogeneous holdings in an asset, and to then compute the asset's expected return. Since it involves numerical integration, this calculation takes a few minutes of computing time; this is fast enough for the application we consider in this paper.

4 Anomalies and Model Parameter Values

The model of Section 3 generates quantitative predictions about the cross-section of average returns when investors evaluate risk according to prospect theory. We now use the model to answer a basic but long-standing question: Can prospect theory shed light on stock market anomalies? To be as comprehensive as possible, subject to the computational constraints we face, we consider 23 anomalies. They are listed in Table 1, along with the abbreviations we use in subsequent tables to refer to them; in Section F 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 the 11 anomalies studied by Stambaugh, Yu, and Yuan (2012) and then add 12 more from the 97 anomalies studied by McLean and Pontiff (2016). 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 6.2, we check that our results are not specific to the 23 anomalies in Table 1 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 $N = 1000$ stocks; each anomaly decile therefore contains 100 stocks. We number the stocks so that, in the case of the value anomaly, say, stocks 1 to 100 belong to decile 1, which contains stocks with low book-to-market ratios; stocks 101 to 200 belong to decile 2; and so on. All stocks in a given decile are identical: they have the same characteristics, namely, the empirical characteristics of the typical stock in that anomaly decile. For each decile in turn, we choose one stock at random and compute our model's prediction for its expected return. Since all stocks in a given decile are identical, this immediately tells us the expected return of all the stocks in that decile. Our model can help explain the value anomaly if the expected return it predicts for the randomly-chosen stock in decile 10 is significantly higher than the expected return it predicts for the randomly-chosen stock in decile 1.[16]

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

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 do so, we use the value anomaly as an example; the process is the same for all the anomalies we consider.

Each month from July 1963 to December 2014, we rank all stocks listed on the NYSE, Amex, or Nasdaq on their book-to-market ratio and then group them into deciles. (For each of the other anomalies, we instead rank stocks on the relevant anomaly characteristic – for example, on their idiosyncratic volatility in the case of the volatility anomaly.) Decile 1 corresponds to stocks with the lowest book-to-market ratios, and decile 10 to stocks with the highest book-to-market ratios. Suppose that, in some particular month, each decile contains 100 stocks. Take decile 4, say. To compute the beta of the typical stock in decile 4 in this month, we calculate the betas of each of the 100 stocks in the decile and average them.[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 this month. At the end of this process, we have four quantities in hand for each anomaly decile in this month: the volatility, skewness, gain overhang, and beta for the typical stock in that decile.

We repeat the above exercise for each month in our sample. This gives us, for each book-to-market decile, a time series for each of the four quantities: return volatility, return skewness, gain overhang, and beta. In the final step, we compute the mean of each time 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]

reality, any given stock makes up only a small fraction of the overall market. Setting $N = 1000$ allows us to do this. Since all 100 stocks in each decile are identical, the computational burden in the case of $N = 1000$ is the same as for $N = 10$.

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

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

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 dataset of individual investor trading, portfolio turnover is on the order of 50% per year (Barber and Odean, 2000); this 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 this, namely 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: when trying to estimate a stock's skewness, say, they may take a set of stocks with similar characteristics and then check how often stocks in this set post an extreme right-tail return. The cross-sectional volatility and skewness are fairly stable from month to month, which means that investors can learn them from even a short sample of data.[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, 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 computed both measures of gain overhang and find that they lead to similar quantitative estimates and similar model predictions. We therefore pick one – the Grinblatt and Han (2005) measure, because it is easier to compute and accounts for both individual

than decile 10.

[19]Barber and Odean (2000) report that the median annual portfolio turnover is 32% while the mean turnover is 78%. Both numbers are informative for our purposes, but the median turnover is more so.

[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 then average these quantities across the 100 stocks. We can then use the volatility and skewness of daily returns to make inferences about the volatility and skewness of annual returns. However, this last step is challenging, particularly in the case of skewness, 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.

and institutional investors – and stick with it throughout.[21]

Table 2 presents the results of the above empirical exercise. 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 definition of what an anomaly is, these average returns differ in a way that is not explained by beta. 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 2. First, for most of the anomalies, the typical stock in decile 1 differs substantially from the typical stock in decile 10 in its return volatility, return skewness, and gain overhang – in other words, in the three characteristics that, aside from beta, determine expected returns in our model. Consider, for example, the size anomaly: the typical stock in decile 1 has an annual standard deviation of 76%, while the typical stock in decile 10 has an annual standard deviation of just 25%. Similarly, while the typical small-cap stock has an annual return skewness of 4.3, the typical large-cap stock has an annual return skewness of just 0.7. And while the typical small-cap stock has a negative gain overhang of −15%, the typical large-cap stock has a positive gain overhang of 17%.

The second, more striking, observation is that the three characteristics – standard deviation, skewness, and gain overhang – are strongly correlated across anomaly deciles: for 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 in the figure. Each asterisk in the graph corresponds to an anomaly decile; since there are 23 anomalies, this makes for a total of 230 asterisks. The horizontal and vertical axes in the graph measure the standard deviation and skewness, respectively, of the typical stock in an anomaly decile.

[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 graph shows the positive correlation between these two quantities. In a similar way, the other two graphs show the negative correlation, across anomaly deciles, between standard deviation and gain overhang, and between skewness and gain overhang.[22]

The empirical patterns in Figure 2 point to the necessity of our quantitative approach. Suppose that, for one of the extreme decile portfolios – decile 1, say – the typical stock in that decile has a higher return skewness, higher return volatility, and lower gain overhang than the typical stock in the other extreme decile, decile 10; of the 23 anomalies, 21 follow this pattern. It is then impossible to tell, without a quantitative model, whether prospect theory can explain the anomaly. The reason is that there are counteracting forces. Decile 1 stocks have more volatile returns than decile 10 stocks. Since prospect theory investors are loss averse, this will lead them, all else equal, to require a higher average return on decile 1 stocks than on decile 10 stocks. However, decile 1 stocks also have more skewed returns than decile 10 stocks. Since prospect theory investors exhibit probability weighting, this will lead them, all else equal, to charge a lower average return on decile 1 stocks. Finally, decile 1 stocks trade at a loss, while decile 10 stocks trade at a gain. Due to diminishing sensitivity, this will lead prospect theory investors, all else equal, to require a lower average return on decile 1 stocks. Since two of these forces go in one direction and the other goes in the opposite direction, we need a quantitative model to determine whether prospect theory can explain the anomaly.

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

4.1 Parameter values

To see if our model can capture a particular anomaly, we proceed as follows. We consider an economy with $N = 1000$ stocks, and assign 100 of these stocks to each anomaly decile: stocks 1 to 100 belong to anomaly decile 1, stocks 101 to 200 to anomaly decile 2, and so on. For any given decile, we take all the stocks in the decile to be identical: they have the same standard deviation, skewness, gain overhang, and beta, namely the empirical standard deviation, skewness, gain overhang, and beta of the typical stock in that anomaly decile, computed as described above. We set the parameters S_i and ζ_i of the skewed t distribution,

[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.

the capital gain g_i, and the beta β_i to capture these empirical values. We then search for a location parameter μ_i so that the conditions for equilibrium described in Section 3.2 around equation (20) are satisfied. Our model's prediction for a stock's expected return is then given by (13). Note that all stocks in a given decile will have the same 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; $\{\beta_i\}$, 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; $\widehat{b}_0$, the importance of the prospect theory term in investor preferences; $(\alpha, \delta, \lambda)$, the prospect theory preference parameters; and $\{\theta_{i,-1}\}$, investors' prior allocations to the N stocks.[23]

We start with the asset-level parameters. We set $\nu = 7.5$, which represents a reasonable degree of fat-tailedness in stock returns; our results are not very sensitive to the value of 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,

$$\begin{aligned}
\mathrm{Std}(\widetilde{R}_i) &= \left[\frac{\nu}{\upsilon-2}S_i + \frac{2\nu^2}{(\nu-2)^2(\nu-4)}\zeta_i^2\right]^{0.5} && (23)\\
\mathrm{Skew}(\widetilde{R}_i) &= \frac{2\zeta_i\sqrt{\nu(\nu-4)}}{\sqrt{S_i}(2\nu\zeta_i^2/S_i+(\nu-2)(\nu-4))^{\frac{3}{2}}}\left[3(\nu-2)+\frac{8\nu\zeta_i^2}{S_i(\nu-6)}\right]. && (24)
\end{aligned}$$

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 plug them into the left-hand side of equations (23) and (24). These equations then allow us to solve for the two unknowns, S_i and ζ_i. For example, in the case of the size anomaly, stocks 1 to 100 belong to the lowest market capitalization decile. From Table 2, we see that the empirical standard deviation and skewness for the typical stock in this decile are 0.76 and 4.27, respectively. Accordingly, for this anomaly, to set the values of S_i and ζ_i for

[23]An alternative to calibrating the model is to estimate it – specifically, to estimate the values of α, δ, λ, $\widehat{b}_0$, and $\widehat{\gamma}$. However, without the use of approximations, estimation is not feasible: it requires generating expected returns for the 230 anomaly deciles for thousands of different sets of parameter values; because of the numerical integration involved, generating the 230 expected returns even for *one* set of parameter values takes a substantial amount of time. One appealing feature of our calibration approach is that it allows us to see how well the model performs when parameterized using data that are *independent* of the anomalies it is trying to explain.

$i \in \{1, \ldots, 100\}$, 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 displayed in Table 2 for the two extreme deciles of each anomaly. We set σ_M to 0.25 and the gross risk-free rate R_f to 1.

We set $\{\theta_{M,i}\}$, the stocks' market weights, to match empirical market weights. Take, for example, the volatility anomaly. In each month of our sample, we compute the fraction of the total market value of all stocks in our sample in that month that is made up by the stocks in each volatility anomaly decile. We then compute the time-series averages of these fractions. We find that, on average, volatility decile 1 makes up 29.6% of total stock market value. Since, in our model, there are 100 identical stocks in decile 1, we set $\theta_{M,i} = 0.296/100$ for all stocks in decile 1, in other words, for $i = 1, \ldots, 100$. 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 6.2. We set $\widehat{\gamma}$, the scaled portfolio risk aversion in (21), and $\widehat{b}_0$, the scaled weight on the prospect theory term in (22), to generate an aggregate equity premium of 6%. There are many pairs $(\widehat{\gamma}, \widehat{b}_0)$ that produce an equity premium of 6%. How do we choose one? As we increase $\widehat{b}_0$, we increase not only the equity premium, but also investors' degree of under-diversification. We therefore choose, from among the $(\widehat{\gamma}, \widehat{b}_0)$ pairs that generate a 6% equity premium, a pair with a value of $\widehat{b}_0$ that is consistent with the degree of under-diversification estimated in actual portfolios by Calvet, Campbell, and Sodini (2007). This is the pair $(\widehat{\gamma}, \widehat{b}_0) = (0.6, 0.6)$.[24] For this value of $\widehat{b}_0$, the prospect theory

[24]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 $(\widehat{\gamma}, \widehat{b}_0) = (0.6, 0.6)$, the average

term in equation (9) makes up a modest 12% of the value of the overall objective function. In Section G of the Internet Appendix, we also check that, for this $\widehat{b}_0$, the investors' objective function generates sensible attitudes to large and small gambles: it 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 are almost 30 years old and are based on a small number of participants. Given that the values we assign to these parameters play a significant role in our results, it seems prudent to base these values on a wide range of studies, not just one.

Tversky and Kahneman's (1992) results have led to the widespread view that the degree of loss aversion λ is approximately 2. However, recent studies suggest that the true level of loss aversion in the population is significantly lower. In a meta-analysis of experimental estimates of loss aversion, Walasek, Mullett, and Stewart (2018) find that the median estimate of 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 $\lambda = 0.98$ for their median participant, albeit somewhat higher for those with greater cognitive ability. To reflect these findings, in a conservative way, we set $\lambda = 1.5$.

Booij, van Praag, and van de Kuilen (2010) compile a list of experimental estimates of α and δ. The median estimate of δ is close to Tversky and Kahneman's (1992) estimate; we therefore maintain $\delta = 0.65$. Experimental estimates of α span a fairly wide range; most lie between 0.5 and 0.95. We set α near the midpoint of this range, at 0.7. Our preference parameter values are therefore[25]

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

Sharpe ratio of investors' portfolios in our economy is 12.5% lower than that of the market portfolio – a similar level of under-diversification.

[25]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). Meanwhile, even the lower level of loss aversion in (27) is strong enough to generate a high equity premium and non-participation in the stock market – two prominent applications of loss aversion in finance.

4.2 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 stocks in momentum decile 1, namely stocks 1 to 100, investors have identical holdings, each holding the market supply of each stock, $\theta_{M,i}$. For example, for stock 1, there is a value of μ_1, the location parameter for stock 1, such that the function in (20), for $i = 1$, has a unique global maximum at $\theta_1 = \theta_{M,1} = 1.86 \times 10^{-4}$. 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 our model, investors usually have identical holdings in stocks that are trading at a loss and that, in part because of this, have a low, even negative, expected return. As an investor increases his allocation to stock 1 from zero, utility initially rises: he is in the convex region of the value function, where he is risk-seeking. As he further increases his allocation, utility falls: a larger fraction of time 1 outcomes bring the investor to the right of the kink in the value function, increasing his risk aversion; the stock's low expected return and the greater portfolio volatility caused by the larger allocation to the stock decrease utility further.

Figure 3 also shows that no investor will choose to short stock 1. Shorting the stock gives the investor a high expected return; however, it also exposes him to strong negative skewness which, due to probability weighting, he finds aversive. This explains why, as the investor decreases his allocation from zero, the objective function initially rises but then falls.

For stocks in momentum decile 10, namely stocks 901 to 1000, investors have heterogeneous holdings. For example, for stock 901, there is a value of μ_{901} such that the function in (20), for $i = 901$, has two global maxima that straddle the 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 $\theta_{901}^{**} = 0.119 > \theta_{M,901}$. Since these optima straddle the market supply, we can clear the market by allocating most investors to the first optimum and the rest to the other optimum.

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, 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 average 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 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, in turn, allows the mean-variance component to create a second global maximum at $\theta^{**}_{901} = 0.119$.[26]

Figure 4 also shows that no investor will choose to short stock 901. Shorting the stock gives the investor both a low expected return and exposes him to negative skewness. This is an unattractive combination and explains why the objective function in Figure 4 falls sharply as the investor lowers his allocation from zero.

The above intuition 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 holdings; 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, in turn, 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 which 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 gave above for the shape of the solid line in Figure 4 helps explain why this structure arises as commonly as it does.

In Section H of the Internet Appendix, we discuss some other aspects of the equilibrium

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

structure. We explain why, for the stocks in momentum decile 10, there is no equilibrium with identical holdings; and why, for the 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).

5 Application

We now use our model to answer a long-standing question: Can the risk attitudes captured by prospect theory shed light on the prominent anomalies in Table 1?

To determine whether the model can help explain an anomaly, we focus on anomaly alphas. For any given anomaly, we compute the empirical alphas for the ten anomaly deciles over our 1963-2015 sample – these are value-weighted CAPM alphas computed from a monthly regression and annualized – and denote them as $\alpha^d(1), \ldots, \alpha^d(10)$, where the "$d$" superscript stands for "data." We then compute the alphas predicted by the model for each of the ten deciles, namely $\alpha^m(1), \ldots, \alpha^m(10)$, where "$m$" stands for "model." Since, within each decile, all stocks are identical and hence have the same expected return and alpha, we can compute the alpha of decile l as the alpha of any stock in that decile – for example, as the alpha of stock $100l$:

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

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

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

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

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

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

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

We find that the model is helpful for thinking about a strikingly large number – a majority, in fact – of the anomalies we consider; we review these anomalies in Section 5.1. In Section 5.2, we discuss the anomalies where the model performs poorly. And in Section 5.3, we note the anomalies where the model does not make a strong prediction. Finally, in Section 5.4, we compare our model's performance to that of some widely-used factor models.

5.1 Anomalies where the model performs well

We present the model's predictions in Table 3. The horizontal lines in the table divide the 23 anomalies into three groups. The upper group is the anomalies that the model performs well on; for these anomalies, the conditions in (28) are satisfied. The middle group is the anomalies that the model does not make a strong prediction about; for these anomalies, the condition in (30) holds. Finally, the lower group is the anomalies that the model performs poorly on; 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, O-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 take eight of the 14 anomalies and present additional results about them. The top-left graph in Figure 5 shows the results for the momentum anomaly. The horizontal axis corresponds to the ten decile portfolios, 1 through 10. The dashed line plots the empirical alpha of each decile. The solid line plots the alphas predicted by the model. The other graphs in Figures 5 and 6 have the same structure. Table 3 and Figures 5-6 show that, for many of the 14 anomalies, the model can explain a large fraction of the spread in empirical alphas between anomaly deciles 1 and 10.[27]

The intuition for why our model helps to explain the 14 anomalies is the following. For these anomalies, the extreme decile with the lower empirical alpha – for example, decile 1 in the case of the momentum anomaly and decile 10 in the case of the volatility anomaly – contains stocks with more volatile returns, more skewed returns, and a more negative capital gain overhang. On the one hand, the higher volatility of these stocks leads the investors in our economy to charge a higher average return on them. On the other hand, the stocks' higher skewness and more negative gain overhang leads investors to charge a lower average return on them. The latter force dominates, so that the model predicts a low average return on these stocks, consistent with the data.

Figures 5 and 6 show that the model can explain not only the alphas for the extreme deciles, but also those 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 return that prospect theory predicts for momentum deciles. Our analysis shows that, once we take all three characteristics – volatility, skewness, and gain overhang – into account, prospect theory can indeed explain the momentum anomaly, but it is only through a quantitative framework like the one we develop in this paper that this conclusion can be drawn.

[27]For six of the 14 anomalies, the model explains a sufficiently large fraction of the empirical spread that there is no statistical difference between the empirical and model-predicted spreads. These are the gain overhang, idiosyncratic skewness, failure probability, O-Score, gross profitability, and return on assets anomalies.

Similarly, prior work has linked the probability-weighting component of prospect theory to some of the anomalies we consider – specifically, to the idiosyncratic volatility, idiosyncratic skewness, and failure probability anomalies (Campbell, Hilscher, and Szilagyi, 2008; Boyer, Mitton, and Vorkink, 2010; Conrad, Kapadia, and Xing, 2014). The idea is that, for these anomalies, stocks in decile 10 have positively-skewed returns, which, due to probability weighting, leads investors to charge a lower average return on them. Our analysis confirms this mechanism, but also indicates that the argument is incomplete. Highly-skewed stocks also tend to have more volatile returns and to trade at a loss – characteristics that, in a prospect theory framework, also affect their average returns. We show that, when all three characteristics are taken into account, prospect theory can explain the above anomalies – but, again, it is only through our quantitative approach that this can be confirmed.

The results in Table 3 also draw a connection between prospect theory and a number of anomalies that, to our knowledge, has not previously been noted. For example, they show that prospect theory is helpful for thinking about the gross profitability, return on assets, external finance, composite equity issuance, net stock issuance, and difference of opinion anomalies – again, anomalies that prospect theory has not previously been linked to.

5.2 Anomalies where the model performs poorly

For seven of the 23 anomalies – the seven anomalies in the bottom group in Table 3 – the model performs poorly, in that, as laid out in the conditions in (29), it predicts a substantial difference between the alphas for deciles 1 and 10, but of the wrong sign. These are the size, value, long-term reversal, short-term reversal, accrual, asset growth, and investment anomalies. We discuss two of these – the size and value anomalies – in more detail, and the others more briefly.

Table 3 shows that the model incorrectly predicts a negative alpha for decile 1 of the size anomaly, which contains the stocks with the lowest market capitalizations. These stocks are very volatile, which, all else equal, leads investors to charge a high average return on them. However, they also have positively-skewed returns and trade at a loss, which leads investors to charge a low average return on them. Our quantitative analysis shows that the second effect overwhelms the first one; as a result, the model makes an incorrect prediction.

Table 3 shows that the model also fails to explain the value anomaly. The reason is that value stocks have more positively-skewed returns than growth stocks and trade at a larger loss; this leads investors to charge a lower average return on value stocks. It is true that value stocks are also more volatile than growth stocks, which, all else equal, leads investors to charge a higher average return on value stocks. However, the second effect is overwhelmed by the first one, and the model makes the wrong prediction. Table 3 also reports results for five other anomalies that the model performs poorly on – the long-term 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 I 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. This, in turn, raises the possibility that many anomalies can be put into one of two categories: "preference-based" anomalies that are driven by investor risk attitudes of the kind captured by prospect theory, and "belief-based" anomalies driven by incorrect beliefs.[28]

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 3 and Figures 5-6, we assumed that investors have accurate beliefs about stocks' return volatility, return skewness, and gain overhang. However, for certain types of stocks, investors' beliefs about these characteristics may be incorrect. For example, when generating the model's predictions for the value anomaly, we assumed that investors *know* that value stocks have more positively-skewed returns than growth stocks; this, in turn, leads the model to predict a lower average return on value stocks. However, if investors think that it is growth stocks that 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 J 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.[29]

[28]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.

[29]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 over-estimate the skewness of growth stocks and under-estimate the skewness of value stocks.

5.3 Anomalies where the model does not make a strong prediction

For two other anomalies – the middle group in Table 3, 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.

5.4 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. Specifically, 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, say, is computed in the usual way as the intercept in a regression of the long-short portfolio return on the three Fama-French factors.

Table 4 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 to the four-factor model, given that the latter was developed in full knowledge of the size, value, and momentum anomalies. By contrast, our model was developed independently of any information about the anomalies: prospect theory itself was designed with no knowledge of any anomalies, and we calibrated the model based on experimental evidence that is again independent of the anomalies. Despite being handicapped in this way, our model is able to match the four-factor model in its performance. It is not surprising that our model underperforms the five- and six-factor models as these were developed in full knowledge of multiple anomalies.

6 Discussion

In this section, we discuss some additional issues raised by the analysis in Section 5.

6.1 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 answer this, we take the 14 anomalies that the model can help explain, and, for each one, we decompose the model-predicted alpha spread between deciles 1 and 10 into three pieces 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 only loss aversion, and not diminishing sensitivity or probability weighting, so that the preference parameters equal $(\alpha, \delta, \lambda) = (1, 1, 1.5)$. 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 results 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

$$\begin{aligned}\alpha^m(10) - \alpha^m(1) \;\; = \;\; & \alpha^{LA}(10) - \alpha^{LA}(1) \\ & +[(\alpha^{LA,PW}(10) - \alpha^{LA,PW}(1)) - (\alpha^{LA}(10) - \alpha^{LA}(1))] \\ & +[(\alpha^m(10) - \alpha^m(1)) - (\alpha^{LA,PW}(10) - \alpha^{LA,PW}(1))], \qquad (31)\end{aligned}$$

where $\alpha^{LA}(j)$ and $\alpha^{LA,PW}(j)$ are, respectively, the alpha predicted by the model for decile j when investors exhibit only loss aversion, and only loss aversion and probability weighting. 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.[30] The

[30]The decomposition for all 14 anomalies is reported in Table A1 in the Internet Appendix.

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 loss aversion typically predicts an alpha spread of the opposite sign from the empirical alpha spread.

These numbers indicate the importance of diminishing sensitivity. However, they understate the role of probability weighting. In addition to the above analysis, 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, the introduction of probability weighting almost doubles the number of anomalies we can explain. The way to reconcile this 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. Probability weighting lowers the expected return on all anomaly deciles. This means that, for a stock with a prior loss at time 0, many of its 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 failure probability 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 study 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, $\alpha^m(10) - \alpha^m(1)$, for different values of the diminishing-sensitivity parameter α, while keeping the other preference parameters 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 A1 in the Internet Appendix. The figure shows that, by the criterion in (28), the model can help explain the volatility anomaly for a large range of values of the preference parameters.

Our model assumes that, when making their allocation decision at time 0, investors merge their prior gain or loss with their potential future gain or loss and derive utility from the integrated gain or loss. It also assumes that investors engage in narrow framing: they derive utility from stock-level gains and losses. Both of these assumptions have experimental support. However, 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 potential future gain or loss, without merging it with their prior gains; this corresponds to setting $g_i = 0$ 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 we adopted in Section 3, 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 1 is the one studied by Grinblatt and Han (2005) and Li and Yang (2013), namely, that 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 3. The reason is that it features a mechanism similar to that 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: 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.[31]

6.2 Initial holdings and an alternate set of anomalies

In this section, we examine the robustness of our results on two dimensions: to endogenizing investors' initial holdings; and to considering an alternate set of anomalies.

When computing the expected return of a stock i, we take investors' initial holdings of 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

[31] In Section K of the Internet Appendix, we show that our conclusions are also robust to allowing for heterogeneity across investors in their prior gains and losses on a stock.

market at time -1 and take investors' initial holdings of stock i to be the maxima of 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 3; 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 L of the Internet Appendix.

We chose the 23 anomalies in Table 1 because they have received a lot of attention from researchers. However, we want to be sure that the conclusion we draw in Section 5 – that prospect theory is helpful for thinking about stock market anomalies – is not special to this set of anomalies.

After completing the analysis in Section 5, 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 alternate set of 23 anomalies and report the results in Section M 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 anomalies. This is a very similar proportion to that reported in Section 5, where the model explained 14 of the 23 anomalies in Table 1. As such, it reinforces our conclusion that prospect theory is helpful for understanding stock market anomalies.

6.3 Comparison with other prospect theory models

Our model accounts for all the elements of prospect theory – all of 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 1 than our model does.

We put the earlier prospect theory models into three groups and consider each group in turn. 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 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 in principle accommodate such an assumption, and doing so may allow us to explain a wider set of phenomena, but we do not pursue this approach: we want our model to reflect only the elements of prospect theory, in part for discipline, and in part because these elements have far more experimental support than does time-varying loss aversion.

Grinblatt and Han (2005), Barberis and Xiong (2009), and Li and Yang (2013) consider models that account for both loss aversion and diminishing sensitivity, and also for investors' prior gains and losses. Between them, the papers show that these 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 the elements of prospect theory but not for investors' prior gains and losses. Between them, these models make the prediction that an asset's idiosyncratic skewness will be priced, and also that some investors 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 1 than our model. To be clear, none of the previous papers makes predictions, either qualitatively or quantitatively, 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 the preference parameter values $(\alpha, \delta, \lambda) = (1, 1, 1.5)$ – explains only three of the 23 anomalies. A model that, like Grinblatt and Han (2005), Barberis and Xiong (2009), and Li and Yang (2013), features loss aversion and diminishing sensitivity but not probability weighting – this corresponds to the parameter values $(\alpha, \delta, \lambda) = (0.7, 1, 1.5)$ – can explain just eight of the 23 anomalies. Finally, as noted in the previous section, a model that, like Barberis and Huang (2008a), Ingersoll (2014), or Barberis, Mukherjee, and Wang (2016), captures all 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 the elements of prospect theory *and* account for prior gains and losses that we can explain a full 14 of the 23 anomalies.

6.4 Time-variation in alphas and interaction effects

In Section 5, we used 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 N of the Internet Appendix, we show that it can. We divide the full sample into four equal subperiods and use the procedure of Section 4 to compute, separately for each subperiod, the volatility, 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. It attributes this variation to changes in stocks' volatility, skewness, and gain overhang from one subperiod to the next.

In Section P of the Internet Appendix, we consider another aspect of time variation in anomaly alphas. We show that the anomalies where 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 where the model performs poorly. One interpretation of this is that the anomalies where the model performs well have a stronger theoretical foundation, one rooted in prospect theory; this in turn explains their better out-of-sample performance.

The model can also help explain three interaction effects related to the anomalies in Table 1. 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 stock 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 Q of the Internet Appendix, we provide formal support for this claim: we show that our model can help explain all three interaction patterns.

6.5 Prospect theory and real-world investors

The experimental evidence that motivates the elements of prospect theory is drawn primarily from studies of ordinary individuals. 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 note three points 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 individuals, despite being a minority, trade in a correlated way, according to prospect theory, thereby exerting demand pressure on stocks; and if institutional investors are not able to fully absorb this demand. There is reason to think that this will be the case because of the particular type of mispricing that prospect theory generates, namely, overpricing of volatile, skewed stocks. 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 involves 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. The 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 et al. (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 et al. (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 et al. (2016) find evidence of loss aversion in the trading of 151 financial professionals. Moreover, studies of option prices – assets that are traded by more sophisticated investors – exhibit price patterns that are 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 will exhibit the elements of prospect theory in their decision-making. Kachelmeier and Shehata (1992) and Post et al. (2008) find support for 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, we find that, for the value of $\widehat{b}_0$ that we assume in our calculations – recall that $\widehat{b}_0$ 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.

7 Conclusion

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

8 References

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

Abdellaoui, M., Bleichrodt, H., and H. Kammoun (2013), "Do Financial Professionals Behave

According to Prospect Theory? An Experimental Study," *Theory and Decision* 74, 411-429.

An, L., Wang, H., Wang, J., and J. Yu (2020), "Lottery-related Anomalies: The Role of Reference-dependent Preferences," *Management Science* 66, 473-501.

Andrikogiannopoulou, A. and F. Papakonstantinou (2020), "History-dependent Risk Preferences: Evidence from Individual Choices and Implications for the Disposition Effect," *Review of Financial Studies* 33, 3674-718.

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

Baele, L., Driessen, J., Ebert, S., Londono, J., and O. Spalt (2019), "Cumulative Prospect Theory, Option Returns, and the Variance Premium," *Review of Financial Studies* 32, 3667-3723.

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

Barber, B. and T. Odean (2000), "Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," *Journal of Finance* 55, 773-806.

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

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

Barberis, N. and M. Huang (2008b), "The Loss Aversion / Narrow Framing Approach to the Equity Premium Puzzle," in R. Mehra (Ed.), *Handbook of the Equity Risk Premium*, Elsevier.

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

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

Barberis, N. and W. Xiong (2009), "What Drives the Disposition Effect? An Analysis of a

Long-standing Preference-based Explanation," *Journal of Finance* 64, 751-784.

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

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

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

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

Boyer, B. and K. Vorkink (2014), "Stock Options as Lotteries," *Journal of Finance* 69, 1485-1527.

Calvet, L., Campbell, J.Y., and P. Sodini (2007), "Down or Out: Assessing the Welfare Costs of Household Investment Mistakes," *Journal of Political Economy* 115, 707-747.

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

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

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

Coval, J. and T. Shumway (2005), "Do Behavioral Biases Affect Prices?," *Journal of Finance* 60, 1-34.

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

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

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

Haigh, M. and J. List (2005), "Do Professional Traders Exhibit Myopic Loss Aversion? An Experimental Analysis," *Journal of Finance* 60, 523-534.

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

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

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

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

Kachelmeier, S., and M. Shehata (1992), "Examining Risk Preferences Under High Monetary Incentives: Experimental Evidence from the People's Republic of China," *American Economic Review* 82, 1120-41.

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

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

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

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

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

Novy-Marx, R. and M. Velikov (2016), "A Taxonomy of Anomalies and Their Trading Costs," *Review of Financial Studies* 29, 104-147.

Polkovnichenko, V. and F. Zhao (2013), "Probability Weighting Functions Implied in Option Prices," *Journal of Financial Economics* 107, 580-609.

Pope, D. and M. Schweitzer (2011), "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes," *American Economic Review* 101, 129-57.

Post, T., Van den Assem, M., Baltussen, G., and R. Thaler (2008), "Deal or No Deal? Decision-making Under Risk in a Large-Payoff Game Show," *American Economic Review* 98, 38-71.

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

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

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

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

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

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

Table 1. Stock market anomalies.

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

Table 2. Empirical properties of anomaly deciles. The first column lists 23 anomalies; the acronyms are defined in Table 1. 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.

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
OSC	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 3. The first column lists 23 anomalies; the acronyms are defined in Table 1. 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 alpha spread between the two deciles. The horizontal lines in the table divide the anomalies into three groups: the 14 anomalies where the model performs well (top group); the two anomalies where the model does not make a strong prediction (middle group); and the seven anomalies where 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
OSC	-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 4. The table reports the average absolute monthly pricing error across the 23 anomalies for the prospect theory model of Section 3 and for five factor models: the CAPM; the Fama-French three-factor model; the Carhart four-factor model; the Fama-French five-factor model; and a six-factor model that augments the five-factor model with a momentum factor. For a given factor model, we compute the alphas of the 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.

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

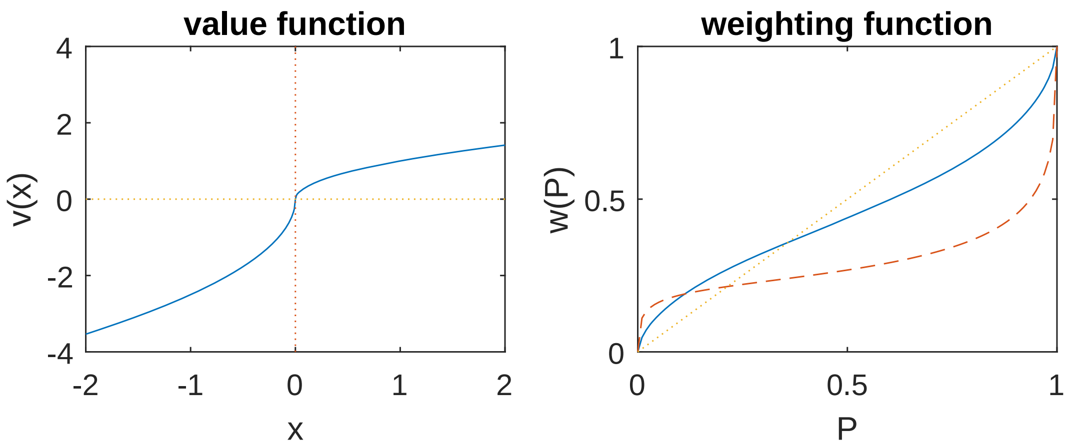

Figure 1. The left panel plots the value function proposed by Tversky and Kahneman (1992) as part of their cumulative prospect theory, namely, $v(x) = x^{\alpha}$ for $x \geq 0$ and $v(x) = -\lambda(-x)^{\alpha}$ for $x < 0$, for $\alpha = 0.5$ and $\lambda = 2.5$. The right panel plots the probability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

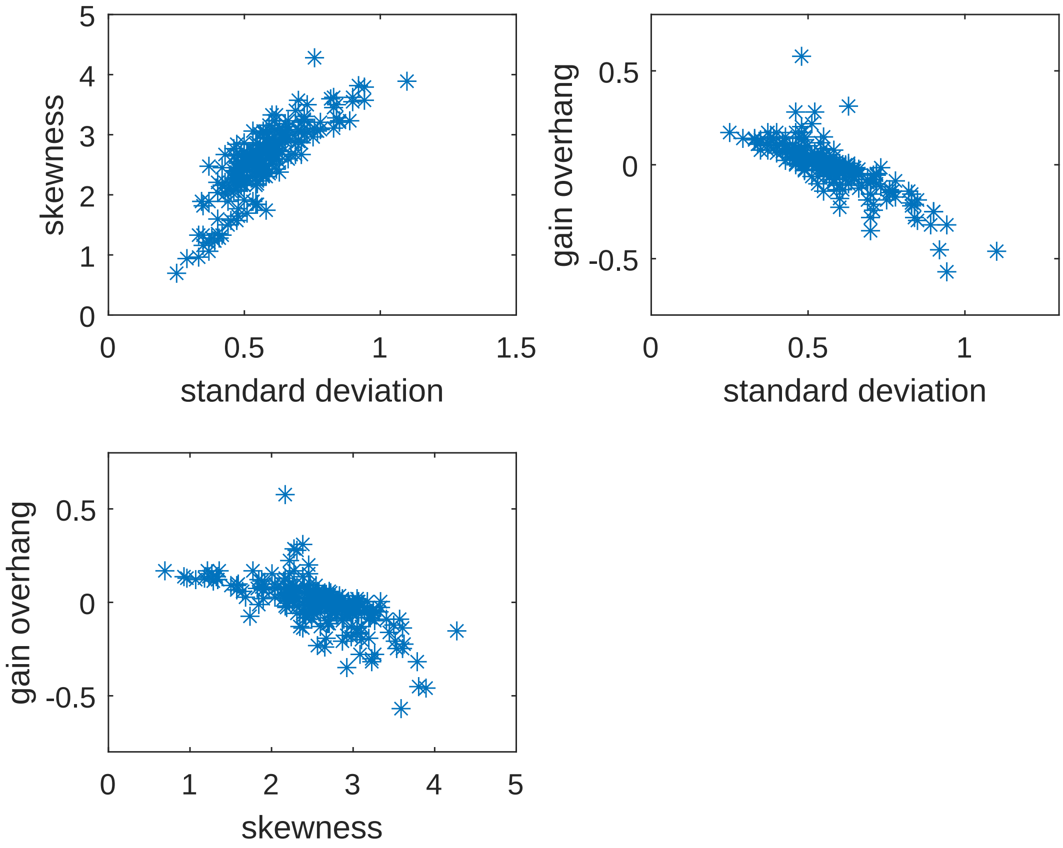

Figure 2. Each graph plots 230 asterisks, where each asterisk corresponds to one of 10 deciles for one of 23 anomalies. In the top-left graph, a given asterisk that corresponds to some anomaly decile marks the standard deviation of returns and the return skewness of the typical stock in that anomaly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

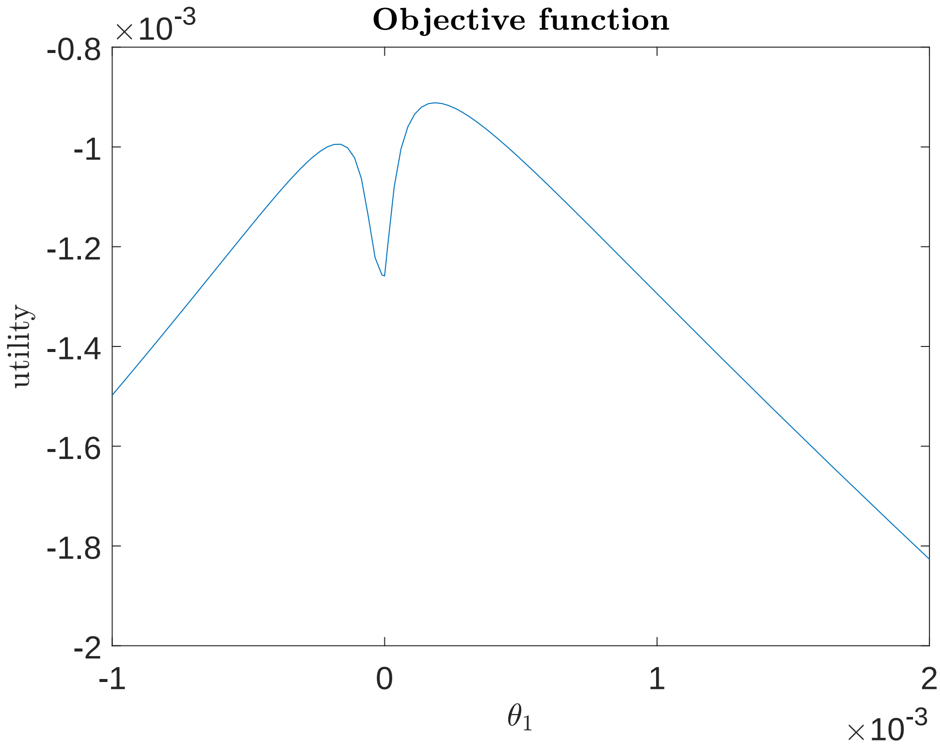

Figure 3. The graph shows that investors have identical holdings of each stock in momentum decile 1. The solid line plots the value of an investor's objective function in equilibrium as a function of θ_1, the (scaled) fraction of the investor's portfolio allocated to stock 1, which belongs to momentum 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}.

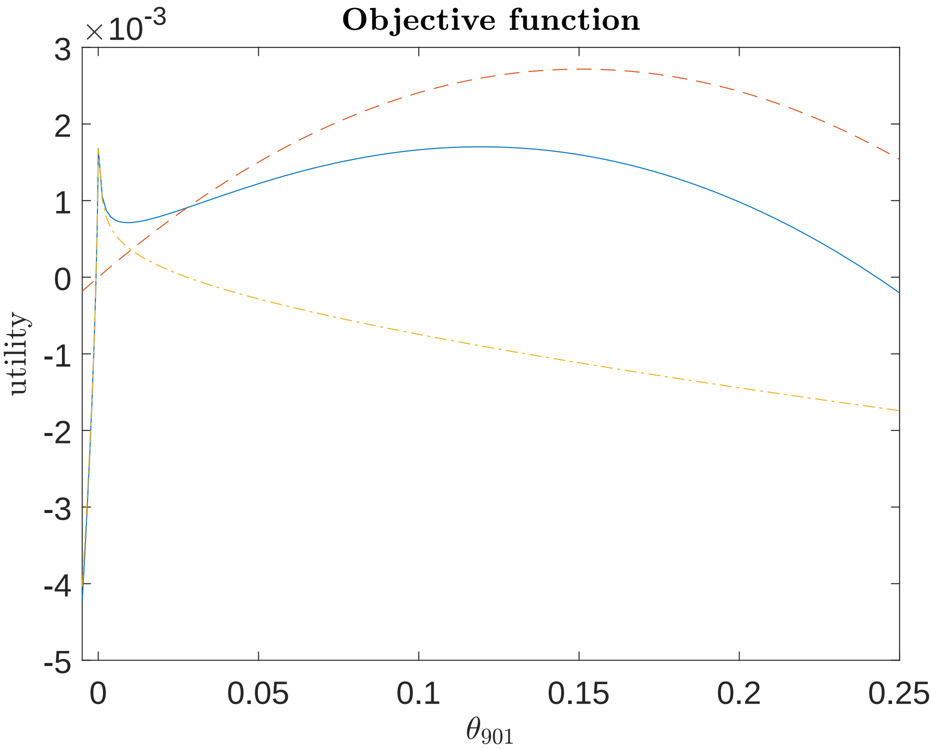

Figure 4. The graph shows that investors have heterogeneous holdings of each stock in momentum decile 10. The solid line plots the value of an investor's objective function in equilibrium as a function of θ_{901}, the (scaled) fraction of the investor's portfolio allocated to stock 901, which belongs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

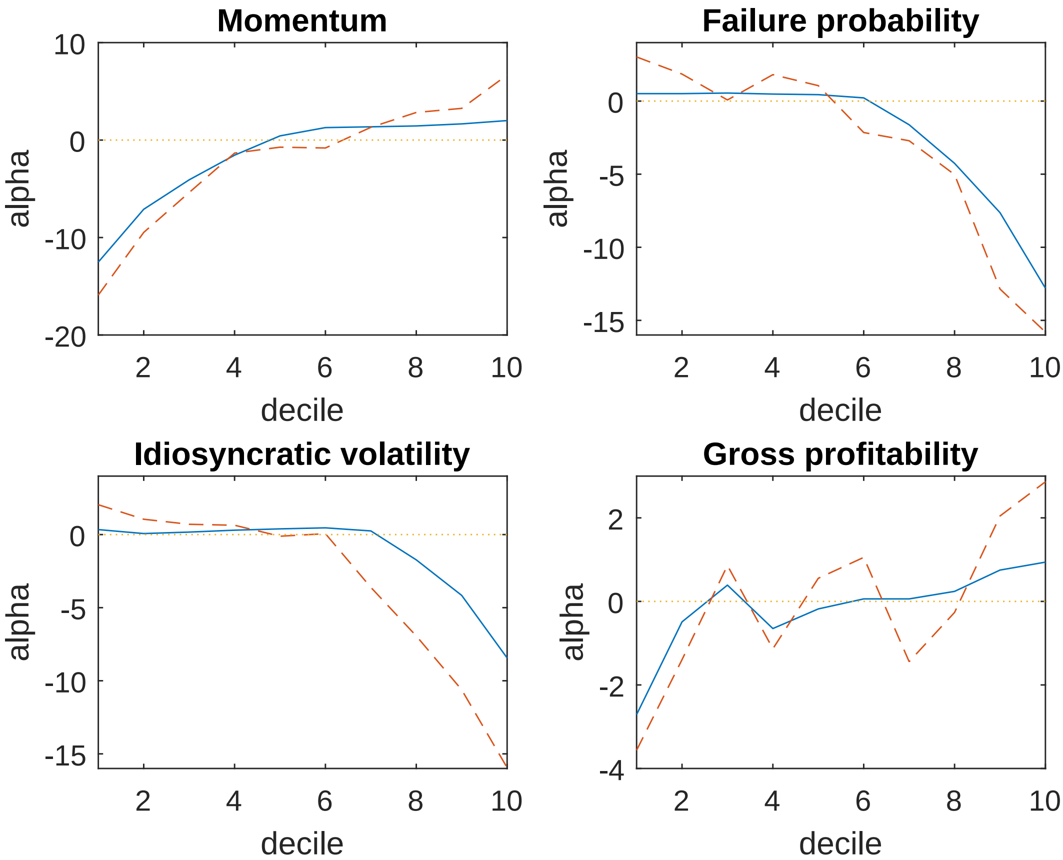

Figure 5. The graphs present results for four of the anomalies where the prospect theory model performs well: the momentum, failure probability, idiosyncratic volatility, and gross profitability anomalies. The dashed lines in the graphs plot the historical annual alpha of each anomaly decile. The solid lines plot the alphas predicted by the model.

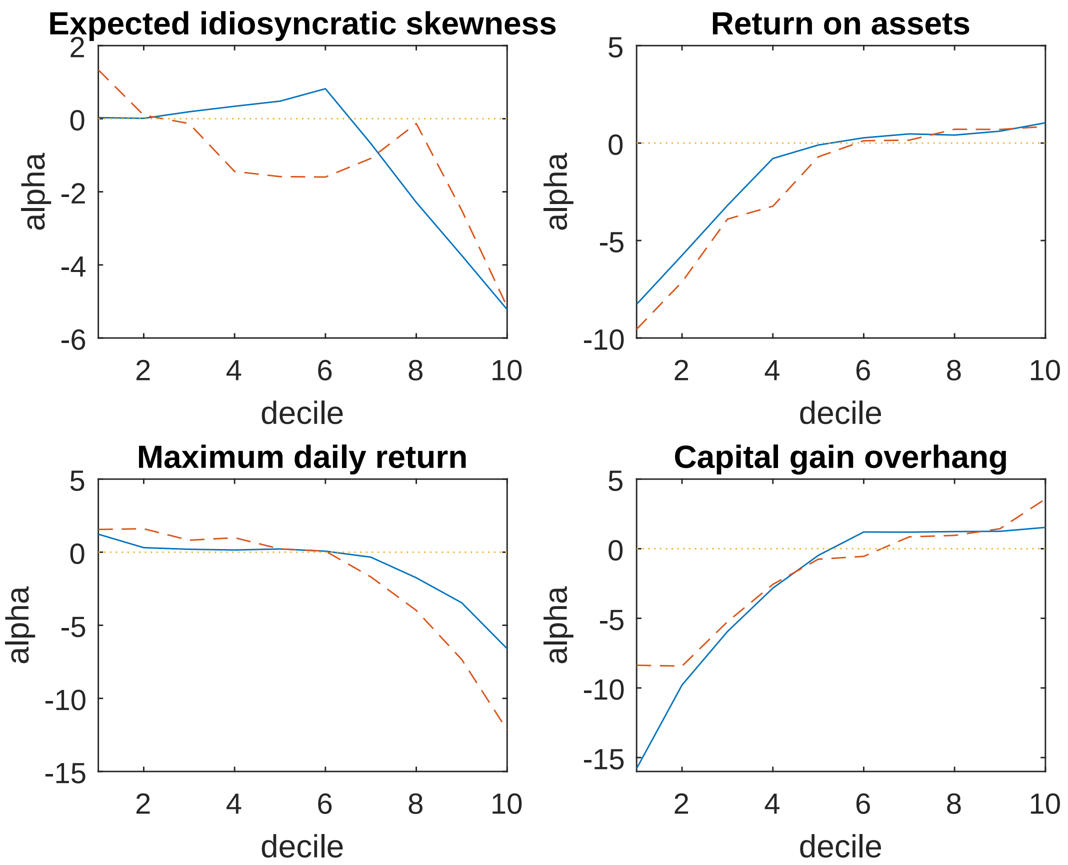

Figure 6. The graphs present results for four of the anomalies where the prospect theory model performs well: the expected idiosyncratic skewness, return on assets, maximum daily return, and capital gain overhang anomalies. The dashed lines in the graphs plot the historical annual alpha of each anomaly decile. The solid lines plot the alphas predicted by the model.

INTERNET APPENDIX

A. The probability weighting terms

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

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

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

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

where the probability density function $p(R_i)$ is given in (12). Similarly,

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

B. The form of the utility function in the case of shorting

In the model, investors are allowed to sell short. If an investor sells short risky asset i, so that $\Theta_i < 0$ and hence $\theta_i < 0$, the expression in equation (11) is modified as

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

In the same way, the prospect theory terms in equation (20) become

$$\begin{aligned} & \widehat{b}_0 \int_{-\infty}^{R_f - \theta_{i,-1} g_i/\theta_i} (\theta_i(R_i - R_f) + \theta_{i,-1} g_i)^\alpha dw(P(R_i)) \\ & + \lambda \widehat{b}_0 \int_{R_f - \theta_{i,-1} g_i/\theta_i}^{\infty} (\theta_i(R_f - R_i) - \theta_{i,-1} g_i)^\alpha dw(1-P(R_i)). \end{aligned}$$

C. The Generalized Hyperbolic (GH) Skewed t distribution

In our model, the vector of asset returns $\widetilde{R} = (\widetilde{R}_1, \ldots, \widetilde{R}_N)'$ follows an N-dimensional GH skewed t distribution. The density function for this distribution is

$$\begin{aligned} p(R) &= \frac{2^{1-\frac{\nu+N}{2}}}{\Gamma(\frac{\nu}{2})(\pi\nu)^{\frac{N}{2}}|S|^{\frac{1}{2}}} . \frac{K_{\frac{\nu+N}{2}}(\sqrt{(\nu+(R-\mu)'S^{-1}(R-\mu))\zeta'S^{-1}\zeta}) \exp((R-\mu)'S^{-1}\zeta)}{\left(\sqrt{(\nu+(R-\mu)'S^{-1}(R-\mu))\zeta'S^{-1}\zeta}\right)^{-\frac{\nu+N}{2}}(1+(R-\mu)'S^{-1}(R-\mu)/\nu)^{\frac{\nu+N}{2}}}, \\ & \quad \text{for } \zeta \neq 0 \\ p(R) &= \frac{\Gamma(\frac{\nu+N}{2})}{\Gamma(\frac{\nu}{2})(\pi\nu)^{\frac{N}{2}}|S|^{\frac{1}{2}}} .(1+(R-\mu)'S^{-1}(R-\mu)/\nu)^{-\frac{\nu+N}{2}}, \text{ for } \zeta = 0, \end{aligned} \tag{32}$$

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

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

$$\mu + \frac{\nu}{\nu-2}\zeta. \tag{33}$$

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

D. Rescaling the decision problem

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

$$\begin{aligned} &\theta_i(\mu_i + \frac{\nu\zeta_i}{\nu-2} - R_f) - \frac{\widehat{\gamma}}{2}(\theta_i^2\sigma_i^2 + 2\sum_{j\neq i}\sigma_{ij}\theta_i\theta_{M,j}) \\ &-\lambda\widehat{b}_0\int_{-\infty}^{R_f-\theta_{i,-1}g_i/\theta_i}(\theta_i(R_f-R_i)-\theta_{i,-1}g_i)^\alpha dw(P(R_i)) \\ &-\widehat{b}_0\int_{R_f-\theta_{i,-1}g_i/\theta_i}^{\infty}(\theta_i(R_i-R_f)+\theta_{i,-1}g_i)^\alpha dw(1-P(R_i)), \end{aligned} \tag{34}$$

where

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

It follows that if Θ_i maximizes (16), then $\theta_i = \Theta_i/\Theta_{M,R}$ maximizes (34), and conversely that if θ_i maximizes (34), then $\Theta_i = \theta_i \Theta_{M,R}$ maximizes (16). Maximizing (16) is therefore equivalent to maximizing (34).

The rescaling also allows us to simplify the variance term in the first row of (34). Specifically, the quantity

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

can be rewritten as

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

where $\widetilde{R}_M$ is the return on the market portfolio of all risky assets, σ_M is the standard deviation of this return, and β_i is the beta of asset i. Substituting this into (34) leads to the expression in (20).

E. Procedure for computing expected returns

The equilibrium structure we consider – a bounded-rationality equilibrium with heterogeneous holdings – consists of a location vector $(\mu_1, \ldots, \mu_N)'$ such that, for each i, the expression in (20) has either a unique global maximum at $\theta_i^* = \theta_{M,i}$ or else two global maxima at θ_i^* and θ_i^{**} with $\theta_i^* < \theta_{M,i} < \theta_i^{**}$.

We now explain how we compute this heterogeneous-holdings equilibrium. For a given risky asset i, we first check whether investors have identical holdings in that asset, in other words, whether they all hold the same per-capita market supply $\theta_{M,i}$. To do this, we take the first derivative of (20), substitute in $\theta_i = \theta_{M,i}$, and set the resulting expression to 0. This gives[32]

$$\begin{aligned} 0 = {} & (\mu_i + \frac{\nu\zeta_i}{\nu-2} - R_f) - \widehat{\gamma}\beta_i\sigma_M^2 \\ & -\alpha\lambda\widehat{b}_0 \int_{-\infty}^{R_f - \theta_{i,-1}g_i/\theta_i} (\theta_{M,i}(R_f - R_i) - \theta_{i,-1}g_i)^{\alpha-1}(R_f - R_i)dw(P(R_i)) \\ & -\alpha\widehat{b}_0 \int_{R_f - \theta_{i,-1}g_i/\theta_i}^{\infty} (\theta_{M,i}(R_i - R_f) + \theta_{i,-1}g_i)^{\alpha-1}(R_i - R_f)dw(1 - P(R_i)). \end{aligned} \tag{35}$$

[32]As discussed in Section 4, we take $\theta_{i,-1} = \theta_{M,i}$, which further simplifies (35).

We then check whether, for the μ_i that solves (35), the function in (20) has a global maximum at $\theta_i = \theta_{M,i}$, as opposed to only a local maximum or a local minimum. If $\theta_i = \theta_{M,i}$ indeed corresponds to a global maximum, then all investors have identical holdings of risky asset i, each holding the per-capita supply of the asset, namely $\theta_{M,i}$. If the function in (20) does not have a global maximum at this θ_i, then, in equilibrium, investors do not have identical holdings of asset i. We must instead look for an equilibrium with heterogeneous holdings.

To find an equilibrium with heterogeneous holdings of asset i, we search for a value of μ_i such that the maximum value of the function in (20) in the range $(-\infty, \theta_{M,i})$, attained at $\theta_i = \theta_i^*$, say, is equal to the maximum value of the function in the range $(\theta_{M,i}, \infty)$, attained at $\theta_i = \theta_i^{**}$. If we find such a μ_i, then there is an equilibrium where investors have heterogeneous holdings of the asset, with some investors allocated to $\theta_i = \theta_i^*$ and others allocated to $\theta_i = \theta_i^{**}$. The value of μ_i for which this holds is typically close to the value of μ_i that solves the first-order condition (35). Therefore, if we find that, for the value of μ_i that solves (35), the function in (20) does not have a global maximum at $\theta_i = \theta_{M,i}$, we search in the neighborhood of that μ_i for an equilibrium with heterogeneous holdings.

F. Stock market anomalies

Here, we list the predictor variable associated with each of the 23 anomalies that we study.

IVOL. Idiosyncratic volatility. Standard deviation of the residuals from a firm-level regression of daily stock returns on the daily Fama-French three factors using data from the past month. See Ang et al. (2006).

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

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

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

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

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

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

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

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

ACC. Accruals. See Sloan (1996).

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

PROF. Gross profitability. Measured as revenue minus cost of goods sold at time t, divided by assets at time $t-1$. See Novy-Marx (2013).

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

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

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

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

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

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

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

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

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

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

CGO. Capital gain overhang. The average percentage capital gain since purchase across investors in a stock, computed as in Grinblatt and Han (2005).

G. Attitudes to gambles

When we implement the model, we set investors' scaled portfolio risk aversion $\widehat{\gamma}$ and scaled weight on the prospect theory term $\widehat{b}_0$ to produce reasonable values for two market data points: the equity premium and the level of under-diversification in household portfolios. It is useful to check that the values of these parameters, namely $(\widehat{\gamma}, \widehat{b}_0) = (0.6, 0.6)$, also generate sensible attitudes to small- and large-scale gambles.

Suppose that, at time 0, an investor in our economy is endowed with a symmetric gamble: a 50:50 bet to gain $\$M$ or lose $\$M$. What is the largest cash "premium" the investor would be willing to pay to avoid this gamble? For simplicity, we assume that the premium payment and gamble outcomes occur at time 1. If the investor takes the gamble, his time 0 utility changes by

$$-\frac{\gamma}{2}M^2 - b_0(\lambda - 1)w\left(\frac{1}{2}\right)M^{\alpha}, \tag{36}$$

where $\gamma = \widehat{\gamma}/W_0\Theta_{M,R}$ and $b_0 = \widehat{b}_0/W_0^{\alpha-1}\Theta_{M,R}^{\alpha-1}$. If he instead pays a premium π at time 1, his utility changes by

$$-\pi - b_0\lambda\pi^{\alpha}. \tag{37}$$

The maximum amount he is willing to pay to avoid the gamble is therefore the value of π that equates (36) and (37).

Using equations (36) and (37), we confirm that, for $\widehat{\gamma} = \widehat{b}_0 = 0.6$, the model satisfies the restrictions on attitudes to small- and large-scale gambles proposed by Barberis and Huang (2008b).

We can also examine investor attitudes to the gamble $(-\$M, \frac{1}{2}; \$X, \frac{1}{2})$. If an investor takes this gamble, his utility changes by

$$\frac{X-M}{2} - \gamma\frac{(X+M)^2}{8} - b_0 w\left(\frac{1}{2}\right)(\lambda M^\alpha - X^\alpha). \tag{38}$$

By setting this expression equal to zero, we can compute how high X needs to be, for a given M , for the investor to be indifferent to the gamble. We confirm that, for $\widehat{\gamma} = \widehat{b}_0 = 0.6$, the model captures the restrictions on attitudes to small- and large-scale gambles proposed by Barberis, Huang, and Thaler (2006): that, for a wide range of wealth levels, the investor rejects the gamble $(-\$500, \frac{1}{2}; \$550, \frac{1}{2})$ and accepts the gamble $(-\$10000, \frac{1}{2}; \$20,000,000, \frac{1}{2})$.

Finally, we confirm, using a similar methodology, that for $\widehat{\gamma} = \widehat{b}_0 = 0.6$, the model captures the basic experimental evidence in Kahneman and Tversky (1979) that motivates prospect theory, including the rejection of $(-\$100, \frac{1}{2}; \$110, \frac{1}{2})$; the preference for $(-\$1000, \frac{1}{2})$ over $-\$500$; and the preference for $(\$5000, 0.001)$ over \$5.

H. Additional details about the equilibrium structure

Here, we present some additional information about the equilibrium structure illustrated in Figures 3 and 4. We explain why, for some stocks – for example, for stocks in momentum decile 10 – there is no equilibrium with identical holdings for these stocks. We also explain why, for other stocks – for example, stocks in momentum decile 1 – there is no equilibrium with heterogeneous holdings for these stocks. Finally, we contrast the heterogeneous holdings that arise in our model with those in the model of Barberis and Huang (2008a).

No equilibrium with identical holdings for momentum decile 10 stocks. The solid line in Figure 4 shows that, for any stock in momentum decile 10 – for example, for stock 901 – investors have heterogeneous holdings in the stock. To see why there is no equilibrium with identical holdings in this stock, consider Figure A2. The solid line in this figure is the same as the solid line in Figure 4: it plots the objective function in (20) for the location parameter $\mu_{901} = 0.5853$. The dashed line in the figure plots the function in (20) for $\mu_{901} = 0.588$. For this higher value of μ_{901}, the function has a unique global maximum at $\theta_{901} = 0.14 > \theta_{M,901} = 7.26 \times 10^{-4}$. Since demand for stock 901 exceeds supply at this value of μ_{901}, it appears that, to clear the market, we need to lower the value of μ_{901}. However, as we do so, the value of θ_{901} at which the function attains its maximum jumps discontinuously downward: the dash-dot line, which plots the objective function for a slightly lower value of μ_{901}, namely 0.582, shows that the unique maximum is now at $\theta_{901} = 0 < \theta_{M,901}$. As such, there is no value of μ_{901} such that the function in (20), for $i = 901$, has a unique global maximum at $\theta_{901} = \theta_{M,901}$. Instead, the market clears only by way of the heterogeneous-holdings structure represented by the solid line.

No equilibrium with heterogeneous holdings for momentum decile 1 stocks. For a stock in momentum decile 1 – for example, for stock 1 – the investors in our model have identical holdings. Why do heterogeneous holdings not arise in this case? This stock has a loss at time 0 and, in part because of this, has a low expected return. One could then imagine a heterogeneous-holdings structure of the following form. Some investors choose a positive allocation in the stock: since they are in the convex region of the prospect theory value function, they are keen to take risk. Other investors choose a negative allocation in the stock to exploit its low expected return.

Figure 3 shows that the model comes close to delivering such an equilibrium: the objective function has two local maxima, one in the positive domain and one in the negative domain, which are driven by precisely the forces in the previous paragraph. However, the local maximum in the negative domain has a lower utility level, and so there is a unique global maximum. The reason it has a lower utility level is that, by shorting the stock, the investor is exposing himself to high negative skewness, which he finds aversive.

Can a heterogeneous-holdings structure be achieved by lowering the stock's expected return below the level used in Figure 3? Doing so can indeed give the two local maxima the same utility level. However, the market then no longer clears. As we lower the expected return, the allocation corresponding to the local maximum in the positive domain falls below the market supply. Since the allocations corresponding to the two local maxima are both lower than the market supply, the market no longer clears. In equilibrium, then, the objective function has a unique global maximum and all investors have identical holdings in the stock.

Comparison with the heterogeneous holdings in Barberis and Huang (2008a). An equilibrium with two global maxima also arises in the prospect theory model of Barberis and Huang (2008a). However, the economic forces underlying that equilibrium are different from those that drive the two global maxima in our model. Similar to our model, the model of Barberis and Huang (2008a) features all the elements of prospect theory; however, it differs from our model in that it does not account for investors' prior gains and losses and assumes broad, rather than narrow, framing. Barberis and Huang (2008a) show that, if a positively-skewed security is introduced into the economy, it earns a negative expected excess return. The two global maxima correspond to a zero allocation in this skewed asset and a large undiversified position in it. The economic force underlying the zero allocation is the skewed asset's low expected return; the economic force underlying the high allocation is that, by adding a significant position in the skewed asset to his portfolio, an investor can make his portfolio return more positively skewed, something that, due to probability weighting, he finds attractive.

The mechanism behind the equilibrium structure in Barberis and Huang (2008a) does not arise in our model because it hinges on the broad framing assumption; we assume narrow framing. Importantly, the mechanism in our paper does not arise in Barberis and Huang (2008a). The lower optimum in Figure 4 reflects investors' desire to preserve a prior gain. This does not apply in the model of Barberis and Huang (2008a) because that model does not account for investors' prior gains and losses.

I. Earnings announcement analysis

The model performs poorly for some anomalies. One possible reason for this is that some anomalies – we label them "belief-based" anomalies – may be driven not by investor risk attitudes but by incorrect beliefs about firms' future prospects. Below, we identify the anomalies that are more likely to be belief-based by computing, for each anomaly, the fraction of the return spread between deciles 1 and 10 that is earned in the five days around firms' earnings announcement dates – in other words, on days when investors' erroneous beliefs would be corrected by realized earnings; we call an anomaly belief-based if this fraction is large. We conjecture that the prospect theory model will do poorly for anomalies that are belief-based and well for anomalies that are not.

We find strong support for this conjecture. For five of the seven anomalies where the model performs poorly – the value, investment, long-term reversal, accrual, and asset growth anomalies – a large fraction of the anomaly return spread comes around earnings announcement dates. Meanwhile, for several of the anomalies where the model performs well, little or none of the overall spread comes around these dates.

Our analysis is summarized in Table A2. Column A reports the monthly return spread for each anomaly – the return difference between decile 10 and decile 1 – in the full sample from 1963 to 2015. Column B reports the return spread predicted by our model. Column C quantifies the performance of the model – how well it explains an anomaly – as the ratio of column B to column A: a positive number indicates that the model is helpful for thinking about an anomaly; a negative number means that it is not.

Columns D to H summarize the earnings announcement results. Due to data availability, we conduct this analysis in a shorter sample which starts in the second quarter of 1983. Column D reports the equal-weighted return spread of each anomaly in this sample period. Column E reports, for each anomaly, the difference between decile 10 and decile 1 in the cumulative abnormal returns of the underlying firms in the five days around their earnings announcements – put simply, the part of the return spread that comes around earnings announcements. To do this calculation, we obtain earnings announcement dates from the quarterly Compustat and Institutional Brokers' Estimate System (IBES). When the announcement dates in the two databases are not the same, we follow DellaVigna and Pollett (2009) and use the earlier of the two dates.

Column F reports the ratio of column E to column D, in other words, the fraction of the return spread that comes in the days around earnings announcements; the higher this number, the more likely that the anomaly is belief-based. We note two outliers in this column, for SIZE and EISKEW; the extreme values for these anomalies are due to their low return spreads in the post-1983 sample period. In column G, we winsorize the entries for these two anomalies, although our results do not depend on this. Finally, in column H, we construct a dummy variable which takes the value 1 if the number in column G exceeds 0.5, in other words, if more than half of an anomaly's return spread comes around earnings announcement dates. Using our terminology, a value of 1 indicates an anomaly that is likely to be belief-based.

Our conjecture is that the prospect theory model will perform better for anomalies that are not belief-based, and worse on anomalies that are. Put differently, the numbers in column C will be negatively correlated with the numbers in column G. We find strong support for this conjecture. The correlation is -0.571, which differs from zero in a highly significant way.

J. Beliefs about skewness and the value and size anomalies

In Section I of the Internet Appendix, we discuss one reason why the model performs poorly for some anomalies: in most cases, these anomalies appear to be driven primarily by incorrect beliefs about firms' future performance rather than by risk attitudes of the kind captured by prospect theory.

There is another possible explanation for the model's poor performance on some anomalies, one that applies within the context of the model. When computing the model's predictions for anomaly alphas – the predictions in Table 3 and Figures 5-6 – we assumed that investors have sensible beliefs about the key model inputs: stocks' volatility, skewness, and gain overhang. However, investors may have incorrect beliefs about these quantities, which may explain the model's poor performance on some anomalies.

We examine this idea in the context of the value and size anomalies. Empirically, value stocks are more volatile than growth stocks, more skewed than growth stocks, and trade at a larger loss than growth stocks. As a consequence, our model fails to explain the value anomaly: the greater volatility of value stocks leads the model to predict a higher average return on them, but their higher skewness and more negative gain overhang leads the model to predict a lower average return on them, and the latter effect dominates.

However, real-world investors may think that it is growth stocks that are more highly skewed than value stocks. If the model takes account of such incorrect beliefs, it will predict a higher (lower) average return on value (growth) stocks, bringing its prediction more in line with the empirical facts. We now examine just how incorrect beliefs about skewness would need to be to generate a substantially positive value premium.

The empirical skewness levels of the typical stocks in the ten value anomaly deciles are given by the vector

$$skew\_value = [1.85; 1.92; 2.05; 2.43; 2.05; 2.2; 2.34; 2.31; 2.97; 2.66];$$

value stocks (decile 10) have higher skewness than growth stocks (decile 1). We now construct a vector of incorrect beliefs about the skewness of the typical stocks in the ten anomaly deciles, namely

$$skew\_value\_incorrect = skew\_value + k\ [1; \frac{7}{9}; \frac{5}{9}; \frac{3}{9}; \frac{1}{9}; -\frac{1}{9}; -\frac{3}{9}; -\frac{5}{9}; -\frac{7}{9}; -1],$$

where $k > 0$. This construction leads investors to over-estimate the skewness of growth stocks and to under-estimate the skewness of value stocks while leaving their belief about the skewness of the average stock unaffected.

We search for the lowest value of k that leads the model to generate a substantially positive value premium, in other words, one that, by the criterion in (28), would lead us to say that the model can help explain the value premium. We find that this value of k is approximately 1.8. For this k, investors believe that the skewness of decile 1 growth stocks is 3.65 and that the skewness of decile 10 value stocks is 0.86. If investors have sufficiently distorted perceptions of stocks' skewness, then, the model can generate a substantially positive value premium. We leave it to future research to determine whether real-world investors hold such distorted beliefs about skewness.

We conduct a similar exercise for the size anomaly. Empirically, small-cap stocks are much more highly skewed than large-cap stocks, and this is a major reason why the model incorrectly predicts a negative size premium. However, it is possible that investors under-estimate the skewness of small-cap stocks, and that, once we take this into account, the model will predict a positive size premium. Here, we investigate how distorted investors' perception of the relative skewness of small-cap and large-cap stocks would need to be to generate a positive size premium.

The empirical skewness levels of the typical stocks in the ten size anomaly deciles are given by

$$skew\_size = [4.27; 1.89; 1.65; 1.51; 1.28; 1.07; 1.17; 0.96; 0.93; 0.69];$$

small-cap stocks (decile 1) have much higher skewness than large-cap stocks (decile 10). We now construct a vector of incorrect beliefs about the skewness of the typical stocks in the ten anomaly deciles, namely

$$skew\_size\_incorrect = skew\_size + k\ [1; \frac{7}{9}; \frac{5}{9}; \frac{3}{9}; \frac{1}{9}; -\frac{1}{9}; -\frac{3}{9}; -\frac{5}{9}; -\frac{7}{9}; -1],$$

where $k < 0$. This construction leads investors to under-estimate the skewness of small-cap stocks and to over-estimate the skewness of large-cap stocks while leaving their belief about the skewness of the average stock unaffected.

We search for the highest value of k that leads the model to predict a positive size premium of 1.5% – a size premium that, by the criterion in (28), would lead us to say that the model can help explain the size anomaly. We find that this value of k is approximately -1.9. For this k, investors believe that the skewness of decile 1 small-cap stocks is 2.37 and that the skewness of decile 10 large-cap stocks is 2.59. If investors have sufficiently distorted perceptions of stocks' skewness – if they perceive small-cap and large-cap stocks to have fairly similar skewness levels – then the model can generate a substantially positive size premium. Again, we leave it to future research to determine whether real-world investors hold such distorted beliefs.

K. Investor heterogeneity

To keep the model tractable, we assume that, at time 0, all investors have the same prior gain g_i in stock i. What happens if we allow for heterogeneity across investors in their prior gains? While it is not possible to study heterogeneity in a fully general way, our model does allow us to consider simple forms of heterogeneity. The outcome of this exercise is reassuring: the model predictions are quite robust to heterogeneity in investors' prior gains.

Take, for example, momentum decile 10. The gain overhang of the typical stock in this decile is 30.98%. Accordingly, in our main analysis, we assume that, for this stock, all investors have the same prior gain of 30.98%, an assumption that leads the model to predict an expected return of 8.54% for the stock. Suppose instead that, for this stock, half of the investors have a gain of $30.98 - 10 = 20.98\%$ while the other half have a gain of $30.98 + 10 = 40.98\%$. There is then an equilibrium in which the objective function for the 40.98-gain investors has a single global optimum at $\theta_i^* = 1.19 \times 10^{-4} < \theta_{M,i} = 7.26 \times 10^{-4}$, while the objective function for the 20.98-gain investors has two global optima at $\theta_i^* = 6.02 \times 10^{-5} < \theta_{M,i}$ and $\theta_i^{**} = 0.109 > \theta_{M,i}$. In equilibrium, the stock has an expected return of 8.42%. As such, heterogeneity affects the expected return, but in a minor way.

For momentum decile 1, where investors have identical holdings of each stock, the impact of heterogeneity is even smaller. The gain overhang of the typical stock in this decile is -45.02%. In our main analysis, we assume that all investors have the same prior gain of -45.02% in the stock, and this leads to a predicted expected return of -6.80%. Suppose that we instead assume that half of the investors have a gain of $-45.02 - 10 = -55.02\%$ while the other half have a gain of $-45.02 + 10 = -35.02\%$. In this case, there is an equilibrium in which the objective function of the -55.02-gain investors has a single global optimum at $\theta_i^* = 2.26 \times 10^{-4} > \theta_{M,i} = 1.86 \times 10^{-4}$, while that for the -35.02-gain investors has a single global optimum at $\theta_i^* = 1.46 \times 10^{-4} < \theta_{M,i}$. The predicted expected return is *the same* as in the equilibrium with identical investors, namely -6.80%.

The fact that heterogeneity in g_i has a relatively small impact on our results immediately implies that heterogeneity in $\Theta_{i,-1}$, investors' time -1 allocation to risky asset i, will also have a small impact: equation (10) shows that allowing for heterogeneity in Θ_{-1} is mathematically equivalent to allowing for heterogeneity in g. This helps explain why, as discussed in Section L of the Internet Appendix, our conclusions are robust to endogenizing investors' initial holdings.

L. Endogenizing investors' initial holdings

When we compute the model's predicted alphas, we set investors' initial stock holdings at time -1 equal to stocks' market weights: $\theta_{i,-1} = \theta_{M,i}$ for all i. Are our conclusions robust to endogenizing the initial holdings? The answer turns out to be "yes".

To illustrate the results, we take an anomaly that the model can help explain, but not by a wide margin: the gross profitability anomaly. For this anomaly, conditions (28) are satisfied, but only narrowly so. We want to see if, for this anomaly, conditions (28) continue to hold even after we endogenize the initial holdings.

We start with anomaly decile 1. To endogenize the initial holdings of a stock i in this decile, we search for a value of the location parameter μ_i that clears the market at time -1

for the objective function in (20) when $g_i = 0$; the last condition captures the fact that, when an investor first buys the stock, he has no prior gain in it. We find that the market-clearing location parameter is $\mu_i = 0.3682$. For this value of μ_i, the objective function in (20) has two global maxima, $\theta_i^* = 0$ and $\theta_i^{**} = 0.0466$. These straddle the market supply of the stock, allowing us to clear the market by allocating some investors to the first maximum and the remaining investors to the second maximum. These, then, are investors' initial holdings.

We now move to time 0. The typical stock in decile 1 of this anomaly has a capital loss at this time. We find that $\mu_i = 0.3682$ no longer clears the market. The reason is that the investors with an initial holding of 0.0466 now have a substantial prior loss; since they are firmly in the convex region of the prospect theory value function, they want to take a large position in the stock – so large that the market no longer clears. Through a manual search, we find that the market now clears for a different location vector, $\mu_i = 0.3407$. For this value of μ_i, the objective function in (20) has a single global maximum at $\theta_i = 0$ when $\theta_{i,-1} = 0$, and a single global maximum at $\theta_i = 0.0466$ when $\theta_{i,-1} = 0.0466$. The investors are therefore happy to maintain their time -1 positions and the market clears. The predicted alpha is -1.89%.

We now turn to decile 10. To endogenize the initial holdings of a stock i in this decile, we again search for a value of the location parameter μ_i that clears the market at time -1 for the objective function in (20) when $g_i = 0$. We find that the market clears for $\mu_i = 0.5432$. For this value of μ_i, the objective function has two global maxima at $\theta_i^* = 0$ and $\theta_i^{**} = 0.063$, which become investors' initial holdings. We now move to time 0. At this time, the typical stock in decile 10 has a capital gain. We find that the market clears for the same value of μ_i, namely $\mu_i = 0.5432$. For this μ_i, the investors who had initial holdings of $\theta_{i,-1} = 0.063$ now choose $\theta_i = 0$: they have large gains that they want to secure. Meanwhile, the investors with initial holdings of $\theta_{i,-1} = 0$ are still indifferent between $\theta_i = 0$ and $\theta_i = 0.063$. By allocating some of them to each maximum, we can clear the market. The alpha is 0.83%.

The overall alpha spread for this anomaly with endogenized initial holdings is therefore 2.73%, and so the conditions in (28) remain satisfied: the model again helps to explain this anomaly.

After conducting this exercise for all 23 anomalies, we obtain the same conclusion as for the benchmark model: the model can again explain 14 of the 23 anomalies – the 14 anomalies in the top group in Table 3. Moreover, the model continues to perform well in terms of average absolute pricing error: it has a pricing error of 0.594, which is again better than the pricing error for the CAPM (0.82) or the three-factor model (0.83) and similar to the pricing error for the four-factor model (0.55). These results reflect the fact that the model's mechanisms, and hence its predictions, are not very sensitive to investors' initial holdings.

M. Testing the model on another set of anomalies

In Section 5, we use the prospect theory model to generate predictions about the 23 anomalies in Table 1 and conclude that the model is helpful for thinking about a majority of the anomalies. However, we want to be sure that this conclusion is not special to this particular set of anomalies.

After completing the analysis in Section 5, we came across another set of 23 anomalies – one constructed by Novy-Marx and Velikov (2016) [NMV] to study transaction costs, in other words, for reasons that have nothing to do with prospect theory. To see if our conclusion in Section 5 is robust, we use our model to generate predicted alphas for the 230 NMV anomaly deciles. The results reinforce our conclusion in Section 5. By the criterion in (28), the model can help explain 13 of the 23 NMV anomalies, a very similar fraction to that reported in Section 5, where we find that the model can help explain 14 of the 23 anomalies in Table 1.

The left column of Table A3 lists the 23 NMV anomalies. The NMV anomalies and the anomalies in Table 1 have 12 anomalies in common; we identify these common anomalies by listing their acronyms in the second column. The third and fourth columns in the table report, for each anomaly, the model-predicted alpha spread – the difference in alphas between anomaly decile 10 and anomaly decile 1, $\alpha^m(10) - \alpha^m(1)$ – and also the empirical alpha spread, $\alpha^d(10) - \alpha^d(1)$. The 13 anomalies that the model can help explain, per the criterion in (28), are in the top part of the table; the remaining anomalies are in the lower part of the table.

N. Explaining time variation in anomaly alphas

To predict the expected return of the typical stock in an anomaly decile, our model requires three main inputs: the stock's volatility, skewness, and gain overhang. In Sections 4 and 5, we estimate these inputs over the full sample from 1963 to 2015 and then compare the model-predicted alphas to the empirical alphas in this sample period. We find that the model is helpful for thinking about 14 of the 23 anomalies.

Can the model also explain variation over time in the alphas of these 14 anomalies? To answer this, we divide the full sample into four equal subperiods. For each anomaly decile, we use the procedure described in Section 4 to estimate the volatility, skewness, and gain overhang of the typical stock in this decile for each of the four subperiods – in other words, using only data from the first subperiod; only data from the second subperiod; and so on. For each subperiod in turn, we use these inputs to generate model-predicted alphas for each anomaly decile in that subperiod.

To see if our model can explain some of the time variation in the empirical alphas, we proceed as follows. For each of the 14 anomalies, we compute its empirical alpha spread – the difference in the alphas of anomaly decile 10 and anomaly decile 1 – in each of the four subperiods, and also its model-predicted alpha spread in each of the four subperiods. We then run a regression of the empirical spreads on the model-predicted spreads. In order to focus on the model's ability to explain time variation in the alphas, we include anomaly fixed effects. The regression coefficient is positive and highly statistically significant, showing that the model can indeed explain time variation in anomaly alphas.

Table A4 reports the model-predicted and empirical alpha spreads for the 14 anomalies in each of the four subperiods. Looking over the table, we can see even without a formal analysis that the model is helpful for thinking about the time variation in the alphas. For example, several of the anomalies have substantially higher alpha spreads in the third subperiod, a pattern that the model is often able to capture and one that it ascribes to more extreme values of stocks' volatility, skewness, and gain overhang in this subperiod.

P. Pre- and post-publication alphas

There is currently a lot of interest in out-of-sample anomaly performance – how anomalies perform after the publication of the journal articles that document them. In this section, we compare the out-of-sample performance of the anomalies where the prospect theory model performs well with the out-of-sample performance of the anomalies where it performs poorly. One hypothesis is that, if the anomalies where the prospect theory model performs well can be said, as a result, to have a sounder theoretical foundation, they should perform better out-of-sample than the anomalies where the model performs poorly.

To investigate this, we take the 14 anomalies where the prospect theory model performs well and, for each one, identify the date on which the article documenting the anomaly was published. For each anomaly, we compute its alpha spread – the difference in alpha between its better-performing extreme decile and its worse-performing extreme decile – in its pre-publication period and also in its post-publication period. Averaged across the 14 anomalies, the pre- and post-publication alpha spreads are 13.8% and 7.86%, respectively, as shown in the table below. We do the same computation for the seven anomalies where the model performs poorly; the pre- and post-publication alpha spreads, averaged across the seven anomalies, are 9.28% and 1.76%, respectively.

From a casual inspection of the table, we see that the anomalies where the prospect theory model does well "hold up better" in the post-publication period. For these anomalies, their post-publication alpha is on average 63% of their pre-publication alpha, while for the anomalies where the model does poorly, the post-publication alpha is just 26% of the pre-publication alpha. (To be clear, the 0.63 is the anomaly-by-anomaly ratio of post-publication alpha to pre-publication alpha, averaged across anomalies – it is not 7.86% divided by 13.8%, although the difference is small.) These results are consistent with the hypothesis we started with: if the anomalies where the prospect theory model performs well can be said to have a stronger theoretical foundation, they should perform better out of sample. The results are also consistent with the anomalies where prospect theory performs well being harder to exploit. As discussed in Section 6.5, this is plausible in that the main mispricing generated by prospect theory investors is the overpricing of volatile, skewed, small-cap stocks – a mispricing that is hard to arbitrage.

	pre-publication alpha	post-publication alpha	post/pre ratio	winsorized ratio
14 anomalies where the model does well	13.8%	7.86%	0.63	0.53
7 anomalies where the model does poorly	9.28%	1.76%	0.26	0.24

We have examined the statistical significance of the above results. The difference between the post-/pre-publication ratios, i.e. the difference between 0.63 and 0.26, is not statistically significant. However, if we apply a reasonable winsorization – one where we constrain the post-publication alpha spread to be at most 100% and at least 0% of the pre-publication alpha spread, a restriction that affects four anomalies, namely EISKEW, ACC, SIZE, and STREV – then the post-/pre-publication ratios become 0.53 and 0.24, which *are* statistically different with a t-statistic of 2.16. In addition, the post-publication alpha spreads for the two groups of anomalies – the 7.86% and the 1.76% – are also statistically different with a t-statistic of 2.56.

Q. Interaction effects

In this section, we examine whether the prospect theory model can help explain three interaction effects related to the anomalies in Table 1.

Wang, Yan, and Yu (2017) sorts stocks into quintiles based on their gain overhang, where quintile 1 contains the lowest overhang stocks and quintile 5 the highest overhang stocks. They then further sort stocks into quintiles based on idiosyncratic volatility, computed as in Ang et al. (2006), where quintile 5 contains the stocks with the highest volatility. Each stock in the cross-section can then be placed into one of 25 categories, which we label using the notation (i, j), where $i, j \in \{1, 2, 3, 4, 5\}$, so that category (i, j) corresponds to the i'th overhang quintile and j'th idiosyncratic volatility quintile. The authors show that, for stocks in the lowest overhang quintile, there is a negative relationship between idiosyncratic volatility and return: the stocks in $(1, 5)$ have a lower average return and alpha than the stocks in $(1, 1)$. However, for stocks in the highest overhang quintile, there is a positive relationship between idiosyncratic volatility and return: the stocks in $(5, 5)$ have a higher average return and alpha than the stocks in $(5, 1)$.

An et al. (2020) conduct a similar double-sort exercise, this time sorting stocks on their gain overhang and on a measure of skewness such as expected idiosyncratic skewness. They again create 25 categories, which we again label with the notation (i, j), which corresponds to the i'th overhang quintile and j'th idiosyncratic skewness quintile. The authors find that, for the quintile of stocks with the lowest overhang, there is a negative relationship between idiosyncratic skewness and return: the stocks in $(1, 5)$ have a lower average return and alpha than the stocks in $(1, 1)$. However, for the stocks in the highest overhang quintile, there is a positive relationship between idiosyncratic skewness and return: the stocks in $(5, 5)$ have a higher average return and alpha than those in $(5, 1)$.

Frazzini (2006) also sorts stocks on two dimensions: their gain overhang and the size of the surprise in their most recent earnings announcement. The 25 resulting categories can again be labeled with the notation (i, j), which now denotes the i'th overhang quintile and j'th earnings surprise quintile. The author shows that a strategy that buys stocks with a positive earnings surprise and high overhang and shorts stocks with a negative earnings surprise and low overhang has a significantly higher alpha than a strategy that buys stocks with a positive earnings surprise and low overhang and shorts stocks with a negative earnings surprise and high overhang. In other words, the alpha of $(5, 5)$ stocks minus the alpha of $(1, 1)$ stocks is significantly higher than the alpha of $(1, 5)$ stocks minus the alpha of $(5, 1)$ stocks.[33]

We now examine whether our model can help explain these empirical patterns. Consider the first pattern – the volatility-overhang interaction described by Wang, Yan, and Yu (2017).

[33]Frazzini (2006) measures "earnings surprise" using cumulative abnormal stock returns around earnings announcement dates. We measure it instead as the difference between realized earnings and expected earnings derived from a seasonal random walk model. We do not expect that this difference in methodology will have a material impact on our results.

We use the methodology of Section 4 to compute the volatility, skewness, and gain overhang of the typical stock in each of the 25 volatility-overhang categories. We then plug these inputs into our model to see what it predicts for the expected returns and alphas of the 25 categories. We proceed in a similar way for the other two interaction patterns.

We find that the model can help explain all three empirical interactions and illustrate the results in Figure A3. The top-left graph in the figure corresponds to the volatility-overhang interaction. The horizontal axis marks the idiosyncratic volatility quintiles, 1 through 5. The lower line in the graph plots the alphas predicted by the model for categories $(1,1)$ to $(1,5)$, in other words, the alphas as we increase idiosyncratic volatility within the lowest overhang quintile. The upper line plots the predicted alphas for categories $(5,1)$ to $(5,5)$, in other words, the alphas as we increase idiosyncratic volatility within the highest overhang quintile. As in the data, the lower line is downward sloping while the upper line is upward sloping.

The top-right graph in the figure corresponds to the skewness-overhang interaction. The horizontal axis marks the idiosyncratic skewness quintiles, 1 through 5. The lower line in the graph plots the alphas predicted by the model for categories $(1,1)$ to $(1,5)$, in other words, the alphas as we increase idiosyncratic skewness within the lowest overhang quintile. The upper line plots the predicted alphas for categories $(5,1)$ to $(5,5)$, in other words, the alphas as we increase idiosyncratic skewness within the highest overhang quintile. As in the data, the lower line is downward sloping while the upper line is upward sloping.

Finally, the bottom-left graph in the figure corresponds to the earnings surprise-overhang interaction. The horizontal axis marks the earnings surprise quintiles, 1 through 5. The lower line in the graph plots the alphas predicted by the model for categories $(1,1)$ to $(1,5)$, in other words, the alphas as we increase the earnings surprise within the lowest overhang quintile. The upper line plots the predicted alphas for categories $(5,1)$ to $(5,5)$, in other words, the alphas as we increase earnings surprise within the highest overhang quintile. As in the data, the difference between the alphas for categories $(5,5)$ and $(1,1)$ is much higher than the difference between the alphas for categories $(1,5)$ and $(5,1)$.

Wang, Yan, and Yu (2017), An et al. (2020), and Frazzini (2006) suggest, using informal arguments, that the risk attitudes captured by prospect theory can help explain the interaction effects they document. Our results provide formal support for this claim.

Table A1. The contribution of each element of prospect theory. The left-most column lists the 14 anomalies that the prospect theory model can help explain; the acronyms are defined in Table 1. The next three columns report, for each anomaly, the percentage of the alpha spread – the difference in the model-predicted alphas for deciles 1 and 10 – that is generated by the loss aversion, probability weighting, and diminishing sensitivity elements of prospect theory, respectively.

Anomaly	Loss aversion	Probability weighting	Diminishing sensitivity
IVOL	–26	58	68
MOM	–13	34	79
FPROB	–24	52	73
OSC	–43	85	58
PROF	–12	50	62
EISKEW	–62	121	41
MAX	–11	37	74
DOP	–28	45	84
CEI	–23	34	88
XFIN	–4	18	85
PEAD	3	6	91
ROA	–12	33	79
NSI	–23	60	64
CGO	–22	38	84

Table A2. Earnings announcement analysis. The left-most column lists 23 anomalies; the acronyms are defined in Table 1. Column C reports the fraction of an anomaly's empirical return spread that is explained by the prospect theory model. Column F reports the fraction of the return spread that comes around earnings announcement dates; these numbers are winsorized in column G. The remaining columns are defined in Section I of the Internet Appendix. The table confirms that the prospect theory model performs better for anomalies where less of the return spread comes around earnings announcements: the numbers in column C are negatively correlated with those in column G.

	A	B	C	D	E	F	G	H
Anomaly	Empirical return spread	Model predicted return spread	Model perfor-mance	Empirical return spread 1983 on	Return spread around earnings	Earnings ratio	Earnings ratio winsorized	Earnings ratio dummy
IVOL	–1.211	–0.591	0.488	–0.671	0.423	–0.63	–0.63	0
SIZE	–0.255	0.334	–1.309	–0.007	–0.283	38.139	1	1
VAL	0.516	–0.627	–1.214	1.507	1.155	0.767	0.767	1
EISKEW	–0.333	–0.447	1.341	–0.087	0.579	–6.635	–1	0
MOM	1.836	1.27	0.692	1.022	0.706	0.69	0.69	1
FPROB	–1.346	–1.109	0.824	–0.778	–0.108	0.139	0.139	0
OSC	0.776	0.389	0.501	0.543	0.656	1.209	1.209	1
NSI	–0.691	–0.117	0.169	–1.05	–0.714	0.68	0.68	1
CEI	–0.549	–0.079	0.144	–0.883	–0.622	0.705	0.705	1
ACC	–0.607	0.302	–0.498	–0.772	–0.71	0.92	0.92	1
NOA	–0.604	–0.044	0.073	–1.305	–0.382	0.293	0.293	0
PROF	0.423	0.293	0.694	0.656	0.878	1.338	1.338	1
AG	–0.59	0.344	–0.583	–1.304	–0.845	0.648	0.648	1
ROA	0.699	0.708	1.013	1.336	0.579	0.433	0.433	0
INV	–0.591	0.211	–0.357	–1.221	–0.808	0.661	0.661	1
MAX	–0.764	–0.402	0.527	–0.865	0.208	–0.241	–0.241	0
ORGCAP	0.456	–0.093	–0.205	0.738	0.667	0.904	0.904	1
LTREV	–0.423	0.729	–1.722	–0.77	–1.019	1.323	1.323	1
XFIN	–0.687	–0.202	0.294	–1.27	–1.027	0.809	0.809	1
STREV	–0.508	0.565	–1.111	–2.07	–0.259	0.125	0.125	0
DOP	–0.411	–0.02	0.049	–0.761	–0.663	0.872	0.872	1
PEAD	0.563	0.25	0.444	1.43	0.372	0.26	0.26	0
CGO	0.74	1.526	2.063	0.425	–0.332	–0.78	–0.78	0

Table A3. The left-most column lists the 23 anomalies studied by Novy-Marx and Velikov (2016). For those anomalies that are also in the set of anomalies in Table 1, the second column lists their acronyms. The third and fourth columns report, for each anomaly, the alpha spread between anomaly decile 10 and decile 1 predicted by the prospect theory model and the empirical alpha spread. The anomalies for which the model performs well are in the upper part of the table; those for which the model performs poorly or does not make a strong prediction are in the lower part.

Anomaly	Abbreviation	Model alpha spread	Empirical alpha spread
Momentum	MOM	14.52	23.26
Failure probability	FPROB	–13.3	–18.82
Return on book equity		9.31	9.91
Idiosyncratic volatility	IVOL	–8.78	–17.99
Value, profitability, and momentum combination		6.44	17.96
Value and momentum combination		6.39	14.69
Three-day earnings announcement return		5.36	16.5
Piotroski F-score		4.68	3.72
Seasonality		3.99	9.71
Gross profitability	PROF	3.65	6.43
Post-earnings announcement drift	PEAD	3.26	7.12
Net stock issuance	NSI	–2.21	–9.23
Industry momentum		1.77	11.6
Industry relative reversal (low volatility)		–0.53	–13.6
Investment	INV	1.84	–7.8
Market capitalization	SIZE	2.34	–1.76
High-frequency combination		–2.8	21.22
Accrual	ACC	2.89	–8.35
Asset growth	AG	2.93	–8.28
Value and profitability combination		–3.68	9.24
Industry relative reversal		6.35	–7.73
Short-term reversal	STREV	6.79	–3.57
Value	VAL	–6.85	5.79

Table A4. Subperiod analysis. The left-most column lists the 14 anomalies for which the prospect theory model can help explain the full-sample anomaly alphas; the acronyms are defined in Table 1. The next four colums report, for each anomaly, the model-predicted alpha spread between anomaly decile 10 and anomaly decile 1 for each of four subperiods. The last four columns report, for each anomaly, the empirical alpha spread for each of the four subperiods.

Anomaly	Model Period 1	Model Period 2	Model Period 3	Model Period 4	Data Period 1	Data Period 2	Data Period 3	Data Period 4
IVOL	–8.4	–5.73	–13.53	–6.67	–9.53	–21.55	–37.89	–1.29
MOM	14.15	14.77	18.74	10.74	26.24	24.51	31.67	10.39
FPROB	–10.62	–16.77	–16.04	–9.64	–12.25	–20.42	–30.91	–10.62
OSC	5.34	5.61	8.48	2.28	5.1	15.45	18.74	5.24
PROF	2.26	1.61	4.54	4.9	3.87	0.32	12.56	10.67
EISKEW	–4.80	3.05	–10.03	–6.04	2.19	–1.27	–19.13	–7.61
MAX	–7.06	–5.18	–12.53	–5.34	–8.33	–13.41	–28.55	–4.86
DOP	–2.93	–2.63	–2.17	–0.62	–9.87	–10.95	–1.85	–11.36
CEI	–2.69	–0.45	–5.06	–2.78	–6.93	–12.2	–12.02	–4.89
XFIN	–1.92	–1.71	–6.39	–3.18	–6.2	–10.87	–22.21	–5.76
PEAD	2.9	2.85	4.77	2.56	11.89	4.79	5.53	6.6
ROA	10.98	9.93	11.07	5.72	2.43	10.96	20.13	10.03
NSI	–1.36	0.27	–4.56	–3.66	–8.92	–9.41	–8.2	–10.51
CGO	17.7	16.56	20.6	15.01	3.0	18.3	20.88	5.05

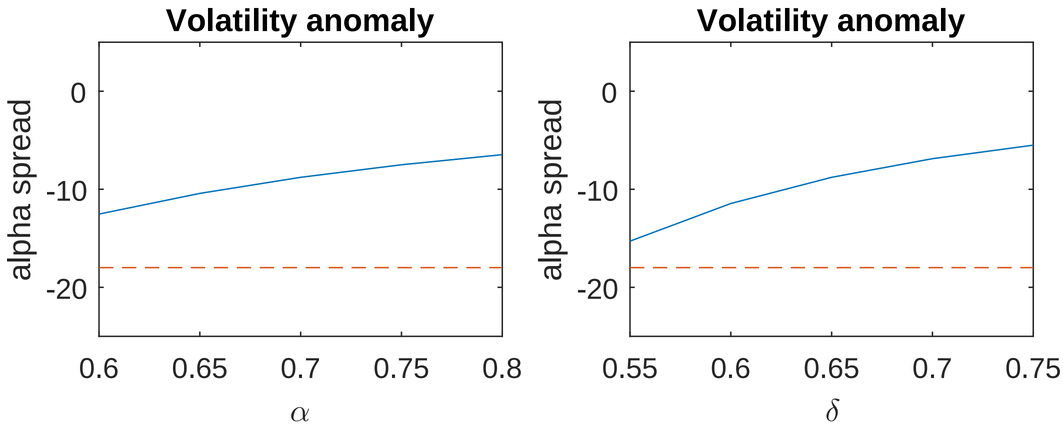

Figure A1. The solid line in the left graph shows how the model-predicted alpha spread for the idiosyncratic volatility anomaly depends on the value of the diminishing-sensitivity parameter α; the baseline value is 0.7. The solid line in the right graph shows how the predicted alpha spread for this anomaly depends on the value of the probability-weighting parameter δ; the baseline value is 0.65. The dashed lines in the graphs mark the empirical alpha spread.

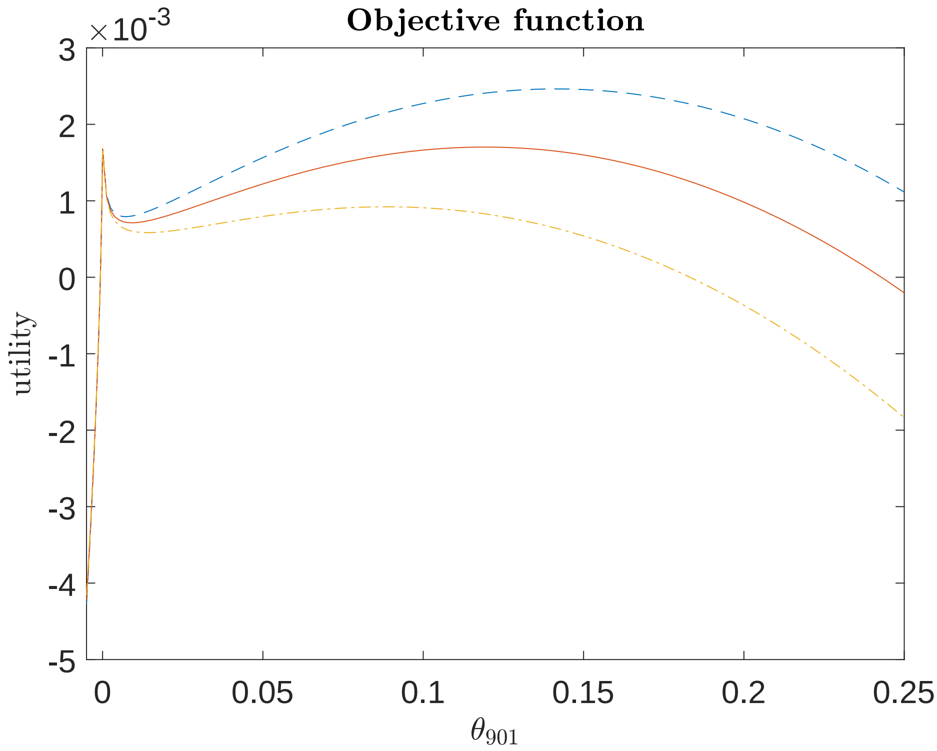

Figure A2. The graph shows that investors have heterogeneous holdings of each stock in momentum decile 10. The solid line plots the value of an investor's objective function in equilibrium as a function of θ_{901}, the (scaled) fraction of the investor's portfolio allocated to stock 901, which belongs to momentum decile 10. The function has two global maxima which straddle the weight of stock 901 in the market portfolio, namely 7.26×10^{-4}. The dashed line plots the objective function for a higher expected return on the stock, while the dash-dot line plots the objective function for a lower expected return.

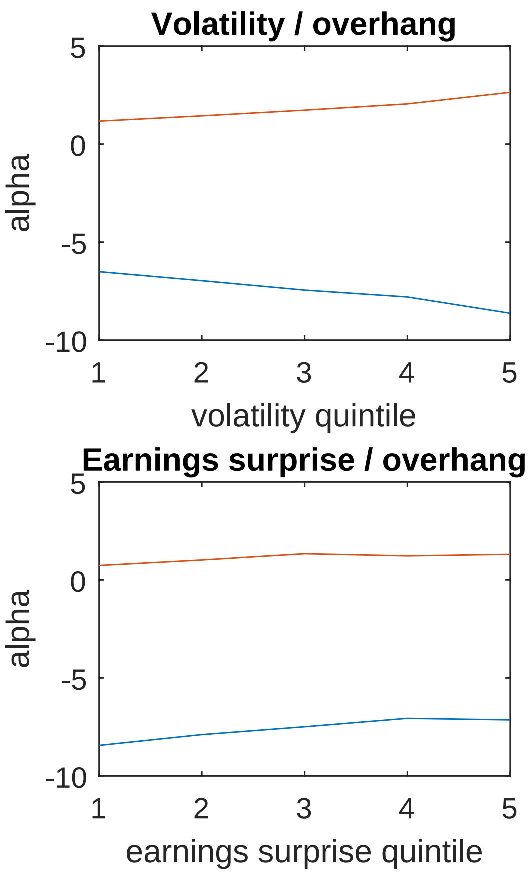

Figure A3. The graphs show that the prospect theory model can help explain three anomaly interaction effects. The lower lines in the three graphs plot the model-predicted alphas for stocks with large negative overhang and various levels of volatility, skewness, and earnings surprise, respectively. The upper lines in the three graphs plot the model-predicted alphas for stocks with large positive overhang and various levels of volatility, skewness, and earnings surprise, respectively.