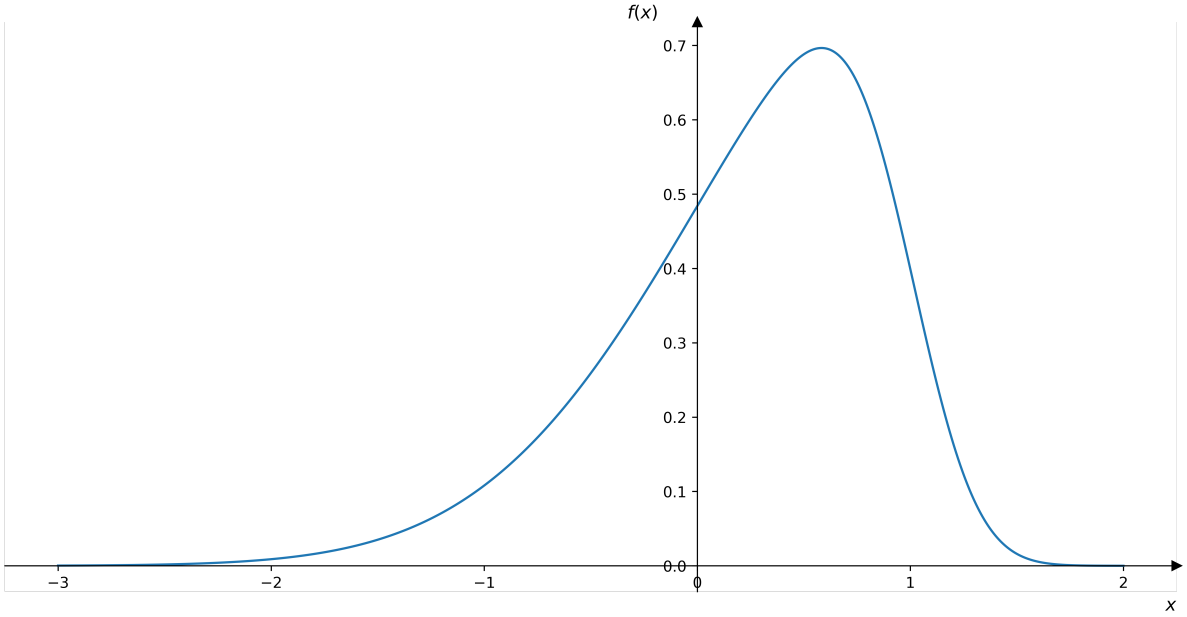
Prospect Theory and Stock Market Anomalies

- Author
  - Nicholas Barberis et al.
- Publication
  - JF 2021

- Prospect Theory and Stock Market Anomalies
  - Introduction
  - Model
  - Equilibrium
  - Param Values
  - Empirical Test
  - Advantages and Robustness

Introduction

- **Contributions**
  - Use 3 forces to quantitively explain anomalies
    - Investors require higher average returns when
      1. returns are more volatile;
      2. returns are more negatively skewed;



         Note that investors overweight the tails, which means they dislike the left tail while the right tail is attractive.
      3. there are more prior gains.
         According to the utility function's concavity over gains and convexity over losses (will be introduced below).
  - Combine the mean-variance optimization with cumulative prospect theory
  - Use an equlibrium condition to solve expected returns to explain anomalies' alpha

- **Anomalies**

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

- **Statistics of Anomaly Deciles**

Anomaly	Average Return Decile 1	Average Return Decile 10	Standard Deviation Decile 1	Standard Deviation Decile 10	Skewness Decile 1	Skewness Decile 10	Gain Overhang Decile 1	Gain Overhang Decile 10
IVOL	11.9	−3.2	36.8	94.3	2.47	3.79	10.5	−31.6
SIZE	14.0	10.6	76.0	24.6	4.27	0.69	−15.4	17.0
VAL	10.7	17.7	54.1	71.0	1.85	2.66	12.1	−24.1
EISKEW	12.4	8.0	33.3	90.3	1.33	3.54	13.7	−24.8
MOM	−2.8	20.9	91.8	63.1	3.81	2.39	−45.0	31.0
FPROB	17.5	0.1	51.5	110.4	2.31	3.9	28.2	−46.2
ZSC	3.8	13.9	84.5	51.6	3.51	2.56	−20.6	6.1
NSI	16.0	6.8	54.2	70.8	2.71	3.2	5.7	−8.4
CEI	14.0	6.8	42.1	66.5	2.46	2.68	7.4	−11.2
ACC	15.4	7.4	85.0	72.0	3.2	3.0	−18.7	−5.4
NOA	14.8	6.9	66.4	65.3	3.13	2.95	−2.6	−4.9
PROF	8.8	14.4	73.0	62.6	3.49	2.68	−12.0	1.1
AG	14.8	7.0	82.5	70.7	3.1	3.07	−20.6	−5.6
ROA	4.1	13.1	85.3	55.3	3.22	2.22	−29.9	15.1
INV	15.9	8.0	81.8	72.2	3.6	3.31	−14.0	−5.4
MAX	10.9	1.2	46.3	90.4	2.76	3.61	3.2	−25.1
ORGCAP	10.2	16.3	54.5	77.7	2.17	3.21	−1.7	−8.6
LTREV	16.7	11.0	88.9	47.7	3.23	1.77	−32.1	17.1
XFIN	13.0	4.1	58.7	83.4	3.17	3.44	−0.5	−15.6
STREV	14.4	7.6	83.1	73.3	3.63	3.03	−22.2	−1.9
DOP	15.5	10.0	40.2	57.9	1.36	1.74	16.6	−7.8
PEAD	9.2	16.8	58.3	58.3	2.49	2.3	−9.4	7.5
CGO	5.8	15.5	93.9	47.8	3.58	2.17	−57.4	57.7

  - We can see the phenomenon in all anomalies except NOA: if decile 1 has a higher standard deviation than decile 10, then it also has a higher skewness and more negative gain

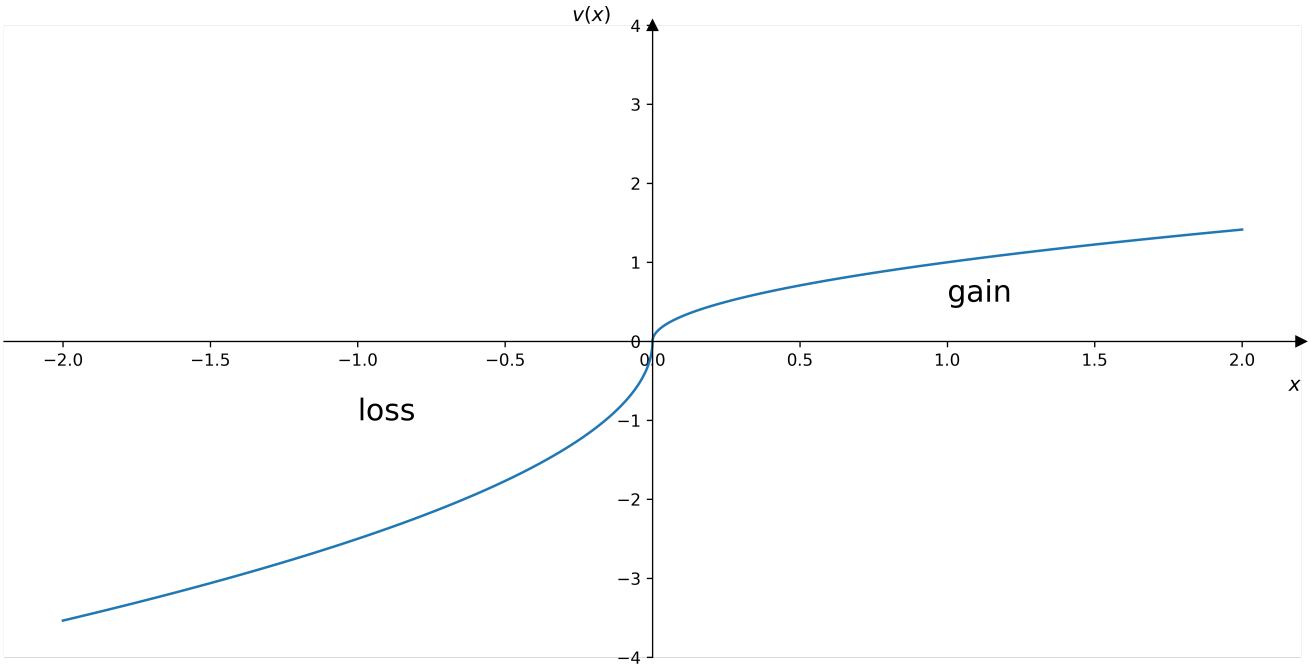
overhang, and vice versa.

- This means there exist counteracting forces since investors require higher average return if there is higher standard deviation while they require lower average return if there is higher skewness and more negative gain overhang.
- Thus, we need a quantitive model to determine which force dominates.

- **Cumulative Prospect Theory (CPT)**
  - Value Function

$$v(x) = \begin{cases} x^{\alpha} & , \ x \geqslant 0 \\ -\lambda(-x)^{\alpha} & , \ x < 0 \end{cases}$$

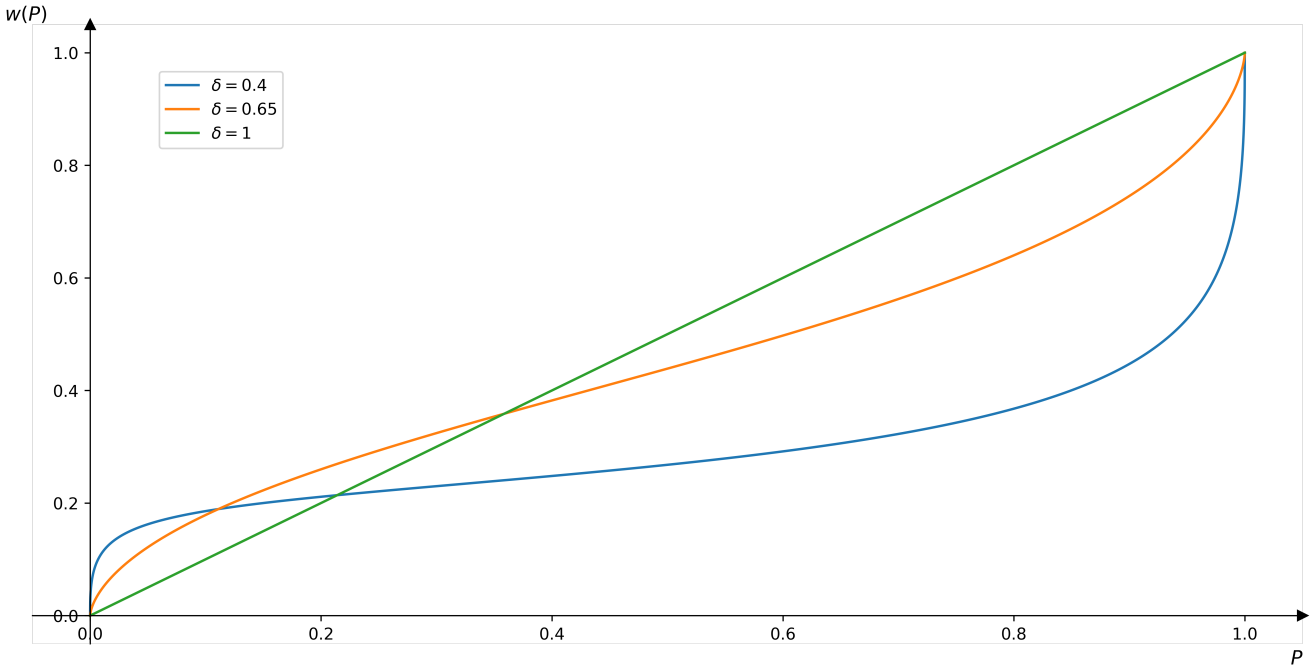
where $\alpha \in (0,\ 1)$ and $\lambda > 1$.



    - The value function models the concave utility over gains and convex utility over losses of investors by setting α.
    - The value function considers the phenomenon that investors are more sensitive to losses by setting λ.
  - Probability Weighting Function

$$w(P) = \frac{P^{\delta}}{(P^{\delta} + (1-P)^{\delta})^{\frac{1}{\delta}}}$$

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



- The probability weighting function describes the phenomenon that investors overweight the tails by setting δ.

- Discrete Form
  - Consider payoffs in ascending order with corresponding probabilities

$$(x_{-m},\ P_{-m};\cdots;x_{-1},\ P_{-1};x_0,\ P_0;x_1,\ P_1;\cdots;x_n,\ P_n)$$

    where $x_0 = 0$.
  - Similar to the traditional (discrete) expected utility framework, i.e. $\sum_{i=-m}^{n} P_i U(W + x_i)$ where U is the utility function and W is the current wealth, the discrete CPT is

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

    where $\pi_i = \begin{cases} w(p_i + \cdots + p_n) - w(p_{i+1} + \cdots + p_n) & ,\ 0 \leqslant i \leqslant n \\ w(p_{-m} + \cdots + p_i) - w(p_{-m} + \cdots + p_{i-1}) & ,\ -m \leqslant i < 0 \end{cases}$ is the pmf-like function.

- Continuous Form
  - Similar to the traditional (continuous) expected utility framework, i.e., $\int_\Omega U(W + x)\ \mathrm{d}P$ where Ω is a set of probabilities that has one-to-one correspondence with payoffs, the continuous CPT is

$$\int_{\Omega'} v(x)\ \mathrm{d}w(P)$$

    where Ω' is a set of weights that has one-to-one correspondence with payoffs.

- **Narrow Framing**
  - A investor derive his utility from individual assets in his portfolio separately instead of from the whole portfolio.
  - The converse is broad framing.

Model

- **Model Assumptions**
  - Consider 3 time states: -1, 0 and 1 that represent past, now and future respectively.
  - For simplicity, assume all investors hold the market portfolio at time -1, and thus they would have the same prior gain or loss in each asset.
  - Also for simplicity, assume that the total wealth does not change from time -1 to 0.
    - It can be set as the change with average market return, which has little impact on the results.
  - Assume narrow framing.
  - Assume investors derive utility from paper gains and losses rather than from realized gains and losses, i.e., investors do not derive utility from selling at a gain at time 0, instead, they derive utility from accumulated gain at time 1.

- **Optimization Objective**

$$\max_{\Theta} \quad \mathrm{E}(\widetilde{W}_1) - \frac{\gamma}{2}\mathrm{Var}(\widetilde{W}_1) + b\sum_{i=1}^{N} V(\widetilde{G}_i)$$

where Θ is the weight vector for the portfolio, $\widetilde{W}_1 = W_0((1-\Theta^{\mathsf{T}}e)R_f + \Theta^{\mathsf{T}}\widetilde{R})$ is the future wealth (W_0 is the current wealth, $e = \left[1,\ \cdots,\ 1\right]^{\mathsf{T}}$, R_f is the gross risk-free rate and $\widetilde{R}$ is the gross return of risky assets), γ is the risk aversion coefficient, b is the importance coefficient of CPT, N is the num of assets, $V(\cdot)$ is the continuous CPT value and $\widetilde{G}_i = W_{-1}\Theta_{i,\,-1}g_i + W_0\Theta_i(\widetilde{R}_i - R_f)$ is the simple addition of prior and future gains or losses (W_{-1} is the past wealth, $\Theta_{i,\,-1}$ is the past weight and g_i is the prior gain or loss).

  - Expansion of $V(\widetilde{G}_i)$ for $\Theta_i > 0$ (the case for $\Theta_i < 0$ is similar):

$$-\lambda W_0^{\alpha}\int_{\Omega_1}\left(\Theta_i\left(R_f - R_i\right) - \Theta_{i,\,-1}g_i\right)^{\alpha}\mathrm{d}w(P(R_i))$$
$$-W_0^{\alpha}\int_{\Omega_2}\left(\Theta_i\left(R_i - R_f\right) + \Theta_{i,\,-1}g_i\right)^{\alpha}\mathrm{d}w(1-P(R_i))$$

    where Ω_1 is a set of weights that has one-to-one correspondence with losses, i.e. $R_i \in (-\infty,\ R_f - \frac{\Theta_{i,\,-1}}{\Theta_i})$ and Ω_2 is a set of weights that has one-to-one correspondence with gains, i.e. $R_i \in (\ R_f - \frac{\Theta_{i,\,-1}}{\Theta_i},\ \infty)$.
  - Return Distribution
    - One-dimensional GH skewed t distribution for each asset i to model the skewness and fat tails

- pdf

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

for $\zeta_i \neq 0$

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

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

- Params
  - μ helps determine the mean of returns
  - S helps determine the variance
  - ξ helps determine the skewness
  - ν helps determine the heaviness of tails
- Moments

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

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

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

  - Expansion of the whole objective (scaled):

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

$$\Theta_{M,R}=\sum_{i=1}^{N}\Theta_{M,i}$$

$$\theta_i=\Theta_i/\Theta_{M,R}$$
$$\theta_{M,i}=\Theta_{M,i}/\Theta_{M,R}$$
$$\theta_{i,-1}=\Theta_{i,-1}/\Theta_{M,R}$$

$$\widehat{\gamma}=\gamma W_0\Theta_{M,R}$$
$$\widehat{b}=bW_0^{\alpha-1}\Theta_{M,R}^{\alpha-1}$$

- **Target**
  - The target is not to use mean and variance estimations to get a optimized weight vector. Instead, estimate the needed params by some infomation and the final target is to get μ under an equilibrium condition, i.e., we try to find a μ that let the optimized weight vector equal the market portfolio.

Equilibrium

- **Full rationality with homogeneous holdings**
  - All investors are identical and they would choose the optimized portfolio which is also the market portfolio at time 0.
  - It is hard or even impossible to get a μ to meet this equilibrium condition.

- **Full rationality with heterogeneous holdings**
  - There exist more than one optimized portfolio, which means there should be multiple global maxima. We can clear the market by allocating investors to different optimized portfolios.
  - Since μ is a N-dimensional vector for N assets, if we consider 100 candidate values for each component of μ, there will be 100^N combinations. It is computaionally infeasible to meet this equilibrium.

- **Bounded rationality with heterogeneous holdings**
  - To overcome the difficulties described above, authors introduce a method: hold other assets' weight as the market weight while optimizing some asset.
  - Actually, the solution cannot be the best using this method, but it do simplify the problem.
  - After optimization using this method, results show that there may be one unique solution or two global maxima for each asset. If there are two global maxima, one would be higher than the market weight while the other is slightly lower than the market weight.

Param Values

- **Market Weight**
  - The market weight of a decile for a particular anomaly is calculated by the percentage that the decile accounts of the total market value.
  - Assume the stocks in a decile for a particular anomaly are identical. Thus, the market weight of a typical stock is just the market weight of the decile divided by num of stocks in that decile.

- **Distribution Params**
  - $\nu = 7.5$, which is a reasonable choice for fat tails. The results are not sensitive to ν.
  - S and ξ are derived by empirical variance and skewness. The (cross-sectional) empirical variance and skewness are calculated through annual returns over the next year (future data).
  - Why using future data
    - It is about what a rational, forward-looking investor is interested in;
    - Also, it is a way to get real-world investors' judgement of future.
  - Why using annual returns
    - In the best-known data set of individual investor trading, portfolio turnover is about 50% per year, which means investors hold assets for 2 years on average. Since authors assume investors buy at time -1, change weights at time 0 and sell all at time 1, it is reasonable that the time between 0 and 1 is one year.

- **CPT Params**
  - The scaled risk aversion coefficient $\hat{\gamma} = 0.6$.
  - The scaled importance coefficient of CPT $\hat{b} = 0.6$.
  - Why $(0.6,\ 0.6)$
    - To generate an aggregate equity premium of 6%.

    > Consider a symmetric gamble: a fifty-fifty bet to gain M or lose M. According to the utility model with CPT, the utility changes by $-\frac{\gamma}{2}M^2 - b(\lambda - 1)w\left(\frac{1}{2}\right)M^{\alpha}$. If paying π to avoid this gamble, then the utility changes by $-\pi - b\lambda\pi^{\alpha}$. When these 2 changes are the same, π is called the premium.

    - There are many combinations of $(\hat{\gamma},\ \hat{b})$ that satisfy the first requirement. Empirical results show that the average Sharpe ratio of the household portfolios is 19% lower than

that of a market benchmark and $(0.6,\ 0.6)$ can derive a similar result, i.e., the Sharpe ratio of investors' portfolios is 12.5% lower than that of the market portfolio.

- Based on lots of literature before, $(\alpha,\ \delta,\ \lambda) = (0.7,\ 0.65,\ 1.5)$.

Empirical Test

- **Procedure**
  - Date range: 1963/7 ~ 2014/12
  - For each anomaly,
    1. compute the annualized value-weighted CAPM α's $\alpha^d(l)$ for each decile l (d represents "data");
    2. compute the model alpha using the formula

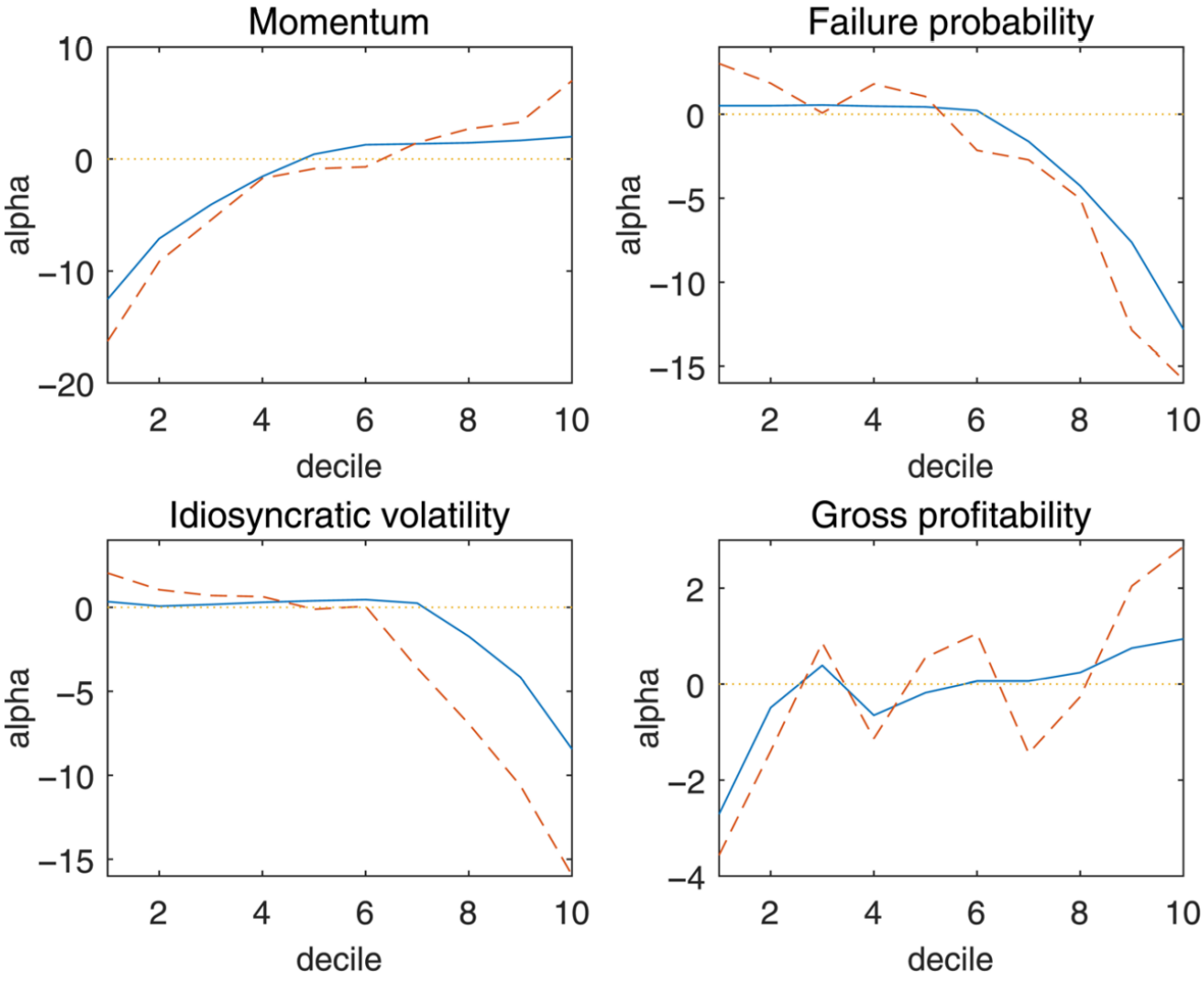
       $$\alpha^m(l) = R_l - (R_f + \beta_l(R_M - R_f))$$

       where m represents "model", R_l is the gross return of a typical stock in decile l, R_M is the gross return of market portfolio and β_l is averaged among market β's calculated using daily returns over the next year (future data). Note that the β's in step 1 and step 2 are different;
    3. get the α spread, i.e. $\alpha^d(10) - \alpha^d(1)$ and $\alpha^m(10) - \alpha^m(1)$;
    4. if $\text{sign}(\alpha^d(10) - \alpha^d(1)) = \text{sign}(\alpha^m(10) - \alpha^m(1))$ and $|\alpha^m(10) - \alpha^m(1)| > 0.015$, then the model can help explain that anomaly.

- **Results**
  - α spread

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

    - 14 anomalies above the blue lines are predicted well, 2 anomalies between the blue lines are predicted in the right direction but the predictions are not strong, while the model performs poorly for 7 anomalies below the blue lines.
    - For 5 of the 7 anomalies that are predicted poorly (VAL, INV, LTREV, ACC and AG), a large fraction of their return comes around earnings announcement dates, which means these anomalies are driven incorrect beliefs instead of prospect theory risk attitudes. Thus, anomalies can be categorized into 2 classes: preference-based and belief-based.
- Decile α's



where the blue solid lines are model α's and the red dot lines are data α's.

- The model α can approximately track the data α.
- Combined with the anomalies' statistics in *Introduction*, we can find that the force of skewness and the force of gain overhang dominate.

- **Performance Comparation**
  - Average Absolute Pricing Error

$$\left(\overline{R}^{d}(1)-\overline{R}^{d}(10)\right)-\left(\overline{R}^{m}(1)-\overline{R}^{m}(10)\right)$$

where $\overline{R}(l)$ is the average return of the stocks in anomaly decile l.

  - Comparation

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

- This model is similar to the four-factor model and better than CAPM as well as the three-factor model. The six-factor model is the best.
- Although the comparation seems to imply that the CPT model is not that good, we should notice that factor models are developed in full knowledge of multiple anomalies while the CPT model are not.

Advantages and Robustness

- **Importance of CPT Characteristic**
  - Diminishing sensitivity composes the most part of the α spread.
  - Probability weighting doubles the anomalies that can be predicted.

- **Importance of Assumptions**
  - If changing some of the assumptions, the model will perform poorly.

- **Robustness**
  - Results are not sensitive to params.
  - If using an alternative set of anomalies, the performance is similar.
  - If changing the sample period, the results are also similar.