What drives behavioral or mispricing-based factors

Dijun Liu*
Universitat Autonoma de Barcelona
Barcelona School of Economics

ABSTRACT

Factor models are important to explain returns and to calculate the required rate of return. We analyze the risk associated with the outstanding mispricing factors proposed recently that outperform other factor models in explaining a broad range of return anomalies. Present variance decomposition suggests that cash flow news is the main driver for all of these factors and their mean-variance efficient portfolios. The explanatory power of these factors stems from the ability of their cash flow news to explain anomalies' unexpected return news. The beta analysis also indicates that the cross-sectional variation of market beta comes from the variation in cash flow components of beta.

Keywords: Factor model, Anomalies, Mispricing factor, Behavioral factor, Present value decomposition, Good beta bad beta.

*International Doctorate in Economic Analysis (IDEA) at Universitat Autònoma de Barcelona (UAB) and Barcelona School of Economics (BSE). Email: dijun.liu@uab.cat. I am indebted to my advisor Abhay Abhyankar for his continuous encouragement and advice. I am also grateful to my tutor Michael Creel and Jordi Caballé for their comments and suggestions.

1 Introduction

The last two decades have seen the emergence of a surging number of anomalies newly discovered. As documented by Harvey and Liu (2019), the production of factors is out of control, leading to nearly 400 factors published in top journals by January 2019. Despite the possibility of data mining, these anomalies do present a challenge for the asset pricing models as these models fail to explain anomalies' returns. In the meantime, some new factor models are proposed and the authors claim that their models perform better as they expand the mean-variance frontier and explain more anomalies (see Chen and Zhang (2010), Fama and French (2015), Hou, Xue, and Zhang (2015), Stambaugh and Yuan (2017), Barillas and Shanken (2018) et. al). The past dominant factor models are mainly characteristics-based, using firms' accounting information like BM, size, etc. But in recent years, some authors propose behavioral or mispricing-based factor model which not only expands the mean-variance frontier but also explains many well-documented anomalies. Given the power of these factors, it is essential to know what drives them, as this understanding not only contributes to theoretical models but also has implications for practical investment. However, as noted by Scientific Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2013, "A weakness of the (Fama-French) three-factor model is that it is primarily an empirical model that describes stock returns, but it is silent on the underlying economic reasons for why these risk factors have nonzero prices." Similarly, we know little about these new factors, do they truly stand for a kind of risk that requires a higher premium, or do they just stand for the animal spirits which are driven by sentiment, and over/under-reaction? There are many papers in the literature trying to explore the reasons. We contribute to the literature by applying the present value decomposition approach to the newly proposed behavioral or mispricing-based factors from Stambaugh and Yuan (2017) and Daniel, Hirshleifer, and Sun (2020). We also find the source of the predictability of the factors from the perspective of news, rather than return regression-based results from previous studies which only look at the coefficients or intercepts. Finally, we provide evidence supporting the fundamental/risk-based explanations for the factors.

Lochstoer and Tetlock (2020) have done similar work to the five well-known anomalies—value, size, profitability, investment, and momentum, which are most frequently included in factor models though there are some minor differences in construction like updating frequency. What they find is that the systematic cash flow news drives the returns of anomaly portfolios, as well as their mean-variance-efficient (MVE) portfolio. We focus on the behavioral or mispricing-based factors as they are from predominant factor models and they compose other anomalies when being formed.

In this paper, I use the present value decomposition approach to analyze the news components of these behavioral or mispricing-based factors. To this end, we first construct the factors following the same procedure or logic. Although the original papers provide factor returns, we need to reconstruct them for two reasons. First, in order to aggregate firm-level news to the other levels, like portfolio level, and anomaly level, it is necessary to get their characteristics so that we know which portfolio the stock is grouped into. Second, some of the factors, say the PEAD factor from Daniel et al. (2020), are rebalanced every month. Yet the cash flow (CF hereafter) news, discount rate (DR) news and unexpected return news components are at annual frequency due to the availability of accounting information. Therefore, the best we can do is to mimic the logic of these monthly updated factors and reconstruct them at the yearly frequency to keep in accordance with the news. It is not surprising that it will harm the explanatory power of these factors to some extent, but we show in the empirical results that it is acceptable. Next, we extract the cash flow news and discount rate news at the aggregate market level and the firm-level. This is done through vector autoregression. Then we aggregate the firm-level news to the desired level like portfolio level or anomaly level in the same procedure of constructing the returns except that we use stock news in place of stock returns. The above enables the analysis of what drives the variation in returns, which news components empower the explanatory ability of these factors, and the comparison of the roles played by news in systematic risk.

Our main results question the underlying behavioral models as they all broadcast the significant effects of cash flow news that represents fundamental risks, instead of sentiment risk or misvaluation proxied by discount rate news.

First, the variance decomposition suggests that the variation in cash flow news takes up the most variation in the unexpected return news across all the misbehavioral or mispricing-based factors. The same applies to the firm-level news decomposition, and the anomalies we considered in this paper. This indicates that though agents' slow reaction, or managers' market timing to exploit the advantage over outside investors, among other explanations based on behavioral theory may play a role in the abnormal returns of these factors/anomalies, it is the cash flow news that capture real risk that drives the variation of the returns. Take FIN factor from Daniel et al. (2020) and the momentum factor from Jegadeesh and Titman (1993) as examples. The FIN factor describes the stock issuance which is empirically shown negatively related to future stock returns. The momentum factor describes the fact that past medium-term (six to twelve months) winners will still outperform in the future. Two strands of reasoning are plausible: risk-based or behavioral-based. From the point of view of the risk explanation, the high returns are compensation for the high risk associated with underlying stocks. On the other hand, the behavioral explanation regards it possible to have free lunches due to misvaluation caused by asymmetric information, market friction etc. High momentum stocks earn high future returns because these stocks are more likely to see their surging prices plunge back to earth, and therefore investors require risk premium—higher returns for bearing this additional risk. It can also be the case that investors are slow to respond to the news, and therefore it takes much longer time for the price to increase to the true intrinsic value, leading the winners to win, and the losers to lose. For a high net issuance, it can be caused by a better investment opportunity or by managers' intention to sell the overpriced stock, both leading to lower future returns. But the former reflects fair pricing while the latter indicates an inefficient market.

Second, we regress anomalies' unexpected return news on factors' cash flow news, discount rate news, or unexpected return news. The adjusted R squared is much larger when we include cash flow news of these factors as an independent variable than including discount rate news alone. Therefore, the predictability of these factors stems mainly from the cash flow news component. Lastly, we regress the news components of the portfolios sorted and grouped by the mispricing or behavioral characteristics on the news components of the market to analyze the composition of the market beta of these groups. The market beta reveals the systematic risk which cannot be diversified away. We are interested in the variation of the beta across portfolios. The pattern in our results leads to the conclusion that though the covariance between discount rate news of the portfolio and the market constitutes a significant portion of the market beta, it is the covariance relevant to cash flow news that results in the variation of beta across portfolios.

The rest parts of this paper are organized as follows. Section 2 provides a literature review of the factor model and return variance decomposition. Section 3 talks about the underlying theoretical framework which inspires the empirical design. Section 4 records the data sources and the variable construction. We provide summary statistics and regression results testing the validity of these reconstructed factors. Section 5 presents all the results and section 6 concludes.

2 Literature Review

My paper builds on two branches of literature, the factor models and present value decomposition. The former motivates our study into the newly proposed behavioral or mispricing-based anomalies, while the latter provides us with the tool to explore them in a new approach.

2.1 Factor models

In the field of asset pricing, the central question is what determines asset price. The most intuitive way to look at the price is that it should be an expected discount value of tomorrow's total payoff. So determining the price falls to resolving the stochastic discount factor (SDF). In the most famous consumption-based Capital Asset Pricing Model (CCAPM), the SDF is just the discounted marginal rate of substitute on consumption.

However, since the end of the last century, both the academic and empirical worlds have recorded many anomalies and claimed the failure of some classic asset pricing models like CCAPM—up to today, there is even an anomaly zoo in which over 400 anomalies are found. After the notion of "factor zoo" from Cochrane (2011), Harvey, Liu, and Zhu (2016), Harvey (2017), Harvey and Liu (2019) also call attention on the finding of new anomalies. Are they true anomalies or merely the outcome of data mining or wrong statistics or other reasons? How could we manage to explain them? We have some factor models developed to rescue this situation. For example, Fama-French three-factor model (Fama and French (1992)) use market excess returns, and the long-short returns on the BM, size portfolio as factors in their model to explain the cross-section of stock return. Fama-French five-factor model (Fama and French (2015)) add a profitability factor and investment factor to their previous three-factor model. Hou et al. (2015), propose the q-factor model inspired by the investment-based asset pricing and it contains market excess returns, investment, and roe factors. These factors are mainly based on the firms' accounting characteristics.

To reduce the space of anomalies to parsimonious factors or to compare among factors, there are also papers applying the statistical or machine learning models like LASSO, Principle Component Analysis, similarity analysis, and proposing new Bayesian models, see Feng, Giglio, and Xiu (2020), Bryzgalova, Huang, and Julliard (2019), Kozak, Nagel, and Santosh (2020), Lettau and Pelger (2020), and Giglio, Kelly, and Xiu (2022) among others.

Diverging from rational asset pricing, some authors propose new models with factors motivated by behavior or mispricing and they claim that it works better in spinning the mean-variance frontier or explaining more anomalies. Daniel et al. (2020) propose a financing factor (FIN) to capture the long-run mispricing based on the managers' decisions to issue or repurchase equity, and post-earning announcement drift factor (PEAD) to capture the short-run mispairing based on investors' inattention and slow responding to the news, along with a market factor in their model. The FIN factor is constructed using 1-year net-share-issuance (NSI), and 5-year composite-share-issuance (CSI) measures to account for the managers' timing to exploit the stock's mispricing in the interest of other investors. The PEAD factor is formed using the earnings surprise after firms' announcements to address investors' limited attention. Stambaugh and Yuan (2017), on the other hand, include market factor, size factor and a MISPRICING (UMO, the underpriced minus over priced in their paper) factor which aggregates information across 11 prominent anomalies by averaging their rankings[1]. By doing so, they aim to achieve a less noisy measure of a stock's mispricing.

In this paper, instead of proposing new factors, we take a different point of view to look at these factors. When explaining the cause of these factors, papers rely on the relation of these factors' characteristics with other variables like Gormsen and Lazarus (2021), if the sentiment index predicts the factors like Stambaugh and Yuan (2017), or if the factor or underlying characteristics predict future cash flows like Chen (2017) among others. The motivation for my using the present value decomposition is due to the fact that in the end, the returns are driven either by cash flow news or discount rate news. Many characteristics like ROE, investment rate, O-score, and so on contain information about either cash flow news, or discount rate news, or both. We would like to see if the variance decomposition could add more value to the shrinkage of anomalies or the comparison between factor models.

2.2 Present value decomposition

The present-value decomposition is based on the work of Campbell and Shiller (1988) who prove a log-linear return approximation between the returns, dividend yield, and dividend growth starting from the definition of returns. In order to study the sources of variation in the returns, Campbell (1991) further shows that the unexpected return news can be decomposed into cash flow news which contains information on unexpected future dividend growth and discount rate news which contains information on unexpected future returns. He finds that the future excess returns' volatility takes up to around 70% of the total variation in the unexpected return news for the aggregate market in the period from 1952 to 1988. However, since this decomposition requires information on dividend growth, while most firms do not pay dividends, it prevents us to apply this approach to the firm level. To solve this problem, Vuolteenaho (2002), uses the accounting identity to reach a similar return decomposition where he uses return on equity (roe) instead of dividend growth. It, therefore, enables decomposition at the firm level. Vuolteenaho (2002) finds that at the firm level, the CF news is the main driver, but when aggregating to the market level, the cash-flow news gets diversified, leaving the discount rate news as the main driver, which is consistent with the conclusion from Campbell (1991).

The present value decomposition is widely applied when studying the predictability of returns, systematic risk, and the driving elements of returns' variance. It also gets new development in recent years. For example, Cochrane (2008) discusses the predictability of returns and dividend growth. Maio and Xu (2020) generalize the Campbell-Shiller decomposition and study the prediction power of aggregate earnings yield. Campbell and Mei (1993), Campbell and Vuolteenaho (2004), Campbell, Polk, and Vuolteenaho (2010) use the news components to analyze the sources of market betas. Mao and Wei (2014) explain price and earnings momentum by investigating dynamics of cash flow (CF) news and discount rate (DR) news. Chen and Zhao (2009), and Engsted,

[1]They also propose a four-factor model with two "mispricing" factors. But to save space, we only use their three-factor model

Pedersen, and Tanggaard (2012) discuss the VAR-based decomposition. Callen and Segal (2004) extends the decomposition by adding an accruals news, while Cho, Kremens, Lee, and Polk (2022) take into account the investment in stock issuance and present a new CF news brought by that. De La O and Myers (2021), instead of using VAR approach to estimate the news components, employ the subjective cash flow and discount rate expectations from survey forecasts and find that cash flow growth expectations explain 93% and 63% of the variation in the S&P 500 price-dividend and price-earnings ratios. Gao and Martin (2021) exploit a measure of dividend yield to derive a new decomposition that resembles the Gordon growth model more closely and has certain other advantages. Gonçalves (2021) relates the stock return with equity strips (i.e. dividends with different maturities) and develops a term structure return decomposition. He finds roughly 60% of equity volatility comes from the present value of dividends with maturities beyond 20 years and that cash flow shocks drive volatility in short-term present values whereas discount rate news is responsible for volatility in long-term present value.

The decomposition is important as it helps us to understand what drives the price fluctuation or even payout policy and its underlying economic implication. The cash-flow news is regarded as a permanent shock and is the fundamental component of firm returns, while the discount rate news is viewed as a temporary shock, related to the investor's risk aversion or sentiment. Michaely, Rossi, and Weber (2021) argue that the CF news drives payout policy, and payout policy conveys information about future cash-flow volatility.

While previous researches mainly apply this decomposition to firm-level and aggregate market-level, and seldom the anomaly level, Lochstoer and Tetlock (2020) study the five well-known anomalies (which form the basis factors of the traditional factor models) under this approach, and they find that CF news is the main driver. I extend this literature by studying the newly proposed behavior factors.

3 Theoretical settings

In this part, we briefly exhibit the theoretical backgrounds to support our empirical approaches. The present value decomposition explains why return news equals CF news minus DR news and thus enables the variance decomposition. The beta decomposition enables us to explore the systematic beta by looking separately at the risks brought by CF or DR news.

3.1 PV Decomposition

Starting from the definition of returns, Campbell and Shiller (1988) reaches a dividend-ratio model i.e. dynamic Gordon model where they show mathematically that log dividend price ratio equals to a constant plus the sum of expected discounted value of all future one-period "growth-adjusted discount rates" ($e_{t+j} - \Delta d_{t+j}$). Build on this work, Campbell (1991) shows that unexpected return news equals CF news minus DR news as follows:

$$\begin{aligned} r_{t+1} - E_t r_{t+1} &= (E_{t+1} - E_t) \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} + (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} \\ &= N_{CF,t+1} - N_{DR,t+1} \end{aligned} \tag{1}$$

where Δd_{t+1+j} is the dividend growth in period $t+1+j$, $(E_{t+1}-E_t)x_j = E_{t+1}[x_j] - E_t[x_j]$, and r_t is the log return at period t.

The most inspiring step is the Taylor first-order expansion which enables one to write $log(A+B)$ as a linear combination of $log(A)$ and $log(B)$.

However, we now document the deduction from Vuolteenaho (2002) because this approach makes it possible to decompose returns at the firm level instead of the aggregate level by using return on equity (ROE) instead of dividend growth as the basic cash-flow fundamental (as many firms do not pay dividends and this prevents us from constructing dividend growth needed in equation (1) and it also sheds light on the variables included in the vector autoregression system.

By assuming zero equity issuance[2], we have the clean-surplus identity:

$$B_{t+1} = B_t + Y_{t+1} - D_{t+1}N_t \tag{2}$$

where B_t is the book value of equity at time t, Y_t is the total earnings at t, D_t is the dividend per share, and N_t is the number of shares.

Multiplying $\frac{N_t}{B_t}$ on both sides of the definition equation on return $P_t = \frac{1}{1+R_{t+1}}(D_{t+1}+P_{t+1})$ gives us:

$$\begin{aligned}\frac{P_tN_t}{B_t} &= \frac{1}{1+R_{t+1}}(D_{t+1}+P_{t+1})\frac{N_t}{B_t}\\ \frac{M_t}{B_t} &= \frac{1}{1+R_{t+1}}(D_{t+1}+P_{t+1})\frac{N_t}{B_t}\end{aligned} \tag{3}$$

as $M_t = P_tN_t$ and it is the market value of equity.

Multiplying $\frac{N_t}{N_{t+1}}$ which equals 1 to the left-hand side of equation (2) and solve for N_t we can obtain N_t as $N_t = \frac{B_t+Y_{t+1}}{D_{t+1}+B_{t+1}/N_{t+1}}$. Plugging it into (3) leads to:

$$\begin{aligned}\frac{M_t}{B_t} &= \frac{1}{1+R_{t+1}}(D_{t+1}+P_{t+1})\frac{1}{B_t}\frac{B_t+Y_{t+1}}{D_{t+1}+B_{t+1}/N_{t+1}}\\ &= \frac{1}{1+R_{t+1}}\frac{B_t+Y_{t+1}}{B_t}\frac{D_{t+1}+P_{t+1}}{D_{t+1}+B_{t+1}/N_{t+1}}\\ &= \frac{1}{1+R_{t+1}}\frac{B_t+Y_{t+1}}{B_t}\frac{1}{D_{t+1}+B_{t+1}/N_{t+1}}\left[D_{t+1}+\frac{P_{t+1}N_{t+1}B_{t+1}/N_{t+1}}{B_{t+1}}\right]\\ &= \frac{1}{1+R_{t+1}}\left(\frac{B_t+Y_{t+1}}{B_t}\right)\left(\frac{D_{t+1}}{D_{t+1}+B_{t+1}/N_{t+1}}+\frac{P_{t+1}N_{t+1}}{B_{t+1}}\times\frac{B_{t+1}/N_{t+1}}{D_{t+1}+B_{t+1}/N_{t+1}}\right)\\ &= \frac{1}{1+R_{t+1}}\left(\frac{B_t+Y_{t+1}}{B_t}\right)\left(1-\frac{B_{t+1}/N_{t+1}}{D_{t+1}+B_{t+1}/N_{t+1}}+\frac{P_{t+1}N_{t+1}}{B_{t+1}}\times\frac{B_{t+1}/N_{t+1}}{D_{t+1}+B_{t+1}/N_{t+1}}\right)\end{aligned} \tag{4}$$

[2]That is, if N_t is the number of shares, then we have $N_t = N_{t-1} = N_{t+1}$. This assumption ignores the future book equity investment through share issuance or repurchase. Cho et al. (2022) relax this assumption by allowing the change in the issuance. They then introduce a new variable as a source of cash flow and divide the cash flow news into two components, one brought by investment and another by profitability. We do not follow this approach because it makes CF news more complex, and we consider other variables in the VAR to account for the investment.

Define $\Lambda_{t+1} = \frac{B_{t+1}/N_{t+1}}{D_{t+1}+B_{t+1}/N_{t+1}}$ which is the plowback ratio, then (4) becomes:

$$\begin{aligned}\frac{M_t}{B_t} &= \frac{1}{1+R_{t+1}} \times (1+ROE_{t+1}) \times \left(1 - \Lambda_{t+1} + \frac{M_{t+1}}{B_{t+1}}\Lambda_{t+1}\right) \\ &= \frac{1}{1+R_{t+1}} \times (1+ROE_{t+1}) \times \left(1 + \left(\frac{M_{t+1}}{B_{t+1}} - 1\right)\Lambda_{t+1}\right)\end{aligned} \tag{5}$$

where $ROE_{t+1} = Y_{t+1}/B_t$ is the return on equity.

Taking log on both sides of (5):

$$mb_t = -r_{t+1} + roe_{t+1} + log[1 + (exp(mb_{t+1}) - 1)exp(\lambda_{t+1})] \tag{6}$$

where $mb_t = log(M_t/B_t)$, $r_{t+1} = log(1+R_{t+1})$, $roe_{t+1} = log(a+ROE_{t+1})$, $\lambda_{t+1} = log(\Lambda_{t+1})$.

Apply the first-order Taylor approximation[3], we can get the log-linear present-value identity in Vuolteenaho (2002):

$$mb_t \approx -r_{t+1} + roe_{t+1} + \rho mb_{t+1} \tag{7}$$

where ρ is set to 0.967.

Iterating mb forward and assuming $lim_{j\to\infty}\rho^j mb_{t+j} = 0$, we can obtain:

$$\begin{aligned}mb_t &\approx -r_{t+1} + roe_{t+1} + \rho mb_{t+1} \\ &= -r_{t+1} + roe_{t+1} + \rho(-r_{t+2} + roe_{t+2} + \rho mb_{t+2}) \\ &= (-r_{t+1} - \rho r_{t+2}) + (roe_{t+1} + \rho roe_{t+1}) + \rho^2 mb_{t+2} \\ &= (-r_{t+1} - \rho r_{t+2}) + (roe_{t+1} + \rho roe_{t+2}) + \rho^2(-r_{t+3} + roe_{t+3} + \rho mb_{t+3}) \\ &= (-r_{t+1} - \rho r_{t+2} + \rho^2 r_{t+3}) + (roe_{t+1} + \rho roe_{t+2} + \rho^2 roe_{t+3}) + \rho^3 mb_{t+3} \\ &= (-r_{t+1} - \rho r_{t+2} + \rho^2 r_{t+3} + ...) + (roe_{t+1} + \rho roe_{t+2} + \rho^2 roe_{t+3} + ...) + \rho^\infty mb_{t+\infty} \\ &= -\sum_{j=1}^{\infty}\rho^{j-1}r_{t+j} + \sum_{j=1}^{\infty}\rho^{j-1}roe_{t+j}\end{aligned} \tag{8}$$

Using (8), we have the following difference between the conditional expectation of bm_{t-1} at

[3] Recall the first order Taylor approximation of $f(x,y)$ around (x_k, y_k):

$$f(x,y) \approx f(x_k,y_k) + (x-x_k)f'_x(x_k,y_k) + (y-y_k)f'_y(x_k,y_k)$$

Taking the first order Taylor approximation of $log[1+(exp(x)-1)exp(y)]$ around $x=0$ and $y = log(\rho)$ gives us $log[1+(exp(x)-1)exp(y)] \approx log[1+(exp(0)-1)exp(log(\rho))] + (x-0)\left[\frac{exp(x)exp(y)}{1+(exp(x)-1)exp(y)}\right]_{x=0,y=log(\rho)} + (y - log(\rho))\left[\frac{(exp(x)-1)exp(y)}{1+(exp(x)-1)exp(y)}\right]_{x=0,y=log(\rho)} = log(1+0) + x\left[\frac{1\times\rho}{1+0}\right] + (y-log(\rho)) \times 0 = x\rho$. Use mb_{t+1} to substitute x, and we end at $log[1+(exp(mb_{t+1})-1)exp(\lambda_{t+1})] \approx mb_{t+1}\rho$.

time t and t-1:

$$\begin{aligned}
&bm_{t-1} = -mb_{t-1} \approx \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j-1} - \sum_{j=1}^{\infty} \rho^{j-1} roe_{t+j-1} = \sum_{j=0}^{\infty} \rho^{j} r_{t+j} - \sum_{j=0}^{\infty} \rho^{j} roe_{t+j} \\
&E_t\left[bm_{t-1}\right] - E_{t-1}[bm_{t-1}] = 0 = E_t\left[\sum_{j=0}^{\infty} \rho^{j} r_{t+j} - \sum_{j=0}^{\infty} \rho^{j} roe_{t+j}\right] - E_{t-1}\left[\sum_{j=0}^{\infty} \rho^{j} r_{t+j} - \sum_{j=0}^{\infty} \rho^{j} roe_{t+j}\right] \\
&0 = E_t\left[r_t + \sum_{j=1}^{\infty} \rho^{j} r_{t+j} - \sum_{j=0}^{\infty} \rho^{j} roe_{t+j}\right] - E_{t-1}\left[r_t + \sum_{j=1}^{\infty} \rho^{j} r_{t+j} - \sum_{j=0}^{\infty} \rho^{j} roe_{t+j}\right] \\
&E_t(r_t) - E_{t-1}(r_t) = (E_t - E_{t-1}) \sum_{j=0}^{\infty} \rho^{j} roe_{t+j} - (E_t - E_{t-1}) \sum_{j=1}^{\infty} \rho^{j} r_{t+j}
\end{aligned} \tag{9}$$

It gives us a similar decomposition as in equation (1):

$$\begin{aligned}
r_{t+1} - E_t r_{t+1} &= (E_{t+1} - E_t) \sum_{j=0}^{\infty} \rho^{j} \Delta d_{t+1+j} + (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^{j} r_{t+1+j} \\
&= N_{CF,t+1} - N_{DR,t+1}
\end{aligned} \tag{10}$$

which implies that the unexpected stock return is due to two components: news about future cash flow (roe) and news about the future discount rate.

3.2 Beta Decomposition

Though the CAPM does not price many assets well, its idea of describing stock returns as a risk measure times risk premium is influential, and we still care about the market beta as it measures the asset's risk relative to the aggregate market. Campbell and Vuolteenaho (2004) decompose the market beta into good beta and bad beta to explain the size and value anomalies. Campbell et al. (2010) further decompose the bad beta and good beta into four beta components, which breaks the returns of asset and market into CF and DR news components, and then analyzes their covariance.

The market beta of asset i is:

$$\beta_{i,M} \equiv \frac{\mathrm{Cov}_t\left(r_{i,t+1}, r_{M,t+1}\right)}{\mathrm{Var}_t\left(r_{M,t+1}\right)} = \frac{\mathrm{Cov}_t\left(N_{i,t+1}, N_{M,t+1}\right)}{\mathrm{Var}_t\left(r_{M,t+1}\right)} \tag{11}$$

where $r_{i,t+1}$, $r_{M,t+1}$ are asset excess return and market factor return at time $t+1$ respectively, $N_{i,t+1}$ is the unexpected return news of asset i at time $t+1$ and $N_{M,t+1}$ is the unexpected return news of market factor at time $t+1$. We use the return news in place of returns because the expectation is just a constant.

As we can decompose return news as CF news minus DR news shown in the last section. Breaking market return news gives us bad cash flow beta ($\beta_{i,CFM}$) and good discount rate beta

($\beta_{i,DRM}$):

$$\beta_{i,CFM} \equiv \frac{\mathrm{Cov}_t\left(N_{i,t+1}, N_{M,CF,t+1}\right)}{\mathrm{Var}_t\left(r_{M,t+1}\right)} \tag{12}$$

$$\beta_{i,DRM} \equiv \frac{\mathrm{Cov}_t\left(N_{i,t+1}, -N_{M,DR,t+1}\right)}{\mathrm{Var}_t\left(r_{M,t+1}\right)} \tag{13}$$

where $N_{M,CF,t+1}$ and $N_{M,DR,t+1}$ are the market CF news and DR newsat time $t+1$

We can further decompose asset return news similarly and get the four beta components:

$$\beta_{CFi,CFM} = \frac{\mathrm{Cov}_t\left(N_{i,CF,t+1}, N_{M,CF,t+1}\right)}{\mathrm{Var}_t\left(r_{M,t+1}\right)} \tag{14}$$

$$\beta_{DRi,CFM} = \frac{\mathrm{Cov}_t\left(-N_{i,DR,t+1}, N_{M,CF,t+1}\right)}{\mathrm{Var}_t\left(r_{M,t+1}\right)} \tag{15}$$

$$\beta_{CFi,DRM} = \frac{\mathrm{Cov}_t\left(N_{i,CF,t+1}, -N_{M,DR,t+1}\right)}{\mathrm{Var}_t\left(r_{M,t+1}\right)} \tag{16}$$

$$\beta_{DRi,DRM} = \frac{\mathrm{Cov}_t\left(-N_{i,DR,t+1}, -N_{M,DR,t+1}\right)}{\mathrm{Var}_t\left(r_{M,t+1}\right)} \tag{17}$$

Through this process, we decompose the market beta into four components:

$$\beta_{i,M} = \beta_{i,CFM} + \beta_{i,DRM} = \beta_{CFi,CFM} + \beta_{DRi,CFM} + \beta_{CFi,DRM} + \beta_{DRi,DRM} \tag{18}$$

4 Empirical Design

We document the data source, the construction of variables, sample filtration, and the procedure to extract news in this section.

4.1 Data and Factor Constructions

Various sources of data are employed. In summary, we get common stocks information in NYSE, Nasdaq, and AMEX, and inflation to calculate returns from CRSP, firms' accounting information to get variables used in VAR system, and some anomaly characteristics from Compustat, Fama-French three-factors, and value spread from French's data library, aggregate predictors provided by Welch and Goyal (2008), historical book equity data from Davis, Fama, and French (2000), certain anomaly characteristics provided by Chen and Zimmermann (2021), and treasury yield fromLochstoer and Tetlock (2020).

For variables to be contained in the VAR, i.e. returns, return on equity (ROE), book-to-market ratio (BM), Profitability (Prof), Investment (Inv_M5), five-year change in log market equity ME_D5), and six-month momentum (Mom6), we construct them exactly the same as in Lochstoer and Tetlock (2020) by modifying their codes to establish the final sample. Stock returns are annualized from July to June next year, adjusted by deducting the inflation rate. ROE is the return on equity, calculated as earnings available for common over last year's book equity. BM is the book equity defined by Fama and French (1992) in December of year t[4] divided by the market capitalization in June of year $t + 1$. When the book value of equity is missing, the historical book equity is used to supplement. Prof is annual revenues minus costs of goods sold, interest expense, and selling, general, and administrative expenses, divided by book equity. Inv_M5 is five-year growth of the total asset. ME_D5 is the five-year change in market equity of June. Mom6 is the six-month cumulative return from January to June. We take log to returns and Mom6, and transform the other variables mentioned above by adding one and taking log. To avoid problems of extreme values when taking log (say, when ROE is near -1), we define pseudo-firms as a portfolio of 90 percent common stock and 10 percent Treasury bills, and adjust the variables accordingly.

For variables used to construct the anomalies, we obtain the announcement return (announcementreturn), five-year composite issuance (compequiss), accruals (accruals), net operating asset (noa), one-year asset growth (assetgrowth), distress probability (failureprobability), O-score (OScore), momentum in last 12 months (mom12), gross profitability (gp), return on asset (roaq) from the data provided by Chen (2022). We construct the net stock issues (lnNS) as annual log change in split-adjusted outstanding shares following Fama and French (2008), and investment-to-asset (ioa) as changes in gross property, plant, and equipment plus changes in inventory, divided by lagged total assets following Stambaugh and Yuan (2017). We carefully align these variables with annual returns to ensure data availability and timeliness when constructing anomalies.

Several restrictions are applied to the sample. We require firms not to get delisted in June, and have trading information, have market capitalization larger than 10 million, and have BM between 0.01 and 100 in June. We also only keep firms with the fiscal-year end in December, with prices larger than or equal to five, and exclude tiny firms which are in the bottom NYSE quintile. These filtrations are common, especially when forming anomalies to avoid microstructure effects. Our final sample starts from 1964 to 2019.

All the twelve anomalies (which are named exactly the same as the characteristic variables used to construct them) are value-weighted long-short portfolio returns. We group all stocks into ten groups but only use stocks listed on NYSE to get the breakpoints. For factors FIN, PEAD from Daniel et al. (2020), and MISPRICING, SIZE from Stambaugh and Yuan (2017) we follow the same procedure of sorting stock into 2×3 portfolios (except SIZE factor) and calculate the value-weighted returns. The most distinctive difference is that instead of a monthly updated PEAD, MISPRICING, SIZE, we rebalance the factors yearly at the end of June each year. The reason for doing so is that the news components are extracted by regressing VAR which consists of accounting information at the yearly frequency. If we have factors that are monthly rebalanced, what is the corresponding monthly news? It is not available. To investigate the potential adverse consequences of this practice, we provide the summary statistics for the factors in table 1, and the regression intercepts of using these models to explain anomaly returns in table 2.

From table 1, MISPRCING (misp) delivers the highest average return of 0.0704, SIZE (size) delivers the lowest return of 0.0323. All of the returns are statistically different from zero as their t values are larger than 2.38. We check the explanatory power of the mispricing model of Stambaugh and Yuan (2017) in panel A of table 2, and of the behavioral model of Daniel et al. (2020) in panel B. In each column, we regress the anomaly returns on the factors included in each model. If the model explains the anomaly, the intercept should be non-significant from

[4] We restrict our sample to firms with a fiscal-year end in December to make the variables in the VAR logical in timing, and it simplifies the time alignment of book value as a bonus.

zero. The mispricing factors we construct cannot explain announcement return and accruals, and the behavioral factors we construct cannot explain net issuance, accruals, and asset growth. Many papers document that using more timely information benefits the factors (see Asness and Frazzini (2013), Barillas and Shanken (2018)), given that we only update our factors at a yearly frequency, it is reasonable and acceptable to see certain anomalies not explained. So we regard the construction of these factors as satisfactory and continue explorations on them.

4.2 Recover News at Different Levels

In order to extract the news components at the portfolio, anomaly, or mean-variance efficient portfolio levels, there are two possible approaches. One is to find the corresponding portfolio-level variables included in the VAR system and then get the news components directly at the portfolio level. Another is to aggregate firm-level news components to portfolio-level. The potential problem with the first approach, as explained in Lochstoer and Tetlock (2020), is that the cash flows and discount rates of rebalanced portfolios can differ substantially from those of the underlying firms in the portfolios. If we are only interested in the properties of the portfolio level instead of its underlying firms' properties, the concern is minor. But the main issue which hinders us to do so is that we do not know how to construct the corresponding variables in the long-short portfolio. What is the book-to-market ratio, or the dividend yield for a portfolio which is long in firms with high book-to-market ratios and short in firms with low book-to-market ratios? If it is reasonable to calculate the book-to-market ratio for a portfolio that takes a long position in its constituting stocks as the sum of the book value of all these stocks over the sum of the market capitalization, or as the simple-weighted, or value-weighted averages of the stocks constituting the portfolio, it by no means is sensible to do so for a long-short portfolio. This leaves us nothing else to do but to apply the second approach—aggregating firms' news into the portfolios' news following Lochstoer and Tetlock (2020).

To this end, we first need to extract the firm-level news in the following three steps.

First, estimate a time-series Vector Autoregression (VAR) for the aggregate market-level return decomposition. We can do so because the composition of the market portfolio is not rebalanced unless a firm quits the stock market.

$$Z_{t+1} = \mu^{agg} + A^{agg} Z_t + \varepsilon_{t+1}^{agg}$$

where $Z_t = [r_t^{agg}, roe_t^{agg}, bm_t^{agg}, prof_t^{agg}, inv_t^{agg}, me_t^{agg}, mom6_t^{agg}]'$ with each variable as the cross-sectional value-weighted average of the firm-level variables (all in log form). We do not restrict the roe_t^{agg} only as a dependent variable as Lochstoer and Tetlock (2020) do because our sample period is after 1964 and we can distinguish between ROE and profitability.

Then the unexpected return news, discount rate news, and cash flow news at the aggregate level are calculated as[5]:

$$r_{t+1}^{agg} - E_t r_{t+1}^{agg} = e_1' \varepsilon_{t+1}^{agg}$$

[5] To see how we obtain the formula for calculating DR news, see Callen and Segal (2010). But shortly speaking, the components (difference in expectations for each period) form a geometric series with a common ratio of $\kappa \times A^{agg}$.

$$DR_{t+1}^{agg} = E_{t+1}\sum_{j=2}^{\infty}\kappa^{j-1}r_{t+j}^{agg} - E_t\sum_{j=2}^{\infty}\kappa^{j-1}r_{t+j}^{agg} = e_1'\kappa A^{agg}\left(I_7 - \kappa A^{agg}\right)^{-1}\varepsilon_{t+1}^{agg}$$

$$CF_{t+1}^{agg} = r_{t+1}^{agg} - E_t\left[r_{t+1}^{agg}\right] + DR_{t+1}^{agg} = e_1'\left(I_7 + e_1'\kappa A^{agg}\left(I_7 - \kappa A^{agg}\right)^{-1}\right)\varepsilon_{t+1}^{agg}$$

where e_1 is a 7×1 column vector with one as its first element and zeros elsewhere, I_7 is a 7×7 identity matrix, $\kappa = 0.967$ as in Vuolteenaho (2002). The CF news, as implied in the equations above, is calculated residually from the present-value identity which states that the unexpected return news equals to the CF news minus DR news. There exists the possibility that the unexpected return news contains shocks other than CF and DR news. Therefore, the CF news we get may be larger than actual CF news, considering that other shocks are included in the CF news. Another approach that may solve this problem is to calculate CF news directly as we do with the DR news. But as explained in Lochstoer and Tetlock (2020), it does not capture the CF for stockholders correctly.

Second, we estimate a panel VAR for the demeaned firm-level return decomposition.

$$Z_{i,t+1} = \mu^{ma} + A^{ma}Z_{i,t} + \varepsilon_{i,t+1}^{ma}$$

where $Z_t = [r_{i,t}^{ma}, roe_{it}^{ma}, bm_{it}^{ma}, prof_{it}^{ma}, inv_{it}^{ma}, me_{it}^{ma}, mom6_{it}^{ma}]'$ with each variable as the firm-level variable demeaned by the cross-sectional value-weighted average of that variable, say $r_{it}^{ma} = r_{it} - r_t^{agg}$.

Then the demeaned firm-level news components are calculated similarly as at the aggregate level:

$$r_{i,t+1}^{ma} - E_t r_{i,t+1}^{ma} = e_1'\varepsilon_{i,t+1}^{ma}$$

$$DR_{i,t+1}^{ma} = E_{t+1}\sum_{j=2}^{\infty}\kappa^{j-1}r_{i,t+j}^{ma} - E_t\sum_{j=2}^{\infty}\kappa^{j-1}r_{i,t+j}^{ma} = e_1'\kappa A^{ma}\left(I_7 - \kappa A^{ma}\right)^{-1}\varepsilon_{i,t+1}^{ma}$$

$$CF_{i,t+1}^{ma} = r_{i,t+1}^{ma} - E_t\left[r_{i,t+1}^{ma}\right] + DR_{i,t+1}^{ma} = e_1'\left(I_7 + e_1'\kappa A^{ma}\left(I_7 - \kappa A^{ma}\right)^{-1}\right)\varepsilon_{i,t+1}^{ma}$$

where e_1, I_7, κ are the same defined vectors or scalar. We use the inverse of the number of firms in each year as a weight to each firm in that year, so the weighted least square regression weights each year equally following Vuolteenaho (2002).

Third, the firms' total news components are defined as the sum of the corresponding aggregate-level news and demeaned firms' news:

$$\begin{aligned}
r_{i,t+1} - E_t r_{i,t+1} &= (r_{t+1}^{agg} - E_t r_{t+1}^{agg}) + (r_{i,t+1}^{ma} - E_t r_{i,t+1}^{ma}) \\
&= e_1'\varepsilon_{t+1}^{agg} + e_1'\varepsilon_{i,t+1}^{ma} \\
DR_{i,t+1} &= DR_{t+1}^{agg} + DR_{i,t+1}^{ma} \\
CF_{i,t+1} &= CF_{t+1}^{agg} + CF_{i,t+1}^{ma}
\end{aligned}$$

As Lochstoer and Tetlock (2020) explains, this procedure allows the VAR coefficients to differ for the common movement (A^{agg}) and firms' idiosyncratic movement (A^{ma}) in order to match the data.

Next, we need to aggregate the firm-level news components into the different portfolio levels.

To form portfolio/anomaly/factor level news, we follow the same procedure of calculating returns except that we use firms' news to replace firms' returns. At the end of June of year t, all information needed to form portfolios is known. We need to assume that all the (long-short) portfolios are held for one year. The assumption is necessary because the availability of firms' CF or DR news is subject to financial information in the VAR system and they only reflect yearly news instead of monthly or quarterly news. Therefore, for factors that are rebalanced every month or every quarter like those in Chen and Zhang, there is no way to organize the analysis in this way unless we only keep the construction spirit but lower the updating frequency to yearly.

For example, we sort firms on their book-to-market ratios and group them evenly into ten portfolios in June of year t. Then the ten portfolios' returns are the simple-weighted or value-weighted returns using next year's returns. The calculation for the news is similar. The CF news of each portfolio is the simple-weighted or value-weighted CF news of the firms in each portfolio in the next holding year. The unexpected return news and DR news are calculated in the same way.

For the long-short portfolio, we use the news of the portfolio in the long position minus the news of the portfolio in the short position. For the factor, it follows the same logic. Throughout this paper, we use NYSE breakpoints to divide all stocks in the sample into ten groups and use market capitalization as the weights. For factors, we follow the exact procedure in the original papers except that we only update them once a year, instead of monthly updated version of PEAD factor from Daniel et al. (2020), for example.

For the news components at the MVE level, we need to calculate the weights for each factor included by maximizing the full-sample Sharpe ratio. For example, for the MVE portfolio composed of FIN, PEAD, Market factors, the corresponding weight is calculated as the inverse of their covariance matrix times the mean of these factors. The MVE news components are then the value-weighted sum of the factor news components, say the cash flow news of the MVE portfolio is $CF_t^{MVE} = \sum_{i\in\{FIN,PEAD,Market\}} w_i \times CF_t^i$. where w_i is the weight associated with factor i to achieve the maximal Sharpe ratio, and CF_t^i is the CF news for factor i.

The output from all the aggregation is a time-series data of news for each portfolio.

5 Empirical Results

Now we present the estimation results for the VAR specified above, decompose the returns to CF news and DR news to study the main driver for the variation of returns at different levels, explore the source of the behavioral or mispricing-based factors' explanatory power, and analyze the systematic risk associated with these factors.

5.1 VAR Estimation

As Chen and Zhao (2009) point out, the specification of VAR affects the relevant importance of CF news and DR news. The conclusion may contradict each other under different model settings. Our specification of the variables included in the vector follows the main specification in Lochstoer and Tetlock (2020) except that we do not restrict the ROE to be on the dependent variable, as explained in the empirical design part. Returns, book-to-market ratio, and ROE are necessary to be contained in the system, especially for the firm-level decomposition, because they are specified in the deduction of return decomposition in Vuolteenaho (2002). The other variables add value to the prediction of returns. Lochstoer and Tetlock (2020) shows that this specification provides a reasonable approximation o the long-run dynamics of returns and earnings. At the aggregate level, the more frequently seen variables are dividend growth, dividend yield, and some other variables at the aggregate level, say value spread, eqis (the ratio of equity issuing activity as a fraction of total issuing activity), etc. However, they mainly apply the decomposition to the stock index like the value-weighted NYSE stock index where we have well-defined relevant variables available. In our paper, it is necessary to study the firm-level returns, and as most firms do not pay dividends, we can only apply the firm-level decomposition from Vuolteenaho (2002). Besides, we are interested in the common movement composed of firms in our sample, so what we do is to use the value-weighted variables we use to learn firms' return decomposition. But we also include other variables: term yield spread, the default yield spread, and the small stock defined in the data part as a robustness check, and it turns out they do not affect the results that much.

Table 3 shows coefficients, and t statistics (in parenthesis) for the time-series aggregate VAR, and table 4 shows those for the panel VAR. The sample period for VAR estimation is from 1964 to 2019 to enlarge the data used in the estimation instead of from 1974 when some anomalies like PEAD are available due to data issues.

From table 3 we can see that the accounting variables—ROE, BM, profitability, investment, and market capitalization, as usually more persistent, are better predicted using past information. The adjusted R^2s for these variables are moderate or large. For return variables, i.e. the value-weighted real returns, and the value-weighted momentum returns, the adjusted R^2s are much smaller, 6% and 5% respectively. It reveals the fact the returns are less predictable, at least for the short term. Last year's return does not convey any information for this year's return. Among all the other dependent variables, only investment (asset growth in the past five-year) exhibits a significant coefficient, which manifests the long-run predictability of market returns.

At the firm-level panel regression, we adopt the common practice to weight each year equally by using weighted least squares. As firm-level variables contain more idiosyncratic information, the adjusted R^2s are smaller than in the aggregate VAR. Only 2% variation in returns is explained by the other characteristics, which is routine when predicting firm returns. For the value-weighted market returns, the signs of the coefficients on dependent variables are consistent with previous studies, though some of them are not significant.

5.2 Decomposing Returns

Once we obtained the coefficients and errors from the VARs, we can compute the news at our desired level and analyze the relevant importance of the news components for the variation of unexpected return news, and also for the explanatory power.

Table 5 shows the variance decomposition for firm-level and market-level returns. We see from the first row that consistent with previous studies like Campbell (1991) and Vuolteenaho (2002) among others, the discount rate news is the main driver of the market. The variance of DR news over the variance of unexpected return news is 85.49%, quite a large proportion. The variance of CF news, or the covariance components only takes up to 17.88% and -3.37% respectively. In total, the sum of them equals one as we use the present value identity to extract the CF news. The correlation between CF news and DR news is only 0.0432. It is quite reasonable as at the aggregate market level, the CF news, which is usually idiosyncratic, gets diversified away, and what is left is the variation in the DR news which usually stands for the market sentiment. The second row shows that for the demeaned firm-level returns, CF news accounts for 95.74% of the total variation in the unexpected returns, leaving DR news and their covariance accounting for less than 10% of the change. The correlation between CF news and DR news is again vary small, with a number of less than 0.1. The decomposition for total firm news is shown in the third row, where we observe that though the contribution of DR news increases as we add back the common movement components among firms, it is still the CF news that is the main driver of returns.

We also look at the anomalies' variance decomposition. For the underlying eleven anomalies which are used to construct the behavioral or mispricing-based factors, we report the results in table 6. All anomalies are rebalanced yearly, using NYSE breakpoints to be divided into ten groups, and formed using the long-short value-weighted method. Consistent with Lochstoer and Tetlock (2020) in which they study five well-known traditional anomalies—value, size, profitability, investment, and momentum, and find that the systematic CF news drives the returns of anomaly portfolios, we find similar results using a different set of anomalies. The ratio of the variance of CF news over the variance of unexpected news ranges from 84.51% to 118.34%. The variation of DR news ranges from 3.34% for the composite equity issuance anomaly to 12.74% for the momentum anomaly. The correlation between the CF news and DR news ranges from 0.0049 to 0.40 in absolute values.

When it comes to our most concerning factors and their MVE portfolios in this paper, table 7 presents the results for the decomposition.

Panel A shows the decomposition for every single factor. For FIN which stands for the long-run managers' equity issuance, CF news takes up nearly 90% of the variation in the unexpected return news. There are many possible explanations for the negative relationship between stock issuance of future returns. From the behavioral-based view, one possible interpretation is that firms' managers who know the intrinsic value of their stocks better as insiders, are willing to issue stocks when they think their stock is currently overpriced and can earn money at the expense of outside investors, and therefore future stock return decreases. A possible rational-based explanation is that when firm has steady cash flow revenues, managers have spare money to pay back to the equity holders by repurchasing stocks (a form of payout policy that is more often applied nowadays to pay dividends, see Farre-Mensa, Michaely, and Schmalz (2014)), thus the issuance decreases but the future return increases. Our result of CF news driving the most variation in returns provides evidence favoring the latter view, though it does not reject the existence of behavioral motivation. The PEAD factor, usually, is regarded as an outcome of short-run investors' underreaction to market information. However, there are also rational explanations for this, see Fink (2021) for a review of all these explanations. In our table, the CF news takes up 77.01% of the variation in the return news. As CF news epitomizes the fundamental risk, it indicates that the risk-based explanation plays a large role for the presence of this anomaly.

The MISP factor, which synthesizes the underlying eleven anomalies, aggregates the information contained in management and firms' performance. Stambaugh and Yuan (2017) argues that this factor is consistent with a mispricing interpretation as it is predicted by sentiment. However, the return decomposition says that the CF news is the main driver of the return news with a portion of 92.71% in the variation of return news. For the SIZE factor from Stambaugh and Yuan (2017), they construct it using a subset of stocks which are most likely not subject to mispricing. We see that again, the CF news is the main driver. The correlation between CF news and DR news is negative for individual anomalies, and the magnitude is larger compared with firm-level and market-level decomposition.

Panel B exhibits the decomposition for MVE portfolios in which the market factor is always included. The row "All Factors" means that we incorporate FIN, PEAD, MISP, SIZE factors, and market factors in the composition of the MVE portfolio. The row "Behavioral Factors" implies that we have FIN, PEAD, and market factors included, while the row "Mispricing Factors" means the MVE portfolio only includes MISP, SIZE, and market factors. Across all the specifications, the variation of CF news takes up 68.25% to 94.95% in the variance of return news. The DR news plays a more important role in the portfolio consisting of behavioral factors, as they are more prone to be affected by the market timing of managers or the underreaction of investors compared with mispricing factors. Panel C presents the results for MVE portfolio without the market factor. The pattern is similar to that in panel B. When the market factor whose DR news contributes most of the return's variation, is excluded, we see that the var(DR) decreases in general.

Along with the discussion of the results, we see that the CF news is always the significant driver for returns at every factor level and MVE portfolio level except the market factor. This trend highlights the common fundamental risk contained in these anomalies/factors and thus supports the risk-based explanation for them.

Next, we turn to the analysis of the explanatory power of these factors. Previous studies usually argue that their factor model is better because it expands the efficient frontier spanned by the previous factor model, and because they explain many anomalies' returns as delivering non-significant intercepts in the factor model regression. Yet nobody knows where their explanatory power comes from. In this paper, since the unexpected returns are decomposed into CF news and DR news components, it enables us to investigate from which news components these factors explain the anomalies.

Table 8 demonstrates the adjusted R^2 when regressing anomalies' return news on the CF news, DR news, and return news of the factors in each factor model. The CAPM in which only the market factor is included does not explain these anomalies. As a result, the news components of the market factor have little explanatory power as shown in panel A. We take the noa anomaly as an illustration. When we regress the unexpected return news of noa on the CF news of the market factor, the adjusted R^2 is even negative, with a statistic of -1.88%. When we regress the same return news on the DR news of the market factor, 0.02% of the variation is explained. If the independent variable is the return news of the market, then the adjusted R^2 is 0.39%. The explanatory power of the CF news of the market factor ranges from -2.16% to 8.29%. The power of the DR news ranges from -2.17% to 7.87%, and the power of the total return news ranges from -2.13% to 9.02%. It implies that consistent with the failure of CAPM to explain anomalies, the news components of the market factor in general do not explain much of the return news of these anomalies.

Panel B shows the adjusted R^2 for the behavioral model. We regress anomaly news on the news components of FIN, PEAD, and market factors. This model explains many anomalies much better than the CAPM does. Taking stock's net issuance (lnNS), for example, the CF (DR) news of these factors explains 62.05% (14.6%) of the return news of this anomaly, and the total return news of them explains 64.94% of the anomaly. This large explanatory power is not only attributable to the close relation between this anomaly and the FIN factor which uses the information of issuance anomaly. The assetgrowth anomaly is not involved in the construction of the behavioral factors, yet it is still explained by them. The portion of return news of this anomaly explained by the CF news, DR news, and total return news of the factors is 40.3%, 13.4%, and 43.11% respectively. The observation that the CF news of factors contributes most of the explanatory power of their return news to explain the anomaly's return news is widespread. All the patterns shown in panel B also hold in panel C where we use news components from the mispricing factors, and in panel D in which we use news components of all the factors (FIN, PEAD, MISPRICING, SIZE, Market). The adjusted R^2s are the highest when we combine all factors except for anomaly announcementreturn, and roaq.

What we conclude from this section is that the CF news is the main driver of all these factors we study though they are considered more subject to behavioral or mispricing phenomena than the traditional factors which are more subject to risk-based explanations. Their explanatory power mainly comes from the CF news components as well. Notice that it is not caused by the large portion of CF news variation over return news variation, as in table 8 the adjusted R^2 captures the covariance between CF news and return news of each anomaly.

5.3 Systematic Risks

We now turn to study the systematic risks of these factors with respect to the market return as Campbell et al. (2010) who explore the value anomaly. The factors are also anomalies. The beta decomposition allows us to delve into why the portfolios sorted on these factor characteristics have different systematic risks. Table 9, 10, 11 reports for PEAD, FIN, and MISPRICING factors separately.

Portfolios are indicated in the first row of each table. We do not sort on factor characteristics directly and group stocks into five portfolios based on the quintiles as usual because the construction of these factors sorts them into three groups—L, M, H, and also because the grouping of FIN factor is determined by the interaction of two variables, and hard to sort FIN into five groups. Portfolio 1 refers to the portfolio of stocks with the lowest factor characteristic; portfolio 3 refers to the one with the highest factor characteristic; portfolio 2 is the group of stocks that are not used when forming the factor returns and portfolio 3-1 refers to the long-short portfolio. One thing I would like to note here is that for FIN, and PEAD, the factor characteristics are positively related to future stock returns, so the long-short portfolios are also the factor portfolios. For the MISPRICING, however, the mispricing measure is negatively related to future stock returns, and therefore the long-short portfolio 3-1 shown in the last column of table 11 is the reverse of forming the factor returns.

i, CFm represents the bad cash-flow beta, which reflects the covariance of the portfolio's returns news and the CF news of the market factor over the variance of market returns. We obtain the estimated bad beta by regressing each portfolio's return news on the scaled CF news of the market by timing $\frac{var(r_m^e}{var(CF_m)}$ where $var(r_m^e)$ is the variance of the market factor we construct in this paper and $var(CF_m)$ is the variance of CF news of the market factor we extract. The coefficient associated with market CF news is the beta desired. We further decompose the bad beta into two components driven either by the portfolio's DR news (DRi, CFm) or CF news (CFi, DRm) by regressing respectively the DR news and the CF news of the portfolio on the scaled market CF news.

i, DRm represents the good discount-rate beta, which reflects the covariance of the portfolio's returns news and the DR news of the market factor over the variance of market returns. We obtain the estimated bad beta by regressing each portfolio's return news on the scaled DR news of the market by timing $\frac{var(r_m^e}{var(DR_m)}$ where $var(DR_m)$ is the variance of DR news of the market factor we extract. Similarly, it is further decomposed into DRi, DRm and CFi, DRm.

Table 9 displays all the beta components associated with factor PEAD. We observe the following patterns:

First, comparing bad beta and good beta in portfolios except for the long-short portfolio, we see that consistent with Campbell et al. (2010), the good discount-rate beta takes up a larger portion of the total beta. For portfolio 1, the bad beta is 0.1124 while the good beta is 0.6480, which gives a total CAPM beta of 0.7604. The magnitude of good beta takes up around 85% of the total beta. All these betas are significant in each portfolio. Yet the good beta does not exhibit much variation across portfolios. The long-short factor's good beta, 0.0372 is not significantly different from 0 as its standard error is 0.0429. The bad beta for the long-short factor, on the other hand, is 0.0335 with a standard error of 0.0191.

Second, when we look at the elements (betas of DRi, CFm, and of CFi, CFm) of the bad beta (betas of i, CFm), we see that the magnitude of the portfolio's CF-driven beta is larger than that of DR-driven beta across portfolio 1, 2, and 3. The betas of DRi_CFm across portfolios 1, 2, and 3, are not only small but also non-significant. Besides, they do not vary much across portfolios and lead to a non-significant number 0.0008 with a standard error of 0.0054 in the long-short portfolio. The betas of CFi, CFm, on the contrary, are large, and significant in all portfolios, long-short one included. It means that the cash flows of stocks with large PEAD characteristic are particularly sensitive to permanent movements in aggregate stock prices proxied by the CF news of the market factor.

Third, if we turn to the elements (betas of DRi, DRm, and of CFi, DRm) of the good beta (betas of i, DRm), we see that the magnitude of the portfolio's DR-driven beta is larger than that of CF-driven beta across portfolio 1, 2, and 3. The betas of DRi_DRm across portfolios 1, 2, and 3 are large and significant, but there is little variation across portfolios. The betas of CFi, DRm, though smaller and not significant across portfolios 1, 2, and 3, take up most variation (0.0391/0.0372 v.s. -0.0019/0.0372) in the good beta in the long-short portfolio.

These patterns also hold for FIN factor as shown in table 10. For the MISPRICING factor, we have a changing pattern and one additional phenomenon calls our attention in table 11. First, the good beta (betas of i, DRm), and the beta components of good beta—betas of DRi, DRm also show variation across portfolios now, leading to a significant number of 0.1276, 0.0369 in the long-short portfolio. Second, unlike PEAD and FIN whose good betas and bad betas contribute to the total beta in the same direction, now we see that the bad beta for the MISPRICING factor is 0.0495 (negative of returns in 3-1 portfolio as noted earlier in this section), while the good beta for the MISPRICING factor is -0.1276, leading to a negative total beta. The reason why MISPRICING factor earns a positive return though its total beta is negative corresponds exactly to the explanations in Campbell and Vuolteenaho (2004) where they argue that the price for bad beta is higher than for good beta. The good beta does not benefit the returns of the factor.

Figure 1 draws the total beta and its four subcomponents. The variation across portfolios is now more apparent. The trend of the total beta is mainly driven by the betas of CFi, CFm or of CFi, DRm, which suggests that the variation of beta stems from the fundamental part of the stocks.

In sum, we have three comments. First, the bad beta always exhibits variation across portfolios while the good beta, though constitutes a large portion of the total beta, either is stable or does not contribute to the positive returns of the factor. Second, when analyzing the components of bad beta or good beta, the magnitude of betas associated with portfolios' CF news is substantial. Third, the variation of total beta is driven by the beta components relevant to the firms' fundamental CF news.

6 Conclusion

We explore what drives the newly proposed behavioral or mispricing-based factors. These factors, which are actually long-short portfolio returns, are shown to better describe the variation in returns for different stocks or portfolios. However, it is controversial about the underlying economic reasons for these factors. Are they proxies for fundamental risk or for misvaluation, and what is the source of their systematic exposure to the market returns? This paper tries to answer these questions by decomposing the factor returns into cash flow news and discount rate news, and to see which component drives the most variation in these factor returns, and doing regressions using news components.

Our results reveal the following conclusions. First, the CF news is the main contributor to the variance of returns for not only these factors but also their MVE portfolio. Second, the explanatory power largely stems from their cash flow news component to explain the returns return news of the other anomalies. Third, though the good beta constitutes a large portion of the market beta, it is always the CF component of the factors whose variation contributes to the returns.

The insight of these results is that, though the behavioral or mispricing factors are more likely to be driven by market sentiment, investor's underreaction to market information, or managers' timing to exploit the overpricing of their stocks at the expense of investors, among others that reflect an inefficient market, we find evidence supporting that they are actually primarily driven by the CF fundamentals. A possible future extension to this work is to see if this approach could add more value to the shrinkage of anomalies by delving into news components of the broad range of anomalies.

Table 1: Summary Statistics for Factor Returns

This table presents the mean (Mean), standard deviation (SD), minimum (Min), maximum (Max), and Newey-West corrected t-statistics with six lags (t) for annual value-weighted market excess returns (mkt), annual size factor returns constructed in the spirit of Stambaugh and Yuan (2017) (size), annual mispricing factor returns constructed in the spirit of Stambaugh and Yuan (2017) (misp), annual financing factor constructed in the spirit of Daniel et al. (2020) (fin), and annual post-earnings announcement drift factor constructed in the spirit of Daniel et al. (2020) (pead). The sample period is from 1972 July to 2020 June mainly due to availability of accounting information needed to construct factors.

	mkt	size	misp	fin	pead
Mean	0.0609	0.0323	0.0704	0.0645	0.0361
SD	0.1399	0.0942	0.0929	0.1503	0.0582
Min	-0.2687	-0.2302	-0.0736	-0.2215	-0.0782
Max	0.4601	0.2481	0.3963	0.5627	0.2139
t	3.0157	2.3773	5.2512	2.9733	4.3036

Table 2: Explanatory Power of Factor Model

This table shows the coefficients of regressing twelve anomalies separately on mispricing factor model (Panel A) and on behavioral model (Panel B). The numbers in parenthesis are Newey-West corrected t-statistics with six lags.

	Panel A: Mispricing Factor Model											
	lnNS	announcementreturn	compequiss	accruals	noa	assetgrowth	ioa	failureprobability	OScore	mom12m	gp	roaq
mkt	-0.1604	0.0463	0.4820***	0.0709	0.1703	-0.0655	0.0271	-0.1543	-0.0986	0.0898	0.0861	-0.2584
	(-1.5350)	(0.5505)	(4.3404)	(1.0604)	(1.5958)	(-0.5326)	(0.3833)	(-1.0173)	(-1.2621)	(0.7095)	(0.6319)	(-1.9538)
misp	0.6355***	-0.0804	0.1132	-0.2802**	0.4906**	0.2858**	0.3400	0.5152**	0.7393***	0.5122**	0.7390***	0.8105***
	(5.5149)	(-0.8109)	(0.8198)	(-2.2912)	(2.4757)	(2.2729)	(1.8948)	(2.1424)	(5.4018)	(2.0348)	(3.1763)	(3.1587)
size	-0.0472	-0.3719	-0.6582***	-0.3439**	0.2967	0.5788***	0.6539***	-0.3829	-0.4804***	-0.2835	-0.0215	-0.3795
	(-0.3129)	(-1.9265)	(-3.7198)	(-2.1018)	(1.7634)	(4.5087)	(8.5419)	(-1.6477)	(-2.8616)	(-1.2614)	(-0.0720)	(-1.6574)
_cons	0.0310	0.0499**	0.0240	0.0552***	-0.0094	0.0128	-0.0397	0.0179	-0.0300	0.0294	-0.0310	-0.0255
	(1.6161)	(2.5893)	(0.8699)	(3.6385)	(-0.3478)	(0.6144)	(-1.5082)	(0.5160)	(-1.7548)	(0.8329)	(-0.9621)	(-0.9300)
	Panel B: Behavioral Factor Model											
	lnNS	announcementreturn	compequiss	accruals	noa	assetgrowth	ioa	failureprobability	OScore	mom12m	gp	roaq
mkt	-0.0833	-0.1016	0.3027***	-0.0046	0.2542**	0.0638	0.1075	-0.1247	-0.1192	-0.1098	0.1013	-0.2892
	(-1.0038)	(-1.5693)	(3.0637)	(-0.0692)	(2.5634)	(0.6367)	(1.1683)	(-0.6661)	(-0.9797)	(-0.8766)	(0.5932)	(-1.6929)
pead	-0.2514	1.2961***	1.4280***	0.2775	-0.0130	-0.6628***	-0.0495	-0.7394	-0.0333	1.4876***	-0.5429**	0.2931
	(-1.7149)	(8.2312)	(4.6118)	(0.9063)	(-0.0482)	(-3.5977)	(-0.1897)	(-1.7106)	(-0.1472)	(3.5993)	(-2.3421)	(0.4673)
fin	0.4987***	-0.1704***	-0.1222	-0.2494***	0.4589***	0.3029**	0.2777**	0.1794	0.2861**	-0.1653	0.1509	0.3412**
	(8.5186)	(-4.2873)	(-0.7071)	(-3.1846)	(3.2066)	(2.1454)	(2.6019)	(1.0609)	(2.4335)	(-0.7265)	(1.1479)	(2.3953)
_cons	0.0465***	0.0054	-0.0221	0.0350**	0.0005	0.0482***	-0.0157	0.0551	-0.0095	0.0253	0.0293	-0.0114
	(4.4573)	(0.4174)	(-0.9144)	(2.1576)	(0.0324)	(2.8858)	(-0.7940)	(1.5524)	(-0.4038)	(0.8893)	(1.0143)	(-0.3852)

t statistics in parentheses

** $p<0.05$ *** $p<0.01$"

Table 3: Aggregate VAR
This table presents the time-series aggregate VAR results. The variables in the first row are dependent variables while the variables in the fisrt column are independent variables. Each column represents a regression. The variables are all value-weighted averages of the corresponding firm-level variables. The sample is from 1964 to 2019. Heteroskedasticity-adjusted (White) standard errors appear in parentheses.

	lnRealRet_Jun	lnROE_V02	lnBM	lnProf	lnInv_M5	lnME_D5	lnMom6
lag1_lnRealRet_Jun	-0.0137	-0.00586	0.0998	0.00988	-0.00882	0.0510	-0.0723
	(-0.09)	(-0.24)	(0.77)	(0.57)	(-0.81)	(0.21)	(-0.52)
lag1_lnROE_V02	-1.575	0.242	1.226	-0.285^{***}	0.220^{***}	0.136	-0.923
	(-1.39)	(1.34)	(1.34)	(-3.44)	(3.29)	(0.09)	(-0.99)
lag1_lnBM	0.147	-0.00175	0.854^{***}	0.00415	-0.0105	0.0407	0.142^{**}
	(1.24)	(-0.17)	(8.72)	(0.45)	(-1.92)	(0.25)	(2.29)
lag1_lnProf	1.018	0.417^{***}	-0.582	1.076^{***}	0.151^{***}	1.887	0.934
	(0.87)	(2.92)	(-0.67)	(13.65)	(2.93)	(1.35)	(1.35)
lag1_lnInv_M5	-1.625^{**}	-0.212	1.567^{**}	-0.120	0.739^{***}	-2.731^{**}	-1.484^{**}
	(-2.36)	(-1.56)	(2.50)	(-1.37)	(9.73)	(-2.17)	(-2.20)
lag1_lnME_D5	-0.0156	-0.00874	-0.0597	-0.00501	0.0102^{**}	0.753^{***}	0.0694
	(-0.18)	(-0.84)	(-0.81)	(-0.54)	(2.18)	(4.97)	(1.30)
lag1_lnMom6	0.133	0.0192	0.0259	-0.0376	-0.000965	-0.109	-0.0629
	(0.56)	(0.52)	(0.13)	(-1.55)	(-0.06)	(-0.29)	(-0.31)
_cons	0.220	0.0145	-0.221	0.0323	-0.0455^{***}	-0.0399	0.122
	(1.03)	(0.57)	(-1.29)	(1.69)	(-4.05)	(-0.12)	(0.86)
N	56	56	56	56	56	56	56
adj. R^2	0.060	0.396	0.797	0.778	0.903	0.444	0.054

t statistics in parentheses

** $p < 0.05$, *** $p < 0.01$

Table 4: Panel VAR

This table presents the panel VAR results. The variables in the first row are dependent variables while the variables in the first column are independent variables. Each column represents a regression. The variables are all market adjusted by deducting the cross-sectional value-weighted averages. The sample is from 1964 to 2019. Heteroskedasticity-adjusted (White) standard errors appear in parentheses.

	lnRealRet_Jun	lnROE_V02	lnBM	lnProf	lnInv_M5	lnME_D5	lnMom6
lag1_lnRealRet_Jun	0.0563	0.107***	0.102***	0.0449***	0.00982***	0.194***	0.0190
	(1.47)	(7.33)	(2.78)	(6.27)	(4.34)	(4.11)	(0.71)
lag1_lnROE_V02	0.0638***	0.256***	0.109***	-0.0262	0.0333***	-0.148**	0.0198
	(2.79)	(10.41)	(4.09)	(-1.48)	(7.05)	(-2.59)	(1.87)
lag1_lnBM	0.0187**	-0.0218***	0.932***	-0.00754***	-0.00319***	-0.00848	0.00930
	(2.32)	(-9.64)	(93.04)	(-3.98)	(-6.49)	(-0.75)	(1.59)
lag1_lnProf	0.0910***	0.219***	0.0228	0.617***	-0.00271	0.147***	0.0559**
	(2.95)	(12.71)	(0.93)	(23.76)	(-0.85)	(3.25)	(2.62)
lag1_lnInv_M5	-0.160***	-0.117***	0.129***	-0.0803***	0.712***	0.00663	-0.0497**
	(-6.18)	(-9.81)	(4.17)	(-9.32)	(77.98)	(0.14)	(-2.26)
lag1_lnME_D5	-0.0201**	0.00528***	0.0445***	-0.000281	0.0203***	0.719***	-0.0157**
	(-2.35)	(3.97)	(5.60)	(-0.20)	(20.02)	(51.81)	(-2.26)
lag1_lnMom6	0.101**	0.0248	-0.0318	0.0283***	-0.0000739	0.0437	0.0288
	(2.64)	(1.67)	(-0.91)	(3.36)	(-0.03)	(0.92)	(1.28)
_cons	0.000896	-0.0126***	0.0350***	-0.0143***	-0.00291***	-0.0222	0.00732
	(0.12)	(-6.59)	(4.03)	(-10.14)	(-4.14)	(-1.83)	(1.37)
N	80233	80570	80571	80571	80571	80571	80571
adj. R^2	0.026	0.211	0.779	0.396	0.774	0.561	0.009

t statistics in parentheses

** $p < 0.05$, *** $p < 0.01$

Table 5: Variance Decomposition for Firm-Level and Market Return

The table displays the variance decomposition of firm- and market-level real returns. var(CF) stands for the portion of variance of cash flow news on variance of unexpected return news, var(DR) stands for the portion of variance of discount rate news on variance of unexpected return news, -2cov(CF,DR) stands for the portion of covariance between cash flow news and discount rate news on variance of unexpected return news, corr(CF,DR) stands for the correlation between cash flow news and discount rate news. mkt return refers to the decomposition from aggregate VAR. firm mkt-adj return refers to the decomposition from panel VAR. firm return refers to the decomposition by combining components of market returns and firm market-adjusted returns. The sample is from 1972 to 2019.

	var(CF)	var(DR)	-2cov(CF,DR)	corr(CF,DR)
mkt return	17.88%	85.49%	-3.37%	4.32%
firm mkt-adj return	95.74%	8.48%	-4.21%	7.40%
firm return	85.74%	16.93%	-2.67%	7.40%

Table 6: Variance Decomposition for Anomaly-Level Return
The table displays the variance decomposition of anomaly-level (anomalies indicated in the first column) returns. var(CF) stands for the portion of variance of cash flow news on variance of unexpected return news, var(DR) stands for the portion of variance of discount rate news on variance of unexpected return news, -2cov(CF,DR) stands for the portion of covariance between cash flow news and discount rate news on variance of unexpected return news, corr(CF,DR) stands for the correlation between cash flow news and discount rate news. The sample is from 1972 to 2019.

	var(CF)	var(DR)	-2cov(CF,DR)	corr(CF,DR)
lnNS	89.86%	3.49%	6.65%	-18.76%
announcementreturn	92.87%	9.92%	-2.79%	4.60%
compequiss	84.51%	3.34%	12.15%	-36.17%
accruals	90.34%	3.62%	6.04%	-16.71%
noa	87.26%	6.32%	6.42%	-13.66%
assetgrowth	91.60%	3.30%	5.10%	-14.68%
ioa	92.53%	4.77%	2.70%	-6.42%
failureprobability	93.65%	6.11%	0.23%	-0.49%
OScore	100.49%	3.50%	-3.99%	10.64%
mom12m	118.34%	12.74%	-31.08%	40.02%
gp	85.06%	3.78%	11.17%	-31.15%
roaq	104.59%	6.23%	-10.82%	21.20%

Table 7: Variance Decomposition for Factor-Level and Mean Variance Efficient Portfolio Return

The table displays the variance decomposition of individual factor-level returns (Panel A), MVE portfolios with market factor (Panel B), and MVE portfolios without market factor (Panel C). var(CF) stands for the portion of variance of cash flow news on variance of unexpected return news, var(DR) stands for the portion of variance of discount rate news on variance of unexpected return news, -2cov(CF,DR) stands for the portion of covariance between cash flow news and discount rate news on variance of unexpected return news, corr(CF,DR) stands for the correlation between cash flow news and discount rate news. "All Factors" means that we include all individual factors in Panel A and with or without market factor. "Behavioral Factors" include "FIN" and "PEAD". "Mispricing Factors" include "MISP" and "SIZE". The sample is from 1972 to 2019.

	var(CF)	var(DR)	-2cov(CF,DR)	corr(CF,DR)
Panel A: Individual Factors				
FIN	89.63%	2.52%	7.85%	-26.14%
PEAD	77.01%	7.35%	15.64%	-32.87%
MISP	92.71%	2.92%	4.37%	-13.27%
SIZE	82.97%	5.60%	11.43%	-26.53%
Panel B: MVE portfolios with market factor				
All Factors	93.91%	10.99%	-4.90%	7.63%
Behavioral Factors	68.25%	22.97%	8.79%	-11.10%
Mispricing Factors	94.95%	15.01%	-9.96%	13.19%
Panel C: MVE portfolios without market factor				
All Factors	98.48%	4.71%	-3.19%	7.40%
Behavioral Factors	85.67%	8.06%	6.27%	-11.93%
Mispricing Factors	94.20%	2.70%	3.10%	-9.74%

Table 8: Explanatory Power of Factor News

This table shows the adjusted squares of regressing twelve anomalies unexpected return news separately on different components of market news (Panel A), on different components of behavioral factor model news (Panel B), on different components of mispricing factor model news (Panel C), on different components of all factors' news (Panel D). Row name "CFnews" means that we regress total news of anomaly on the cash flow news components of factors indicated by panels, say in panel B, it implies that we regress unexpected anomaly news on cash flow news of market factor cash flow news, fin factor cash flow news, and pead factor cash flow news.

	lnNS	announcementreturn	compequiss	accruals	noa	assetgrowth	ioa	failureprobability	OScore	mom12m	gp	roaq
Panel A: Anomaly News Explained by Components of Market News												
CFnews	2.88%	-0.79%	4.44%	-1.66%	-1.88%	0.29%	-1.84%	-2.09%	-2.16%	8.29%	-0.15%	-1.21%
DRnews	0.85%	-1.57%	2.13%	-1.86%	0.02%	-2.17%	-2.03%	3.95%	2.61%	2.21%	-0.80%	7.87%
Totalnews	4.37%	-2.13%	6.86%	-2.13%	0.39%	-1.68%	-1.82%	3.66%	1.75%	-1.85%	-1.94%	9.02%
Panel B: Anomaly News Explained by Components of Behavioral Factor News												
CFnews	62.05%	50.55%	31.23%	13.91%	25.53%	40.30%	14.98%	3.02%	19.33%	26.47%	8.20%	13.48%
DRnews	14.66%	11.54%	22.60%	1.53%	13.93%	13.40%	5.00%	2.86%	0.52%	3.88%	-0.14%	5.97%
Totalnews	64.94%	54.34%	38.09%	12.01%	29.63%	43.11%	14.15%	6.57%	19.04%	24.00%	2.69%	20.14%
Panel C: Anomaly News Explained by Components of Mispricing Factor News												
CFnews	37.41%	-1.24%	26.02%	17.09%	9.52%	28.42%	35.16%	31.11%	60.70%	19.29%	33.32%	37.62%
DRnews	3.64%	12.23%	5.40%	-2.10%	1.55%	-4.72%	-4.73%	3.39%	4.67%	1.29%	2.04%	6.87%
Totalnews	32.62%	0.83%	26.32%	16.44%	15.44%	22.52%	32.36%	32.72%	61.69%	11.70%	35.11%	38.63%
Panel D: Anomaly News Explained by Components of All Factor News												
CFnews	66.53%	50.90%	44.05%	19.27%	25.49%	48.38%	37.74%	36.29%	64.90%	53.92%	38.30%	34.88%
DRnews	15.87%	19.04%	24.36%	-0.88%	11.39%	9.34%	0.86%	4.59%	4.30%	0.85%	4.93%	5.47%
Totalnews	69.50%	54.26%	48.96%	19.45%	30.68%	48.37%	37.11%	37.42%	66.94%	54.81%	42.85%	36.37%

Table 9: Beta Analysis for PEAD Factor

This table reports firm-level news components of bad beta (i,CFm) and good beta (i,DRm) measured for pead-sorted portfolios. From portfolio 1 to portfolio 3, the pead measure increases. These components are $\beta_{DRi,CFm}$, $\beta_{CFi,CFm}$, $\beta_{DRi,DRm}$, $\beta_{CFi,DRm}$ obtained by regressing corresponding news component of PEAD factor on corresponding news component of market return.

β	1	2	3	3-1
i,CFm	0.1124	0.1091	0.1458	0.0335
se	0.0483	0.0477	0.0512	0.0191
DRi,CFm	0.0181	0.0170	0.0189	0.0008
se	0.0462	0.0457	0.0460	0.0054
CFi,CFm	0.0943	0.0921	0.1269	0.0326
se	0.0154	0.0156	0.0189	0.0167
i,DRm	0.6480	0.6587	0.6852	0.0372
se	0.0577	0.0518	0.0675	0.0429
DRi,DRm	0.6759	0.6720	0.6740	-0.0019
se	0.0174	0.0140	0.0162	0.0117
CFi,DRm	-0.0278	-0.0134	0.0112	0.0391
se	0.0452	0.0452	0.0582	0.0375

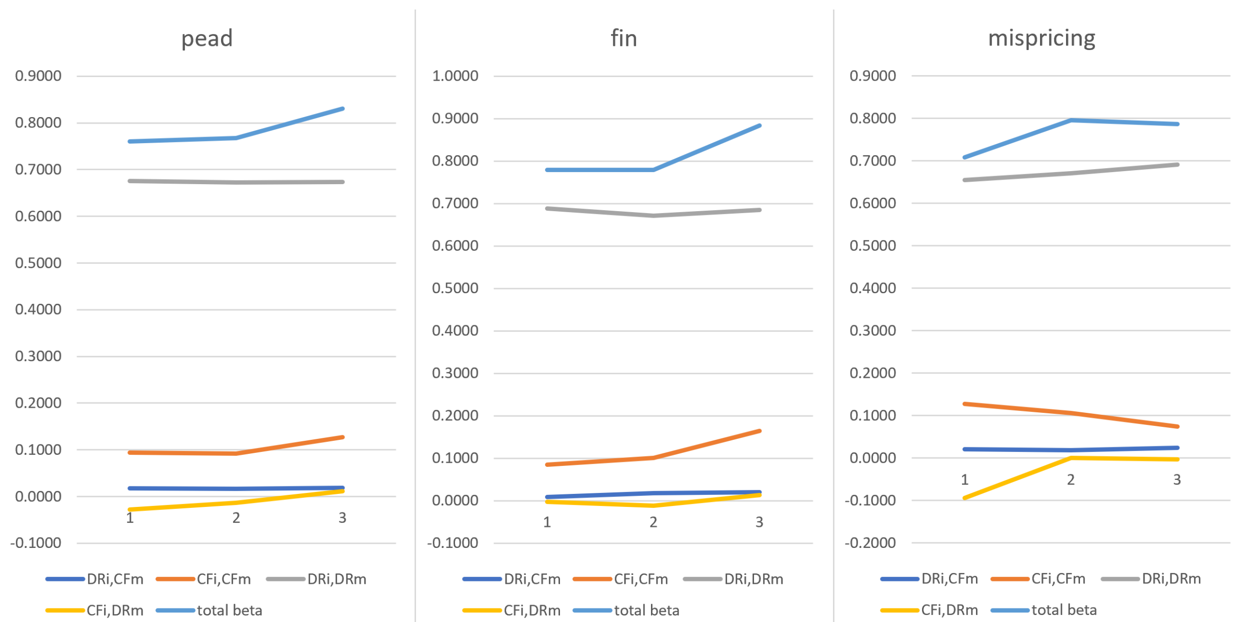

Figure 1. Cross-sectional variation in the components of beta for different factors
This figure plots the total beta and its four components for pead factor (the left one), fin factor (the middle one) and for mispricing factor (the right one).

Table 10: Beta Analysis for FIN Factor

This table reports firm-level news components of bad beta (i,CFm) and good beta (i,DRm) measured for fin-sorted portfolios. From portfolio 1 to portfolio 3, the fin measure increases. These components are $\beta_{DRi,CFm}$, $\beta_{CFi,CFm}$, $\beta_{DRi,DRm}$, $\beta_{CFi,DRm}$ obtained by regressing corresponding news component of FIN factor on corresponding news component of market return.

β	1	2	3	3-1
i,CFm	0.0938	0.1186	0.1851	0.0913
se	0.0593	0.0473	0.0611	0.0620
DRi,CFm	0.0085	0.0177	0.0204	0.0119
se	0.0475	0.0457	0.0469	0.0099
CFi,CFm	0.0853	0.1009	0.1647	0.0794
se	0.0362	0.0132	0.0345	0.0589
i,DRm	0.6858	0.6604	0.6990	0.0131
se	0.0865	0.0515	0.1038	0.1386
DRi,DRm	0.6880	0.6714	0.6851	-0.0029
se	0.0222	0.0144	0.0184	0.0220
CFi,DRm	-0.0022	-0.0109	0.0139	0.0161
se	0.0837	0.0434	0.0922	0.1312

Table 11: Beta Analysis for MISPRICING Factor

This table reports firm-level news components of bad beta (i,CFm) and good beta (i,DRm) measured for mispricing-sorted portfolios. From portfolio 1 to portfolio 3, the mispricing measure increases. These components are $\beta_{DRi,CFm}$, $\beta_{CFi,CFm}$, $\beta_{DRi,DRm}$, $\beta_{CFi,DRm}$ obtained by regressing corresponding news component of MISPRICING factor on corresponding news component of market return.

β	1	2	3	3-1
i,CFm	0.1479	0.1235	0.0983	-0.0495
se	0.0423	0.0488	0.0569	0.0399
DRi,CFm	0.0204	0.0180	0.0238	0.0033
se	0.0445	0.0457	0.0474	0.0069
CFi,CFm	0.1274	0.1055	0.0746	-0.0528
se	0.0192	0.0148	0.0298	0.0382
i,DRm	0.5608	0.6718	0.6884	0.1276
se	0.0633	0.0563	0.0788	0.0866
DRi,DRm	0.6544	0.6712	0.6913	0.0369
se	0.0149	0.0150	0.0203	0.0141
CFi,DRm	-0.0936	0.0006	-0.0030	0.0907
se	0.0571	0.0469	0.0695	0.0843

References

Asness, C., Frazzini, A., 2013. The devil in hml's details. The Journal of Portfolio Management 39, 49–68.

Barillas, F., Shanken, J., 2018. Comparing asset pricing models. The Journal of Finance 73, 715–754.

Bryzgalova, S., Huang, J., Julliard, C., 2019. Bayesian solutions for the factor zoo: We just ran two quadrillion models. Available at SSRN 3481736 .

Callen, J. L., Segal, D., 2004. Do accruals drive firm-level stock returns? a variance decomposition analysis. Journal of accounting Research 42, 527–560.

Callen, J. L., Segal, D., 2010. A variance decomposition primer for accounting research. Journal of Accounting, Auditing & Finance 25, 121–142.

Campbell, J. Y., 1991. A variance decomposition for stock returns. The Economic Journal 101, 157–179.

Campbell, J. Y., Mei, J., 1993. Where do betas come from? asset price dynamics and the sources of systematic risk. The Review of Financial Studies 6, 567–592.

Campbell, J. Y., Polk, C., Vuolteenaho, T., 2010. Growth or glamour? fundamentals and systematic risk in stock returns. The Review of Financial Studies 23, 305–344.

Campbell, J. Y., Shiller, R. J., 1988. The dividend-price ratio and expectations of future dividends and discount factors. The Review of Financial Studies 1, 195–228.

Campbell, J. Y., Vuolteenaho, T., 2004. Bad beta, good beta. American Economic Review 94, 1249–1275.

Chen, A. Y., Zimmermann, T., 2021. Open source cross-sectional asset pricing. Critical Finance Review, Forthcoming .

Chen, H. J., 2017. Do Cash Flows of Growth Stocks Really Grow Faster? Journal of Finance 72, 2279–2330.

Chen, L., Zhang, L., 2010. A better three-factor model that explains more anomalies. Journal of finance 65, 563–595.

Chen, L., Zhao, X., 2009. Return decomposition. The Review of Financial Studies 22, 5213–5249.

Chen, Z., 2022. Inferring Stock Duration Around FOMC Surprises: Estimates and Implications. Journal of Financial and Quantitative Analysis 57, 669–703.

Cho, T., Kremens, L., Lee, D., Polk, C., 2022. Scale or yield? a present-value identity .

Cochrane, J. H., 2008. The dog that did not bark: A defense of return predictability. The Review of Financial Studies 21, 1533–1575.

Cochrane, J. H., 2011. Presidential address: Discount rates. The Journal of finance 66, 1047–1108.

Daniel, K., Hirshleifer, D., Sun, L., 2020. Short-and long-horizon behavioral factors. The review of financial studies 33, 1673–1736.

Davis, J. L., Fama, E. F., French, K. R., 2000. Characteristics, covariances, and average returns: 1929 to 1997. The Journal of Finance 55, 389–406.

De La O, R., Myers, S., 2021. Subjective cash flow and discount rate expectations. The Journal of Finance 76, 1339–1387.

Engsted, T., Pedersen, T. Q., Tanggaard, C., 2012. Pitfalls in var based return decompositions: A clarification. Journal of Banking & Finance 36, 1255–1265.

Fama, E. F., French, K. R., 1992. The cross-section of expected stock returns. the Journal of Finance 47, 427–465.

Fama, E. F., French, K. R., 2008. Dissecting anomalies. The Journal of Finance 63, 1653–1678.

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of financial economics 116, 1–22.

Farre-Mensa, J., Michaely, R., Schmalz, M., 2014. Payout policy. Annu. Rev. Financ. Econ. 6, 75–134.

Feng, G., Giglio, S., Xiu, D., 2020. Taming the factor zoo: A test of new factors. The Journal of Finance 75, 1327–1370.

Fink, J., 2021. A review of the post-earnings-announcement drift. Journal of Behavioral and Experimental Finance 29, 100446.

Gao, C., Martin, I. W., 2021. Volatility, valuation ratios, and bubbles: An empirical measure of market sentiment. The Journal of Finance 76, 3211–3254.

Giglio, S., Kelly, B., Xiu, D., 2022. Factor models, machine learning, and asset pricing. Annual Review of Financial Economics 14, 337–368.

Gonçalves, A., 2021. What moves equity markets? a term structure decomposition for stock returns. Kenan Institute of Private Enterprise Research Paper No.20-08 .

Gormsen, N. J., Lazarus, E., 2021. Duration-driven returns. Available at SSRN 3359027 .

Harvey, C. R., 2017. Presidential address: The scientific outlook in financial economics. The Journal of Finance 72, 1399–1440.

Harvey, C. R., Liu, Y., 2019. A census of the factor zoo. Available at SSRN 3341728 .

Harvey, C. R., Liu, Y., Zhu, H., 2016. … and the cross-section of expected returns. The Review of Financial Studies 29, 5–68.

Hou, K., Xue, C., Zhang, L., 2015. Digesting anomalies: An investment approach. The Review of Financial Studies 28, 650–705.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. The Journal of finance 48, 65–91.

Kozak, S., Nagel, S., Santosh, S., 2020. Shrinking the cross-section. Journal of Financial Economics 135, 271–292.

Lettau, M., Pelger, M., 2020. Estimating latent asset-pricing factors. Journal of Econometrics 218, 1–31.

Lochstoer, L. A., Tetlock, P. C., 2020. What drives anomaly returns? The Journal of Finance 75, 1417–1455.

Maio, P., Xu, D., 2020. Cash-flow or return predictability at long horizons? the case of earnings yield. Journal of Empirical Finance 59, 172–192.

Mao, M. Q., Wei, K. J., 2014. Price and earnings momentum: An explanation using return decomposition. Journal of Empirical Finance 28, 332–351.

Michaely, R., Rossi, S., Weber, M., 2021. Signaling safety. Journal of Financial Economics 139, 405–427.

Stambaugh, R. F., Yuan, Y., 2017. Mispricing factors. The review of financial studies 30, 1270–1315.

Vuolteenaho, T., 2002. What drives firm-level stock returns? The Journal of Finance 57, 233–264.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. The Review of Financial Studies 21, 1455–1508.