Essays on Subjective Expectations in Finance

Eugene Larsen-Hallock

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2023

Abstract

Essays on Subjective Expectations in Finance

Eugene Larsen-Hallock

In chapter one, I examine the predictive content of subjective return expectations derived from price targets issued by equity analysts. Equity price targets are an ubiquitous feature of the financial information landscape, but it is not clear how informative they actually are. In this chapter, I show that the cross-section of price-target implied subjective return expectations contains rich informational content for forecasting returns. In-sample, I find that expected returns correlate strongly with average cross-sectional returns to a large panel of portfolios formed on the basis of observable firm characteristics. In out-of-sample exercises, forecasting models using subjective expectations are shown to offer more accurate predictions for portfolio returns than several other commonly employed, cross-sectional predictors, including the book-to-market and dividend-price ratios, momentum, and forward-looking cash-flow measures. Furthermore, these differences are shown to be economically relevant, with conditional portfolios formed on the basis of subjective expectations offering substantially improved risk-adjusted returns compared to many of the other predictors considered. The relative informational content, as well as the production by analysts, of subjective return expectations is found, however, to peak during recessions, with negligible predictive advantage discernible in expansions.

In chapter two, my coauthors (Adam Rej, with CFM; David Thesmar, with MIT, CEPR, and

NBER) and I empirically analyze a large panel of firm sales growth expectations. We find that the relationship between forecast errors and lagged revision is non-linear. Forecasters underreact to typical (positive or negative) news about future sales, but overreact to very significant news. To account for this non-linearity, we propose a simple framework, where (1) sales growth dynamics have a fat-tailed high frequency component and (2) forecasters use a simple linear rule. This framework qualitatively fits several additional features of data on sales growth dynamics, forecast errors, and stock returns.

In chapter three, my coauthor (Ken Teoh, with Columbia) and I construct a novel text-based measure of firm-level attention to macroeconomic conditions and document that stocks associated with higher macroeconomic attention earn lower returns. Moving from the bottom decile to top decile of macroeconomic attention decreases a stock's average return by 11.6% per year. We propose a risk-based explanation in which stocks with higher macroeconomic attention contribute less idiosyncratic cash flow risk to the investor's portfolio, hence earn lower expected returns. Decomposing the unexpected returns of macroe- conomic attention-sorted portfolios into cash flow and discount rate news, we find that portfolios with higher macroeconomic attention stocks have lower cash flow risk.

Table of Contents

List of Figures

List of Tables

Dedication

This work is dedicated to my wife, Jieun, who tirelessly tended to the joy and flourishing of our small family during all the years I spent pursuing my curiosity.

Chapter 1: Forecasting Returns from Subjective Expectations[1]

[1] I thank Serena Ng, Harrison Hong, and David Thesmar for their invaluable guidance on various iterations of this project. I also thank the participants of the Columbia Economics Department Econometrics and Financial Economics colloquia for the useful feedback they provided on early versions of this chapter.

1.1 Overview

Forecasts produced by equity analysts can provide rich insight into the expectations of the market, but despite the academic attention given to the informational content and properties of subjective cash flow expectations, relatively little interest has been shown in expectations of returns. The wide availability of target price and dividend forecast information through the information sources commonly available to traders of all sorts suggest that this information is likely relied on to some degree by many, if not most, market participants. In recent decades, target price forecasts have become available for nearly all publicly-traded U.S. firms (as well as nearly all firms of note in other countries), and are often featured prominently in market data sources used by both finance professionals and retail traders. Equity analysts, the producers of these forecasts, are generally well educated and well paid (Groysberg, Healy, and Maber, 2011), and the production of subjective forecasts at scale represents a substantial investment on the part of brokers and equity research firms.

Despite these efforts, however, price targets (and the expectations for future returns they imply) have been largely neglected in academic asset pricing. This is surprising, both because of the prodigious volume of work on cash flow expectations and stock recommendations that has appeared since the 1980s, and also because all evidence suggests that market participants believe price targets to be an informative product. The challenge, however, has been identifying just what the information conveyed in market prices might be. Writing around the turn of the millennium, Brav and Lehavy (2003) note that "Despite the increasing prominence of target prices, their role in conveying information to market participants and their contribution to the formation of equity prices have remained largely unexplored." Despite the passage of more than two decades, the situation is largely unchanged.

The straightforward goal of this chapter is to examine the informational content of subjective return expectations for the forecasting of future equity returns. Constructing a measure of expected stock-level returns from equity analyst price targets, I show in simulated out-of-sample exercises that subjective expectations forecast the returns of a wide panel of characteristic-sorted portfolios with greater accuracy than several other commonly employed cross-sectional predictors. I provide suggestive evidence that, while subjective returns are poor predictors of future price movements, they are nonetheless informative about relative returns. I find, furthermore, that this increased predictability allows for the construction of conditional portfolios with greater risk-adjusted returns than can be constructed using the other predictors considered. I also find, however, that there is significant time-variation in the ability of models estimated on subjective return expectations to predict returns, with the greatest relative informational advantages being observed during crisis period.

While the forecasting objective that motivates this chapter draws on the long literature examining return predictability, I draw most heavily on the insights of papers such as Kelly and Pruitt (2013) that address the problem of forecasting returns from the cross-section of a predictive signal, such as the book-to-market ration, as in that paper. The factor modeling approach I pursue also employs a partial least-squares method similar to that in Kelly and Pruitt (2013) and Kelly and Pruitt (2015), but I combine dimensional reduction methods (PCA, PLS) with other regularization techniques commonly associated with the machine learning literature to further increase the predictive performance of my model. Naturally, the use of reduced-rank factor models to forecast returns also speaks to the broader literature on forecasting from "diffusion indices" that originated in Stock and Watson (2002a), and has been further developed in papers such as Ludvigson and Ng (2007). The context of this chapter differs from those, however, in that the basic forecasting problem considered is one in which the cross-section of a single predictive variable is used as the set of predictors.

I believe that this chapter also represents a significant step in understanding the informational content of analyst price targets. Most of the early work in this area concerned the obvious biases that price targets evinced, and the market responses to the information price targets conveyed. While Brav and Lehavy (2003) confirmed that equity prices are responsive to price target revisions, they also observed that price targets are "unrealistically optimistic," with forecast one-year price growth for U.S. equities averaging 28% in the late 90s. Price target revisions are often released simultaneously with a great deal of other information, but Asquith, Mikhail, and Au (2005) find that price targets do contain independent information not contained in other sources, such as analyst recommendations and earnings forecasts. Bradshaw, Brown, and Huang (2013) examine the evidence for differential forecasting ability across analysts, finding that, while there is evidence that some analysts are superior than others at forecasting earnings, the same does not appear to be true for price targets. They additionally note that analyst compensation and career outcomes are positively associated with their accuracy in forecasting earnings, but uncorrelated with accuracy in forecasting prices.

Recently, a small number of papers is beginning to take the asset pricing implications of subjective expectations of returns more seriously. Greenwood and Shleifer (2014) considered a number of forward looking proxies for expectations of market returns. They find that subjective expectations of returns are strongly negatively correlated with past returns and with statistically estimated measures of expected returns. They suggests this is evidence many investors form expectations via an extrapolative process, a la Barberis et al. (2015), in which investors place excessive weight on the recent history of returns. O and Myers (2021) examines the covariance between subjective returns, as measured in the Graham-Harvey survey, and price ratios. Employing a variation on the famous variance decomposition of Campbell (1991), but employing forward-looking subjective return and cash flow expectations, they find that subjective expectations for market returns exhibit low volatility, relative to both future returns dividend expectations, and little correlation with future returns. They conclude from this that most of the variation in prices is due to variation in expectations of future cash flows. Needless to say, none of these results suggest we should be optimistic that expectations of returns will be useful for return forecasting.

While the Graham-Harvey survey, the only direct measure of expected returns considered in those papers, is a rich resource on the expectations and practices of CFOs, it is only quarterly in frequency and the questions asked about return expectations are limited to the expected one- and ten-year return of the S&P 500. The I/B/E/S derived measure studied here is effectively limited to a one-year horizon for returns, but is observable at monthly or greater frequencies, and contains forecasts for nearly all publicly traded equities in the U.S. This same data source is also examined in O, Han, and Myers (2022). In that paper, however, return expectations derived from analyst price targets are assumed to be direct measures of future expected returns, and are then employed in a Campbell (1991)-type decomposition to study cross-sectional variation in valuation ratios.

The forecasting exercise I pursue only requires that the subjective expectations of analysts and future price movements be correlated, and does not require assuming that analyst forecasts are at all representative of market participant beliefs. As suggested by papers such as Hong and Stein (2007), even if the distribution of subjective expectations across market participants can be measured accurately, trading frictions may result in prices reflecting only a biased subset of those expectations. This is only one reason, however, for why one should hesitate before regarding analyst forecasts as an accurate description of the "market" expectations driving price formation.

Kothari, So, and Verdi (2016) provides an overview of the literature on sell-side analyst forecasts in asset pricing and accounting. In particular, they provide a review of the use of analyst earnings forecasts in constructing measures of expected returns. provides a summary of the first few decades of research into analyst forecasts. While a large number of papers have examined the accuracy of and market reactions to analyst forecasts, particularly EPS forecasts, very little work has been done to consider whether whether analyst forecasts have useful content for predicting future returns.[2] While this literature is too broad to cover here, several consistent themes appear with regularity. One possibility is that analysts may face economic incentives to bias their forecasts. They might, for example, feel pressured by their employers to generate trading commissions, or to provide positive coverage for investment banking deals the employers underwrite. Alternatively, analysts may engage in quid-pro-quo arrangements with firm management in exchange for increased access or private information, although the sign of the bias this would create is unclear and may be potentially time-varying, e.g. management may favor high forecasts initially, but more tempered forecasts prior to results announcements.

[2] Kothari, So, and Verdi (2016) and Bradshaw (2011) offer concise surveys of this literature.

In addition to financial incentives, there is also reason to suppose that analysts are subject to cognitive limitations that may result in predictable biases. Concerns of these sorts data back to at least Abarbanell (1991) and Abarbanell and Bernard (1992), which find that analysts underreact to both recent price changes and recent earnings releases. More recently, Bordalo et al. (2019) confirms the classic finding of La Porta (1996) that equities with higher long-term earnings growth forecasts tend to have lower realized future returns, and postulates that this is the result of overreaction in long-term expectations to recent earnings surprises. It is likely that further analysis of the data I consider here may be useful in further illuminating the important question of how observed subjective expectations relate to future returns, but that will ultimately fall beyond the scope of the current chapter.

1.2 Data

1.2.1 Subjective expectations

To measure firm-level subjective return expectations, I rely on data provided by I/B/E/S on price targets and dividend forecasts issued by equity analysts. I/B/E/S is considered the gold-standard data source for firm-level data on price targets, as well as other cash flow and accounting variables. The forecast data distributed by I/B/E/S is widely used by finance practitioners, and has been employed in numerous academic studies. As noted above, however, nearly all of the academic work on subjective expectations has exclusively used cash flow forecasts, while studies using price targets have been rare.

I construct two measures for expected returns: one using only analyst price targets, and one combining price targets with forecasts for future dividend payments. The reason for considering two different measures is that it is ambiguous whether or not analysts take projected dividends into account when they are issuing their price targets. Informal discussions with equity analysts suggest that the treatment of dividends may, in fact, vary across analysts (for example, with the type of price target model employed) or across firms. Which should be preferred is thus an empirical question, although it will turn out that both measures are extremely highly correlated and yield almost identical results.

In constructing these measures, I employ the consensus mean estimates for price targets and dividends provided in the I/B/E/S "summary" files. These estimates are provided by I/B/E/S at a monthly frequency, and provide broad coverage for both price targets and dividends starting from 2003. To ensure that adjustments for outstanding shares are consistent when calculating expected returns, and the consistency of forecast and price timing, I use prices provided by I/B/E/S. For the many firms that do not regularly pay a dividend, dividend forecasts are generally not provided in I/B/E/S. In those cases where a price target forecast is available for a firm/month, but no dividend forecast is provided, I therefore assume forecast dividends are equal to zero.

One obstacle in the construction of expected returns is the staggered forecasting horizons for dividends across firms. Horizons for forecasts of dividends and earnings are timed around firm fiscal periods. Thus, the I/B/E/S one-year ahead dividend forecast is the forecast for dividends paid out in the current fiscal year—the end of which is typically less than one calendar year into the future from the month in which the dividend was issued. In order to construct constant horizon return expectations, I interpolate across forecast horizons in the data to obtain expected dividends for constant 12-month horizons.

With few exceptions, price targets are issued for constant one-year horizons, so no interpolation is necessary for this component of expected returns. To ensure that equity prices are measured simultaneously with future price and dividend expectations, and that consistent adjustments for shares outstanding are applied, I use the prices provided by I/B/E/S in the summary data files to calculate all forward returns, price growth, and valuation ratios. Similarly, the forecast 1-year forward earnings yield is interpolated in the same way to construct a "constant" 1-year horizon expectation for earnings that are expected to be generated over the following twelve months. While this quantity is not amongst the accounting variables commonly reported by firms, and therefore of limited use in the forecasting of realized firm cash flows, which are only reported quarterly, it is nonetheless a well-defined quantity from a valuation perspective and potentially useful in that context.

1.2.2 Returns

I use Center for Research in Securities prices (CRSP) monthly individual stock returns for common stock equities (share code 11, 12, or 13) of all firms listed on the three major U.S. exchanges: NYSE, AMEX, and NASDAQ. In order to be retained, equities must have a valid prices. When a delisting code is present, but no delisting returns are specified, an expected delisting return is imputed following Shumway (1997) and Shumway and Warther (1999). Dividends per share are calculated "top down' using the difference between simple and ex-dividend monthly returns, multiplied by adjusted shares outstanding.

In constructing portfolio signals and returns, I make two key restrictions. The first is to exclude financial firms, which I classify as those with a Standard Industry Classification code between 6000-6999 (variable `siccd` in the WRDS CRSP-Compustat combined dataset). The primary motivation for this restriction is to obtain a more consistent sample across the various data sources I employ. As the WRDS Financial Ratios dataset excludes financial firms, enforcing the same restriction throughout my data ensures that differences in robust checks performed using my alternative dataset for forming characteristic portfolios, from Green, Hand, and Zhang (2017), are not due to the inclusion of financial firms in one dataset and not in another. Additionally, financial firms differ from non-financials in significant ways that make comparison across financial and non-financial firms difficult for many of the characteristics considered—the classic example being the high leverage ratios that are normal for financial firms, but unusual for firms in other industries (Fama and French, 1992).

The second restriction I make is to restrict portfolio composition to firms with market caps greater than the 20th-percentile for NYSE-listed firms in the month of portfolio formation. One reason for doing this is to eliminate micro-cap firms, which are largely traded on AMEX (now NYSE American). The importance of this restriction is pointed in Hou, Xue, and Zhang (2020). Over my sample period, firms with market capitalizations smaller than the NYSE 20th-percentile firm in a given month make up 53.6% of all observations in CRSP, but only 1.89% of the total market capitalization across those observations. In particular, return predictability in micro-cap stocks is often difficult to exploit in practice due to transaction costs. (Novy-Marx and Velikov, 2016) Additionally, the capacity of any trading strategy is, in general, limited by the market capitalization of the smallest equities, restricting to only firms larger than the NYSE 20th-percentile firm further helps to ensure the economic relevance of my findings.

Another reason is that limited coverage of small-caps in I/B/E/S results in higher rates of missingness in subjective return proxies for the smallest, while coverage rates are consistent in larger size deciles. Price targets are available for only 47.1% of firm-months for firms with market capitalizations below the NYSE 10th percentile. This rises to 80.7% for firms between the 10th and 20th deciles, and 88.8% for firms between the 20th and 30th deciles. For firms larger than the 30th decile, the coverage rate is between 90-95% in each decile. The coverage rate for firms larger than the NYSE 90th percentile is 91.9%. This lack of any significant difference in coverage for larger firms suggest that restricting to only those firms larger than the NYSE 20th percentile is sufficient to ensure that any remaining missingness are not due to differences in firm size.

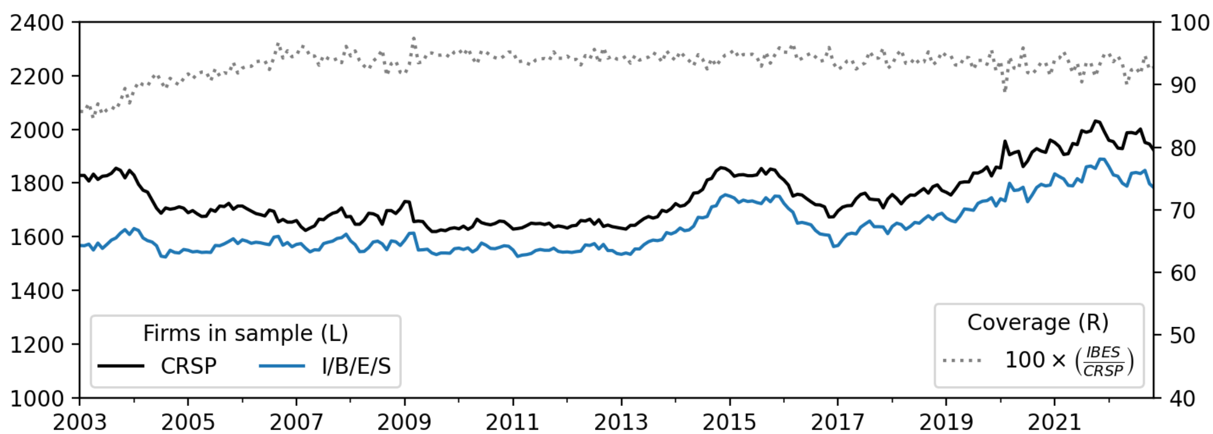


Figure 1.1: I/B/E/S coverage over time

Note: This figure plots the number of firms for which I observe returns (CRSP) and subjective expectations (I/B/E/S) in each month of my sample, after applying the sample restrictions described in the body. The dotted line shows the I/B/E/S coverage rate, relative to the total number of firms observed in CRSP.

Figure 1.2 describes the issuance patterns of new price targets over time. One important fact to note is that, even as the number of firms covered has risen and fallen over time, the number of forecasts issued by individual analysts has tended to rise over my sample. Furthermore, the production of forecasts is highly seasonal, with forecast production rising during periods when the release rate of new information from firms is highest, as when quarterly results are announced. When exceptional events occur, forecast production goes into overdrive, with, for example, nearly 50% more forecasts than usual being estimated in the wake of the Lehman collapse and the first cases of COVID-19 in the U.S.

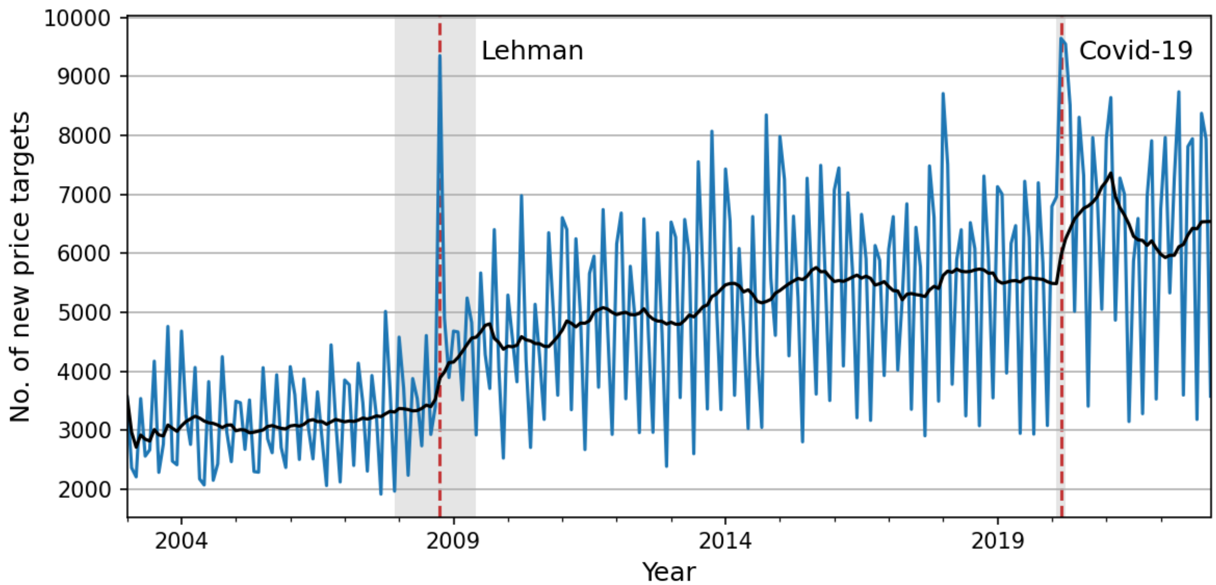


Figure 1.2: Price target issuance, individual forecasts

Note: This figure shows the rate of issuance of new price target forecasts by individual analysts for each month in my sample after application of the sample restrictions described in the text. Raw issuance count is plotted in blue, while a 12-month moving average is shown in black. Vertical lines show the collapse of Lehman Brothers (2008/10) and the reporting of the first confirmed COVID-19 cases in the United States (2020/3). Appendix figure A.1 shows the same figure, but adjusted for seasonality and a linear time trend.

1.2.3 Characteristic-sorted portfolios

In my empirical exercises, I test the predictors and methods proposed here against a broad panel of characteristic-sorted test portfolios. This is done to avoid the critique raised by Lewellen, Nagel, and Shanken (2010) that common sets of test assets, such as the 25 Fama-French size-B/M portfolios, that are known to exhibit a strong low-dimensional factor structure, provide too low of a bar for meaningful testing of return predictors.

Using characteristic-sorted portfolios as test assets is equivalent to assuming that equity-level factor exposures are functions of time-varying firm characteristics. This approach is inspired by long literature, going back to Rosenberg (1974) , that models conditional factor exposures as constant functions of firm characteristics that are allowed to vary over time. Kelly, Pruitt, and Su (2019) and Chen, Roussanov, and Wang (2021), amongst other studies, find that returns to managed characteristic portfolios are more predictable than those of individual equities.

The primary data source employed for the formation of characteristic-managed portfolios is the Financial Ratios Suite from WRDS. This recently introduced data source brings together many of the financial ratios most commonly employed in empirical financial and accounting research, in a format that allows for consistent construction of those variables across studies, and is already being employed in asset pricing research, for example Kozak, Nagel, and Santosh (2020). A further advantage of the WRDS financial ratios dataset is the availability of accurate timestamps of when the information upon which a ratio was based became public knowledge. This information eliminates the need to guess about the delay between financial period end dates and when the information for that period reaches the market ensuring that the measured values of characteristics are as timely as possible.

For all portfolios, I construct zero-cost long-short portfolios via rank transformations of the underlying characteristics, as in Asness, Moskowitz, and Pedersen (2013). Letting z_{ijt} be the value of characteristic j for firm i at time t, the weight placed on each equity in characteristic-sorted porfolios is given by:

$$w_{ijt} = c_{jt}\left(rank\left(z_{ijt}\right) - \sum_i \frac{rank\left(z_{ijt}\right)}{N}\right) \tag{1.1}$$

where c_{jt} is a normalizing constant chosen such that $\sum_i \left|w_{ijt}\right| = 1$, and ranks are taken cross-sectionally in eaach period. Continuous rank transforms have become increasingly popular in the recent empirical asset pricing literature, and have been employed by Kozak, Nagel, and Santosh (2020) and Freyberger, Neuhierl, and Weber (2020), to name just a few examples. This transformation ensures weight scaling is consistent across all characteristics, and reduces the influence of outliers. Except for some initial descriptive exercises, the forecasting and portfolio formation exercises I carry out will focus exclusively on portfolio returns and signals. Returns to the portfolio formed on the basis of characteristic j are formed from the returns to the underlying equities, indexed by i, as:

$$r_{j,t+1} = \sum_i w_{ijt} r_{i,t+1}$$

Return proxy signals for each characteristic portfolio are formed identically to portfolio returns, as weighted averages of the equity level signals:

$$z_{jt} = \sum_i w_{ijt} z_{it}$$

To lessen the influence of occasionally extreme values, all predictors are clipped at the 1$^{\text{st}}$ and 99$^{\text{th}}$ percentiles across firms in each month before aggregation into portfolios.

1.2.4 Predictors considered

My primary predictors of interest are two measures of subjective expectations for excess returns: one that combines both price growth and dividend yield forecasts, and one that omits expected dividend yields and uses only price growth. It is necessary to consider these alternative specifications, as the treatment of dividends by analysts in forming price target forecasts is ambiguous. If forecasters deduct expected dividends in setting their price targets, a full accounting of returns would require that both price growth and dividends be summed to obtain forecasts of returns. On the other hand, if forecasters disregard dividend forecasts when issuing price targets, adding per-share dividends to the target price would resulting in a double counting of dividends. Informal discussions with equity forecasters regarding common practice suggest that conventions may differ across sectors and with the type of valuation method used (e.g discounted FCF or multiples based methods). In practice, both measures are extremely highly correlated with no clear advantage coming from using either one over the other.

To assess the relative performance of my proposed predictor, I also consider several alternative predictors that are commonly employed in the empirical asset pricing literature as predictors of returns, and which have generally been found to be useful in equity return forecasting.

One set of comparison predictors I consider are a variety of financial ratios. The first of these, the price-dividend ratio, has a long history in the return prediction literature, dating at least to Shiller (1981), Fama and French (1988), and Campbell and Shiller (1988). The other trailing financial ratio I consider is the book-to-market ratio. This ratio has been employed to examine the cross-section of expected returns in Vuolteenaho (1999), Vuolteenaho (2002), and Cohen, Polk, and Vuolteenaho (2003). More recently, Haddad, Kozak, and Santosh (2020) has found that portfolios formed from the first few PCs of asset returns are robustly predictable by the book-to-market ratios of those portfolios, but do not consider cross-portfolio predictability, as I do here. I also consider two forward-looking price ratios constructed from consensus I/B/E/S forecasts of future firm cash flows. The dividend-price ratio and earnings-price ratios are constructed from constant one-year horizon consensus forecasts, as described above, and have previously been explored in papers such as O and Myers (2021) .

For characteristic portfolios, the predictor signals for the portfolio are calculated as weighted averages of the stock-level signals, using the same weights assigned to each stock in the formation of the portfolio.[3] This means that, for example, the dividend-price ratio of a characteristic portfolio should not be confused for a weighted sum of stock-level dividends over a weighted sum of stock-level prices. Additionally, given that half of the equities in a given portfolio receive negative weight, it is possible for ratios that cannot take a negative value at the stock level (such as the dividend-price ratio) to be negative at the level of the portfolio.

Finally, I also look at two measures constructed from past returns. The first is a 12-month momentum signal, constructed following Jegadeesh and Titman (1993) as the sum of returns over the previous 12 months, excluding returns in the last month. This predictor has been studied in, for example, Fama and French (1996) and Carhart (1997), and is amongst the most exhaustively researched phenomena in financial economics. Recent work studying momentum strategies has

[3] When observations are missing for a predictor, the signal weights are re-scaled so that their absolute values continue to sum to one.

shown that momentum is also a robust feature of returns to factors and characteristic-managed portfolios, with much of the momentum in equities ascribable to momentum in underlying factors (Ehsani and Linnainmaa, 2019). However, as shown in Daniel and Moskowitz (2016), momentum strategies can be subject to sudden reversals, particularly during rebounds after steep market declines. The last predictor I consider is the short-term reversal (i.e. prior month return) signal initially proposed by Jegadeesh (1990), and found by numerous studies since to be a robust predictor of equity returns. As was observed in Falck, Rej, and Thesmar (2020), I find that, for the characteristic portfolios I use, returns in one month are positively correlated with returns in the next, but I will continue to refer to this signal as the "reversal" signal, as that is how it is most commonly known.

The formulas used to calculate each of these measures are shown in table 1.1. Subjective expectations (consensus analyst forecasts) are denoted by $\widetilde{\mathrm{E}}_t\left[\cdot\right]$. To simplify notation, however, forward looking variables will generally be denoted by $\widetilde{x}_{it}$, with the time index t referring to the time a forecast was observed. As all forecasts are for constant 12-month horizons, forecast horizon will not generally be separately indicated.

Type	Proxy	Var.	Construction
Expected ret.	Returns	$\widetilde{r}_{it}$	$\frac{\widetilde{\mathrm{E}}_t[p_{i,t+12}]+\widetilde{\mathrm{E}}_t[d_{i,t+12}]}{p_{it}} - r_t^f - 1$
	Price change	$\widetilde{\Delta p}_{it}$	$\frac{\widetilde{\mathrm{E}}_t[p_{i,t+12}]}{p_{i,t}} - r_t^f - 1$
Forward ratios	Dividend yield	$\widetilde{dp}_{it}$	$\frac{\widetilde{\mathrm{E}}_t[d_{i,t+12}]}{p_{it}}$
	Earnings yield	$\widetilde{ep}_{it}$	$\frac{\widetilde{\mathrm{E}}_t[e_{i,t+12}]}{p_{it}}$
Trailing ratios	Dividend yield	dp_{it}	$\frac{d_{it}}{p_{it}}$
	Book-to-market	bm_{it}	$\frac{Book_{it}}{Market_{it}}$
Past returns	Momentum	mom_{it}	$\sum_{h=-1}^{11} r_{i,t-h}$
	Short-term reversal	rev_{it}	r_{it}
	Trailing mean	$\overline{r}_{it}$	$\sum_{h=1}^{T-t} r_{i,t-h}$

Note: Subjective expectations (consensus analyst forecasts) are denoted by $\widetilde{\mathrm{E}}_t\left[\cdot\right]$.

Table 1.1: Predictor signal construction

1.3 The informativeness of return expectations

I first consider a couple of simple correlations describing the relationship between return expectations and realized returns in the cross-section and in the time-series. Table 1.2 shows the estimates from two pooled, two cross-sectional, and two-time-series regressions. The pooled specification estimates:

$$r_{i,t+1} = \alpha + \beta \widetilde{r}_{it} + \epsilon_{i,t+1} \tag{1.2}$$

In this table, I only show results for $x_{it} = \widetilde{r}_t$. Results for expected price growth are nearly identical, and are omitted. In the cross-sectional regressions, both returns and expected returns are averaged over the time-dimension before estimating the following specification:

$$\overline{r}_i = \alpha + \beta \overline{\widetilde{r}}_i + \epsilon_i \tag{1.3}$$

where $\overline{r}_i = \frac{1}{T} \sum_t r_{i,t+1}$ and $\overline{\widetilde{r}}_i = \frac{1}{T} \sum_t \widetilde{r}_{it}$.

	(1) Pooled		(2) Cross-section		(3) Time-series	
	Equities	Portfolios	Equities	Portfolios	Equal wgt.	Value wgt.
α	0.001	0.009	-0.010	0.012	0.007	0.008
	(0.183)	(0.000)	(0.000)	(0.000)	(0.472)	(0.388)
β	0.113	0.795	-1.085	1.651	0.443	0.091
	(0.000)	(0.000)	(0.000)	(0.000)	(0.549)	(0.920)
R^2	0.000	0.005	0.040	0.557	0.004	0.000
Obs	381,636	16,969	4,576	71	239	239

Note: Cross-sectional (2) regressions average returns over the time dimension before estimating the specification in body equation 1.3. Time-series (3) regressions average equity returns over the cross-sectional dimension, with either equal or value weighting, before estimating the specification in equation 1.4. Realized returns are monthly frequency. 12-month expected returns are re-scaled to 1-month returns prior to averaging. P-values are in parentheses and have been calculated using Huber-White heteroskedasticity consistent standard errors.

Table 1.2: Return expectations and returns, descriptive regressions

In the time-series regressions, I average the returns and expected returns of individual equities over the cross-section and estimate:

$$\overline{r}_{t+1} = \alpha + \beta \overline{\widetilde{r}}_t + \epsilon_t \tag{1.4}$$

where $\overline{r}_{t+1} = \frac{1}{N}\sum_i r_{i,t+1}$ and $\overline{\widetilde{r}}_t = \frac{1}{N}\sum_i \widetilde{r}_{it}$ for the equal-weighted specification. In the value-weighted specification, I weight those same averages by firm market capitalization. Given the long-short nature of the characteristic portfolios I use in my analysis, the time series average of returns across portfolios is not informative and is therefore not shown.

The first two columns of table 1.2 show the results for estimating the "pooled" regression specification (1.2) on both individual equities and characteristic sorted portfolios. While the regression coefficients are positive and significant, the fraction of variance explained is close to zero. For this table, one-year return expectations have been re-scaled to obtain implied monthly expected returns. Even if subjective return expectations were efficient predictors of returns, however, analyst expectations are almost certainly a noisy measure of market expectations. It is that not surprising that the β coefficients estimated in both the pooled and time-series regressions are significantly below one.

The cross-sectional regressions are more interesting. One striking observation is the strong negative relationship between average firm-level return expectations and realized returns. In fact, average cross-sectional returns decline approximately one-for-one with average return expectations. Whatever biases give rise to this phenomenon, however, largely wash out when returns and expectations are aggregated to long-short portfolios. While portfolio return expectations are, on average, significantly too high and too extreme, subjective return expectations can explain a remarkable amount of the cross-sectional variation in portfolio returns. This suggests that even as analysts struggle to identify high-return and low-return assets in individual equities, even over a relatively long period, they nonetheless successfully identify the average differences in returns expected from the cross-sectional factors proxied for by the firm characteristics used in the portfolio sorts.

This latter finding suggests that, while analysts may be unsuccessful in forecasting variation in returns over time, especially, for individual equities, subjective return forecasts may still be useful for dynamic portfolio formation decisions if they are informative about expected factor "winners" and "losers." Figure 1.3 provides suggestive evidence that analysts are able to do this, not only on average across the entire sample, but dynamically, in each period—at least for managed portfolios. In this figure, a binned scatterplot, I rank observations cross-sectionally by expected returns and returns in each month. I then bin observations by expected return decile in month t and take the average return percentile in month $t + 1$ for observations within each bin. All percentiles are presented as deviations from the median, such that, for example, 1 on the vertical axis corresponds to the 51$^{\text{st}}$ percentile of returns.

If analysts had no ability to pick the winners and losers in each month, we would expect average return rank in the next month to fall near the median (zero in figure 1.3) , as forecast return rank would be no better than a random guess. Insofar as analysts are successful in separating each period's winners from losers, however, we should expect average return rank to be increasing with expected return rank in the prior month.

Similar to the cross-sectional findings from table 1.2, figure 1.3 shows that, while expected returns at the equity level appear largely uninformative about the winners and losers in the next period, expectations for managed characteristic portfolios are more informative. Figure 1.3 is a binned scatter plot showing the median (cross-sectional) percentile of returns in month $t+1$ over the percentile of expected returns in month t. All percentiles are differences from the median (i.e. the 50$^{\text{th}}$ percentile), such that a median return percentile of 1 for a given bin indicates that the median return within that bin is 1 percentile rank higher than the median return in the observation month. For individual equities (left panel), future return rank is nearly uncorrelated with expected return rank, indicating that analysts are generally unsuccessful in identifying the winners and losers each period at the level of individual equities. The Spearman rank correlation coefficient (ρ_{corr}) for individual equities is 0.009. For characteristic portfolios, however, the picture is rather different, and the median portfolio in the top decile of expected returns is nearly 12 percentiles higher than

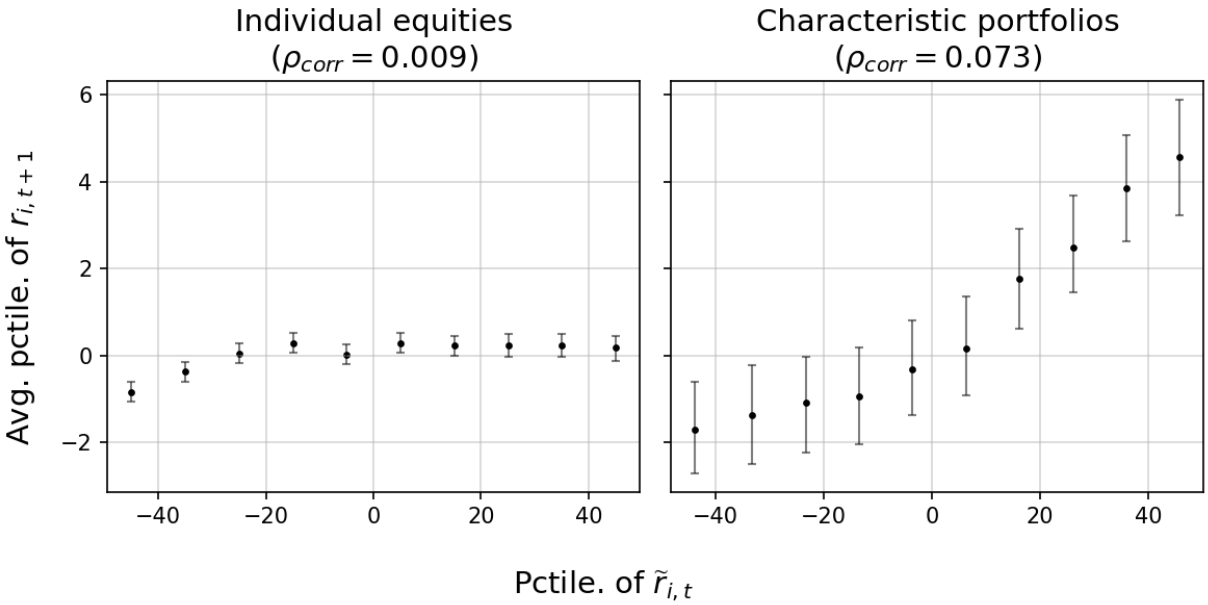


Figure 1.3: Return expectations and returns, rank correlations

Note: This figures depicts the cross-sectional percentile rank of expected returns in month t (horizontal axis) against the mean cross-sectional percentile rank for returns in month $t + 1$ (vertical axis). Observations have been binned into deciles of expected returns. All percentiles are presented as differences from the median, such that a mean return percentile of 1 for a given bin indicates that the median return within that bin is 1 percentile rank higher than the mean return in the month of the observation. Vertical lines depict 90% confidence intervals bootstrapped from 1,000 draws.

the median portfolio in the lowest decile, while ρ_{corr} is 0.073.

Examining the returns to a simple rank-sorted strategy trading on subjective return expectations provides an alternative mechanism for observing the relative informativeness of those expectations for the returns of individual equities versus managed portfolios. In table 1.3, I show the Sharpe and Sortino ratios[4], for strategies taking rank-weighted positions in either individual equities or characteristic portfolios. Individual equities (top two rows) or characteristic portfolios (bottom two rows) are ranked according to the value of the predictor indicated in each column. The weighting scheme used is identical to that used to form characteristic portfolios, which is shown in equation (1.1).

[4]Where the Sharpe ratio of a portfolio is defined as the mean return over the standard deviation of returns $\frac{\bar{r}_j}{\sigma_j}$, the Sortino ratio is defined as the mean return over the standard deviation of *downside* returns $\frac{\bar{r}_j}{\sigma_j^d}$, where σ_j^d is the standard deviation calculated using returns only less than zero. Where the Sharpe ratio penalizes volatility coming from upward movements, as well as downward, the Sortino ratio only penalizes downward. For portfolios with symmetric distributions, the Sharpe and Sortino ratios are identical. For portfolios, however, that are skewed to the positive side, the Sortino ratio will tend to be greater.

As it is indeterminate for many predictors whether the long leg of the strategy should correspond to high or low values of the predictor, the signs of the weights are chosen month-by-month such that average returns to the strategy for a given predictor would have been positive using the chosen weights. I.e. if going long assets with high values of $\widetilde{ep}_t$ resulted in average negative returns in the past (as is the case, on average, in my sample), then the rank-weighted strategy would assign positive weights to assets with low values of $\widetilde{ep}_t$ and negative weights to assets with high values. The period from 2003-2008 is taken as an initial in-sample period for the estimation of portfolio signs and is excluded. For both individual equities and characteristic portfolios, the weights assigned to both $\widetilde{r}_t$ and $\widetilde{\Delta p}_t$ are positive throughout the post-2008 period. From table 1.3 it can be

Assets	**Ratio**	**Predictor**							
		$\widetilde{r}_t$	$\widetilde{\Delta p}_t$	$\widetilde{dp}_t$	$\widetilde{ep}_t$	dp_t	mom_t	rev_t	bm_t
Equities	Sharpe	0.094	0.099	0.076	0.042	-0.011	0.047	0.133	0.064
	Sortino	0.155	0.163	0.096	0.052	-0.014	0.100	0.236	0.130
Portfolios	Sharpe	0.187	0.169	0.107	0.116	0.090	0.006	0.180	0.074
	Sortino	0.289	0.260	0.144	0.135	0.141	0.011	0.286	0.147

Note: This table shows the Sharpe and Sortino ratios to strategies formed by rank-weighting individual equities or characteristic portfolios according to the value of the predictor indicated above each column. In forming these strategies, the direction of each strategy is chosen month-by-month such that the average returns to the strategy using the chosen direction would have been positive in the period up that month. Cumulative log returns for the strategies in this table are shown in appendix figure A.2. All ratios are monthly values.

Table 1.3: Out-of-sample Sharpe and Sortino ratios, rank-weighted portfolios

seen that, even employing a relatively simplistic model for portfolio formation, we observe patterns in returns that are consistent with the descriptive results examined above. Here as well, I find that the $\widetilde{r}_t$ and $\widetilde{\Delta p}_t$ predictors are more successful at picking the winner and losers in each period after equities have been formed into portfolios, with both the Sharpe and Sortino ratios for the strategy using characteristic portfolios being nearly double those of the strategy using individual equities. Similar results also generally obtain for the other predictors I consider, reflecting the generally higher cross-sectional predictability of characteristic-sorted portfolio returns relative to those of individual equities. Remarkably, comparing across predictors, I find that the strategies formed on the basis of subjective return expectations consistently outperform those formed from the other predictors, with only the rev_t strategy offering similarly high risk-adjusted returns for characteristic portfolios.

The results presented here show that, even before any estimation is done on the data, subjective return expectations appear to be more informative than previously considered—at least with regards to the cross section of returns. Working with individual equities, however, tends to obscure the informational content of return expectations, and much clearer results are obtained when equity returns are first formed into portfolios based on characteristics that are likely to covary with expected risk exposures. In further sections, I will apply more structure to the problem of forecasting returns to examine the usefulness of subjective return expectations for forecasting the level of returns, and whether the predictive content of return expectations can be used to form strategies that improve on the naive strategies considered in table 1.3.

1.4 Predictive model estimation

Next I explore whether the informational content of subjective return expectations can be leveraged to more accurately forecast returns. In the analysis that follows, I explore variations on the following predictive regression:

$$r_{j,t+1} = \alpha_j + \beta_j^{\intercal} x_t + \epsilon_{j,t+1} \tag{1.5}$$

The predictors x_t are not assumed to be centered, so the inclusion of a constant α_j in this regression is purely practical, and has no asset pricing significance. The key assumption of this predictive model is that the returns of all test asset portfolios can be forecast from a common set of predictors. In the broadest case I consider, x_t will be taken to be the full panel of predictor signals for all characteristic sorted portfolios, with the number of predictors equal to the number of assets being forecast. The motivation behind this modeling decision is the simple intuition that, insofar as the predictor signals for different portfolios contain independent information, the value of a predictor for one characteristic portfolio may be of use in forecasting returns for another.

The challenge, however, is that, even having reduced the cross-section of returns from a panel of thousands of firms to less than one hundred characteristic-sorted portfolios, estimation of (1.5) via OLS is infeasible due to high correlations between predictors and the limited time series available. Given the wealth of data available in recent decades, this is now a common problem in the empirical finance and economics literatures and numerous solutions have been proposed, ranging from simple shrinkage and selection methods (ridge regression, LASSO, LAR, elastic nets) and linear dimensional reduction techniques (PCA, PLS), to elaborate machine learning techniques (neural nets, random forests).

Here, I have attempted to error on the side of simplicity. In part, this decision was made to highlight more clearly the advantages of the data I am considering, which do not require especially complicated methodologies to reveal. Practical experience, however, also suggests that a simpler, but more carefully optimized approach may be superior to more powerful techniques that are difficult to tune, either because they require the selection of numerous hyperparameters, or because their computational costs make repeated estimation challenging.

With that in mind, I employ three tools to tame the large panels of predictors I work with. The first is L_2 regularization in the estimation of the regression coefficients a_j and β_j. Commonly referred to as "ridge regression'," this is a shrinkage method that pulls the estimated coefficients towards zero as the regularization penalty is increased. I also explore regularization by compressing the panel of predictors into low-rank approximations. In the context I examine, many predictors are highly correlated and well approximated by a small number of low-dimensional factors, which suggests that rank reduction and selection on factors will be more successful than methods that select individual predictors for inclusion or exclusion from my model. The techniques I consider here are PCA, which is well known in the economics forecasting literature following the introduction of "diffusion forecasting" in Stock and Watson, 2002a; Stock and Watson, 2002b and PLS, which has recently been examined in Kelly and Pruitt (2013) and Kelly and Pruitt (2015). Finally, as both ridge regression and PCA/PLS require the estimation of hyperparameters, I employ a K-fold cross-validation procedure for hyperparameter selection and model averaging as a final regularization step.

1.4.1 Shrinkage

The primary regularization mechanism I employ is shrinkage L_2 penalty on the norm of both sets of coefficients. In all out-of-sample forecasting exercises, a and B are thus chosen to solve the "ridge" regression objective:

$$\min_{\alpha_j,\beta_j} \sum_{t=1}^{T} \left(r_{j,t+1} - \alpha_j - \beta_j^\intercal x_t \right)^2 + \phi \left(\alpha_j^2 + \beta_j^\intercal \beta_j \right)$$

The hyperparameter ϕ controls the strength of the regularization, with higher ϕ shrinking the estimates of α_j and β_j towards zero.

1.4.2 Rank reduction

In can be shown, as in Hastie, Tibshirani, and Friedman (2009), that the shrinkage induced by L_2 regularization is concentrated in the smallest singular values of the predictor matrix. PCA and PLS, on the other hand, can be understood as singular value *thresholding* methods: PCA retains only the first K singular vectors of the covariance matrix of the predictors, while PLS (as implemented here) retains the first K (left) singular vectors of the cross-covariance matrix between predictors and returns.

For what follows, it will be convenient to work with matrix forms of returns and predictors. Let $R_{T\times N}$ be the matrix of returns, $X_{T\times M}$ the matrix of predictors, and $E_{T\times N}$ a matrix of regression errors. In practice, X will consist simply of the matrix of signal portfolios, such that $M = N$. Both PCA and PLS will be estimated on cross-sectionally demeaned data. These cross-sectionally demeaned data matrices will be denoted $\dot{R}$ and $\dot{X}$.[5]

Using the full panel of predictors to forecast returns is equivalent to the following time-series regression:

$$r_{t+1} = a + Bx_t + e_{t+1} \tag{1.6}$$

where B is an $N \times M$ matrix of prediction coefficients, and a is a column vector of intercepts.

[5]To be concrete, $\dot{R} \equiv (I_T - \frac{1}{T}J_T)R$, where J_T is a $T \times T$ matrix of ones. $\dot{X}$ is defined in the same way.

The objective of both PCA and PLS is to estimate a projection matrix $\Gamma_{M\times K}$ that condenses X into a reduced-rank matrix consisting of K linear combinations of the original predictors. Using this reduced-rank set of predictors, equation 1.6 becomes:

$$r_{t+1} = a + B\left(\Gamma^{\intercal} x_t\right) + e_{t+1} \tag{1.7}$$

with the dimension of B being similarly reduced to $N \times K$. The key difference between PCA and PLS is the methodology used to select the projection matrix Γ_K for constructing the reduced-rank predictors. In PCA, the matrix Γ is chosen to maximize the fraction of variance explained in the predictor matrix by the transformed predictors. In the factor estimation step, Γ is chosen to maximize the objective:

$$\begin{aligned} &\max_{\Gamma} \quad \left(\dot{X}\Gamma\right)^{\intercal} \dot{X}\Gamma \\ &s.t. \quad \Gamma^{\intercal}\Gamma = I \end{aligned}$$

Intuitively, the factors identified by PCA are the K combinations of the underlying signal portfolios that best explain the variance in those portfolios. A potential limitation of the procedure used by PCA to form the set of low-rank predictors is that the factor estimation procedure does not depend on returns. This means that, in forming the set of reduced-rank predictors, no consideration is given to the ultimate objective, which is the forecasting of returns. The success of PCA thus hinges on the predictive content of the predictors being concentrated in the first few principal components of the signal matrix.

PLS, on the other hand, estimates Γ to maximize the covariance between the reduced-rank predictors and returns:

$$\max_{\Gamma,\Pi} \ \left(\dot{X}\Gamma\right)^{\intercal} \dot{Y}\Pi$$
$$s.t. \ \ \Gamma^{\intercal}\Gamma = \Pi^{\intercal}\Pi = I$$

where $\Pi_{N\times K}$ is a matrix that projects the columns of $\dot{Y}$ into the same low-rank space as $\dot{X}\Gamma$. Alternatively, PLS can be interpreted as minimizing the square norm of the difference between returns and predictors when both are projected into that low-rank space:

$$\max_{\Gamma,\Pi} \ \left(\dot{X}\Gamma - \dot{Y}\Pi\right)^{\intercal} \left(\dot{X}\Gamma - \dot{Y}\Pi\right)$$
$$s.t. \ \ \Gamma^{\intercal}\Gamma = \Pi^{\intercal}\Pi = I$$

Several algorithms have been proposed for solving the PLS objective—NIPLS, SIMPLS, etc.—most of which require iterative estimation of the columns of Γ and Π.[6] In practice, however, I find that the differences between the solutions obtained appear to be negligible. For this reason, I solve the PLS objective via singular value decomposition of the cross-covariance matrix of $\dot{X}^{\intercal}\dot{Y}$, taking the first K right singular values as my estimate of Γ, and the first K left singular values as my estimate of Π . It should be stressed that PLS factors are not estimated separately for individual portfolio return series. Instead, the full cross-section of returns are used for the target matrix when extracting factors from the predictors.[7]

The result of both the PCA and PLS procedures is an estimated weighting matrix $\widehat{\Gamma}^{(a)}$, where $a \in \{pca, pls, ridge\}$ denotes the factor model used. When $a \in \{pca, pls\}$, $\widehat{\Gamma}^{(a)}$ will be understood to refer to the reduced rank matrix of loadings derived by either of those model, while

[6]For a detailed consideration of the PLS problem and its solution methods, see, for example, De Bie, Cristianini, and Rosipal (2005)

[7]These two estimation methods are sometimes referred to as PLS1 (single target) and PLS2 (multiple target). The estimation here is PLS2.

$a = ridge$ will simply imply that $K = M$ such that $\widehat{\Gamma}^{(a)} = I$. To keep notation concise later, I will use $\widehat{g}_t^{(a)} \equiv \widehat{\Gamma}^{(a)\intercal} x_t$ to denote the estimated predictor factors from each model.

1.4.3 Hyperparameter selection and model averaging

The key challenge in both the L_2 penalization and dimension reduction methods described above is the selection of relevant hyperparameters. In the case of ridge regression, the strength of the regularization parameter ϕ must be chosen, while in the case of PCA and PLS, the rank of the reduced-rank predictors must be selected.

I employ K-fold cross-validation to tune the selection of these hyperparameters. K-fold cross-validation divides the data into D non-overlapping subsets (or "folds"). During training, one of these subsets is left out of the training, while model parameters are estimated on the remaining $D - 1$ folds. Hyperparameters are then selected to minimize MSE in the validation set. In order to maintain the temporal structure of my data, I divide the in-sample period into D non-overlapping, sequential blocks. For each fold, $\widehat{\phi}^{(a_d)}$ is selected to minimize MSE on that fold, and the regression parameters $\widehat{\alpha}_i^{(a_d)}$ and $\widehat{\beta}_i^{(a_d)}$ (which depend on $\widehat{\phi}^{(a_d)}$) are estimated on the training folds. The output from this procedure is D predictive models of the form:

$$\widehat{r}_{j,t+1}^{(a_d)} = \widehat{\alpha}_j^{(a_d)} + \widehat{\beta}_j^{(a_d)\intercal} \widehat{g}_t^{(a)} \tag{1.8}$$

where $\widehat{r}_{j,t+1}^{(a_d)}$ is the forecast for $r_{j,t+1}$ using model a, estimated using the d^{th} fold as the validation set and the remaining folds for training. Given these D models, the final estimate for model a (without d subscript) is formed as the equal weighted average of the models estimated on each train/validation split. This sort of model averaging helps to improve the stability of parameter estimates, and is commonly employed in the machine learning literature, where stochastic training algorithms This technique is frequently employed in machine learning contexts, and has been previously used in the estimation of empirical asset pricing models papers such as Gu, Kelly, and

Xiu (2021):

$$\widehat{r}_{j,t+1}^{(a)} = \frac{1}{D}\sum_{d=1}^{D}\left[\widehat{\alpha}_j^{(a_d)} + \widehat{\beta}_j^{(a_d)\top}\widehat{g}_t^{(a)}\right] \tag{1.9}$$

One minor point that should be noted is that, for the factor models, the optimal number of factors K, up to a maximum of $K = 10$, is selected to minimize the average MSE across all of the train-validation splits of the data. As each of the models being combined is identical in structure and numbers of parameters, only estimated on a different sample and optimized against a different validation set, it is thus justified to use a simple equal weighting scheme in combining the predictions of the models. In practice, I set $D = 4$.

I simulate the out-of-sample performance of each model by estimating $\widehat{g}_t^{(a)}$ and the parameters in (1.9) over an expanding window, taking the first 60 months of my sample as an initial in-sample period. For each month t, factor weights $\widehat{\Gamma}^{(a)}$ are estimated over months $1, \ldots, t-1$. Given the in-sample factor estimates obtained, the optimal L_2 penalty for each train/validation pair is estimated via the cross-validation procedure described above, regression parameters in (1.8 are estimated, and those models are averaged to obtain the final estimate (1.9). The final observed value of the predictor factors $\widehat{g}_t^{(a)}$ is then used to forecast the following month's returns.

Despite the regularization imposed by compressing the predictors into reduced-rank factors, I nonetheless find that both model averaging and the addition of L_2 regularization in estimating the predictive regression can be beneficial. Figure 1.4 shows the variation in simulated out-of-sample R^2 achieved for different numbers of cross-validation folds D used to estimate models for averaging, and L_2 penalty. This example, which shows results using the PCA$_6$ model and $\widetilde{r}_t$, suggests that both model averaging and L_2 regularization contribute to the final performance of the model, with the combination being more effective than either alone.[8]

[8]It should be stressed that this exercise assumes that both the regularization penalty and number of factors are held fixed across all train/validation splits and periods. In practice, however, both the strength of the L_2 regularization and the optimal number of factors are hyperparameters that need to be estimated. This is reflected in my simulated other simulated out-of-sample exercises, so the results displayed in this figure are not directly comparable to the results shown in table 1.5.

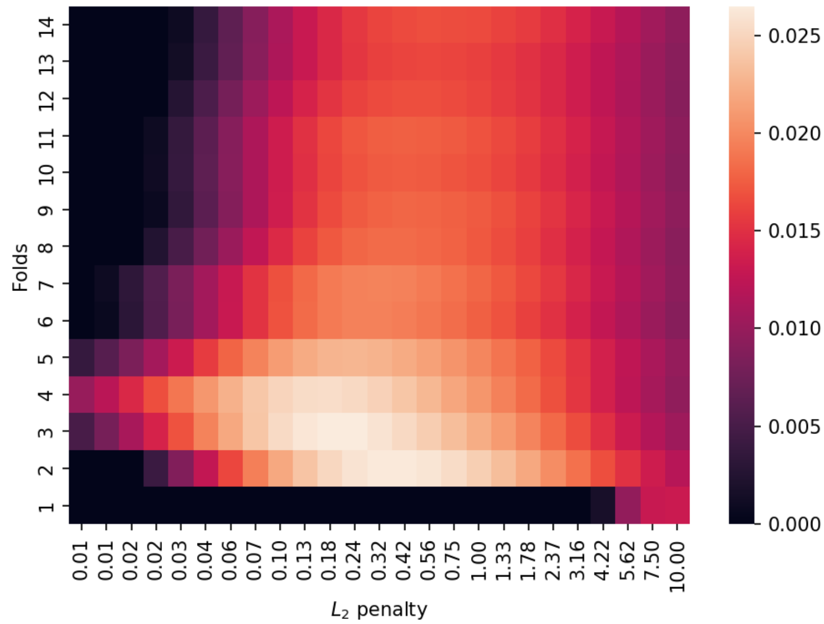


Figure 1.4: The impact of L_2 regularization and model averaging

Note: This figure shows the predictive R^2 achieved by varying the L_2 regression penalty when estimating the forecasting equation (1.9), and the number of folds over which the model is estimated and averaged. Results shown here are for the PCA_6 model, using $\widetilde{r}_t$ as the predictor.

1.5 Predicting returns

1.5.1 In-sample predictive regressions

In sample, I consider two basic predictive models. In the first (labeled "OLS" in table 1.4), returns for each characteristic portfolio are regressed on only the predictor signal for the same portfolio in a univariate time-series regression:

$$r_{j,t+1} = \alpha_j + \beta_j x_{jt} + \epsilon_{j,t+1} \tag{1.10}$$

In the second specification, returns are regressed on the first K factors estimated from the panel of predictor signal portfolios, via either PCA or PLS:

$$r_{j,t+1} = \alpha_j + \beta_j^{\intercal} g_t^{(a)} + \epsilon_{j,t+1} \tag{1.11}$$

Table 1.4 shows the in-sample fit achieved by each of these models, measured by R^2, which is calculated in the usual way. From this table, several observations can be made.

First, insofar as each portfolio's own predictor signal is an adequate summary of the information in the predictor panel relevant for that portfolio, it should be expected that there should be little benefit to allowing for cross-predictability. The univariate regression of each portfolio's returns on its own predictor signal does not, however, consistently explain more of the variation in returns than the first PCA factor, and explains much less of the variation in returns than the first PLS factor. This latter finding suggests that the cross section of predictors has substantial informational content beyond that contained in each portfolio's own predictor signal. While this is not surprising in the case of those predictors that are not themselves forecasts of subjective returns, and is a fact that has previously been explored in Kelly and Pruitt (2013), it is interesting to observe that this is also true for the expected return and expected price growth predictors. These predictors represent, after all, the best estimate of market analysts for what the returns to a given portfolio will be. For those predictors, the substantial difference in fit between what is achievable using only a portfolio's own signal and a single factor extracted from the cross-section suggests that either consensus forecasts are not representative of market expectations (as, for example, if the market predominantly reflects the expectations of optimists à la Miller, 1977), or that forecasts do not efficiently incorporate all available information in a timely fashion (due, for example, to delays in information diffusion as in Hong and Stein, 1999).

I also observe that the R^2 statistics for $\widetilde{r}_t$ and $\widetilde{\Delta p}_t$ are frequently well above those of other predictors, across a wide range of model specifications. The exceptions to this pattern come in the OLS model, and the PCA models with relatively few factors. For the OLS model, the own portfolio B/M ratio achieves the highest R^2 by a substantial margin. The same holds true for PCA_1. Moving from PCA_1 to PCA_2 then sees a large jump in model fit for dp_t . Above PCA_3, however, the fit achievable using the expected return predictors dominates the others, suggesting that return expectations contain richer cross-sectional information than any of the other individual predictors. For the PLS models, the R^2 for $\widetilde{r}_t$ and $\widetilde{\Delta p}_t$ are already as good as those of any other predictor with only a single factor, and decisively greater by PLS_3 and above. Of course, the success of these models in sample does not guarantee success out of sample, so I next simulate out-of-sample estimation of all models to see how they perform. Given the clear evidence that cross-sectional information is valuable in explaining returns, I will drop the own-signal OLS model from further consideration.

Model	K	**Predictor**							
		$\widetilde{r}_t$	$\widetilde{\Delta p}_t$	$\widetilde{dp}_t$	$\widetilde{dp}_t$	dp_t	mom_t	rev_t	bm_t
OLS	0	0.010	0.009	0.007	0.008	0.013	0.004	0.010	0.031
	1	0.006	0.005	0.019	0.006	0.007	0.008	0.011	0.036
	2	0.030	0.029	0.026	0.025	0.051	0.033	0.017	0.051
	3	0.050	0.050	0.037	0.034	0.056	0.053	0.019	0.057
PCA_k	4	0.086	0.085	0.047	0.051	0.063	0.055	0.024	0.065
	5	0.100	0.100	0.051	0.056	0.071	0.058	0.029	0.067
	6	0.107	0.107	0.063	0.060	0.073	0.062	0.039	0.069
	1	0.047	0.047	0.038	0.019	0.047	0.022	0.013	0.044
	2	0.055	0.054	0.048	0.054	0.061	0.028	0.020	0.055
	3	0.114	0.114	0.065	0.069	0.076	0.057	0.037	0.063
PLS_k	4	0.131	0.131	0.090	0.099	0.094	0.066	0.046	0.075
	5	0.139	0.140	0.102	0.116	0.112	0.097	0.063	0.090
	6	0.147	0.148	0.111	0.127	0.121	0.116	0.077	0.097

Note: This model shows the in-sample R^2 obtained by regressing portfolio returns on either the predictor series for the same portfolio ("OLS"), or on the first k factors of the predictor panel ("PCA_k" and "PLS_k").

Table 1.4: In-sample predictive R^2

1.5.2 Out-of-sample predictability

First, I consider the ability of each model to predict returns out of sample. Given the panel structure of the model I calculate the out-of-sample R^2_{oos} as follows:

$$R^2_{oos} = 1 - \frac{\sum_{i=1}^{N} \sum_{t=61}^{T} \left(r_{i,t+1} - \widehat{r}^{(a)}_{i,t+1} \right)^2}{\sum_{i=1}^{N} \sum_{t=61}^{T} \left(r_{i,t+1} - \overline{r}_{it} \right)^2}$$

It should be noted that in calculating this statistic, I use the squared difference between realized return and the trailing mean. It has recently become common in the empirical asset pricing literature to instead use only squared returns $r^2_{i,t+1}$ in the denominator when evaluating out-of-sample predictive performance. The typical justification given for this is that, in the typical samples and for the assets considered in those studies, demeaning by the trailing mean tends to inflate the denominator and estimated R^2_{oos}, making square returns a more conservative measure. I find, however, that, for the sample and assets I consider, that measured R^2_{oos} is greater when using an uncentered denominator, and so, following the principle of using the more conservative statistic, I use demeaned returns in the denominator.

Sample	Model	Predictor							
		$\widetilde{r}_t$	$\widetilde{\Delta p}_t$	$\widetilde{dp}_t$	$\widetilde{dp}_t$	dp_t	mom_t	rev_t	bm_t
	Ridge	0.012	0.012	0.006	-0.005	**0.012**	0.000	-0.002	**0.017**
	PCA_1	0.003	0.003	0.002	-0.003	0.008	-0.003	0.001	**0.019**
Full	PCA_k	**0.017**	0.013	0.009	-0.005	**0.011**	0.000	-0.006	**0.017**
	PLS_1	**0.027**	**0.029**	-0.002	-0.002	**0.024**	-0.001	-0.007	**0.024**
	PLS_k	0.014	0.013	-0.051	-0.023	-0.006	-0.003	-0.019	-0.012
	Ridge	**0.047**	**0.048**	0.004	**0.029**	**0.016**	0.009	-0.012	0.012
	PCA_1	**0.026**	**0.026**	-0.010	**0.039**	0.010	-0.009	-0.012	0.020
Rec.	PCA_k	**0.064**	**0.053**	0.017	**0.028**	**0.021**	-0.001	-0.030	0.023
	PLS_1	**0.091**	**0.100**	-0.005	**0.068**	**0.059**	0.001	-0.041	0.040
	PLS_k	0.085	**0.090**	-0.074	-0.008	**0.042**	0.009	-0.044	-0.014
	Ridge	-0.001	-0.001	0.006	-0.017	**0.011**	-0.003	0.002	**0.019**
	PCA_1	-0.005	-0.006	0.006	-0.018	0.007	-0.001	0.006	**0.018**
Exp.	PCA_k	0.000	-0.002	0.006	-0.017	0.008	0.000	0.003	0.015
	PLS_1	0.004	0.004	-0.001	-0.027	0.011	-0.001	0.006	**0.019**
	PLS_k	-0.011	-0.014	-0.043	-0.029	-0.024	-0.007	-0.010	-0.012

Note: Out-of-sample predictive R^2_{oos}. Bolded values indicate that the improvement in forecast error over the trailing mean is significant at a better than 0.1 significance level. All p-values are calculated using the modified Diebold-Mariano test described in the body.

Table 1.5: R^2_{oos} by model and predictor

To assess the statistical significance of the observed differences in predictive performance, I employ a variant of the test statistic of Diebold and Mariano (2002) that was proposed for comparing the performance of competing models across a cross-section of portfolios in Gu, Kelly, and Xiu (2020). In the modified Diebold-Mariano statistic of Gu, Kelly, and Xiu (2020), the difference in squared forecast errors from two competing models are reduced to a single time series by cross-sectionally averaging that difference in each period.

Letting $\widehat{e}^{(a)}_{j,t+1} = r_{j,t+1} - \widehat{r}_{j,t+1}$ be the estimated one-period ahead forecast error in the returns of portfolio i under forecast model a, the modified Diebold-Mariano statistic for testing whether a is a better predictor of returns than an alternative model b is:

$$DM^{(ab)} \equiv \frac{\frac{1}{T-61}\sum_{t=61}^{T}\left[\overline{d}_t^{(ab)}\right]}{\widehat{\sigma}\left(\overline{d}_t^{(ab)}\right)}$$

The numerator of this statistic is the time-series average of a cross-sectional average squared forecast error difference $\overline{d}_t^{(ab)}$, defined as:

$$\overline{d}_t^{(ab)} \equiv \frac{1}{N}\sum_{j=1}^{N}\left[\left(\widehat{e}_{j,t+1}^{(a)}\right)^2 - \left(\widehat{e}_{j,t+1}^{(b)}\right)^2\right]$$

The denominator of the Diebold-Mariano $\widehat{\sigma}\left(\overline{d}_t^{(ab)}\right)$ is the standard deviation of $\overline{d}_t^{(ab)}$, appropriately adjusted for auto-correlation. As with the usual Diebold-Mariano statistic, in practice, this statistic is estimated as the t-score in a time-series regression of $\overline{d}_t^{(ab)}$ on a constant. To account for potential auto-correlation in $\overline{d}_t^{(ab)}$, I use Newey-West standard errors with one lag.

Results for the estimation and testing of R^2_{oos} are shown in table 1.5. I find that over the full out-of-sample period, the maximum R^2_{oos} values obtained using $\widetilde{r}_t$ or $\widetilde{\Delta p}_t$ as predictors is somewhat greater than that obtained using any of the other predictors, with bm_t or dp_t being the strongest performers amongst those others. Examining only the performance of these predictors over the full sample, however, masks striking differences in performance over sub-samples consisting of only recession and expansion periods. During recessions, expected returns and price growth show exceptional predictive performance, easily outperforming the strongest models estimated using any of the other predictors. During expansions, however, the best performance achievable using either expected returns or price growth is no better than the naive trailing mean forecast. The portfolio book-to-market ratio proves to be a more consistent predictor, exhibiting similar performance in both recessions and expansions.

Figure 1.5 shows this time-variation in predictive performance graphically, and depicts the cumulative difference in squared forecast error (CDSFE), relative to the trailing mean. CDFSE was suggested in Goyal and Welch (2003) and Goyal and Welch (2008) and provides an intuitive view

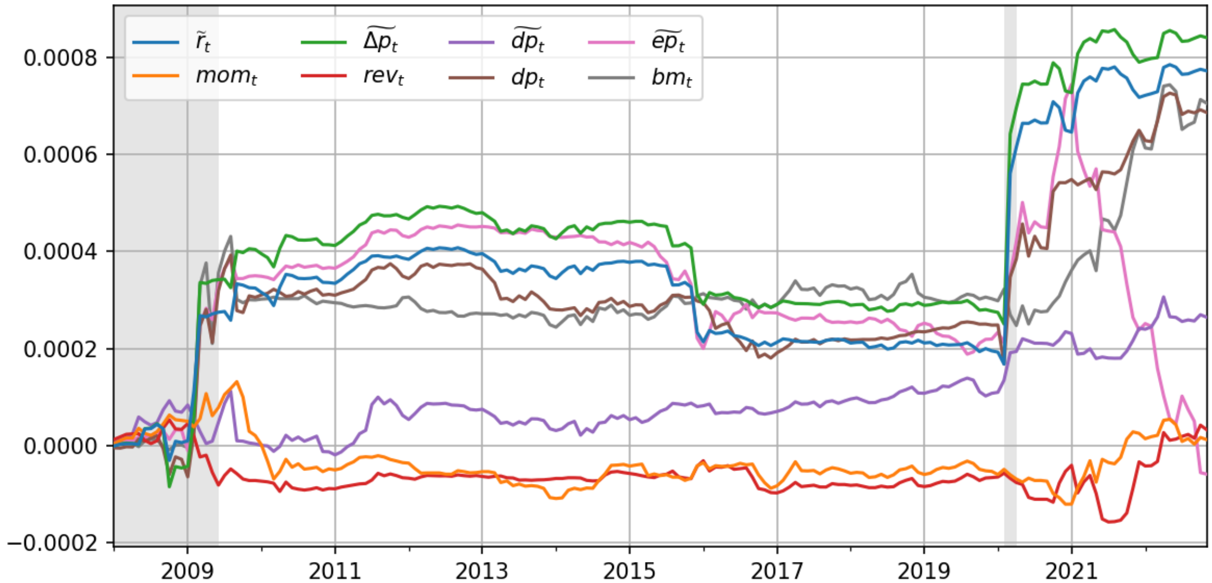


Figure 1.5: Cumulative difference in squared forecast error, relative to trailing mean

Note: This figure plots the cumulative difference in squared forecast error (CDFSE), relative to the trailing mean. For each predictor, I plot the best performing model according to average predictive R^2 from table 1.5.

of the relative forecasting performance of a number of models over time. Given the cross-sectional dimension of my data, I report the average CDSFE across test assets, for the best performing model for each predictor. The CDFSE for model a up to period t is defined:

$$CDFSE_t^{(a)} = \frac{1}{N}\sum_{j=1}^{N}\sum_{\tau=61}^{t}\left(r_{j,\tau} - \bar{r}_{j,\tau-1}\right)^2 - \frac{1}{N}\sum_{j=1}^{N}\sum_{\tau=61}^{t}\left(r_{j,\tau} - \widehat{r}_{j,\tau-1}^{(a)}\right)^2$$

The most striking fact that emerges from this figure is that the predictive advantage of the expected return and expected price growth models, relative to those of other predictors, is greatest during the 2009 and 2020 crisis periods. Given the short observation window available, caution should be taken in reading too much into this observation, but it is nonetheless suggestive that the forward looking content in subjective returns may be most useful in times of great uncertainty. As was observed above, analyst forecast production also exhibited peaks around these same periods, and this suggests that the additional informational production in those periods may result in relatively more informative forecasts than in other times.

I find that the maximum average (across portfolios) predictive R^2_{oos} is achieved by the PLS1 model estimated using expected price growth. Figure 1.6 examines the predictability of the varying characteristic portfolios using this model.[9] Dividing those portfolios into categories, I find that predictability is greatest for the subset of valuation ratio portfolios formed on the basis of various price-earnings ratios. The characteristic portfolio with the highest predictive R^2_{oos} ("capei"), is the Shiller cyclically adjusted P/E ratio, with other variants on the price-earnings ratio (prefix "pe") exhibiting similar levels of predictability. In addition to valuation ratios, high predictive R^2_{oos} values are also observed for portfolios formed on the basis of profitability.[10]

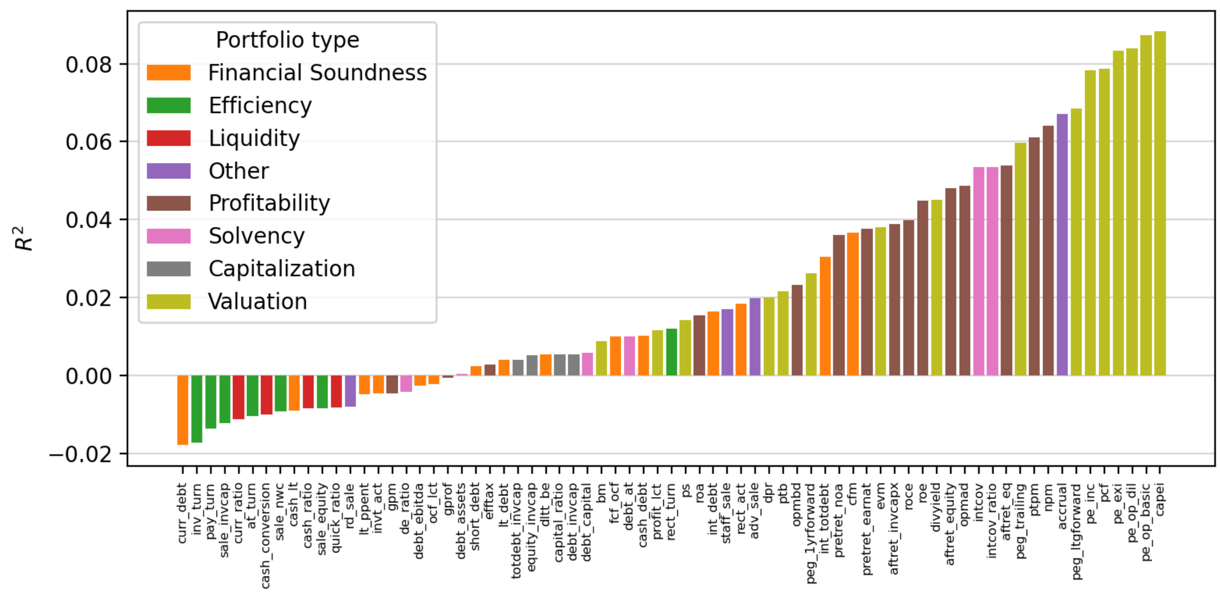


Figure 1.6: R^2_{oos} by portfolio characteristic

Note: This figure shows R^2_{oos} by characteristic portfolio for the PLS_1 model using $\widetilde{\Delta p_t}$ as the predictor. The vertical axis of this figure shows predictive R^2_{oos} for each characteristic over the out-of-sample period. Bar color indicates characteristic type classification, as assigned by WRDS. For interpretation of individual variable labels, refer to appendix table A.1

[9]The second-highest average predictive R^2_{oos} out-of-sample is achieved by the PLS1 model estimated with expected returns. The equivalent to figure 1.6 for that model/predictor combination is nearly identical, so I do not present it here.

[10]Interestingly, while I do not reproduce this figure for all models/predictors here, I observe that valuation ratio and profitability portfolios tend to be amongst the most predictable portfolios, irrespective of the predictor and model used.

1.5.3 Do expected returns allow for better portfolio allocations?

Because the greatest predictive advantage available using expected returns occurs during times of economic turmoil, however, taking full advantage of that increased predictability seems likely to be challenging. To see why, it is informative to look at the mean-variance efficient solution to forming portfolios from the conditional expected returns estimated from each model. The estimation of some form of mean-variance efficient portfolio, however, brings with it all of the usual challenges. Even with the reduced cross-sectional dimension achieved by working with portfolios in place of equities, the short length of the sample I am able to observe makes accurate estimation of the covariance matrix of returns challenging.

In order to reduce the noise that results from the estimation of the covariance matrix for the full panel of assets, I use a low-rank PCA approximation of the covariance matrix of returns in estimating mean-variance efficient asset weights. These are not, it should be emphasized, the same factors estimated from the matrix of predictors. Rather, after estimating forecasts for returns at time $t+1$ from data up to time t, I perform a second factor decomposition, but this time only using the covariance matrix of returns observed until that point. Forecasts for individual portfolios and then combined into a further reduced set of portfolios that lines up with the first few eigenvectors in returns. Theoretical motivation for only retaining the largest eigenvalues of returns, and then forming conditional factor timing portfolio weights is provided in Kozak, Nagel, and Santosh (2020) and Haddad, Kozak, and Santosh (2020), which show that the absence of near arbitrage opportunities requires predictable variation in returns be concentrated in the largest eigenvalue factors. The factor loadings estimated from returns are all normalized to have unit length, ensuring that total position size at this step remains constrained. As these portfolios are linear combinations of the zero-cost long-short portfolios I use as test assets, they are themselves self-financing long-short portfolios, so the net position taken in the market is constant at zero.

Letting $\widehat{\Lambda}_{N\times L} = \left[\widehat{\lambda}_1, \ldots, \widehat{\lambda}_L\right]$ be the the matrix of loading vectors estimated via PCA decomposition of the covariance matrix of portfolio returns. Period $t+1$ returns for factor $l \in 1, \ldots, L$

are then formed as:

$$f_{l,t+1} = \widehat{\lambda}_l^{\intercal} r_{t+1}$$

The time t forecast for factor l under model a is denoted $\widehat{f}_{l,t+1}^{(a)}$ and estimated from portfolio forecasts as:

$$\widehat{f}_{l,t+1}^{(a)} = \widehat{\lambda}_l^{\intercal} \widehat{r}_{t+1}^{(a)}$$

Letting $\widehat{\Sigma}_f = diag(\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_L^2)$ be the factor covariance matrix. The orthogonality of the factors then implies the conditional mean-variance efficient portfolio weight for each factor is given by:

$$w_{lt} = \frac{1}{L} \frac{\widehat{f}_{l,t+1}^{(a)}}{\widehat{\sigma}_l^2}$$

and mean-variance efficient portfolio returns are:

$$r_{t+1}^{mv} = \sum_{l=1}^{L} w_{lt} f_{l,t+1}$$

It should be noted, however, that no other constraints are imposed in the formation of these portfolio weights. Implicitly, the portfolio returns observed under this weighting scheme are those that are achievable by an investor in a frictionless market, with the ability to scale both the long and short legs of their portfolio to any size. It also does not take into consideration factors such as turnover, which is likely to be high for several of underlying characteristic portfolios.

While the Sortino ratio is not often seen in academic finance, it is informative in this case due to the tendency of the portfolios formed on the basis of expected returns to "crash up" during economic crisis periods. As the Sharpe ratio implicitly assumes a symmetric distribution of returns, and penalizes upward and downward movements equally, it gives a rather incomplete picture of those portfolios. The Sortino ratio, on the other hand, only penalizes variation in the negative direction.

Table 1.6 shows the Sharpe and Sortino ratios for each predictor/model combination. By both the Sharpe and Sortino measures, I find that the expected return measures $\widetilde{r}_t$ and $\widetilde{\Delta p}_t$ achieve the

Ratio	Model	Predictor							
		$\widetilde{r}_t$	$\widetilde{\Delta p}_t$	$\widetilde{dp}_t$	$\widetilde{dp}_t$	dp_t	mom_t	rev_t	bm_t
	Ridge	0.178	0.176	0.148	0.064	0.141	0.137	0.100	0.102
	PCA_1	0.136	0.136	0.103	0.088	0.094	0.126	0.076	0.158
Sharpe	PCA_k	0.207	0.181	0.180	0.044	0.114	0.102	0.023	0.138
	PLS_1	0.198	0.204	-0.009	0.091	0.143	0.115	0.091	0.156
	PLS_k	0.136	0.171	0.034	0.063	0.074	0.142	0.173	0.033
	Ridge	0.363	0.351	0.181	0.080	0.238	0.178	0.134	0.115
	PCA_1	0.202	0.198	0.094	0.125	0.099	0.133	0.098	0.229
Sortino	PCA_k	0.434	0.320	0.230	0.052	0.188	0.109	0.025	0.214
	PLS_1	0.898	0.949	-0.010	0.163	0.245	0.147	0.101	0.299
	PLS_k	0.210	0.451	0.042	0.074	0.102	0.239	0.236	0.034

Note: This table shows the out-of-sample Sharpe and Sortino ratios for each combination of model and predictor. Details of the out-of-sample estimation procedure are described in the text. PCA_1 and PLS_1 models are restricted to only using a single factor in all periods. For the PCA_k and PLS_k models, the number of factors employed is chosen by cross-validation, up to a maximum of ten. All ratios are monthly values.

Table 1.6: Out-of-sample Sharpe and Sortino ratios, unconstrained weights

highest risk-adjusted returns of any of the predictors. Similar to the results observed in table 1.5, the PLS_1 model achieves the best performance, with Sortino ratios that are more than triple those achieved by any other predictor.

There is also a large discrepancy between the Sharpe and Sortino ratios for $\widetilde{r}_t$ and $\widetilde{\Delta p}_t$, especially for the PLS_1 strategy. This is due to upward crashes in the returns to that strategy during crises. This can be seen in figure 1.7, which plots the log cumulative returns for the maximum Sortino ratio model estimated for each predictor. While the magnitude of the jumps observed in the $\widetilde{r}_t$ and $\widetilde{\Delta p}_t$ portfolios is striking, it is consistent with the jumps in predictive performance observed during those periods, and reflects the model making very large directional bets during recessions.

Figure 1.8 plots the absolute position size taken by the maximum Sortino ratio model for each predictor over time. Because the base portfolios that are being blended are self-financing, long-short portfolios, the net position and leverage across portfolios is constant. The total absolute size of the position, however, varies over time as forecast returns to portfolio rise and fall. In some cases these differences can be dramatic. The maximum Sortino ratio model is the PLS_1 model estimated for expected price growth. For this model, however, the maximum position size is

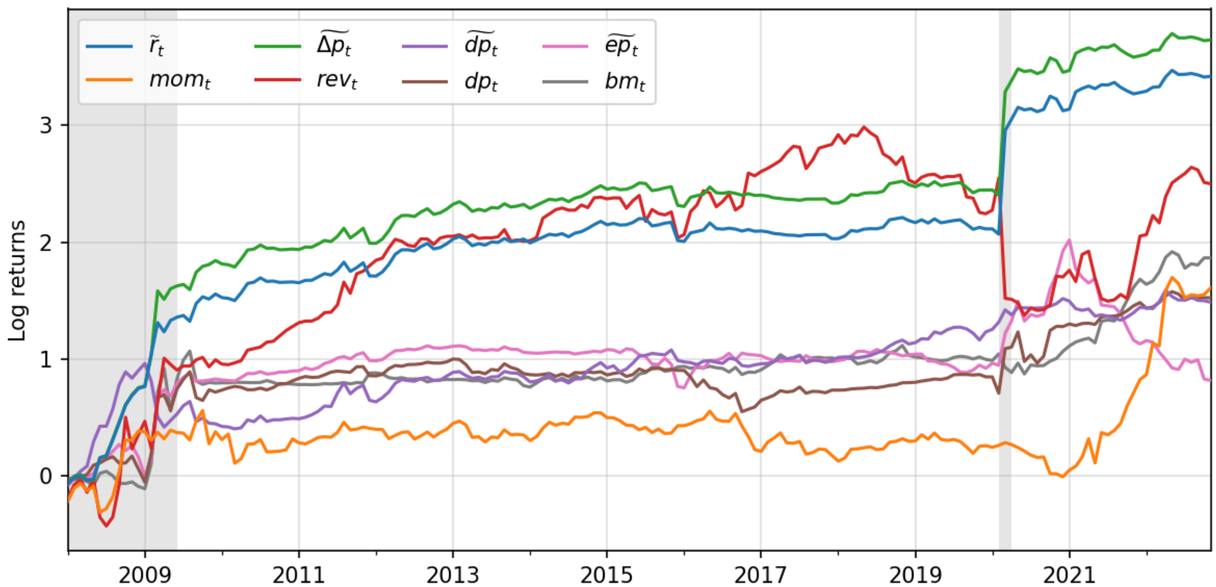


Figure 1.7: Cumulative log returns, unconstrained MV-optimizing portfolio

Note: This model shows the cumulative log returns for the maximum Sortino ratio model for each predictor, according to table 1.6. Total absolute factor portfolio weights in this table are allowed to vary over time, although the net position of the portfolio is always zero. The portfolio estimation procedure is described in the text. Shading indicates NBER recession periods.

reached in the month following the first reported cases of COVID-19 in the United States, when the total position adopted by this strategy swells to more than 10 times the average position size held during non-recession months. A similar phenomenon occurs in the months following the collapse of Lehman Brothers. It is these implausibly large bets that drive the remarkable returns observed during recessions for the models using expected returns as predictors.[11]

Thus, while the expected returns models are seen to perform extremely well during crisis periods, the position scaling required to achieve those gains is almost certainly unrealistic as those periods are precisely the times when capital is most scarce, uncertainty high, and short sale constraints most binding. For this reason, the results reported in table 1.6 should be seen as something of a stylized example.

A more realistic portfolio construction scheme would restrict position size to be constant over

[11] As suggested by the difference between the Sharpe and Sortino ratios, a similar phenomenon is also observed in many of the other models estimated using $\widetilde{r}_t$ and $\widetilde{\Delta p}_t$, such as the ridge regression and PCA$_k$ models.

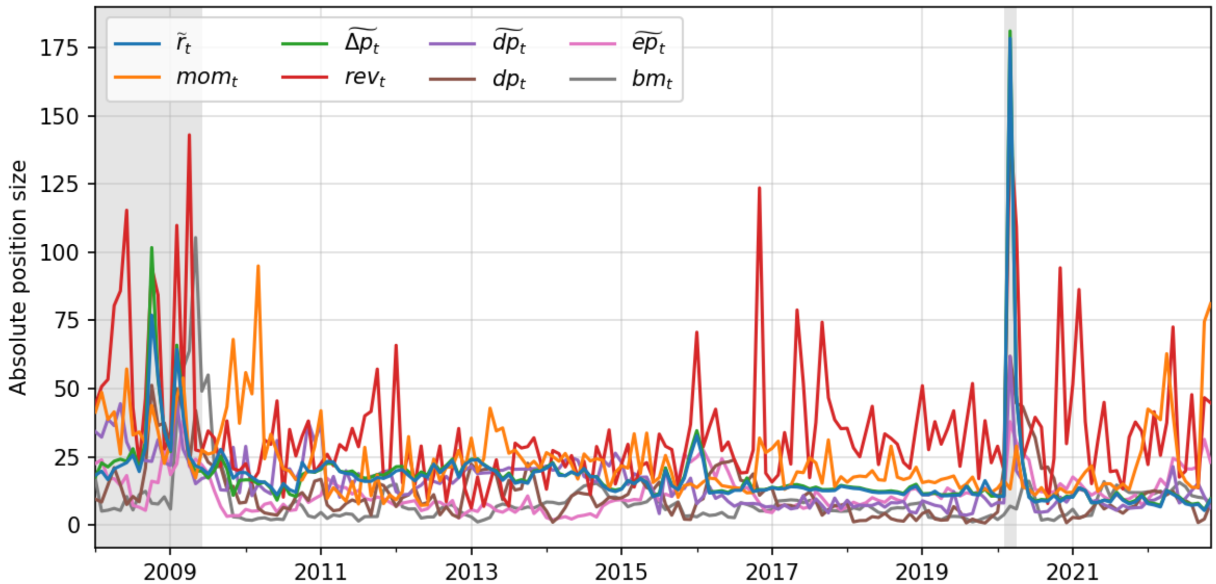


Figure 1.8: Absolute position size over time

Note: This model shows the absolute sum of the portfolio weights for the maximum Sortino ratio model for each predictor, according to table 1.6. Shading indicates NBER recession periods.

time. As all underlying portfolios have equally sized long and short legs, maintaining a consistent portfolio size requires a constraint on the sum of the *absolute* sizes of the weights. These constrained portfolio weights, denoted $\widetilde{w}_{kt}$, are identical to those for the unconstrained model w_{kt}, but normalized such that their absolute values sum to one:

$$\widetilde{w}_{kt} = \frac{w_{kt}}{\sum_{k=1}^{K} |w_{kt}|}$$

Table 1.7 shows the Sharpe and Sortino ratios for these constrained portfolios. Even under this more restrictive portfolio formation rule, the models with the highest Sharpe ratios use expected returns or price growth. The maximum Sharpe ratio achieved using forecasts based on expected returns or price growth has actually increased, relative to the unconstrained portfolio. This increase in Sharpe ratio, however, is achieved through the moderation of the large upward price movements these models exhibit during recession periods. So this gain in Sharpe ratio comes at the cost of lower total returns. This is illustrated in the substantial reduction in the maximum possible Sortino ratio now attainable using those predictors.

Ratio	Model	Predictor							
		$\widetilde{r}_t$	$\widetilde{\Delta p}_t$	$\widetilde{dp}_t$	$\widetilde{dp}_t$	dp_t	mom_t	rev_t	bm_t
	Ridge	0.231	0.233	0.182	0.135	0.188	0.187	0.135	0.136
	PCA_1	0.150	0.146	0.167	0.100	0.136	0.140	0.090	0.152
Sharpe	PCA_k	0.272	0.266	0.222	0.090	0.165	0.153	0.063	0.110
	PLS_1	0.240	0.246	0.105	0.057	0.227	0.181	0.140	0.191
	PLS_k	0.169	0.184	0.079	0.106	0.156	0.133	0.222	0.091
	Ridge	0.408	0.435	0.273	0.198	0.315	0.307	0.272	0.192
	PCA_1	0.220	0.214	0.225	0.154	0.171	0.176	0.131	0.208
Sortino	PCA_k	0.502	0.508	0.321	0.132	0.296	0.200	0.117	0.144
	PLS_1	0.406	0.439	0.159	0.081	0.384	0.275	0.196	0.358
	PLS_k	0.262	0.288	0.107	0.149	0.221	0.189	0.395	0.139

Note: This table shows the out-of-sample Sharpe and Sortino ratios for each combination of model and predictor. Total absolute factor portfolio weights in this table are constrained to be constant over time. Details of the out-of-sample estimation procedure are described in the text. PCA_1 and PLS_1 models are restricted to only using a single factor in all periods. For the PCA_k and PLS_k models, the number of factors employed is chosen by cross-validation, up to a maximum of ten. All ratios are monthly values.

Table 1.7: Out-of-sample Sharpe and Sortino ratios, constrained weights

1.6 Discussion

In this chapter, I have examined the informational content of the cross-section of subjective return expectations, which I measure through analyst price target forecasts. Despite the general neglect of price targets in the academic finance and accounting literatures, I find that the return expectations they imply are at least as useful as many other popular and well-studied cross-sectional return predictors for forecasting returns, and for improving cross-sectional portfolio allocations. In particular, subjective return expectations are observed to be especially informative about near-future winners and losers during times of economic crisis.

Through this examination, I have also considered several of the practical issues involved in the estimation of financial forecasting models when the number of predictors is large, and many of the predictors are highly correlated. While the approaches explored here make no claim to optimality, the results obtained suggest integrating simple regularization techniques commonly employed in the machine learning literature—model averaging and L_2 regularization, in particular—can complement otherwise standard linear factor forecasting models, producing large improvements with

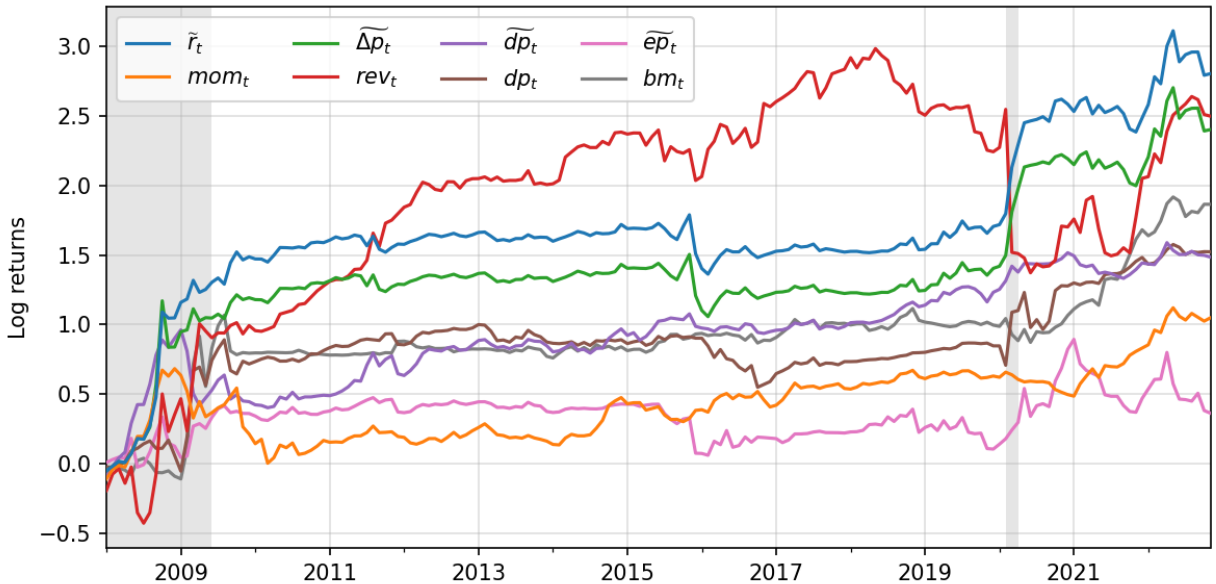


Figure 1.9: Cumulative log returns over time, constrained MV-optimizing portfolio

Note: This figure shows the cumulative log returns for the maximum Sortino ratio model for each predictor, according to table 1.7. The portfolio estimation procedure is described in the text. Shading indicates NBER recession periods.

minimal computing cost.

There is clearly much left to consider, however, in the relationship between financial returns and their subjective expectations. In particular, I do not explore the source of the return predictability I find. Understanding, for example, whether the correlations between subjective expectations and future returns are driven by a genuine understanding of the market's expected compensation for risk on the part of analysts, by behavioral interactions between the biases of forecasters and market participants, or some combination thereof, would have tremendous import for financial theory.

There is also much left to uncover in the data sources I consider. While I examine a broad panel of characteristic portfolios, there are many others that might be considered. I have also taken a very granular view of my data source. New price targets are being issued constantly, with thousands of new forecasts appearing for U.S. equities every week. This rapid arrival rate means that forecasts can, in principle, be measured at a weekly, or potentially daily, frequency. In working with only consensus forecasts, I have also completely neglected the information contained in the higher moments of the distribution of forecasts across individual analysts. I have also not touched upon the substantial data available for non-U.S. firms.

Chapter 2: Expectations Formation with Fat-tailed Processes[1]

[1]Joint work with Adam Rej (CFM) and David Thesmar (MIT, CEPR, NBER). My coauthors and I thank seminar participants at CFM and the Columbia Department of Economics for their constructive feedback.

2.1 Introduction

Expectations formation is a core question in economics. In recent years, a strain of literature in macroeconomics and finance has been collecting empirical regularities using survey data on subjective forecasts. It finds that forecasts largely deviate from the full information model that predominates in economic modelling: forecast errors are biased and predictable using past errors and past revisions. Two types of explanations for this have been put forward. The first one is that the data-generating process (DGP) is simple and known to forecasters, but forecasting rules are irrational but linear, featuring for instance under-reaction (Bouchaud et al., 2019) or overreaction (Bordalo et al., 2019; Bordalo et al., 2018; Afrouzi et al., 2020). The second approach to explaining observed biases is the tenet that the data-generating process is too complex to be known by forecasters. Thus, they use a misspecified model calibrated on the data they observe. This may come from the fact that the DGP is hard to learn (for recent contributions along these lines see Kozlowski, Veldkamp, and Venkateswaran, 2015; Farmer, Nakamura, and Steinsson, 2021), or alternatively from bounded rationality of the forecasters. They can only use simple forecasting rules (Fuster, Laibson, and Mendel, 2010; Gabaix, 2018). In any case, forecast errors are predictable because forecasters use an imperfect model. In this chapter, we find evidence consistent with the second view, i.e. that, facing complex (non-Gaussian) processes, forecasters use simple rules.

We use data on some 63,601 analyst forecasts of corporate revenue growth and their realizations. An advantage of focusing on revenue growth (instead of EPS as the literature typically does) is that revenue is always positive so that growth rate is always well defined. We first show that the relationship between forecast revisions and future forecast error is non-linear, a feature that is not reported in the existing literature. In some settings, revisions linearly and *positively* predict forecast errors, a feature commonly interpreted as evidence of under-reaction (Coibion and Gorodnichenko, 2015). In others, revisions linearly and *negatively* predict forecast errors, which is considered as evidence of overreaction (Bordalo et al., 2019; Bordalo et al., 2018). In our sample, which is much larger than the existing studies, and which focuses on a rather new object, sales growth, we find evidence of both. For intermediate values of revisions, forecasters underreact to news (an increasing relation between revisions and errors). For large values of revisions, forecaster overreact (a decreasing relation between revisions and errors). This non-linearity is robust. It holds in U.S. data and international data. It holds across most industry groups.

The remainder of the chapter is dedicated to explaining this fact. Our framework is based on the simple assumption that forecasters use a linear rule to forecast sales growth, but that this rule is misspecified because the true DGP is more complex. Taking inspiration from the literature on firm size distribution (in particular, Axtell, 2001; Bottazzi and Secchi, 2006), we posit that sales growth distribution may be modelled by the sum of a low-frequency and high-frequency shock. The low frequency shock is Gaussian, while the high-frequency shock is non-Gaussian. It may have a very large (positive or negative) realizations. With such a model, the optimal forecast of future growth, conditional on current growth, is non-linear. A perfectly rational forecaster anticipates more reversion to the mean when realizations are extreme and more persistence when realizations are intermediate. We assume, however, that agents stick to a linear rule to make their forecasts. The fact that agents use a misspecified model may be grounded in bounded rationality (i.e., agents use a simple rule even if the process is complex, as in Fuster, Laibson, and Mendel, 2010) or the difficulty of learning about complex processes (shocks with multiple frequencies are hard to learn Farmer, Nakamura, and Steinsson, 2021; shocks with fat tails also Kozlowski, Veldkamp, and Venkateswaran, 2015).

Combined, these two assumptions (linear forecasting rule but short-term non-Gaussian shocks) are enough to generate the non-linear relation between forecast errors and past revisions that observe empirically. The mechanism is intuitive. When revisions are large, the rational forecaster should anticipate mean reversion, but the linear forecaster won't. She overreacts to big positive (or negative) news. When fitting her forecasting rule to the data, she does, however, take this overreaction into account, and optimally attenuates the sensitivity of her forecast to recent observations in the bulk of the distribution. As a result, she underreacts to news of lesser significance.

We then qualitatively test four additional predictions of the model. We start with two natural predictions of the data-generating process. The first such prediction is that the distribution of sales growths has fat tails, a fact that holds strongly in the data (and previously shown by Bottazzi and Secchi, 2006). In particular, we check that this fact is not driven by an alternative model of firm dynamics, where firms have heterogeneous volatility, but Gaussian dynamics. In such a setting, large growth shocks could be generated by the subset of firms who are more volatile than average (Wyart and Bouchaud, 2003). We thus rescale sales growths by estimates of firm-level standard deviation and find that the resulting distribution still has very fat tails, suggesting that growth shocks occur within firms, not across firms.

The second prediction from our DGP is that, conditional on past growth, future growth should follow a S-shaped pattern as discussed above. We show that this holds in the data, whether we normalize sales growth by firm-level standard deviation or not.

The third prediction is on forecast errors. A natural prediction of our forecasting model is that the autocorrelation of forecast errors should have the same non-linear relation as the relation between errors and lagged revision. In our model, where the forecasting rule is linear, they are the same. Large past errors are equivalent to big shocks and therefore transient ones: This leads to overreaction, as in the error-revision relation. We find that this pattern holds in the data: forecast error are positively correlated for intermediate values and negatively for large absolute values.

Our fourth and last prediction is on stock returns. Assuming risk-neutral pricing and that equity cash-flows follow a dynamic similar to revenues, it is easy to show that our model predicts that the autocorrelation of returns should have a shape similar to the autocorrelation of forecast errors. For intermediate values of past returns, momentum should dominate, but for extreme values of returns, stock returns should mean revert. We find this pattern to hold in the data. Our findings line up with recent research from Schmidhuber, 2021, who also finds evidence of momentum for "normal past returns" and mean-reversion for extreme values of returns. We conclude from this analysis that the risk-adjusted performance of momentum strategies would be considerably improved by excluding stocks whose past returns have been large in absolute value.

This chapter contributes to the recent empirical literature on expectations formation. Most papers in this space focus on linear and Gaussian data-generating processes. Forecasting rules may, or may not, be optimal, but are in general linear, so that the relationship between forecast errors and past revisions (or past errors) is also linear. This chapter emphasizes the non-linearity of such a relation, and as a result, suggests an different modelling approach for the data-generating process to account for this non-linearity. We emphasize non-Gaussian dynamics in firms' growth (as Kozlowski, Veldkamp, and Venkateswaran, 2015, have done in a different setting and in their case with a focus on Bayesian learning).

In doing this we also connect the expectations formation literature with the empirical literature on firm dynamics, which has since Axtell, 2001 emphasized the omnipresence of power laws in the distribution of firms sizes (see Gabaix, 2009, for a survey of power laws in economics). That sales *growths* (rather than log sales) have fat tails is a less well-known fact, although it was first uncovered by Bottazzi and Secchi, 2006.

Last, our assumption that forecasters use a simple, linear, forecasting rule that is misspecified is inspired by the literature on bounded rationality, which assumes economic agents have a propensity to use oversimplified models to minimize computation costs (Fuster, Laibson, and Mendel, 2010; Fuster, Hebert, and Laibson, 2012; Gabaix, 2018). Such models are correct on average, they are fitted on available data, but their misspecification gives rise to predictability in forecast errors.

Section 2.2 describes the data we use: publicly available data on analyst forecasts (IBES) and confidential data on international stock returns from CFM. Section 2.3 documents the main fact: future errors are a S-shaped function of past revision. Section 2.4.1 lays out the simple framework that we build in order to explain this novel pattern. Section 2.5 tests four additional predictions from this model. Section 2.6 concludes.

2.2 Data

2.2.1 Analyst forecast data

This chapter focuses on firm revenue (sales) forecasts made by analysts. Analyst forecasts come from IBES Adjusted Summary Statistics files, which are available both in the U.S. and internationally. Summary statistics files contains "current" estimates as of the third Wednesday of each month. While Earnings per Share forecasts have received greater attention in the literature, sales forecasts are, in fact, better populated in the data than EPS forecasts in recent years. Another advantage of revenue forecasts is that they are never negative, so that we can easily calculate sales growth. A downside of EPS is that it is frequently negative or small rendering the calculation of EPS growth forecast impractical. Thus, the literature on EPS forecasts studies a variable that is, in essence, non-stationary (typically, future EPS normalized by current stock price).

For each firm i and each year t, we denote sales by R_{it}, and $F_t R_{it+1}$ the forecast made in year t for the future realization of R_{it+1}. We compute $F_t R_{it+1}$ as the consensus three months after the end of fiscal year t (i.e. nine months prior to the end of fiscal year $t+1$) to ensure that sales results for fiscal year t are available when the forecast for $t+1$ is formed. Similarly, the two-year ahead forecast $F_{t-1}R_{it+1}$ is measured three months after the end of fiscal year $t-1$. Finally, we retrieve the realization of R_{it+1} from the IBES actual files, which is designed to recover the realization of the quantity actually forecast by analysts.

In this chapter, we focus on log sales growth and forecast of log sales growth. We define $g_{it+1} = \log R_{it+1} - \log R_{it}$ the log-growth of this quantity. The growth forecast is defined as $F_t g_{it+1} = \log F_t R_{it+1} - \log R_{it}$ for the one-year ahead growth forecast, and $F_{t-1} g_{it+1} = \log F_{t-1} R_{it+1} - \log F_{t-1} R_{it}$ for the two-year ahead forecast of annual growth.

Finally, in the spirit of the expectations formation literature (Coibion and Gorodnichenko, 2015; Bouchaud et al., 2019), we construct two empirical variables: the forecast error $ERR_{it+1} = g_{it+1} - F_t g_{it+1}$ and the forecast *revision* $R_t g_{it+1} = F_t g_{it+1} - F_{t-1} g_{it+1}$. These two variables will be the main focus of our analysis.

To ensure forecast quality and improve sample consistency when we examine returns, restrict our analysis of forecasts to firms that belong to one of the major global stock indexes.[2]. Further, we restrict ourselves to firm-year observations for which both the forecast error ERR_{it+1} and the revision $R_t g_{it+1}$ are available. We give more details about the number of observations and the start

[2]The list of stock markets used consists of: AEX, AS5, CAC, DAX, HSC, HSI, IBE, IND, KOS, MID, NDX, NIF, NKY, OMX, SMI, SPT, RAY, SX5, TOP, TPX, TWY, UKX

date in Table 2.1.

2.2.2 International Data on Stock Returns

In examining returns we restrict our sample to equities included in a major national index. We rely on proprietary return data purchased and maintained by CFM. The start of data availability differs by index and is shown in Table 2.2. For all indexes data has been obtained through January, 2022. Each observation is a ticker-month, and returns are log returns.

Index	Total	2000	2005	2010	2015	2020
AEX	533	0	32	19	30	28
AS5	3228	48	122	167	196	161
CAC	921	0	40	45	49	48
DAX	680	0	29	38	37	35
HSC	972	7	24	41	75	74
HSI	572	15	24	28	29	29
IBE	715	0	35	37	40	35
IND	746	1	38	38	41	39
KOS	1540	34	29	30	101	124
MID	13016	10	586	818	782	646
NDX	1174	1	47	67	72	61
NIF	1037	13	24	47	66	64
NKY	4959	207	206	226	224	233
OMX	605	19	26	31	32	31
RAY	15923	4	525	995	1057	881
SMI	479	8	21	21	27	23
SPT	998	0	34	56	67	63
SX5	215	0	10	11	13	11
TOP	493	0	18	14	40	32
TPX	10836	372	486	531	504	574
TWY	1314	13	40	71	80	74
UKX	2645	82	110	131	142	119

Note: This table shows the total observation count, by exchange, for our study of sales growth forecasts, as well as observation counts for select years. As not all sample years are shown, the counts for individual years do not sum up to the total in the leftmost column.

Table 2.1: Sample size by exchange, annual sales growth

2.3 Motivating Facts

In this section we describe new evidence on expectations formation and document a strong non-linearity in the link between forecast error and revisions.

Since (Coibion and Gorodnichenko, 2015) many papers in the expectations formation literature estimate the following linear relationship between forecast errors and revision:

$$ERR_{it+1} = \alpha + \beta R_t g_{it+1} + \epsilon_{it+1} \tag{2.1}$$

which is intuitive to interpret. Full information rationality predicts $\beta = 0$ for consensus forecasts (Coibion and Gorodnichenko, 2015). Plain rationality predicts $\beta = 0$ for individual forecasts. $\beta > 0$ is typically interpreted as evidence of information frictions (Coibion and Gorodnichenko, 2015), or, if run at the forecaster level, plain under-reaction (Bouchaud et al., 2019, study EPS forecasts; Ma et al., 2020, study the revenue forecasts of managers). In contrast, $\beta < 0$ is interpreted as evidence of overreaction (Bordalo et al., 2019, study long-term EPS growth forecasts; Bordalo et al., 2019 focus on macroeconomic expectations). All these papers restrict their analyses to linear functional forms, as in equation (2.1).

In this section we show that for revenue growth forecasts this relationship is actually non-linear. In Figure 2.1, we represent the relationship in a non-parametric way through a binned scatter plot where the x-axis are the revisions $R_t g_{it+1}$ and the y-axis is the average forecast error FE_{it+1}. Each black dot corresponds to a centile of the distribution of revisions, with the x coordinate being the average revision in this bin and the y coordinate being the average forecast error. The grey shaded area shows a bootstrapped 95% confidence interval. The blue line shows the predicted error from a local polynomial regression (or LOESS) model estimated at the center of each percentile of lagged revision. The kernel for this local regression model is Gaussian with the bandwidth set equal to the average of the distances between the centers of the 1st and 2nd, and the 99th and 100th, percentiles.

For revisions of relatively small to moderate magnitude we find that errors are increasing in revision. Thus, forecasters are *under-reacting* in response to moderately-sized news shocks. This

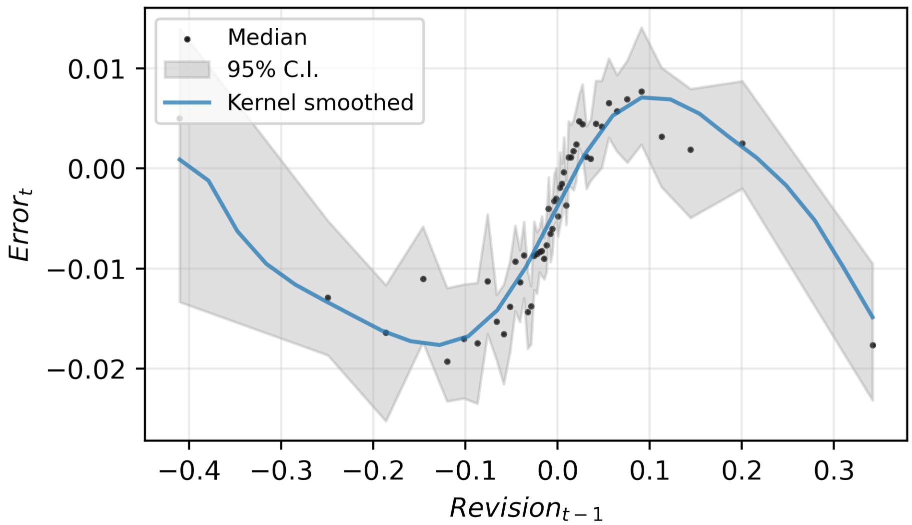


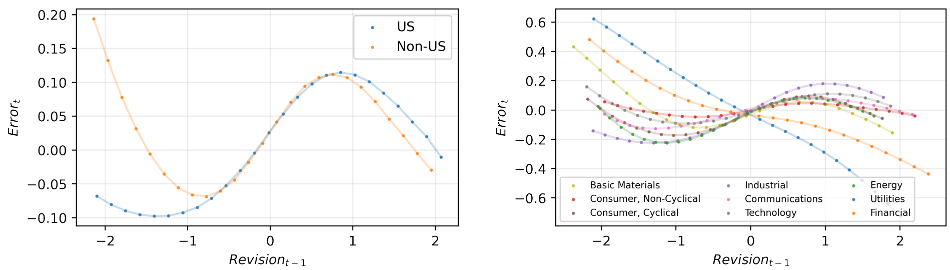
Figure 2.1: Revenue forecast error conditional on past revision

Note: In this figure we use our international sample of firm revenue expectations to report the binned scatter plot of future log forecast errors $g_{t+1} - F_t g_{t+1}$ as a function of past revision $F_t g_{t+1} - F_{t-1} g_{t+1}$. The blue line is a local polynomial approximation, centered in the middle of each bin. The shaded area depicts a 95% confidence interval based on 1,000 bootstrap samples.

consistent with evidence from Bouchaud et al., 2019 on EPS forecasts in United states. Ma et al., 2020 find similar evidence on revenue forecasts from managers' expectations in the U.S. (using guidance data) and Italy (using a survey from the Bank of Italy). Their samples are, however, much smaller than ours (a few 10,000 observations at most), which precludes observing the tails of the distribution of revisions.

The key difference is in the tails of the distribution of revisions, for which this relationship is reversed. In the face of exceptionally bad news, forecasters are *over-reacting*: a larger, positive revision leads to more negative surprises. A similar non-linearity is marginally observable in U.S. EPS forecasts in Bouchaud et al., 2019, but the S shape in not complete there.

We then explore the robustness of this relationship across sub-categories in Figure 2.2. This figure has two panels: one that splits between U.S. and non-U.S. firms (Panel A) and one that splits the sample into industries (Panel B). In both cases we only show the prediction from flexible polynomial approximation. In both subcategories the S-shaped function emerges. In particular, it is visible both in US and non-US firms, although more pronounced among U.S. firms.



Panel A: US v RoW

Panel B: By Industry

Figure 2.2: Error conditional on past revision, by sub-sample

Note: In this figure we use our international sample of firm revenue expectations to report mean future log forecast errors $g_{t+1} - F_t g_{t+1}$ as a function of past revision $F_t g_{t+1} - F_{t-1} g_{t+1}$. In Panel A, we split the sample between U.S. and non U.S. observations. In Panel B, we split the sample into 1 digit GICS industries.

Overall, the evidence on log forecast errors and revisions points towards a different treatment of large v. smaller shocks. Such evidence is hard to square with established models of expectations formations, which feature linear DGPs (typically, AR1 models) and linear expectations models. In what follows, we set up a simple model that features extreme (i.e. non-Gaussian) shocks and linear expectations formation in order to captures the above non-linearity.

2.4 Model

In this section we develop a parsimonious model that features extreme shocks and linear expectations in order to capture the non-linear behavior of expectation errors of Figure 2.1.

2.4.1 Modeling Sales Growth

The first piece of the model is the data-generating process. We will omit the firm index i for clarity's sake and assume that log sales growth, g_{it+1}, evolves according to:

$$g_{t+1} = \underline{g}_{t+1} + \epsilon_{t+1} \tag{2.2}$$

$$\underline{g}_{t+1} = \overline{g} + \phi(\underline{g}_t - \overline{g}) + u_{t+1} \tag{2.3}$$

where $\underline{g}_{t+1}$ is the unobservable latent state that follows the classic linear-Gaussian AR1 dynamics. The key difference with most existing models of expectations formation is that ϵ_{t+1} follows a probability distribution density that has heavy tails. Because it fits the data quite well (as we document below), we assume that ϵ_{t+1} follows a Student's t distribution with ν degrees of freedom. Thus:

$$\epsilon_{t+1} \sim \text{Student-t}(0, 1, \nu)$$

$$u_{t+1} \sim \text{Normal}(0, 1)$$

Note that, although we analyze the cross-section of firms, we assume a single process for all firms. In this chapter, we do not explore the consequences of firm heterogeneity for forecasting biases. For instance, such biases could arise from forecasters using one single forecasting model for firms following different processes. We believe such an avenue is interesting, but beyond the scope of this chapter, which focuses on one single deviation from the classical model, i.e. that temporary shocks have fat tails. In order to bring the data closer to the model however, we will conduct all of our analysis with "normalized growth" data, thereby ensuring that all firms have the same growth volatility. We discuss this adjustment extensively in Section 2.5.

In our simple model the conditional expectation $E\left(g_{it+1}|g_{it}\right)$ is non-linear. We show this numerically in Figure 2.3. For different values of ν, we numerically simulate the process and compute the conditional expectation $E\left(g_{it+1}|g_{it}\right)$ on simulated data. As shown in Figure 2.3 this relationship is indeed quite linear in the body of the distribution, but experiences "reversals" in the tails. While not visible in Figure 2.3, all finite values of ν leads to such reversals in the tail, but as ν gets larger (and ϵ is closer to being Gaussian) they get pushed out farther into the tails and are very sharp and localized.

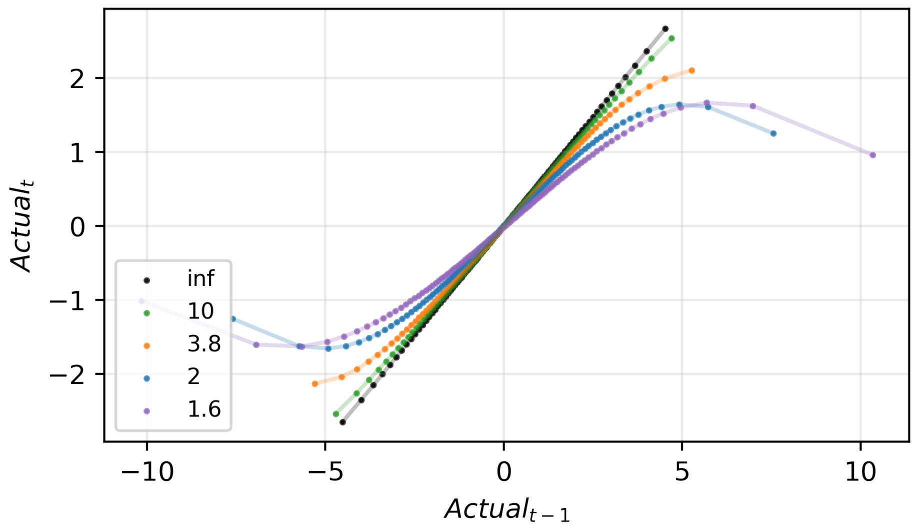


Figure 2.3: Growth conditional on past growth, simulated model

Note: We simulate the model over T periods with u following a normal distribution $N(0,1)$, and ϵ following Student-t$(0,1,\nu)$. We then show local polynomial regressions of g_t on g_{t-1} estimated at the center of each percentile of lagged realization g_{t-1}. We explore values of ν from 1.6 (fat tailed) to ∞ (Gaussian).

The economic intuition is simple. Since the underlying state variable is Gaussian, extreme negative or positive realizations are more likely to come from the transitory process ϵ than the persistent one $\underline{g}$, since it features more extreme shocks. As a result, a large sales growth realization today is unlikely to translate into future large sales growth tomorrow: This suggests the presence of "reversals" in the tails, as we see in Figure 2.3.

The above process has several predictions about the distribution of growth rates, one of them being that the cross-sectional distribution of growth rates should have fat tails. We will explore these predictions in Section 2.5.

2.4.2 Expectations Formation

The second building block of the model is the formation of expectations. Our core assumption is that forecasters fail to perceive the non-linearity of true expectations $E\left(g_{it+1}|g_{it}\right)$ and use a linear rule. This assumption is based on the idea that economic agents use simplified, "sparse", model of reality to formulate expectations (Fuster, Hebert, and Laibson, 2012; Gabaix, 2018). Agents assume g_{it} follows a linear AR(p) model, estimate it on data and use this model to form forecasts. One advantage of this representation is that the term structure of forecasts is naturally defined, as agents calculate mathematical expectations under the AR(p) model. Hence, we assume that the forecaster believes growth follows the following AR(p) model:

$$g_{t+1} = \underline{g} + \sum_{s=0}^{p-1} \beta_k \left(g_{t-k} - \underline{g}\right) + u_{t+1}$$

We denote the subjective expectation operator by $F_t g_{t+k} \equiv EL(g_{t+k}|\underline{g_t})$.

We assume that this prior is dogmatic. The forecaster is willing to re-estimate the model's parameters as new data comes in, but does not explore models outside of the AR(p) set-up. As a result, the agent does not really formulate rational expectations since she does not estimate the right DGP, as in Fuster, Laibson, and Mendel, 2010. One foundation for such dogmatism is that learning is extremely slow in non-Gaussian, non-linear environments, so that it takes many periods to modify the prior about the model (in recent literature, see Kozlowski, Veldkamp, and Venkateswaran, 2015 and Farmer, Nakamura, and Steinsson, 2021).

Thus the agent estimates the parameters of the misspecified model using OLS on expanding windows – using all information until date t. Let $\widehat{\underline{g}}$ and $\widehat{\beta_k}$ be these estimates. The one-period ahead forecast and the revision are given by:

$$F_t g_{t+1} = \widehat{\underline{g}} + \sum_{s=0}^{p-1} \widehat{\beta_k} \left(g_{t-k} - \widehat{\underline{g}}\right)$$

2.4.3 Predictions of the Model: Errors on Revisions

We now check that our model indeed generates the non-linear relation between revenue forecast errors and revenue forecast revisions shown in Figure 2.1.

In Figure 2.4, we report results from simulations, assuming that forecasts are based on a fitted AR(2) model. We vary the thickness of the tail of the temporary shock ϵ, which is governed by ν. $\nu = +\infty$ corresponds to a normal distribution, while $\nu = 1.6$ is the thickest tail possible.

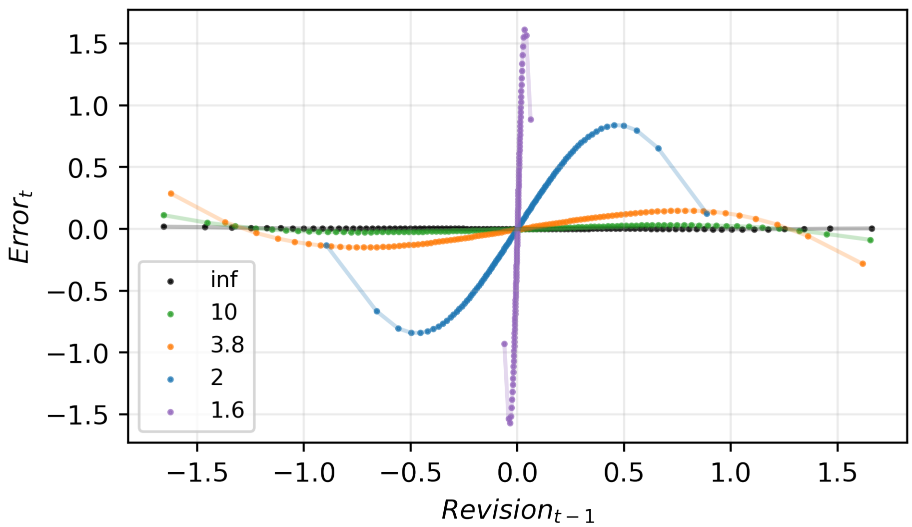


Figure 2.4: Forecast error conditional on past revision, simulated model

Note: We simulate the model over T periods with u following a normal distribution $N(0, 1)$, and ϵ following Student-t$(0, 1, \nu)$. We then show local polynomial regressions of error $ERR_{t+1} = g_{t+1} - F_t g_{t+1}$ on revision $R_t g_{t+1} = F_t g_{t+1} - F_{t-1} g_{t+1}$ estimated at the center of each percentile of revision. We explore values of ν from 1.6 (fat tailed) to ∞ (Gaussian). Forecasters are assumed to employee an AR(3) model when predicting dividend growth rates.

Figure 2.4 shows that, as long as the temporary shock has sufficiently fat tails, the linear expectations model generates predictable forecast error that display a non linear pattern similar to Figure 2.1. This is quite intuitive. As shown previously, the true conditional expectation is non-linear (see Figure 2.3). When the past realization of revenue growth is large, it is likely that it was driven by the temporary fat-tailed process. As a result, the rational forecaster would expect some mean-reversion, but the linear forecaster does not. This creates overreaction to large shocks. In contrast, when past realizations are moderate, there is underreaction. This comes from the fact that the linear forecaster is on average rational: She fits a linear relation on the S-shaped data of Figure 2.3. The slope of forecasts for smaller realizations incorporates some of the overreaction in the tails.

To gain further insights in Figure 2.3 we vary the fatness of the tail ν. The less thick-tailed the innovation process, the less predictable errors are. When $\nu = +\infty$, the temporary shock ϵ is Gaussian and forecast errors are very close to zero for all lagged realizations (the black dots line up on the x-axis). This is because in this case the linear AR2 forecasting rule is nearly rational. Indeed, in this case, the rational expectation is a Kalman filter:

$$K_t g_{t+h} = \underline{g} + \phi^h G \sum_{s=0}^{+\infty} (1-G)^s \left(g_{t-s} - \underline{g} \right)$$

where G is the Kalman "gain". The AR2 process is close enough to the above equation that forecast errors are nearly zero in our simulations.

The bottom line of this analysis is that the non-linear structure of expectations errors can easily arise when forecasters use linear models when the data generating process has temporary shocks that have fat tails. Indeed in this case, the optimal forecasting rule is non-linear, even though the process is itself linear.

2.4.4 An Additional Prediction: Error on Lagged Error

The empirical expectations literature also investigates a different moment: The autocorrelation of expectation errors (for instance, Ma et al., 2020 and Farmer, Nakamura, and Steinsson, 2021 among many others).

In our model the autocorrelation of errors is equivalent to the error-revision coefficient. This happens because revisions are directly proportional to current forecast errors:

$$\underbrace{F_t g_{t+1} - F_{t-1} g_{t+1}}_{\equiv R_t g_{t+1}} = \widehat{\beta}_0 \cdot (g_t - F_{t-1} g_t) \tag{2.4}$$

which means that a positive surprise translates into a positive revision about future growth. The fact that the prior is linear makes this relationship linear, whatever the number of lags p.

As a result, we expect the non-linear relation between errors and lagged revision of Figure 2.1 to also hold between error and lagged *errors*. We test this additional prediction in Section 2.5.

2.4.5 Model Prediction on Returns: Building Intuition

We also derive predictions on stock returns. Our simple model, as we will see, predicts that momentum occurs for intermediate returns and mean-reversion occurs for extreme returns.

In the spirit of Bouchaud et al., 2019, we assume stock prices are given by:

$$P_t = \sum_{s=1}^{\infty} \frac{F_t D_{t+s}}{(1+r)^s} \tag{2.5}$$

where $F_t D_{t+s}$ is based on the forecasting rule described above in Section 2.4.2. Hence, the stock is priced by investors who form expectations based on a linear AR2 fitted on past realizations, while we assume dividends to follow the process described in Section 2.4.1. We also assume for simplicity that investors are risk-neutral, so that the discount rate is fixed at r.

In this very simple asset pricing model we expect returns to be a non-linear function of past returns, similar to what we documented for the error-revision relation in Figure 2.1. Before we discuss simulation results and economic intuition, it is worth showing the algebra. The standard first order Campbell-Shiller approximation is written:

$$r_{t+1} - F_t r_{t+1} \approx (F_{t+1} - F_t) \sum_{s=0}^{\infty} \rho^s g_{t+1+s} - (F_{t+1} - F_t) \sum_{s=1}^{\infty} \rho^s r_{t+1+s}$$

where we denote log dividend growth as g with a slight abuse of notation (g stands for log revenue growth in the rest of the chapter). Equation (2.5) assumes constant expected returns $F_t r_{t+k} = r$

(investors may be biased but are risk neutral), so that the CS decomposition simplifies into:

$$
\begin{aligned}
r_{t+1} - r &= (F_{t+1} - F_t) \sum_{s=0}^{\infty} \rho^s g_{t+1+s} \\
&= g_{t+1} - F_t g_{t+1} + \rho \left(F_{t+1} - F_t\right) \sum_{s=0}^{\infty} \rho^s g_{t+2+s}
\end{aligned}
$$

It then remains to describe the infinite sum of discounted dividend growth. In this chapter, we assume that forecasters (mistakenly) estimate dividend growth as an AR(p) process:

$$
g_t - \underline{g} = \sum_{s=1}^{p} \beta_s \left(g_{t-s} - \underline{g}\right) + \epsilon_t
$$

We can then stack the estimated AR(p) coefficients $\beta_1, \ldots, \beta_p$ into "companion" form:

$$
\begin{bmatrix} g_t - \underline{g} \\ g_{t-1} - \underline{g} \\ \vdots \\ g_{t-p} - \underline{g} \end{bmatrix} = \begin{bmatrix} \beta_1 & \cdots & \beta_{p-1} & \beta_p \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & 0 & 0 \\ 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} g_{t-1} - \underline{g} \\ g_{t-2} - \underline{g} \\ \vdots \\ g_{t-p-1} - \underline{g} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}
$$

or more compactly:

$$
\mathcal{G}_t = \mathbf{B}\mathcal{G}_{t-1} + \epsilon_t
$$

Such that time t forecasts for $g_{t+s} - \underline{g}$ can be written:

$$
F_t(g_{t+s} - \underline{g}) = \mathbf{e}_1' \mathbf{B}^s \mathcal{G}_t
$$

where $\mathbf{e}_1$ is a "selector" vector picking out the first element of the following vector. The infinite sum of discounted forecast dividend growth is given by:

$$\begin{aligned} F_t \sum_{s=0}^{\infty} \rho^s (g_{t+1+s} - \underline{g}) &= \sum_{s=0}^{\infty} \rho^s \mathbf{e}_1' \mathbf{B}^{s+1} \mathcal{G}_t \\ &= \mathbf{e}_1' \mathbf{B} (\mathbf{I} - \mathbf{B})^{-1} \mathcal{G}_t \end{aligned}$$

We then plug this formula into the CS decomposition and obtain:

$$\begin{aligned} r_{t+1} - r &= \mathbf{e}_1' \left(\mathcal{G}_{t+1} - \mathbf{B}\mathcal{G}_t\right) + \rho \mathbf{e}_1' \mathbf{B}(\mathbf{I} - \mathbf{B})^{-1} \left(\mathcal{G}_{t+1} - \mathbf{B}\mathcal{G}_t\right) && (2.6) \\ &= \mathbf{e}_1' \left(\mathbf{I} + \rho \mathbf{B}(\mathbf{I} - \mathbf{B})^{-1}\right) \underbrace{\left(\mathcal{G}_{t+1} - \mathbf{B}\mathcal{G}_t\right)}_{=ERR_{t+1}\mathcal{G}_{t+1}} && (2.7) \end{aligned}$$

The above expression shows that under the AR(p) assumption, returns are a linear function of past forecast errors ($ERR_{t+1}\mathcal{G}_{t+1}$ is the vector of past p forecast error). In this simple asset-pricing model, returns are only predictable if dividend growth forecast errors are predictable. Under rational expectations (i.e. if the true DGP for dividends is an AR(p)), they are not. But if dividends are driven by a thick-tailed state variable, the true DGP is far from AR(p) as we have documented. Thus, expected forecast errors are non-linear functions of past errors, and the same should hold for returns and past returns. So our model predicts that returns should be a non-linear function of past returns, in other words, momentum should only be present for intermediate values of past returns.

2.4.6 Model Prediction on Returns: Simulations

In order to check that this prediction also holds without CS approximation, we proceed to simulate our model. On simulated data, we build returns as $R_{t+1} = (P_{t+1} + D_t - P_t)/P_t$. We then plot average future returns by bins of past returns in Figure 2.5. In this very simple asset pricing model, returns are predictable as soon as the dividend process

The intuition is the same one as before. Very high past returns likely emerge from surprises due to a large thick-tailed temporary shock. Linear forecasters overreact: large past return events are likely to be situations where dividend realization a one-off boon. As a result, linear forecasters overestimate the level of future dividends: The stock price rises too much and future returns are lower. Intermediate past returns, however, are likely generated by standard shocks. There, the linear forecaster under-reacts to small dividend news and the price does not respond enough to these news. Future returns are thus positively correlated with past returns for "smaller" absolute values of returns. Note that when $\nu = +\infty$, temporary shocks are Gaussian and, as discussed previously, linear expectations are quasi optimal – a true Kalman filter would be perfectly rational – and past returns do *not* predict future returns.

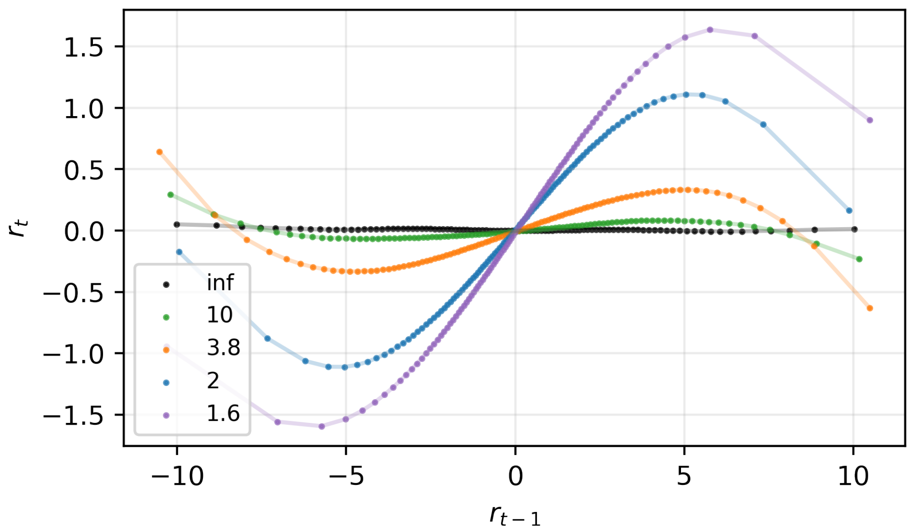


Figure 2.5: Returns conditional on past returns, simulated model

Note: We simulate the model over many periods with u following a normal distribution $N(0, 1)$, and ϵ following Student-t$(0, 1, \nu)$. We then show local polynomial regressions of R_{t+1} on R_t estimated at the center of each percentile of lagged realization $R_t g_{t+1}$. We explore values of ν from 1.6 (fat tailed) to ∞ (Gaussian). Forecasters are assumed to employee an AR(3) model when predicting dividend growth rates.

2.5 Testing the Model's Predictions

2.5.1 Predictions of Growth Rate Dynamics

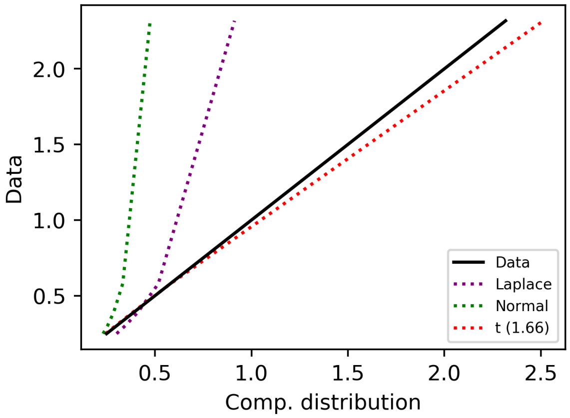
In this section we discuss two key predictions of our data-generating process (2.2)-(2.3). The first one is that the distribution of firms *revenue growths* should have fat tails. The second one is that the conditional expectation of g_{t+1} on g_t should be non-linear as in Figure 2.3.

First, a key prediction of our DGP (2.2)-(2.3) is that the distribution of firm size growth has fat tails. Many variables relevant for finance and economics are not normally distributed Gabaix, 2009. It is for instance well-known that the distribution of firm sizes follows a Zipf law (Axtell, 2001). Less well-known is also the fact that the distribution of firm *growth rates* has fat tails. Bottazzi and Secchi, 2006 show that the distribution of Compustat firms follows a Laplace distribution. Here, we provide similar evidence from our sample.

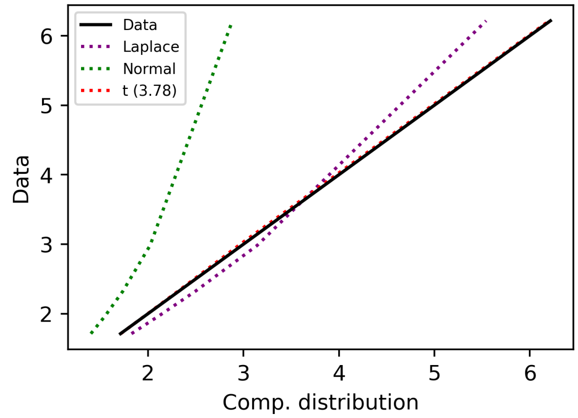
In Panel (a) of Figure 2.6 we show the Q-Q plot of the sales growth distribution in our sample, along some other textbook distributions. This Q-Q plot focuses on observations above the 90^{th} percentile distribution of $|g_{it}|$, the *absolute* value of sales growth (so both negative and positive shocks). Each point of this chart corresponds to one quantile of the data distribution. For a given quantile q and a c.d.f. F, the y coordinate of the point is the average value of absolute growth at quantile q of the data distribution. Since we focus on the top 10% of absolute sales growth, this number is positive (so the y axis does not start at zero). The x coordinate of that point is the average value of the *same quantile* of the chosen distribution F, or $F^{-1}(q)$. The closer is F to the data distribution, the more the chart will look like a 45 degree line (the black line on the Figure).

Looking at Panel (a) of Figure 2.6 it is clear that the distribution of growth rates is very different from normal in the tail. The green line increases faster than the 45 degree line meaning that the sales growth distribution has much heavier tails than normal distribution. The best fit is obtained by fitting a Student distribution.

Our model crucially assumes that this distribution comes from temporary shocks occurring *within* firms. Wyart and Bouchaud, 2003 suggest an alternative explanation for such thick tails:



(a) Un-normalized



(b) Normalized by MAD_i

Figure 2.6: Revenue growth tail distribution

Note: This figures displays Q-Q plots log sales growth vs. some textbook distributions (Laplace, Normal and Student). Panel (a) shows the tail above the 90$^{\text{th}}$-percentile of the distribution of $|g_{it}|$; Panel (b) of a normalized versaion of the same variable $|g_{it} - \overline{g}_i|/MAD_i$. On the x-axis, we report the value of the quantile $(F^{-1}(q))$ in the comparison distribution. On the y-axis, we report the value of the same quantile for the data distribution $(F_{\text{data}}^{-1}(q))$. By construction, the "data" line is the 45 degree line.

sales growth has a normal distribution at the firm-level, but that the standard deviation of this process varies across firms. In this case, extreme growth rates would typically occur among firms that have very volatile growth rates (for instance, smaller firms). This alternative interpretation does not explain our findings, but it is worthwhile to analyze its validity.

In order to do that we normalize growth rates by a measure of firm-level "volatility". To do this we compute the mean absolute deviation of log sales for each firm. This measure of volatility has the advantage of being more immune to fat tails in the growth distribution (since variance may not exist in such cases). For firm i, we thus compute:

$$MAD_i = \frac{1}{T_i} \sum_{t=0}^{T_i} |g_{it} - \overline{X}_i|$$

where T_i is the number of observations for the firm, and $\overline{X}_i$ is the average sales growth at the firm level.

In Panel (b) of Figure 2.6 we show a Q-Q plot of the distribution of normalized growth rates $\frac{|g_{it} - \overline{g}_i|}{MAD_i}$. If heavy tails are driven by firms with larger growth variance, this adjustment should significantly reduce the fat tails of the data. The Q-Q plot shows that the distribution is still strongly non-normal, though now the fit of the Laplace distribution is much better (consistent with Bottazzi and Secchi, 2006) and the one of the student distribution is nearly perfect.

From this analysis we draw the conclusion that sales growth distribution has heavy tails and that the conditional expectation $E\left(g_{it+1}|g_{it}\right)$ is non-linear. We will postulate below a model of growth dynamics that fits these two facts, and show how it can explain the non-linear relation between revisions and errors.

From now on we report results using the above normalization by the mean absolute distance. This allows to account for fat-tails effects stemming from heterogeneous growth variance. Also quite importantly our model (2.2)-(2.3) assumes homoskedasticity, so it is important to rescale the data so that they have the same property.

A second key prediction of our Data is that the conditional expectation $E_t\left(g_{t+1}|g_t\right)$ should be non-linear, as shown in Figure 2.3. As mentioned previously the intuition is that large revision presumably come from large shocks to firm sales growth. Since in our model large shock are transitory, the rational forecaster should not expect that large shocks should persist going forward. Smaller shocks are, however, much more likely to stem from the permanent component of revenue, and therefore the rational forecaster should expect them to persist. We now check whether this relationship holds in the data.

In Figure 2.7 we construct a binned scatter plot of sales growth against lagged sales growth. To make sure all firms have the same growth volatility (as in the model), we normalize growth by our estimate of the firm level standard deviation MAD_i. Each black dot on this figure represents a centile of the distribution of lagged log growth of sales. The x axis shows the average lagged growth and the y axis measures the average current growth. This chart shows that the relationship between current and lagged growth is far from being linear and looks like the S curve shown in Figure 2.1. For intermediate levels of growth (between 0 and 1 standard deviation), past growth translates into higher future growth, with a coefficient of about 0.3. The relationship does, however, become much flatter for high growth (with a slightly negative slope in the tail). For negative growth

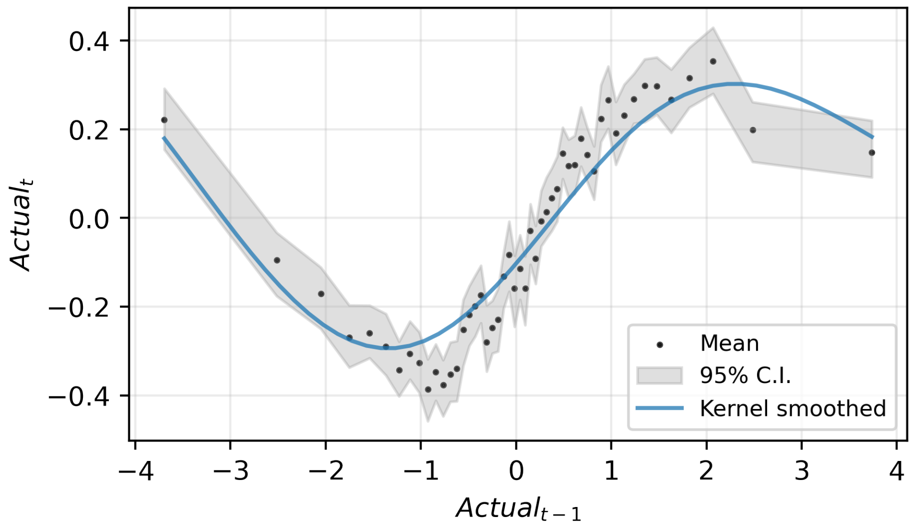


Figure 2.7: Revenue growth conditional on past growth (normalized)

Note: This figure depicts a binned scatter plot of future growth conditional on past growth. The data has been normalized such that each firm has zero mean and unit MAD. Each bin corresponds to a centile of the past growth distribution. The blue line is a local polynomial approximation centered around each one of these centiles.

the slope becomes strongly negative. The lower the past growth, the higher the future growth will be, which is consistent with the idea of a rebound. Conditional on survival, very poor past performance predicts strong future growth, as in our model.

2.5.2 Predicting Forecast Errors

The model was designed to predict that forecast errors be a non-linear function of past revisions. Given that the model assumes that all firms have the same variance of shocks, it is natural to check that our main empirical results holds after rescaling by firm-level variance. Another reason why it is important to perform such a robustness check is discussed is discussed in the previous Section. Assume, following Wyart and Bouchaud, 2003, that large news purely come from a separate group of firms (those with more volatile, but still Gaussian, shocks). Then, if forecasters use a firm-level linear forecasting rule, then their forecast errors should be close to unpredictable (to the extent that the AR(p) model they use mimics the optimal Kalman filter).[3]

Figure 2.8 shows that normalized error and normalized revisions follow the same relationship as in our headline Figure 2.1. In this Figure, we simply show the binned scatter plot of future log forecast errors $\frac{g_{t+1}-F_t g_{t+1}}{MAD_i}$ as a function of past revision $\frac{F_t g_{t+1}-F_{t-1} g_{t+1}}{MAD_i}$. This suggests that the non-linear relationship does not stem from firm volatility heterogeneity.

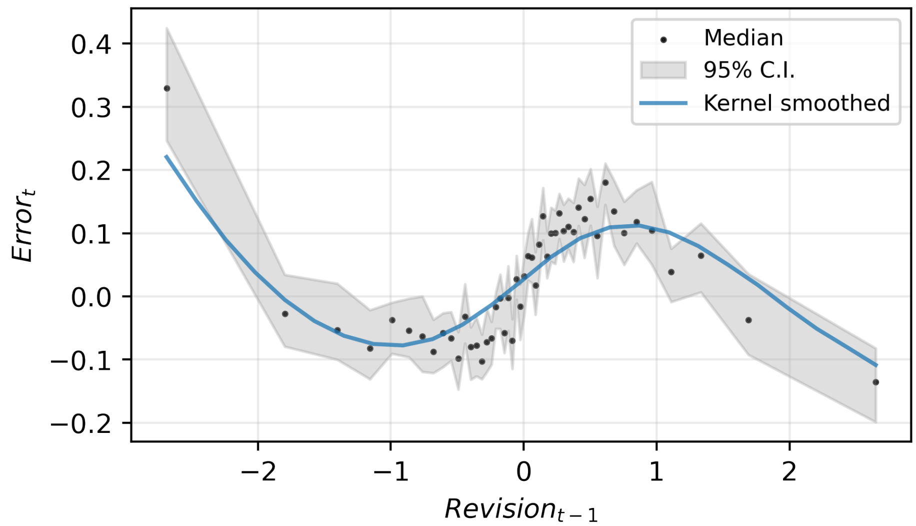


Figure 2.8: Revenue forecast error conditional on past revision (normalized)

Note: In this figure we use our international sample of firm revenue expectations to report the binned scatter plot of future log forecast errors $\frac{g_{t+1}-F_t g_{t+1}}{MAD_i}$ as a function of past revision $\frac{F_t g_{t+1}-F_{t-1} g_{t+1}}{MAD_i}$. The blue line is a local polynomial approximation, centered in the middle of each bin.

Another natural prediction of our forecasting model is that current and past forecast errors should follow a similar relationship. This comes from the fact that in our linear forecasting model, errors and revisions are proportional (equation 2.4). Thus, it mechanically follows that if error and lag revisions are linked by the S-shaped curve of Figure 2.8, then error and lagged error should follow the same relationship.

We look at the relation between error and lagged error in Figure 2.9. It shows the binned

[3] If, however, forecasters were to use a global forecasting rule (a single rule estimated on all firms), the non-linear shape may be predicted. Indeed, assume all shocks are Gaussian, but firms differ in the volatility of their *temporary* shock ϵ. In this case, a unique forecasting rule would overestimate the persistence of large shocks. We do not explore this lead in this chapter since we have documented in the previous Section that normalized growth is far from being normally distributed.

scatter plot of future log forecast errors $\frac{g_{t+1}-F_t g_{t+1}}{MAD_i}$ as a function of past error $\frac{g_t-F_{t-1}g_t}{MAD_i}$, both of them normalized by firm-level volatility. As can be seen from this figure, the forecast errors follow a linear relationship for intermediate values (until about 1 unit of volatility), but the relationship reverses for larger past errors.

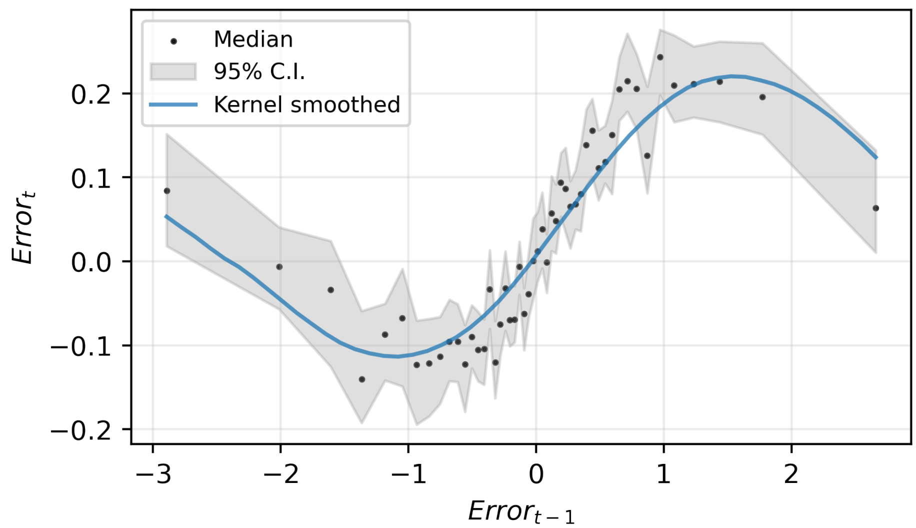


Figure 2.9: Revenue forecast error conditional on past error (normalized)

Note: In this figure we use our international sample of firm revenue expectations to report the binned scatter plot of future log forecast errors $\frac{g_{t+1}-F_t g_{t+1}}{MAD_i}$ as a function of past error $\frac{g_t-F_{t-1}g_t}{MAD_i}$. The blue line is a local polynomial approximation centered in the middle of each bin.

2.5.3 Evidence from Returns

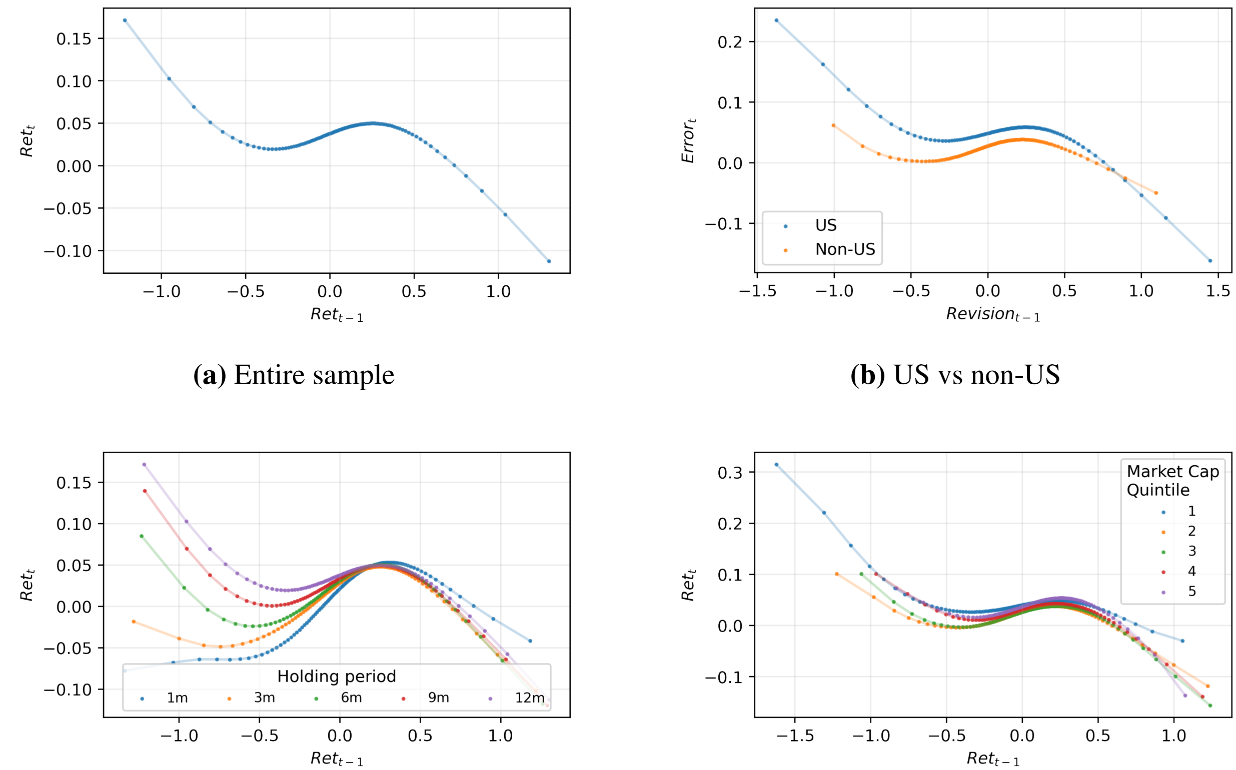
Our model in Section 2.4.5 predicts that past returns should predict future returns in a non-linear way. We now provide evidence on returns based on CFM's international monthly stock returns data described in the Data section of this chapter.

In Figure 2.10 we first show a smoothed binscatter plot of future returns on past returns. Future returns are monthly and past returns are calculated over the past 12 months excluding the last month of returns, as is common in the literature on stock momentum. The only difference here with the standard literature is that we take the log of returns (this is done because this analysis tends to focus on extreme past returns).

Index	Total	2000	2005	2010	2015	2020
AEX	5892	0	290	300	278	283
AS5	47252	0	2127	2340	2164	2361
CAC	9946	0	475	474	474	480
DAX	6531	0	360	360	360	351
HSC	5946	0	0	281	435	585
HSI	8949	0	0	522	582	597
IBE	8663	0	398	419	410	419
IND	5377	0	0	360	360	348
KOS	36595	0	0	2381	2345	2369
MID	120017	4585	4773	4770	4583	4750
NDX	21624	0	1171	1200	1234	1197
NIF	6794	0	0	52	599	600
NKY	60783	2642	2674	2698	2676	2698
OMX	7565	0	348	360	360	360
RAY	830673	29792	32936	33833	31915	33247
SMI	3715	0	0	233	240	228
SPT	10200	0	0	707	708	720
SX5	9744	0	0	600	600	587
TOP	5229	0	0	5	504	452
TPX	518734	16490	19074	19783	21722	25588
TWY	13663	0	0	355	1168	1028
UKX	29775	1086	1180	1196	1180	1187

Note: This table shows the total observation count, by exchange, for our study of stock returns, as well as observation counts for select years. As not all sample years are shown, the counts for individual years do not sum up to the total in the leftmost column.

Table 2.2: Sample size by exchange (returns)



(a) Entire sample

(b) US vs non-US

(c) By holding period

(d) By market cap

Figure 2.10: Returns conditional on past returns, by sample and holding-period

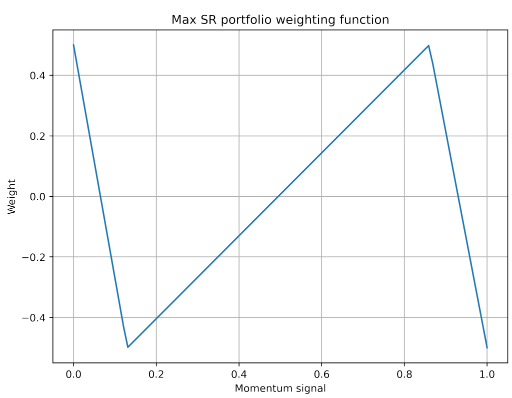
Note: These 4 panels represent smoothed binned scatter plots of future log returns as a function of log past cumulative returns of the past 12 months excluding the last month. Panel (a) is the entire sample; Panel (b) splits the sample into US and non-US stocks; Panel (c) computes future returns using different holding periods; and Panel (d) splits the sample by size quintile, using home exchange breakpoints.

Figure 2.10 shows the binned scatter plot for different splits of the data. Panel A looks at the entire dataset and provides a picture consistent with our prediction: There is momentum for most levels of past returns, but for extreme values it is mean-reversion that prevails. Panel B shows that this pattern holds both on US data and non-US returns. Panel C investigates the role of various holding periods, i.e. looking for future returns over the following 1, 3, 6, 9 and 12 months. We find that the S-shaped curve emerges as soon as this holding period is longer than one month. In panel D, we sort stocks into market cap quintiles at the index-month level. Even when examining different sizes of stocks, the S-shaped pattern is to be seen everywhere.

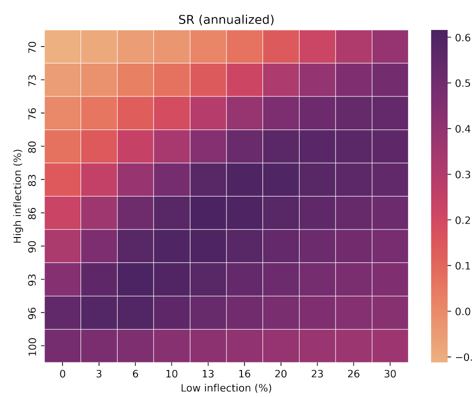
This finding suggests that the performance of traditional momentum strategies could be “boosted” by allowing for a region of reversal in both tails. To test this, we consider a self-financing strategy that goes long on momentum for moderate values of the momentum signal, and goes short on a momentum (i.e. long on reversal) for more extreme values of the signal. Let $s_{i,t-1}$ be a momentum signal calculated from past returns. Specifically, we calculate the momentum signal as the cross-sectional rank transform of cumulative returns over 11 months from $t-12$ to $t-1$, normalized such that $s_{i,t-1} = 0.5$ for firm with the greatest past returns, and $s_{i,t-1} = -0.5$ for the firm with the least. A portfolio with weights $w(s_{it})$ is then formed at time t as follows:

$$w_{it} = \begin{cases} 0.5 - \frac{s_{i,t-1}}{a} & \text{if } s_{i,t-1} \leq a \\ \frac{s_{i,t-1}-b}{b-a} - 0.5 & \text{if } a < s_{i,t-1} \leq b \\ 0.5 - \frac{s_{i,t-1}-b}{1-b} & \text{if } b < s_{i,t-1} \end{cases}$$

where a and b are constant "inflection" points at which our strategy flips from reversal to momentum, and then from momentum to reversal.



(a) Portfolio weights



(b) SR by inflection point

Figure 2.11: Tail-reversal strategy

Note: Panel (a) shows example tail-reversal strategy weights for the Sharpe ratio maximizing inflection points a and b. Panel (b) shows the Sharpe ratio obtained varying a and b. Portfolio formed across all firms in our equity sample, both U.S. and non-us. See text for details of strategy construction.

Our findings are depicted in figure 2.11. The inflection points of our strategy, a and b are chosen to optimize the Sharpe ratio of our strategy over the sample period. Panel A shows the

Sharpe ratio maximizing weighting function, while panel B shows how the Sharpe ratio varies for different upper and lower inflection points. Note that a strategy with a low inflection point of 0 and upper inflection point of 100 corresponds to a traditional momentum strategy, with no tail reversal. Clearly, our approach contains some look ahead bias, as the coefficients a and b are estimated on the entire sample. A more systematic investigation of these returns is beyond the scope of this chapter.

In line with our earlier empirical results, we find that the Sharpe ratio of our strategy is maximized when the lower inflection point is set at the 13th cross-sectional percentile of the normalized momentum signal, and the upper inflection point is set at the 86th percentile. The maximum Sharpe ratio to this strategy (0.61) is 1.27x the Sharpe ratio we observe for a pure momentum strategy (0.48).

2.6 Conclusion

In this chapter we emphasize that boundedly rational agents, when faced with fat-tailed processes, will make predictable mistakes. In order to explore such processes, we need large samples. Our empirical research here leverages the international version of IBES which gives us a large panel of sales growth forecasts. Consistently with the firm demographics literature, we find that sales growth dynamics are well described by the sum of a short-run and and long-run processes. The long-run process is a simple Gaussian, AR1 process, but the short-run process has fat tails. As a result, a simple, linear filtering rule will not be optimal. This simple model of expectations formation matches a lot of the key features of the data.

Natural extensions of our work consists in exploring alternative forecasts data. Macro forecasts are unlikely to provide us with non-Gaussian processes and are in general too sparse to measure the tails of the DGP with enough accuracy. Within IBES studying EPS forecasts is another natural research direction, although it presents a scaling challenge. Growth cannot be computed for a large number of firms. Internal sales forecast from large companies could be another path.

Chapter 3: Macroeconomic attention and expected returns[1]

[1]Joint work with Ken Teoh (Columbia). My coauthor and I are grateful to Stephanie Schmitt-Grohé, Serena Ng, and Laura Veldkamp for invaluable guidance, support, and suggestions. We also thank Hassan Afrouzi, Andres Drenik, Émilien Gouin-Bonenfant, Gustavo Cortes, Se- ungki Hong, Sophia Kazinnik, Jennifer La'O, Simon Lee, José Luis Montiel Olea, Bernard Salanié, Jesse Schreger, Mani Sethuraman, Martín Uribe, Michael Woodford for constructive discussions, as well as participants at the 27th International Conference in Computational Economics and Finance, Columbia University's Economic Fluctuations and Econometrics Colloquia for helpful suggestions. Both authors acknowledge the generous financial support of Columbia University's Program for Economic Research.

3.1 Introduction

This chapter investigates whether attention to the macroeconomy is a source of risk that is priced in the cross section of expected returns. We find that firms with higher macroeconomic attention subsequently earn returns that are lower on average than firms that pay less attention to the macroeconomy. Sorting stocks into portfolios based on their degree of attentiveness to the macroeconomy, we document that stocks in the top decile (most attention) have returns that are on average 1.02 percent per month (11.6 percent per year) lower than stocks in the bottom decile (least attention). A portfolio that longs the top decile portfolio and shorts the bottom decile portfolio earns an absolute annualized Sharpe ratio of 1.34, which is sizeable relative to the market portfolio's Sharpe ratio over the same sample period.

If macroeconomic attention proxies for a stock's exposure to aggregate risk, then the excess returns of macroeconomic attention sorted portfolios should be fully explained by the portfolios' market betas. We find that the excess return of the portfolios persists even after controlling for market betas, as well as other factors and firm characteristics known to predict the cross section of asset returns, including the stock's log market capitalization, book-to-market ratio, exposure to aggregate volatility and idiosyncratic volatility. In each of these cases, we find that returns of portfolios decrease as the average macroeconomic attention of stocks in the portfolios increase. Our findings are also robust to measuring the effects of macroeconomic attention on returns at the individual stock level. Hence, an explanation that relies solely on exposure to aggregate risk does not fully account for the negative macroeconomic attention premium we observe.

Measuring firm-level macroeconomic attention is challenging given that we do not directly observe what firm managers pay attention to. Recent empirical work draws on survey responses of firm managers and finds rich heterogeneity in the degree to which managers are paying attention to macroeconomic conditions (Kumar et al., 2015; Coibion, Gorodnichenko, and Kumar, 2018). We contribute to this literature by proposing an alternative measure based on the attention CEOs and CFOs allocate to discussing macroeconomic conditions in earnings calls. To identify macroeconomic-related discussions in calls, we train a neural-network classifier on a library of labeled Reuters news articles published in the same period as the calls, and then use the classifier to label sentences in earnings calls transcripts as either macroeconomic- or firm-related.

To explain the negative macroeconomic premium, we first link our measure of macroeconomic attention to firm fundamentals through a model of analysts whose objective is to accurately forecast future earnings, which in turn is affected by unobserved macroeconomic and firm-specific shocks.[2] Information about these shocks is conveyed through the content of the call, which is naturally limited by the length of the call and the finite attention span of analysts listening in on the call. Prior literature offers evidence of such limitations, including underreaction of prices to earnings announcements made on Fridays as opposed to other weekdays (Dellavigna and Pollet, 2009) and lower consumption of earnings calls on days where more firms hold calls (Heinrichs, Park, and Soltes, 2018). The model predicts that attention to the macroeconomy is increasing in the share of earnings news explained by the macroeconomic component.

Using the return decomposition framework of Campbell (1991), we then show that firms with a greater share of cash flow news explained by the macroeconomic component is associated lower cash flow risk, hence lower risk premium. The reason is that these stocks contribute less undiversified cash flow risk to the investor's portfolio, hence earn lower expected returns. As the price of risk on cash flow news is different from that of discount rate news (Campbell and Vuolteenaho, 2004), market betas do not fully explain heterogeneity in cash flow risk. At the level of portfolio returns, we find evidence that stocks with higher macroeconomic attention have lower cash flow betas. This finding is consistent with the explanation that different levels of macroeconomic attention we observe reflects different exposures to aggregate cash flow risk, and supports a rational explanation for the macroeconomic attention premia.

[2] The model builds on a rich literature that studies the implications of rational inattention, in particular Maćkowiak and Wiederholt (2009). See Maćkowiak and Wiederholt (2009), Maćkowiak, Moench, and Wiederholt (2009), and Mackowiak, Matejka, and Wiederholt (2018) for models of attention allocation in a linear Gaussian quadratic framework, Van Nieuwerburgh and Veldkamp (2009), Van Nieuwerburgh and Veldkamp (2010), and Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016) for models of attention allocation in a noisy rational expectations equilibrium framework.

3.1.1 Related literature

Our analysis is an application of neural network models, which has gained prominence in natural language application in recent years.[3] Complex prediction problems such as parts-of-speech tags, question answering, and machine translation leverage sophisticated networks with convolutional and recurrent structures. While these structures allow for a richer representation of meaning to be extracted from the text data, we find that the nonlinearity of neural network models provides significant accuracy gains for our classification tasks without the need to introduce more complexity into our prediction model. As such, we focus on the simplest of neural network models for our application.

Closely related to our work, Flynn and Sastry (2020) measures macroeconomic attention in firm disclosure data. Our findings complement theirs in that we find that our measure of macroeconomic attention is strongly countercyclical. However, our analysis differs in several regards. First, we apply a supervised machine learning method for parsing the content of earnings calls related to the macroeconomy, which is more suitable to our application given our interest in predicting returns. Second, their analysis focuses on the implications of attention allocation for the volatility of output growth, whereas our analysis focuses on the effects of attention allocation for the cross section of stock returns.

Our work also contributes to a growing literature in macroeconomics using earnings call transcripts.[4] Hassan et al. (2019) constructs a novel measure of political risk from the text of earnings calls and finds that their measure strongly predicts investment, hiring, lobbying and political donation activities of firms in a manner highly indicative of political risk. Hassan et al. (2021) constructs a measure of country-level risk using earnings calls transcripts. They find that increased perceptions of country riskiness is associated with capital outflows and fall in asset prices within that country, which provides novel insight into cross-border contagion through firm-level exposures to

[3] See Goldberg (2016) for a survey of application of neural network models in textual analysis.

[4] Gentzkow, Kelly, and Taddy (2019) provides a recent and comprehensive survey of the methods in textual analysis with emphasis on its application to economics and finance, Li (2010) and Loughran and McDonald (2016) provide comprehensive surveys of the use of textual analysis in the accounting literature.

country specific risks.

Finally, our work contributes to a long tradition in finance studies characteristics that predict expected returns in the cross-section of stocks.[5] We add to this literature by documenting a characteristic that has economically meaningful and robust effects on stock returns, particularly in the recent sample period. Furthermore, we draw on the contributions of Campbell and Mei (1993), Vuolteenaho (2002), Campbell and Vuolteenaho (2004), and Campbell, Polk, and Vuolteenaho (2009) to explain the macroeconomic attention risk premium that we observe in the data.

The rest of the chapter proceeds as follows. Section 3.2 describes our methodology for measuring macroeconomic attention, and Section 3.3 reports our empirical results on the effects of macroeconomic attention on stock returns. In Section 3.4, provide our explanation for the negative macroeconomic attention that we observe in the data in Section . Finally, Section 3.5 concludes.

3.2 Measuring firm-level macroeconomic attention

3.2.1 Data

Our laboratory is the set of transcripts of earnings conference calls held by publicly listed firms in the United States from 2002Q1 to 2020Q2, transcribed and published by FactSet. Our subsequent analysis focuses on calls of US public listed firms, excluding firms in the financial and utilities sectors, which take place between 2005Q1 and 2019Q4.[6]

Earnings calls are typically held once per quarter, and serve as a medium for firms to discuss their most recent earnings results and disclose material information to market participants. Earnings calls typically consist of a management discussion section where senior management discusses the company's most recent financial results and a questions and answers section where they field questions from a selected group of analysts.

Heinrichs, Park, and Soltes (2018) documents that earnings call participants are primarily buy-

[5] See Lewellen (2015), Mclean and Pontiff (2016), Chen (2020), and Feng, Giglio, and Xiu (2020) for detailed surveys of cross-sectional predictors of asset returns.

[6] We drop calls in 2020 given that we have access to data only until May 2020, and drop calls before 2005 due to poor coverage of publicly listed US firms. See Figure B.1 in the Appendix for share of public listed US firms with matched earnings calls over time.

side investors, and that participants are approximately equally split between common stock holders and non-holders. They find that buy-side non-holders are significantly more likely to invest in the firm's shares in the quarter following its earnings calls, after controlling for the firm's earnings performance and other firm and call characteristics. Their evidence lends support that conference calls contain material information that influence the investment decisions of market participants, complementing prior evidence that calls materially affect stock returns, both through the content of the discussions as well as in the use of vocal cues (Matsumoto, Pronk, and Roelofsen, 2011).

As training data, we use a collection of Reuters news articles that we label as either related to the macroeconomy or firm-specific news. We gather our library of Retuers news articles by systematically collecting the top-ranked Google Search results Reuters news every week from January 2004 to May 2020. Our search query is

$$
\begin{aligned}
\text{query} = &\text{``site: reuters.com''} \\
&+ \text{``economy''or ``[company name]''} \\
&+ \text{``after: [start date]''} + \text{``before: [end date]''}
\end{aligned}
$$

where [company name] is a placeholder for the name of companies from the S&P1500. This search query algorithm allows us to collect both news articles relevant to economic conditions, as well as articles related to company-specific news.

We then proceed to systematically label our news articles using topic codes in the metadata of the articles collected, which are assigned by Reuters for search engine optimization[7] Specifically, we classify an article as "Macro" news if it has keywords "United States" and "Economy" but not "Company News". Conversely, an article is classified as "Firm" news if it has keywords "United States" and "Company News" but not "Economy". This ensures we obtain non-overlapping training libraries for macro and firm-specific words. Our final sample consists of 44,835 Reuters news articles, with 12.7 percent of articles labeled as related to the macroeconomy.

[7] See Reuters metadata guide (`https://liaison.reuters.com/tools/metadata-guide`) for a complete list of topic codes.

3.2.2 Preprocessing and feature selection

Prior to estimation, we preprocess the text of both Reuters news articles and earnings calls transcripts similarly with the purpose of reducing the vocabulary to a set of meaningful terms informative about underlying content of interest. The preprocessing procedure that we adopt is commonly applied in the literature on textual analysis (Gentzkow, Kelly, and Taddy, 2019), and involves removing stop words such as "the", "a", "an", which convey little meaning in our application, and stemming words down to their root using the Porter stemmer algorithm (Porter, 1980).

We divide the news articles and call transcripts into a collection of sentences, which in turn consists of a collection of words. Formally, denote the vocabulary $\mathcal{V}$ as the unique set of words w in the training data, and a document $d_{s,i,t}$ as a $|\mathcal{V}| \times 1$ vector of counts for each word in the vocabulary, indexed by the sentence s from the earnings call held by firm i in quarter t. Denote corpus $\mathcal{D} = \cup_{s,i,t} d_{s,i,t}$ as a collection of documents across sentences s, firms i, and quarter t. Our corpus has a bag-of-words representation, which is an $|\mathcal{D}| \times |\mathcal{V}|$ matrix of word counts. While this characterization of the text ignores the rich complexity conveyed by the grammatical structure and co-occurrence of words in each document, it is shown to be an effective representation of text data in many economics and finance applications (Gentzkow, Kelly, and Taddy, 2019).

To construct the vocabulary list, we compile the set of individual words or unigrams from the text, as well as bi- and trigrams, such as "capit expenditur" and "foreign currenc exchang", which we construct by the patterns-of-speech algorithm adopted in Hansen, McMahon, and Prat (2018). We retain bigrams that occur at least 100 times in the corpus and trigrams that occur at least 50 times.

To further reduce the weights of common and rare words in subsequent analysis, we assign a document-specific measure of importance to each word in a given document known as the term frequency-inverse document frequency (tf-idf) score. Formally, for word v in document d_{sit}, define the term frequency-inverse document frequency (tf-idf) score $w_{d,v}$ as

$$w_{d,v} = f_{d,w} \times \log\left(\frac{1}{b_{d,w}}\right)$$

where $f_{d,v}$ is the fraction of times word w occurs in document d, and $b_{d,v}$ is the fraction of documents word v appears across the entire corpus $\mathcal{D}$. A document d is thus represented as a $|\mathcal{V}| \times 1$ vector of tf-idf scores $w_d = [w_{d,1}, ..., w_{d,|\mathcal{V}|}]'$, and the document-term matrix is the $|\mathcal{D}| \times |\mathcal{V}|$ matrix of tf-idf scores, where $w = [w'_1, ..., w'_{|\mathcal{D}|}]'$. To further reduce the dimensions of our data, we restrict the set of vocabulary terms to top 5000 terms by average tf-idf score across documents in the training sample.

3.2.3 Classification problem

We are interested in mapping the representation of the text w to the outcome variable of interest. For our analysis, this outcome variable is the attention to the macroeconomy in earnings calls of firm i in quarter t. In our library of Reuters news article sentences, this variable takes on a binary value of 0 or 1, where 1 indicates a sentence from an article labeled as "Macro" and 0 otherwise. Our set of earnings call transcripts, however, do not have such naturally assigned labels. As such, we are interested in estimating a function that allows us to predict the label m_d of the document from from its tf-idf score T_d. Formally, for each document d, we want to estimate the function

$$m_d = h(w_{d,1}, ..., w_{d,|\mathcal{V}|})$$

where $h(\cdot)$ is a function that maps the tf-idf score w_d to the probability that document d has label 1, $m_d \in [0, 1]$. As an application in supervised machine learning, we train the parameters of the function $h(\cdot)$ on our training data for which we have labels assigned (Reuters news articles), and then use the trained function to predict labels on data for which we do not have labels (earnings call transcripts).

Our preferred approach is to approximate the function $h(\cdot)$ using an artificial neural-network. Artificial neural networks (ANN) encompass a large class of machine learning models with widespread applications in many fields, and is one of the preferred approaches for complex prediction problems such as computer vision, natural language processing, and speech recognition (Hastie, Tibshirani, and Friedman, 2009). This is largely due to the model's ability to approximate any continuous function arbitrarily well for a large enough number of hidden features.Hornik, Stinchcombe, and White (1989)

Formally, the ANN model is given by

$$m_d = h(w_{d,1}, ..., w_{d,|\mathcal{V}|}) \approx \theta_0 + \sum_{n=1}^{N} \theta_n \sigma\Big(\beta_{0,n} + \beta_n' w_d\Big) \tag{3.1}$$

where the first layer consists of N linear models with parameters $(\beta_{0,n}, \beta_0)$, and the second layer is a linear combination of the N models with parameters (θ_0, θ_n), after applying a suitable non-linear transformation to each model. There are many possible non-linear transformations commonly applied in estimating the ANN. Given that the outcome variable is $\{0, 1\}$, choosing the sigmoid function $\sigma(x) = 1/(1 + e^{-x})$, which restricts the output space to be $[0, 1]$, is appropriate. Multi-layer ANN generalize the above structure by allowing for multiple layers of hidden units, where each layer consists of a linear combination of the previous layer's hidden units after applying a nonlinear transformation. The specification shown is a single hidden-layer ANN.

3.2.4 Inference

For the single-layer artificial neural network, the model parameters consists of $N(|\mathcal{V}|+1)$ first-layer weights $\{\beta_{0,n}, \beta_{n,w}; n = 1, ..., N, w = 1, ..., |\mathcal{V}|\}$ and $(N+1)$ output-layer weights $\{\theta_0, \theta_n; n = 1, ..., N\}$. As an application of supervised learning, we train the parameters of the model $\Theta = \{\theta_0, \theta_n, \beta_{0,n}, \beta_{n,w}\}$ on our library of Reuters news articles. Formally, the parameters are estimated by minimizing the cross-entropy loss function

$$\min J(\Theta) = \frac{1}{|\mathcal{D}|} \sum_{d=1}^{|\mathcal{D}|} \mathcal{L}(\hat{m}_d, m_d)$$

$$\mathcal{L}(\hat{m}_d, m_d) = -m_d \log(\hat{m}_d) - (1 - m_d) \log(1 - \hat{m}_d) \tag{3.2}$$

where $\hat{m}_d$ denotes the prediction of document d's label from the neural net model (we drop the subindices of documents (s, i, t) for notational convenience). The generic approach to minimizing $J(\Theta)$ is by gradient descent, which involves computing the predicted values $\hat{m}_d$ using the forward propagation equation (3.1), computing the local gradients of each parameter given the lost function (3.2), and then updating each parameter by the size of the computed gradients subject to a learning rate adjustment. This process is done iteratively until the parameter estimates converge as the gradients tend to zero with each update. In general, computing the gradients using all training examples $|\mathcal{D}|$ is computationally expensive, and a common solution is to employ stochastic gradient descent in which the gradient of parameters are updated using a random subset of the training data. The stochastic nature of the descent requires that the step size of the updates shrinks to zero as the gradient approaches zero so that the noise from random sampling does not dominate the directional signal of the gradient. A critical tuning parameter that governs the step size is the learning rate, and we adopt the Adam algorithm proposed by Kingma and Ba (2017) to adaptively control the learning rate.

Fitting a neural-net model also involves choosing the number of hidden layers and number of hidden features per layer. A common approach is to estimate the optimal number of units and layers by cross-validation. This approach evaluates the fit of the model on a validation dataset not used to estimate the model parameters, and chooses the number of units and layers that yields the best fit. Our preferred specification based on a five-fold cross validation procedure is a single hidden-layer neural-net with $N = 64$.

Having obtained a prediction of the macroeconomic relevance of each sentence, we construct a measure of attention to the macroeconomy for firm i in quarter t as the share of sentences that are macroeconomic-relevant. Formally, define C_{it} to be the set of sentences d in earnings call for firm i in quarter t. The macroeconomic attention of firm i in quarter t is given by

$$MacroAttn_{it} = \frac{1}{|C_{it}|} \sum_{d \in C_{it}} 1\{m_d \geq c\}$$

where c is a threshold to be specified. In our baseline specification, we choose $c = 0.5$, but find that our results remain robust to variation in this threshold. As a validation exercise, we report the top-20 terms that occur in the earnings call transcripts labeled as "Macro" as well as "non-Macro" in Appendix Table B.2. We find interpretable differences in the terms found in earnings call transcript sentences labeled as "Macro" and "non-Macro", which suggest that our algorithm works as intended.

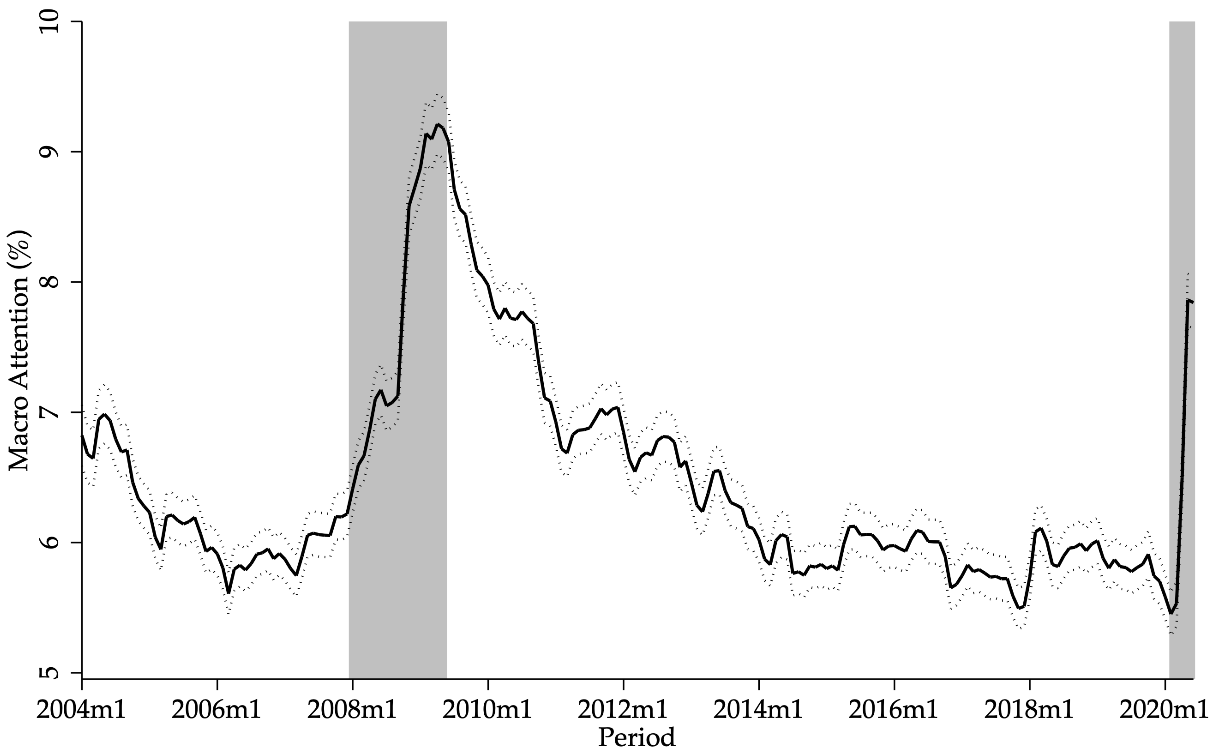


Figure 3.1: Attention to the macroeconomy over time.

Note: The sample period is from January 2004 to May 2020. Dotted lines show the 95 percent confidence interval of the estimated average. Shaded bar denotes NBER recession months.

3.2.5 Discussion

Figure 3.1 plots the cross-sectional average log macroeconomic attention each month over the sample period for which we have earnings calls transcripts – from January 2004 to May 2020.[8] The

[8]To construct monthly observations for each firm, the quarterly macroeconomic estimates are carried forward from the month the call was held until the month before next call is observed.

average attention to the macroeconomy is strongly countercyclical; The two peaks in the attention index occurs in April 2009 and May 2020, which coincides with the Great Financial Recession and the Covid-19 pandemic-induced recession. The countercyclical discussion of macroeconomic conditions is consistent with the notion of attention in models of rational attention allocation. In particular, heightened aggregate uncertainty during recessions draw more attention away from firm-specific conditions to macroeconomic conditions (Maćkowiak and Wiederholt, 2009; Kacperczyk, Van Nieuwerburgh, and Veldkamp, 2016).

Another plausible explanation for the observed countercyclical nature of macroeconomic discussions is that firms managers may be more inclined to blame the economy when they perform poorly relative to expectations. Recent empirical work finds evidence that managers engage in strategic disclosures that are influenced by earnings performance, for example by limiting information in the presence of unfavorable condition (Chen, Matsumoto, and Rajgopal, 2011) or disproportionately calling on analysts with positive views of the firm (Cohen and Malloy, 2016). We explore the robustness of the relationship between macroeconomic uncertainty and attention in the following regression specification

$$log(MacroAttn)_{ic} = \alpha_i + \beta_1 log(MacroUnc)_c + \beta_2 SUE_{ic} + \delta X_{ic} + \epsilon_{ic} \tag{3.3}$$

where $log(MacroAttn)_{ic}$ is log macro attention of firm i in call c, $log(MacroUnc)_c$ is the log macroeconomic uncertainty in the month call c was held, SUE_{ic} is firm i's earnings in the fiscal quarter associated with call c relative to market expectations, X_{ic} are firm-specific characteristics in the fiscal quarter associated with call c, and α_i controls for firm fixed effects. To proxy for uncertainty about macroeconomic shocks, we use the 12-month ahead total macro uncertainty index (JLN Uncertainty) of Jurado, Ludvigson, and Ng (2015). As controls, we include firm specific controls including the log market capitalization (Size), debt-to-asset ratio (Leverage), log book-to-market equity ratio (Book-to-market), and number of analysts issuing forecasts (NumAnalyst). We also control for the length of the call using the number of sentences, and the dictionary-based

	log(MacroAttn)
log(MacroUnc)	0.986***
	(32.93)
Earnings Surprise	0.000132
	(0.04)
Controls	Yes
Firm FE	Yes
R-squared	0.439
Nobs	84260

t statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Table 3.1: Macroeconomic attention, uncertainty, and earnings performance.

Note: This table reports the regression estimates of macroeconomic attention on macroeconomic uncertainty, earnings surprise, time-varying firm controls, and firm fixed effects at the firm-quarter level. Controls include the firm size, leverage, book-to-market ratio, analyst coverage, and excess returns over the fiscal quarter, and number of sentences in the call. The sample period is from 2005Q1 to 2019Q4.

sentiment measure of the call constructed from the Loughran and McDonald (2011) sentiment dictionary.

If our text-based measure of macroeconomic attention is related to overall uncertainty about macroeconomic conditions, we expect $\beta_1 > 0$ after controlling for other potential explanations. In words, an increase in macroeconomic uncertainty corresponds to more time allocated to discussing macroeconomic conditions, controlling for firm-specific characteristics. Furthermore, finding $\beta_2 < 0$ would be evidence suggestive that firms which underperform market expectations are more likely to blame the economy.

Table 3.1 reports the coefficient estimates of the regression specification (3.3). We find that our measure of macroeconomic discussions covary positively with macroeconomic uncertainty ($\beta_1 > 0$), even after controlling for the firm's operating performance, time-varying characteristics and firm fixed effects. This result provides suggestive evidence that discussions of macroeconomic conditions are higher during periods of higher macroeconomic uncertainty. We do not find evidence that firms are more likely to discuss macroeconomic conditions when earnings underperform relative to analyst expectations ($\beta_2 = 0$).

3.3 Macroeconomic attention and expected returns

In this section, we examine whether our measure of macroeconomic attention predicts the cross-section of stock returns. Our analysis is based on two well-established methodologies for testing predictors of the cross section of expected returns: (1) examining the excess returns of portfolios sorted on the basis of macroeconomic attention, and (2) a regression of returns on firm characteristics in the spirit of Fama and MacBeth (1973).[9]

To improve our predictor's signal-to-noise ratio, we fit a predictive model of macroeconomic attention using past values of macroeconomic attention, and use the out-of-sample predicted value of macroeconomic attention as our preferred measure of the characteristic in a given quarter. Formally, for firm i and quarter $t \in \{1, ..., T\}$, we estimate a linear predictive model given by $log(MacroAttn)_{i,t} = \alpha_i + \sum_{j=1}^{4} \beta_j log(MacroAttn)_{i,t-j} + \varepsilon_{i,t}$, where α_i are firm fixed effects, using only data up to quarter T. Having estimated the model parameters, we then predict the out-of-sample value of macroeconomic attention in quarter $T + 1$, $log(\widehat{MacroAttn})_{i,T+1}$. We repeat this procedure on a rolling basis to generate predicted values of macroeconomic attention for each firm over the sample period.

Prior empirical studies have found a large number of cross-sectional factors that have explanatory power for the cross-section of returns (Lewellen, 2015; Mclean and Pontiff, 2016; Chen, 2020; Feng, Giglio, and Xiu, 2020), and the challenge lies in showing that our characteristic provides incremental information about the cross section of returns. We control for well-known factors and characteristics known to predict returns, including exposure to the market factor (beta), log market capitalization (size), book-to-market ratio (BM), as well as aggregate volatility and idiosyncratic volatility betas from Ang et al. (2006). We also examine whether returns of portfolios sorted on the basis of macroeconomic attention can be explained by the market model, Cahart four-factor model, Fama-French three- and five-factor models (Fama and MacBeth, 1973). As an additional robustness exercise, we include additional controls by considering the 15 characteristics found by

[9]To reduce the effects of outlier observations in subsequent analysis, we restrict our sample to firm-month observations with stock prices greater than 5 and less than 1000, ordinary common shares incorporated inside the US (CRSP share codes 10 and 11).

	Macro Attention	
$\beta(MKT)$	0.003	(0.091)
$\beta(SMB)$	-0.070	(0.000)
$\beta(HML)$	0.060	(0.000)
$\beta(VIX)$	0.006	(0.001)
Size	0.191	(0.000)
Book-to-market	0.086	(0.000)
Lagged returns (12 mths)	-0.052	(0.000)
Idio vol	-0.191	(0.000)
Issuances (36 mths)	-0.210	(0.000)
Accruals	-0.132	(0.000)
Return on asset	0.291	(0.000)
Dividend yield	-0.096	(0.000)
Asset growth	-0.180	(0.000)
Lagged returns (36 mths)	-0.070	(0.000)
Issuances (12 mths)	-0.181	(0.000)
Turnover	-0.083	(0.000)
Net debt-to-Price	0.132	(0.000)
Sale-to-Price	0.110	(0.000)

p-values in parentheses

Table 3.2: Correlation between macroeconomic attention and other predictors of stock returns.

Note: See Appendix Table B.3 for variable definitions.

Lewellen (2015) to be important predictors of returns. Table 3.2 reports the contemporaneous correlation of our measure of macroeconomic attention with these characteristics, which we control for in subsequent analysis.

3.3.1 Portfolio analysis

We sort stocks into ten portfolios based on the firm's macroeconomic attention at the beginning of each month, and compute portfolio return as an equally weighted average of monthly returns of stocks within each portfolio. Portfolio sorts, as opposed to regression of individual stock returns on characteristics, mitigate idiosyncratic noise in returns and provide stable estimates of risk prices, particularly in the presence of time-varying loadings of characteristics (Feng, Giglio, and Xiu, 2020).

Figure 3.2 plots the average monthly returns of each decile portfolios sorted on macroeconomic

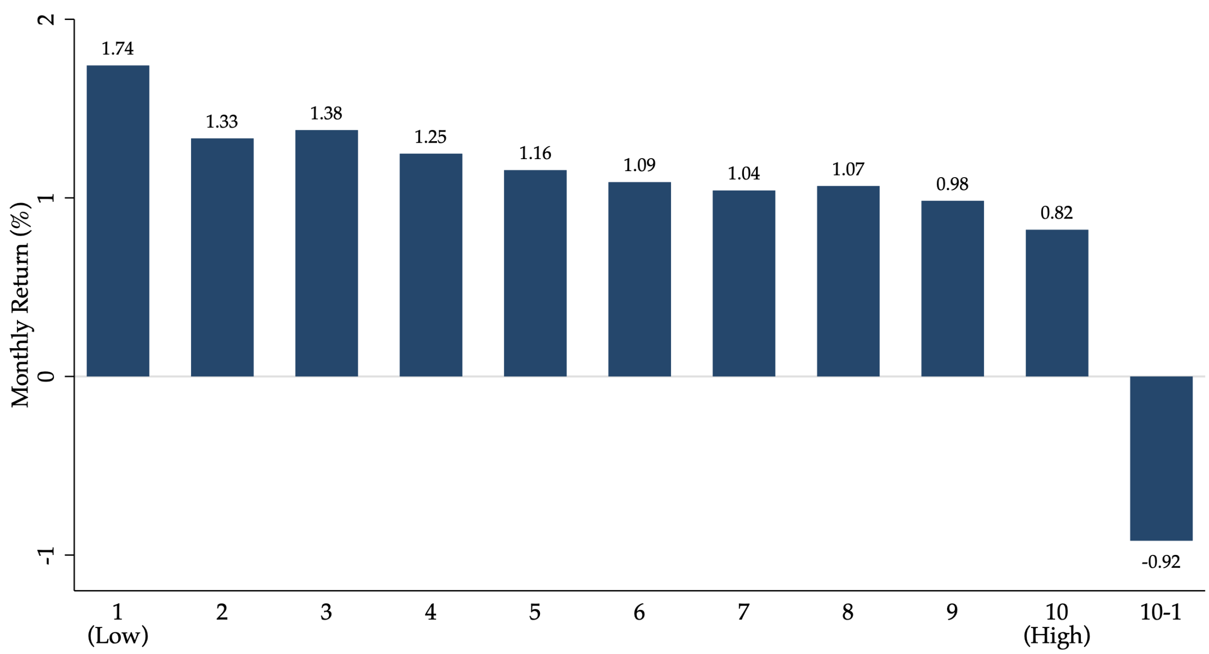


Figure 3.2: Average monthly returns of macroeconomic attention sorted portfolios.

Note: Stocks are sorted into ten equally-weighted portfolios using NYSE breakpoints based on the out-of-sample predicted macroeconomic attention in month , and then compute returns for month . "10-1" refers to a portfolio with a long position in portfolio 10 (high attention) and a short position in portfolio 1 (low attention). The sample period is from January 2005 to December 2019.

attention in excess of the risk free rate. The last column shows the excess returns of a cash-neutral portfolio that takes a long position in Portfolio 10 and a short position in Portfolio 1. Going from the bottom decile (lowest macroeconomic attention) to the top decile (highest macroeconomic attention), average returns declines from 2.00% per month to 0.98% per month. The long-short portfolio generates a return of -1.02% (-11.6%) per month (annum).

Table 3.3 reports the statistical significance of the average excess returns and alphas of the macroeconomic attention sorted portfolios. We examine whether the excess returns of the macroeconomic attention sorted portfolios are predicted by the factors of several asset pricing models from the literature. Across each of the model we consider, we find that the alphas of the long short portfolio remains negative and significant at the 1 percent level. This suggest that macroeconomic attention likely captures common variation in returns that is not fully explained by standard asset pricing models.

	AvgRet		CAPM		FF-3		Carhart-4		FF-5		FF-3 + FVIX	
	α	t	α	t	α	t	α	t	α	t	α	t
1	2.00***	(4.42)	1.11***	(5.13)	1.20***	(10.63)	1.61***	(6.42)	1.33***	(12.09)	1.19***	(10.65)
2	1.49***	(3.44)	0.61***	(3.33)	0.71***	(7.59)	0.92***	(4.08)	0.74***	(7.74)	0.70***	(8.04)
3	1.45***	(3.23)	0.58***	(3.48)	0.68***	(5.83)	0.87***	(3.26)	0.70***	(6.02)	0.68***	(6.28)
4	1.29***	(2.98)	0.42**	(2.38)	0.55***	(4.55)	0.88***	(3.13)	0.53***	(4.48)	0.55***	(4.63)
5	1.28***	(2.98)	0.38**	(2.19)	0.50***	(4.45)	0.76***	(3.69)	0.48***	(4.27)	0.50***	(4.48)
6	1.29***	(3.05)	0.41***	(2.66)	0.53***	(6.27)	0.79***	(6.25)	0.48***	(6.11)	0.53***	(6.30)
7	1.16***	(2.79)	0.29*	(1.73)	0.42***	(3.76)	0.69***	(2.89)	0.32***	(3.40)	0.42***	(3.74)
8	1.22***	(2.85)	0.34**	(2.16)	0.47***	(4.45)	0.60***	(2.86)	0.39***	(4.12)	0.47***	(4.49)
9	1.05**	(2.44)	0.15	(0.96)	0.29***	(2.71)	0.47**	(2.54)	0.23**	(2.02)	0.29***	(2.69)
10	0.98**	(2.23)	0.07	(0.47)	0.23*	(1.92)	0.37**	(2.21)	0.16	(1.42)	0.23*	(1.92)
10-1	-1.02***	(-6.15)	-1.04***	(-5.69)	-0.97***	(-6.74)	-1.09***	(-5.05)	-1.17***	(-8.29)	-0.96***	(-6.57)

Table 3.3: Average excess returns and alphas in monthly percentages for equal-weighted decile macroeconomic attention portfolios.

Note: Stocks are sorted into equal-weighted decile every month based on predicted macroeconomic attention, where Portfolio 1 are stocks with the lowest attention and Portfolio 10 are stocks with the highest attention. “10-1” refers to the difference in monthly returns between Portfolio 10 and Portfolio 1. “AvgRet” refers to average portfolio returns in excess of the risk free rate, “CAPM” is the alpha (intercept) of regressing average portfolio excess return on the market factor, “FF-3” is the alpha from the Fama and French (1993) 3-factor model, "Carhart-4" is alpha from the Carhart (1997) 4-factor model, "FF-5" is alpha from the Fama and French (2015) 5-factor model, and "FF-3+FVIX" is alpha from the Fama-French 3-factor model and VIX factor from Ang et al. (2006). Within each column, subcolumn refer to the estimated alpha and is the associated t-statistics, which incorporate Newey-West correction with four lags. The sample period is from January 2005 to December 2019. * p<0.10, ** p<0.05, *** p<0.01.

As an exercise in robustness, we control for the pricing effects of characteristics known to predict the cross section of returns. We consider four well-known characteristics from the literature: sensitivity to the market factor (beta), log market capitalization (size), book-to-market ratio (BM), sensitivity to aggregate volatility (Agg Vol) and idiosyncratic volatility of returns (Ang et al., 2006). To control for the effects of these characteristics, we construct double sorted portfolios where we first sort stocks into five portfolios based on the control characteristic using NYSE breakpoints, and then for each characteristic sorted portfolio, we further divide stocks into deciles based on macroeconomic attention using NYSE breakpoints. Finally, we take the simple average of returns of portfolios in similar macroeconomic attention deciles.

Table 3.4 reports the alphas for the double sorted portfolios along the macroeconomic attention decile, with respect to the Fama-French three factor model. Across characteristics we control for, we find that alphas consistently declines going from the lowest attention to the highest attention portfolio. Long-short portfolio constructed from the double sorted portfolios generates negative alpha that is significant at the 1 percent level.

3.3.2 Fama-MacBeth linear regressions

The Fama and MacBeth (1973) regressions for the firm i at end-of-month t takes the form of

$$r_{it} = c + \gamma log(\widehat{MacroAttn})_{it} + \lambda'_{\beta}\beta_{i,t} + \lambda'_{z}z_{i,t} + \epsilon_{it} \tag{3.4}$$

where r_{it} is the stock's excess return from month $t-1$ to month t, $log(\widehat{MacroAttn})_{it}$ is the predicted value of macroeconomic attention for month t, $\beta_{i,t}$ is a vector of factor loadings over the month t, and $z_{i,t}$ the vector of firm charactersitics available prior to month t. The regression coefficient of interest is γ, which is the price of a one-unit exposure to macroeconomic attention risk. We test whether this coefficient is significantly different from zero, which would sugges that macroeconomic attention provides incremental information for predicting returns, over and above the information in the characteristics we control for. In the Fama-MacBeth approach, this linear

	Beta		Size		BM		Agg Vol		Idio Vol	
	α	t	α	t	α	t	α	t	α	t
1	0.81***	(8.24)	0.58***	(6.83)	0.78***	(8.15)	0.83***	(8.91)	0.73***	(8.57)
2	0.72***	(7.30)	0.55***	(6.75)	0.63***	(6.52)	0.68***	(7.76)	0.60***	(6.87)
3	0.59***	(5.39)	0.52***	(4.54)	0.60***	(5.47)	0.69***	(5.44)	0.57***	(5.31)
4	0.52***	(5.06)	0.35***	(4.01)	0.43***	(4.05)	0.45***	(4.74)	0.45***	(5.46)
5	0.42***	(3.92)	0.41***	(4.10)	0.50***	(4.07)	0.44***	(4.62)	0.42***	(4.63)
6	0.43***	(4.24)	0.30***	(3.46)	0.45***	(4.69)	0.42***	(4.04)	0.43***	(4.69)
7	0.44***	(4.28)	0.40***	(4.99)	0.39***	(4.36)	0.50***	(4.46)	0.42***	(4.54)
8	0.33***	(2.88)	0.28***	(2.76)	0.35***	(2.83)	0.34***	(2.79)	0.34***	(3.07)
9	0.41***	(4.05)	0.33***	(3.41)	0.39***	(4.13)	0.37***	(3.92)	0.36***	(3.79)
10	0.30**	(2.41)	0.22**	(1.98)	0.27**	(2.16)	0.30**	(2.39)	0.30**	(2.51)
10-1	-0.51***	(-3.89)	-0.35***	(-2.64)	-0.52***	(-4.14)	-0.53***	(-3.97)	-0.43***	(-3.28)

Table 3.4: Alphas for double sorted equally-weighted portfolios on a control characteristic and macroeconomic attention.

Note: To construct the double-sorted portfolios, stocks are first sorted into two portfolios based on their value of a given control characteristic each month. For each characteristic portfolio, stocks are then sorted into ten portfolios based on their value of macroeconomic attention. The rows report the alphas of portfolios in similar macroeconomic attention deciles with respect to Fama and French (1993) three factor model. Each column is the characteristic that is controlled for. Beta is exposure to the market factor, Size is log market capitalization, BM is the log book-to-market ratio, Agg Vol is exposure to innovations in the VIX index, and Idio Vol is the idiosyncratic volatility with respect to the market factor model. Within each column, subcolumn refer to the estimated alpha and is the associated t-statistics, which incorporate Newey-West correction with four lags. The sample period is from January 2004 to May 2020. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

model is first estimated each month across firms, followed by a second stage test of whether the time series of γ estimated is significantly different from zero. The advantage of Fama and MacBeth (1973) regressions in comparison to portfolio sorts is that they allow for controls for multiple factor loadings and characteristics.

Our vector of contemporaneous factor loadings β_{it} are the loadings from the Fama and French (1993) model using the value-weighted market excess returns (MKT) , size (SMB), and value (HML) factors in addition to the loadings on the aggregate volatility ΔVIX factor (Ang et al. 2006). In addition to the factor loadings, we also include a vector of firm characteristics z_{it} known at the beginning of month t. These include log market capitalization (size), book-to-market ratios, cumulative returns from the past 12 to 2 months (lagged returns), and the stock's idiosyncratic volatility (idio vol).

Table 3.5 reports the coefficient estimates of the Fama and MacBeth (1973) regressions. Across all specifications, we find that higher macroeconomic attention in the past month predicts lower returns. The Newey-West heteroskedastic and autocorrelation robust t-statistic indicates that the slope is significant at the 1 percent level across all specifications. In Column (9), jointly controlling for the stock's loadings on relevant factor loadings and firm characteristics, we find a slope coefficient of -0.301 on macroeconomic attention. This translates to a -0.39% (-4.6%) per month (annum) decrease in average return moving from the bottom decile to the top decile of macroeconomic attention ($-0.301 \times (-2.18 - (-3.49)) = -0.39\%$).

As a robustness check, we include additional return predictors studied in Lewellen (2015). Similar to Asness, Frazzini, and Pedersen (2019), we perform a simple rank transformation of the characteristics to ensure we have sufficient observations when all control variables are included. Specifically for characteristic c_{it} of firm i in month t, we compute the cross-sectional rank as $f(c_{i,t}) = rank(c_{i,t})/(N_t + 1)$ where $N_t \equiv \max_i c_{i,t}$. We set the value of missing observations to 0, which is equivalent to setting the value of missing observations to the cross-sectional mean of the characteristic each period. This approach has the advantage of retaining information conveyed through the cross-sectional distribution of the characteristic, at the same time reducing the effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MacroAttn	-0.676***	-0.610***	-0.650***	-0.671***	-0.572***	-0.816***	-0.803***	-0.616***	-0.302***
	(-5.01)	(-4.94)	(-5.04)	(-5.11)	(-4.36)	(-6.31)	(-5.51)	(-4.96)	(-2.82)
$\beta(MKT)$	-0.346*								-0.325*
	(-1.79)								(-1.82)
$\beta(SMB)$		0.0497							-0.107
		(0.83)							(-1.28)
$\beta(HML)$			0.151*						0.214**
			(1.84)						(2.24)
$\beta(VIX)$				9.470					49.62*
				(0.30)					(1.72)
Size					-0.376***				-0.131***
					(-6.61)				(-3.41)
Book-to-market						0.114			-0.120
						(1.01)			(-1.62)
Lagged returns (12 mths)							-0.0931		0.420*
							(-0.32)		(1.97)
Idio vol								44.78***	7.930*
								(4.61)	(1.69)
Observations	312680	312660	312783	312926	313191	301129	310379	312839	274268
R^2	0.0181	0.0153	0.0167	0.0122	0.0123	0.0115	0.0136	0.0138	0.0755

t statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Note: For each month, we regress individual stock returns in excess of the one-month Treasury bill rate on a constant, predicted macroeconomic attention for month, and a set of factor exposures and characteristics as control, and then compute the average across coefficients estimated for each characteristic. R^2 reports the average cross-sectional R^2's. The sample period is from January 2005 to December 2019. t-statistics incorporate Newey-West correction with four lags.

Table 3.5: Fama-MacBeth regressions.

of outlier observations. Table 3.6 reports the coefficient estimates using this approach. The coefficient estimates on macroeconomic attention remains negative and significant at the 1 percent level.

	(1)		(2)	
MacroAttn	-0.17***	(-4.53)	-0.13***	(-4.45)
$\beta(MKT)$	-0.11	(-1.10)	-0.12	(-1.33)
$\beta(SMB)$	-0.18*	(-2.55)	-0.19**	(-2.81)
$\beta(HML)$	0.13	(1.53)	0.10	(1.26)
$\beta(VIX)$	-0.05	(-0.73)	-0.05	(-0.79)
Size	-0.44***	(-7.93)	-0.48***	(-8.30)
Book-to-market	-0.10	(-1.96)	-0.22***	(-4.17)
Lagged returns (12 mths)	0.04	(0.68)	0.03	(0.64)
Idio vol	0.26***	(4.30)	0.19***	(4.11)
Issuances (36 mths)			-0.02	(-0.65)
Accruals			0.14***	(4.72)
Return on asset			-0.29***	(-5.28)
Asset growth			-0.08**	(-3.22)
Lagged returns (12 mths)			0.08*	(2.11)
Issuances (12 mths)			-0.02	(-0.56)
Turnover			0.08	(1.52)
Sale-to-price			0.09	(1.78)
Net debt-to-price			0.07	(1.79)
Dividend yield			-0.04	(-1.05)
Observations	318152		318152	
R^2	0.0582		0.0779	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.6: Fama-MacBeth regressions, with additional predictors.

Note: For each month, we regress individual stock returns in excess of the one-month Treasury bill rate on a constant, predicted macroeconomic attention for month, and a set of factor exposures and characteristics as control, and then compute the average across coefficients estimated for each characteristic. R^2 reports the average cross-sectional R^2's. The sample period is from January 2005 to December 2019. t-statistics incorporate Newey-West correction with four lags.

3.4 Conceptual framework

In the previous section, we find that firms with higher macroeconomic attention is associated with lower returns that are both economically and statistically meaningful. If our measure of macroeconomic attention is a proxy for a firm's exposure to cash flow risk, then we expect returns of the macroeconomic attention-sorted portfolios to be fully explained by its market beta. However, we find that the abnormal returns of the portfolios to persist even after controlling for exposure to market risk as well as other characteristics and factors known to predict returns. In this section, we develop a risk-based explanation to account for the abnormal returns associated with macroeconomic attention. Through the lens of a model of optimal attention allocation, we show that optimal attention to the macroeconomy varies with a firm's exposure to aggregate and idiosyncratic cash flow risk. We then use the returns decomposition framework of Campbell (1991) to explain how firms with greater macroeconomic attention earn lower returns.

3.4.1 Modeling attention allocation

We build on Maćkowiak and Wiederholt (2009) and outline a model where analysts learn about an aggregate and firm-specific shock with the objective of minimizing forecast errors on future dividends. Information about these shocks is conveyed through the content of the call, which is naturally limited by the length of the call and the finite attention span of analysts listening in on the call. Prior literature offers evidence of such limitations, including underreaction of prices to earnings announcements made on Fridays as opposed to other weekdays (Dellavigna and Pollet, 2009) and lower consumption of earnings calls on days where more firms hold calls (Heinrichs, Park, and Soltes, 2018).

We consider an analyst covering firm i who receives payoffs $-(d_{it+1} - \hat{d}_{it+1})^2$ at the end of period t, where $\hat{d}_{it+1}$ is the forecast of dividends in period $t+1$, d_{it+1}. Dividends are revealed at the end of each period, hence d_{it} is observed when the analyst forecasts d_{it+1}. We assume that firm i's dividend growth follows the process

$$\Delta d_{i,t+1} = z_{it} + u_{it+1} \tag{3.5}$$

$$u_{it+1} = \eta_{t+1} + \nu_{it+1}$$

where z_{it} are predictors of the firm's dividend growth in $t+1$ known at time t, and the unexpected component u_{it+1} is the sum of an aggregate component $\eta_{t+1} \sim N(0, \sigma_\eta^2)$ and an idiosyncratic component $\nu_{i,t+1} \sim N(0, \varphi_i \sigma_\nu^2)$. Here, φ_i is a parameter that scales the firm-specific component of dividend growth for firm i. In period t, the analyst attends firm i's earnings call and receives signals s_{it} given by

$$s_{it}^\eta = \eta_{t+1} + \epsilon_{it}^\eta$$

$$s_{it}^\nu = \nu_{i,t+1} + \epsilon_{it}^\nu$$

where signal noises are distributed as $\epsilon_{it}^\eta \sim N(0, \sigma_{\epsilon,\eta}^2)$ and $\epsilon_{it}^\nu \sim N(0, \sigma_{\epsilon,\nu}^2)$, which are mutually independent with each other and with η_{t+1} and $\nu_{i,t+1}$ for all i and t. There is an active discussion in the literature on whether attention to different sources of shocks are reasonably modeled as separate activities (see Afrouzi, 2016; Miao and Su, 2019). We adopt this assumption simply because we are able to identify discussions of macroeconomy separately from firm-specific conditions in earnings calls.

The analyst for firm i to choose which signal to pay more attention to, subject to a limitation on the information that the analyst is able to learn from earnings calls. This could either arise from limitations in the analyst's information processing capacity or a limitation on the firm manager's ability to convey precise information about the firm's future cash flows. Formally, the analyst's problem is to choose signal noise precision $(\sigma_{\epsilon,\eta}^2, \sigma_{\epsilon,\nu}^2)$ to minimize forecast error on the firm's dividends. Since dividends are revealed at the end of each period, we can rewrite the analyst's payoffs as $-(\Delta d_{it+1} - \Delta\hat{d}_{it+1})^2$. We assume that the analyst have priors that are identical to the actual data generating process. That is, her priors over unobserved shocks η_{t+1} and $\nu_{i,t+1}$ are given by $\eta_{t+1} \sim N(0, \sigma_\eta^2)$ and $\nu_{i,t+1} \sim N(0, \varphi_i \sigma_\nu^2)$. As a Bayesian, the optimal update of earnings expectations, given the specified prior and signal structure is given by

$$\Delta \hat{d}_{it+1} = z_{it} + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_{\epsilon,\eta}^2} s_{it}^\eta + \frac{\varphi_i \sigma_\nu^2}{\varphi_i \sigma_\nu^2 + \sigma_{\epsilon,\nu}^2} s_{it}^\nu$$

As standard in the rational inattention literature, we assume a mutual information constraint on information processing capacity. Specifically, the analyst solves the optimization problem

$$\max_{\sigma_{\epsilon,\eta}^2, \sigma_{\epsilon,\nu}^2} -E_t\left[(\Delta d_{it+1} - \Delta \hat{d}_{it+1})^2\right]$$

subject to information processing constraint

$$\underbrace{\frac{1}{2}\log_2\left(1 + \frac{\sigma_\eta^2}{\sigma_{\epsilon,\eta}^2}\right)}_{\kappa_\eta} + \underbrace{\frac{1}{2}\log_2\left(1 + \frac{\varphi_i \sigma_\nu^2}{\sigma_{\epsilon,\nu}^2}\right)}_{\kappa_\nu} \leq \kappa$$

The analyst's problem is similar to the setup in Section IV of Maćkowiak and Wiederholt (2009), and the unique solution to the analyst's attention problem is given by

$$\kappa_{\eta,i} = \begin{cases} \kappa & \text{if } x_i \geq 2^{2\kappa} \\ \frac{1}{2}\kappa + \frac{1}{4}\log_2(x_i) & \text{if } x_i \in [2^{-2\kappa}, 2^{2\kappa}] \\ 0 & \text{if } x_i \leq 2^{-2\kappa} \end{cases}$$

where $x_i = \sigma_\eta^2 / (\varphi_i \sigma_\nu^2)$.

Definition 1 *The analyst's attention to macroeconomic conditions in the earnings call of firm i, $MacroAttn_i$, is the share of information about the aggregate component of cash flows relative to* *total information about cash flows*

$$MacroAttn_i \equiv \frac{\kappa_{\eta,i}}{\kappa} = \frac{1}{2} + \frac{1}{4\kappa}\log_2\left(\frac{\sigma_\eta^2}{\varphi_i \sigma_\nu^2}\right) \tag{3.6}$$

3.4.2 Return decomposition

Denote $r_{i,t}$ to be the log market return of asset i in time t, and $d_{i,t}$ is the dividends of firms in period t. Using the Campbell (1991) return decomposition, we can characterize the unexpected log return on stock i, $r_{i,t+1} - E_t r_{i,t+1}$ as two components, revisions in expected future dividends $N^{CF}_{i,t+1}$ and revisions in discount rates $N^{DR}_{i,t+1}$

$$r_{i,t+1} - E_t r_{i,t+1} = N^{CF}_{i,t+1} - N^{DR}_{i,t+1}$$

where $N^{CF}_{i,t+1} = (E_{t+1} - E_t)\sum_{j=0}^{\infty} \rho^j d_{i,t+1+j}$ and $N^{DR}_{i,t+1} = (E_{t+1} - E_t)\sum_{j=1}^{\infty} \rho^j r_{i,t+1+j}$.

Using the definition of dividends $d_{i,t}$ from 3.5 and given the assumption that analysts do not receive signals about dividends beyond the next period $t+1$ from earnings calls, cash flow news is given by

$$N^i_{CF,t+1} \equiv (E_{t+1} - E_t)\sum_{j=0}^{\infty} \rho^j d_{i,t+1+j} = \eta_{t+1} + \nu_{i,t+1}$$

where $\eta_{t+1} \sim N(0, \sigma_\eta^2)$ and $\nu_{i,t+1} \sim N(0, \varphi_i \sigma_\nu^2)$. We further assume that each firm's discount rate news is uncorrelated with the cash flow news, and is uncorrelated across firms, formally $N^{DR}_{i,t+1} \sim_i N\left(0, \sigma_\omega^2\right)$.

Definition 2 *The diversified cash flow risk of firm i, D_i, is the share of variance of cash flow news attributed to the aggregate component*

$$D_i = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \varphi_i \sigma_\nu^2} \tag{3.7}$$

The higher the share of cash flow risk that is diversified, the lower the exposure of an investor who holds asset i to its idiosyncratic cash flow risk. Under optimal attention allocation as described *in Section 3.4.1, stocks with a greater share of diversified cash flow risk is associated with higher attention to the macroeconomy.*

Corollary 3 *Attention to macroeconomic conditions is higher for firms with greater diversified cash flows, D_i.*

Proof. *See Appendix B.1.*

Consider a market portfolio constructed as an equal-weighted portfolio consisting of all assets in the economy with returns $r_{m,t+1} = 1/M \sum_{i=1}^{M} r_{i,t+1}$.[10] The unexpected log return on the market portfolio can be similarly written as $r_{m,t+1} - E_t r_{m,t+1} = N_{m,t+1}^{CF} - N_{m,t+1}^{DR}$. The cash flow news component of the market portfolio is given by $N_{m,t+1}^{CF} = \frac{1}{M} \sum_{i=1}^{M} N_{i,t+1}^{CF}$and the discount rate component is $N_{m,t+1}^{DR} = \frac{1}{M} \sum_{i=1}^{M} N_{i,t+1}^{DR} \sim N\left(0, \frac{1}{M}\sigma_{\omega}^{2}\right)$. Given a representative investor with Epstein-Zin preferences and who holds the market portfolio Campbell and Vuolteenaho (2004), the risk premium for any stock i is given by

$$rp_{i,t} = \gamma Cov(r_{i,t+1} - E_t r_{i,t+1}, N_{m,t+1}^{CF}) + Cov(r_{i,t+1} - E_t r_{i,t+1}, -N_{m,t+1}^{DR}) \tag{3.8}$$

where γ is risk aversion coefficient. As the diversified cash flow risk D_i increases (φ_i decreases), the risk premium of stock i decreases.

Corollary 4 *The risk premium on asset i's returns is decreasing in the share of diversified cash flow risk.*

Proof. *See Appendix B.1.*

Following Campbell and Vuolteenaho (2004), we define cash flows and discount rate betas of

[10]Under the assumption that the size of a stock is uncorrelated to its idiosyncratic variance, this is equivalent to a value-weighted portfolio of all assets.

asset i as

$$\beta_{i,m}^{DR} = \frac{Cov(r_{i,t+1} - E_t r_{i,t+1}, -N_{M,t+1}^{DR})}{Var(r_{M,t+1})}$$
$$\beta_{i,m}^{CF} = \frac{Cov(r_{i,t+1} - E_t r_{i,t+1}, N_{M,t+1}^{CF})}{Var(r_{M,t+1})} \tag{3.9}$$

whcih sums up to the total market beta $\beta_{i,m} = \beta_{i,m}^{CF} + \beta_{i,m}^{DR}$. Using this definitions, we rewrite the risk premium of asset i in (3.8) as

$$rp_i = \gamma\sigma_m^2\beta_{i,m}^{CF} + \sigma_m^2\beta_{i,m}^{DR} \tag{3.10}$$

The insight from Campbell and Vuolteenaho (2004) is that the price of risk on the cash flow beta $\beta_{i,m}^{CF}$ is γ times greater that the price of risk on the discount rate beta $\beta_{i,m}^{DR}$, where γ is the risk aversion coefficient that is greater than 1. Hence, the composition of the market betas matter. Consider two assets A and B with identical market betas $\beta_{i,m}$. If asset A has higher cash flow beta $\beta_{i,m}^{CF}$ than asset B, asset A will earn a higher risk premia than asset B, which will not be explained by a single-factor market beta model. Hence, Campbell and Vuolteenaho (2004) terms cash flow beta $\beta_{i,m}^{CF}$ as the bad beta and discount rate beta $\beta_{i,m}^{DR}$ as the good beta, given that the latter has a lower price of risk.

Given our assumption about the data generating process for cash flow news, we have the following testable implication.

Proposition 5 *Stocks with higher diversification factor have lower cash flow betas, and no systematic difference in discount rate betas. The risk premium unaccounted for the stock's market beta (the CAPM alpha) is decreasing in the stock's cash flow beta.*

Proof. *See Appendix B.1.*

3.4.3 Measuring cash flow betas

In this section, we empirically evaluate the testable implication of the model outlined above, namely that the stocks with larger macroeconomic attention have lower cash flow risk due to a lower share of firm-specific risk. The estimation strategy closely follows Vuolteenaho (2002), Campbell and Vuolteenaho (2004), and Campbell, Polk, and Vuolteenaho (2009). The data generating process follows a first-order VAR model

$$z_{t+1} = a + \Gamma z_t + u_{t+1} \tag{3.11}$$

where z_{t+1} is an $m \times 1$ state vector with return r_{t+1} as the first element, a and Γ are respectively a $m \times 1$ vector and $m \times m$ matrix of VAR parameters to be estimated, and u_{t+1} is an $m \times 1$ vector of shocks with variance-covariance Σ. We specify the remaining state variables in the VAR specification shortly below. Given the above specification, the $t + 1$ cash flow and discount rate news are given by

$$\begin{aligned} N_{t+1}^{DR} &= \lambda' u_{t+1} \\ N_{t+1}^{CF} &= (e1' + \lambda') u_{t+1} \end{aligned}$$

where $e1$ is a vector whose first element equals unity and the remaining elements are zero, and $\lambda' \equiv e1' \rho \Gamma (I - \rho \Gamma)^{-1}$. In words λ captures the effect of each individual VAR shock on the discount-rate expectations, which increases in the value of coefficient in Γ as well as its persistence $(I - \rho \Gamma)^{-1}$. Following Campbell, Polk, and Vuolteenaho (2009), we estimate separate VARs for aggregate news and firm-specific news and set $\rho = 0.95$. This approach is consistent with the empirical literature documenting different sources of risk driving returns in the aggregate and cross-section. We estimate both VARs using annual data over the sample period from 1928 to 2019, and treat the parameters (a, Γ, Σ) as constant across the sample period. As such, the parameters are estimated by separate equation-by-equation pooled regression.

	var(DR)	var(CF)	-2cov(DR,CF)
Market return	59	16	25
Firm market-adjusted return	1	88	10
Portfolio market-adjusted return	0	96	4

Table 3.7: Variance decomposition of market, firm, and portfolio-level unexpected annual returns.

Note: Firm returns are market-adjusted by subtracting the market return from individual firm returns, and portfolio returns are the simple average of firm market-adjusted returns of stocks within each portfolio. Returns are decomposed into discount rate news (DR) and cash flow news (CF) using the panel VAR specified in 3.11 estimated using annual returns from 1928 to 2019. Variance decomposition is reported for the sample period from 2005 to 2019 for which we observe macroeconomic attention data.

The aggregate VAR consists of four state variables ($m = 4$): the log market return, the term yield spread, log smoothed price-earnings ratio, and the small-stock value spread. The firm-level VAR consists of three state variables ($m = 3$): the log return (r_i) on a firm's common stock equity, log book-to-market ratio of unlevered equity, and the long term profitability. All variables in the firm-level VAR are cross-sectionally demeaned by subtracting the value-weighted market return in the case of firm-returns and average value of each variable each year for the other state variables.

From the cash flow and discount rate news components of individual stock returns, we proceed to estimating the constructing the cash flow and news series of the macroeconomic attention portfolios. Portfolios are constructed by sorting stocks into deciles based on the average value of their macroeconomic attention each year. We first compute the portfolio level market-adjusted news series by taking a simple average of the respective news components of individual firms within each portfolio. Given that we only have macroeconomic attention data over the sample period for which we observe a firm's earnings calls, we have 15 annual observations of news for each portfolio over the sample period from 2005 to 2019.[11] In subsequent analysis, we report robust standard errors constructed from 1000 bootstrap samples of the observations.

Variance decomposition

Table 5 reports the share of variance explained by cash flow and discount rate news, as well as two times the covariance between the two news components. For the market portfolio consisting of a value-weighted portfolio of all stocks, we find that 59% of total return variation is attributed to discount rate news, 16% to cash flow news, and the remaining 25% to the covariance component. This result is consistent with the findings in the prior literature such as Campbell (1991). The middle row reports the decomposition for firm-level market adjusted returns. In our sample period from 2005 to 2019, we find that cash flow news explains 88% of firm-level market-adjusted returns, with a much smaller share attributed to discount rate news. This result again is consistent with the findings of Vuolteenaho (2002).

In the last row, we show the decomposition for portfolio-level market adjusted returns. Interestingly, we find that cash flow news explains a much larger share of returns relative to firm-level market adjusted returns, accounting for 96% of total return variation. We interpret this result as consistent with the explanation that our anomaly returns is largely due to cash flow shocks, rather than shocks to investor behavior that affects prices but does not affect the earnings of firms. This result is also consistent with the findings from Lochstoer and Tetlock (2020), who show that cash flow news account for most of the variation in returns across key anomaly portfolios including value, size, profitability, investment, and momentum.

Portfolio betas and its components

Our theoretical framework suggests that stocks with higher macroeconomic attention are better diversified stocks, hence expose investors to lower firm-specific cash flow risk. As described in the previous section, portfolios with higher macroeconomic attention should have lower betas with respect to cash flow news. In this subsection, we examine the prediction by computing the betas of the portfolios using the news series estimated from the VAR model. The betas are constructed as

[11] The VAR model from which the news components are estimated have parameters estimated on a longer sample of annual data from 1929 to 2019.

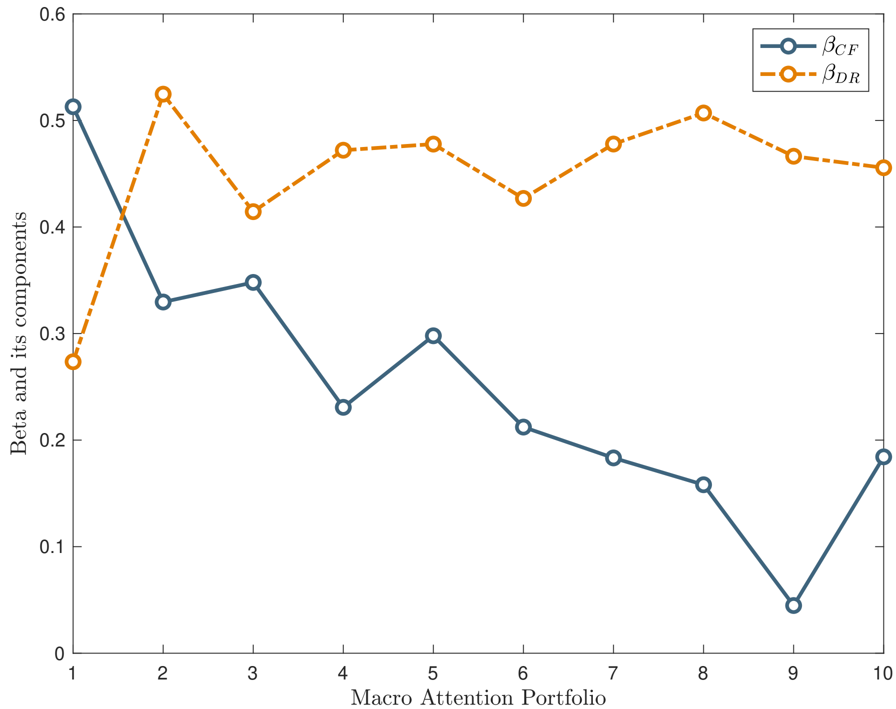


Figure 3.3: Cash flow and discount rate betas of macroeconomic attention portfolios.

Note: Betas are constructed as in (3.9) by regressing the unexpected returns of each portfolio on the cash flow news and discount rate news of the market portfolio. The sample period is from January 2005 to December 2019.

in (3.9), by regressing the respective portfolio's news components on a constant and the cash flow and discount rate news components of the market portfolio.

Figure 3.3 shows the cash flow and discount rate betas for each portfolio. Portfolio 1 are stocks in the lowest decile of macroeconomic attention each year, whereas Portfolio 10 are stocks in the highest decile. The key takeaway from the figure is that, as our theoretical framework would suggest, the cash flow betas of portfolios declines as macroeconomic attention increases, whereas we do not observe systematic patterns in discount rate betas in our sample.[12] Overall, our results provide some evidence that is consistent with our explanation (Prediction 1) that the risk premia of our macroeconomic attention portfolio captures varying exposures to the macroeconomic and idiosyncratic components of cash flow risk.

[12]Table B.4 in the Appendix provides the beta components estimates with relevant t-statistics. While we observe a systematic sorting of cash flow betas consistent with the prediction of our model, the difference across portfolios is not statistically significant. This is likely due to limited sample availability given that betas of each portfolio are computed only from 15 annual observations.

3.5 Conclusion

The idea that economic agents choose what to pay attention to has important implications for a variety of economic phenomena. In this chapter, we focus on implications of macroeconomic attention on the cross section of stock returns. We quantify the amount of attention to the macroeconomy at the firm-level from the text of earnings calls transcripts, and find that this measure strongly predicts returns of stocks in the cross section. In particular, we find that firms associated with higher attention to the macroeconomy earn on average lower returns than those associated with lower attention to the macroeconomy.

We provide a risk-based explanation for the negative risk premia that we observe. In a model where analysts allocate attention optimally to learn about macroeconomic and firm-specific cash flow news, attention to the macroeconomy is increasing in the share of cash flow news variation explained by macroeconomic risk. All else equal, firms with higher macroeconomic attention exposes investors to less firm-specific fundamental risk. As such, investors demand higher premia for stocks with lower macroeconomic attention. We find empirical support that portfolios of stocks with higher macroeconomic attention have lower cash flow risk, hence justifying their lower expected returns.

References

Abarbanell, Jeffery S. (June 1991). "Do analysts' earnings forecasts incorporate information in prior stock price changes?" In: *Journal of Accounting and Economics* 14.2, pp. 147–165.

Abarbanell, Jeffery S and Victor L Bernard (1992). "Tests of Analysts' Overreaction/Underreaction to Earnings Information as an Explanation for Anomalous Stock Price Behavior". In: *Journal of Finance*, p. 28.

Afrouzi, Hassan (2016). "Endogenous Firm Competition and Cyclicality of Markups". In: *Working Paper*. Publisher: Columbia University.

Afrouzi, Hassan, Spencer Yongwook Kwon, Augustin Landier, Yueran Ma, and David Thesmar (Oct. 2020). *Overreaction and Working Memory*. w27947. Cambridge, MA: National Bureau of Economic Research, w27947.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang (Jan. 2006). "The Cross-Section of Volatility and Expected Returns". In: *The Journal of Finance* 61.1. Publisher: Wiley, pp. 259–299.

Asness, Clifford S, Andrea Frazzini, and Lasse Heje Pedersen (2019). "Quality minus junk". In: *Review of Accounting Studies* 24.1. Publisher: Springer, pp. 34–112.

Asness, Clifford S, Tobias J Moskowitz, and Lasse Heje Pedersen (2013). "Value and Momentum Everywhere". In: *Journal of Finance* 68.3, pp. 929–985.

Asquith, Paul, Michael B Mikhail, and Andrea S Au (2005). "Information content of equity analyst reports". In: *Journal of Financial Economics*, p. 38.

Axtell, Robert L. (Sept. 7, 2001). "Zipf Distribution of U.S. Firm Sizes". In: *Science* 293.5536, pp. 1818–1820.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer (2015). "X-CAPM: An Extrapolative Capital Asset Pricing Model". In: *Journal of Financial Economics* 115, pp. 1–24.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer (Aug. 2018). *Over-reaction in Macroeconomic Expectations*. w24932. Cambridge, MA: National Bureau of Economic Research, w24932.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer (Dec. 2019). "Diagnostic Expectations and Stock Returns". In: *The Journal of Finance* 74.6, pp. 2839–2874.

Bottazzi, Giulio and Angelo Secchi (June 2006). “Explaining the distribution of firm growth rates”. In: *The RAND Journal of Economics* 37.2, pp. 235–256.

Bouchaud, Jean-Philippe, Philipp Krueger, Augustin Landier, and David Thesmar (2019). “Sticky Expectations and the Profitability Anomaly”. In: *Journal of Finance* 74.2. Publisher: Wiley Online Library, pp. 639–674.

Bradshaw, Mark T., Lawrence D. Brown, and Kelly Huang (Dec. 2013). “Do sell-side analysts exhibit differential target price forecasting ability?” In: *Review of Accounting Studies* 18.4, pp. 930–955.

Bradshaw, Mark Thomas (2011). “Analysts’ Forecasts: What Do We Know after Decades of Work?” In: *SSRN Electronic Journal*.

Brav, Alon and Reuven Lehavy (Oct. 2003). “An Empirical Analysis of Analysts’ Target Prices: Short-term Informativeness and Long-term Dynamics”. In: *The Journal of Finance* 58.5, pp. 1933–1967.

Campbell, John Y. (Mar. 1991). “A Variance Decomposition for Stock Returns”. In: *The Economic Journal* 101.405. Publisher: Oxford University Press (OUP), p. 157.

Campbell, John Y. and Jianping Mei (July 1993). “Where Do Betas Come From? Asset Price Dynamics and the Sources of Systematic Risk”. In: *Review of Financial Studies* 6.3. Publisher: Oxford University Press (OUP), pp. 567–592.

Campbell, John Y., Christopher Polk, and Tuomo Vuolteenaho (Apr. 2009). “Growth or Glamour? Fundamentals and Systematic Risk in Stock Returns”. In: *Review of Financial Studies* 23.1. Publisher: Oxford University Press (OUP), pp. 305–344.

Campbell, John Y. and Robert J. Shiller (July 1988). “The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors”. In: *Review of Financial Studies* 1.3, pp. 195–228.

Campbell, John Y. and Tuomo Vuolteenaho (Nov. 2004). “Bad Beta, Good Beta”. In: *American Economic Review* 94.5. Publisher: American Economic Association, pp. 1249–1275.

Carhart, Mark M. (Mar. 1997). “On Persistence in Mutual Fund Performance”. In: *The Journal of Finance* 52.1, pp. 57–82.

Chen, Qihui, Nikolai Roussanov, and Xiaoliang Wang (Dec. 13, 2021). “Semiparametric Conditional Factor Models: Estimation and Inference”. In: *arXiv:2112.07121 [econ, math, stat]*. arXiv: `2112.07121`.

Chen, Shuping, Dawn Matsumoto, and Shiva Rajgopal (Feb. 2011). “Is silence golden? An empirical analysis of firms that stop giving quarterly earnings guidance”. In: *Journal of Accounting and Economics* 51.1. Publisher: Elsevier BV, pp. 134–150.

Chen, Tao (Mar. 2020). “Does news affect disagreement in global markets?” In: *Journal of Business Research* 109, pp. 174–183.

Cohen, Lauren and Christopher J. Malloy (2016). “Playing Favorites: How Firms Prevent the Revelation of Bad News”. In: *SSRN Electronic Journal*. Publisher: Elsevier BV.

Cohen, Randolph B., Christopher Polk, and Tuomo Vuolteenaho (Apr. 2003). “The Value Spread”. In: *The Journal of Finance* 58.2, pp. 609–641.

Coibion, Olivier and Yuriy Gorodnichenko (Aug. 1, 2015). “Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts”. In: *American Economic Review* 105.8, pp. 2644–2678.

Coibion, Olivier, Yuriy Gorodnichenko, and Saten Kumar (Sept. 2018). “How Do Firms Form Their Expectations? New Survey Evidence”. In: *American Economic Review* 108.9, pp. 2671–2713.

Daniel, Kent and Tobias J. Moskowitz (Nov. 2016). “Momentum crashes”. In: *Journal of Financial Economics* 122.2, pp. 221–247.

De Bie, Tijl, Nello Cristianini, and Roman Rosipal (2005). “Eigenproblems in Pattern Recognition”. In: Corrochano, Eduardo Bayro. *Handbook of Geometric Computing*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 129–167. ISBN: 978-3-540-20595-1 978-3-540-28247-1.

Dellavigna, Stefano and Joshua M. Pollet (Apr. 2009). “Investor Inattention and Friday Earnings Announcements”. In: *The Journal of Finance* 64.2, pp. 709–749.

Diebold, Francis X and Robert S Mariano (Jan. 2002). “Comparing Predictive Accuracy”. In: *Journal of Business & Economic Statistics* 20.1, pp. 134–144.

Ehsani, Sina and Juhani Linnainmaa (Feb. 2019). *Factor Momentum and the Momentum Factor*. w25551. Cambridge, MA: National Bureau of Economic Research, w25551.

Falck, Antoine, Adam Rej, and David Thesmar (2020). “Is Factor Momentum More than Stock Momentum?” In: *SSRN Electronic Journal*.

Fama, Eugene F. and Kenneth R. French (Oct. 1988). “Dividend yields and expected stock returns”. In: *Journal of Financial Economics* 22.1, pp. 3–25.

— (June 1992). “The Cross-Section of Expected Stock Returns”. In: *The Journal of Finance* 47.2, pp. 427–465.

— (Feb. 1993). “Common risk factors in the returns on stocks and bonds”. In: *Journal of Financial Economics* 33.1. Publisher: Elsevier BV, pp. 3–56.

Fama, Eugene F. and Kenneth R. French (Mar. 1996). "Multifactor Explanations of Asset Pricing Anomalies". In: *The Journal of Finance* 51.1, pp. 55–84.

— (Apr. 2015). "A five-factor asset pricing model". In: *Journal of Financial Economics* 116.1, pp. 1–22.

Fama, Eugene F. and James D. MacBeth (May 1973). "Risk, Return, and Equilibrium: Empirical Tests". In: *Journal of Political Economy* 81.3, pp. 607–636.

Farmer, Leland E., Emi Nakamura, and Jón Steinsson (2021). "Learning About the Long Run". In: *SSRN Electronic Journal*.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu (June 2020). "Taming the Factor Zoo: A Test of New Factors". In: *The Journal of Finance* 75.3, pp. 1327–1370.

Flynn, Joel P. and Karthik Sastry (2020). "Attention Cycles". In: *SSRN Electronic Journal*. Publisher: Elsevier BV.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber (May 1, 2020). "Dissecting Characteristics Nonparametrically". In: *The Review of Financial Studies* 33.5. Ed. by Andrew Karolyi, pp. 2326–2377.

Fuster, Andreas, Benjamin Hebert, and David Laibson (2012). "Natural Expectations, Macroeconomic Dynamics, and Asset Pricing". In: *NBER Macroeconomics Annual* 26.1. Publisher: University of Chicago Press Chicago, IL, pp. 1–48.

Fuster, Andreas, David Laibson, and Brock Mendel (2010). "Natural Expectations and Macroeconomic Fluctuations". In: *Journal of Economic Perspectives* 24.4. Publisher: American Economic Association, pp. 67–84.

Gabaix, Xavier (Sept. 2, 2009). "Power Laws in Economics and Finance". In: *Annual Review of Economics* 1.1, pp. 255–294.

— (2018). "Behavioral Inattention". In: *Handbook of Behavioral Economics*.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy (Sept. 1, 2019). "Text as Data". In: *Journal of Economic Literature* 57.3, pp. 535–574.

Goldberg, Yoav (Nov. 20, 2016). "A Primer on Neural Network Models for Natural Language Processing". In: *Journal of Artificial Intelligence Research* 57, pp. 345–420.

Goyal, Amit and Ivo Welch (May 2003). "Predicting the Equity Premium with Dividend Ratios". In: *Management Science* 49.5, pp. 639–654.

Goyal, Amit and Ivo Welch (July 2008). “A Comprehensive Look at The Empirical Performance of Equity Premium Prediction”. In: *Review of Financial Studies* 21.4, pp. 1455–1508.

Green, Jeremiah, John R. M. Hand, and X. Frank Zhang (Dec. 1, 2017). “The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns”. In: *The Review of Financial Studies* 30.12, pp. 4389–4436.

Greenwood, Robin and Andrei Shleifer (Mar. 2014). “Expectations of Returns and Expected Returns”. In: *Review of Financial Studies* 27.3, pp. 714–746.

Groysberg, Boris, Paul M. Healy, and David A. Maber (Sept. 2011). “What Drives Sell-Side Analyst Compensation at High-Status Investment Banks?: analyst compensation at high-status banks”. In: *Journal of Accounting Research* 49.4, pp. 969–1000.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu (May 1, 2020). “Empirical Asset Pricing via Machine Learning”. In: *The Review of Financial Studies* 33.5. Ed. by Andrew Karolyi, pp. 2223–2273.

— (May 2021). “Autoencoder asset pricing models”. In: *Journal of Econometrics* 222.1, pp. 429–450.

Haddad, Valentin, Serhiy Kozak, and Shrihari Santosh (May 1, 2020). “Factor Timing”. In: *The Review of Financial Studies* 33.5. Ed. by Stijn Van Nieuwerburgh, pp. 1980–2018.

Hansen, Stephen, Michael McMahon, and Andrea Prat (May 1, 2018). “Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach*”. In: *The Quarterly Journal of Economics* 133.2, pp. 801–870.

Hassan, Tarek A., Stephan Hollander, Laurence van Lent, and Ahmed Tahoun (Aug. 2019). “Firm-Level Political Risk: Measurement and Effects\ast”. In: *The Quarterly Journal of Economics* 134.4. Publisher: Oxford University Press (OUP), pp. 2135–2202.

Hassan, Tarek Alexander, Jesse Schreger, Markus Schwedeler, and Ahmed Tahoun (Nov. 2021). *Sources and Transmission of Country Risk*. w29526. Cambridge, MA: National Bureau of Economic Research, w29526.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY: Springer New York. ISBN: 978-0-387-84857-0 978-0-387-84858-7.

Heinrichs, Anne, Jihwon Park, and Eugene F. Soltes (Aug. 2018). “Who Consumes Firm Disclosures? Evidence from Earnings Conference Calls”. In: *The Accounting Review* 94.3. Publisher: American Accounting Association, pp. 205–231.

Hong, Harrison and Jeremy C Stein (1999). “A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets”. In: *The Journal of Finance*, p. 42.

Hong, Harrison and Jeremy C Stein (2007). "Disagreement and the Stock Market". In: *Journal of Economic Perspectives*, p. 38.

Hornik, Kurt, Maxwell Stinchcombe, and Halbert White (Jan. 1989). "Multilayer feedforward networks are universal approximators". In: *Neural Networks* 2.5. Publisher: Elsevier BV, pp. 359–366.

Hou, Kewei, Chen Xue, and Lu Zhang (May 1, 2020). "Replicating Anomalies". In: *The Review of Financial Studies* 33.5, pp. 2019–2133.

Jegadeesh, Narasimhan (1990). "Evidence of Predictable Behavior of Security Returns". In: *Journal of Finance*.

Jegadeesh, Narasimhan and Sheridan Titman (1993). "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency". In: *Journal of Finance* 48.1, pp. 65–91.

Jurado, Kyle, Sydney C. Ludvigson, and Serena Ng (Mar. 2015). "Measuring Uncertainty". In: *American Economic Review* 105.3, pp. 1177–1216.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp (2016). "A Rational Theory of Mutual Funds' Attention Allocation". In: *Econometrica* 84.2, pp. 571–626.

Kelly, Bryan and Seth Pruitt (Oct. 2013). "Market Expectations in the Cross-Section of Present Values: Market Expectations in the Cross-Section of Present Values". In: *The Journal of Finance* 68.5, pp. 1721–1756.

— (June 2015). "The three-pass regression filter: A new approach to forecasting using many predictors". In: *Journal of Econometrics* 186.2, pp. 294–316.

Kelly, Bryan T., Seth Pruitt, and Yinan Su (Dec. 2019). "Characteristics are covariances: A unified model of risk and return". In: *Journal of Financial Economics* 134.3, pp. 501–524.

Kingma, Diederik P. and Jimmy Ba (Jan. 29, 2017). *Adam: A Method for Stochastic Optimization*. arXiv: `1412.6980[cs]`.

Kothari, S.P., Eric So, and Rodrigo Verdi (Oct. 23, 2016). "Analysts' Forecasts and Asset Pricing: A Survey". In: *Annual Review of Financial Economics* 8.1, pp. 197–219.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh (Feb. 2020). "Shrinking the cross-section". In: *Journal of Financial Economics* 135.2, pp. 271–292.

Kozlowski, Julian, Laura Veldkamp, and Venky Venkateswaran (2015). "The Tail that Wags the Economy: Belief-Driven Business Cycles and Persistent Stagnation". In: *SSRN Electronic Journal*.

Kumar, Saten, Hassan Afrouzi, Olivier Coibion, and Yuriy Gorodnichenko (Dec. 2015). *Inflation Targeting Does Not Anchor Inflation Expectations: Evidence from Firms in New Zealand*. w21814. Cambridge, MA: National Bureau of Economic Research, w21814.

La Porta, Rafael (Dec. 1996). "Expectations and the Cross-Section of Stock Returns". In: *The Journal of Finance* 51.5, pp. 1715–1742.

Lewellen, Jonathan (June 2015). "The Cross-section of Expected Stock Returns". In: *Critical Finance Review* 4.1. Publisher: Now Publishers, pp. 1–44.

Lewellen, Jonathan, Stefan Nagel, and Jay Shanken (2010). "A skeptical appraisal of asset pricing tests". In: *Journal of Financial Economics*.

Li, Feng (2010). "Survey of the Literature". In: *Journal of accounting literature* 29, pp. 143–165.

Lochstoer, Lars A. and Paul C. Tetlock (Feb. 2020). "What Drives Anomaly Returns?" In: *The Journal of Finance* 75.3. Publisher: Wiley, pp. 1417–1455.

Loughran, Tim and Bill McDonald (Feb. 2011). "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks". In: *The Journal of Finance* 66.1, pp. 35–65.

— (June 2016). "Textual Analysis in Accounting and Finance: A Survey". In: *Journal of Accounting Research* 54.4. Publisher: Wiley, pp. 1187–1230.

Ludvigson, Sydney C. and Serena Ng (Jan. 2007). "The empirical risk-return relation: A factor analysis approach". In: *Journal of Financial Economics* 83.1, pp. 171–222.

Ma, Yueran, Tiziano Ropele, David Sraer, and David Thesmar (Mar. 2020). *A Quantitative Analysis of Distortions in Managerial Forecasts*. w26830. Cambridge, MA: National Bureau of Economic Research, w26830.

Mackowiak, Bartosz, Filip Matejka, and Mirko Wiederholt (2018). *Rational inattention: A disciplined behavioral model*. CEPR Discussion Paper.

Matsumoto, Dawn, Maarten Pronk, and Erik Roelofsen (2011). "What Makes Conference Calls Useful? The Information Content of Managers' Presentations and Analysts' Discussion Sessions". In: *The Accounting Review* 86.4. Publisher: American Accounting Association, pp. 1383–1414.

Maćkowiak, Bartosz, Emanuel Moench, and Mirko Wiederholt (2009). "Sectoral price data and models of price setting". In: *Journal of Monetary Economics* 56 (S), S78–S99.

Maćkowiak, Bartosz and Mirko Wiederholt (May 2009). "Optimal Sticky Prices under Rational Inattention". In: *American Economic Review* 99.3. Publisher: American Economic Association, pp. 769–803.

Mclean, R. David and Jeffrey Pontiff (Feb. 2016). “Does Academic Research Destroy Stock Return Predictability?: Does Academic Research Destroy Stock Return Predictability?” In: *The Journal of Finance* 71.1, pp. 5–32.

Miao, Jianjun and Dongling Su (July 2019). *Asset Market Equilibrium under Rational Inattention*. Boston University - Department of Economics - Working Papers Series WP2019-09. Boston University - Department of Economics.

Miller, Edward M. (1977). “Risk, Uncertainty, and Divergence of Opinion”. In: *Journal of Finance*, p. 19.

Novy-Marx, Robert and Mihail Velikov (Jan. 2016). “A Taxonomy of Anomalies and Their Trading Costs”. In: *Review of Financial Studies* 29.1, pp. 104–147.

O, Ricardo De la, Xiao Han, and Sean Myers (2022). “The Cross-section of Subjective Expectations: Understanding Prices and Anomalies”. In: *SSRN Electronic Journal*.

O, Ricardo De la and Sean Myers (June 2021). “Subjective Cash Flow and Discount Rate Expectations”. In: *The Journal of Finance* 76.3, pp. 1339–1387.

Porter, Martin F (1980). “An algorithm for suffix stripping”. In: *Program*. Publisher: MCB UP Ltd.

Rosenberg, Barr (Mar. 1974). “Extra-Market Components of Covariance in Security Returns”. In: *The Journal of Financial and Quantitative Analysis* 9.2, p. 263.

Schmidhuber, Christof (Mar. 2021). “Trends, reversion, and critical phenomena in financial markets”. In: *Physica A: Statistical Mechanics and its Applications* 566, p. 125642.

Shiller, Robert J (1981). “Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?” In: *American Economic Review* 71.3. Publisher: American Economic Association, pp. 421–436.

Shumway, Tyler (Mar. 1997). “The Delisting Bias in CRSP Data”. In: *The Journal of Finance* 52.1, pp. 327–340.

Shumway, Tyler and Vincent A. Warther (Dec. 1999). “The Delisting Bias in CRSP’s Nasdaq Data and Its Implications for the Size Effect”. In: *The Journal of Finance* 54.6, pp. 2361–2379.

Stock, James H and Mark W Watson (Dec. 2002a). “Forecasting Using Principal Components From a Large Number of Predictors”. In: *Journal of the American Statistical Association* 97.460, pp. 1167–1179.

— (Apr. 2002b). “Macroeconomic Forecasting Using Diffusion Indexes”. In: *Journal of Business & Economic Statistics* 20.2, pp. 147–162.

Van Nieuwerburgh, Stijn and Laura Veldkamp (June 2009). "Information Immobility and the Home Bias Puzzle". In: *The Journal of Finance* 64.3, pp. 1187–1215.

— (Apr. 2010). "Information Acquisition and Under-Diversification". In: *Review of Economic Studies* 77.2, pp. 779–805.

Vuolteenaho, Tuomo (1999). "Understanding the Aggregate Book-to-Market Ratio". In: *SSRN Electronic Journal*.

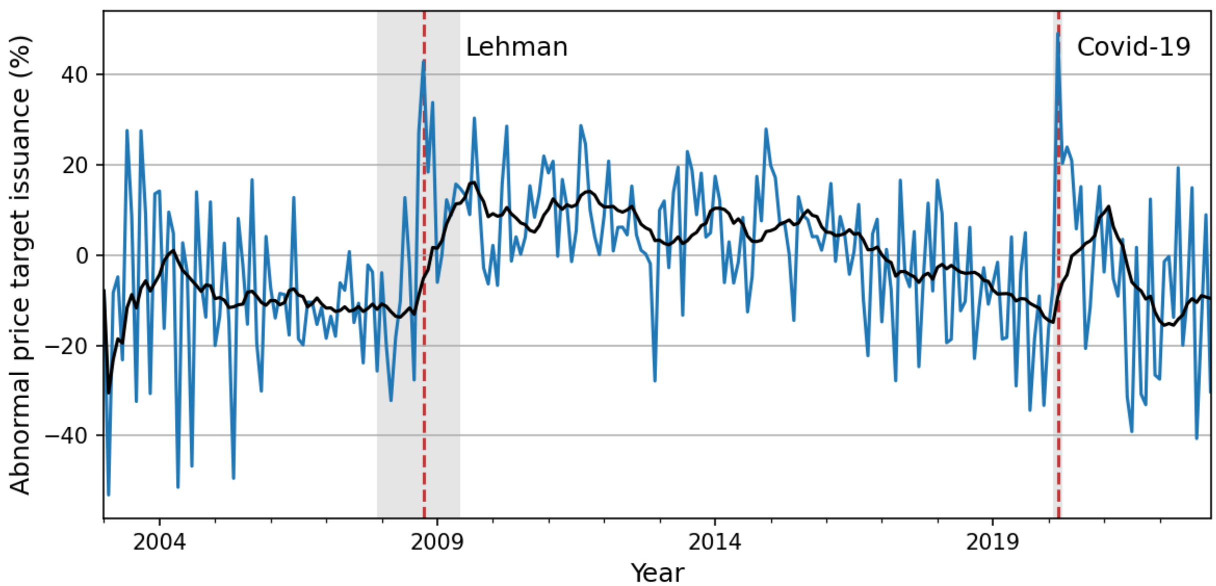
— (2002). "What Drives Firm-Level Stock Returns?" In: *The Journal of Finance* 57.1, pp. 233–264.

Wyart, Matthieu and Jean-Philippe Bouchaud (Aug. 2003). "Statistical models for company growth". In: *Physica A: Statistical Mechanics and its Applications* 326.1, pp. 241–255.

Appendices

Appendix A: Appendix to Chapter 1

A.1 Additional figures



Note: This figure depicts the abnormal issuance of target forecasts, relative to a baseline model with a linear time trend and monthly fixed effects. Units are percent deviations from the baseline model. Vertical lines show the collapse of Lehman Brothers (2008/10) and the reporting of the first confirmed COVID-19 cases in the United States (2020/3).

Figure A.1: Abnormal issuance of target forecasts

A.2 Characteristic definitions

Definitions of WRDS Financial Ratios

Financial Ratio	**Variable**	**Category**	**Formula**
Capitalization Ratio	capital_ratio	Capitalization	Total Long-term Debt as a fraction of the sum of Total Long-term Debt, Common/Ordinary Equity and Preferred Stock
Common Equity/Invested Capital	equity_invcap	Capitalization	Common Equity as a fraction of Invested Capital
Long-term Debt/Invested Capital	debt_invcap	Capitalization	Long-term Debt as a fraction of Invested Capital
Total Debt/Invested Capital	totdebt_invcap	Capitalization	Total Debt (Long-term and Current) as a fraction of Invested Capital
Asset Turnover	at_turn	Efficiency	Sales as a fraction of the average Total Assets based on the most recent two periods
Inventory Turnover	inv_turn	Efficiency	COGS as a fraction of the average Inventories based on the most recent two periods
Payables Turnover	pay_turn	Efficiency	COGS and change in Inventories as a fraction of the average of Accounts Payable based on the most recent two periods

Financial Ratio	Variable	Category	Formula
Receivables Turnover	rect_turn	Efficiency	Sales as a fraction of the average of Accounts Receivables based on the most recent two periods
Sales/Stockholders Equity	sale_equity	Efficiency	Sales per dollar of total Stockholders' Equity
Sales/Invested Capital	sale_invcap	Efficiency	Sales per dollar of Invested Capital
Sales/Working Capital	sale_nwc	Efficiency	Sales per dollar of Working Capital, defined as difference between Current Assets and Current Liabilities
Inventory/Current Assets	invt_act	Financial Soundness	Inventories as a fraction of Current Assets
Receivables/Current Assets	rect_act	Financial Soundness	Accounts Receivables as a fraction of Current Assets
Free Cash Flow/Operating Cash Flow	fcf_ocf	Financial Soundness	Free Cash Flow as a fraction of Operating Cash Flow, where Free Cash Flow is defined as the difference between Operating Cash Flow and Capital Expenditures
Operating CF/Current Liabilities	ocf_lct	Financial Soundness	Operating Cash Flow as a fraction of Current Liabilities
Cash Flow/Total Debt	cash_debt	Financial Soundness	Operating Cash Flow as a fraction of Total Debt

Financial Ratio	Variable	Category	Formula
Cash Balance/Total Liabilities	cash_lt	Financial Soundness	Cash Balance as a fraction of Total Liabilities
Cash Flow Margin	cfm	Financial Soundness	Income before Extraordinary Items and Depreciation as a fraction of Sales
Short-Term Debt/Total Debt	short_debt	Financial Soundness	Short-term Debt as a fraction of Total Debt
Profit Before Depreciation/Current Liabilities	profit_lct	Financial Soundness	Operating Income before D&A as a fraction of Current Liabilities
Current Liabilities/Total Liabilities	curr_debt	Financial Soundness	Current Liabilities as a fraction of Total Liabilities
Total Debt/EBITDA	debt_ebitda	Financial Soundness	Gross Debt as a fraction of EBITDA
Long-term Debt/Book Equity	dltt_be	Financial Soundness	Long-term Debt to Book Equity
Interest/Average Long-term Debt	int_debt	Financial Soundness	Interest as a fraction of average Long-term debt based on most recent two periods
Interest/Average Total Debt	int_totdebt	Financial Soundness	Interest as a fraction of average Total Debt based on most recent two periods

Financial Ratio	Variable	Category	Formula
Long-term Debt/Total Liabilities	lt_debt	Financial Soundness	Long-term Debt as a fraction of Total Liabilities
Total Liabilities/Total Tangible Assets	lt_ppent	Financial Soundness	Total Liabilities to Total Tangible Assets
Cash Conversion Cycle (Days)	cash_conversion	Liquidity	Inventories per daily COGS plus Account Receivables per daily Sales minus Account Payables per daily COGS
Cash Ratio	cash_ratio	Liquidity	Cash and Short-term Investments as a fraction of Current Liabilities
Current Ratio	curr_ratio	Liquidity	Current Assets as a fraction of Current Liabilities
Quick Ratio (Acid Test)	quick_ratio	Liquidity	Quick Ratio: Current Assets net of Inventories as a fraction of Current Liabilities
Accruals/Average Assets	Accrual	Other	Accruals as a fraction of average Total Assets based on most recent two periods
Research and Development/Sales	RD_SALE	Other	R&D expenses as a fraction of Sales
Avertising Expenses/Sales	adv_sale	Other	Advertising Expenses as a fraction of Sales

Financial Ratio	Variable	Category	Formula
Labor Expenses/Sales	staff_sale	Other	Labor Expenses as a fraction of Sales
Effective Tax Rate	efftax	Profitability	Income Tax as a fraction of Pretax Income
Gross Profit/Total Assets	GProf	Profitability	Gross Profitability as a fraction of Total Assets
After-tax Return on Average Common Equity	aftret_eq	Profitability	Net Income as a fraction of average of Common Equity based on most recent two periods
After-tax Return on Total Stockholders' Equity	aftret_equity	Profitability	Net Income as a fraction of average of Total Shareholders' Equity based on most recent two periods
After-tax Return on Invested Capital	aftret_invcapx	Profitability	Net Income plus Interest Expenses as a fraction of Invested Capital
Gross Profit Margin	gpm	Profitability	Gross Profit as a fraction of Sales
Net Profit Margin	npm	Profitability	Net Income as a fraction of Sales
Operating Profit Margin After Depreciation	opmad	Profitability	Operating Income After Depreciation as a fraction of Sales
Operating Profit Margin Before Depreciation	opmbd	Profitability	Operating Income Before Depreciation as a fraction of Sales

Financial Ratio	**Variable**	**Category**	**Formula**
Pre-tax Return on Total Earning Assets	pretret_earnat	Profitability	Operating Income After Depreciation as a fraction of average Total Earnings Assets (TEA) based on most recent two periods, where TEA is defined as the sum of Property Plant and Equipment and Current Assets
Pre-tax return on Net Operating Assets	pretret_noa	Profitability	Operating Income After Depreciation as a fraction of average Net Operating Assets (NOA) based on most recent two periods, where NOA is defined as the sum of Property Plant and Equipment and Current Assets minus Current Liabilities
Pre-tax Profit Margin	ptpm	Profitability	Pretax Income as a fraction of Sales
Return on Assets	roa	Profitability	Operating Income Before Depreciation as a fraction of average Total Assets based on most recent two periods

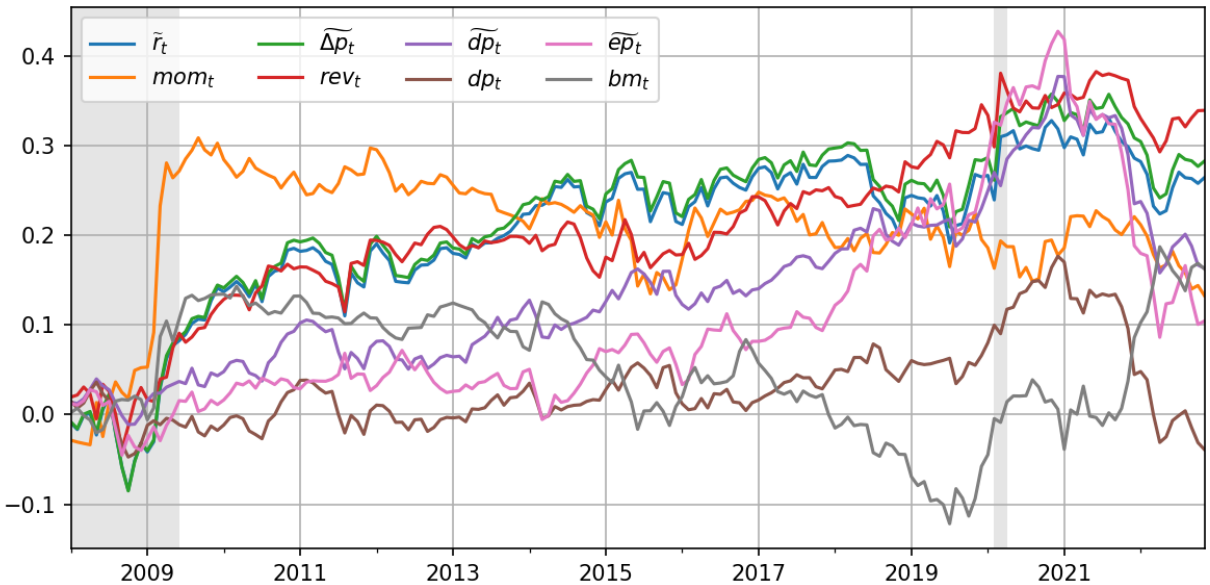
Financial Ratio	**Variable**	**Category**	**Formula**
Return on Capital Employed	roce	Profitability	Earnings Before Interest and Taxes as a fraction of average Capital Employed based on most recent two periods, where Capital Employed is the sum of Debt in Long-term and Current Liabilities and Common/Ordinary Equity
Return on Equity	roe	Profitability	Net Income as a fraction of average Book Equity based on most recent two periods, where Book Equity is defined as the sum of Total Parent Stockholders' Equity and Deferred Taxes and Investment Tax Credit
Total Debt/Equity	de_ratio	Solvency	Total Liabilities to Shareholders' Equity (common and preferred)
Total Debt/Total Assets	debt_assets	Solvency	Total Debt as a fraction of Total Assets
Total Debt/Total Assets	debt_at	Solvency	Total Liabilities as a fraction of Total Assets

Financial Ratio	**Variable**	**Category**	**Formula**
Total Debt/Capital	debt_capital	Solvency	Total Debt as a fraction of Total Capital, where Total Debt is defined as the sum of Accounts Payable and Total Debt in Current and Long-term Liabilities, and Total Capital is defined as the sum of Total Debt and Total Equity (common and preferred)
After-tax Interest Coverage	intcov	Solvency	Multiple of After-tax Income to Interest and Related Expenses
Interest Coverage Ratio	intcov_ratio	Solvency	Multiple of Earnings Before Interest and Taxes to Interest and Related Expenses
Dividend Payout Ratio	dpr	Valuation	Dividends as a fraction of Income Before Extra. Items
Forward P/E to 1-year Growth (PEG) ratio	PEG_1yrforward	Valuation	Price-to-Earnings, excl. Extraordinary Items (diluted) to 1-Year EPS Growth rate
Forward P/E to Long-term Growth (PEG) ratio	PEG_ltgforward	Valuation	Price-to-Earnings, excl. Extraordinary Items (diluted) to Long-term EPS Growth rate

Financial Ratio	Variable	Category	Formula
Trailing P/E to Growth (PEG) ratio	PEG_trailing	Valuation	Price-to-Earnings, excl. Extraordinary Items (diluted) to 3-Year past EPS Growth
Book/Market	bm	Valuation	Book Value of Equity as a fraction of Market Value of Equity
Shillers Cyclically Adjusted P/E Ratio	capei	Valuation	Multiple of Market Value of Equity to 5-year moving average of Net Income
Dividend Yield	divyield	Valuation	Indicated Dividend Rate as a fraction of Price
Enterprise Value Multiple	evm	Valuation	Multiple of Enterprise Value to EBITDA
Price/Cash flow	pcf	Valuation	Multiple of Market Value of Equity to Net Cash Flow from Operating Activities
P/E (Diluted, Excl. EI)	pe_exi	Valuation	Price-to-Earnings, excl. Extraordinary Items (diluted)
P/E (Diluted, Incl. EI)	pe_inc	Valuation	Price-to-Earnings, incl. Extraordinary Items (diluted)
Price/Operating Earnings (Basic, Excl. EI)	pe_op_basic	Valuation	Price to Operating EPS, excl. Extraordinary Items (Basic)

Financial Ratio	**Variable**	**Category**	**Formula**
Price/Operating Earnings (Diluted, Excl. EI)	pe_op_dil	Valuation	Price to Operating EPS, excl. Extraordinary Items (Diluted)
Price/Sales	ps	Valuation	Multiple of Market Value of Equity to Sales
Price/Book	ptb	Valuation	Multiple of Market Value of Equity to Book Value of Equity

(a) Individual equities



(b) Characteristic portfolios

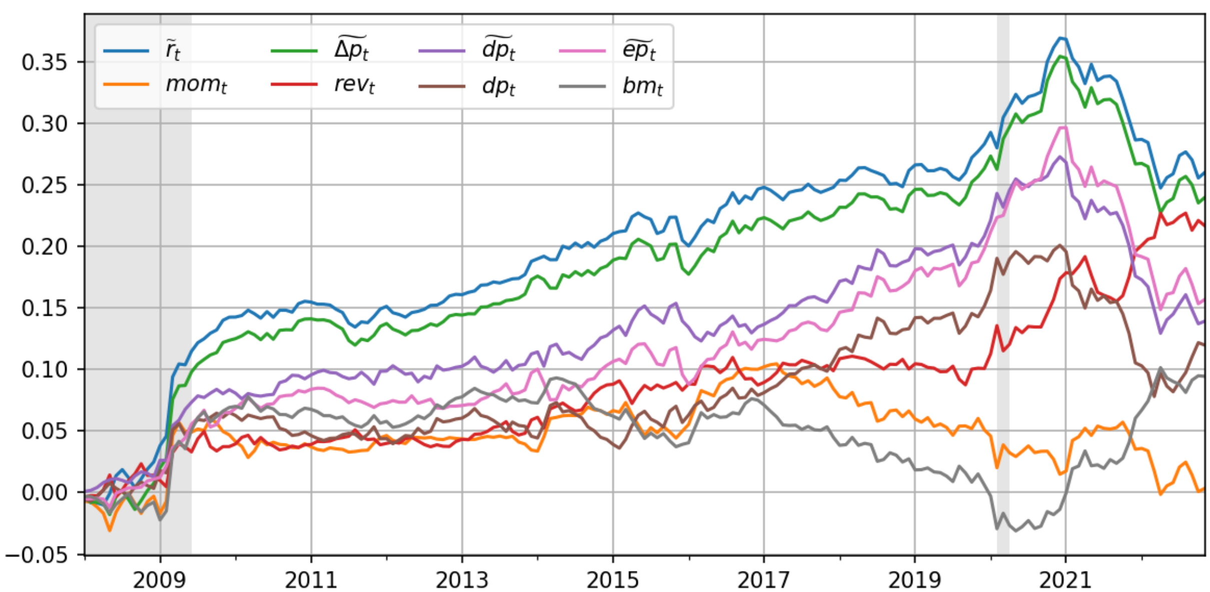


Figure A.2: Cumulative log returns, rank-weighted strategy

Note: These figures show the log cumulative returns to strategies formed by rank-weighting individual equities or characteristic portfolios according to the value of the predictor indicated above each column. These are the same strategies discussed in table 1.3. In forming these strategies, the direction of each strategy is chosen month-by-month such that the average returns to the strategy using the chosen direction would have been positive in the period up that month.

Appendix B: Appendix to Chapter 3

B.1 Proofs

Corollary 3. *Attention to macroeconomic conditions is higher for firms with greater diversified cash flows, D_i.*

Proof. Substitute the definition of the diversification factor D_i from (3.7) into (3.6), we have

$$MacroAttn_i = \frac{1}{2} + \frac{1}{4\kappa}\log_2\left(\frac{D_i}{1-D_i}\right)$$

Differentiating with respect to D_i and given that $D_i \in [0,1]$, we see that macroeconomic attention is increasing in D_i.

□

Corollary 4. *The risk premium on asset i's returns is decreasing in the share of diversified cash flow risk.*

Proof. Rewrite the unexpected log return of asset i as $r_{it+1} - E_t r_{i,t+1} = N^{CF}_{i,t+1} - N^{DR}_{i,t+1}$. We then can write the first term in (3.8) as

$$\begin{aligned}
Cov(r_{i,t+1} - E_t r_{i,t+1}, N^{CF}_{m,t+1}) &= Cov(N^{CF}_{i,t+1}, N^{CF}_{m,t+1}) - Cov(N^{DR}_{i,t+1}, N^{CF}_{m,t+1}) \\
&= Cov(\eta_{t+1} + \nu_{it+1}, \eta_{t+1} + \frac{1}{M}\sum_i \nu_{it+1}) - Cov(N^{DR}_{i,t+1}, \frac{1}{M}\sum_i N^{CF}_{i,t+1}) \\
&= \sigma^2_\eta + \frac{1}{M}\varphi_i \sigma^2_\nu
\end{aligned}$$

where the second equality follows from definitions of the news terms $N^{CF}_{i,t+1}$ and $N^{CF}_{m,t+1}$ and the third equality follows from our assumption of independence of cash flow and discount rate news.

The second term in (3.8) can likewise be written as

$$Cov(r_{i,t+1} - E_t r_{i,t+1}, -N^{DR}_{m,t+1}) = Cov(-N^{DR}_{i,t+1}, -\frac{1}{M}\sum_i N^{DR}_{i,t+1}) = \frac{1}{M}\sigma^2_\omega$$

Combining the terms we have

$$rp_{i,t} = \gamma\left(\sigma^2_\eta + \frac{1}{M}\varphi_i\sigma^2_\nu\right) + \frac{1}{M}\sigma^2_\omega$$

By chain rule, the effect on risk premium is given by

$$\frac{\partial rp_{i,t}}{\partial D_i} = \frac{\partial rp_{it}}{\partial \varphi_i}\frac{\partial \varphi_i}{\partial D_i} = -\frac{\gamma}{M}\frac{(\sigma^2_\eta + \varphi_i\sigma^2_\nu)^2}{\sigma^2_\eta} < 0$$

□

Prediction 5. *Stocks with higher diversification factor have lower cash flow betas, and no systematic difference in discount rate betas. The risk premium unaccounted for the stock's market beta (the CAPM alpha) is decreasing in the stock's cash flow beta.*

Proof. By chain rule and the definition of the betas, we have

$$\frac{\partial \beta^{CF}_{i,m}}{\partial D_i} = \frac{1}{Var(r_{M,t+1})}\frac{\partial Cov(r_{i,t+1} - E_t r_{i,t+1}, N^{CF}_{M,t+1})}{\partial \varphi_i}\frac{\partial \varphi_i}{\partial D_i} = -\frac{1}{\sigma^2_M}\frac{\gamma}{M}\frac{(\sigma^2_\eta + \varphi_i\sigma^2_\nu)^2}{\sigma^2_\eta} < 0$$

$$\frac{\partial \beta^{DR}_{i,m}}{\partial D_i} = \frac{1}{Var(r_{M,t+1})}\frac{\partial Cov(r_{i,t+1} - E_t r_{i,t+1}, -N^{DR}_{M,t+1})}{\partial \varphi_i}\frac{\partial \varphi_i}{\partial D_i} = 0$$

Substitute the identity $\beta_{i,m} = \beta^{CF}_{i,m} + \beta^{DR}_{i,m}$ into (3.10) yields

$$rp_i = \sigma^2_m(\gamma - 1)\beta^{CF}_{i,m} + \beta\sigma^2_m$$

It follows that the component of risk premium unexplained by the market beta (CAPM alpha) is given by

$$\alpha^{CAPM} \equiv rp_i - \beta_{i,m}\sigma_m^2 = \sigma_m^2(\gamma - 1)\beta_{i,m}^{CF}$$

Given that $\beta_{i,m}^{CF}$ is decreasing in a stock's diversification factor D_i, it follows that α^{CAPM} is decreasing in D_i as well.

□

B.2 Additional tables and figures

	Mean	Std dev	AR(1)	P10	P90	Sharpe ratio
Macro Attention	-2.80	0.52	0.46	-3.49	-2.18	1.34
$\beta(MKT)$	1.01	0.85	0.12	0.04	2.03	0.23
$\beta(SMB)$	0.72	1.28	0.09	-0.72	2.35	0.05
$\beta(HML)$	0.08	1.55	0.11	-1.69	1.86	0.51
$\beta(VIX)$	-0.00	0.01	-0.01	-0.01	0.01	0.14
Size	14.08	1.55	0.88	12.16	16.23	1.62
Book-to-market	-0.90	0.74	0.86	-1.90	-0.02	0.45
Lagged returns (12 mths)	0.14	0.35	0.72	-0.29	0.57	0.17
Idio vol	0.02	0.01	0.43	0.01	0.03	1.27
Issuances (36 mths)	0.06	0.20	0.89	-0.11	0.28	0.65
Accruals	0.06	0.07	0.83	-0.01	0.14	1.29
Return on asset	0.11	0.14	0.91	0.00	0.24	1.30
Dividend yield	0.02	0.01	0.86	0.01	0.04	0.38
Asset growth	-2.20	1.30	0.80	-3.87	-0.54	0.13
Lagged returns (36 mths)	0.35	0.49	0.81	-0.25	0.99	0.23
Issuances (12 mths)	0.02	0.09	0.77	-0.04	0.09	0.59
Turnover	2.07	1.42	0.92	0.67	3.96	0.25
Net debt-to-Price	0.16	0.48	0.84	-0.23	0.64	0.08
Sale-to-Price	1.12	1.29	0.84	0.18	2.55	0.77

Table B.1: Summary statistics of return predictors from 2005Q1 to 2019Q4.

Note: The table reports the average (Mean), standard deviation (Std dev), AR(1) coefficient of firm-level observations (AR(1)), 10th percentile (P10), 90th percentile (P90), and Sharpe ratio of long-short portfolio constructed from decile sorts on characteristics using monthly observations for each variable. See Table B.3 in the Appendix for variable definitions.

Macro terms	Non-macro terms
economi	compani
inflat	acquisit
budget	ebitda
hous	technolog
central	brand
recess	store
slow	patient
pace	platform
foreign exchang	launch
euro	execut
gdp	facil
export	network
moder	sharehold
uncertainti	client
exchang rate	strateg
germani	excit
read	dividend
labor	digit
steadi	capabl
crisi	equiti

Table B.2: Macroeconomic and firm-specific terms occurring in earnings call transcripts.

Note: The table reports the top 20 vocabulary terms occurring in earnings call transcripts labeled as "Macro" (Left) and "Non-Macro" (Right). Terms are first stemmed to root (Porter, 1980) and common stopwords are removed, and then ranked by the average TF-IDF score across all earnings calls transcripts. Terms that occur in both sets of labeled call transcripts are excluded.

Name	Description	Year	Authors
beta(MKT)	Market factor beta	1973	Fama and MacBeth
beta(SMB)	Small-minus-big factor beta	1993	Fama and French
beta(HML)	High-minus-low factor beta	1993	Fama and French
beta(VIX)	Aggregate volatility factor beta	2006	Ang and Hodrick and Xing and Zhang
Size	Log market capitalization	1981	Banz
Book-to-market	Log book-to-market ratio	1980	Stattman
Lagged returns (12 mths)	Stock returns from mo -12 to mo -1	1985	De Bondt and Thaler
Idio vol	Volatility of residuals from regressing daily returns on market factor in past mo	2006	Ang and Hodrick and Xing and Zhang
Issuances (36 mths)	Log growth in split-adjusted shares outstanding from mo -36 to -1	2006	Daniel and Titman
Accruals	Change in non-cash net working capital in prior fiscal year	1996	Sloan
Return on asset	Income before extraordinary items divided by average total assets in prior fiscal year	2008	Soliman
Asset growth	Log growth in total assets in the prior fiscal year	2004	Titman and Wei and Xie
Lagged returns (36 mths)	Log stock returns from mo -36 to mo -13	1985	De Bondt and Thaler
Issuances (12 mths)	Log growth in split-adjusted shares outstanding from mo -12 to -1	2006	Daniel and Titman
Turnover	Average monthly turnover (shares traded div shares outstanding) from mo -12 to mo -1	2000	Lee and Swaminathan
Sale-to-price	Sales in the prior fiscal year divided by market value at the end of the prior mo	1994	Lakonishok and Shleifer and Vishny
Net debt-to-price	Short-term plus long-term debt net of cash divided by market value at the end of the prior mo	2007	Penman and Richardson and Tuna
Dividend yield	Dividends per share over the prior 12 mo divided by price at end of the prior mo	1982	Litzenberger and Ramaswamy

Table B.3: Definition of predictors of stock returns.

	(1) CFi-CFm		(2) CFi-DRm		(3) DRi-CFm		(4) DRi-DRm		(5) Ri-CFm		(6) Ri-DRm	
	β	t	β	t	β	t	β	t	β	t	β	t
1	1.84	(4.2)	-0.52	(-4.1)	-1.12	(-1.9)	0.72	(141.7)	0.72	(1.1)	0.20	(1.6)
2	1.61	(3.9)	-0.36	(-3.2)	-1.15	(-1.9)	0.74	(175.1)	0.46	(0.6)	0.38	(3.4)
3	1.63	(5.0)	-0.43	(-4.4)	-1.14	(-2.0)	0.73	(117.0)	0.49	(0.8)	0.30	(3.1)
4	1.47	(4.6)	-0.39	(-4.4)	-1.15	(-2.0)	0.73	(107.6)	0.32	(0.5)	0.34	(4.0)
5	1.57	(5.5)	-0.39	(-4.4)	-1.15	(-2.0)	0.74	(161.0)	0.42	(0.7)	0.34	(3.9)
6	1.44	(4.3)	-0.42	(-3.4)	-1.14	(-1.9)	0.73	(161.7)	0.30	(0.5)	0.31	(2.5)
7	1.42	(5.5)	-0.39	(-4.7)	-1.16	(-2.0)	0.74	(115.8)	0.26	(0.5)	0.34	(4.3)
8	1.40	(5.4)	-0.37	(-4.7)	-1.18	(-2.0)	0.74	(116.7)	0.22	(0.4)	0.36	(4.7)
9	1.26	(4.5)	-0.40	(-4.4)	-1.19	(-2.1)	0.74	(98.1)	0.06	(0.1)	0.34	(3.7)
10	1.43	(4.0)	-0.40	(-3.9)	-1.17	(-2.0)	0.73	(96.1)	0.26	(0.5)	0.33	(3.1)
10-1	-0.41	(-0.7)	0.12	(0.7)	-0.05	(-0.1)	0.01	(1.2)	-0.46	(-0.5)	0.13	(0.8)

Table B.4: Decomposing portfolio returns into cash flow and discount rate news.

Note: The table reports the estimates of regressing portfolio cash flow (CFi) and discount rate (DRi) news on the market portfolio's cash flow (CFm) and discount rate (DRm) news. Ri refers to the portfolio unexpected return, computed as the difference in cash flow and discount rate news. Rows refer to portfolios sorted on the basis of predicted macroeconomic attention, with 1 as the lowest attention portfolio and 10 the highest. "10-1" refers to the difference of the beta of Portfolio 10 relative to Portfolio 1. Within each column, refers to the estimated betas, and refers to the corresponding t-statistic computed from bootstrap standard errors. The sample period is from 2005 to 2019.

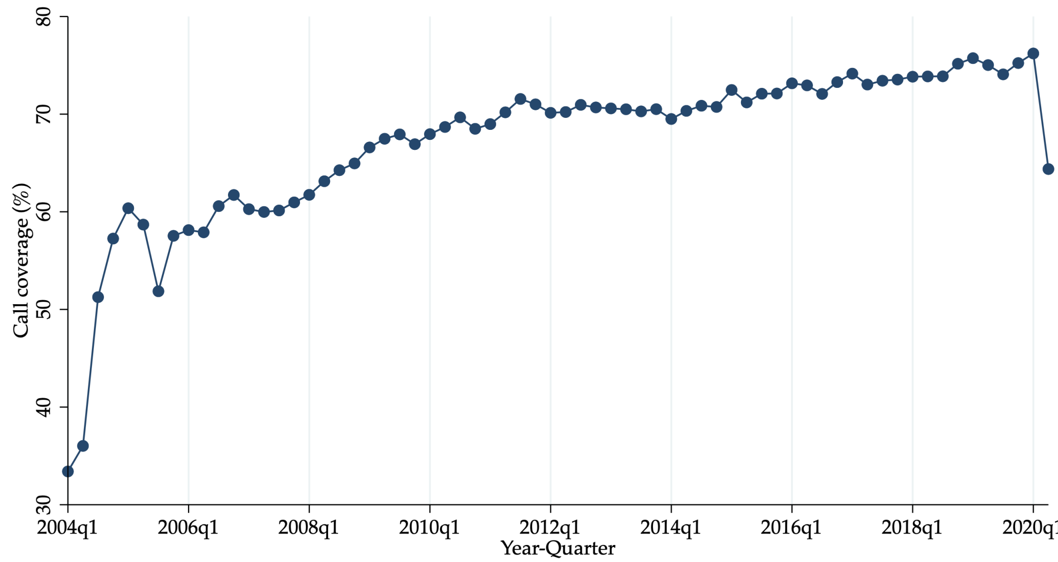


Figure B.1: Call coverage from 2004Q1 to 2020Q1.

Note: Call coverage is the percentage of unique CUSIPs with monthly stock returns data from CRSP and matched earnings call transcripts from Factset.

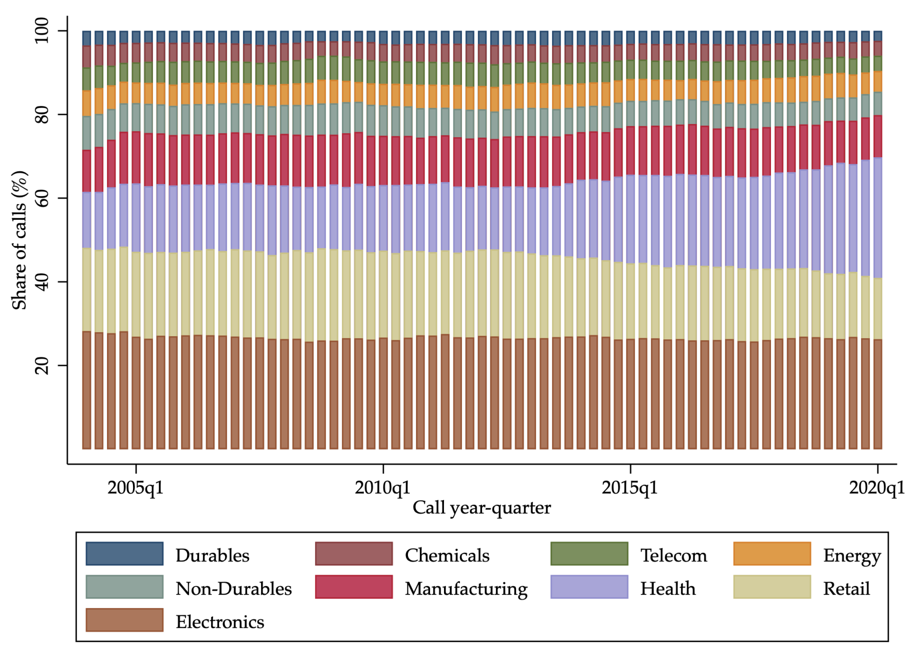


Figure B.2: Composition of calls by industry from 2004Q1 to 2020Q1.

Note: Industry classification based on Fama and French 12 industry portfolios, excluding firms in the utilities (SIC 4900-4999) and financial (SIC 6000-6999) sectors.

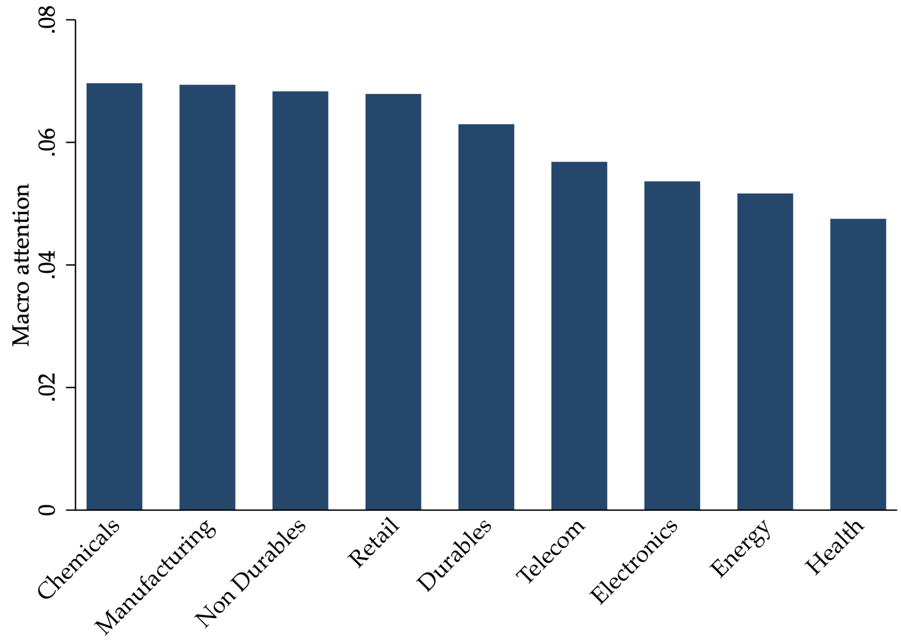


Figure B.3: Average macroeconomic attention by industry.

Note: Industry classification based on Fama and French 12 industry portfolios, excluding firms in the utilities (SIC 4900-4999) and financial (SIC 6000-6999) sectors.

B.3 VAR estimation

The estimation strategy closely follows Vuolteenaho (2002), Campbell and Vuolteenaho (2004), and Campbell, Polk, and Vuolteenaho (2009). The data generating process follows a first-order VAR model

$$z_{t+1} = a + \Gamma z_t + u_{t+1}$$

where z_{t+1} is an $m \times 1$ state vector with return r_{t+1} as the first element, a and Γ are respectively a $m \times 1$ vector and $m \times m$ matrix of VAR parameters to be estimated, and u_{t+1} is an $m \times 1$ vector of shocks with variance-covariance Σ. We specify the remaining state variables in the VAR specification shortly below. Following Campbell, Polk, and Vuolteenaho (2009), we estimate separate VARs for aggregate news and firm-specific news and set $\rho = 0.95$. This approach is consistent with the empirical literature documenting different sources of risk driving returns in the aggregate and cross-section. We estimate both VARs using annual data over the sample period from 1928 to 2019, and treat the parameters (a, Γ, Σ) as constant across the sample period. As such, the parameters are estimated by separate equation-by-equation pooled regression.

The aggregate VAR consists of four state variables $(m = 4)$. The first is the log market return r_M defined as the value-weighted average log return of all common equity from the end of June in year $t-1$ to end of June in year t. The second state variable is the term yield spread (TY), which is the difference between the ten-year Treasury bond log yield in June of year t and the 90-day secondary market Treasury bill log yield in June of year t, using data from the Global Financial database. The third variable is the log smoothed price-earnings ratio (PE) in June of year t taken from Professor Shiller's website, and the last variable is the small-stock value spread (VS) defined as the difference between the log book-to-market (BM) ratio of the small high BM portfolio and the log BM ratio of the small low BM portfolio in June of year t, which comes from Professor French's website.

The firm-level VAR consists of three state variables $(m = 3)$. The first state variable is the log return (r_i) on a firm's common stock equity from the end of June in year $t-1$ to the end of June in year t. Following Vuolteenaho (2002), we use un-levered returns by defining a stock as consisting of 90% firm's common stock and 10% Treasury bill. The second variable is the log book-to-market ratio of unlevered equity, BM, which is defined as $BM = \log(0.9BE+0.1ME)-\log ME$, where BE is the book equity at the end of calendar year $t-1$ and ME is the market equity at the end of May of year t. The third variable is the long term profitability ROE, defied as the training average of earnings divided by the trailing five-year average of $(0.9BE + 0.1ME)$. Following Campbell et al. (2010), earnings X_t is defined from the identity $X_t = [(1+r_i)ME_{t-1} - D_t]/ME_t \times BE_t - BE_{t-1} + D_t$. We use this rather than recorded earnings given that the quantity is measured with less precision in the early sample period. All variables in the firm-level VAR are cross-sectionally demeaned by subtracting the value-weighted market return in the case of firm-returns and average value of each variable each year for the other state variables.