The Optimal Stock Valuation Ratio*

Sebastian Hillenbrand†1 and Odhrain McCarthy‡2

[1]Harvard Business School
[2]New York University

This Version: November 2022

Draft updated regularly, please click here for latest version

Abstract

Stock valuation ratios contain expectations of returns, yet, their performance in predicting returns has been rather dismal. This is because of an omitted variable problem: valuation ratios also contain expectations of cash flow growth. Time-variation in cash flow volatility and a structural shift towards repurchases have magnified this omitted variable problem. We show theoretically and empirically that scaling prices by forward measures of cash flows can overcome this problem yielding optimal return predictors. We construct a new measure of the forward price-to-earnings ratio for the S&P index based on earnings forecasts using machine learning techniques. The out-of-sample explanatory power for predicting one-year aggregate returns with our forward price-to-earnings ratio ranges from 7% to 11%, thereby beating all other predictors and helping to resolve the out-of-sample predictability debate (Goyal and Welch, 2008).

Keywords: Stock market valuation, return prediction, out-of-sample prediction, machine learning

*We thank Lukas Kremens for helpful comments.

†shillenbrand@hbs.edu
‡otm210@nyu.edu

1 Introduction

Is it possible to predict stock market returns? This is one most of the important questions for finance applications. Unfortunately, researchers have not come to a conclusive answer on this fundamental question. For out-of-sample prediction – which is arguably the only relevant test for investors' asset-allocation decisions – the outlook is rather bleak. Goyal and Welch (2008) show that most predictors proposed by prior studies fail to predict returns out-of-sample (for an update see Goyal, Welch, and Zafirov (2021)). A number of studies have since developed tools to "resurrect" out-of-sample return predictability, but still find rather small explanatory power (e.g. Campbell and Thompson, 2008, etc.).[1] Yet, the fact that we struggle to predict returns out-of-sample is unsatisfactory given the widely-held view that discount rates move around over time (Cochrane, 2011). And this fact is even more puzzling considering that commonly-used price-ratios, such as the price-dividend (PD) or price-earnings (PE) ratio, contain expectations of returns (Campbell and Shiller, 1988). Unless price-ratios perfectly forecast cash flow growth, they must also forecast returns (Cochrane, 2008).

Our paper aims to make progress on this issue. To do so, we first theoretically outline the issues that have resulted in the poor return-prediction performance of previously-used price ratios in the literature and the optimal stock market price ratio that circumvents these issues and yields the highest possible R-squared performance in predicting future long-run returns. Second, we then empirically test our theoretical predictions and show they are borne out in data. Most saliently, when we construct a proxy for the optimal stock market price ratio we find we can obtain an in-sample and out-of-sample R-squared in predicting annual stock market returns of 12.7% and 10.8%, respectively, thereby beating – by a wide margin – all other forecasting variables that have been put forth in the literature to date (Goyal and Welch, 2008; Goyal, Welch, and Zafirov, 2021). These results imply that our method extracts a clean measure of future expected returns from stock prices, similar to the measure of expected corporate bond returns constructed by Gilchrist and Zakrajšek (2012). Our results have important relevance for both academics and practitioners alike, for example for investors' asset-allocation and firms' investment decisions.

We start by drawing on insights put forth in prior work (e.g., Campbell and Shiller, 1988; Lettau and Ludvigson, 2005, etc.) that trailing price-ratios, i.e. ratios of prices to a trailing measure of cash flows such as dividends or earnings, contain both expected returns and expected cash flow growth, and

[1]Some of these studies have focused on imposing model restrictions based on economic priors (Campbell and Thompson, 2008; Ferreira and Santa-Clara, 2011). Other studies have allowed for time-varying model parameters to take into account regime shifts (Dangl and Halling, 2012). Another approach is combine individual forecast (Rapach, Strauss, and Zhou, 2010). In contrast to these paper, we construct a price-ratio whose ability to predict return in- and out-of-sample is stable over time, i.e., we completely eliminate concerns about model uncertainty and parameter instability.

that not accounting for expected cash flow growth leads to an omitted variable problem. Building on this work, we show that the in-sample explanatory power of trailing price-ratios for predicting future returns is, in almost all cases, negatively related to two components of time-varying expected cash flow growth. First, it is negatively related to the volatility of expected cash flow growth. Intuitively, if expectations of cash flow growth are very volatile, then these expectations drive most of the variation in trailing price-ratios undermining their ability to predict returns. Second, it is negatively related to structural changes in mean expected cash flow growth. Intuitively, these changes will result in shifts in the mean of trailing price-ratios without necessarily any change in expected returns undermining the in-sample regression's ability to predict returns. In addition to these two considerations, a final related issue is the extent to which the trailing cash flow measure used to scale prices captures all payout methods (Boudoukh, Michaely, Richardson, and Roberts, 2007). We show if there is a regime change in the payout method (e.g., from dividends to repurchases) and the trailing cash flow measure (e.g., dividends) does not capture the new payout method it will again lead to a structural shift in the price-ratio and undermine the in-sample regressions ability to predict returns.

While this omitted variable problem impacts in-sample return predictions, it is even more important for out-of-sample predictions. More formally, we show that the delta between the in-sample and out-of-sample R-squared from using a price-ratio to forecast returns over a period that spans two distinct regimes which have different population regression coefficients (i.e., different population intercepts and slopes) increases in the wedge between the two intercept coefficients and the wedge between the two slope coefficients. Accordingly, regime changes in cash-flow return volatility, expected cash flow growth or payout methods (to the extent the trailing cash flow measure used does not include them) will cause structural shifts in the intercept and slope coefficients which will drive a wedge between the (potentially already) low in-sample R-squared and the even lower out-of-sample R-squared.

Armed with these theoretical insights, we proceed to delineate empirically why trailing price-ratios struggle to predict returns. For the PE ratio, we show that its poor predictive performance is related to the high and time-varying volatility of earnings. Especially over the last three decades, annual earnings volatility was extremely high at 44% (Hillenbrand and McCarthy, 2022). We split the data into three periods – namely, 1871-1944, 1945-1989 and 1990-2022 – where the first and last periods were high earnings volatility regimes and where the intermittent period was a low earnings volatility regime. In line with the theory, we show the in-sample R-squared of the PE ratio in predicting returns for the two high-volatility periods is dismal, ranging between 1% and 4%, while we find the R-squared was much higher (14%) for the low-volatility period. In addition, the out-of-sample R-squared is even lower and negative over the entire sample period. For the PD ratio, we show that its poor predictive performance is related to a shift in firms' payout policy (Boudoukh, Michaely, Richardson, and Roberts, 2007) undermining the use of dividends as an accurate measure of total cash flows. In particular, firms started distributing cash flows to shareholders via repurchases (instead of dividends) in the 1980s and markedly increase their repurchases in the 1990s. This shift towards repurchases meant prices (which are based off total cash flows) are being scaled by a comparatively smaller amount of dividends overtime causing an upward shift in the PD ratio. It furthermore implies that, when we regress future returns on the PD ratio, the regression intercept will invariably change over time. Consistent with this, we find that the out-of-sample R-squared using the PD ratio begins to deteriorate around 1990 coinciding with structural shift in firms' payout-method. By contrast, the predictive power of the PD ratio is much higher and stable when we split the sample based on the different payout policy regimes.

Ultimately, trailing PD and the PE ratios fail to predict returns, because the former does not accurately capture total cash flows over the forecasting sample while both do not adequately account for expected future cash flow growth. Accordingly, we show that the optimal valuation ratio for predicting expected returns is one which uses a measure of total cash flows (e.g., earnings) but then removes expected cash flow growth and we call such a price-ratio the "Optimal price-ratio". We show the Optimal price-ratio yields the highest R-squared in predicting long-run returns and is equal to the ratio of the variance of expected returns to the variance of realized returns. We then delineate that, in an economy in which cash flow growth shocks last for Tyrs (before reverting to a constant expected growth process), removing expected cash flow growth over the next Tyrs will result in the Optimal price-ratio. By corollary, in such an economy, removing T-1yrs, T-2yrs, ..., 1yrs of expected cash flow growth will be progressively sub-optimal at accounting for expected cash flow growth, and therefore, forecasting long-run returns. We then derive three additional theoretical results that guide our empirical analyses. First, we show that the Optimal price-ratio is also optimal for one-period returns if expected returns follow an autoregressive structure (e.g. Van Binsbergen and Koijen, 2010). Second, we show that it is optimal to use rational, full-information expectations of cash flows, instead of (potentially biased) market expectations. Intuitively, the Optimal price-ratio will pick up on both market expected returns and any cash flow news that arises from biased market expectations. Finally, we show that we can empirically proxy for the Optimal price-ratio by scaling prices by optimal forecast of cash flows ("forward price-ratios"). Intuitively, this is because forward price-ratios circumvent the aforementioned issues that not accounting for expected cash flow growth give rise to.

To construct optimal forward price-ratios we need to construct optimal forecasts of aggregate future cash flows. To do so, we build on a recent burgeoning literature using machine learning algorithms to construct optimal forecasts of macroeconomic quantities (e.g., Van Binsbergen, Han, and Lopez-Lira, 2020; De Silva and Thesmar, 2021; Bianchi, Ludvigson, and Ma, 2022, etc.). In particular, we use machine learning algorithms to predict firm-level earnings growth and then aggregate these forecasts up to the market-level. To construct optimal, out-of-sample firm-level forecasts two ingredients are key. First, we need a large set of firm-level accounting information. To achieve this, we use firm-level annual report data collected from the Morningstar Industrial Manuals (Graham, Leary, and Roberts, 2015; Graham and Leary, 2018; Graham, Kim, and Leary, 2020), and thereafter use data from Compustat. This process yields a large set of annual firm-level variables that spans the period from 1927 to 2020. Second, we need to ensure that this information would be available to forecasters at the time of forecast. To achieve this, we assume a six-month lag between the annual report date and its release date.

We then train our machine learning algorithm to predict firm-level earnings growth over 1-year, 2-year, 3-year and 5-year horizons, crucially allowing for the necessary gaps between training and forecasting to ensure there is no look-ahead bias. We aggregate these firm-level growth forecasts to the S&P 500 level and compound trailing S&P 500 earnings with the aggregate earnings growth forecasts. Finally, we scale the S&P 500 index by these earnings forecasts, thereby yielding 1-year, 2-year, 3-year and 5-year forward PE ratios. These forward PE ratios are constructed only using information that would be available to forecasters in real-time, so that investors could feasibly use them for real-time capital-allocation decisions.

With these forward PE ratios in hand, we test their ability to predict returns out-of-sample. We find that the out-of-sample ("OOS") R-squared using the forward PE ratios ranges between 7.4% and 10.8%. This is substantially higher than price-ratios previously used in the literature, such as the PD ratio, the PE ratio, and the CAPE ratio, for which the OOS R-squared ranges between -6.7% and -0.9%. The predictive power is far better than any single predictor variable tested in the framework of Goyal and Welch (2008) (see also Goyal, Welch, and Zafirov (2021) for an update of the initial analysis).

Lastly, we test whether our empirical results confirm additional theoretical predictions. First, we find the OOS R-squared of the 5-year forward PE ratio is 10.8%, which is higher than the OOS R-squared of the 3-year ratio (8.8%), which is higher than the OOS R-squared of the 2-year ratio (8.4%) and so forth. This is exactly in line with theoretical predictions that price-ratios using longer-term cash flow forecasts should be better predictors of returns as they are better in isolating expected returns. Second, we compare the deltas between in-sample and OOS R-squared for different return predictors. Consistent with our theory, we show that the deltas for the forward 1-year, 3-year and 5-year PE ratios are fairly close to 0% over the sample period and converge toward 1.5%, 1% and 1.9%, respectively. This indicates that forward PE ratios account for any structural shifts in the economy over the sample and so their intercept and slope coefficients are stable. By contrast, the deltas for the PD and PE ratios are much larger at 9.9% and 5.5%, respectively, and diverge exactly when there is a structural shift in firms' payout method and firms' earnings volatility, respectively, in line with the theory. For the CAPE the divergence occurs early on and is constant throughout the sample suggesting that taking an average of past earnings consistently yields unstable intercept and slope coefficients. Finally, we test whether the forward PE ratio using objective machine learning forecasts can outperform the forward PE ratio using subjective market forecasts. To proxy for the market's expectations of future earnings, we use forecasts of equity analysts. We find the forward PE ratio constructed with the machine learning forecasts outperforms the one constructed using market forecasts by nearly 3%. This indicates market expectations of future cash flows are not fully optimal – in line with prior work (e.g., Bordalo, Gennaioli, LaPorta, and Shleifer, 2022; Bianchi, Ludvigson, and Ma, 2022, etc.) — and that the forward PE ratios are extracting both market expected returns and the cash flow news that arises from biased market expectations.

The rest of this paper is organized as follows. Section 2 shows theoretically and empirically why traditional price-ratios have failed to predict returns. Section 3 delineates the theory behind why forward price-ratios are the optimal valuation ratio for predicting returns. Section 4 discusses how we construct forward price-ratios. Section 5 uses forward price-ratios to predict returns. Section 6 concludes.

2 Return prediction and the omitted variable problem

2.1 Some basic theory

Price and return decompositions. We start with some rather basic price and return decompositions that are useful for deriving some of our theoretical results. The Campbell and Shiller (1988) identity states that the PD ratio can be decomposed into (for readability we ignore constants in this paper, because they are irrelevant for the results)

$$pd_t = \Delta d_{t+1} - r_{t+1} + \kappa pd_{t+1}, \tag{1}$$

where $\kappa = \frac{\exp(\overline{pd})}{1+\exp(\overline{pd})}$. Iterating forward on pd_{t+1} we get (under the assumption that the no-bubble condition holds)

$$pd_t = \sum \Delta d_{t+1} - \sum r_{t+1}, \tag{2}$$

where we use the short-hand notation $\sum \Delta d_{t+1} = \sum_{j=0}^{\infty} \kappa^j \Delta d_{t+1+j}$ and $\sum r_{t+1} = \sum_{j=0}^{\infty} \kappa^j r_{t+1+j}$. Taking expectations yields

$$pd_t = \sum \mathbb{E}_t \Delta d_{t+1} - \sum \mathbb{E}_t r_{t+1} \tag{3}$$

where we have defined similarly $\sum \mathbb{E}_t \Delta d_{t+1} = \sum_{j=0}^{\infty} \kappa^j \mathbb{E}_t \left[\Delta d_{t+1+j}\right]$ and $\sum \mathbb{E}_t r_{t+1} = \sum_{j=0}^{\infty} \kappa^j \mathbb{E}_t \left[r_{t+1+j}\right]$. Note that equation (3) must hold for any investor who understands the present-value identity.

We can also use the Campbell-Shiller equation to decompose realized returns into expected returns, cash flow news and discount rate news (Campbell, 1991):

$$r_{t+1} = \mathbb{E}_t r_{t+1} + \eta_{CF,t+1} - \eta_{DR,t+1} \tag{4}$$

where $\eta_{CF,t+1} = (\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=0}^{\infty} \kappa^j \Delta d_{t+1+j}$ and $\eta_{DR,t+1} = (\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=1}^{\infty} \kappa^j \Delta r_{t+1+j}$. Summing over all future periods yields

$$\sum r_{t+1} = \sum \mathbb{E}_t r_{t+1} + \sum \eta_{CF,t+1} - \sum \eta_{DR,t+1} \tag{5}$$

where $\sum \eta_{CF,t+1} = \sum_{j=0}^{\infty} \kappa^j \eta_{CF,t+1}$ and $\sum \eta_{DR,t+1} = \sum_{j=0}^{\infty} \kappa^j \eta_{DR,t+1}$.

The summation term for cash flows is given by

$$\sum \eta_{CF,t+1} = \sum_{j=0}^{\infty} \kappa^j \left(\mathbb{E}_{t+1+j} - \mathbb{E}_{t+j}\right) \left[\sum_{i=0}^{\infty} \kappa^i \Delta d_{t+1+j+i}\right] = \sum_{j=0}^{\infty} k^j (\Delta d_{t+1+j} - \mathbb{E}_t \left[\Delta d_{t+1+j}\right]) \tag{6}$$

and similarly for the summation of discount rate news

$$\sum \eta_{DR,t+1} = \sum_{j=0}^{\infty} \kappa^j \left(\mathbb{E}_{t+1+j} - \mathbb{E}_{t+j}\right) \left[\sum_{i=1}^{\infty} \kappa^i r_{t+1+j+i}\right] = \sum_{j=1}^{\infty} k^j (\mathbb{E}_{t+j} \left[r_{t+1+j}\right] - \mathbb{E}_t \left[r_{t+1+j}\right]). \tag{7}$$

We will use these price and return decompositions throughout the paper.

Return prediction. Return prediction using linear regressions is a simple problem. One regresses future returns, i.e. return that occur beyond time t, on a single predictor variable that is formed as of time t. For analytical tractability, we mostly focus on long-horizon returns $\sum r_{t+1}$. This could approximate the perspective of investors who have a very long holding period. Let us predict long-horizon returns with predictor variable x_t, then the long-horizon return regression is given by

$$-\sum r_{t+1} = a_x + b_x x_t + \epsilon_t \tag{8}$$

The (population) regression coefficient b_x is then determined by

$$b_x = \frac{Cov(-\sum r_{t+1}, x_t)}{\sigma^2(x_t)} = \frac{Cov(-\sum \mathbb{E}_t r_{t+1}, x_t)}{\sigma^2(x_t)}, \tag{9}$$

where we use the fact that (i) we can use equation (5) to decompose the realized returns and (ii) future cash flows news and future discount rate news are unpredictable as of time t if the expectations $\mathbb{E}$ are formed rationally. Formally, this means $Cov(-\sum \eta_{CF,t+1}, x_t) = Cov(-\sum \eta_{DR,t+1}, x_t) = 0$

The R-squared – our main object of interest in this paper – is the squared correlation coefficient (see the appendix for the derivation as well derivations of the following results)

$$R_x^2 = Corr(-\sum r_{t+1}, x_t)^2 = \frac{Cov(-\sum r_{t+1}, x_t)^2}{\sigma^2(-\sum r_{t+1})\sigma^2(x_t)} = \frac{Cov(-\sum \mathbb{E}_t r_{t+1}, x_t)^2}{\sigma^2(-\sum r_{t+1})\sigma^2(x_t)}. \tag{10}$$

Because only expected returns are predictable, the maximum R-squared can be obtained by a regressor that is perfectly correlated with expected returns, i.e. $\tilde{x}_t = \gamma \sum \mathbb{E}_t r_{t+1}$. Such a regressor would yield the maximum R-squared of

$$R^2_{\max} = \frac{Cov(-\sum \mathbb{E}_t r_{t+1}, \gamma \sum \mathbb{E}_t r_{t+1})^2}{\sigma^2(-\sum r_{t+1})\sigma^2(\gamma \sum \mathbb{E}_t r_{t+1})} = \frac{\sigma^2(\sum \mathbb{E}_t r_{t+1})}{\sigma^2(\sum r_{t+1})}. \tag{11}$$

2.2 The omitted variable problem

Price ratios that scale prices by a measure of past cash flows, such as the PD or PE ratio, can be such a predictor variable x_t. From the Campbell-Shiller decomposition given by equation (3), we can immediately see why the PD ratio is theoretically appealing for forecasting returns: it contains expected returns. Re-arranging the equation yields

$$-\sum \mathbb{E}_t r_{t+1} = pd_t - \sum \mathbb{E}_t \Delta d_{t+1} \tag{12}$$

We can see why there is an omitted variable problem when forecasting returns with the PD ratio: Expected cash flow growth show up on the right-hand side of this equation. Or in other words, the price-dividend ratio not only contains expected returns, but also expected dividend growth. That intuition is well known (Lettau and Ludvigson, 2005; Cochrane, 2008). While the PD ratio must forecast returns, the fact that we are not controlling for cash flow growth weakens the predictive power of the PD ratio. One potential way to overcome this issue is to estimate a joint system of expected returns and cash flow growth (Van Binsbergen and Koijen, 2010, e.g.).

Let us examine of what happens if we simply run a regression of returns on the PD ratio as does most prior studies (Goyal and Welch, 2008, e.g.,), i.e., we use pd_t as predictor in equation (8). The (population) slope coefficient is given by (see appendix for the derivation)

$$b_{pd} = \frac{1 - \rho_{r,\Delta d} \frac{\sigma(\sum \mathbb{E}_t \Delta d_{t+1})}{\sigma(\sum \mathbb{E}_t r_{t+1})}}{1 + \frac{\sigma^2(\sum \mathbb{E}_t \Delta d_{t+1})}{\sigma^2(\sum \mathbb{E}_t r_{t+1})} - 2\rho_{r,\Delta d} \frac{\sigma(\sum \mathbb{E}_t \Delta d_{t+1})}{\sigma(\sum \mathbb{E}_t r_{t+1})}} \tag{13}$$

where $\rho_{r,\Delta d}$ is the correlation between $\sum \mathbb{E}_t \Delta d_{t+1}$ and $\sum \mathbb{E}_t r_{t+1}$. Thus, the slope coefficient depends crucially on two factors. First, it depends on the correlation between expected returns and expected dividend growth. Second, it depends on the ratio of the volatility of expected cash flow growth relative to the volatility of expected returns. Thus, if these two parameters change over time – and we show empirical evidence below which suggests that this is likely the case –, then this leads to a change in the slope coefficient. Of course, this will make out-of-sample prediction hard, as it relies on parameter estimates that are estimated in prior data samples. We can see that expected cash flow growth would be constant, then would obtain a slope coefficient of one.

There is also an economic interpretation of the slope coefficient b_{pd}. We know from the Campbell-Shiller identity that all movements in the PD ratio must either come from movements in expectation of cash flows or movements. We can start with equation (3), take the covariance of both sides with pd_t and divide by the variance of pd_t. It follows that b_{pd} measures the variation of the PD ratio that is driven by cash flow expectations. Vice versa, $1 - b_{pd}$ then gives the variation of the PD ratio that is driven by return expectations.

Returning to our regression, the intercept is given by (using equation (3))

$$a_{pd} = (1 + b_{pd}) \sum \mathbb{E}_t r_{t+1} - b_{pd} \sum \mathbb{E}_t \Delta d_{t+1}. \tag{14}$$

Because this depends on the slope coefficient, b_{pd}, the intercept also depends on the correlation between expectations of cash flows and expectations of returns as well as the volatility ratio of these two expectations. In addition, it also depends on the level of expected cash flow growth $\sum \mathbb{E} \Delta d$. Thus, if there are changes in the expectation about long-run cash flow growth over time, then the regression intercept is going to change. This is particularly important (and problematic) in light of structural shift towards repurchases which we will highlight in more detail below.

The (in-sample) R-squared is given by

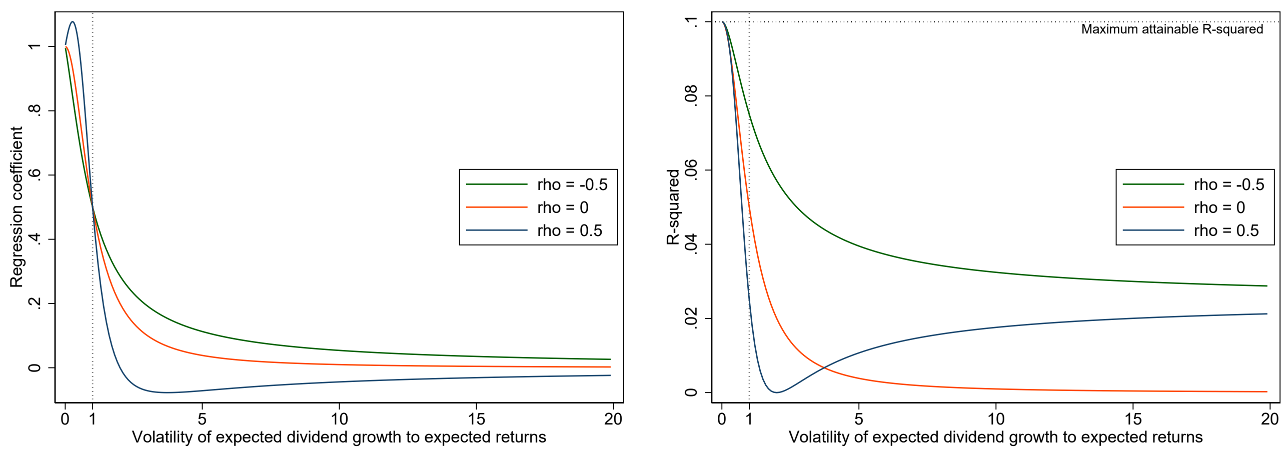
$$R^2_{pd} = \frac{\sigma^2(\sum \mathbb{E}_t r_{t+1})}{\sigma^2\left(\sum r_{t+1}\right)} \cdot \frac{\left[1 + \rho^2_{r,\Delta d} \frac{\sigma^2(\sum \mathbb{E}_t \Delta d_{t+1})}{\sigma^2(\sum \mathbb{E}_t r_{t+1})} - 2\rho_{r,\Delta d} \frac{\sigma(\sum \mathbb{E}_t \Delta d_{t+1})}{\sigma(\sum \mathbb{E}_t r_{t+1})}\right]}{\left[1 + \frac{\sigma^2(\sum \mathbb{E}_t \Delta d_{t+1})}{\sigma^2(\sum \mathbb{E}_t r_{t+1})} - 2\rho_{r,\Delta d} \frac{\sigma(\sum \mathbb{E}_t \Delta d_{t+1})}{\sigma(\sum \mathbb{E}_t r_{t+1})}\right]} \tag{15}$$

Thus, the in-sample predictive power depends also correlation between expectations of cash flows and expectations of returns $\rho_{r,\Delta d}$ as well as the volatility ratio of these two expectations. While the in-sample R-squared is not affected by any changes in long-run expected cash flow growth, the out-of-sample R-squared will also be.

Figure 1: The omitted variable problem

(A) Slope coefficient

(B) In-sample R-squared



Note: This figure shows the regression coefficient and in-sample R-squared when predicting long-run returns $\sum r_{t+1}$ with the PD ratio pd_t for (i) varying correlations between expectations of cash flows and expectation of returns $\rho_{r,\Delta d}$ and (ii) varying volatility ratios $\sigma(\sum \mathbb{E}_t \Delta d)/\sigma(\sum \mathbb{E}_t r)$.

Figure 1 shows the slope coefficient and the in-sample R-squared for different values of the correlation $\rho_{r,d}$ and the volatility ratio $\frac{\sigma(\sum \mathbb{E}_t \Delta d)}{\sigma(\sum \mathbb{E}_t r)}$. We can see that the R-squared drops sharply when the volatility ratio rises from 0 to about 1. Intuitively, this means that more of the variation in the PD ratio is driven by expectation of cash flows growth and therefore the PD ratio will be a noiser measure of expected returns and the R-squared goes down.

Out-of-sample prediction. The problem that investors face is an out-of-sample problem. They want to forecast returns with the data that is available up to period t. In out-of-sample regression, we are therefore forecasting returns with regression coefficients that have been estimated in prior data.

Let's imagine that these estimated coefficients are $\tilde{a}$ and $\tilde{b}$. Then, we run the regression in the unseen data using these estimates. To fix the idea, let us say the optimal coefficient, i.e. the in-sample coefficients, for this new sample are a and b. By definition, the coefficient a and b are the coefficient that maximize the R-squared (as they minimize the squared errors). Thus, using the previously estimated coefficients $\tilde{a}$ and $\tilde{b}$ will inherently decrease the predictive power of the PD ratio.

To see how much the prediction gets worse, we can compare the out-of-sample R-squared, i.e. the R-squared from a prediction using $\tilde{a}$ and $\tilde{b}$, with the in-sample R-squared, i.e. the R-squared from a prediction using a and b. Let us define $\Delta a = \tilde{a} - a$ and $\Delta b = \tilde{b} - b$, then we can show that

$$\frac{R^2_{pd} - R^2_{OOS,pd}}{R^2_{pd}} = \frac{(\Delta a)^2 + (2\Delta a + \Delta b)\,\Delta b \cdot \mathbb{E}\left[pd_t\right] + (\Delta b)^2 \cdot \sigma^2(pd_t)}{b^2\sigma^2(pd_t)}. \tag{16}$$

What do we learn from this exercise? The equation says that the R-squared difference increases in the difference of the estimated coefficient and the optimal in-sample coefficients. Thus, if the underlying economy changes a lot over time, then the out-of-sample predictive power goes down. There is one additional message compared to the in-sample R-squared. While the in-sample R-squared is not affected by $\sum \mathbb{E}_t \Delta d_{t+1}$, the out-of-sample R-squared is (through the changes in the intercept Δa).

We derive three main insights from this theoretical analysis. First, price ratios, such as the PD ratio, are natural predictors for returns given that they contain expectations of returns. Second, while are natural predictors, they entail an omitted variable problem: they also contain expectations of cash flows. Depending on the characteristics of cash flow and return expectations, this decreases the in-sample predictive power for returns. Third, time-variation in these characteristics amplifies this problem when we try to predict returns out-of-sample.

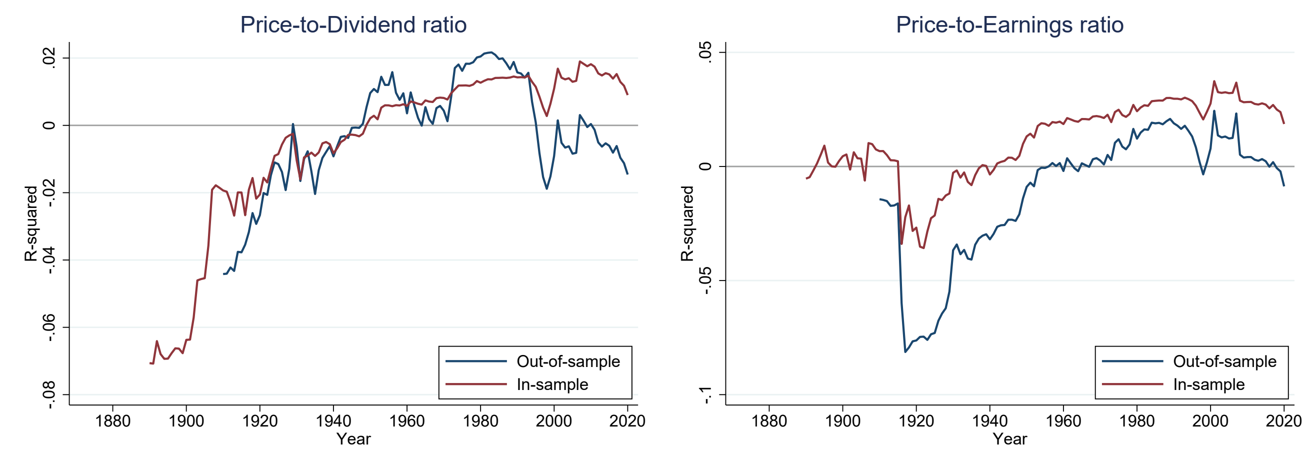
2.3 Why trailing price ratios fail to predict returns

In this section we try to use these insights to understand why commonly-used price ratio, such as the PD and the PE ratio, struggle to predict returns (out-of-sample). Because these price ratios scale prices by measures of trailing cash flows, we call them "trailing price ratios". Figure 2 repeats the exercise of Goyal and Welch (2008) and shows the poor performance of trailing price ratios. It shows that the R-squared for the PD and PE ratio is close to zero and slightly negative. Why do these price ratios have such a hard time forecasting returns despite their apparent appeal?

The shift towards repurchases and the PD ratio. Firm earnings can either be paid out to shareholders (in the form of dividends or share repurchases) or can be retained to invest into the firm to generate future payouts to shareholders. Since 1980 there is a strong trend towards firms paying out their earnings in the form of repurchases and not in the form of dividends. Boudoukh, Michaely, Richardson, and Roberts (2007) documents this trend until 2003, but Figure 3 shows that this trend has only strengthened in the recent two decades. Especially, since the 2000s, firms pay out only half of their entire payouts in the form of dividends (panel A). If firms still pay out the same total amount as they did in the past, then this shift towards repurchases implies that dividends have grown slower over the past decades. This means that scaling prices by "raw" dividends leads to an extreme increase in the price-dividend ratio

Figure 2: Empiricial performance of trailing price ratios

(A) Price-to-Dividend Ratio (B) Price-to-Earnings Ratio

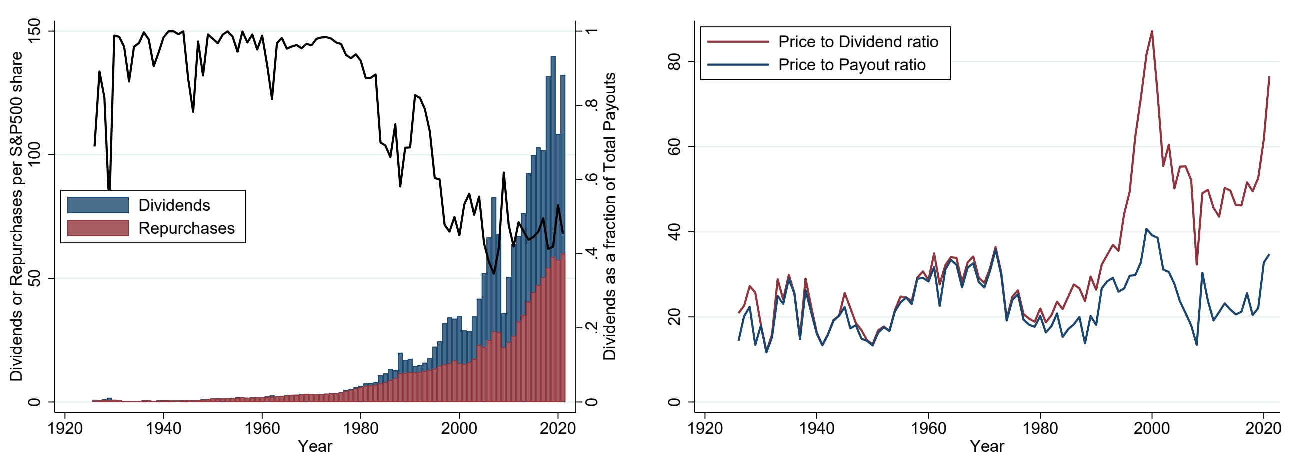


Note: The figure shows the predictive power when predicting one-year log returns with the log PD ratio in Panel (A) and the log PE ratio in Panel (B) for the S&P 500 index. One-year returns, the S&P 500 index, earnings and dividends prior 1926 are from Robert Shiller's website. Dividends after 1926 are from CRSP. For the out-of-sample exercise, we require 20 years of data to fit the regression and we plot the R-squared after predicting 20 years of returns out-of-sample. We use an expanding window to fit the regression.

(panel B). If instead we scale prices by total payouts, then we see that the price ratio looks more stable.

Figure 3: Repurchases and price ratios

(A) Dividends and Repurchases (B) Price ratios



Note: This figure shows the trend towards repurchases. Panel A shows dividends and repurchases at the S&P 500 level (on a per share basis). The black line shows the fraction of total payouts (dividends plus repurchases) that are paid out as dividends. Panel B shows price ratios scaling the price by trailing dividends and total payouts.

How does the shift towards repurchases impact return prediction? Let us assume that the firm payouts a fraction α_t of the total payouts O_t as dividends D_t and fraction $1 - \alpha_t$ as repurchases. Scaling log prices by log dividends now yields

$$pd_t = \log\left(\frac{P_t}{D_t}\right) = \log\left(\frac{P_t}{\alpha_t O_t}\right) = po_t - log(\alpha_t). \tag{17}$$

Similarly we can decompose the PD ratio into

$$pd_t = \sum \mathbb{E}_t\left[\Delta\alpha_{t+1}\right] + \sum \mathbb{E}_t\left[\Delta o_{t+1}\right] - \sum \mathbb{E}_t\left[r_{t+1}\right], \tag{18}$$

where $\sum \Delta\alpha_{t+1} = \sum_{j=0}^{\infty} \kappa^j \Delta\alpha_{t+1+j}$ and $\sum \Delta o_{t+1} = \sum_{j=0}^{\infty} \kappa^j \Delta o_{t+1+j}$. Thus, a shift towards repurchases would show up in $\Delta\alpha$. Thus, even if $\sum \Delta o_{t+1}$ was stable over time, the shift towards repurchases would cause an omitted variable problem when predicting returns. This undermines the predictive power of the PD ratio over long samples and when trying to predict returns out-of-sample. By contrast, focusing on in-sample prediction for shorter samples produces better results (Dybvig and Zhang, 2018).

Earnings volatility and the PE ratio. Dividends are not a good measure of firm payouts, i.e. cash flows to shareholders, since there has been a shift to repurchases. Alternatively, we can scale prices by earnings to get the PE ratio.

Let's start with earnings and define the log payout ratio $\delta_t = d_t - e_t$. Using this, we can rewrite the Campbell-Shiller identity

$$pe_t = \Delta e_{t+1} - r_{t+1} + (1-\kappa)\delta_{t+1} + \kappa pe_{t+1} \tag{19}$$

Iterating forward on pe_{t+1} we obtain

$$pe_t = (1-\kappa)\sum \delta_{t+1} + \sum \Delta e_{t+1} - \sum r_{t+1} \tag{20}$$

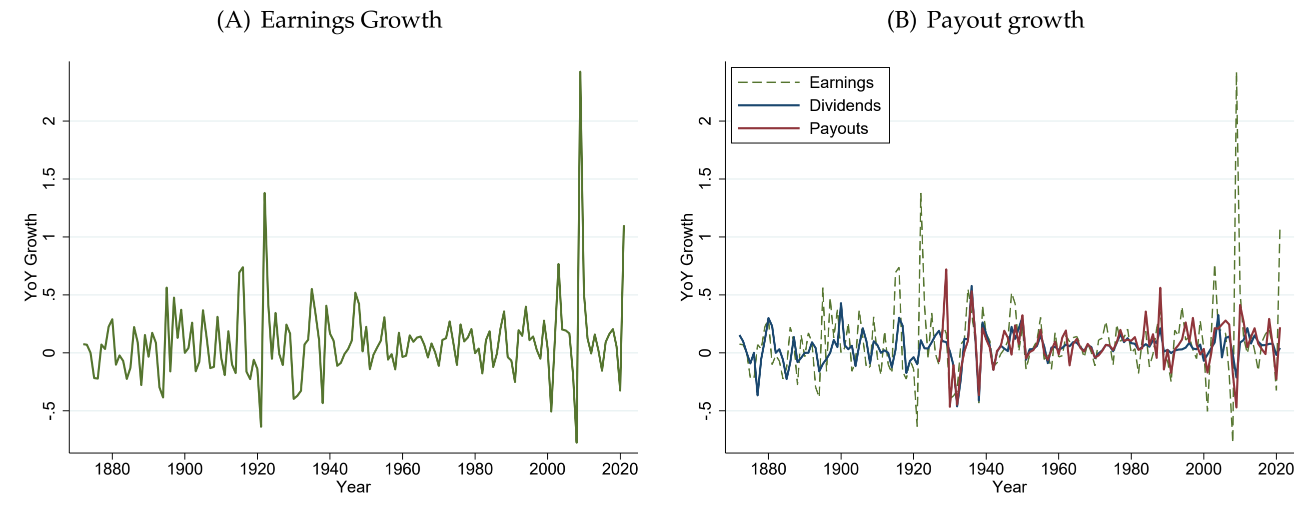
which we can alternatively re-write as

$$pe_t = \delta_t + \sum \Delta\delta_{t+1} + \sum \Delta e_{t+1} - \sum r_{t+1}. \tag{21}$$

Thus, under this alternative definition the PE ratio is a function of expected earnings growth, expected changes in the payout ratio and expected returns.

Unfortunately, using the PE ratio does not solve the return predictability problem either. The reason is that the volatility of earnings has shifted dramatically over the last decades. Especially, under the most commonly used accounting standard, i.e. generally accepted accounting principles (GAAP), earnings have become massively volatile over the last three decades (Hillenbrand and McCarthy, 2022).

Figure 4: The volatility of earnings and payouts



Note: This figure shows the time-variation in earnings and payout volatility. Earnings and dividends prior to 1926 are from Robert Shiller's website. Dividends since 1926 and total payouts (the sum of dividends and repurchases) are from CRSP.

Figure 4 documents the secular trends in earnings volatility over the last 150 years. It shows that there are approximately three regimes: a regime with high earnings volatility up until 1945, followed by a regime of subdued earnings volatility, followed again by a regime of high volatility. Based on the theory outlined above, we know that a high earnings volatility entails a large omitted variable problem for any return prediction. The theory predicts that the in-sample predictive power would be particularly low for the periods with high earnings volatility. The table documents that this prediction is exactly borne out in the data. The in-sample R-squared is below 4% for the periods 1987 – 1944 and 1990 – 2022, but is much higher (15%) in the period 1945 – 1989. Additionally, the table reports that the fraction of the PE ratio movements that can be explained by one-year earnings growth is 41% in the sample from 1871 to 1944 and 56% in the sample from 1990 to 2022. This numbers rise to 49% and 87%, respectively, when looking at two-year earnings growth. Thus, much of the variation in the PE ratio does not come from expected returns explaining the dismal in-sample R-squared.

Did the earnings volatility trickle down to volatility in payouts to shareholders? Table 1 shows that the volatility regimes also holds when we look at total payouts (dividends plus repurchases) as well as for dividends. Unsurprisingly, we then the find same pattern for the in-sample R-squared that we documented for earnings. Similarly, the fraction of price ratio that is explained by cash flow growth varies substantially over the subsamples. Our facts presented here are consistent with Chen (2009) who shows that dividend growth predictability varies significantly over time.

This section was mostly concerned with illustrating the omitted variable problem theoretically and

Table 1: The volatility of earnings and payouts

	1871-1944	1945-1989	1990-2022
Earnings			
Mean of log earnings growth	1.2%	7.2%	6.7%
Volatility of log earnings growth	0.29	0.14	0.44
In-sample R-squared when predicting 1-year returns	0.01	0.15	0.04
Fraction price-to-earnings ratio explained by 1-year earnings growth	0.41	0.08	0.56
Fraction price-to-earnings ratio explained by 2-year earnings growth	0.49	0.09	0.87
Payouts (=Dividends + Repurchases)			
Mean of log payout growth	0.8%	7.1%	6.4%
Volatility of log payout growth	0.19	0.11	0.20
In-sample R-squared when predicting 1-year returns	0.00	0.25	0.03
Fraction price-to-payout ratio explained by 1-year payout growth	0.52	-0.00	0.33
Fraction price-to-payout ratio explained by 2-year payout growth	0.64	-0.09	0.39
Dividends			
Mean of log dividend growth	1.2%	6.5%	5.1%
Volatility of log dividend growth	0.17	0.07	0.09
In-sample R-squared when predicting 1-year returns	0.02	0.25	0.14
Fraction price-to-dividend ratio explained by 1-year dividend growth	0.47	-0.05	0.06
Fraction price-to-dividend ratio explained by 2-year dividend growth	0.62	-0.11	0.02

Note: This table shows the time-variation in earnings, dividend and payout volatility and its implication for return predictability. For each cash flow measure, the third row reports the in-sample R-squared when predicting one-year log returns using the respective price ratio, i.e., the PE ratio for earnings. We also report the fraction of the price ratio that can be explained by one-year and two-year realized cash flow growth, i.e., for the PE ratio and one-year earnings growth, we report the regression coefficient $Cov(\Delta e_{t+1}, pe_t)/Var(\Delta pe_t)$. One-year returns, the S&P 500 index, earnings and dividends prior to 1926 are from Robert Shiller's website. Dividends since 1926 and total payouts (the sum of dividends and repurchases) are from CRSP.

documenting why it has a serious impact on the ability of trailing price ratios, such as the PD or the PE ratio, to predict returns. The next section tries to make progress by solving or at least mitigating the omitted variable problem.

3 Theory: The optimal valuation ratio

In this section we derive theoretical results that help us to come up with the optimal valuation ratio for predicting returns. We show that the optimal predictor for long-run returns is the long-run forward price ratio which takes into account all future expected cash flows. Intuitively, this means that once we take into future cash flows, the only variation in the price ratio that is left comes from expected returns. Putting it differently, this implies that the long-run forward price ratio achieves the maximum attainable R-squared for out-of-sample return prediction. We then derive a number of additional theoretical results that guide the empirical analysis that follows.

3.1 The optimal return predictor: long-run forward price ratios.

In this section, we will show formally that using forward price ratios, i.e. ratios that scale prices by forward-looking fundamentals are the best long-run return predictors.

Realized cash flow growth. To gain some intuition, let us start by subtracting the entire stream of future *realized* cash flow growth, $\sum \Delta d_t$, from the pd_t ratio, to construct the predictor $pd_t - \sum \Delta d_t$. Equation (2) implies

$$pd_t - \sum \Delta d_{t+1} = -\sum r_{t+1}. \tag{22}$$

Thus, by subtracting the cash flow growth, all that we are left are future returns. Using the left-hand size as predictor variable x_t in the return prediction regression (xx) leads to regression statistics given by

$$b_{pd_t - \sum \Delta d_{t+1}} = \frac{Cov(-\sum r_{t+1}, pd_t - \sum \Delta d_{t+1})}{\sigma^2(pd_t - \sum \Delta d_t)} = \frac{Cov(-\sum r_{t+1}, -\sum r_{t+1})}{\sigma^2(-\sum r_{t+1})} = 1 \tag{23}$$

$$R^2_{pd_t - \sum \Delta d_{t+1}} = Corr^2(-\sum r_{t+1}, pd_t - \sum \Delta d_{t+1}) = \frac{Cov^2(-\sum r_{t+1}, -\sum r_{t+1})}{\sigma^2(-\sum r_{t+1})\sigma^2(-\sum r_{t+1})} = 1. \tag{24}$$

Thus, perhaps unsurprisingly to the reader, we obtain the maximum R-squared of 1 when we run this regression.[2] Of course, one would not even need to estimate a regression to get this results as equation (22) directly reveals this result. Putting it differently, we do not need to estimate the relationship between future realized returns and our predictor as the relationship is already known.

Of course, the issue with this predictor is that an investors (or an any econometrician) would not

[2]This must hold in the data except for any approximation errors in the Campbell-Shiller approximation.

have access to this information, as realized cash flow growth after time t becomes only observable after time t. Running this regression is therefore a purely hypothetical exercise.

Expected cash flow growth. Instead of substracting the realized cash flows, we now want to subtract the entire stream of future expected cash flow growth $\sum \mathbb{E}_t \Delta d_{t+1}$. Let's define the long-run forward price ratio pd_t^∞ as

$$pd_t^\infty = pd_t - \sum \mathbb{E}_t \Delta d_{t+1}. \tag{25}$$

Equation (3) implies

$$pd_t^\infty = -\sum \mathbb{E}_t r_{t+1}. \tag{26}$$

Intuitively, this equation says that all variation in pd_t^∞ comes from expected returns. The results that now follow rest on this intuition.

Let us re-run the predictive regression with the long-run forward price ratio, i.e. use pd_t^∞ as the predictor variable x_t. The slope coefficient is given by

$$b_{pd^\infty} = \frac{Cov(-\sum r_{t+1}, pd_t^\infty)}{\sigma^2\left(pd_t^\infty\right)} = \frac{Cov(-\sum \mathbb{E}_t r_{t+1}, pd_t^\infty)}{\sigma^2\left(pd_t^\infty\right)} = \frac{Cov(-\sum \mathbb{E}_t r_{t+1}, -\sum \mathbb{E}_t r_{t+1})}{\sigma^2\left(-\sum \mathbb{E}_t r_{t+1}\right)} = 1 \tag{27}$$

and the R-squared for this regression is given by

$$R^2_{pd^\infty} = \frac{\sigma^2(\sum \mathbb{E}_t r_{t+1})}{\sigma^2\left(\sum r_{t+1}\right)}, \tag{28}$$

where we use the fact that $Cov(-\sum \eta_{CF,t+1}, pd_t^\infty) = Cov(-\sum \eta_{DR,t+1}, pd_t^\infty) = 0$ which holds if the expectations are formed rationally. Intuitively, forecast errors for cash flows and discount rates are not predictable as of time t. The following proposition states that the long-run forward price ratio is the optimal valuation ratio for forecasting returns.

Proposition 1. *The maximum attainable R-squared when forecasting long-run horizon returns is given by* $R^2 = \sigma^2(\sum \mathbb{E}_t r_{t+1})/\sigma^2\left(\sum r_{t+1}\right)$*. The long-run forward price ratio is therefore the best predictor for returns, i.e. it is the predictor that maximizes the R-squared when forecasting long-horizon returns.*

The proposition follows immediately from equation (26), but we also provide a formal proof in the appendix. This simple result is the key result in our study. The theoretical analysis that follows provides more guidance for the empirical analyses that follow later.

3.2 Rational vs. market expectations

Investment professionals' expectations ("market expectations") of future earnings are now readily available, for example in the widely-used Thomson Reuters I/B/E/S data. However, research suggest that these expectations biased and more importantly, that the bias could vary over time De Silva and Thesmar (2021); Bordalo, Gennaioli, LaPorta, and Shleifer (2022).

Is it better to use market or rational expectations when predicting returns? When the two are the same, i.e. market expectations are rational, then this question is meaningless. However, when the market deviates from rational expectations, then it is important to understand which is a better measure to scale prices.

To make progress on this question, let us denote the market expectation by $\mathbb{E}^M$ and define the market forward PD ratio as $pd_t^{M,\infty} = pd_t - \sum \mathbb{E}_t \Delta d_{t+1}$. If the market expectations satisfy the Campbell-Shiller identity (3), then we can write

$$pd_t^{M,\infty} = \sum \mathbb{E}_t^M r_{t+1}$$

Proposition 2. *When the market expectation deviate from rational expectation, the forward price ratio constructed using rational forecasts outperforms the market forward price ratio. That is, it has a higher R-squared when predicting returns.*

This can be seen immediately from proposition 1. We give a formal proof that illustrates the intuition of why this is the case.

Proof. Because we have assumed that the Campbell-Shiller identity also holds for the market expectation, we can use equation (3) to write

$$\sum \mathbb{E}_t \Delta d_{t+1} - \sum \mathbb{E}_t r_{t+1} = pd_t = \sum \mathbb{E}_t^M \Delta d_{t+1} - \sum \mathbb{E}_t^M r_{t+1}$$

Re-arranging this equation we get

$$\sum \mathbb{E}_t r_{t+1} = \sum \mathbb{E}_t^M r_{t+1} + \left(\sum \left[\mathbb{E}_t - \mathbb{E}_t^M \right] \Delta d_{t+1} \right) \tag{29}$$

Intuitively, the rational expectations of returns consist of two parts: (1) the market expectations of returns plus (2) the forecast bias of the market expectations. Following on this, using rational expectations of cash flows we can extract both terms, while using market expectations we can only extract the first component. To show this formally we need that predicting returns with $pd_t^{M,\infty}$ produces an R-squared lower than $\sigma^2(\sum \mathbb{E}_t r_{t+1})\sigma^2 / (\sum r_{t+1})$, i.e. the R-squared when predicting returns with pd_t^{∞}:

$$
\begin{aligned}
R^2_{pd^{M,\infty}} &= \frac{Cov^2(-\sum r_{t+1}, pd_t^{M,\infty})}{\sigma^2\left(\sum r_{t+1}\right)\sigma^2\left(pd_t^{M,\infty}\right)} = \frac{Cov^2(\sum \mathbb{E}_t r_{t+1}, \sum \mathbb{E}_t^M r_{t+1})}{\sigma^2\left(\sum r_{t+1}\right)\sigma^2\left(\sum \mathbb{E}_t^M r_{t+1}\right)} \\
&= Corr^2(\sum \mathbb{E}_t r_{t+1}, \sum \mathbb{E}_t^M r_{t+1})\frac{\sigma^2(\sum \mathbb{E}_t r_{t+1})}{\sigma^2\left(\sum r_{t+1}\right)} < \frac{\sigma^2(\sum \mathbb{E}_t r_{t+1})}{\sigma^2\left(\sum r_{t+1}\right)}
\end{aligned} \tag{30}
$$

where the second equality uses $Cov(-\sum \eta_{CF,t+1}, pd_t^{M,\infty}) = Cov(-\sum \eta_{DR,t+1}, pd_t^{M,\infty}) = 0$. □

3.3 Short-horizon vs. long-horizon returns

So far, we have only considered long-horizon returns (where we sum all returns until infinity). However, in practice investment managers have a finite holding period or investment horizon, and would be therefore interested in forecasting a finite sum of returns. In addition, many prior studies has considered short-term returns, e.g. one-year returns. How well does the long-run forward price ratio when we predict one-year returns?

Let us decompose the term structure of expected returns into two components, a short-run (one-year) component and a long-run components $R_1 = \mathbb{E}_t r_{t+1}$ and $R_2 = \sum_{j=1}^{\infty} \kappa^j \mathbb{E}_t\left[r_{t+1+j}\right]$, which means

$$
R_1 + R_2 = \sum \mathbb{E}_t r_{t+1} \tag{31}
$$

Let's imagine we run the one-year return prediction (where the tilde denotes statistics for the one-year return prediction)

$$
r_{t+1} = \tilde{\beta}_0 + \tilde{\beta}_{pd^{\infty}} pd_t^{\infty} + \tilde{\varepsilon}_t \tag{32}
$$

Then we can derive the R-squared to be (see the appendix for the derivation)

$$
\widetilde{R}^2_{pd^{\infty}} = \frac{\sigma^2\left(R_1\right)}{\sigma^2\left(r_{t+1}\right)} \cdot \frac{1 + 2 \cdot \rho_{R_1R_2} \cdot \frac{\sigma(R_2)}{\sigma(R_1)} + \rho^2_{R_1R_2} \cdot \frac{\sigma^2(R_2)}{\sigma^2(R_1)}}{1 + 2 \cdot \rho_{R_1R_2}\frac{\sigma(R_2)}{\sigma(R_1)} + \frac{\sigma^2(R_2)}{\sigma^2(R_1)}} \tag{33}
$$

where $\rho_{R_1R_2}$ is the correlation between R_1 and R_2. If we assume that this correlation equals one, then second term would equal one implying that the R-squared we obtain is $\frac{\sigma^2(R_1)}{\sigma^2(r_{t+1})}$.

Proposition 3. *If short-term expected returns (1year returns) and long-horizon expected returns (everything beyond 1-year) are perfectly correlated, then the long-run forward price ratio pd_t^{∞} is the optimal predictor.*

We can use the same logic as in proposition 1 to conclude that the maximum R-squared when forecasting one-year returns is R_1. In other words, any predictor that is perfectly correlated with one-year expected returns yields the maximum attainable predictive power. The fact that the short-run and the long-run component are perfectly correlated is an assumption that is very common in the literature. In particular, it is very common in the literature to assume that expected returns following autoregressive processes (Pástor and Stambaugh, 2009; Van Binsbergen and Koijen, 2010, e.g.,).

What if expectations of short-run and long-run expectation are not perfectly related? For example, Cochrane (2011) argues that adding a high-frequency variables such as the cay factor (Lettau and Ludvigson, 2001a,b, 2005) to the predictive regression can increase the in-sample explanatory power. We can get some simple intuition from the rather extreme case where R_1 and R_2 are uncorrelated. In this case, the R-squared is given by

$$R^2_{pd^\infty} = \frac{\sigma^2\left(R_1\right)}{\sigma^2\left(r_{t+1}\right)} \cdot \frac{\sigma^2\left(R_1\right)}{\sigma^2\left(R_1\right) + \sigma^2\left(R_2\right)}.$$

We can see this R-squared gets closer to the optimal R-squared if more of the variation in prices comes from short-term expected returns rather than long-term expected returns. And vice versa, the more of the variation comes from expectations of long-term returns, the less strong is the predictive power of pd^∞ for one-period returns.

3.4 Short-run vs. long-run forward price ratios

Obtaining or making forecasts for the entire term structure of expected cash flows is infeasible. In our empirical analysis that follows, we obviously use finite long-run forecast to construct the forward price ratio. When is this ratio still the optimal predictor?

For illustration, we focus on the one-year forward PD ratio. Subtracting $\mathbb{E}_t\left[\Delta d_{t+1}\right]$ from both sides of the Campbell-Shiller equation (3) yields the 1-year forward PD ratio pd_t^1:

$$pd_t^1 = p_t - \mathbb{E}_t\left[d_{t+1}\right] = \sum_{j=2}^{\infty} \kappa^{j-1} \mathbb{E}_t\left[\Delta d_{t+j}\right] - \sum \mathbb{E}_t\left[r_{t+1}\right]. \tag{34}$$

We can re-arrange this to get

$$\sum \mathbb{E}_t\left[r_{t+1}\right] = pd_t^1 - \sum_{j=2}^{\infty} \kappa^{j-1} \mathbb{E}_t\left[\Delta d_{t+j}\right] \tag{35}$$

Thus, the one-year forward price ratio is a worse predictor than the long-run forward price ratio as it eliminates less of the omitted variable problem. Putting it differently, it is a less clean measure of expected returns because it also contains expectations of cash flow growth.

When is the one-year forward price ratio the optimal predictor? Imagine that the cash flow expectations beyond the next period, i.e., beyond $t+1$, are equal to a constant g, i.e. $\mathbb{E}_t\left[\Delta d_{t+j}\right] = g \quad \forall j \geq 2$. Putting it differently, dividend growth beyond the next year is unpredictable. Formally, this means

$$\sum_{j=1}^{\infty} \kappa^{j-1} \mathbb{E}_t\left[\Delta d_{t+j}\right] = \mathbb{E}_t\left[\Delta d_{t+j}\right] + \sum_{j=2} \kappa^{j-1} g = \mathbb{E}_t\left[\Delta d_{t+j}\right] + \frac{\kappa}{1-\kappa} g.$$

Thus, all variation in expected dividend growth comes from the cycle, i.e. next-periods growth,

$$\sigma^2\left(\sum \mathbb{E}_t \Delta d_{t+1}\right) = \sigma^2\left(\mathbb{E}_t \Delta d_{t+1}\right)$$

Corollary 1. *If cash flow growth only contains a one-period cycle component, i.e. cash flow growth is unpredictable beyond the next year, then the one-year forward price ratio is the optimal predictor for long-horizon returns.*

This follows because the maximum R-squared pd_t^1 is equal to the maximum attainable R-squared of $\sigma^2(\sum \mathbb{E}_t r_{t+1})/\sigma^2\left(\sum r_{t+1}\right)$. Intuitively, the one-year forward PD ratio removes the expected cash flow terms and therefore eliminates the omitted variable problem.

What if cash flow growth is predictable beyond year one? Then, we would need to use longer-term forecast to scale prices.

Corollary 2. *If cash flows growth predictability persist for some time into the future, then a longer-run forward price ratio outperforms a shorter-run forward price ratio.*

We finish our section on the construction of the optimal valuation ratio with this useful insight. A three-year forward price ratio should outperform a two-year forward price ratio which should outperform a one-year forward price ratio. We now turn to constructing an empirical optimal valuation ratio using our theoretical insights from this section.

4 Construction of forward price ratios

To construct optimal forward valuation ratios we need to construct optimal forecasts of future cash flow growth. To do so, we use Machine Learning algorithms conditions on the large set of information in the economy that would be available to researchers at the time of forecast. With these cash flow growth forecasts at the 1yr, 2yr, 3yr and 5yr horizon we compound the most recently available S&P index earnings at these growth forecasts to construct 1yr, 2yr, 3yr and 5yr aggregate earnings forecasts. Finally, we scale the SP500 index by these earnings forecasts to construct 1yr, 2yr, 3yr and 5yr forward price-to-earnings ("Xyr Fwd. PE") ratios.

4.1 Data sources

To obtain firm-level data prior to 1950, we use annual report data as collected from the Morningstar Industrial Manuals and used in series of papers (Graham, Leary, and Roberts, 2015; Graham and Leary, 2018; Graham, Kim, and Leary, 2020). This gives us nineteen independent forecasting variables for the pre-1950 period that capture firm's balance sheet and income statement dynamics overtime. For the post-1950 data period, we use data from Compustat and CRSP yielding a set of fourty-seven forecasting variables. The precise variables we use pre- and post-1950 are given in Appendix C.

We apply the following cleaning procedures before feeding the data into the machine learning algorithm. First, we drop financial firms.[3] Second, for the pre- and post-1950 periods, we drop firm-year observations which are missing the set of pre- and post-1950 required variables as laid out in Appendix C. This yields an average of 411 and 2,689 yearly firm-level pre- and post-1950. Third, we scale all firm-level variables by assets to transform them to stationary series. Additionally, to capture firm-size in the cross-section we standardize firm assets each year. Finally, we replace any missing observations with a value of zero.

For the dependent variable, we need a measure of (annualized) firm-level earnings growth over the 1yr, 2yr, 3yr and 5yr horizons. For this, we use the (annualized) growth in net income excluding excluding special, extraordinary, and non-recurring tax items over the 1yr, 2yr, 3yr and 5yr horizons. We windsorize firm-level growth at the 5%-95% level each year. To ensure no-look ahead bias we forecast earnings growth in June of every year assuming we have access to the annual firm-level data from the prior-year's annual report (i.e., we assume a minimum of six-months lag between a firm's annual report date and it's release data).

4.2 Machine learning algorithm to forecast earnings

To construct optimal forward valuation ratios we need to construct optimal forecasts of future cash flow growth. To do so, we build on a recent burgeoning literature using machine learning algorithms to construct optimal forecasts of macroeconomic quantities (e.g., Nagel, 2021; De Silva and Thesmar, 2021; Bianchi, Ludvigson, and Ma, 2022, etc.).

We use a Random Forest 5-fold cross validation machine learning procedure. We employ this procedure each year from the initial training set and forecast set until the end of the sample. To do so, we follow five steps. First, every June we forecast annualized firm-level earnings growth for the 1yr, 2yr, 3yr and 5yr forecast horizons allowing for 1yr, 2yr, 3yr and 5yr gaps between training and forecasting, respectively, to ensure no look-ahead bias in our forecasts. Second, within each training and forecast

[3]Specifically, we drop firms which have 2-digit SIC codes between 60 and 67.

Table 2: Regression Results: Firm-Level earnings growth predictions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Realized 1yr		Realized 2yr		Realized 3yr		Realized 5yr	
ML forecasted 1yr	0.96^{***}	0.99^{***}						
	(0.06)	(0.06)						
ML forecasted 2yr			0.87^{***}	0.95^{***}				
			(0.05)	(0.05)				
ML forecasted 3yr					0.67^{***}	0.76^{***}		
					(0.06)	(0.05)		
ML forecasted 5yr							0.24^{***}	0.29^{***}
							(0.05)	(0.06)
Firm FE		✓		✓		✓		✓
Regression R-squared (%)	8	14	6	14	6	15	1	13
N	120,875	120,875	120,875	120,875	120,875	120,875	120,875	120,875

Note: This table displays the results from regressing firm-level future realized earnings growth on firm-level predicted earnings growth as generated by the Machine Learning procedure. Columns (1), (3), (5) and (7) give the regression results for the 1yr, 2yr, 3yr and 5yr horizons, respectively, and columns (2), (4), (6) and (8) repeat these regression results controlling for firm-year fixed effects. Standard errors are clustered at the firm-year level and are shown in parantheses. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

loop, we mitigate the impact of outliers by replacing each independent variable observation in both the training set and forecast set with five times it's inter-quartile range for those observations which are outside this range, where the interquartile range is calculated from the training set. Third, within each training and forecast loop, we standardize each independent variable in both the training and forecast set, where the mean and variances are calculated from the training set. Fourth, we then train the algorithm where we grid search for the optimal hyper-parameters using the same hyper-parameter ranges as in De Silva and Thesmar (2021). Finally, we use two-hundred fifty ensembles and average the forecasts over all ensembles to produce our final forecast. To train the algorithm we use an expanding window where we set the initial training window to be such that the out-of-sample forecasts start 15yrs after the beginning of the sample for the 1yr, 2yr, 3yr and 5yr forecast horizons.[4] Furthermore, we expand the set of forecasting variables used by the machine as we progress through the sample given our set of information on the economy increases overtime.[5]

This procedure yields firm-level 1yr, 2yr, 3yr and 5yr earnings growth forecasts starting in 1946Q2. To display the forecasting power of the machine at predicting future firm-level earnings growth, Table

[4]This implies our initial training windows are 14yrs, 13yrs, 12yrs and 10yrs for the 1yr, 2yr, 3yr and 5yr forecast horizons, respectively.

[5]Per section 4.1, prior to 1950 we use data collected from Morningstar Industrial Manuals allowing us to construct roughly twenty forecasting variables but from 1950 onwards we have access to large array of accounting variables available in Compustat allowing us to construct roughly fifty forecasting variables.

2 gives the results of regressing future realized firm-level earnings growth on the Machine Learning generated firm-level earnings growth forecasts. Starting in column (1), we see the machine can significantly predict 1yr-ahead firm-level earnings level growth with a regression coefficient close to one (as one would expect if the machine forecasts are unbiased). Similarly, we find the machine is almost equally good at yielding unbiased forecasts for the 2yr, 3yr-ahead earnings growth per columns (3) and (5), respectively. When we extend the forecast horizon to 5rs, we see the performance of the machine deteriorates somewhat given the increased role for unforeseeable shocks at longer horizons (Bianchi, Ludvigson, and Ma, 2022), but we note the machine can still significantly forecast future 5yr earnings growth. Ultimately, we are more interested in forecasting time-series variation in the firm-level earnings growth (relative to cross-sectional variation). Accordingly, in columns (2), (4), (6) and (8) we repeat the regressions in columns (1), (3), (5) and (7), respectively, also controlling for firm fixed effects. We see the results are similar but that the regressions coefficients in all cases are closer to one.

To obtain proxies for earnings growth forecasts at the market-level, we aggregate the firm-level earnings growth forecasts by taking their market-capitalization weighted average each year.[6]

4.3 Forward PE ratios

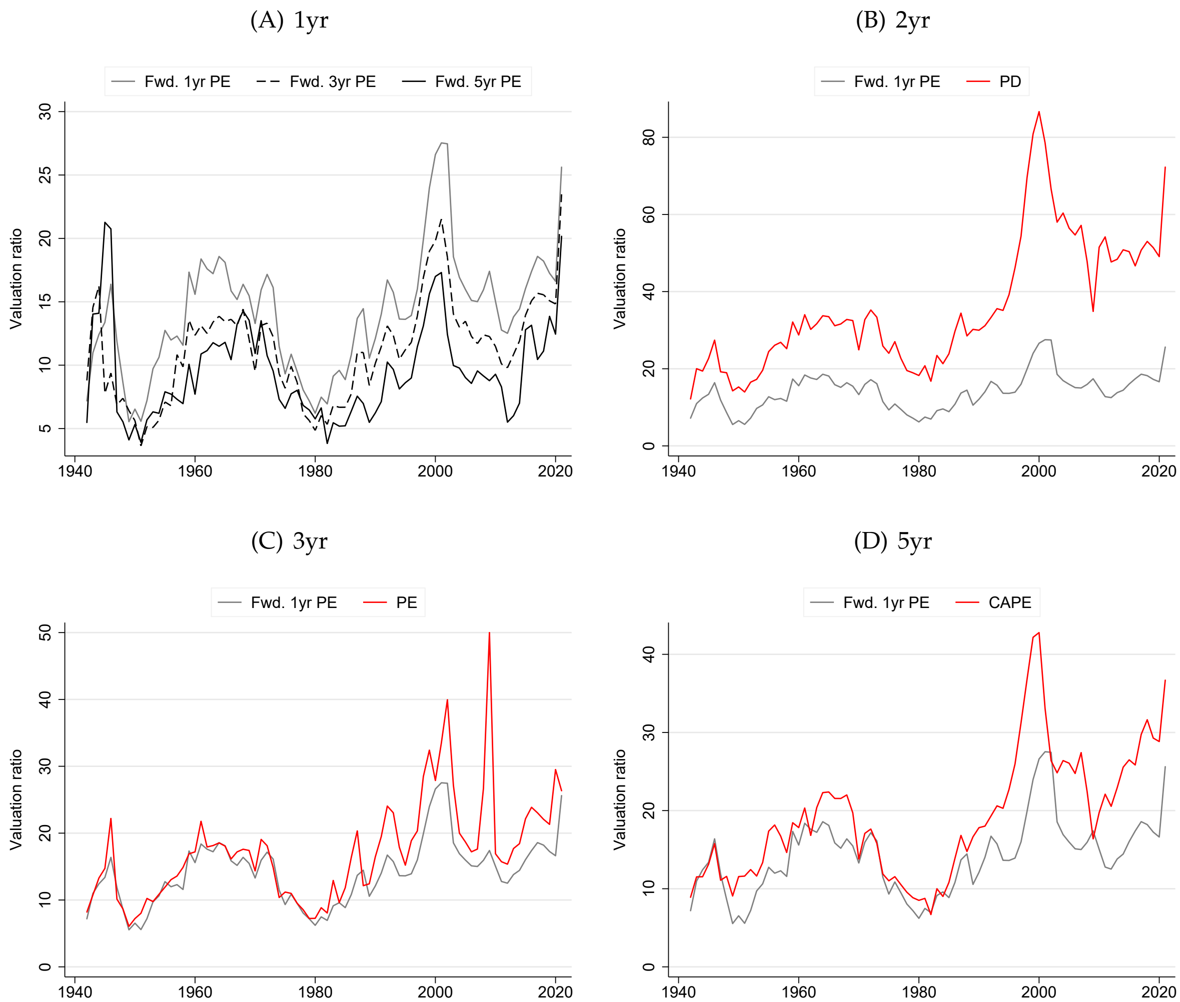
We construct market forward price ratios in three steps. First, we use realized S&P500 GAAP earnings per share (excluding extraordinary, discontinued, special and non-recurring tax items) over the past year to proxy for realized earnings per share ("Actual EPS") of the stock market.[7] Second, we compound Actual EPS at the Machine Learning generated 1yr, 2yr, 3yr and 5yr (annualized) earning growth forecasts to construct earnings per share forecasts at the 1yr, 2yr, 3yr and 5yr horizons, respectively. Third, we take the log difference between the S&P500 index and the log of the corresponding earnings forecasts, yielding 1yr, 2yr, 3yr and 5yr forward PE ratios.

Figure 5 Panel (A) plots the 1yr, 3yr and 5yr forward PE ratios against one another. We see the forward valuation ratios are stationary overtime fluctuation around their mean values. In Panel (B), we compare the Forward 1yr PE against the PD ratio where we see their is a structural upward shift in the latter ratio reflecting the structural shift by firms' away from dividends toward repurchases to pay out cash flows. In Panel (C) we plot the Forward 1yr PE against the PE ratio where we see the latter ratio is much more volatile and spikes in both the dot-com bust and financial crisis because of the transitory earnings items that hit firms' income statements in crises periods. Finally, in Panel (D) we plot the 1yr

[6]Results are robust if we take the earnings weighted average each year instead.

[7]We exclude extraordinary, discontinued, special and non-recurring tax to remove transitory items related to realized earnings which has contributed to the increased volatility in GAAP earnings in the past three decades(Hillenbrand and McCarthy, 2022).

Figure 5: Optimal Valuation Ratios



Note: Panels (A), (B), (C) and (D) plot the 1yr, 2yr, 3yr and 5yr forward (log) PE ratios against both the (log) price-to-dividend and the (log) price-to-earnings ratio, respectively. The time period is 1942 to 2020.

forward PE ratio against the (shiller) CAPE where we see the CAPE is subject to an upward structural shift post-2000. Indeed, the CAPE has been constructed to smooth prior earnings by using an average of earnings over the 10yrs but this creates an upward shift in the latter period because it includes large depressed earnings report (i.e., in the financial crisis) in its denominator which arguably the market does not believe will re-occur in the near future.

Finally, we report summary statistics for the aggregate earnings growth, forward PE ratios and traditional PE ratios (i.e., PD, PE and CAPE) in Appendix Table D.

5 Return prediction with forward PE ratios

5.1 Empirical framework

To compute a measure for the valuation ratios forecasting power at predicting future realized 1yr returns, we calculate their in-sample and out-of-sample r-squared following Goyal and Welch (2008). Concretely, the in-sample r-squared for valuation ratio x_t over the forecasting sample is calculated by

$$\text{In-sample } R^2 = 1 - \sum_{t=0}^{T} \frac{(r_{t+1} - \hat{r}_{t+1}^{IS})^2}{(r_{t+1} - \bar{r})^2} \tag{36}$$

where T is the number of years in the forecasting sample, $\bar{r} = \sum_{t=0}^{T} \frac{r_{t+1}}{T}$ is the mean future realized 1yr return over the forecasting sample and $\hat{r}_{t+1}^{IS}$ is the predicted future realized 1yr return generated by running the following in-sample regression over the full forecasting period

$$r_{t+1} = b_0 + b_1 x_t + \epsilon_{t+1} \tag{37}$$

The out-of-sample R^2 is computed analogously, except we re-run equation (37) for every year t using only data that was available prior to and including year t, thereby generating (i) a series of out-of-sample 1yr return predictions by running forecasting regression (37) ($\hat{r}_{t+1}^{OOS}$) and (ii) a series of mean realized 1yr returns by calculating $\bar{r}_t = \sum_{j=0}^{t-1} \frac{r_{j+1}}{t-1}$, thereby yielding the out-of-sample R^2 statistic given by

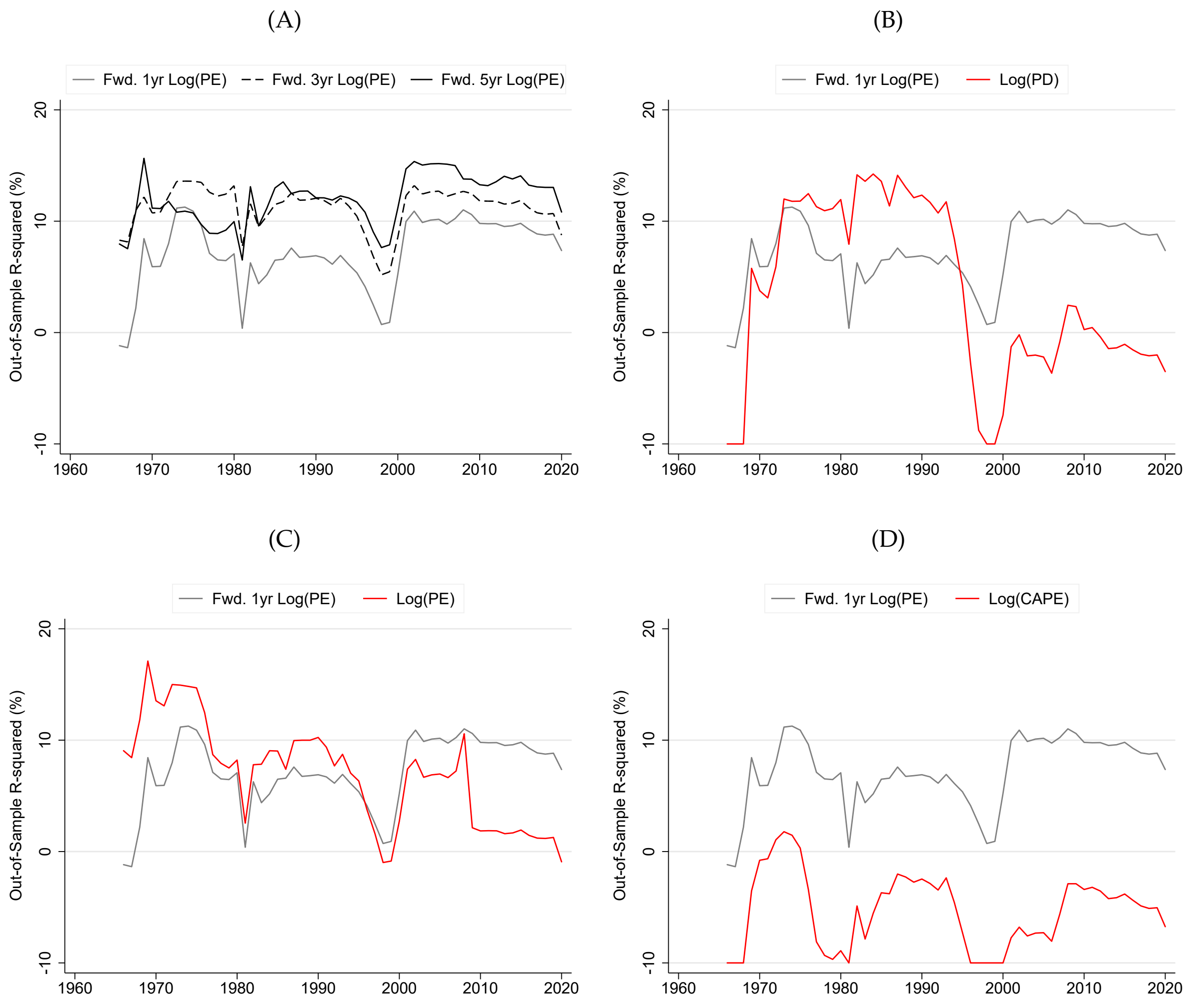
$$\text{Out-of-sample } R^2 = 1 - \sum_{t=0}^{T} \frac{(r_{t+1} - \hat{r}_{t+1}^{OOS})^2}{(r_{t+1} - \bar{r}_t)^2} \tag{38}$$

Finally, in the empirical implementation we start calculating return predictions only once we have at least ten years of data to generate regression coefficients. Given we have data on forward PE ratios starting in 1942, this implies we start calculating out-of-sample 1yr ahead return predictions in 1951.

5.2 Empirical Results

Out-of-sample return prediction. Figure 6 plots the evolution of the out-of-sample r-squared performance of the forward PE ratios against the PE, PD and CAPE ratios. Starting in Panel (A), we see the out-of-sample ("OOS") r-squared of the ML forward PE ratios range between 7.4% and 10.8% over the full-sample and are consistently above zero. These OOS r-squared statistics are significantly above that of any predictor variable tested in Goyal and Welch (2008)'s comprehensive review. Furthermore, the OOS r-squared of the 5yr forward PE ratio is 10.8%, which is higher than the OOS r-squared of the 3yr ratio (8.8%) which is higher than than the 1yr ratio (7.4%) exactly in line with our theoretical predic-

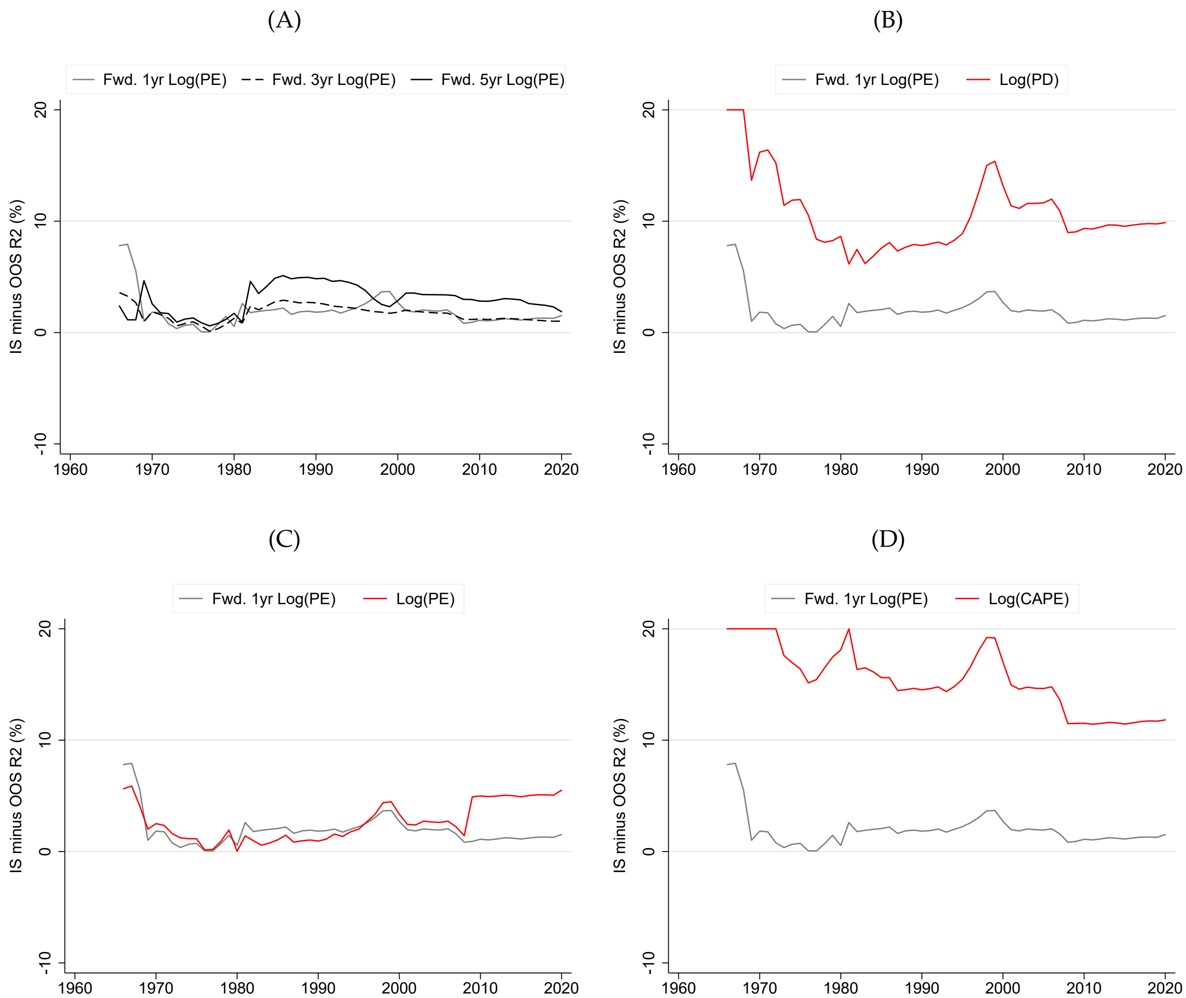
Figure 6: Out-of-sample R-squared



Note: This figure plots the out-of-sample R-squared when we predict one-year aggregate returns as in Goyal and Welch (2008). Panel (A) plots the out-of-sample r-squared performances of the (machine learning constructed) forward PE ratios for the 1yr, 3yr and 5yr horizons. Panels (B), (C) and (D) compares the out-of-sample r-squared performance of the forward 1yr PE ratio against the PD, PE and (shiller) CAPE ratios, respectively. We start calculating return forecasts in 1951 (i.e., 10 years post sample-start) and allow for fifteen years to pass before reporting out-of-sample r-squared results. Accordingly, the time period for out-of-sample r-squared statistics spans 1966 to 2020.

tion that, in the presence cash flow shocks that propagate for many years, longer-term forward PE ratios which better account for these cash flow shocks will outperform shorter-term forward PE ratios. In Panel (B), we compare the OOS performance of the forward 1yr PE ratio against the PD ratio where we see the performance of the PD ratio deteriorates rapidly around 1995 coinciding with the commencement of large firm repurchase programs per Figure 3 Panel (A) in line with the theory that not accounting for structural shifts in payouts magnifies the omitted variable problem. Similarly in Panel (C), we plot the OOS r-squared of the forward 1yr PE ratio against the PE ratio where we see the PE ratio's performance deteriorates rapidly in the financial crisis due to increased earnings volatility during this period again

Figure 7: In-sample vs out-of-Sample R-squared



Note: This figure plots the difference between the in-sample and out-of-sample r-squared overtime for each valuation ratio. Panel (A) plots this differential r-squared performance for the (machine learning constructed) forward PE ratios for the 1yr, 3yr and 5yr horizons. Panels (B), (C) and (D) compares the differential r-squared performance of the forward 1yr PE ratio against the PD, PE and (shiller) CAPE ratios, respectively. We start calculating return forecasts in 1951 (i.e., 10 years post sample-start) and allow for fifteen years to pass before reporting r-squared results. Accordingly, the time period for the r-squared statistics spans 1966 to 2020.

in line with the theoretical predictions. Finally, in Panel (D) we plot the forward 1yr PE OOS r-squared against the CAPE where we see the CAPE performs poorly throughout the entire sample because it's denominator averages over backward looking earnings which is not reflective of current normalized earnings because of structural shifts in earnings growth and earnings volatility across multi-decade periods per Table 1.

In-sample vs. out-of-sample return prediction. Per equation (16), our framework predicts that the difference between the in-sample ("IS") and OOS performance for a valuation ratio scaled by current cash flows will diverge in the presence of structural shifts in cash flow growth (e.g., caused by structural

Table 3: R-squared overview

	(1) Fwd 1yr PE (%)	(2) Fwd 2yr PE (%)	(3) Fwd 3yr PE (%)	(4) Fwd 5yr PE (%)	(5) PD (%)	(6) PE (%)	(7) CAPE (%)
IS r-squared	8.9	9.3	9.7	12.7	6.4	4.6	5.0
OOS r-squared	7.4	8.4	8.8	10.8	-3.5	-0.9	-6.7
IS minus OOS r-squared	1.5	0.9	1.0	1.9	9.9	5.5	11.8

Note: This table provides summary statistics of the in-sample, out-of-sample and in-sample minus out-of-sample r-squared performances for various valuation ratios over the sample period. Columns (1), (2) and (3) report the results for the (machine learning constructed) forward PE ratios over the 1yr, 3yr and 5yr horizons, respectively. Columns (1), (2) and (3) report the results for the price-to-dividend (PD), price-to-earnings (PE) and (shiller) 10yr cyclically adjusted price-to-earnings (CAPE) ratios, respectively. The sample period is from 1942 to 2020.

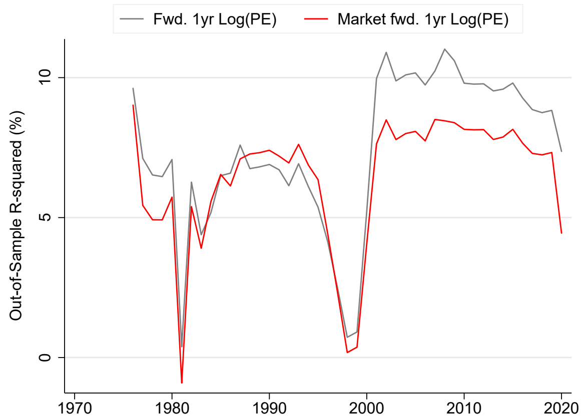
shifts in productivity or firm's payout methods) or changes in earnings volatility regimes. Moreover, to the extent constructed forward PE ratios adequately account for these changes the difference between the IS and OOS performance of the forward PE ratio should converge toward zero. To empirically test these predictions, we calculate the evolution of the IS and OOS performances of the forward PE ratios alongside the PD, PE and CAPE ratios. Results are plotted in Figure 7.

Per Panel (A), we find the delta between the IS and OOS performance of the forward 1yr, 3yr and 5yr PE ratios converge toward 1.5%, 1% and 1.9%, respectively. This indicates our forward PE ratios are adequately accounting for any structural changes embedded in current earnings and are close to optimal. Furthermore, we note these deltas consistently fluctuate close to zero over the entire sample. In Panels (B), (C) and (D) we plot the differential r-squared performance of the 1yr forward PE ratio against the PD, PE and CAPE ratios, respectively. We see for the PD, PE and CAPE ratio the IS r-squared performance significantly beats the corresponding OOS r-squared by 9.9%, 5.5% and 11.8% consistent with the intuition that these ratios undergo structural changes over the sample period for which the IS regression can better account for relative to its OOS counterpart.

Machine vs. market forecasts. Per Proposition 2, if the market's expectation of future cash flows deviates from rational expectation, the forward price ratio constructed using rational forecasts should outperform the market forward price ratio. That is, it has a higher r-squared when predicting returns. Intuitively, this is because the rationally constructed forward PE ratio will capture not only the market's expected returns but also forecastable cash flow news due to the market's biased expectations.

To proxy for the market's expectations of aggregate future earnings, we use investment professionals' expectations per the widely-used Thomson Reuters I/B/E/S data. Concretely, we construct a measure of the market's 1yr ahead earnings by aggregating equity analysts' 1yr ahead firm-level earnings forecasts at each forecast date (Hillenbrand and McCarthy, 2022, see also). We then scale the S&P500

Figure 8: Machine vs market: Out-of-sample R-squared



Note: This figure plots the evolution of the out-of-sample r-squared performance of the (machine learning constructed) 1yr forward PE ratio against a measure of the Market's 1yr forward PE ratio. We construct the latter measure by scaling the S&P500 index by equity analysts' 1yr ahead earnings per share forecasts at the S&P500 level. This Market 1yr forward PE ratio is only available from 1976 onwards and so for the pre-1976 period we set the Market forward PE ratio to be equal to the (machine learning constructed) 1yr forward PE measure. The time period is from 1976 to 2020.

index by this market measure of 1yr ahead earnings per share at the S&P500 level, yielding the "Market forward 1yr PE ratio".

A line of prior research indicates that market expectations are biased (De Silva and Thesmar, 2021; Bordalo, Gennaioli, LaPorta, and Shleifer, 2022; Bianchi, Ludvigson, and Ma, 2022), suggesting that the (machine learning constructed) forward 1yr PE ratio will outperform the Market forward 1yr PE ratio. To test this, Figure 8 plots the evolution of the OOS r-squared performance of the forward 1yr PE ratio against the Market forward 1yr PE ratio from 1976 onwards.[8] We find the forward 1yr PE ratio outperforms the Market forward 1yr PE by nearly 3% over the sample, and the out-performance is concentrated during both the dot-com bust and financial crisis, suggesting the bias in market expectations is pronounced during these periods.

6 Conclusion

This paper outlines both theoretically and empirically why commonly-used price ratios have struggled to predict return. In short, they do not adequately account for expected cash flow growth thereby creating an omitted variable problem, a problem which we show is amplified when there are structural shifts in firms' payout method, shifts in the level of expected earnings growth or shifts in the expected earnings volatility. We show how to circumvent the omitted variable problem by a constructing forward

[8]The Market 1yr forward PE ratio is only available from 1976 onwards (given data on equity analyst forecasts as collected I/B/E/S start then) and so for the pre-1976 period we set the Market forward PE ratio to be equal to the (machine learning constructed) 1yr forward PE measure.

price ratios which adequately accounts for expected cash flow growth, thereby by yielding a price ratio which has the highest r-squared in forecasting long-run returns and is therefore optimal. When we construct empirical proxies for the optimal price-ratio we find an in-sample and out-of-sample r-squared for predicting one-year aggregate returns of 12.7% and 10.8%, respectively, beating all other predictors and resolving the return-predictability debate (Goyal and Welch, 2008). We conclude by asserting that our work has large relevance for economic agents, who use real-time expected return measures to make economic decisions, impacting firms' capital budgeting decisions, portfolio theory, investment professionals' capital-allocation decisions, macroeconomics and so forth.

References

Bianchi, Francesco, Sydney C. Ludvigson, and Sai Ma, 2022. Belief distortions and macroeconomic fluctuations. American Economic Review 112, 2269–2315.

Bordalo, Pedro, Nicola Gennaioli, Rafael LaPorta, and Andrei Shleifer, 2022. Belief overreaction and stock market puzzles. .

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R Roberts, 2007. On the importance of measuring payout yield: Implications for empirical asset pricing. The Journal of Finance 62, 877–915.

Campbell, John Y, 1991. A variance decomposition for stock returns. The economic journal 101, 157–179.

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

Campbell, John Y, and Samuel B Thompson, 2008. Predicting excess stock returns out of sample: Can anything beat the historical average?. The Review of Financial Studies 21, 1509–1531.

Chen, Long, 2009. On the reversal of return and dividend growth predictability: A tale of two periods. Journal of Financial Economics 92, 128–151.

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

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

Dangl, Thomas, and Michael Halling, 2012. Predictive regressions with time-varying coefficients. Journal of Financial Economics 106, 157–181.

De Silva, Tim, and David Thesmar, 2021. Noise in expectations: Evidence from analyst forecasts. .

Dybvig, Philip H, and Huacheng Zhang, 2018. That is not my dog: Why doesn't the log dividend-price ratio seem to predict future log returns or log dividend growths?. .

Ferreira, Miguel A, and Pedro Santa-Clara, 2011. Forecasting stock market returns: The sum of the parts is more than the whole. Journal of Financial Economics 100, 514–537.

Gilchrist, Simon, and Egon Zakrajšek, 2012. Credit spreads and business cycle fluctuations. American economic review 102, 1692–1720.

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

——— , and Athanasse Zafirov, 2021. A comprehensive look at the empirical performance ofequity premium prediction ii. Available at SSRN 3929119.

Graham, John R, Hyunseob Kim, and Mark Leary, 2020. Ceo-board dynamics. Journal of Financial Economics 137, 612–636.

Graham, John R, and Mark T Leary, 2018. The evolution of corporate cash. The Review of Financial Studies 31, 4288–4344.

——— , and Michael R Roberts, 2015. A century of capital structure: The leveraging of corporate america. Journal of financial economics 118, 658–683.

Greenwald, Daniel L, Martin Lettau, and Sydney C Ludvigson, 2016. Origins of stock market fluctuations. Discussion paper, National Bureau of Economic Research.

Hillenbrand, Sebastian, and Odhrain McCarthy, 2022. Heterogeneous beliefs and stock market fluctuations. Working paper.

Lettau, Martin, and Sydney Ludvigson, 2001a. Consumption, aggregate wealth, and expected stock returns. the Journal of Finance 56, 815–849.

——— , 2001b. Resurrecting the (c) capm: A cross-sectional test when risk premia are time-varying. Journal of political economy 109, 1238–1287.

Lettau, Martin, and Sydney C Ludvigson, 2005. Expected returns and expected dividend growth. Journal of Financial Economics 76, 583–626.

Nagel, Stefan, 2021. Machine learning in asset pricing. in Machine Learning in Asset Pricing Princeton University Press.

Pástor, L'uboš, and Robert F Stambaugh, 2009. Predictive systems: Living with imperfect predictors. The Journal of Finance 64, 1583–1628.

Rapach, David E, Jack K Strauss, and Guofu Zhou, 2010. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. The Review of Financial Studies 23, 821–862.

Van Binsbergen, Jules H, Xiao Han, and Alejandro Lopez-Lira, 2020. Man vs. machine learning: The term structure of earnings expectations and conditional biases. .

Van Binsbergen, Jules H, and Ralph SJ Koijen, 2010. Predictive regressions: A present-value approach. The Journal of Finance 65, 1439–1471.

Appendix for "The Optimal Stock Valuation Ratio"

A Derivations for Section 2

R-squared and correlation coefficient. The R-squared from a regression of y on x is equal to the correlation coefficient between y and x:

$$
\begin{aligned}
R^2 &= \frac{\text{Explained sum of squares}}{\text{Total sum of squares}} = \frac{\text{Var}(\hat{y})}{\text{Var}(y)} = \frac{\text{Var}\left(\beta_0 + \beta_1 x\right)}{\text{Var}(y)} = \beta_1^2 \frac{\text{Var}(x)}{\text{Var}(y)} \\
&= \left(\frac{\text{cov}(x,y)}{\text{Var}(x)}\right)^2 \cdot \frac{\text{Var}(x)}{\text{Var}(y)} = \frac{\text{cov}(x,y)^2}{\text{Var}(x)^2} \cdot \frac{\text{Var}(x)}{\text{Var}(y)} \\
&= \frac{\text{cov}(x,y)^2}{\text{Vav}(x) \cdot \text{Var}(y)} = \left(\frac{\text{cov}(x,y)}{\sqrt{\text{Vav}(x) \cdot \text{Var}(y)}}\right)^2 = \rho_{xy}^2
\end{aligned}
\tag{A.1}
$$

Derivation of regression coefficient and R-squared for pd_t. Then the regression coefficient is given by

$$
\begin{aligned}
b_{\sum r} &= \frac{Cov(-\sum \mathbb{E}_t^M r, pd_t)}{\sigma^2\left(pd_t\right)} \\
&= \frac{\sigma^2(\sum \mathbb{E}_t^M r) - Cov(\sum \mathbb{E}_t^M r, \sum \mathbb{E}_t^M \Delta d)}{\sigma^2(\sum \mathbb{E}_t^M r) + \sigma^2(\sum \mathbb{E}_t^M \Delta d) - 2Cov(\sum \mathbb{E}_t^M r, \sum \mathbb{E}_t^M \Delta d)}
\end{aligned}
\tag{A.2}
$$

The R-squared is:

$$
\begin{aligned}
R_{\sum r}^2 &= Corr^2(-\sum r, pd_t) \\
&= \frac{Cov^2(-\sum \mathbb{E}_t^M r, pd_t)}{\sigma^2\left(\sum r\right)\sigma^2\left(pd_t\right)} \\
&= \underbrace{\frac{Cov(-\sum \mathbb{E}_t^M r, pd_t)}{\sigma^2\left(pd_t\right)}}_{b_{\sum r}} \frac{Cov(-\sum \mathbb{E}_t^M r, pd_t)}{\sigma^2\left(\sum \mathbb{E}_t^M r + \sum \eta_{CF}^M - \sum \eta_{DR}^M\right)} \\
&= \frac{Cov^2(-\sum \mathbb{E}_t^M r, -\sum \mathbb{E}_t^M r + \sum \mathbb{E}_t^M \Delta d)}{\sigma^2\left(\sum r\right)\sigma^2\left(pd_t\right)} \\
&= \frac{\left[\sigma^2(\sum \mathbb{E}_t^M r) - Cov(\sum \mathbb{E}_t^M r, \sum \mathbb{E}_t^M \Delta d)\right]^2}{\sigma^2\left(\sum r\right)\left[\sigma^2(\sum \mathbb{E}_t^M r) + \sigma^2(\sum \mathbb{E}_t^M \Delta d) - 2Cov(\sum \mathbb{E}_t^M r, \sum \mathbb{E}_t^M \Delta d)\right]} \\
&= \frac{\sigma^2(\sum \mathbb{E}_t^M r)}{\sigma^2\left(\sum r\right)} \frac{\left[1 - \rho_{d,r}\frac{\sigma(\sum \mathbb{E}_t^M \Delta d)}{\sigma(\sum \mathbb{E}_t^M r)}\right]^2}{\left[1 + \frac{\sigma^2(\sum \mathbb{E}_t^M \Delta d)}{\sigma^2(\sum \mathbb{E}_t^M r)} - 2\rho_{d,r}\frac{\sigma(\sum \mathbb{E}_t^M \Delta d)}{\sigma(\sum \mathbb{E}_t^M r)}\right]} \\
&= \frac{\sigma^2(\sum \mathbb{E}_t^M r)}{\sigma^2\left(\sum r\right)} \frac{\left[1 + \rho_{d,r}^2\frac{\sigma^2(\sum \mathbb{E}_t^M \Delta d)}{\sigma^2(\sum \mathbb{E}_t^M r)} - 2\rho_{d,r}\frac{\sigma(\sum \mathbb{E}_t^M \Delta d)}{\sigma(\sum \mathbb{E}_t^M r)}\right]}{\left[1 + \frac{\sigma^2(\sum \mathbb{E}_t^M \Delta d)}{\sigma^2(\sum \mathbb{E}_t^M r)} - 2\rho_{d,r}\frac{\sigma(\sum \mathbb{E}_t^M \Delta d)}{\sigma(\sum \mathbb{E}_t^M r)}\right]}
\end{aligned}
\tag{A.3}
$$

B Derivation and Proofs for Section 3

Proof for proposition 1

Proof. Suppose there is another predictor that predicts returns according to

$$\tilde{x}_t = \sum \mathbb{E}_t r_{t+1} + \epsilon_t$$

Note that we can write this without loss of generality (run a regression of ϵ_t on $\mathbb{E}_t r_{t+1}$) as

$$\tilde{x}_t = \sum \mathbb{E}_t r_{t+1} + \nu_t + \widetilde{\epsilon}_t$$

where ν_t is perfectly correlated with $\mathbb{E}_t r_{t+1}$ and $\widetilde{\epsilon}_t$ is orthogonal to $\mathbb{E}_t r_{t+1}$. Note that the existence of $\widetilde{\epsilon}_t$ immediately implies

$$|Corr(\mathbb{E}_t r_{t+1}, \tilde{x}_t)| < 1$$

Let's start with the case where $\nu_t = 0$, then

$$\begin{aligned} R^2_{\tilde{x}} &= \frac{\sigma^4(\sum \mathbb{E}_t r_{t+1})}{\sigma^2\left(\sum r_{t+1}\right)\sigma^2\left(\tilde{x}_t\right)} = \frac{\sigma^2(\sum \mathbb{E}_t r_{t+1})}{\sigma^2\left(\sum r_{t+1}\right)} \frac{\sigma^2(\sum \mathbb{E}_t r_{t+1})}{\sigma^2\left(\sum \mathbb{E}_t r_{t+1}\right) + \sigma^2\left(\widetilde{\epsilon}_t\right)} = \\ &= \frac{\sigma^2(\sum \mathbb{E}_t r_{t+1})}{\sigma^2\left(\sum r_{t+1}\right)} \frac{1}{1 + \frac{\sigma^2(\widetilde{\epsilon}_t)}{\sigma^2(\sum \mathbb{E}_t r_{t+1})}} < \frac{\sigma^2(\sum \mathbb{E}_t r_{t+1})}{\sigma^2\left(\sum r_{t+1}\right)} \end{aligned} \tag{A.4}$$

Now, let's go to the case where $\nu_t \neq 0$, then

$$R^2_{\tilde{x}} = Corr^2(\mathbb{E}_t r_{t+1}, \tilde{x}_t) \frac{\sigma^2(\sum \mathbb{E}_t r_{t+1})\sigma^2(x_t)}{\sigma^2(\sum r_{t+1})\sigma^2(x_t)} = Corr^2(\mathbb{E}_t r_{t+1}, x_t) \frac{\sigma^2(\sum \mathbb{E}_t r_{t+1})}{\sigma^2(\sum r_{t+1})} < \frac{\sigma^2(\sum \mathbb{E}_t r_{t+1})}{\sigma^2(\sum r_{t+1})} \tag{A.5}$$

where the last line follows from $|Corr(\mathbb{E}_t r_{t+1}, \tilde{x}_t)| < 1$. □

Derivation of the regression coefficient when predicting one-year returns.

$$\begin{aligned} R^2_\infty &= \mathrm{Corr}\left(r_{t+1}, pd^\infty_t\right)^2 = \frac{\mathrm{Cov}\left(R_1, -R_1 - R_2\right)^2}{\sigma^2\left(r_{t+1}\right) \cdot \sigma^2\left(R_1 + R_2\right)} = \frac{\mathrm{Cov}\left(R_1, R_1 + R_2\right)^2}{\sigma^2\left(r_{t+1}\right)\sigma^2\left(R_1 + R_2\right)} \\ &= \frac{\sigma^4\left(R_1\right) + 2 \cdot \sigma^2\left(R_1\right) \cdot \mathrm{Cov}\left(R_1, R_2\right) + \mathrm{Cov}\left(R_{11} R_2\right)^2}{\sigma^2\left(r_{t+1}\right) \cdot \sigma^2\left(R_1 + R_2\right)} = \\ &= \frac{\sigma^4\left(R_1\right) + 2 \cdot \sigma^2\left(R_1\right)\rho_{R_1,R_2} \cdot \sigma\left(R_1\right)\sigma\left(R_2\right) + \rho^2_{R_1,R_2}\sigma^2\left(R_1\right)\sigma^2\left(R_2\right)}{\sigma^2\left(r_{t+1}\right) \cdot \left[\sigma^2\left(R_1\right) + \sigma^2\left(R_2\right) + 2 \cdot \rho_{R_1,R_2} \cdot \sigma\left(R_1\right)\sigma\left(R_2\right)\right]} \\ &= \frac{\sigma^2\left(R_1\right)}{\sigma^2\left(r_{t+1}\right)} \cdot \frac{1}{\frac{\sigma^2(R_1) + \sigma^2(R_2) + 2\rho_{R_1 R_2} \cdot \sigma(R_1)\sigma(R_2)}{\sigma^2(R_1) + \rho^2_{R_1,R_2}\sigma^2(R_2) + 2p_{R_1 R_2} \cdot \sigma(R_1) \cdot \sigma(R_2)}} \\ &= \frac{\sigma^2\left(R_1\right)}{\sigma^2\left(r_{t+1}\right)} \cdot \frac{1}{\frac{1 + 2 \cdot \rho_{R_1 R_2} \frac{\sigma(R_2)}{\sigma(R_1)} + \frac{\sigma^2(R_2)}{\sigma^2(R_1)}}{1 + 2 \cdot \rho_{R_1 R_2} \cdot \frac{\sigma(R_2)}{\sigma(R_1)} + \rho^2_{R_1 R_2} \cdot \frac{\sigma^2(R_2)}{\sigma^2(R_1)}}} \end{aligned} \tag{A.6}$$

Derivation of R-squared for one-year forward price ratio and trend growth

$$
\begin{aligned}
R^2 &= \operatorname{Corr}\left(-\sum r, pd_t^1\right)^2 = \frac{\operatorname{Cov}\left(-\Sigma r, b_1 g_t - \Sigma \mathbb{E} r\right)^2}{\sigma^2(\Sigma r)\cdot\sigma^2\left(b_1 g_t - \Sigma \mathbb{E}_r\right)} = \\
&= \frac{\sigma^4\left(\sum \mathbb{E} r\right) - 2\cdot\operatorname{Cov}\left(\sum \mathbb{E} r, b_1 g_t\right)\cdot\sigma^2\left(\sum \mathbb{E}\right) + \operatorname{Cov}\left(\sum \mathbb{E} r, b_1 g_t\right)^2}{\sigma^2\left(\sum r\right)\cdot\left[\sigma^2\left(\sum \mathbb{E} r\right) - 2\operatorname{Cov}\left(b_1 g_t, \Sigma \mathbb{E} r\right) + \sigma^2\left(b_1 g_t\right)\right]} \\
&= \frac{\sigma^2(\sum \mathbb{E} r)}{\sigma^2(\sum r)}\cdot\frac{1}{\frac{1-2b_1\cdot\rho_{g,\mathbb{E}R}\frac{\sigma(g_t)}{\sigma(\Sigma\mathbb{E})}+b_1^2\frac{\sigma^2(g_t)}{\sigma^2(\Sigma\mathbb{E}_r)}}{1-2b_1\cdot\rho_{g,\mathbb{E}R}\frac{\sigma(g_t)}{\sigma(\Sigma\mathbb{E}r)}+b_1^2\rho_{g,\mathbb{E}r}^2\frac{\sigma^2(g_t)}{\sigma(\Sigma\mathbb{E}r)}}}
\end{aligned}
\tag{A.7}
$$

C Data: Variables Used

Variables (Post-1950)	Compustat Label	Pre-1950	Required Pre-1950	Required Post-1950
Assets	at	✓	✓	✓
Liabilities	lt	✓	✓	✓
Revenue	revt	✓	✓	✓
SG&A	xsga			
R&D	xrd			
Cost-of-goods dold	cogs			✓
Current assets	act	✓		
Current Liabilities	lct	✓		
Cash	ch	✓		
Cash and short-term investemtns	che	✓		
Income tax	txt	✓		
Total long-term debt	dltt	✓		
Total long-term debt due in 1yr	dd1			
Debt in current liabilites	dlc	✓		
D&A	dp			
Ebit	ebit	✓		
Ebitda	ebitda			
Interest & related expense total	xint	✓		
Interest paid-net	intpn			
Capex	capx			
Income taxes payable	txp			
Income taxes paid	txpd			
Income taxes total	txt	✓		
Net income	ni	✓		
Cash dividends on common stock	cdvc	✓		
Purchase of common and preferred stock	prstkc			
Sale of common and preferred stock	sstk			
Subordinated debt	ds			
Gross profit	gp			✓
Operating activities - net cash flow	oancf			
Common shares for diluted eps	cshfd	✓		
Price close (fiscal year-end)	prcc_f	✓	✓	✓
Extraordinary items	xi	✓		
Special items	spi			
Acquisitions	aqc			
Capitalized leases (2yr-5yr)	cld2 – cld5			
Common esop obligation	esopct			
Goodwill (net0	gdwl			
Interest and related income (total)	idit			
Total intangible assets	intan			
Marketable security adjustments	msa			
Property, plant and equipment (net)	ppent	✓	✓	✓
Non-operating income	nopi			
Tax loss carryforward	tlcf			
Pension and retirement expense	xpr			
Preferred stock liquidating value	pstkl	✓		
Extraordinary and discontinued items	xido			
Non-recurring income taxes (after-tax)	nrtxt			

D Data Output

Table A.1: Summary Statistics

	mean	sd	min	max	count
Earnings growth					
Realized 1yr growth	0.13	0.18	−0.17	1.14	75
ML predicted 1yr growth	0.12	0.08	−0.12	0.32	75
Realized 2yr growth	0.12	0.14	−0.10	0.92	75
ML predicted 2yr growth	0.11	0.06	−0.04	0.26	75
Realized 3yr growth	0.12	0.12	−0.07	0.66	75
ML predicted 3yr growth	0.15	0.08	−0.05	0.34	75
Realized 5yr growth	0.13	0.10	−0.02	0.60	75
ML predicted 5yr growth	0.14	0.07	−0.05	0.33	75
Valuation Ratios					
ML fwd. 1yr price-to-earnings	2.56	0.38	1.64	3.41	75
ML fwd. 2yr price-to-earnings	2.47	0.39	1.61	3.35	75
ML fwd. 3yr price-to-earnings	2.32	0.40	1.32	3.16	75
ML fwd. 5yr price-to-earnings	2.15	0.40	1.34	3.03	75
price-to-dividend	3.45	0.45	2.50	4.46	75
price-to-earnings	2.74	0.47	1.80	4.78	75
cape	2.83	0.41	1.90	3.76	75

Note: This table provides summary statistics of the variables used in the paper. All variables are on the aggregate stock market, i.e., S&P500 index, level. The unit of observation is year. The time period is from 1942 to 2016.