THE OHIO STATE UNIVERSITY
FISHER COLLEGE OF BUSINESS

Fisher College of Business
Working Paper Series

Charles A. Dice Center for Research in Financial Economics

Rediscover Predictability: Information from the Relative Prices of Long-term and Short-term Dividends

Ye Li,
The Ohio State University

Cheng Wang,
Yale School of Management

Dice Center WP 2018-16
Fisher College of Business WP 2018-03-016

September 2018

Rediscover Predictability: Information from the Relative Prices of Long-term and Short-term Dividends*

Ye Li† Chen Wang‡

September 30, 2018

Abstract

The ratio of long- to short-term dividend prices, "price ratio" (pr_t), predicts one-year stock market return with an out-of-sample R^2 of 19%. It subsumes the predictive power of price-to-dividend ratio (pd_t). The residual from regressing pd_t on pr_t predicts one-year dividend with an out-of-sample R^2 of 30%. Our results hold outside the U.S. In an exponential-affine model, we show the key to understand these findings is the (lack of) persistence of expected dividend growth. We also characterize the risk of time-varying expected return: (1) the expected return is countercyclical; (2) the response of expected return (rather than expected dividend growth) accounts for the impact of monetary policy on stock price; (3) shocks to pr_t are priced in the cross-section, which serves as an ICAPM test of pr_t as an adequate proxy for the expected return.

*We thank Nick Barberis, Geert Bekaert, Jules H. van Binsbergen, Zhi Da, Kent Daniel, Stefano Giglio, Will Goetzmann, Bryan T. Kelly, Ralph S.J. Koijen, Lars A. Lochstoer, Alan Moreira, Tano Santos, José A. Scheinkman, and Michael Weber for helpful comments. Chen Wang acknowledges financial support from International Center for Finance at Yale School of Management. All errors are ours.

†The Ohio State University. E-mail: li.8935@osu.edu

‡Yale School of Management. E-mail: ch.wang@yale.edu

1 Introduction

Return and cash-flow predictability are at the core of asset pricing studies. Remarkably, a simple variable, the ratio of long- to short-term dividend prices ("price ratio"), brings new insights on both fronts. It predicts one-year stock market return with an out-of-sample R^2 of 19%. The residual from regressing the traditional price-to-dividend ratio on the price ratio predicts one-year dividend with an out-of-sample R^2 of 30%. While enormous efforts have been made to study dividend strips (Binsbergen and Koijen (2017)), we are the first to show that their price ratio contains critical information on the *aggregate* market. Using the exponential affine framework of Lettau and Wachter (2007), we find the key to understanding our findings is the (lack of) persistence in the expected dividend growth.

We further study the variation of expected return. In ICAPM (Merton (1973)), shocks to investment opportunities (e.g., the expected market return) are priced; so are shocks to adequate return predictors as proxy for the expected return. Indeed, shocks to the price ratio are priced in the cross section of stocks. To the best of our knowledge, we are the first to conduct this economic test of return predictor. The estimated price of risk also implies a coefficient of relative risk aversion greater than one (Campbell (1993)). We demonstrate the risk of time-varying expected return by showing its response to monetary policy and its correlations with macroeconomy, financial intermediation capacity, uncertainty, and sentiment.

The return predictive power of the price ratio is also strong outside the U.S., and the variations of expected stock return across countries are almost perfectly synchronized in our sample period from 1988 to 2017.[1] Finally, we examine conditional predictability and find returns are more predictable during market downturns in both the U.S. and other countries.

We start by decomposing the total market valuation into the prices of long- and short-term dividends, so the traditional price-dividend ratio is the sum of two components,

$$\frac{P_t}{D_t} = \frac{\textit{Price of Long-term Dividends}}{D_t} + \frac{\textit{Price of Short-term Dividends}}{D_t}.$$

As in Binsbergen and Koijen (2010) and Kelly and Pruitt (2013), a state-space model shows that the two components contain distinct information on future returns and dividends, so to extract more information beyond the log price-dividend ratio ($pd_t = \ln(P_t/D_t)$), we calculate the log difference of this pair (i.e., the price ratio or pr_t),

$$pr_t = \ln\left(\frac{\textit{Price of Long-term Dividends}}{\textit{Price of Short-term Dividends}}\right).$$

[1] Binsbergen, Brandt, and Koijen (2012) show that futures or option data can be used to calculate dividend strip prices. We use futures because they have a longer sample period starting in 1988.

pr_t is the slope of term structure of dividend trip prices, while pd_t captures the level. pr_t is also a measure of duration. We set "short-term" to be one year. With the value of one-year dividends in the numeraire, pr_t measures how many years of value are there beyond one year.

pr_t strongly predicts market return. A decrease of pr_t by one standard deviation adds 7.3% to the expected return over the next year. Annual forecasting delivers an out-of-sample R^2 equal to 19.2%, which is three times the out-of-sample R^2 of pd_t in our sample. Improvements mainly come from the variation of pr_t at higher frequencies, since pr_t is constructed from market prices that are more responsive to news than accounting variables (e.g., past dividends). This variability in return expectations is difficult to reconcile with state-of-the-art asset pricing models (e.g., Campbell and Cochrane (1999) and Bansal and Yaron (2004)). The market timing strategy using pr_t as signal delivers a Sharpe ratio of 0.84.

We establish the robustness of our prediction results in a number of ways. First, following Hodrick (1992), we adjust our standard error by taking into account the overlapping structure of annual returns. Second, we show that the autocorrelation of pr_t is 91.5%, lower than that of pd_t (98.7%), and that our estimate of predictive coefficient is robust to Stambaugh (1999) bias.[2] Third, we conduct several out-of-sample tests (e.g., Clark and McCracken (2001)). Finally, we also show that in terms of in-sample R^2, out-of-sample R^2, and Hodrick (1992) t-statistic, pr_t outperforms the existing predictors.[3]

The price ratio also predicts return outside the United States. We run a panel predictive regression with pr_t of each country as the predictor. The coefficient is significant and close in magnitude to the coefficient in the U.S. Interestingly, once we add time fixed effect to absorb the global factor in the realized returns of each country, return predictability disappears. This suggests that the variations of expected stock return are synchronized across countries, offering new evidence on global market integration (Karolyi and Stulz (2003)).

To understand our findings, we impose more structure on the state space model following Lettau and Wachter (2007). The prices of dividends at different horizons are exponential-affine functions of the expected return and expected dividend growth, with the latter's coefficient depending on the specific horizon through the persistence of expected dividend growth. If and only if the persistence is zero, the prices of all dividends have the same coefficient on the expected dividend growth, and the log difference between long- and short-term dividend prices ("price ratio") becomes a function of the expected return only – a perfect return predictor. Then pr_t outperforms pd_t in return prediction precisely because pd_t mixes the

[2] Ferson, Sarkissian, and Simin (2003) show a spurious regression bias when both the proposed predictor and the underlying expected return are persistent. pr_t also passes a recent test proposed by Kostakis, Magdalinos, and Stamatogiannis (2015) that explicitly addresses the issue of predictor persistence.

[3] Previously studied predictors typically perform well in-sample but become insignificant out-of-sample, often performing worse than forecasts based on the historical mean (Goyal and Welch (2007)).

information on both the expected return and expected dividend growth (Menzly, Santos, and Veronesi (2004); Lettau and Ludvigson (2005)).

Can we use pr_t to tease out the information on expected dividend growth in pd_t? We regress pd_t on pr_t to remove the information on the expected return and use the residuals to forecast dividend growth. The out-of-sample R^2 is 30%. In contrast, pd_t itself does not predict dividends, as already well documented (Chen (2009); Cochrane (2011); Chen, Da, and Priestley (2012)). Using the pd_t residual as a proxy for the expected dividend growth, we find the autocorrelations are close to zero across countries. The persistence of expected dividend growth is also weak in a state space model fitted to the realized dividends.

To form a better return predictor, recent studies focus on adjusting pd_t to eliminate the variation of expected dividend growth (Campbell and Thompson (2008); Lacerda and Santa-Clara (2010); Da, Jagannathan, and Shen (2014); Golez (2014)). pr_t is simply an alternative combination of pd_t's components, and does not rely on an adjustment model. Moreover, the pr_t-adjusted pd_t (residual) is a stronger dividend predictor, adding to the literature on cash flow predictability (Larrain and Yogo (2008); Binsbergen, Hueskes, Koijen, and Vrugt (2013); Chen, Da, and Zhao (2013)). Many have shown that cash flow expectation is important for understanding key asset pricing patterns (Bansal and Yaron (2004); Beeler and Campbell (2012); Belo, Collin-dufresne, and Goldstein (2015); Collin-Dufresne, Johannes, and Lochstoer (2016)). We further this line of research by pinpointing the return predictive power of pr_t to the persistence of expected dividend growth.

After establishing the evidence on return predictability, we proceed to understand the risk of time-varying expected return. In particular, we examine whether shocks to pr_t are priced in the cross-section of stocks. By the logic of ICAPM, shocks to the expected return should be priced; so should the shocks to pr_t if it is an adequate proxy for the expected return. Therefore, this cross-sectional asset pricing exercise is first and foremost an economic test of pr_t as a return predictor. To the best of our knowledge, we are the first to conduct this test in the literature of return predictability.

We find a significant and negative price of pr_t risk in the cross-section. Consider two assets with one standard-deviation difference in their pr_t beta. The average return of the high-beta asset is 2.1% lower, as it delivers higher returns when pr_t is high and the expected market return is low.[4] This negative (positive) price of pr_t (expected return) risk implies a coefficient of relative risk aversion greater than one (Campbell (1993)) – hedging against deterioration in investment opportunities is more desirable than having more wealth to profit

[4]Note that here we focus on the first moment of market return. It is likely that the investment opportunity improves because the volatility, or covariance with the marginal investors' marginal value of wealth, declines even more than the expected return. However, empirically, the evidence on the correlation between the expected return and expected volatility is mixed (Guo and Whitelaw (2006); Lettau and Ludvigson (2010)).

from improved investment opportunities.

To further characterize the risk of time-varying expected return, we examine how the expected return (proxied by pr_t) responds to monetary policy shocks. The impact of monetary policy on asset prices continues attracting enormous attention (Lucca and Moench (2015); Campbell, Pflueger, and Viceira (2015); Drechsler, Savov, and Schnabl (2017)). Specifically, we regress pr_t on the unanticipated changes in the Federal Funds rate (Cochrane and Piazzesi (2002)), and find a negative coefficient, suggesting that the expected return declines during monetary expansions. In contrast, pd_t, the common proxy for the expected return (e.g., Muir (2017)), does not respond to monetary policy shocks. We also find that monetary easing is associated with a higher contemporaneous realized return, in line with Thorbecke (1997) and Bernanke and Kuttner (2005). Finally, we use the pr_t-adjusted pd_t (residual) to proxy the expected dividend growth, and find it does not respond to monetary policy shocks. In sum, stock price rises in response to expansionary monetary policy, but since the expected return declines, such increase tends to revert over the next year.

Next, we show that the expected stock return is countercyclical. The expected return is positively correlated with unemployment and negatively correlated with consumption growth, fixed investment, and inflation. The expected return shows a very strong negative correlation with broker-dealer leverage (Adrian and Shin (2010)) and intermediaries' net worth (inversely proxied by the broker-dealer CDS spreads). Interestingly, the expected return declines when VIX rises, which has important implications on the dynamics of risk-return trade-off (Lettau and Ludvigson (2010); Moreira and Muir (2017)).The expected return tends to be low when the sentiment (Baker and Wurgler (2006)) is high, even after the sentiment index is orthogonalized to other macro variables.

Last but not least, we study conditional return predictability. We find that the return predictive power of pr_t is asymmetric – it is much stronger following a negative market excess returns in the past twelve months.[5] Such asymmetry holds outside the United States. Related, Rapach, Strauss, and Zhou (2010), Henkel, Martin, and Nardari (2011), Dangl and Halling (2012), and Cujean and Hasler (2017) find that return predictors, such as the price-dividend ratio, work better in economic downturns. We evaluate two asset pricing models, Barberis, Huang, and Santos (2001) and He and Krishnamurthy (2013), that imply strong asymmetry in return predictability. While the return predictive power of pr_t does depend

[5] Our results of conditional return prediction suggest that the expected return is a function of both pr_t and past returns. Thus, our findings are related to the long-standing literature on return autocorrelation (Fama and French (1988); Poterba and Summers (1988); Moskowitz et al. (2012)). When pr_t is at its mean, the return does not show autocorrelation. However, when pr_t is one standard deviation (or more) above the mean, return exhibits momentum at one-year horizon, and when pr_t is one standard deviation (or more) below the mean, return shows reversal. Related, Huang, Jiang, Tu, and Zhou (2017) find that whether the stock market exhibits time-series momentum or reversal depends on the state of the economy.

on the state variables proposed by both models, these state variables alone do not predict future returns, suggesting that these theories still miss important drivers of expected return.

The remainder of the paper is organized as follows. Section 2 presents the data, empirical results, and our structural model of return and dividend predictability. Section 3 focuses on the risk of time-varying expected return, starting with the ICAPM test. Section 4 provides evidence on conditional return predictability. Section 5 concludes. Derivation details and additional results are provided in the appendices.

2 Return Prediction

Roadmap. We motivate pr_t, the ratio of long- to short-term dividend prices, as a return predictor in a state space model, and document its superior predictive power in comparison with pd_t, the price-to-dividend ratio (Table 2), and other predictors (Figure 3). Moreover, the residuals from regressing pd_t on pr_t strongly predict dividends (Table 3). By imposing more structure on the state space model following Lettau and Wachter (2007), we find the key to understand our results is the (lack of) persistence in expected dividend growth. Finally, Table 4 and Figure (4) show the return predictive power of pr_t across countries.

2.1 Decomposing the price-dividend ratio

A motivating model. We consider a state space model of return and cash flow (Cochrane (2008)). Let μ_t denote the expected return from time t to $t+1$, and g_t the expected dividend growth. We assume that the information set at time t is summarized by factors F_t, and the expected return and dividend growth are given by the following linear system[6]

$$\begin{aligned} \mu_t &= \gamma_0 + \boldsymbol{\gamma}' F_t, \\ g_t &= \delta_0 + \boldsymbol{\delta}' F_t. \end{aligned} \tag{1}$$

Following Binsbergen and Koijen (2010) and Kelly and Pruitt (2013), we impose a VAR(1) structure on the factors

$$F_{t+1} = \boldsymbol{\Lambda} F_t + \xi_{t+1}, \tag{2}$$

where $\boldsymbol{\Lambda}$ is a constant matrix with conformable dimensions. Let pd_t denote the log price-dividend ratio of the market at time t, Δd_{t+j} the one-period dividend growth from $t+j-1$

[6] A non-linear model is more general, but this model is only used for motivation, not estimation.

to $t+j$, and r_{t+j} the market return from $t+j-1$ to $t+j$. We can use the present value identity of Campbell and Shiller (1988), i.e.,

$$pd_t = \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1} \mathbb{E}_t \left[\Delta d_{t+j} - r_{t+j} \right], \tag{3}$$

to solve the price-dividend ratio as a function of F_t:

$$pd_t = \phi_0 + \boldsymbol{\phi}' F_t, \tag{4}$$

where ϕ_0 is equal to $\frac{\kappa+\delta_0-\gamma_0}{1-\rho}$, and $\boldsymbol{\phi}'$ is equal to $\boldsymbol{\iota}\boldsymbol{\psi}'\left(1-\rho\boldsymbol{\Lambda}\right)^{-1}$ with $\boldsymbol{\iota}$ being a row vector $(1,-1)$ and $\boldsymbol{\psi}$ equal to $(\boldsymbol{\delta}', \boldsymbol{\gamma}')$. Derivation details are in Appendix I.

By linking the price-dividend ratio to future returns and dividend growth, the present value identity serves as a motivation to use pd_t as a predictor. The factor structure reveals that any predictive power of pd_t comes from a particular linear combination of F_t, a compression of information. Next, we decompose the price-dividend ratio into different components with distinct information content from F_t.

The price ratio. Let S_t denote ex-dividend market value, D_t, the dividend at t, and r_t, the short rate. Under the no-arbitrage condition, there exists a risk-neutral measure, Q, such that the stock price is the expected sum of discounted future dividends:

$$S_t = \sum_{\tau=1}^{\infty} \mathbb{E}_t^Q \left[e^{-\int_t^{t+\tau} r_s ds} D_{t+\tau} \right] = \underbrace{\sum_{\tau=1}^{T} \mathbb{E}_t^Q \left[e^{-\int_t^{t+\tau} r_s ds} D_{t+\tau} \right]}_{P_t^{T-}} + \underbrace{\sum_{\tau=T+1}^{\infty} \mathbb{E}_t^Q \left[e^{-\int_t^{t+\tau} r_s ds} D_{t+\tau} \right]}_{P_t^{T+}},$$

where P_t^{T-} is the price of dividends paid from $t+1$ to $t+T$, i.e., the price of short-term dividends, and P_t^{T+} is the price of long-term dividends. Dividing both sides by D_t, we obtain a decomposition of price-dividend ratio into two valuation ratios, i.e., the ratio of short-term dividend price to D_t, and the ratio of long-term dividend price to D_t:

$$\frac{S_t}{D_t} = \frac{P^{T-}}{D_t} + \frac{P^{T+}}{D_t}. \tag{5}$$

Since the price-dividend ratio is the sum of these two ratios, we construct our predictor by taking the (log) difference so that it contains different information from the pair:

$$pr_t = \ln\left(\frac{P^{T+}}{D_t}\right) - \ln\left(\frac{P^{T-}}{D_t}\right) = \ln\left(\frac{P^{T+}}{P^{T-}}\right) \tag{6}$$

The *price ratio*, "pr_t", is the log ratio of long- to short-term dividend prices. We use the log difference instead of level difference to get rid of D_t, so that pr_t only contains market prices, and thereby, captures the variation of expected return at relatively higher frequencies than pd_t. In the literature, and as in this paper, the current dividend D_t is measured by the sum of dividends paid in the previous year to remove seasonality (Fama and French (1988)), so through D_t, pd_t tends to be more slow-moving than pr_t.

Together, pd_t and pr_t should reflect the information content of $\left(\frac{P^{T+}}{D_t}, \frac{P^{T-}}{D_t}\right)$. We will show that pr_t is a better way to extract information about future returns. Intuitively, the valuation of long-term dividends is more sensitive to discount rate movements than the valuation of short-term dividends. The ratio of the former to the latter tends to increase when the discount rate declines, and decrease when the discount rate rises.

To construct pr_t, we need the short-term dividend price and the long-term dividend price, which are calculated using data of S&P 500 futures and zero-coupon bonds (ZCBs) as follows.[7] Consider any $T > 0$. To calculate P_t^{T+} from futures price and ZCB price, we make the assumption that $\int_t^{t+T} r_s ds$ and S_{t+T} are not correlated under Q measure, so we have

$$P_t^{T+} = \sum_{\tau=T+1}^{\infty} \mathbb{E}_t^Q \left[e^{-\int_t^{t+T} r_s ds} e^{-\int_{t+T}^{t+\tau} r_s ds} D_{t+\tau} \right] = \mathbb{E}_t^Q \left[e^{-\int_t^{t+T} r_s ds} \underbrace{\mathbb{E}_{t+T}^Q \left[\sum_{\tau=T+1}^{\infty} e^{-\int_{t+T}^{t+\tau} r_s ds} D_{t+\tau} \right]}_{S_{t+T}} \right]$$

$$= \mathbb{E}_t^Q \left[e^{-\int_t^{t+T} r_s ds} S_{t+T} \right] = \underbrace{\mathbb{E}_t^Q \left[e^{-\int_t^{t+T} r_s ds} \right]}_{ZCB_t^T} \mathbb{E}_t^Q \left[S_{t+T} \right]. \tag{7}$$

Therefore, we can calculate P_t^T directly from the price of ZCB that matures in T periods, ZCB_t^T, and futures price that is the Q-expectation of future stock price (Duffie (2001)).

2.2 Predicting return

Data and summary statistics. We use monthly data of S&P 500 futures (source: Bloomberg) and zero-coupon bond prices (source: Fama-Bliss database) from January 1988 to June 2017 to construct pr_t.[8] The sample starts in 1988 because after the market crash of October 1987 regulators overhauled several trade-clearing protocols.[9] pd_t is the month-end

[7]Figure 7 in the appendix shows that pr_t constructed from futures data has 88% correlation with pr_t constructed from option data in Binsbergen, Brandt, and Koijen (2012).

[8]Available maturities vary over time, so to obtain futures at constant maturities, we interpolate data. We use shape-preserving piecewise cubic interpolation to preserve the shape of the futures curve.

[9]The stock market crash in October 1987 reveals anomalous trading behavior in the futures market that was largely driven by portfolio insurance (Brady Report (1988)). According to the New York Stock Exchanges current website: "In response to the market breaks in October 1987 and October 1989, the New

Table 1: Summary Statistics

This table reports the number of observations, mean, standard deviation, minimum, maximum, quartiles, and first-order (one-month) autocorrelation (ρ) of our predictor, pr_t (the ratio of long-term dividend price to short-term dividend price) and pd_t (the price-dividend ratio). The correlation matrix is shown at the end of the table. Using Equation (7), we construct long-term dividend price from data of S&P 500 futures price and zero-coupon bond price (source: Bloomberg), and short-term dividend price is the difference between S&P 500 index value and long-term dividend price. pd_t is the month-end price-to-dividend ratio of S&P 500 index (source: Bloomberg).

	obs	mean	std	min	25%	50%	75%	max	ρ	corr.	pr	pd
pr	348	3.992	0.531	2.677	3.630	3.992	4.195	6.631	0.915		1.000	
pd	348	3.873	0.307	3.241	3.594	3.887	4.052	4.551	0.987		0.874	1.000

price-dividend ratio of S&P 500 index (source: Bloomberg). We set T equal to one year, so pr_t is the log ratio of price of dividends paid beyond the coming year to the price of dividends paid within the coming year. Accordingly, we focus on forecasting the return of S&P 500 index at the one-year horizon but also report the forecasting results at the one-month horizon in the appendix.

Table 1 reports the summary statistics of pr_t, and the log price-dividend ratio pd_t for comparison. We can interpret pr_t as a measure of duration. Its median value, 3.992, translates into 54.2 after taking exponential, meaning that the valuation of dividends in two years onward is 54.2 times the valuation of dividends in the next year. In other words, the market has a *valuation duration* of a total 55.2 years. pr_t has a wide range of variation, with a minimum of 2.677 (i.e., 15.5 years) right before the 1990-1991 recession (Jun. 1990) and a maximum of 6.631 (i.e., 759.2 years) near the end of the dot-com boom (Nov. 2000).

pr_t has a lower one-month autocorrelation ("ρ") than pd_t. The persistence of predictors is a major concern in the literature on return forecasting, especially due to the associated small-sample bias (Nelson and Kim (1993); Stambaugh (1999)) and spurious regression when the underlying expected return is persistent (Ferson, Sarkissian, and Simin (2003)).

The correlation between pr_t and pd_t is 0.87. As shown in the cross-spectrum in Figure 1, the high correlation is mainly from low-frequency movements. When forecasting the market return, we will consider pd_t and pr_t separately as univariate predictors, and also examine the predictive power of their components orthogonalized to each other.

Inference and forecasting evaluation. We run the following regression to predict one-year return:

$$r_{t,t+12} = \alpha + \beta x_t + \epsilon_{t,t+12}, \tag{8}$$

York Stock Exchange instituted circuit breakers to reduce volatility and promote investor confidence. By implementing a pause in trading, investors are given time to assimilate incoming information and the ability to make informed choices during periods of high market volatility."

where x_t is a predictor. Twelve-month forecasts use overlapping monthly data, so we adjust our standard errors to reflect the dependence that overlap introduces into error terms. Following Cochrane and Piazzesi (2002), we report Newey and West (1987) standard errors with 18 lags to account for the moving-average structure induced by overlap. Besides, we also calculate Hodrick (1992) standard errors. Hodrick (1992) shows that GMM-based autocovariances correction (e.g., Newey and West (1987)) can have poor small-sample properties, Related to the serial correlation in errors, another concern is the persistence of predictor that induces bias in β estimate. We report the estimate adjusted for Stambaugh (1999) bias. In the appendix, we report the IVX-Wald test of predictive coefficient in Kostakis, Magdalinos, and Stamatogiannis (2015) that accounts for the persistence of predictor (Table 16).

The adjusted R^2 measures in-sample fitness. Several studies have raised concerns over out-of-sample performances of return predictors (Bossaerts and Hillion (1999); Goyal and Welch (2007)). To address these issues, we report the out-of-sample R^2 and two formal tests of out-of-sample performances. We calculate out-of-sample forecasts as a real-time investor, using data up to time t in the predictive regression to estimate β, which is then multiplied by the time-t value of the predictor to form the forecast. Out-of-sample forecasting start from December 1997, when we have at least ten years of data. Out-of-sample R^2 is

$$R^2_{OOS} = 1 - \frac{\sum_t \left(r_{t,t+12} - \hat{r}_{t,t+12}\right)^2}{\sum_t \left(r_{t,t+12} - \overline{r}_t\right)^2},$$

where $\hat{r}_{t,t+12}$ is the forecast value and $\overline{r}$ is the average of twelve-month returns (the first is January-December 1998). The out-of-sample R^2 lies in the range $(-\infty, 1]$, where a negative number means that a predictor provides a less accurate forecast than the historical mean.

We report the p-value of two tests of out-of-sample performance, "ENC" and "CW". ENC is the encompassing forecast test derived by Clark and McCracken (2001), which is widely used in the forecasting literature. We test whether the predictor has the same out-of-sample forecasting performance as the historical mean, and compare the value of the statistic with critical values calculated by Clark and McCracken (2001) to obtain a range of p-value. Besides, Clark and West (2007) adjust the standard MSE t-test statistic to produce a modified statistic (CW) that has an asymptotic distribution well approximated by the standard normal distribution, so for CW, we report the precise p-value.

One-year return prediction. Table 2 presents the results of annual return forecasting. Column (1) shows that our price ratio, pr_t, demonstrates a striking degree of predictability for one-year returns. The in-sample implementation generates a predictive R^2 reaching

Table 2: One-year Return Prediction

This table reports the results of predictive regression (Equation (8)). The left-hand side variable is the return of S&P 500 index in the next twelve months. We consider four the right-hand side variables (i.e., predictors), pr_t, pd_t, the residuals of pr_t after regressing on pd_t (ϵ_t^{pr}), and the residuals of pd_t after regressing on pr_t (ϵ_t^{pd}), and the results are reported in Column (1) to (4) respectively. The β estimate is shown followed by Newey and West (1987) t-statistic (with 18 lags), Hodrick (1992) t-statistic, the coefficient adjusted for Stambaugh (1999) bias, and the in-sample adjusted R^2. We run the regression monthly. Starting from December 1997, we form out-of-sample forecasts of return in the next twelve months by estimating the regression with data up to the current month, and use the forecasts to calculate out-of-sample R^2, ENC test (Clark and McCracken (2001)), and the p-value of CW test (Clark and West (2007)).

	pr_t	pd_t	ϵ_t^{pr}	ϵ_t^{pd}
β	-0.138	-0.193	-0.160	0.098
Newey-West t	(-4.718)	(-3.575)	(-2.233)	(0.848)
Hodrick t	[-2.743]	[-2.217]	[-1.677]	[0.613]
Stambaugh bias adjusted β	-0.127	-0.182	-0.152	0.107
R^2	0.238	0.157	0.076	0.010
OOS R^2	0.192	0.068	0.048	-0.043
ENC	4.052	1.776	2.249	-0.241
$p(ENC)$	< 0.01	< 0.10	< 0.05	> 0.10
$p(CW)$	0.007	0.041	0.111	0.348

23.8%.[10] Out-of-sample forecasts are similarly powerful, delivering an R^2 of 19.2%, significantly outperforming the historical mean as shown by the p-values of ENC and CW.

Campbell and Thompson (2008) calculate a long-term estimate of the market Sharpe ratio ("s_0") equal to 0.374. In the Appendix (see also Kelly and Pruitt (2013)), we show that the Sharpe ratio of a mean-variance investor's market-timing strategy ("s_1") is related to s_0 through $s_1 = \sqrt{\frac{s_0^2 + R^2}{1 - R^2}}$, where R^2 is the out-of-sample R^2 when pr_t is used as annual return predictor. Therefore, an out-of-sample R^2 of 19.2% implies a Sharpe ratio of 0.84, suggesting that the stochastic discount factor is more volatile than implied by state-of-art asset pricing models (e.g., Campbell and Cochrane (1999) and Bansal and Yaron (2004)).

The predictive coefficient is also large in magnitude, indicating high volatility of the expected return. A decrease of pr_t by one standard deviation adds 7.3% to the expected return. Both Newey-West and Hodrick t-statistics are significant at least at the 1% level.

Column (2) reports the results for pd_t. The return predictive power of pd_t is weaker

[10] Foster, Smith, and Whaley (1997) discuss the potential data mining issues that arise from researchers searching among potential regressors. They derive a distribution of the maximal R^2 when k out of m potential regressors are used as predictors, and they calculate the critical value for R^2, below which the prediction is not statistically significant. For instance, when $m = 50$, $k = 5$, and the number of observations is 250, the 95% critical value for R^2 is 0.164.

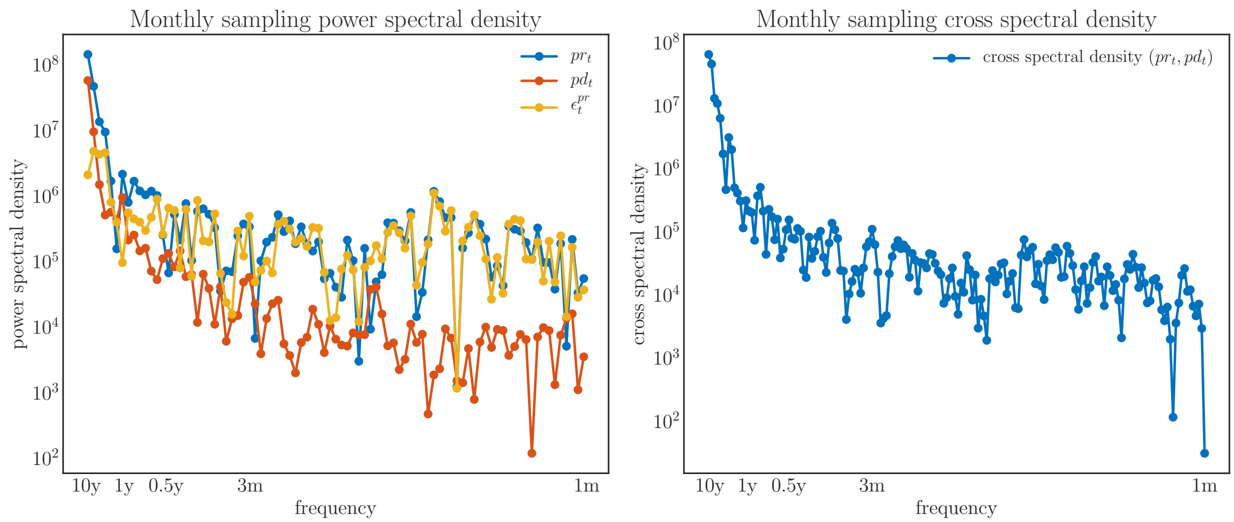


Figure 1: **Spectrum and Cross-spectrum of Price Ratio and Price-Dividend Ratio.** The left panel shows the estimated spectral densities of pr_t, pd_t, and the residuals of pr_t after regressing on pd_t ($\epsilon^p r_t$). The integral of spectral density is equal to the variance. The horizontal line starts from zero and ends at π, but labeled with the corresponding length of a cycle. The right panel shows the cross-spectral density between pr_t and pd_t. The integral of cross-spectral density is equal to the covariance.

than pr_t in all aspects. Its in-sample and out-of-sample R^2 is almost half of those of pr_t. Its coefficient is smaller and less significant. Moreover, an decrease in pd_t by one standard deviation leads to an increase of expected return by 5.8%, implying a less volatile expected return than the one from pr_t. In the appendix, the IVX-Wald test of Kostakis, Magdalinos, and Stamatogiannis (2015) in Table 16 also confirms the significant predictive power of pr_t while rejects the predictive power of pd_t.

Since pr_t and pd_t are highly correlated, we regress pr_t on pd_t to obtain residuals, ϵ_t^{pr}, that are orthogonal to pd_t in sample, and use the residual as a predictor to evaluate the return predictive power of pr_t beyond pd_t. The results are reported in Column (3). pr_t residual still delivers in-sample and out-of-sample R^2 of 9%, showing a very strong incremental predictive power of pr_t. Note that to obtain out-of-sample forecasts, at time t we obtain the residuals ϵ_t^{pr} only using data up to t from the regression of pr_t on pd_t, and then use these residuals to estimate the predictive regression. In Column (4), ϵ_t^{pd}, the residuals from regressing pd_t on pr_t does not show return predictive power, which again confirms pr_t as the superior predictor.

Variation in the frequency domain. To better understand the incremental predictive power of pr_t beyond pd_t, Panel A of Figure 1 shows the spectrum of pr_t, pd_t, and ϵ_t^{pr} (residuals from regressing pr_t on pd_t). The area under spectrum (integral) is the variance, so the spectrum graph provides a variance decomposition in the frequency domain. On the horizontal axis, instead of showing the frequencies from zero to π, we mark the corresponding length of cycle for easier interpretation. Consistent with the fact that pr_t is less persistent

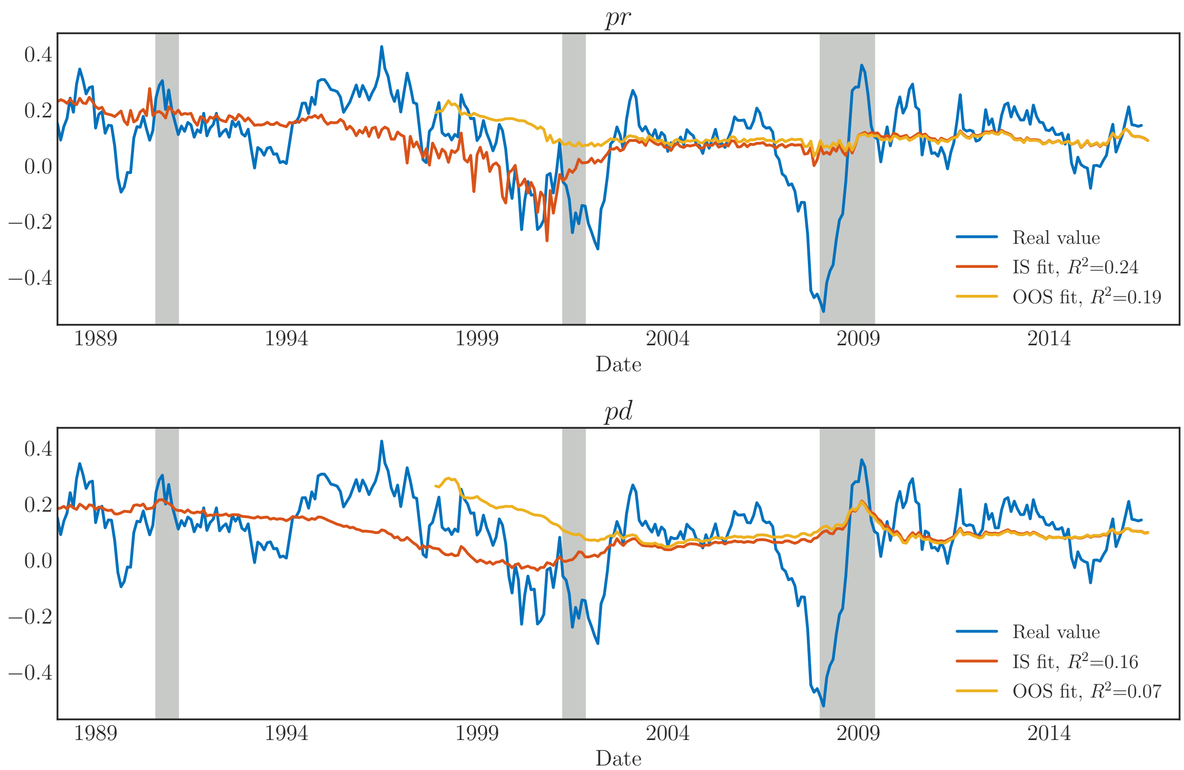


Figure 2: **Expected Return Dynamics.** The graph reports the in-sample fitted value, the out-of-sample forecast, and the realized twelve-month return of S&P 500 index. The date on horizontal axis is the beginning date of the twelve-month period. Starting from December 1997, we form out-of-sample forecasts of return in the next twelve months by estimating the predictive regression with data up to the current month.

than pd_t, its variation is also more concentrated in higher frequencies than the variation of pd_t. Once orthogonalized to pd_t, pr_t's residual varies mainly at frequencies above one year.

Panel B plots the cross-spectrum of pr_t and pd_t. The integral is the covariance between pr_t and pd_t. The correlation between pr_t and pd_t is mainly from low frequencies. This again indicates that it is the high-frequency variation in pr_t that adds the return predictive power.

Expected return dynamics. Figure 2 plots the realized market return, the in-sample fitted value, and the out-of-sample forecast. The horizontal axis shows the *beginning date* of each twelve-month return, i.e., the time when the expectation is formed. As before, out-of-sample forecasts at time t only uses data up to time t to estimate the predictive coefficient. The out-of-sample forecasting starts from December 1997 when we have at least ten years. We plot separately the expected return from pr_t and that from pd_t. For both predictors, in- versus out-of-sample estimates of the expected return are fairly consistent with each other.

The first message from this graph is that in contrast to pd_t, which produces a smooth expected return over time, pr_t reveal variations of expected return at higher frequencies. This observation is consistent with the Figure 1, and the fact that pr_t is less persistent. pr_t is more

responsive to news, as it contains only the market prices of short- and long-term dividends. In contrast, pd_t has a denominator that is a rolling accumulation of past dividends.

Our sample has three recession periods (shaded). Near the end of each recession, the expected return tends to increase, which is in line with studies that document countercyclical equity premium (e.g., Fama and French (1989); Ferson and Harvey (1991)). Related to the high-frequency variation revealed by pr_t, such increase is sharper for the estimate from pr_t than that from pd_t. Another interesting finding is that in the year leading up to the dot-com burst and the global financial crisis, the expected return from pr_t exhibits slump, while the expected return from pd_t barely moves. Also, the expected return from pr_t starts to recover near the end of these recessions, while the expected return from pd_t only shows a smooth upward trend throughout the recession. These new patterns from pr_t as the expected return proxy are informative for constructing asset pricing and macroeconomic models.

Other predictors. How do our market return forecasts compare with predictors proposed in earlier literature? Figure 3 compares the predictive accuracy of our approach with an extensive collection of existing predictors. In the caption, we document the sources. We consider 18 alternative predictors including the price-dividend ratio (pd), the default yield spread (dfy), the inflation rate (infl), stock variance (svar), the cross-section premium (csp), the dividend payout ratio (de), the long-term yield (lty), the term spread (tms), the T-bill rate (tbl), the default return spread (dfr), the dividend yield (dy), the long-term rate of return (ltr), the earnings-to-price ratio (ep), the book to market ratio (bm), the investment-to-capital ratio (ik), the net equity expansion ratio (ntis), the percent equity issuing ratio (eqis), the consumption-wealth-income ratio (cay), the short interests index (SII), the option-implied lower bound of 1-year equity premium (SVIX) and Kelly and Pruitt (2013) factor extracted from 100 book-to-market and size portfolios (kp).

Most predictors are studied in a return predictability survey by Goyal and Welch (2007), and others are proposed more recently, such as short interest index ("SII" in Rapach, Ringgenberg, and Zhou (2016)) and SVIX (Martin (2017)). We report in-sample ("IS") R^2, out-of-sample ("OOS") R^2, the absolute values of Newey-West and Hodrick t-statistics. In our sample, pr_t outperforms other predictors in all aspects. Among the alternatives, the price-dividend ratio and the book-to-market ratio ("bm") deliver the most successful univariate forecasts, while others either fail in the out-of-sample R^2 (e.g., cay, the consumption-wealth ratio of Lettau and Ludvigson (2001)) or in statistical significance (e.g., ik, the investment-capital ratio of Cochrane (1991)). In the appendix, we report the correlation between pr_t and other predictors. pd_t, bm, ik, and dy show significant correlations.

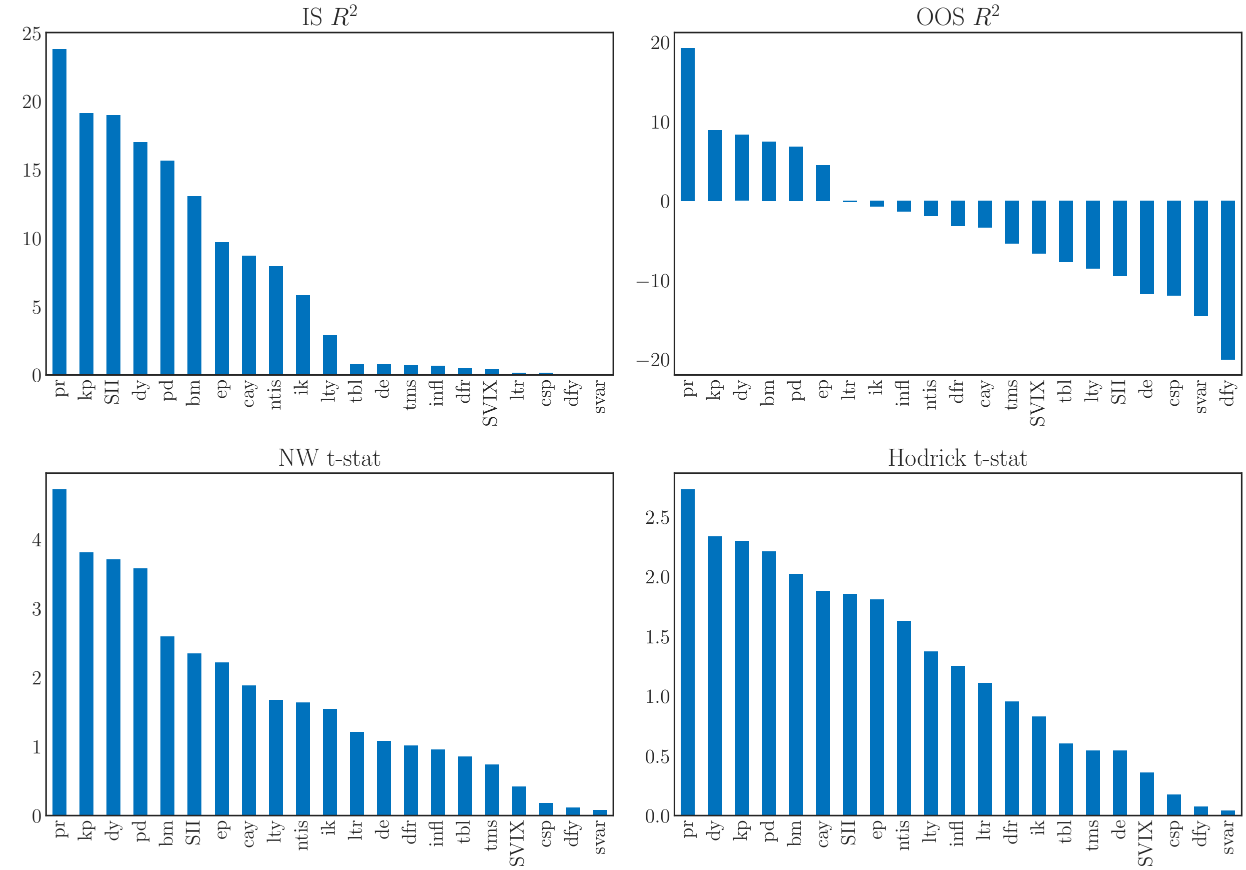


Figure 3: **Comparison with Alternative Return Predictors.** This graphs compares the 1-year return predictive power between pr_t and other commonly studied predictors from 1988 to 2016. Panel A reports the in-sample adjusted R^2. Panel B reports the out-of-sample R^2. Negative out-of-sample R^2 indicates that the predictive power is below historic mean. Panel C reports the absolute values of Newey and West (1987) t-statistic (with 18-month lag). Panel D reports the absolute values of Hodrick (1992) t-statistic. Most alternative predictors are from Goyal and Welch (2007) and include the price-dividend ratio (pd), the default yield spread (dfy), the inflation rate (infl), stock variance (svar), the cross-section premium (csp), the dividend payout ratio (de), the long-term yield (lty), the term spread (tms), the T-bill rate (tbl), the default return spread (dfr), the dividend yield (dy), the long-term rate of return (ltr), the earnings-to-price ratio (ep), the book to market ratio (bm), the investment-to-capital ratio (ik), the net equity expansion ratio (ntis), the percent equity issuing ratio (eqis), and the consumption-wealth-income ratio (cay). SII is the short interests index from Rapach, Ringgenberg, and Zhou (2016) (1988-2014). SVIX is option-implied lower bound of 1-year equity premium from Martin (2017) (1996-2012). kp is the single predictive factor extracted from 100 book-to-market and size portfolios from Kelly and Pruitt (2013).

2.3 Predicting dividend growth

As shown in Equation (4), the price-dividend ratio compresses information about expected return and expected dividend growth. As the return predictive power is concentrated in pr_t, the component of pd_t that is orthogonal to pr_t (i.e., ϵ_t^{pd}) should forecast dividend growth. We

Table 3: Dividend Growth Prediction

This table reports the results of dividend growth forecasting regression. The left-hand side variable is the one-year, non-overlapping dividend growth rate of S&P 500 index defined in Equation (10). We consider four the right-hand side variables (i.e., predictors), the residuals of pd_t after regressing on pr_t (ϵ_t^{pd}), pd_t, pr_t, the equity yield ($\ln\left(\frac{D_t}{P_t^{T-}}\right)$), and the results are reported in Column (1) to (4) respectively. The estimated predictive coefficient (β) is shown followed by Newey and West (1987) t-statistic (with 18 lags), Hodrick (1992) t-statistic, the coefficient adjusted for Stambaugh (1999) bias, and the in-sample R^2. We run the regression monthly. Starting from December 1997, we form out-of-sample forecasts of return in the next twelve months by estimating the regression with data up to the current month, and use the forecasts to calculate out-of-sample R^2, ENC test (Clark and McCracken (2001)), and the p-value of CW test (Clark and West (2007)).

	ϵ_t^{pd}	pd_t	pr_t	$\ln\left(\frac{D_t}{P_t^{T-}}\right)$
β	0.307	0.014	-0.035	-0.127
Newey-West t	(3.204)	(0.247)	(-2.005)	(-3.395)
Hodrick t	[5.153]	[0.642]	[-3.990]	[-6.767]
Stambaugh bias adjusted β	0.316	0.025	-0.024	-0.118
R^2	0.349	0.003	0.057	0.233
OOS R^2	0.304	-0.045	0.046	0.222
$p(ENC)$	< 0.01	> 0.10	< 0.10	< 0.01
$p(CW)$	0.011	0.418	0.054	0.001

measure dividend growth by the ratio of adjacent, non-overlapping cumulative dividends,

$$\Delta D_{t,t+12} = \frac{\sum_{i=1}^{12} D_{t+i}}{\sum_{i=1}^{12} D_{t-12+i}}.^{11} \tag{9}$$

In the predictive regression, we use the logarithm of $\Delta D_{t,t+12}$ as the forecasting target.

Table 3 reports the results of dividend growth prediction.[12] Column (1) shows that ϵ_t^{pd}, the residual of pd_t after regressing on pr_t, exhibits very strong predictive power with in-sample and out-of-sample R^2 of 34.9% and 30.4% respectively. The coefficient has a large magnitude and is statistically significant. One standard-deviation increase of ϵ_t^{pd} is associated with 4.55% increase of dividend growth (i.e., 3.7 standard deviations). Column (2) shows that pd_t itself does not strongly predict dividend growth. Together, Table 2 and 3 show that the information about future return and dividend is mingled together in pd_t. Such information is disentangled, once pd_t is decomposed by cash-flow horizon, and the price ratio

[11] Dividends are calculated from the difference between cum- and ex-dividend S&P index levels.

[12] Forecasting dividend growth has been at the center of asset pricing literature (see Ball and Watts (1972), Campbell and Shiller (1988), Cochrane (1992), Fama and French (2000), Lettau and Ludvigson (2005), Koijen and van Nieuwerburgh (2011), Lacerda and Santa-Clara (2010) and Golez (2014)).

pr_t is constructed to capture the information about expected return. Column (3) shows that in comparison with ϵ_t^{pd}, the dividend predictive power of pr_t is weaker, with an out-of-sample R^2s only 15% of the out-of-sample R^2 of ϵ_t^{pd}. Thus, the decomposition of pd_t into pr_t and ϵ_t^{pd} adequately separates the information on expected return and dividend growth.

Our analysis of return and dividend predictability echoes the observation of Cochrane (2007) that price-dividend ratio must either predict return or dividend growth, but we show an even richer story: the predictive information on return and dividend cancels out each other within pd_t (Lettau and Ludvigson (2005)). Once we tease out the information on the future return, the rest of pd_t predicts dividends better than pd_t itself.

ϵ^{pd_t} is related to the "equity yield" in Binsbergen, Hueskes, Koijen, and Vrugt (2013) in its dividend predictive power. Following that paper, we define equity yield as $\ln\left(\frac{D_t}{P_t^{T-}}\right)$. The following equation directly decomposes pd_t into pr_t and the equity yield:

$$pd_t = \ln\left(1 + e^{pr_t}\right) - \ln\left(\frac{D_t}{P_t^{T-}}\right) \approx \kappa_0 + \kappa_1 pr_t - \kappa_2 \ln\left(\frac{D_t}{P_t^{T-}}\right), \tag{10}$$

where the linearizion coefficients are $\kappa_1 = \frac{\overline{\exp(pr)}}{1+\overline{\exp(pr)}}$, $\kappa_2 = 1$, and $\kappa_0 = \ln\left(1 + \overline{\exp(pr)}\right) - \kappa_2 \overline{pr}$. The upper bar represents long-run means, around which we log-linearize the equation. The correlation between pr_t and the equity yield is 0.86 in our sample, so Equation (10) is only an imperfect decomposition. As shown in Column (4) of Table 3, the equity yield also predicts dividend growth, albeit with a forecasting power less than ϵ_t^{pd}.[13]

2.4 A structural interpretation

The economic intuition behind our results can be understood by imposing more structure on the state space model. Explicitly, we construct the stochastic discount factor and aggregate dividend following Lettau and Wachter (2007), and show that the return predictive power of pr_t depends on the persistence of expected dividend growth. This model is nested by the previous state space model as motivation. To streamline the exposition, here we define one unit of time as one year, instead of one month as in the empirical exercises.

Model and solution. The economy has three independent shocks: one to dividend growth, one to expected dividend growth, and a preference shock. A 3×1 vector ε_{t+1} record these standard normal shocks. The aggregate dividend is assumed to evolve according to

$$\Delta d_{t+1} = g + z_t + \sigma_d \varepsilon_{t+1}, \tag{11}$$

[13] Our sample period differs from Binsbergen, Hueskes, Koijen, and Vrugt (2013) (October 2002 and April 2011) who use dividend derivatives for dividend prices, so the estimates of predictive coefficient are different.

where z_t follows the AR(1) process

$$z_{t+1} = \phi_z z_t + \sigma_z \varepsilon_{t+1}, \tag{12}$$

with $|\phi_z| < 1$. The conditional mean of dividend growth is $g + z_t$. Row vectors σ_d and σ_z multiply the shocks on dividend growth and z_{t+1}. The conditional standard deviation of Δd_{t+1} equals $\|\sigma_d\| = \sqrt{\sigma_d \sigma_d'}$. Similarly, the conditional standard deviation of z_{t+1} equals $\|\sigma_z\| = \sqrt{\sigma_z \sigma_z'}$, and its conditional covariance with Δd_{t+1} is $\sigma_d \sigma_z'$. This cash flow process is also explored by Campbell and Cochrane (1999) and Bansal and Yaron (2004).

The stochastic discount factor is directly specified for this economy. In particular, we assume that the price of risk is driven by a single variable x_t that follows the AR(1) process

$$x_{t+1} = (1 - \phi_x)\,\bar{x} + \phi_x x_t + \sigma_x \varepsilon_{t+1}, \tag{13}$$

where $|\phi_x| < 1$ and σ_x is a 1×3 vector. For simplicity, the risk-free rate, denoted r^f, is constant. Following Lettau and Wachter (2007), and in line with Campbell and Cochrane (1999) and Menzly, Santos, and Veronesi (2004), we assume only dividend risk is priced. Let $\varepsilon_{d,t+1} = \frac{\sigma_d}{\|\sigma_d\|}\varepsilon_{t+1}$ denote the normalized dividend shock. The stochastic discount factor is

$$M_{t+1} = \exp\{-r^f - \frac{1}{2}x_t^2 - x_t \varepsilon_{d,t+1}\}. \tag{14}$$

$\ln E_t\left[M_{t+1}\right] = -r^f$, so it follows from no-arbitrage that r^f is indeed the risk-free rate. In Campbell and Cochrane (1999), the price of risk (x_t here) is perfectly negatively correlated with cash flow growth, which corresponds to $\sigma_x/\|\sigma_x\| = \sigma_d/\|\sigma_d\|$. Here, as in Lettau and Wachter (2007), we allow the conditional correlation to be imperfect, and interpret shocks that are uncorrelated with changes in fundamentals as preference or sentiment shocks.

We solve pr_t, which equals $\ln\left(\frac{S_t}{P_t^{T-}} - 1\right)$. First, we solve P_t^{T-}, the price of one-year dividend following Lettau and Wachter (2007):

$$P_t^{T-} = D_t \exp\{A^{T-} + B^{T-} x_t + C^{T-} z_t\}, \tag{15}$$

where A^{T-} is a constant, and the constant coefficients on x_t and z_t are

$$B^{T-} = -\|\sigma_d\|, \quad \text{and} \quad C^{T-} = 1.$$

Next, we solve the price of all dividends from the next year to the indefinite future, as in Bansal and Yaron (2004), using Campbell and Shiller (1988) approximation of market return

r_{t+1}, i.e., $\kappa_0 + \kappa_1 pd_{t+1} - pd_t + \Delta d_{t+1}$, and no-arbitrage condition (details in the Appendix):

$$S_t = D_t \exp\{A + Bx_t + Cz_t\}, \tag{16}$$

where A is a constant, and the constant coefficients on x_t and z_t are

$$B = -\frac{\frac{\sigma_d \sigma_z'}{\|\sigma_d\|}\kappa_1 C + \|\sigma_d\|}{\frac{\sigma_d \sigma_x'}{\|\sigma_d\|}\kappa_1 + \kappa_1 \phi_x - 1}, \quad \text{and} \quad C = \frac{1}{1 - \kappa_1 \phi_z}.[14]$$

The return predictor pr_t is a function of S_t / P_t^{T-}, which in turn depends on x_t and z_t:

$$\frac{S_t}{P_t^{T-}} = \exp\{A - A^{T-} + \left(B - B^{T-}\right) x_t + \left(C - C^{T-}\right) z_t\}. \tag{17}$$

Finally, we solve the expected market return that only depends on x_t, the price of risk:

$$\mathbb{E}_t\left[r_{t+1}\right] = A^r + B\left(\kappa_1 \phi_x - 1\right) x_t, \tag{18}$$

where A^r is a constant and the coefficient of x_t is a product of B and $(\kappa_1 \phi_x - 1)$.

Equation (11) and (18) show that the state space model that motivates our decomposition of price-dividend ratio nests this structural model. F_t contains x_t and z_t, with the former driving the expected market return and the latter driving the expected dividend growth.

Return predictability and the expected dividend growth. From Equation (18), we know that to predict return, all we need is x_t. pd_t and pr_t depend on both x_t and z_t. The information on future dividends may compromise the return predictive power (Menzly, Santos, and Veronesi (2004); Lettau and Ludvigson (2005)). For example, Lettau and Wachter (2007) calibrate the shock correlation between x_t and z_t to zero, so z_t adds pure noise. In fact, when the expected dividend growth is transient ($\phi_z = 0$), pr_t perfectly reveals x_t, because $\phi_z = 0$ implies $C = 1$ and the coefficient of z_t in S_t/P_t^{T-} $(C - C^{T-})$ is zero.

Proposition 1 *In an economy with stochastic discount factor given by Equation (14) and aggregate dividend growth given by Equation (11), there is a one-to-one mapping from pr_t to x_t if and only if the expected dividend growth is not persistent, i.e., $\phi_z = 0$.*

Our findings on dividend predictability lend further support to this structural interpretation of pr_t being a superior return predictor. In Table 3, we show that ϵ_t^{pd}, the residuals

[14] Lettau and Wachter (2007) solve the total value of dividends using a different approximation that sums up the closed-form prices of dividend strips up to a finite horizon and approximate the residual value by exploiting the fact that strip prices' coefficients on x_t and z_t converge to horizon-independent fixed points.

from regressing pd_t on pr_t, predict dividend growth with an out-of-sample R^2 of 30.4%. When $\phi_z = 0$, pr_t perfectly reveals x_t, so ϵ_t^{pd} precisely maps to z_t. Moreover, pd_t itself does not predict dividends, in line with the existing literature (e.g., Cochrane (2011)). Moreover, $\ln\left(P_t^{T-}/D_t\right)$ predicts dividends, but its power is weaker than ϵ_t^{pd} because, as shown in the model, $\ln\left(P_t^{T-}/D_t\right)$ is $A^{T-} + B^{T-}x_t + C^{T-}z_t$ so the variation of z_t is masked by that of x_t.

Using ϵ_t^{pd} as a proxy for the expected dividend growth, we can examine its persistence directly. Figure 11 in the appendix shows that autocorrelations of ϵ_t^{pd} are not statistically different from zero. Since next we show the return predictive power of pr_t outside the U.S., we report the autocorrelations of ϵ_t^{pd} both in the U.S. and other countries.

Finally, we fit a state space model to aggregate dividends, and report the estimates in Table 15 in the appendix. The null hypothesis of zero persistence in the expected dividend growth cannot be rejected when both S&P 500 dividends and the total cash payout of CRSP NYSE/AMEX/NASDAQ Cap-Based index are used as dividends. Note that in Table 3, the dividend predictive power of pr_t seems stronger than pd_t (still much weaker than ϵ_t^{pd}'s). This is due to the shock correlation between x_t and z_t, which causes an unconditional correlation between pr_t and the expected dividend growth. Our results are robust to different correlations between structural shocks (Figure 10 in the appendix).

When z_t is a constant, pr_t and pd_t are both perfect proxies for x_t, but when z_t varies over time, the return predictive power of pd_t is compromised. Variants of growth-adjusted valuation ratios have been proposed (Campbell and Thompson (2008); Lacerda and Santa-Clara (2010); Da, Jagannathan, and Shen (2014); Golez (2014)). Binsbergen and Koijen (2010), Rytchkov (2012), and Jagannathan and Liu (2015) use state-space models to filter out ans separate the information on expected return and dividend growth. We contribute to this line of research by proposing a *model-free* predictor that is directly constructed from market prices. Moreover, we construct a dividend predictor by simply regressing pd_t on pr_t.

2.5 Predicting return outside the United States.

A potential concern is that our US sample of thirty years (354 monthly observations) is relatively short. We close this section with international evidence on return predictability.

Sample construction. The index return and futures data are obtained from Datastream. Zero coupon bond yields and index dividends are obtained from Bloomberg. We start with all developed countries with index futures, and drop a country from the sample if one of the following criteria is met: 1) futures with maturity larger or equal to one year do not exist (Germany, Hong Kong, Switzerland) or exist for less than five years (Norway); 2) futures

Table 4: International Panel Return Prediction

This table reports the results of return forecasting regression (Equation (19)) using the panel data of Australia, France, Japan, Spain, the United Kingdom, and the United States. The left-hand side variable is the one-month, non-overlapping index return of a country, and for the right-hand side variable, we consider pr_t (Column 1 and 2), pd_t (Column 3 and 4), ϵ_t^{pr} (Column 5 and 6), and ϵ_t^{pd} (Column 7 and 8) in that country. ϵ_t^{pr} is the residual of pr_t after regressing on pd_t, and ϵ_t^{pd} is the residual of pd_t after regressing on pr_t. For each predictor, we report both the results with and without time fixed effects. The estimated predictive coefficient (β) is shown followed by Hodrick (1992) t-statistic. In each column, we report whether country and time fixed effects are included, the number of observations, and adjusted R^2. We drop observations with negative one-year dividend strip prices, so the estimation using pr_t has a shorter sample than that using pd_t.

pr_t	-0.189	-0.051						
	[-4.497]	[-0.740]						
pd_t			-0.109	-0.135				
			[-3.843]	[-1.339]				
ϵ_t^{pr}					-0.404	-0.113		
					[-7.087]	[-1.438]		
ϵ_t^{pd}							-0.063	-0.083
							[-1.961]	[-1.230]
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Time FE		✓		✓		✓		✓
Obs	1,469	1,469	1,553	1,553	1,469	1,469	1,469	1,469

price exhibits strong seasonality (Italy, Netherlands, and Switzerland) or break (Canada).[15] For each country, our sample starts from the earliest date when index return, futures, and dividend data are all available. We end up with 1,469 country-month observations: UK (FTSE100, starting in 1993), France (CAC40, starting in 1998), Spain (IBEX35, starting in 1994), Australia (ASX200, starting in 2002), and Japan (Nikkei225, starting in 1993). We construct pr_t and pd_t, and estimate ϵ_t^{pr} and ϵ_t^{pd} country-by-country.

International return predictability. We supplement the US sample with data from the other five countries, and use this unbalanced panel to test the return predictive power of pr_t.

In the panel data regression, the left-hand side variable is the future stock market return in a country, and the right-hand side variable of interest is pr_t in that country. Instead of running the typical predictive regression with overlapping returns on the left-hand side, we follow the suggestion of Hodrick (1992) and run the following ("reverse") regression to test the return predictive power of pr_t at one-year horizon.

$$12\ r_{t,t+1}^n = \alpha + \beta \left(\frac{1}{12} \sum_{i=0}^{11} x_{t-i}^n \right) + \epsilon_{t,t+1}^n, \tag{19}$$

[15] In the appendix, Figure 9 plots the futures-to-spot ratio for these four countries.

where n represents a country.[16] The dependent variable is the (annualized) next-month market return (non-overlapping), and the predictor is averaged in the most recent twelve months. Hodrick (1992) points out the difficulties in inference when using overlapping observations, especially given the the poor small-sample properties of GMM-based autocovariances correction (e.g., Newey-West standard error), and suggest this reverse regression (19) for drawing inference on long-run prediction.[17] We cluster the standard error by time and country. The model of Equation (19) combines the small-sample properties of Hodrick (1992) standard error and the robustness of clustered standard errors to various error correlations.

Table 4 reports the results. Column (1) shows the strong return predictive power of pr_t after controlling for the heterogeneity in level of equity premium across countries (through country fixed effect). The coefficient estimate is similar to the predictive coefficient in the U.S. sample, and more statistically significant. The comparison between Column (1) and (3) of Table 4 shows that the return predictive power of pr_t is stronger than pd_t. Column (5) shows that the residuals of pr_t after regressing on pd_t strongly predicts return at one-year horizon. Column (7) shows that pr_t largely subsumes the return predictive power of pd_t (as a reminder, ϵ_t^{pd} is the residual from regressing pd_t on pr_t). In contrast to the U.S. results, pd_t now carries some distinct information on future returns.

Cross-country comovement in the expected return. We introduce time fixed effect in Column (2) that absorbs a global factor in the returns of each country. Return predictability disappears, meaning that the return predictive power of pr_t mainly comes from the information it contains regarding the global factor. Note that pr_t is constructed country-by-country. This finding suggests that the expected return across countries comoves, which is in line with the literature on global market integration (Miranda-Agrippino and Rey (2015)). And in Column (4), any return predictive power of pd_t also disappears once the global factor is absorbed by the time fixed effect. A similar result holds in Column (6) for pr_t's residuals.

Figure 8 in the appendix shows the time series of the first three principal components of pr_t in these countries, which together account for more than 80% of the variation. The first principal component (48% of variation) exhibits spikes at the onsets of the global financial crisis and the European sovereign debt crisis, suggesting that a major part of the global comovement of expected stock return comes from crisis periods.

[16] Note that the Hodrick (1992) standard error in Table 2 is not based on such non-overlapping regression. We corrected the standard error of predictive coefficient of overlapping regression following the calculation in Hodrick (1992) who show that under certain assumptions the corrected t-statistic of the overlapping regression equals the t-statistic of the non-overlapping reverse regression.

[17] Note that the adjusted R^2 from the non-overlapping regression of Equation (19) is not comparable to that of the overlapping regression in Table 2, because in Equation (19), we effectively forecast monthly return using the one-year average of predictor, even if the inference we draw from such regression is about the return prediction at the one-year horizon. Thus, we do not report the R^2 of non-overlapping regression.

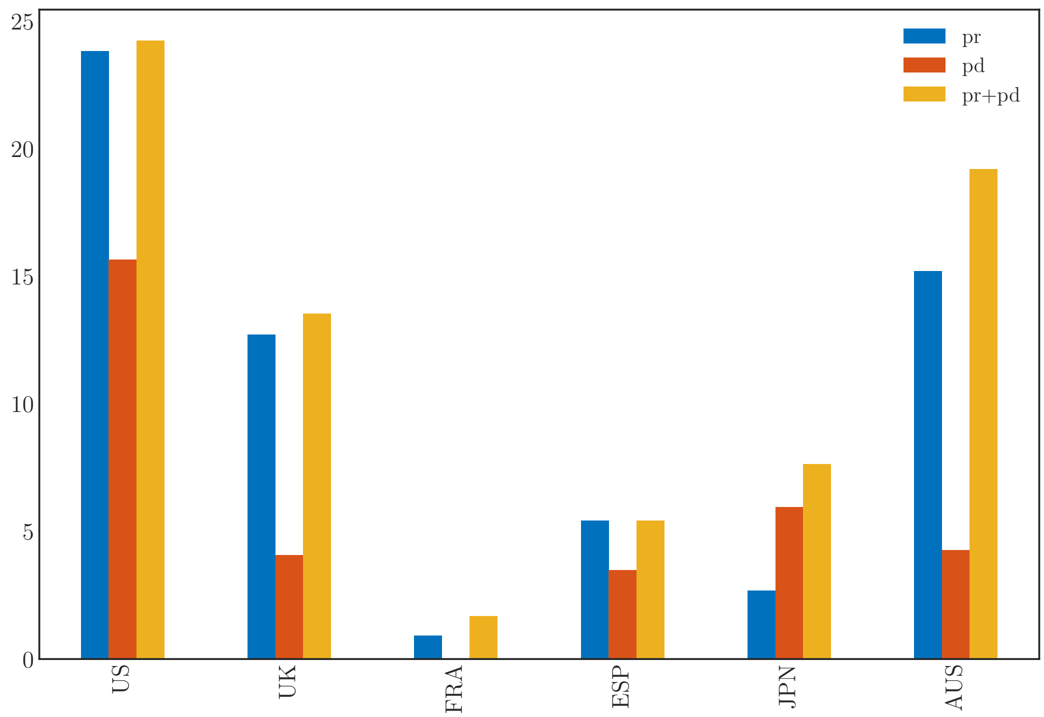


Figure 4: **Return Prediction across Countries.** This graph shows side-by-side the adjusted R^2s of three univariate predictive regressions, with pr_t, pd_t, and pr_t and pd_t together on the right-hand side respectively for each country. The left hand side variable is the total market return in the next twelve months.

Return predictability in each country. Figure 4 reports the adjusted R^2 from predictive regressions in each country using pr_t, pd_t, and pr_t and pd_t together on the right hand side. pr_t outperforms pd_t in all countries but Japan, and adding pd_t does not seem to bring extra return predictive power. Table 12 in the appendix reports the details of estimation results.

3 The Expected Return Risk

Roadmap. We study the risk of time-varying expected return using pr_t as the return forecasting variable. Shocks to pr_t are negatively and significantly priced in the cross-section (Table 5). Stocks that deliver high returns when pr_t is high (the expected market return is low) have low average returns. This exercise also serves as an economic test of pr_t as a return predictor – by the logic of ICAPM, shocks to agents' investment opportunity set are priced, so if pr_t proxies the expected return well, we should be able to identify the price of pr_t shocks in the cross-section. Next, we document the cyclicality of expected return by examining how pr_t responds to monetary policy shocks (Table 6) and by presenting the correlations between pr_t and macroeconomic and financial market conditions (Table 7).

3.1 The price of pr_t risk

Revisiting ICAPM. The expected market return, and more generally, agents' investment opportunity set, varies with pr_t. On the one hand, assets that have negative covariance with shocks to pr_t are desirable because they enable investors to profit from improved investment opportunities – a negative shock to pr_t is a positive shock to the expected market return (i.e., goods news on about future market returns). On the other hand, assets with positive pr_t-beta are desirable because they hedge against deterioration of investment opportunities. As shown by Campbell (1993) using the utility function of Epstein and Zin (1989), the hedging motive dominates, that is when shocks to pr_t are negatively priced in the cross section, when the coefficient of relative risk aversion is greater than one. Next, we estimate the price of pr risk in the cross-section of standard sorted portfolios.

We are the first to perform this asset pricing test in the literature of return predictability. If predictive power is not from spurious relations, shocks to the predictor change agents' investment opportunity set and should be priced. Here, we go beyond standard error correction, bias adjustments, and out-of-sample test. By estimating the price of pr_t risk, we test pr_t as a return predictor using the economic logic of ICAPM (Merton (1973)).

Estimating pr risk price. Our testing assets are the twenty-five Fama-French portfolios (sorted by size and book-to-market ratio), twenty-five momentum portfolios (sorted by size and prior returns), twenty-five investment portfolios (sorted by size and change in total assets), and twenty-five profitability portfolios (sorted by size and operating profitability). The data of monthly portfolio returns are from Kenneth R. French's website.[18] We consider this set of portfolios as good proxy for the U.S. investors' opportunity set.

The first step is to calculate the loadings of assets on shocks to pr_t. As noted by Pástor and Stambaugh (2003), an asset's beta should be defined with respect to shocks (innovations) instead of the level of a state variable, because the expected changes in the state variable and the expected asset return can be correlated, which contaminate beta measures. In our case, pr_t is very likely to correlate with expected asset returns, because it forecasts the market return and the expected asset returns are functions of the expected market return in CAPM or other equilibrium asset pricing models. We measure shocks to pr_t as the first difference. In the appendix (Table 13), we show that results based on AR(1) shocks are similar.

To estimate the price of pr_t shocks, we use two procedures. The first is the Fama-MacBeth method. The second one is GMM, which corrects potential biases in the Fama-MacBeth standard errors (Cochrane (2005)). The parameters are over-identified in GMM. For the weight matrix, we use the two-stage efficient weight matrix (Hayashi (2000)). In both

[18]http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Table 5: The Price of pr_t Risk

This table reports the price of market risk and pr risk estimated using Fama-MacBeth method and Generalized Method of Moments (GMM). We use the two-stage GMM estimator with efficient weight matrix. pr_t shock is measured by the first difference of pr_t. The full asset universe ("All") includes the twenty-five Fama-French portfolios (sorted by size and book-to-market ratio), ten momentum portfolios, ten investment portfolios, and ten profitability portfolios. We also estimate pr risk price using twenty-five value-size, momentum-size, investment-size, and profitability-size portfolios. The data of monthly portfolio returns are from Kenneth R. French's website. Each column corresponds to one set of assets. Each estimated price of risk is followed by the t-statistic in parenthesis. *, **, and *** denote 5%, 2%, and 1% level of statistical significance respectively. We also report mean absolute pricing error (MAPE) and R^2.

	All (1)	Fama-French 25 (2)	Momentum 25 (3)	Investment 25 (4)	Profitability 25 (5)
Fama-MacBeth					
Δpr_t	-0.252^{***}	-0.288^{***}	-0.367^{***}	-0.099	-0.193^{**}
	(-4.707)	(-2.803)	(-3.858)	(-1.047)	(-2.513)
$r_t - r_t^f$	0.009^{***}	0.009^{***}	0.009^{***}	0.010^{***}	0.009^{***}
	(4.025)	(3.893)	(3.772)	(4.214)	(4.023)
MAPE	0.189%	0.172%	0.220%	0.212%	0.228%
R^2	0.563	0.742	0.693	0.769	0.708
GMM					
Δpr_t	-0.359^{***}	-1.354^{***}	-0.296^{***}	-52.494	-0.069^{**}
	(-8.314)	(-2.907)	(-5.428)	(-0.059)	(-2.571)
$r_t - r_t^f$	0.010^{***}	0.010^{***}	0.009^{***}	-0.015	0.013^{***}
	(5.006)	(3.662)	(4.284)	(-0.036)	(5.979)
MAPE	0.091%	0.027%	0.089%	0.066%	0.147%
R^2	0.729	0.678	0.667	0.720	0.727

cases, the cross-sectional pricing equations exclude intercepts, and we include the market excess return as the other risk factor following the equilibrium condition of ICAPM.

Table 5 reports the results of cross-sectional estimations. Each column corresponds to a universe of assets. *, **, and *** denote 5%, 2%, and 1% level of statistical significance respectively. "All" refers to the universe that includes Fama-French twenty-five portfolios, ten momentum, ten investment, and ten profitability portfolios (a total of fifty-five assets). Column (2) to (5) correspond respectively to twenty-five double-sorted portfolios of book-to-market, momentum, investment and profitability interacting with size. The price of risk is reported for both pr_t shock (Δpr_t) and market excess return ($r_t - r_t^f$), followed by the t-statistic. We also report the mean absolute pricing error (MAPE) and R^2.

The price of pr_t risk is negative and statistically significant in the cross-section of all assets, size and book-to-market sorted portfolios, momentum portfolios, and profitability portfolios. The magnitude is similar across asset universes and economically significant. For example, one standard deviation difference in the pr_t beta of two assets corresponds to a difference of $0.252 \times 0.00685 \times 12 = 2.1\%$ in average return per annum. A significant price of

pr risk helps establish pr_t as a return predictor, i.e., a proxy for the expected market return.

Among size (SMB), value (HML), profitability (RMW), investment (CMA) and momentum factors, pr_t shocks have the highest correlation (-12.7%) with the momentum (Table 14 in the appendix), suggesting that the cross-sectional dispersion of pr_t beta are highest among momentum-sorted portfolios. This may explain why the estimated price of pr_t risk is higher and more precise in the momentum universe than the other sets of portfolios.

3.2 The cyclicality of expected return

The fact that shocks to pr_t are priced suggests the expected market return, proxied by pr_t, comoves with macroeconomic variables. The relation between macroeconomic conditions and the expected stock return has always been at the center of asset pricing research (Fama and French (1989); Ferson and Harvey (1991)). In particular, the impact of monetary policy on asset prices continues attracting great attention (Campbell, Pflueger, and Viceira (2015))). Bernanke and Kuttner (2005) find that an unanticipated cut in the Federal funds rate is associated with an increase in stock indexes, and based on the VAR approach proposed by Campbell and Ammer (1993), they show the largest part of stock price response is from changes in the expected return. More recently, Lucca and Moench (2015) show that sizable fractions of realized stock returns are concentrated in the twenty-four hours before the scheduled meetings of the Federal Open Market Committee (FOMC) in recent decades.

Next, using pr_t as a proxy of the expected return (discount rate) in the stock market, we examine how monetary policy affects stock prices through its impact on the discount rate and relate our results to the literature. We also document significant correlations between the expected returns from our predictive regression and macroeconomic variables. In sum, the expected stock market return comoves with monetary policy and the business cycle.

Monetary policy and the expected return. To examine the impact of monetary policy on stock prices, we construct four variables. We define a variable, FOMC Day, that equals one if the day has an FOMC meeting and zero otherwise. We also construct Pre-FOMC Day that equals one if the next day has an FOMC meeting and zero otherwise, and finally, Post-FOMC Day that equals one if it is the previous day has an FOMC meeting and zero otherwise. Finally, we use the monetary policy shocks from Nakamura and Steinsson (2017) to construct a variable, MP Shock, that equals the value of monetary policy shock on days of FOMC meetings and equals zero in non-FOMC days.[19]

[19] Monetary policy shock is calculated using a 30-minute window from 10 minutes before the FOMC announcement to 20 minutes after it. Data of the Federal funds futures is used to separate changes in the target funds rate into anticipated and unanticipated components. For earlier contributions, please refer to Cook and Hahn (1989), Kuttner (2001), and Cochrane and Piazzesi (2002) among others.

Table 6: Realized Return, Expected Return, and Monetary Policy Announcements

This table reports how daily returns and expected returns (proxied by pr_t) changes during FOMC announcement days and respond to monetary policy shocks. Monetary policy shock (MP Shock) is equal to the unanticipated changes in the Federal Funds rate from Nakamura and Steinsson (2017) on the days of FOMC meetings, and zero otherwise. "FOMC day" is the FOMC-day dummy variable. Pre-FOMC day is the pre-FOMC day dummy variable. Post-FOMC day is the post-FOMC day dummy variable. The sample period is 01 Jan 1988 to 31 Dec 2014 (the end of the sample of Nakamura and Steinsson (2017)). We regress contemporaneous returns, pr, the residuals from regressing pr_t on pd_t (i.e., ϵ_t^{pr}), and the residuals from regressing pd_t on pr_t (i.e, ϵ_t^{pd}) on these monetary policy variables. The results are reported in each column for each specification. *, **, and *** denote 5%, 2%, and 1% level of statistical significance respectively.

	r_t	r_t	pr_t	pd_t	ϵ_t^{pd}	ϵ_t^{pr}
MP Shock		-0.075^{***}	-1.326^{**}	-0.251	0.247	-0.951^{**}
		(-5.507)	(-2.329)	(-0.879)	(1.305)	(-2.519)
FOMC day	0.004^{***}	0.003^{***}	0.075	0.042^{*}	0.014	0.013
	(4.513)	(3.443)	(1.935)	(2.138)	(1.052)	(0.492)
Pre-FOMC day	0.001	0.001	0.099^{**}	0.042^{*}	0.002	0.040
	(0.702)	(0.704)	(2.553)	(2.191)	(0.174)	(1.559)
Post-FOMC day	-0.001	-0.001	0.101^{***}	0.045^{**}	0.007	0.035
	(-0.896)	(-0.898)	(2.628)	(2.327)	(0.508)	(1.358)
R^2	0.004	0.009	0.004	0.003	0.000	0.002
Obs	5600	5600	5600	5600	5600	5600

Table 6 reports the results of regressing contemporaneous return, pr_t, pd_t, the residuals of pd_t (w.r.t. pr_t), and the residuals of pr_t (w.r.t. pd_t). Column (1) confirms the results of Lucca and Moench (2015). The day of FOMC on average sees an average positive return of 40 basis points. While Lucca and Moench (2015) argue that most of the realized market returns are concentrated in the twenty-four hours before FOMC announcements (on average 49 basis points per FOMC), a big fraction of the twenty-four hours are on the FOMC day because the time of the release usually varies between 12:30 pm and 2:30 pm. We do not perform an intraday analysis because of the concern over intraday liquidity of S&P futures.

Column (2) of Table 6 shows a tightening of monetary policy (i.e., an increase in MP Shock) decreases stock market returns, in line with the evidence in Thorbecke (1997) and Bernanke and Kuttner (2005). After directly controlling for the monetary policy shock, the relation between stock return and FOMC day is weakened. There is a long tradition in understanding the contemporaneous response of stock price to monetary policy. Rozeff (1974) finds that a substantial fraction of stock return variation is related to monetary news.

Our main interest is on the regression of pr_t on the monetary policy variables because pr_t serves as a proxy for the expected stock return (i.e., the discount rate in the stock market). Column (3) of Table 6 shows a negative response of pr_t to monetary tightening, which translates into an increase in the expected return. In other words, an unanticipated

increase in the Federal Funds rate tends to raise the expected stock return.[20] Therefore, the decline of realized return can be largely attributed to the increase in discount rate. Column (6) delivers the same message using the residuals of pr_t from regressing pr_t on pd_t.

If we use the traditional price-dividend ratio as a proxy for the expected return, we shall not see any response to monetary policy shock (as shown in Column (4)), in particular, because the response of ϵ_t^{pr} to monetary policy shock is missed. Our new return predictor pr_t carries important information about how the expected return varies with monetary policy.

Finally, since the residuals from regressing pd_t on pr_t (i.e, ϵ_t^{pd}) strongly forecast dividend growth (Table 3), we regress it on monetary policy variables. We do not find significant relations. Therefore, monetary policy does not seem to affect the cash flow expectation.

Monetary policy has a strong impact on stock prices mainly through the discount rate – expansionary monetary policy tends to lower the discount rate, and thereby, raise the stock price, leading to higher contemporaneous return. Since the expected return is lower, the impact of monetary policy on stock price tends to revert over time.

Correlation with macroeconomic variables. Table 7 shows that the expected stock return comoves with variables that reflect the conditions of macroeconomy, financial markets, financial intermediaries, uncertainty, and sentiment.

The expected return is countercyclical. It is positively correlated with unemployment and negatively correlated with consumption growth, fixed investments, and GDP deflector, suggesting that a major fraction of variation in the expected return is from the business cycle. The expected return is also positively correlated with the term spread and weakly correlated with the default spread (Baa-Aaa) (Fama and French (1989)). The expected return comoves with *cay*, as suggested by Lettau and Ludvigson (2001), but pr_t outperforms *cay* in return forecasting (Figure 3), especially out of sample. The fact that many of these cyclical variables fail to predict return as strongly as pr_t does is likely because (1) we need better measurements (e.g., see Savov (2011) for consumption), (2) most are only available at lower (quarterly) frequencies, or (3) each of them reflects only a fraction of variation in the expected return but pr_t is a comprehensive measure (sufficient statistics).[21]

The expected return exhibits strong negative correlation with broker-dealer leverage (Adrian and Shin (2010); Adrian, Moench, and Shin (2013)). This indicates that when

[20] Our results are related to Patelis (1997) who documents some return predictive power of monetary policy variables.

[21] For example, Lamont (2000) find that the aggregate nonresidual investment does not forecast returns. He suggests that investment plans are more responsive to variation in risk premia. In contrast, our measure of expected return is highly correlated with aggregate nonresidual investment, suggesting that the findings of Lamont (2000) are likely to be biased by the noise between realized returns and expected returns. Pástor and Stambaugh (2009) propose a predictive system to address the imperfect correlation between expected returns and predictors.

Table 7: The Correlations between Macro Variables and the Expected Return

This table reports the correlation between the in-smaple fitted expected returns and macroecnomic variables. The variables are divided into four categories. 1) Macroeconomic: nominal GDP Growth, Industrial Production Growth ("IP Growth"), Chicago Fed National Activity Index ("CFNAI"), Unemployment Rate, Real Consumption Growth, Total Business Inventories, Nonresidential Fixed Investment (nominal), Residential Fixed Investment (nominal), and GDP Deflator are all from FRED database. 2) Financial: Term Spread and Default Spread ("Baa-Aaa") are from FRED. cay is from Lettau and Ludvigson (2001). 3) Intermediary: Broker/Dealer leverage ("B/D Leverage") is from Adrian, Etula, and Muir (2014); Broker/Dealer 1(5) year average CDS spreads ("B/D 1(5) Year Avg. CDS") is from Gilchrist and Zakrajšek (2012); ROA of banks ("ROA Banks") is from FRED. 4) Uncertainties: CBOE 1-month VIX index ("VIX") and Chauvet and Piger (2008)'s smoothed U.S. recession probabilities estimates for given month ("CP Recession") are from FRED; Economics policy uncertainties ("EPU") is from Baker, Bloom, and Davis (2016); Survey of Professional Forecasters recession probability estimates ("SPF Recession") is from the Philadelphia Fed. 5) Sentiments: Sentiment Index (both raw and orthogonalized against several macro variables), Number of IPOs ("IPO #") and close-end fund NAV discount ("Close-end Discount") are all from Baker and Wurgler (2006).

	$\hat{r}_{pr}$	p-value
Macroeconomic:		
GDP Growth	0.08	(0.38)
IP Growth	0.03	(0.48)
CFNAI	0.07	(0.18)
Unemployment	0.38	(0.00)
Cons. Growth	-0.43	(0.00)
Business Inventories	-0.08	(0.14)
Nonres. Fixed Investment	-0.47	(0.00)
Res. Fixed Investment	-0.31	(0.01)
GDP Deflator	-0.35	(0.00)
Financial:		
Term Spread	0.27	(0.00)
Baa-Aaa	0.09	(0.07)
cay	0.29	(0.00)
Intermediary:		
B/D Leverage	-0.55	(0.00)
B/D 1 Year Avg. CDS	0.21	(0.02)
B/D 5 Year Avg. CDS	0.32	(0.00)
ROA Banks	-0.43	(0.00)
Uncertainties:		
VIX	-0.25	(0.00)
EPU	0.17	(0.00)
CP Recession	-0.03	(0.47)
SPF Recession	0.13	(0.17)
Sentiments:		
Sentiment Index	-0.43	(0.00)
Sentiment Index (orth.)	-0.44	(0.00)
IPO #	0.10	(0.05)
Close-end Discount	0.32	(0.00)

dealer banks increase their leverage to acquire risky assets or to extend credit to hedge funds through prime brokerage services, the expected return tends to be low. The positive correlation between the expected return and broker-dealer CDS spread suggests that the net worth of financial intermediaries may also play a role in the variation of expected return (He and Krishnamurthy (2013); He, Kelly, and Manela (2017)). We also find that the expected return comoves with the profitability of commercial banks. Overall, the expected return is closely associated with conditions of the financial intermediation sector.

Interestingly, the expected return tends to be high when VIX is low. This finding has important implications on the dynamics of risk-return trade-off (Lettau and Ludvigson (2010)).[22] Our finding is related to Moreira and Muir (2017), who show that a trading strategy that scales up when the expected volatility declines tend to generate profits unexplained by common risk factors.[23] The expected return has a positive correlation with policy uncertainty (EPU), but the correlations with recession probabilities are not conclusive.

Finally, we show that the expected return tends to be low when sentiment is high. The sentiment index (raw and orthogonalized to macro factors) is from Baker and Wurgler (2006), together with IPO volume and closed-end fund discount (inversely related to sentiment).

4 Asymmetric Predictability

Roadmap. In this section, we study conditional return predictability. Specifically, we find that predictability is asymmetric – stronger following a down market (Table 8). The results are similar outside the United States (Figure 5). We evaluate related theories that imply an asymmetry in return predictability in Table 9.

4.1 Asymmetric return predictability: evidence

Conditional return prediction. We decompose pr_t into two components: (1) $I_{(r_{t-12,t}<0)} \times pr_t$, the interaction between pr_t and an down-market indicator that equals one if the cumulative market return in the past twelve months falls below the risk-free rate (i.e., the yield of twelve-month zero-coupon bond); (2) $I_{(r_{t-12,t}\geq 0)} \times pr_t$, the interaction between pr_t and the up-market indicator. The return forecasting model now becomes

$$r_{t,t+12} = \alpha + \beta_D I_{\{r_{t-12,t}<r^f_{t-12,t}\}} \times pr_t + \beta_U I_{\{r_{t-12,t}\geq r^f_{t-12,t}\}} \times pr_t + \beta_I I_{\{r_{t-12,t}<r^f_{t-12,t}\}} + \epsilon_{t,t+12}. \quad (20)$$

[22] VIX may not capture the risk of change in investment opportunity set, which can be an important component of risk (Guo and Whitelaw (2006)).

[23] A similar risk-return trade-off manifests itself in the cross-section of stocks, as shown by the profitability of strategies that explore low-risk anomalies, such as idiosyncratic volatility (Ang et al. (2009)), risk parity (Asness, Frazzini, and Pedersen (2012)), and betting against beta (Frazzini and Pedersen (2014)).

Table 8: Conditional Return Prediction

This table reports the results of conditional return prediction. The left-hand side variable of the regression is the return of S&P 500 index in the next twelve months. We run the regression monthly. Column (1) reports the results of the specification of Equation (20). On the right-hand side is the interaction between pr_t and the down-market indicator (equal to one if the past twelve-month market return is below risk-free rate), the interaction between pr_t and the up-market indicator, the down-market indicator, and the intercept (omitted in the table). Column (4) reports the results of the specification of Equation (21). On the right-hand side are pr_t, the interaction between pr_t and the past twelve-month market excess return, the past-twelve month market excess return, and the intercept (omitted in the table). The specifications of Column (2) and (5) has only the down-market indicator and the past twelve-month market excess return on the right-hand side respectively. The specifications of Column (3) and (6) adds pr_t to Column (2) and (5) respectively. For each specification, the β estimate is shown followed by Newey and West (1987) and Hodrick (1992) t-statistics, and the adjusted R^2 is reported in the last row.

		(1)	(2)	(3)	(4)	(5)	(6)
$I_{\{r_{t-12,t}<r^f_{t-12,t}\}} \times pr_t$		-0.180					
	Newey-West t	(-3.810)					
	Hodrick t	[-2.977]					
$I_{\{r_{t-12,t}\geq r^f_{t-12,t}\}} \times pr_t$		-0.108					
		(-2.981)					
		[-1.751]					
$I_{\{r_{t-12,t}<r^f_{t-12,t}\}}$		0.257	-0.038	-0.031			
		(1.000)	(-0.750)	(-0.901)			
		[0.987]	[-0.670]	[-0.558]			
pr_t				-0.137	-0.140		-0.137
				(-5.005)	(-5.496)		(-4.949)
				[-2.735]	[-2.776]		[-2.739]
$\left(r_{t-12,t} - r^f_{t-12,t}\right) \times pr_t$					0.269		
					(1.462)		
					[1.099]		
$r_{t-12,t} - r^f_{t-12,t}$					-1.037	0.083	0.065
					(-1.358)	(0.680)	(0.732)
					[-1.080]	[0.545]	[0.432]
R^2		0.261	0.012	0.246	0.264	0.008	0.243

Thus, the return predictive power of pr_t following a down market is reflected in β_D, and the return predictive power following an up market is captured by β_U.

Column (1) of Table 8 reports the regression results. Following a down market, pr_t strongly predicts the market return at one-year horizon. The predictive power is much weaker following an up market, i.e., when the market outperforms the risk-free benchmark. In fact, β_D is almost twice β_U in both magnitude and the t-statistic. The midpoint between β_D and β_U is very close to the coefficient of pr_t as a univariate predictor. This decomposition by the previous market condition reveals a sharp asymmetry in return predictability.

Column (2) and (3) of Table 8 show that the down-market indicator itself does not predict future returns or together with pr_t. When both the down-market indicator and pr_t

are used as predictors, the predictive coefficient on pr_t is almost identical to the predictive coefficient in the univariate regression, and the t-statistics and R^2s are almost identical.

Column (4) of Table 8 reports the results of an alternative specification,

$$r_{t,t+12} = \alpha + \beta pr_t + \rho_0 \left(r_{t-12,t} - r^f_{t-12,t}\right) + \rho_1 \left(r_{t-12,t} - r^f_{t-12,t}\right) \times pr_t + \epsilon_{t,t+12}. \quad (21)$$

Adding the interaction term and the previous market excess return only changes the predictive coefficient of pr_t by very little (in comparison with Table 2), but makes the coefficient more statistically significant. Column (6) shows that adding the past market excess return itself also does not change the predictive coefficient of pr_t by much, and the previous market excess return does not forecast future return.

Time series momentum and reversal. The regression of Equation (21) also shows that return autocorrelation depends pr_t. This is related to the studies on return autocorrelation (Fama and French (1988); Poterba and Summers (1988)) that find positive return autocorrelations at monthly and shorter horizons, and negative autocorrelations at annual and longer horizons. However, the evidence is not without debate (Kim, Nelson, and Startz (1991)).

Unconditional return autocorrelation is not significant at one-year horizon in Column (5). But as suggested by Column (1) and (4) of Table 8, the relation between past and future returns is a function of pr_t. As shown in Column (4), the return autocorrelation coefficient is a function of pr_t, i.e., $\rho_0 + \rho_1 pr_t$. With the mean of pr_t equal to 3.992 (Table 1), the average of return autocorrelation coefficient is only 0.037. When pr_t is one-standard deviation above its mean, the autocorrelation coefficient increases from 0.037 to 0.180, exhibiting momentum. When pr_t is one-standard deviation below its mean, the autocorrelation coefficient is -0.106, exhibiting reversal. Campbell, Grossman, and Wang (1993) find that daily return autocorrelation depends on volume. Huang, Jiang, Tu, and Zhou (2017) find that one-year autocorrelation of market return differs in good and bad times.[24] Our results suggest that one-year autocorrelation depends on the relative valuation of long- vs. short-term dividends.

Conditional predictability outside the United States. Figure 5 reports the results of conditional prediction (regression of Equation (20)) for different countries. The candle graph shows the estimates of β_D (red) and β_U (blue) with the one Hodrick (1992) standard error band. It is clear that except Japan, the return predictive power of pr_t is more prominent following a down market. Table 12 in the appendix reports the details of estimation results.

Our finding of asymmetric return predictability is related to the evidence on stronger return predictive power of other variables (e.g., price-dividend ratio) during economic down-

[24] Moskowitz, Ooi, and Pedersen (2012) study time-series momentum in futures market.

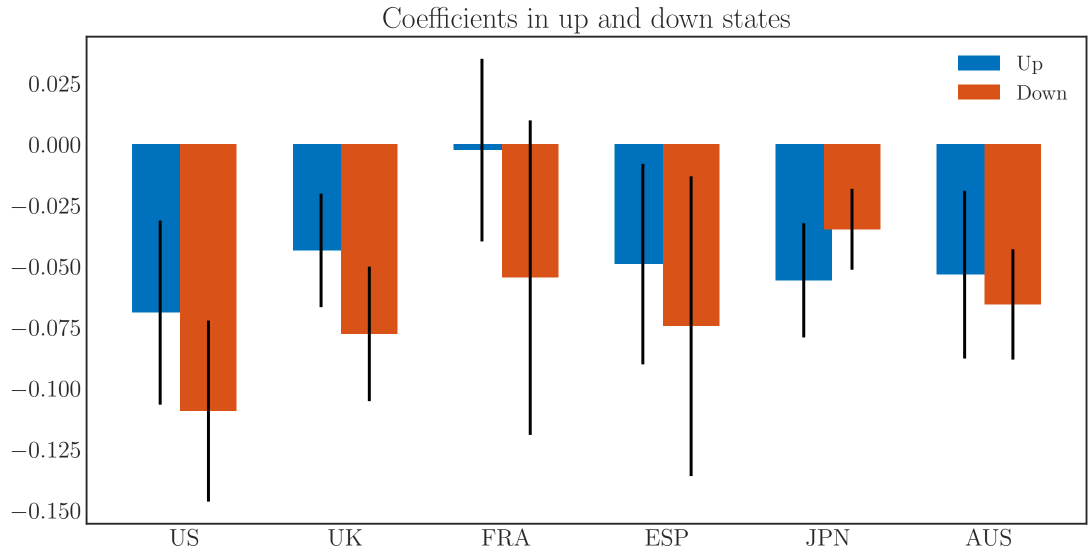


Figure 5: **Conditional predictability across countries.** Figure 5 reports the results of conditional prediction (regression of Equation (20)) for different countries. The candle graph shows the estimates of β_D (red) and β_U (blue) together with the one Hodrick (1992) standard error band.

turns.[25] Henkel, Martin, and Nardari (2011) show that the return predictive power of price-dividend ratio and short rate (Ang and Bekaert (2007)) appear non-existent during business cycle expansions but sizable during contractions. Considering a combination of predictors, Rapach, Strauss, and Zhou (2010) find that during recessions, the return is more predictable. Dangl and Halling (2012) propose a dynamic prediction model with a time-varying coefficient to account for conditional predictability. Farmer, Schmidt, and Timmermann (2018) find return predictability is concentrated in several relatively short periods. Our choice of conditioning variable is also motivated by related theoretical models that exhibit asymmetry of return predictability. Next, we shall discuss the related theories.

4.2 Asymmetric return predictability: related theories

Two theories based on financial intermediation (He and Krishnamurthy (2013)) and behavioral bias (Barberis, Huang, and Santos (2001)) produce asymmetry in return predictability.

Intermediary asset pricing. Panel A of Figure 12 in Appendix II.3 is from He and Krishnamurthy (2013). It plots the risk premium against the state variable, which is the share of aggregate wealth that belongs to financial intermediaries. He and Krishnamurthy (2013) model intermediaries as agents with exclusive access to risky assets. Intermediaries

[25] Cujean and Hasler (2017) build an equilibrium model with counter-cyclical investors' disagreement to explain why stock return predictability is concentrated in bad times.

manage wealth for the rest of economy ("households"), but the delegation capacity is linked to intermediaries' own wealth due to a typical principal-agent problem. When intermediaries are rich, their delegation capacity is sufficient to satisfy the needs of the household, and risk premium varies with the *aggregate* wealth of the economy, showing little variation ("unconstrained region"). When intermediaries are poor, the capacity constraint binds, and the risk premium varies with the wealth of intermediaries, fluctuating widely. The asymmetry of risk premium variation implies the asymmetry of return predictability. Related, our down-market indicator, which spans one year, is motivated by the observation by Benartzi and Thaler (1995) that investors tend to evaluate fund performances annually because they receive most comprehensive fund reports once a year.

Prospect theory. Panel B of Figure 12 is from Barberis, Huang, and Santos (2001). Their model is built upon two ideas. First, investors are subject to loss aversion (Kahneman and Tversky (1979)). Losses and gains from the stock market are defined with the risk-free rate as a benchmark. Second, how loss-averse investors are, depends on their prior gains and losses (Thaler and Johnson (1990)) against certain reference point. As they explain in the paper: "after a prior gain, the agent becomes less loss averse: the prior gains will cushion any subsequent loss, making it more bearable. Conversely, after a prior loss, the agent becomes more loss averse: after being burned by the initial loss, he is more sensitive to additional setbacks." Panel B of Figure 12 shows that the expected return barely moves when z_t is below one. z_t measures the prior losses (if < 1) or gains (if > 1) against a historical benchmark. Therefore, only under prior losses, the expected return exhibits large variation, which in turn implies asymmetric return predictability.

Evaluating related theories. Table 9 compares our results of conditional return prediction with the results from empirical specifications suggested by the theories of He and Krishnamurthy (2013) and Barberis, Huang, and Santos (2001). For each model, we construct negative and positive indicator variable by comparing the value of conditioning variable with a benchmark value, that is zero for our past excess return, average (i.e., $\overline{\eta}$) for the intermediary capital ratio of He, Kelly, and Manela (2017) (an empirical study of He and Krishnamurthy (2013)), and one for z_t of Barberis, Huang, and Santos (2001). Column (1) and (4) repeat the results in Column (1) and (2) in Table 8.

The empirical specifications suggested by both He and Krishnamurthy (2013) and Barberis, Huang, and Santos (2001) produce results that are very similar to those of our model with past excess return as the conditioning variable. The predictive coefficient on the interaction between the negative indicator and pr_t is twice as large as the coefficient of the interaction between the positive indicator and pr_t, and all three specifications in Column (1),

Table 9: Evaluating Related Theories

This table reports the results of annual return prediction conditioning on three different variables: the past twelve-month market excess return, the intermediary net worth η_t in He, Kelly, and Manela (2017), and the z_t constructed following the model of Barberis, Huang, and Santos (2001). We construct negative and positive indicator variables by comparing the three conditioning variable with zero, average, and one (as suggested by the theory) respectively. The specifications of Column (1) to (3) have the interaction terms between indicator variables and pr_t, the negative indicator variable, and the intercept (omitted in the table). The specifications of Column (4) to (6) have the negative indicator variables and the intercept (omitted in the table). For each right-hand side variable, the coefficient estimate is shown followed by Newey and West (1987) and Hodrick (1992) t-statistics. For each specification, the adjusted R^2 is reported in the last row. Note that we use η_t constructed by He, Kelly, and Manela (2017) whose sample ends in 2012.

		$r_{t-12,t} - r^f_{t-12,t}$	$\eta_t - \overline{\eta}_t$	$z - 1$	$r_{t-12,t} - r^f_{t-12,t}$	$\eta_t - \overline{\eta}_t$	$z - 1$
Negative × pr_t		-0.180	-0.247	-0.195			
	Newey-West t	(-3.810)	(-2.340)	(-5.221)			
	Hodrick t	[-2.977]	[-2.101]	[-2.990]			
Positive × pr_t		-0.108	-0.152	-0.101			
		(-2.981)	(-4.549)	(-3.133)			
		[-1.751]	[-2.111]	[-1.714]			
Negative		0.257	0.279	0.382	-0.038	0.023	0.020
		(1.000)	(0.697)	(2.240)	(-0.750)	(0.473)	(0.541)
		[0.987]	[0.549]	[1.296]	[-0.670]	[0.455]	[0.441]
R^2		0.261	0.314	0.265	0.012	0.006	0.004

(2), and (3) have adjusted R^2 of similar magnitude. However, both theories imply that the negative indicator itself should also predict returns, which is not the case in data as shown in Column (5) and (6). This again suggests that our return predictor pr_t is likely to be a sufficient statistic that reflects various economic sources of expected return variation.

5 Conclusion

We find strong evidence of stock return and cash-flow predictability. The ratio of dividend strip prices (pr_t) predicts stock market return, and the pr_t-adjusted price-to-dividend ratio predicts future dividends. Shocks to pr_t are priced in the cross-section as in ICAPM, and the expected return (proxied by pr_t) responds strongly to monetary policy shocks and co-moves with the conditions of the macroeconomy and financial markets. Moreover, return predictability is asymmetric – stronger following market downturns.

In asset pricing literature, return and cash-flow predictability are closely tied to the decomposition of stock market volatility (Campbell and Shiller (1988)). Our findings can be incorporated into this paradigm to understand the relative importance of news on the discount rate and cash flow. Moreover, pr_t can be constructed for individual firms with options

data and used to study the cross-sectional variation of stock returns and the decomposition of firm-level volatility into discount-rate and cash-flow news (Vuolteenaho (2002)).

The price ratios can be constructed in every asset class as long as futures, forwards or options data are available, and for assets without explicitly defined dividends such as foreign currencies, commodities, and cryptocurrencies. Whether the price ratio predicts returns of other types assets, and if so, how it changes our understanding of the discount-rate dynamics in these asset classes, are interesting directions for future research.

Appendix I: Derivation

I.1 Derivation of the state space model

We start with the Campbell-Shiller decomposition of price-dividend ratio

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

By law of iterated expectation, we can replace Δd_{t+j} and r_{t+j} with their time $t + j - 1$ expectations:

$$\begin{aligned} v_t =& \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1} \mathbb{E}_t \left[g_{t+j-1} - \mu_{t+j-1} \right] \\ =& \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1} \mathbb{E}_t \left[(\delta_0 + \boldsymbol{\delta}' F_{t+j-1}) - (\gamma_0 + \boldsymbol{\gamma}' F_{t+j-1}) \right] \end{aligned}$$

Define ϕ_0 as $\frac{\kappa+\delta_0-\gamma_0}{1-\rho}$, and stack the factor coefficients into $\boldsymbol{\psi} = (\boldsymbol{\delta}', \boldsymbol{\gamma}')$. Denote the row vector $(1, -1)$ as $\boldsymbol{\iota}$. We can rewrite the equation

$$\begin{aligned} v_t =& \phi_0 + \sum_{j=1}^{\infty} \rho^{j-1} \boldsymbol{\iota} \boldsymbol{\psi}' \mathbb{E}_t \left[F_{t+j} \right] \\ =& \phi_0 + \sum_{j=1}^{\infty} \rho^{j-1} \boldsymbol{\iota} \boldsymbol{\psi}' \boldsymbol{\Lambda}^j F_t \\ =& \phi_0 + \boldsymbol{\iota} \boldsymbol{\psi}' \left(\sum_{j=1}^{\infty} \rho^{j-1} \boldsymbol{\Lambda}^j \right) F_t \\ =& \phi_0 + \boldsymbol{\iota} \boldsymbol{\psi}' \left(1 - \rho \boldsymbol{\Lambda} \right)^{-1} F_t. \end{aligned}$$

Define $\boldsymbol{\phi}'$ as $\boldsymbol{\iota} \boldsymbol{\psi}' \left(1 - \rho \boldsymbol{\Lambda} \right)^{-1}$. We have the factor decomposition of price-dividend ratio.

I.2 Deriving the Sharpe ratio of market-timing strategy

Following Campbell and Thompson (2008), we assume that the excess return can be decomposed as follows:

$$r_{t+1} = \mu + x_t + \varepsilon_{t+1}$$

where μ is unconditional meanm the predictor x_t has mean 0 and variance σ_x^2, independent from the error term ε_{t+1}. For simplicity, we assume that the mean-variance investor has

relative risk aversion coefficient $\gamma = 1$. When using x_t to time the market, the investor allocates

$$\alpha_t = \frac{\mu + x_t}{\sigma_\varepsilon^2}$$

to the risky asset and on average earns excess return of

$$\mathbb{E}\left(\alpha_t r_{t+1}\right) = \mathbb{E}\left(\frac{\left(\mu + x_t\right)\left(\mu + x_t + \varepsilon_{t+1}\right)}{\sigma_\varepsilon^2}\right) = \frac{\mu^2 + \sigma_x^2}{\sigma_\varepsilon^2}$$

The variance of market-timing strategy is

$$\mathrm{Var}\left(\alpha_t r_{t+1}\right) = \mathrm{Var}\left[\frac{\left(\mu + x_t\right)\left(\mu + x_t + \varepsilon_{t+1}\right)}{\sigma_\varepsilon^2}\right]$$

The (squared) market-timing Sharpe ratio s_1^2 can be written as

$$s_1^2 = \frac{\left[\mathbb{E}\left(\alpha_t r_{t+1}\right)\right]^2}{\mathrm{Var}\left(\alpha_t r_{t+1}\right)} = \frac{\frac{\mu^2+\sigma_x^2}{\sigma_\varepsilon^2}}{\mathrm{Var}\left(\alpha_t r_{t+1}\right)} = A \cdot \frac{\mu^2 + \sigma_x^2}{\sigma_\varepsilon^2}$$

where A is a constant that depends on $Var\left[\left(\mu + x_t\right)\left(\mu + x_t + \varepsilon_{t+1}\right)\right]$.

Given the buy-and-hold Sharpe ratio s_0,

$$s_0^2 = \frac{\mu^2}{\sigma_x^2 + \sigma_\varepsilon^2}$$

and the predictive regression R^2,

$$R^2 = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\varepsilon^2},$$

we obtain the relationship between the buy-and-hold and market-timing Sharpe ratios as

$$s_1^2 = A \cdot \frac{\mu^2 + \sigma_x^2}{\sigma_\varepsilon^2} = A \cdot \frac{\mu^2 + \sigma_x^2}{\left(\sigma_x^2 + \sigma_\varepsilon^2\right)\left(1 - R^2\right)} = A \cdot \frac{s_0^2 + R^2}{1 - R^2}$$

When the predictor has no predictive power, we know that $R^2 = 0$ and $s_0 = s_1$. We therefore pin down the constant $A = 1$ and obtain

$$s_1 = \sqrt{\frac{s_0^2 + R^2}{1 - R^2}}. \tag{22}$$

I.3 Solving the structural model

We follow directly Lettau and Wachter (2007) when solving the price of one-year dividend, so we do not repeat the derivation details here. For the price of all dividends, we first conjecture that the market price-dividend ratio is exponential-affine in the state variables, that is

$$pd_t = \ln(S_t/D_t) = A + Bx_t + Cz_t.$$

Next, we use the log-linearizion of Campbell and Shiller (1988), i.e.,

$$r_{t+1} = \kappa_0 + \kappa_1 pd_{t+1} - pd_t + \Delta d_{t+1},$$

and substitute this log market return into the no-arbitrage condition

$$\mathbb{E}_t \left[M_{t+1} \exp\{r_{t+1}\} \right] = 1.$$

to obtain

$$\mathbb{E}_t \left[\exp\left\{ -r^f - \frac{1}{2}x_t^2 - x\varepsilon_{d,t+1} + \kappa_0 + \kappa_1 pd_{t+1} - pd_t + \Delta d_{t+1} \right\} \right] = 1,$$

where $\Delta d_{t+1} = g + z_t + \sigma_d \varepsilon_{t+1}$ from Equation (11) and $\varepsilon_{d,t+1} = \sigma_d / \|\sigma_d\| \varepsilon_{t+1}$ as in Lettau and Wachter (2007). Note that pd_{t+1} can also be written as a linear combination of state variables at time t and $t+1$ shocks because x_{t+1} and z_{t+1} are given by Equation (13) and (12) respectively. Therefore, we can take all time-t measurable terms outside of the expectation and only leave $t+1$ shocks in it:

$$\begin{aligned} &\exp\left\{ -r^f - \frac{1}{2}x_t^2 + \kappa_0 - pd_t + g + z_t + \kappa_1 A + \kappa_1 B (1 - \phi_x) \bar{x} + \kappa_1 B \phi_x x_t + \kappa_1 C \phi_z z_t \right\} \\ &\mathbb{E}_t \left[\exp\left\{ -x_t \varepsilon_{d,t+1} + \kappa_1 B \sigma_x \varepsilon_{t+1} + \kappa_1 C \sigma_z \varepsilon_{t+1} + \sigma_d \varepsilon_{t+1} \right\} \right] = 1. \end{aligned} \tag{23}$$

Using the Gaussian moment-generating function, we rewrite the term within expectation as

$$\begin{aligned} &\mathbb{E}_t \left[\exp\left\{ -x_t \varepsilon_{d,t+1} + \kappa_1 B \sigma_x \varepsilon_{t+1} + \kappa_1 C \sigma_z \varepsilon_{t+1} + \sigma_d \varepsilon_{t+1} \right\} \right] \\ &= \exp\left\{ \frac{1}{2} (\kappa_1 B \sigma_x + \kappa_1 C \sigma_z + \sigma_d)(\kappa_1 B \sigma_x + \kappa_1 C \sigma_z + \sigma_d)' \right. \\ &\qquad \left. + (\kappa_1 B \sigma_x + \kappa_1 C \sigma_z + \sigma_d) \frac{\sigma_d'}{\|\sigma_d\|} x_t + \frac{1}{2} x_t^2 \right\} \end{aligned}$$

Equation (23) holds only if the coefficient on x_t and z_t are zero, so we have the Collect

coefficients of x_t equal to

$$-B + \kappa_1 B\phi_x + (\kappa_1 B\sigma_x + \kappa_1 C\sigma_z + \sigma_d)\frac{\sigma_d'}{\|\sigma_d\|} = 0,$$

and the coefficient on z_t equal to

$$-C + 1 + \kappa_1\phi_z C = 0.$$

From these two equations, we solve

$$B = \frac{\frac{\sigma_d\sigma_z'}{\|\sigma_d\|}\kappa_1 C + \frac{\sigma_d\sigma_d'}{\|\sigma_d\|}}{1 - \kappa_1\phi_x - \frac{\sigma_d\sigma_x'}{\|\sigma_d\|}\kappa_1},$$

and

$$C = \frac{1}{1-\kappa_1\phi_z}.$$

Note that x_t^2 is canceled out. By setting all the constant terms in the exponential equal to zero, we can solve the constant A in our conjecture of pd_t. Hence, we confirm the conjecture.

Next, we solve the expected market return.

$$\begin{aligned}
\mathbb{E}_t[r_{t+1}] =& \kappa_0 + \kappa_1 E_t[pd_{t+1}] - pd_t + E_t[\Delta d_{t+1}] \\
=& \kappa_0 + \kappa_1 A + \kappa_1 B E_t[x_{t+1}] + \kappa_1 C E_t[z_{t+1}] - A - Bx_t - Cz_t + g + z_t \\
=& \kappa_0 + \kappa_1 A + \kappa_1 B(1-\phi_x)\overline{x} + \kappa_1 B\phi_x x_t + \kappa_1 C\phi_z z_t - A - Bx_t - Cz_t + g + z_t \\
=& [\kappa_0 + (\kappa_1 - 1)A + g + \kappa_1 B(1-\phi_x)\overline{x}] + B(\kappa_1\phi_x - 1)x_t + [C(\kappa_1\phi_z - 1) + 1]z_t \\
=& [\kappa_0 + (\kappa_1 - 1)A + g + \kappa_1 B(1-\phi_x)\overline{x}] + B(\kappa_1\phi_x - 1)x_t.
\end{aligned}$$

Note that the coefficient on z_t equals zero because $C = \frac{1}{1-\kappa_1\phi_z}$.

Appendix II: Additional Results

II.1 Alternative out-of-sample sample splits

In the main text, we report out-of-sample forecasting tests based on a 1988 sample split date, but recent forecast literature suggests that sample splits themselves can be data-mined (see Hansen and Timmermann (2012) and Rossi and Inoue (2012)). To demonstrate the robustness of out-of-sample forecasts to alternative sample splits, Figure 6 plots out-of-sample annual return predictive R^2 as a function of the sample split for a variety of predictors. We consider a sample split as early as 1993. The latest split we consider is Jun 2012 (5-year

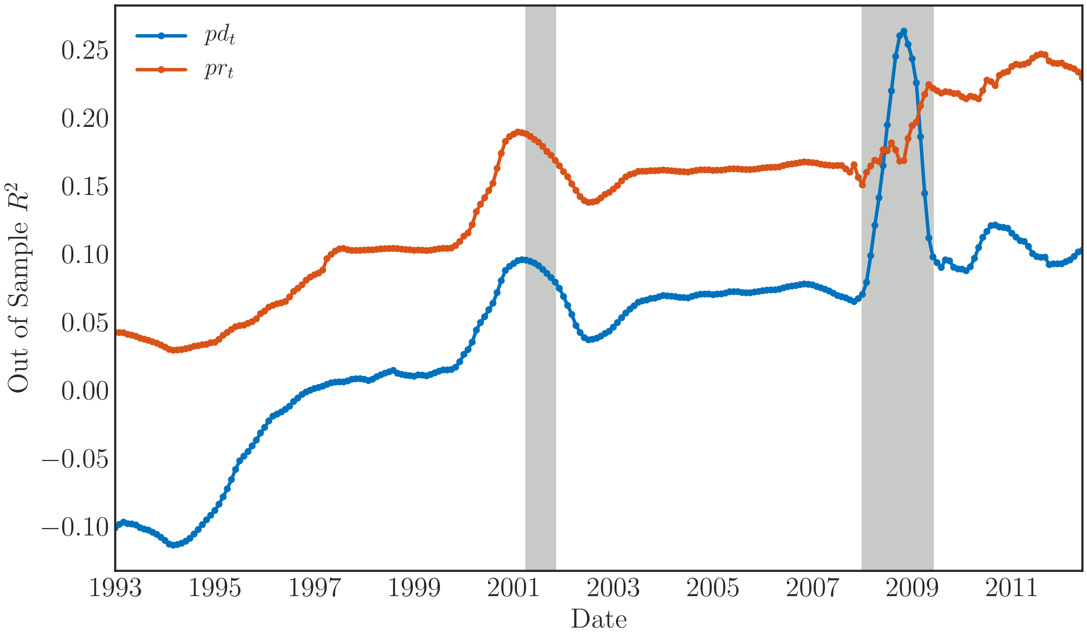


Figure 6: **Out-of-sample R^2 by Sample Split Date.** This graph reports out-of-sample R^2 with different sample split dates of 1-year return prediction. The first and last out-of-sample split date are Jan 1993 and Jun 2012 respectively.

prior to the end of our sample), which uses a 24.5-year training sample.

For early sample splits, for example 1994, the training (i.e., estimation) sample is relatively short, so the precision of coefficient estimate is poor, which contributes to the low out-of-sample R^2 that we see in the early years. As the sample split date progresses, the estimation sample extends, and the evaluation sample starts to exclude more data from earlier dates in the calculation of out-of-sample R^2. Excluding the dotcom burst, i.e. out-of-sample split starting 2002 or later, leads to a relatively low R^2 for both pr_t and pd_t, suggesting that both predictors perform well during the dotcom burst. Using data starting from the 2007-09 crisis, pd_t delivers a higher out-of-sample R^2 than pr_t. The reason is that its denominator, i.e., the rolling sum of dividends, reacts to the crisis sluggishly, so the decrease of pd_t is larger than the decrease of pr_t throughout the crisis, coinciding with the slump of market return. After the financial crisis, pr_t outperforms pd_t out-of-sample.

II.2 Monthly return prediction

One-month return prediction. Table 10 reports the results of one-month return prediction. The predictive coefficient is large in magnitude and statistically significant. A decrease of pr_t by one standard deviation adds 0.53% to the expected monthly return (annualized to 6.55%). The out-of-sample R^2 of 0.9% implies a large improvement in investment perfor-

Table 10: One-month Return Prediction

This table reports the results of predictive regression (Equation (8)). The left-hand side variable is the return of S&P 500 index in the next month. We consider four the right-hand side variables (i.e., predictors), pr_t, pd_t, the residuals of pr_t after regressing on pd_t (ϵ_t^{pr}), and the residuals of pd_t after regressing on pr_t (ϵ_t^{pd}), and the results are reported in Column (1) to (4) respectively. The β estimate is shown followed by Hodrick (1992) t-statistic and the in-s1ample adjusted R^2. We run the regression monthly. Starting from December 1997, we form out-of-sample forecasts of return in the next month by estimating the regression with data only up to the current month, and use the forecasts to calculate out-of-sample R^2. ENC statistic (Clark and McCracken (2001)), and the p-value of CW statistic (Clark and West (2007)).

	pr_t	pd_t	ϵ_t^{pr}	ϵ_t^{pd}
β	-0.010	-0.015	-0.011	0.004
Hodrick t	[-2.197]	[-2.006]	[-0.959]	[0.215]
R^2	0.017	0.014	0.005	0.000
OOS R^2	0.009	0.007	-0.009	-0.012
$p(ENC)$	< 0.10	< 0.10	> 0.10	> 0.10
$p(CW)$	0.078	0.129	0.427	0.170

mance for an investor who rebalances portfolio monthly and uses pr_t to time the market. For a mean-variance investor, Campbell and Thompson (2008) show that in comparison with a buy-and-hold strategy, the proportional increase in the expected return from observing pr_t is $\left(\frac{R^2}{1-R^2}\right)\left(\frac{1+S^2}{S^2}\right)$, where R^2 is the out-of-sample R^2 and S^2 is the squared Sharpe ratio of stock returns. Given the monthly Sharpe ratio of 0.1570 (annualized to 0.544), a monthly out-of-sample R^2 of 0.9% implies a 36.5% proportional improvement of expected return.

The difference in return predictive power between pd_t and pr_t is smaller at one-month forecasting horizon than at one-year horizon. pd_t has an out-of-sample R^2 of 0.7% (Column (2) of Table 10), and the residual of pr_t after regressing on pd_t does not significantly forecast monthly return. This suggests that the additional return predictive power of pr_t beyond pd_t is mainly at longer horizons.

II.3 Additional figures

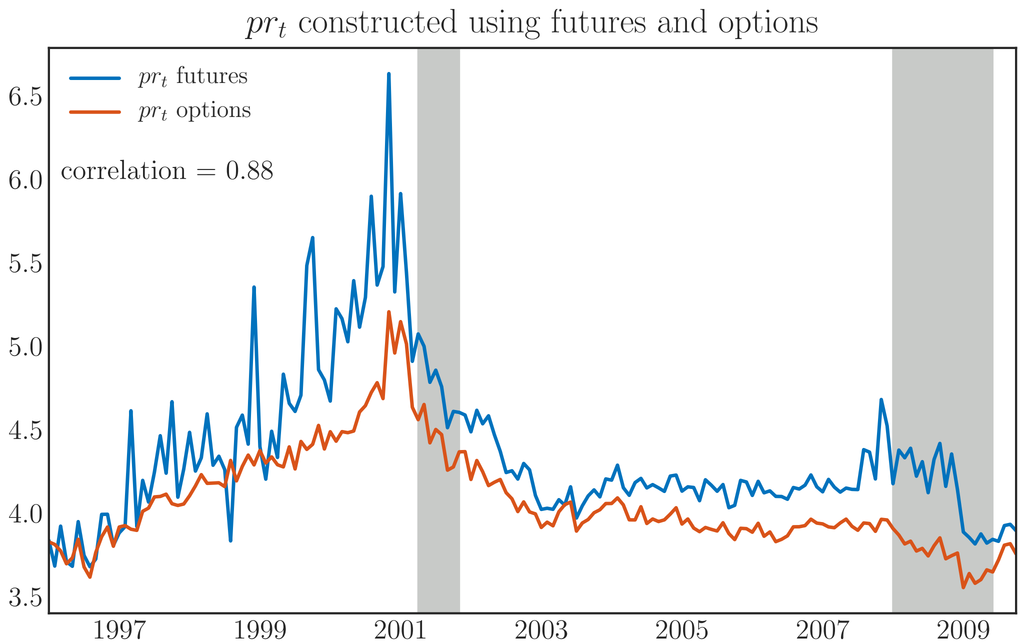


Figure 7: pr_t **from Futures and Option Data.** This graph reports pr_t constructed from futures and option data (from Binsbergen, Brandt, and Koijen (2012) from January 1996 and October 2009).

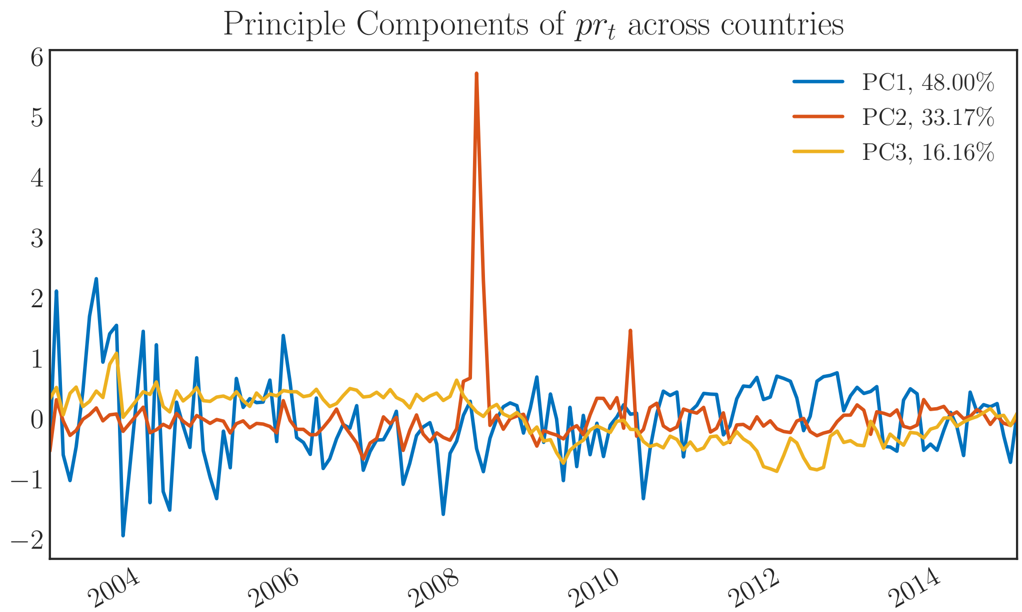


Figure 8: **Principal Components of pr_t.** This figure plots the first three principal components of pr_t in US, UK, France, Spain, Japan and Australia.

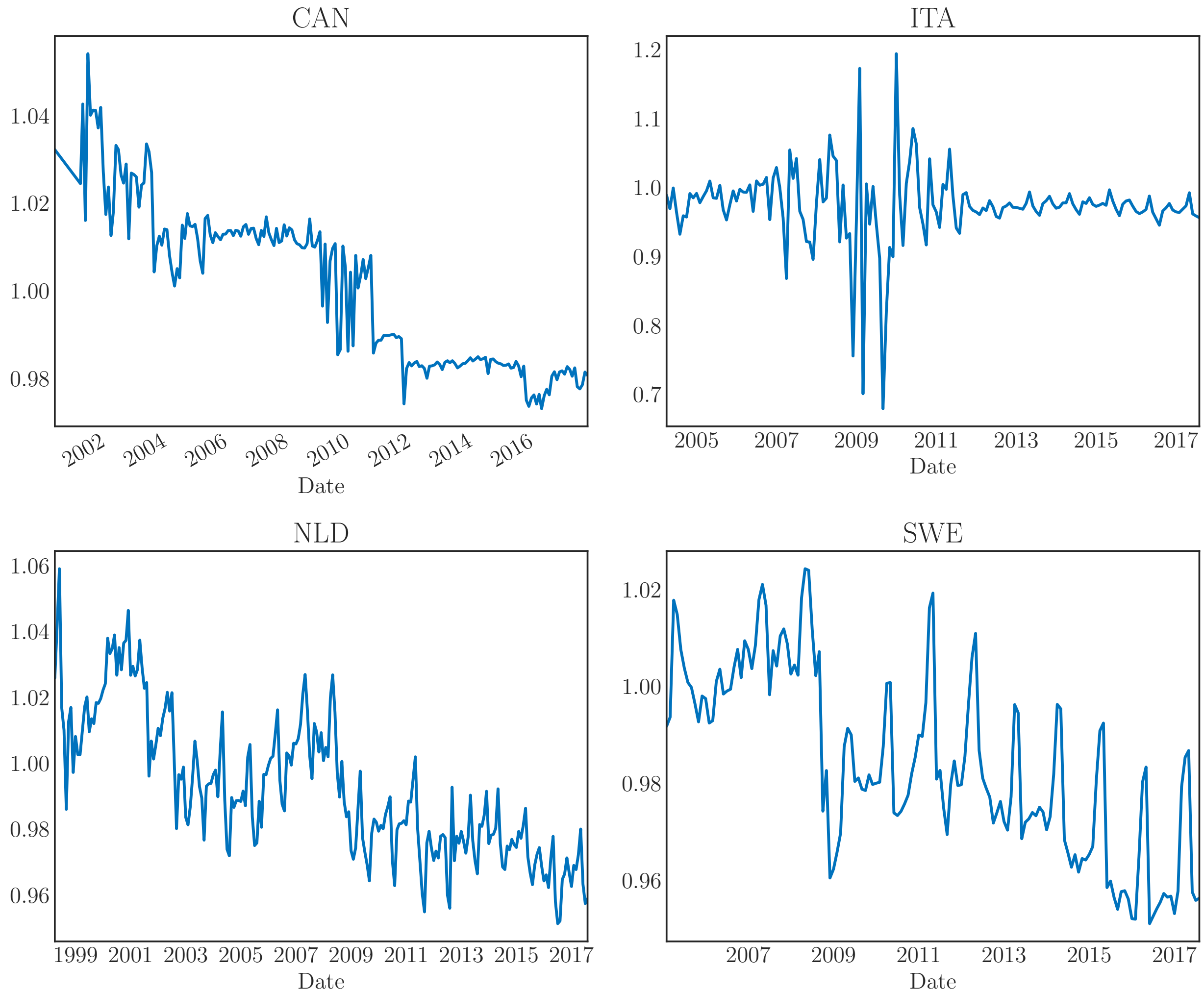
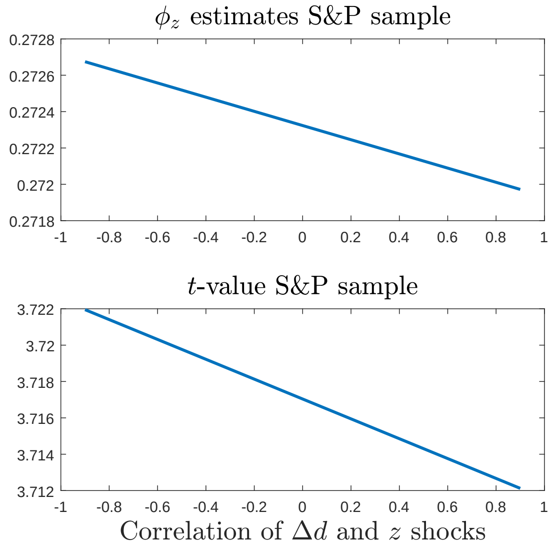
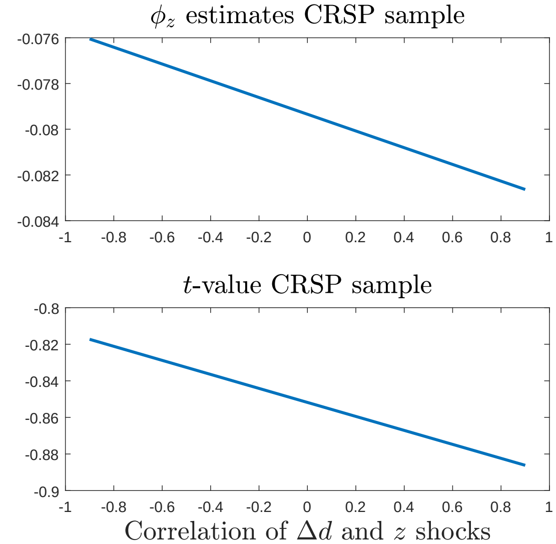


Figure 9: **Futures-to-spot ratio of international stock indices.** This graph plots 1-year futures-to-spot ratio of international stock indices. There are 4 countries, Canada, Italy, Netherlands, and Switzerland.



(a) S&P 500 dividend



(b) CRSP NYSE/NASDAQ/AMEX dividend

Figure 10: **ϕ_z estimates from state space model with correlated shocks.** This figure reports the expected dividend growth ϕ_z point estimates and t-value in an unrestricted state space model as in Section 2.4 with different correlations of Δd and z shocks. The correlations of Δd and z shocks range from -0.9 to 0.9 and the volatility of Δd shock is calibrated to the estimate $\hat{\sigma}_D$ from unrestricted state space model with uncorrelated shocks. Panel (a) uses annual dividend growth (non-overlapping) of S&P 500 index and Panel (b) uses annual dividend growth (non-overlapping) of CRSP NYSE/AMEX/NASDAQ Cap-Based index.

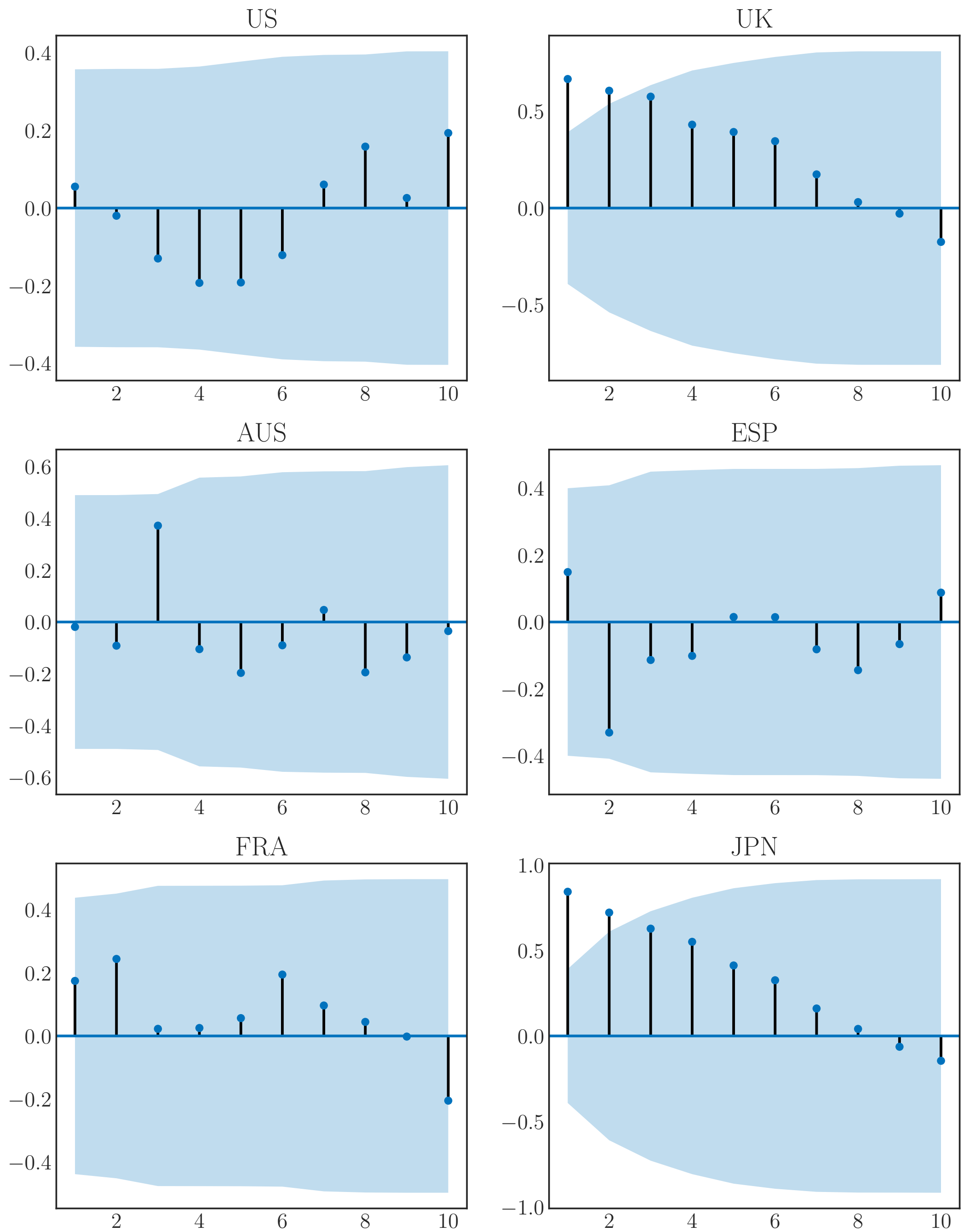


Figure 11: **Autocorrelations of cash-flow expectation, ϵ_t^{pd}, across countries.** This graph plots autocorrelations of ϵ_t^{pd} at lags from one to ten for the U.S., the U.K., Australia, Spain, France, and Japan, which constitute our international sample for return prediction.

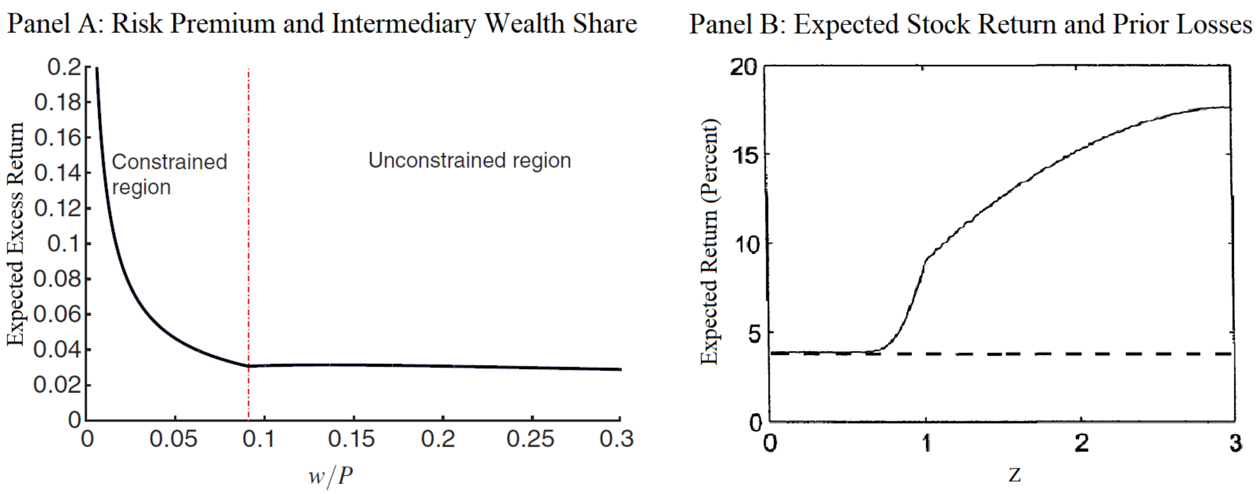


Figure 12: **Expected Return from Asset Pricing Theories.** Panel A is Figure 2 (Panel A) of He and Krishnamurthy (2013). The expected excess return of risky asset is plotted against intermediaries' share of aggregate wealth. A decline of w/P means that intermediaries become relatively undercapitalized due to losses. The dashed line splits the region where intermediaries are unconstrained in raising external funds, and the region where intermediaries are constrained in raising external funds because the principal-agent problem cannot be resolved under low net worth of intermediaries. Panel B is Figure VI (Panel A) of Barberis, Huang, and Santos (2001). The expected market return (in percent) is plotted against z_t that measures prior losses. High values of z_t mean that the representative investor has accumulated prior losses that increase risk aversion. The dashed line shows the constant risk-free rate.

II.4 Additional tables

Table 11: Correlations with other common return predictors

This table shows the correlation of alternative return predictors with both pr_t and pd_t from 1988 to 2016. Most alternative predictors are from Goyal and Welch (2007) that include the default yield spread (dfy), the inflation rate (infl), stock variance (svar), the cross-section premium (csp), the dividend payout ratio (de), the long-term yield (lty), the term spread (tms), the T-bill rate (tbl), the default return spread (dfr), the dividend yield (dy, log difference between current-period dividend and lagged S&P 500 index price), the long-term rate of return (ltr), the earnings-to-price ratio (ep), the book to market ratio (bm), the investment-to-capital ratio (ik), the net equity expansion ratio (ntis), the percent equity issuing ratio (eqis), and the consumption-wealth-income ratio (cay). SII is the short interests index from Rapach, Ringgenberg, and Zhou (2016) (1988-2014). kp is predictive facotr extracted from 100 book-to-market and size portfolios from Kelly and Pruitt (2013). SVIX is option-implied lower bound of 1-year equity premium from Martin (2017) (1996-2012). ZCB1Y is 1-year zero coupon bond yield from Fama-Bliss.

	pr	pd
pr	1.000	0.874
pd	0.874	1.000
bm	-0.790	-0.827
tbl	-0.173	-0.199
lty	-0.378	-0.393
ntis	-0.041	0.134
infl	-0.108	-0.073
ltr	0.005	-0.042
svar	0.161	-0.041
csp	0.355	0.427
cay	-0.384	-0.381
ik	0.616	0.631
ep	-0.554	-0.442
de	-0.247	-0.472
dfy	-0.067	-0.284
dfr	-0.031	-0.012
tms	-0.241	-0.214
dy	-0.883	-0.989
SII	0.047	-0.047
kp	-0.728	-0.642
SVIX	0.047	-0.295
ZCB1Y	-0.205	-0.215

Table 12: Country-by-country unconditional and conditional return predictions

This table reports the results of international country-by-country return predictions for US, UK, France, Spain, Japan and Australia. Panel A and B tabulate the results of unconditional (Equation (8)) and conditional (Equation (20)) return predictions respectively for each country. The coefficients estimates are followed by Newey and West (1987) t-statistic (with 18 lags) and Hodrick (1992) t-statistic. Intercept estimates are untabulated.

		US	UK	FRA	ESP	JPN	AUS
		Panel A: Unconditional predictions					
pr_t		-0.138	-0.195	-0.050	-0.093	-0.040	-0.123
	Newey-West t	(-4.719)	(-2.613)	(-0.889)	(-1.837)	(-2.520)	(-9.938)
	Hodrick t	[-2.743]	[-2.475]	[-0.790]	[-1.275]	[-3.081]	[-2.792]
Obs		344	280	203	261	272	167
R^2		0.238	0.127	0.009	0.054	0.027	0.152
		Panel B: Conditional predictions					
$I_{\{r_{t-12,t}<r^f_{t-12,t}\}}$		-0.012	-0.051	-0.086	-0.102	-0.097	0.020
		(-0.398)	(-1.149)	(-1.156)	(-1.684)	(-1.211)	(0.528)
		[-0.249]	[-0.919]	[-1.026]	[-1.434]	[-1.258]	[0.365]
$I_{\{r_{t-12,t}<r^f_{t-12,t}\}} \times pr_t$		-0.109	-0.078	-0.055	-0.075	-0.035	-0.066
		(-3.828)	(-5.265)	(-0.767)	(-2.588)	(-2.120)	(-9.371)
		[-2.949]	[-2.821]	[-0.856]	[-1.208]	[-2.782]	[-2.904]
$I_{\{r_{t-12,t}>r^f_{t-12,t}\}} \times pr_t$		-0.069	-0.044	-0.002	-0.049	-0.056	-0.053
		(-3.145)	(-2.102)	(-0.825)	(-1.424)	(-2.537)	(-1.400)
		[-1.829]	[-1.866]	[-0.573]	[-1.196]	[-2.600]	[-1.553]
Obs		344	280	203	261	272	167
R^2		0.255	0.162	0.058	0.102	0.066	0.156

Table 13: Risk Prices – AR(1) Shocks

This table reports the price of market risk and *pr* risk estimated using Fama-MacBeth method and Generalized Method of Moments (GMM). We use the two-stage GMM estimator with efficient weight matrix. pr_t shock is measured by AR(1) residual (ϵ_t^{pr}) estimated using the full sample. The full asset universe ("All") includes the twenty five Fama-French portfolios (sorted by size and book-to-market ratio), ten momentum portfolios, ten investment portfolios, and ten profitability portfolios. We also estimate *pr* risk price using twenty five value-size, momentum-size, investment-size, and profitability-size portfolios. The data of monthly portfolio returns are from Kenneth R. French's website. Each column corresponds to one set of assets. Each estimated price of risk is follwed by the t-statistic in parenthesis. *, **, and *** denote 5%, 2%, and 1% level of statistical significance respectively. We also report mean absolute percentage error (MAPE) and R^2.

	All (1)	Fama-French 25 (2)	Momentum 25 (3)	Investment 25 (4)	Profitability 25 (5)
Fama-MacBeth					
ϵ_t^{pr}	-0.202^{***}	-0.203^{*}	-0.382^{***}	0.099	-0.108
	(-4.067)	(-2.203)	(-3.764)	(1.686)	(-1.751)
$r_t - r_t^f$	0.009^{***}	0.010^{***}	0.010^{***}	0.010^{***}	0.010^{***}
	(4.130)	(4.147)	(4.000)	(4.207)	(4.103)
MAPE	0.206%	0.184%	0.238%	0.218%	0.241%
R^2	0.771	0.458	0.941	0.973	0.703
GMM					
ϵ_t^{pr}	-0.343^{***}	-1.096^{***}	-0.312^{***}	4.840	-0.064^{*}
	(-8.770)	(-3.366)	(-5.493)	(0.606)	(-2.312)
$r_t - r_t^f$	0.010^{***}	0.009^{***}	0.010^{***}	0.010^{*}	0.013^{***}
	(4.971)	(3.711)	(4.471)	(2.002)	(5.839)
MAPE	0.088%	0.047%	0.071%	0.071%	0.156%
R^2	0.730	0.678	0.667	0.720	0.726

Table 14: The Correlations Between pr Shocks and U.S. Stock Market Factors

This table documents the correlation between pr_t shocks and market excess return, size factor (SMB), value factor (HML), profitability factor (RMW), investment factor (CMA), and momentum factor. The factor returns are obtain from Kenneth R. French's website. We consider two versions of pr_t shocks, the first difference (Δpr_t) and AR(1) residual (ϵ_t^{pr}) estimated using full sample.

	Mkt-RF	SMB	HML	RMW	CMA	Momentum
Δpr_t	0.104	-0.019	-0.052	-0.057	-0.062	-0.127
ϵ_t^{pr}	0.081	-0.006	-0.031	-0.034	-0.035	-0.113

Table 15: State Space Model of Aggregate Dividends

This table reports the estimation results of (1) unrestricted state space model in Section 2.4, (2) restricted state space model (i.e., $\phi_z = 0$), and for comparison, (3) MA(1) model ($\Delta d_{t+1} = g + \sigma_D \varepsilon_{t+1} + \rho \sigma_D \varepsilon_t$), and (4) AR(1) model ($\Delta d_{t+1} = g + \gamma \Delta d_t + \sigma_D \varepsilon_{t+1}$) of the aggregate dividend growth series. Panel A uses annual dividend growth (non-overlapping) of S&P 500 index and Panel B uses annual dividend growth (non-overlapping) of CRSP NYSE/AMEX/NASDAQ Cap-Based index (i.e., the annual growth of total cash payment to shareholders in the U.S. stock market). Log likelihood ("LogL"), AIC, and BIC are reported. t-stats are in the squared bracket.

	$\hat{\phi}_z$	$\hat{g}$	$\hat{\sigma}_d$	$\hat{\sigma}_z$	$\hat{\rho}$	$\hat{\gamma}$	LogL	AIC	BIC
Panel A: S&P 500 dividend									
Unrestricted	0.27	0.04	0.00	0.12			106.35	-201.31	-189.40
	[1.29]	[3.19]	[0.00]	[2.33]					
Restricted		0.04	0.09	0.09			100.76	-195.51	-186.5
		[4.34]	[0.21]	[0.21]					
MA(1)		0.04	0.11		0.39		109.02	-212.04	-203.1
		[3.28]	[15.04]		[5.78]				
AR(1)		0.03	0.12			0.28	106.42	-206.84	-197.91
		[3.21]	[15.30]			[3.77]			
Panel B: CRSP NYSE/AMEX/NASDAQ dividend									
Unrestricted	-0.08	0.06	0.00	0.15			43.96	-79.92	-69.8
	[-0.06]	[3.86]	[0.00]	[0.12]					
Restricted		0.06	0.11	0.11			43.67	-81.34	-73.8
		[3.62]	[0.10]	[0.10]					
MA(1)		0.06	0.15		-0.09		44.00	-82.00	-74.4
		[3.94]	[6.99]		[-1.02]				
AR(1)		0.06	0.15			-0.08	43.96	-81.93	-74.39
		[3.89]	[6.98]			[-0.87]			

Table 16: Kostakis, Magdalinos, and Stamatogiannis (2015) IVX-Wald Test

This table reports test results on the predictive coefficient β in Table (2). IVX-Wald is the Wald statistic from Kostakis, Magdalinos, and Stamatogiannis (2015) to test $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$. The test is designed to be robust to the persistence of predictor. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	pr_t	pd_t	ϵ_t^{pr}	ϵ_t^{pd}
IVX-Wald	8.08***	1.67	5.55**	0.86

References

Adrian, Tobias, Erkko Etula, and Tyler Muir, 2014, Financial Intermediaries and the Cross-Section of Asset Returns, *Journal of Finance* LXIX, 1–56.

Adrian, Tobias, Emanuel Moench, and Hyun Song Shin, 2013, Dynamic leverage asset pricing, Working Paper 625, Federal Reserve Bank of New York.

Adrian, Tobias, and Hyun Song Shin, 2010, Liquidity and leverage, *Journal of Financial Intermediation* 19, 418 – 437, Risk Transfer Mechanisms and Financial Stability.

Ang, Andrew, and Geert Bekaert, 2007, Stock return predictability: Is it there?, *The Review of Financial Studies* 20, 651–707.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2009, High idiosyncratic volatility and low returns: International and further u.s. evidence, *Journal of Financial Economics* 91, 1 – 23.

Asness, Clifford S., Andrea Frazzini, and Lasse H. Pedersen, 2012, Leverage aversion and risk parity, *Financial Analysts Journal* 68, 47–59.

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *Journal of Finance* LXI, 1645–1680.

Baker, Scott R, Nicholas Bloom, and Steven J. Davis, 2016, Measuring Economic Policy Uncertainty, *The Quarterly Journal of Economics* 131, 1593–1636.

Ball, Ray, and Ross Watts, 1972, Some time series properties of accounting income, *The Journal of Finance* 27, 663–681.

Bansal, Ravi, and Amir Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles, *The Journal of Finance* 59, 1481–1509.

Barberis, Nicholas, Ming Huang, and Tano Santos, 2001, Prospect theory and asset prices, *The Quarterly Journal of Economics* 116, 1–53.

Beeler, Jason, and John Y. Campbell, 2012, The long-run risks model and aggregate asset prices: An empirical assessment, *Critical Finance Review* 1, 141–182.

Belo, Frederico, Pierre Collin-dufresne, and Robert Goldstein, 2015, Endogenous Dividend Dynamics and the Term Structure of Dividend Strips, *Journal of Finance* LXX, 1115–1160.

Benartzi, Shlomo, and Richard H. Thaler, 1995, Myopic loss aversion and the equity premium puzzle, *The Quarterly Journal of Economics* 110, 73–92.

Bernanke, Ben S., and Kenneth N. Kuttner, 2005, What explains the stock market's reaction to federal reserve policy?, *The Journal of Finance* 60, 1221–1257.

Binsbergen, Jules H. Van, Michael W. Brandt, and Ralph S. J. Koijen, 2012, On the Timing and Pricing of Dividends, *American Economic Review* 102, 1596–1618.

Binsbergen, Jules H. Van, Wouter Hueskes, Ralph S. J. Koijen, and Evert B Vrugt, 2013, Equity yields, *Journal of Financial Economics* 110, 503–519.

Binsbergen, Jules H. Van, and Ralph S. J. Koijen, 2010, Predictive Regressions: A Present-Value Approach, *Journal of Finance* LXV, 1439–1471.

Binsbergen, Jules H. Van, and Ralph S. J. Koijen, 2017, The term structure of returns: Facts and theory, *Journal of Financial Economics* 124, 1 – 21.

Bossaerts, Peter, and Pierre Hillion, 1999, Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn? , *Review of Financial Studies* 12, 405–428.

Campbell, John Y., 1993, Intertemporal asset pricing without consumption data, *The American Economic Review* 83, 487–512.

Campbell, John Y., and John Ammer, 1993, What moves the stock and bond markets? a variance decomposition for long-term asset returns, *The Journal of Finance* 48, 3–37.

Campbell, John Y., and John H. Cochrane, 1999, By force of habit: A consumptionbased explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–251.

Campbell, John Y., Sanford J. Grossman, and Jiang Wang, 1993, Trading volume and serial correlation in stock returns*, *The Quarterly Journal of Economics* 108, 905–939.

Campbell, John Y., Carolin Pflueger, and Luis M. Viceira, 2015, Monetary policy drivers of bond and equity risks, Working paper, Harvard University.

Campbell, John Y., and Robert J. Shiller, 1988, The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors, *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.

Chauvet, Marcelle, and Jeremy Piger, 2008, A comparison of the real-time performance of business cycle dating methods, *Journal of Business and Economic Statistics* 26, 4249.

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

Chen, Long, Zhi Da, and Richard Priestley, 2012, Dividend smoothing and predictability, *Management Science* 58, 1834–1853.

Chen, Long, Zhi Da, and Xinlei Zhao, 2013, What drives stock price movements?, *The Review of Financial Studies* 26, 841–876.

Clark, Todd E, and Michael W McCracken, 2001, Tests of equal forecast accuracy and encompassing for nested models, *Journal of Econometrics* 105, 85 – 110, Forecasting and empirical methods in finance and macroeconomics.

Clark, Todd E., and Kenneth D. West, 2007, Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138, 291 – 311, 50th Anniversary Econometric Institute.

Cochrane, John H., 1991, Production-based asset pricing and the link between stock returns and economic fluctuations, *The Journal of Finance* 46, 209–237.

Cochrane, John H., 1992, Explaining the variance of pricedividend ratios, *Review of Financial Studies* 5, 243–280.

Cochrane, John H., 2005, *Asset Pricing*, revised edition edition (Princeton University Press, Princeton, NJ).

Cochrane, John H., 2007, The Dog That Did Not Bark: A Defense of Return Predictability, *Review of Financial Studies* 21, 1533–1575.

Cochrane, John H, 2008, State-space vs. var models for stock returns, *Unpublished paper, Chicago GSB* .

Cochrane, John H., 2011, Presidential Address: Discount Rates, *Journal of Finance* 66, 1047–1108.

Cochrane, John H., and Monika Piazzesi, 2002, The fed and interest rates - a high-frequency identification, *American Economic Review* 92, 90–95.

Collin-Dufresne, Pierre, Michael Johannes, and Lars A. Lochstoer, 2016, Parameter learning in general equilibrium: The asset pricing implications, *American Economic Review* 106, 664–98.

Cook, Timothy, and Thomas Hahn, 1989, The effect of changes in the federal funds rate target on market interest rates in the 1970s, *Journal of Monetary Economics* 24, 331–351.

Cujean, Julien, and Michael Hasler, 2017, Why does return predictability concentrate in bad times?, *The Journal of Finance* forthcoming.

Da, Zhi, Ravi Jagannathan, and Jianfeng Shen, 2014, Growth expectations, dividend yields, and future stock returns, Working Paper 20651, National Bureau of Economic Research.

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

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl, 2017, A model of monetary policy and risk premia, *The Journal of Finance* forthcoming.

Duffie, Darrell, 2001, *Dynamic Asset Pricing Theory*, Princeton Series in Finance, third edition (Princeton University Press, Princeton, N.J.).

Epstein, Larry G., and Stanley E. Zin, 1989, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework, *Econometrica* 57, 937–969.

Fama, Eugene, and Kenneth French, 2000, Forecasting profitability and earnings, *The Journal of Business* 73, 161–75.

Fama, Eugene F., and Kenneth R. French, 1988, Permanent and temporary components of stock prices, *Journal of Political Economy* 96, 246–273.

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23 – 49.

Farmer, Leland, Lawrence Schmidt, and Allan Timmermann, 2018, Pockets of predictability, Discussion Paper DP12885, CEPR.

Ferson, Wayne, and Campbell Harvey, 1991, The variation of economic risk premiums, *Journal of Political Economy* 99, 385–415.

Ferson, Wayne E., Sergei Sarkissian, and Timothy T. Simin, 2003, Spurious regressions in financial economics?, *The Journal of Finance* 58, 1393–1413.

Foster, F. Douglas, Tom Smith, and Robert E. Whaley, 1997, Assessing goodness-of-fit of asset pricing models: The distribution of the maximal r¡sup¿2¡/sup¿, *The Journal of Finance* 52, 591–607.

Frazzini, Andrea, and Lasse Heje Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1 – 25.

Gilchrist, Simon, and Egon Zakrajšek, 2012, Credit spreads and business cycle fluctuations, *American Economic Review* 102, 1692–1720.

Golez, Benjamin, 2014, Expected returns and dividend growth rates implied by derivative markets, *The Review of Financial Studies* 27, 790–822.

Goyal, Amit, and Ivo Welch, 2007, A Comprehensive Look at The Empirical Performance of Equity Premium Prediction, *Review of Financial Studies* 21, 1455–1508.

Guo, Hui, and Robert F. Whitelaw, 2006, Uncovering the riskreturn relation in the stock market, *The Journal of Finance* 61, 1433–1463.

Hansen, Peter Reinhard, and Allan Timmermann, 2012, Choice of sample split in out-of-sample forecast evaluation, Working paper, Stanford University.

Hayashi, Fumio, 2000, *Econometrics* (Princeton University Press, Princeton, NJ).

He, Zhiguo, Bryan Kelly, and Asaf Manela, 2017, Intermediary asset pricing: New evidence from many asset classes, *Journal of Financial Economics* .

He, Zhiguo, and Arvind Krishnamurthy, 2013, Intermediary asset pricing, *American Economic Review* 103, 732–70.

Henkel, Sam James, J. Spencer Martin, and Federico Nardari, 2011, Time-varying short-horizon predictability, *Journal of Financial Economics* 99, 560 – 580.

Hodrick, Robert J., 1992, Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement, *Review of Economic Dynamics* 5, 357–386.

Huang, Dashan, Fuwei Jiang, Jun Tu, and Guofu Zhou, 2017, Forecasting stock returns in good and bad times: The role of market states, Working paper, Singapore Management University.

Jagannathan, Ravi, and Binying Liu, 2015, Dividend dynamics, learning, and expected stock index returns, Working Paper 21557, National Bureau of Economic Research.

Kahneman, Daniel, and Amos Tversky, 1979, Prospect theory: An analysis of decision under risk, *Econometrica* 47, 263–291.

Karolyi, G. Andrew, and Ren M. Stulz, 2003, Chapter 16 are financial assets priced locally or globally?, in *Financial Markets and Asset Pricing*, volume 1 of *Handbook of the Economics of Finance*, 975 – 1020 (Elsevier).

Kelly, Bryan, and Seth Pruitt, 2013, Market Expectations in the Cross-Section of Present Values, *Journal of Finance* LXVIII, 1721–1756.

Kim, Myung Jig, Charles R. Nelson, and Richard Startz, 1991, Mean reversion in stock prices? a reappraisal of the empirical evidence, *The Review of Economic Studies* 58, 515–528.

Koijen, Ralph S. J., and Stijn van Nieuwerburgh, 2011, Predictability of Returns and Cash Flows, *Annual Review of Financial Economics* 3, 467–491.

Kostakis, Alexandros, Tassos Magdalinos, and Michalis P. Stamatogiannis, 2015, Robust econometric inference for stock return predictability, *Review of Financial Studies* 28, 15061553.

Kuttner, Kenneth, 2001, Monetary policy surprises and interest rates: Evidence from the fed funds futures market, *Journal of Monetary Economics* 47, 523–544.

Lacerda, Filipe, and Pedro Santa-Clara, 2010, Forecasting dividend growth to better predict returns, Working paper, Universidade Nova de Lisboa.

Lamont, Owen A., 2000, Investment plans and stock returns, *The Journal of Finance* 55, 2719–2745.

Larrain, Borja, and Motohiro Yogo, 2008, Does firm value move too much to be justified by subsequent changes in cash flow?, *Journal of Financial Economics* 87, 200 – 226.

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

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

Lettau, Martin, and Sydney C. Ludvigson, 2010, {CHAPTER} 11 - measuring and modeling variation in the risk-return trade-off, in YACINE AT-SAHALIA, and LARS PETER HANSEN, eds., *Handbook of Financial Econometrics: Tools and Techniques*, volume 1 of *Handbooks in Finance*, 617 – 690 (North-Holland, San Diego).

Lettau, Martin, and Jessica A. Wachter, 2007, Why Is Long-Horizon Equity Less Risky? A Duration-Based Explanation of the Value Premium, *Journal of Finance* LXII, 55–92.

Lucca, David O., and Emanuel Moench, 2015, The pre-fomc announcement drift, *The Journal of Finance* 70, 329–371.

Martin, Ian, 2017, What is the expected return on the market?*, *The Quarterly Journal of Economics* 132, 367–433.

Menzly, Lior, Tano Santos, and Pietro Veronesi, 2004, Understanding predictability, *Journal of Political Economy* 112, 1–47.

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.

Miranda-Agrippino, Silvia, and Hélène Rey, 2015, World asset markets and the global financial cycle, Working Paper 21722, National Bureau of Economic Research.

Moreira, Alan, and Tyler Muir, 2017, Volatility-managed portfolios, *The Journal of Finance* 72, 1611–1644.

Moskowitz, Tobias J., Yao Hua Ooi, and Lasse Heje Pedersen, 2012, Time series momentum, *Journal of Financial Economics* 104, 228 – 250, Special Issue on Investor Sentiment.

Muir, Tyler, 2017, Financial crises and risk premia, *The Quarterly Journal of Economics* 132, 765–809.

Nakamura, Emi, and Jon Steinsson, 2017, High frequency identification of monetary non-neutrality: The information effect, *Quarterly Journal of Economics* forthcoming.

Nelson, Charles R., and Myung J. Kim, 1993, Predictable stock returns: The role of small sample bias, *The Journal of Finance* 48, 641–661.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

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

Pástor, Ľuboš, and Robert F. Stambaugh, 2003, Liquidity Risk and Expected Stock Returns, *Journal of Political Economy* 111, 642–685.

Patelis, Alex D., 1997, Stock return predictability and the role of monetary policy, *The Journal of Finance* 52, 1951–1972.

Poterba, James M, and Lawrence H Summers, 1988, Mean reversion in stock prices: Evidence and implications, *Journal of Financial Economics* 22, 27 – 59.

Presidential Task Force on Market Mechanisms (Chairman: Nicholas Brady), 1988, *Report of the Presidential Task Force on Market Mechanisms: submitted to The President of the United States, The Secretary of the Treasury, and The Chairman of the Federal Reserve Board* (U.S. Government Printing Office, Washington, D.C.).

Rapach, David E., Matthew C. Ringgenberg, and Guofu Zhou, 2016, Short interest and aggregate stock returns, *Journal of Financial Economics* 121, 46 – 65.

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.

Rossi, Barbara, and Atsushi Inoue, 2012, Out-of-sample forecast tests robust to the choice of window size, *Journal of Business & Economic Statistics* 30, 432–453.

Rozeff, Michael S., 1974, Money and stock prices: Market efficiency and the lag in effect of monetary policy, *Journal of Financial Economics* 1, 245–302.

Rytchkov, Oleg, 2012, Filtering out expected dividends and expected returns, *Quarterly Journal of Finance* 02, 1250012.

Savov, Alexi, 2011, Asset pricing with garbage, *The Journal of Finance* 66, 177–201.

Stambaugh, Robert F., 1999, Predictive Regressions, *Journal of Financial Economics* 54, 375–421.

Thaler, Richard H., and Eric J. Johnson, 1990, Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice, *Management Science* 36, 643–660.

Thorbecke, Willem, 1997, On stock market returns and monetary policy, *The Journal of Finance* 52, 635–654.

Vuolteenaho, Tuomo, 2002, What drives firm-level stock returns, *Journal of Finance* LVII, 233–264.