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

Ye Li† Chen Wang‡
June 4, 2018

Abstract

The prices of dividends at alternative horizons contain critical information on the behavior of aggregate stock market. The ratio between prices of long- and short-term dividends, “price ratio” \( pr_t \), predicts annual market return with an out-of-sample \( R^2 \) of 19%. \( pr_t \) subsumes the predictive power of traditional price-dividend ratio \( pd_t \). After orthogonalized to \( pr_t \), the residuals of \( pd_t \) strongly predict dividend growth. Using an exponential-affine model, we show a one-to-one mapping between \( pr_t \) and the expected market return when the expectation of future cash flow is transient. Moreover, we find that return predictability is stronger after market downturns, and holds outside the U.S. As an economic test, shocks to \( pr_t \) are priced in the cross-section of stocks, consistent with ICAPM. Our measure of expected return declines during monetary expansions, and varies strongly with the conditions of macroeconomy, financial intermediaries, and sentiment.

*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. 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

We provide new evidence on stock return predictability, and show that return predictability is closely related to the (lack of) persistence of cash-flow expectation. Our predictor reveals interesting dynamics of the equity premium. The expected return of the stock market declines in response to expansionary monetary policy, and varies closely with variables that reflect the conditions of macroeconomy, financial intermediary, and sentiment. Moreover, shocks to the expected market return is priced in the cross section of stocks.

Our return predictor is the ratio of long-term dividend price to short-term dividend price. While great efforts have been made to understand the properties of dividend strips, especially their average returns (reviewed by Binsbergen and Koijen (2017)), we are the first to show that the ratio of dividend strip prices contain critical information on the expected return of aggregate market, and explore its economic implications in an exponential-affine framework of Lettau and Wachter (2007). We also find stronger return predictability after the market underperforms the risk-free rate, and evaluate theories that feature such asymmetry. Our results hold outside the United States, and our return predictor gains its power mainly from a common component of discount rates across countries.

We start from a simple identity: the total valuation of the market is equal to the sum of the price of long-term dividends and the price of short-term dividends, so the price-dividend ratio can be decomposed as follows

\[ \frac{P_t}{D_t} = \frac{\text{Price of Long-term Dividends}}{D_t} + \frac{\text{Price of Short-term Dividends}}{D_t}. \]

Using a general state-space model, we show that the two components could contain distinct information on future returns and dividends. To extract information from the pair that is beyond the traditional price-dividend ratio (i.e., \( pd_t = \ln \left( \frac{P_t}{D_t} \right) \)), we calculate the ratio of
long-term to short-term dividend prices (“price ratio” or “pr”)

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

The dividend prices are obtained from derivative markets from January 1988 to June 2017.\(^1\)

Our price ratio can be interpreted as the slope of term structure of dividend prices, while the price-dividend ratio \((pd_t)\) captures the level. Moreover, \(pr_t\) is a measure of duration. In our implementation, “short-term” is defined as one year. Using the valuation of dividends in the coming year as numeraire, \(pr_t\) measures how many years of valuation are there beyond one year. A high value of \(pr_t\) means that the market has a long duration.

We find that \(pr_t\) strongly predicts market return. A decrease of \(pr_t\) by one standard deviation adds 7.3% to the expected return over the next year. When forecasting annual returns, \(pr_t\) produces 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. This degree of 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 forecasting performance of \(pr_t\) can be directly compared to that of the many alternative predictors in the literature. Previously studied predictors typically perform well in-sample but become insignificant out-of-sample, often performing worse than forecasts based on the historical mean return (Goyal and Welch (2007)).

We establish the robustness of our return 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 on the left-hand side of predictive regression. 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. Ferson, Sarkissian, and Simin (2003) show a spurious regression bias when both the proposed predictor and the underlying expected return are persistent. Third, we conduct several out-of-sample tests (e.g., Clark

\(^1\)Binsbergen, Brandt, and Kojien (2012) show that futures or option data can be used to calculate dividend strip prices. We use futures data, because futures have a longer sample than options. Figure 8 in the appendix shows that \(pr_t\) computed from futures and option data have 88% correlation.
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 all the alternative predictors.

Next, we explore the economic implications of our findings by imposing more structure on the state space model. In particular, we construct the dividend process and the stochastic discount factor following Lettau and Wachter (2007). The superior return predictive power of $pr_t$ suggests that it is a strong signal of price of risk. In the model, $pr_t$ is a perfect signal of price of risk if and only if the expected dividend growth is not persistent. Below we present a simplified version of the model to illustrate the intuition. Consider an economy of infinite horizon, where each period (“year”) is index by $t$. Let $D_t$ denote the dividend at $t$, and $g_{t+1}$ the realized dividend growth from $t$ to $t+1$ that has a time-$t$ conditional normal distribution, $g_{t+1} \sim N\left(z_t - \frac{1}{2} \sigma^2_d, \sigma^2_d\right)$, where $z_t$ is the expected growth so $\mathbb{E}_t [D_{t+1}] = \mathbb{E}_t [e^{g_{t+1} D_t}] = e^{z_tD_t}$. To simplify the exposition, we consider a constant expected return $r$, and an i.i.d. normal distribution of $z_t$ with mean $-\frac{1}{2} \sigma^2_z$ and variance $\sigma^2_z$ so that $\mathbb{E} [e^{z_t}] = 1$. Shocks to $g$ and $z$ are independent. We show that given this transient expected dividend growth, $pr_t$ precisely reveals $r$. At $t$, the price of dividends paid in the coming year is simply

$$P_t^- = \frac{\mathbb{E}_t [D_{t+1}]}{1 + r} = \frac{e^{z_tD_t}}{1 + r}. \quad (1)$$

For $j > 1$, $D_{t+j} = e^{(\sum_{k=1}^j g_{t+k})} D_t$, where $\{g_{t+k}\}_{k=1}^j$ are independent, and for $k \geq 2$, each $g_{t+k}$ follows $N\left(-\frac{1}{2} \sigma^2_z - \frac{1}{2} \sigma^2_d, \sigma^2_z + \sigma^2_d\right)$. Using the law of iterated expectation, we have

$$\mathbb{E}_t [D_{t+j}] = \mathbb{E}_t \left[\mathbb{E}_{t+1} \left[e^{(\sum_{k=2}^j g_{t+k})} e^{g_{t+1} D_t}\right]\right] = \mathbb{E}_t [e^{g_{t+1} D_t}] = e^{z_tD_t}. \quad (2)$$

The price of dividends paid beyond one year is

$$P_t^+ = \sum_{t=2}^{+\infty} \frac{\mathbb{E}_t [D_{t+j}]}{(1 + r)^j} = \frac{e^{z_tD_t}}{r (1 + r)}, \quad (3)$$

3
so

\[ pr_t = \ln \left( \frac{P^+}{P^-} \right) = -\ln(r). \]  

The assumption of transient expected dividend growth, i.e., \( i.i.d. \) \( z_t \), is critical. Otherwise, for instance, \( z_{t+1} \sim N(\phi z_t - \frac{1}{2}\sigma^2 z, \sigma^2) \), we have \( \mathbb{E}_t [D_{t+2}] = \mathbb{E}_t \left[ e^{(\beta z_t + \gamma z_{t+1})} D_t \right] = e^{(1+\phi z_t)} D_t \), so dividing \( P^+ \) by \( P^- \) cannot get rid of \( z_t \), and as a result, \( pr_t \) still contains both \( r \) and \( z_t \).

The intuition behind is very simple. When the expected dividend growth is transient, the expected growth in the next period, \( z_t \), increases in the expected level of future dividends by the same amount in spite of the horizon, shifting upward the path of future dividends in a parallel fashion. Therefore, dividing \( P^+ \) by \( P^- \) gets rid of the expected dividend growth, leaving only the expected return in \( pr_t \). Many have shown that the structure of aggregate dividend is important for understanding key asset pricing moments (Bansal and Yaron (2004); Beeler and Campbell (2012); Belo et al. (2015)). We are the first to pin return predictability directly to a single parameter, that is the persistence of cash-flow expectation. By examining the autocorrelation structure, and by directly fitting a state space model to the aggregate dividends, we find that the autoregressive parameter of expected dividend growth is not statistically different from zero.

The structural model also helps explain another intriguing finding: the residuals from regressing \( pd_t \) on \( pr_t \) strongly forecast dividend growth at one year horizon with an out-of-sample \( R^2 \) equal to 30%. \( pd_t \) is solved as function of both the price of risk and the expected dividend growth. Projecting \( pd_t \) on \( pr_t \) takes the variation of price of risk out of \( pd_t \), so what remains should forecast future dividends.\(^2\) In contrast, \( pd_t \) itself does not predict dividends, as already well documented (e.g., Cochrane (2011)). A recent literature focuses instead on adjustments of \( pd_t \) that may eliminate the variation of expected dividend growth, and thus, make the adjusted \( pd_t \) a better return predictor. (Campbell and Thompson (2008); Lacerda and Santa-Clara (2010); Da, Jagannathan, and Shen (2014); Golez (2014)). Our predictor

\(^2\)It has been shown in other settings that information regarding future dividends compromises the return predictive power of \( pd_t \) (e.g., Menzly, Santos, and Veronesi (2004); Lettau and Ludvigson (2005)).
$pr_t$ does not rely on an adjustment model. It is obtained directly from market prices of dividend strips. Since prices respond to news more promptly than fundamentals, $pr_t$ reveals high-frequency variation in the expected return.

The price ratio, $pr_t$, predicts return outside the United States. We run a panel predictive regression with the future realized return of each country on the left hand side, and $pr_t$ of each country on the right hand side. The panel predictive coefficient is strongly significant, and close in magnitude to the coefficient estimated using the U.S. sample. Interestingly, once time fixed effect is added to absorb the global factor in realized returns, return predictability disappears, suggesting that the variation of expected stock return across countries tends to comove, which is in line with Miranda-Agrippino and Rey (2015).

We find that the return predictive power of $pr_t$ is asymmetric – it is much stronger following a down market (i.e., negative market excess returns in the past twelve months). Such asymmetry holds outside the United States. We evaluate two asset pricing models, Barberis, Huang, and Santos (2001) and He and Krishnamurthy (2013), that produce strong asymmetry in return predictability. While the return predictive power of $pr_t$ does depend 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.

To explore the mechanisms behind expected return dynamics, we study how the estimated expected return responds to monetary policy shocks, and its correlation with macroeconomic and financial market conditions. 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)). Using $pr_t$ as an expected return proxy, we show that the expected return declines during monetary expansions. Specifically, 

---

3Our 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)). When $pr_t$ is at its mean, 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, 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, have better predictive power in economic downturns.
we regress $pr_t$ on the unanticipated changes in Federal Funds rate that are calculated following Cochrane and Piazzesi (2002), and find a negative shock to the policy rate (monetary easing) is associated with an increase in $pr_t$, and thus, a decrease in the expected stock return. In contrast, the level of price-dividend ratio, a typical proxy for expected return (e.g., Muir (2017)), does not respond to monetary policy shocks. We also find that monetary easing tends to be associated with higher contemporaneous realized return, in line with Thorbecke (1997) and Bernanke and Kuttner (2005). Thus, stock price rises in response to expansionary monetary policy, but since the expected return declines, such increase tends to revert over the next year. We proxy the expected dividend growth by the residual from regressing $pd_t$ on $pr_t$, and find it does not respond to monetary policy shocks.

Next, we show that the expected stock return, calculated from unconditional and conditional predictive regressions, varies with business conditions, broker-dealer balance sheets, uncertainty, and sentiment. The expected return is positively correlated with unemployment and term spread, 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 a positive correlation with broker-dealer CDS spreads, a proxy for under-capitalization. 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 macro variables.

Last but not least, we estimate the price of risk for shocks to $pr_t$ as an economic test of return predictability. If $pr_t$ is a valid return predictor, shocks to it are shocks to the expected market return. By the logic of ICAPM, shocks to investment opportunity set are priced. Using characteristic-sorted portfolios as the asset universe, we find a negative and statistically significant price of $pr_t$ risk. Two assets with one standard deviation difference in

\footnote{VIX may not reflect the risk from variation in the investment opportunity set, which can be a important component of risk (Guo and Whitelaw (2006)).}
their $pr_t$ beta have 2.1% difference in average annual returns. In contrast, few existing studies on return predictors conduct this economic test, and most studies solely rely statistical tests to establish return predictive power.

The reminder of the paper is organized as follows. Section 2 shows variable construction and documents the evidence of return predictability (in and outside the United States) when $pr_t$ is used as predictor. This section also shows that the residuals from regressing $pd_t$ on $pr_t$ strongly predict dividend growth. An exponential-affine model unveils the economic intuitions behind these findings. Section 3 provides the evidence on the asymmetry of return predictability (in and outside the United States), and evaluates related theories. Section 4 provides evidence on how monetary policy affects the expected return, and how the expected return varies with macro and financial variables. It closes with the estimation of price of $pr_t$ risk. Section 4 concludes. Derivation and additional results are provided in the appendices.

2 Return Prediction

This section starts with a general state space model of return and cash flow that shows the price-dividend ratio (“$pd_t$”) as a compression of information on future return and dividends. To release the information trapped in $pd_t$, we decompose market price into the prices of long- and short-term dividends, and calculate the ratio of the former to the latter, i.e., the “price ratio” ($pr_t$). Table 1 compares summary statistics of $pd_t$ and $pr_t$, and shows that $pr$ is much less persistent. Table 2 shows the return predictive power of $pr_t$, which is much stronger that $pd_t$’s. $pr_t$ also outperforms other predictors in out-of-sample $R^2$ and Hodrick (1992) t-statistics (Figure 3). Next, we explore dividend predictability, In particular, we find that the residuals from regressing $pd_t$ on $pr_t$ strongly predict dividends (Table 3).

To explore the economic intuitions behind our findings, we impose more structure on the state space model in Section 2.4. Specifically, we adopt the exponential-affine framework of Lettau and Wachter (2007), and show the superior return predictive power of $pr_t$ is closely
related to the persistence of expected dividend growth. Finally, we expand our sample, and provide evidence that \( pr_t \) forecasts returns outside of the United States (Table 4).

### 2.1 Decomposing the price-dividend ratio

**State space model.** We consider a state space model of return and cash flow growth (e.g., Cochrane (2008)). Let \( \mu_t \) 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\(^5\)

\[
\begin{align*}
\mu_t &= \gamma_0 + \gamma' F_t, \\
g_t &= \delta_0 + \delta' F_t.
\end{align*}
\]

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

\[
F_{t+1} = \Lambda F_t + \xi_{t+1},
\]

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

\[
\begin{align*}
pd_t &= \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^{j-1} E_t[\Delta d_{t+j} - r_{t+j}],
\end{align*}
\]

(7)

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

\[
pd_t = \phi_0 + \phi' F_t,\]

(8)

---

A non-linear model is more general, but this model is only used for the purpose of motivation, not estimation.
where $\phi_0$ is equal to $\frac{\kappa + \delta_0 - \gamma_0}{1 - \rho}$, and $\phi'$ is equal to $\iota\psi' (1 - \rho \Lambda)^{-1}$ with $\iota$ being a row vector $(1, -1)$ and $\psi$ equal to $(\delta', \gamma')$. 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. Yet, the factor structure reveals that any predictive power of $pd_t$ comes from a particular linear combination of $F_t$, i.e., a compression of information. Therefore, we should be able to release the trapped information by decomposing the price-dividend ratio into different components with distinct information content from $F_t$. Next, we consider a decomposition along cash-flow horizon.

**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 a sum of the expected future dividends discounted by the cumulative short rates:

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

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 $D_t$:

$$\frac{S_t}{D_t} = \frac{P_t^{T-}}{D_t} + \frac{P_t^{T+}}{D_t}.$$ (9)

While the price-dividend ratio is the sum of these two valuation ratios, we construct our predictor by taking the (log) difference so that it may reflect different information from
the pair \( \left( \frac{PT-}{Dt}, \frac{PT+}{Dt} \right) \):

\[
pr_t = \ln \left( \frac{PT+}{Dt} \right) - \ln \left( \frac{PT-}{Dt} \right) = \ln \left( \frac{PT+}{PT-} \right)
\] (10)

Our predictor “\( pr_t \)” is a price ratio, the log ratio of long-term dividend price to short-term dividend price. We use the log difference instead of level difference to get rid of \( Dt \), so that \( pr_t \) has market prices in both its numerator and denominator, 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 sluggish than \( pr_t \), and thus, less responsive to the current conditions of financial markets and the real economy.

Together, \( pd_t \) and \( pr_t \) should reflect the information content of \( \left( \frac{PT-}{Dt}, \frac{PT+}{Dt} \right) \). Our empirical results will show that our price ratio \( pr_t \) is a better way to extract information about future returns than the traditional price-dividend ratio. 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.\(^6\) Consider any \( T > 0 \). To calculate \( P^T_t \) from futures price and ZCB price, we make

\(^6\)Figure 8 shows that \( pr_t \) from futures has 88% correlation with \( pr_t \) from options in Binsbergen, Brandt, and Koijen (2012).
the assumption that $\int_t^{t+T} r_s ds$ and $S_{t+T}$ are not correlated under $Q$ measure, so we have

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

(11)

Therefore, we can calculate $P_{t+T}^T$ 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. To construct $pr_t$, 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. $pd_t$ is the month-end 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 price of dividends paid within the the coming year. Accordingly, we focus on forecasting the return of S&P 500 index at one-year horizon, but also report the forecasting results at one-month horizon in the appendix. market return is replaced by the return of S&P index, and when we use excess returns (results in the Appendix). The sample starts in 1988 because 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)). After the crash, regulators overhauled several trade-clearing protocols.8

---

7 Available maturities vary over time, so to obtain futures at constant maturities, such as one year, we need to interpolate data. We use shape-preserving piecewise cubic interpolation to preserve the shape of the futures curve.

8 According to the New York Stock Exchange’s current website: “In response to the market breaks in October 1987 and October 1989, the New 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.”
Table 1: Summary Statistics

This table reports the number of observations, mean, standard deviation, minimum, maximum, quartiles, and first-order (one-month) autocorrelation ($\rho$) 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 (11), 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).

<table>
<thead>
<tr>
<th></th>
<th>obs</th>
<th>mean</th>
<th>std</th>
<th>min</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>max</th>
<th>$\rho$</th>
<th>corr.</th>
<th>$pr$</th>
<th>$pd$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$pr$</td>
<td>348</td>
<td>3.992</td>
<td>0.531</td>
<td>2.677</td>
<td>3.630</td>
<td>3.992</td>
<td>4.195</td>
<td>6.631</td>
<td>0.915</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$pd$</td>
<td>348</td>
<td>3.873</td>
<td>0.307</td>
<td>3.241</td>
<td>3.594</td>
<td>3.887</td>
<td>4.052</td>
<td>4.551</td>
<td>0.987</td>
<td>0.874</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 reports the summary statistics of $pr_t$, and 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 all the years after the coming year is 54.2 times the valuation of dividends in the coming 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 dot-com boom (Nov. 2000).

$pr_t$ has a lower one-month autocorrelation (“$\rho$”) 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 will be shown later by 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 the residual of $pr_t$ after regressing on $pd_t$ and that of the residual of $pd_t$ after regressing on $pr_t$.

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

\[ r_{t,t+12} = \alpha + \beta x_t + \epsilon_{t,t+12}, \]  

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 \( \beta \) estimate. We report the estimate adjusted for Stambaugh (1999) bias.

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 from the perspective of a real-time investor, using data up to time \( t \) in the predictive regression to obtain the coefficient \( \beta \), 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 defined by

\[ R^2_{OOS} = 1 - \frac{\sum_t (r_{t,t+12} - \hat{r}_{t,t+12})^2}{\sum_t (r_{t,t+12} - \bar{r})^2}, \]

where \( \hat{r}_{t,t+12} \) is the forecast value and \( \bar{r} \) is the average of twelve-month returns (the first return 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 return’s 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
Table 2: One-year Return Prediction

This table reports the results of predictive regression (Equation (12)). 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 $pd_t$ ($\epsilon^p_{pr,t}$), and the residuals of $pd_t$ after regressing on $pr_t$ ($\epsilon^p_{pd,t}$), and the results are reported in Column (1) to (4) respectively. The $\beta$ 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)).

<table>
<thead>
<tr>
<th></th>
<th>$pr_t$</th>
<th>$pd_t$</th>
<th>$\epsilon^p_{pr,t}$</th>
<th>$\epsilon^p_{pd,t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>-0.138</td>
<td>-0.193</td>
<td>-0.160</td>
<td>0.098</td>
</tr>
<tr>
<td>Newey-West $t$</td>
<td>(-4.718)</td>
<td>(-3.575)</td>
<td>(-2.233)</td>
<td>(0.848)</td>
</tr>
<tr>
<td>Hodrick $t$</td>
<td>[-2.743]</td>
<td>[-2.217]</td>
<td>[-1.677]</td>
<td>[0.613]</td>
</tr>
<tr>
<td>Stambaugh bias adjusted $\beta$</td>
<td>-0.127</td>
<td>-0.182</td>
<td>-0.152</td>
<td>0.107</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.238</td>
<td>0.157</td>
<td>0.076</td>
<td>0.010</td>
</tr>
<tr>
<td>OOS $R^2$</td>
<td>0.192</td>
<td>0.068</td>
<td>0.048</td>
<td>-0.043</td>
</tr>
<tr>
<td>ENC</td>
<td>4.052</td>
<td>1.776</td>
<td>2.249</td>
<td>-0.241</td>
</tr>
<tr>
<td>$p(ENC)$</td>
<td>&lt; 0.01</td>
<td>&lt; 0.10</td>
<td>&lt; 0.05</td>
<td>&gt; 0.10</td>
</tr>
<tr>
<td>$p(CW)$</td>
<td>0.007</td>
<td>0.041</td>
<td>0.111</td>
<td>0.348</td>
</tr>
</tbody>
</table>

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 23.8%.\textsuperscript{9} Out-of-sample forecasts are similarly powerful, delivering an $R^2$ of 19.2%, signifi-

\textsuperscript{9}Foster, Smith, and Whaley (1997) discuss the potential data mining issues that arises 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.
cantly outperforming the historic 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 (“s0”) 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 (“s1”) is related to s0 through $s_1 = \sqrt{\frac{s_0^2 + R_t^2}{1 - R_t^2}}$, where $R_t^2$ 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% (Table 2) implies a Sharpe ratio of 0.84, suggesting that the stochastic discount factor is more volatile than what is implied by state-of-art structural asset pricing models (e.g., Campbell and Cochrane (1999) and Bansal and Yaron (2004)).

The predictive coefficient is also large in magnitude, indicating a high volatility of expected return. An 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 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. In the appendix (Figure 9), we show that the predictive coefficient of $pd_t$ is also less stable than $pr_t$. 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 $pr_t$.

Since $pr_t$ and $pd_t$ are highly correlated, we regress $pr_t$ on $pd_t$ to obtain residuals, $\epsilon_{t}^{pr}$, 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 $\epsilon_{t}^{pr}$ 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), we report the prediction results of $\epsilon_{t}^{pd}$, the residuals from regressing $pd_t$ on $pr_t$. $pd_t$ residuals do not exhibit return predictive power, which again confirms $pr_t$ as a superior predictor.
Variation in 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 \( \epsilon^pr_t \), the residuals from regressing \( pr_t \) on \( pd_t \). The area under spectrum, i.e., the 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 \( \pi \), we mark the corresponding length of cycle for easier interpretation. Consistent with the fact that \( pr_t \) is less persistent than \( pd_t \), its variation is also much more concentrated in higher frequencies than the variation of \( pd_t \), and once orthogonalized with respect to \( pd_t \), \( pr_t \)’s residual varies mainly at frequencies higher than one year. Panel B plots the cross-spectrum of \( pr_t \) and \( pd_t \). The integral of cross-spectrum density is the covariance between \( pr_t \) and \( pd_t \). The high 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 has strong return predictive power beyond \( pd_t \).

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-
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.

The first conclusion we draw from this graph is that in contrast to $pd_t$, which produces a very 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 $pd_t$ is more persistent than $pr_t$. $pr_t$ is more responsive to news, as it contains only the market prices of short-term and long-term dividends. In contrast, $pd_t$ has a denominator that is a rolling accumulation of past dividends. As shown in Table 2, the high-frequency variation (captured
by $\epsilon_t^{pr}$) is the main reason that $pr_t$ outperforms $pd$ in return forecasting.

Our sample has three recession periods (shaded). Near the end of recessions, 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 expected return from $pr_t$ than that from $pr_t$. Another interesting fact is that in the year leading up to 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. While 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 a proxy for expected return are very informative for constructing macro-finance models that speaks to both asset pricing and business-cycle fluctuations.

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 alternative predictors considered in the literature. In the caption, we document the sources of these predictors. In particular, we explore forecasts from 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
Figure 3: **Comparison with Alternative Return Predictors.** This graph 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).
in statistical significance (e.g., ik, the investment-capital ratio of Cochrane (1991)). In the appendix we report the correlation between \( pr_t \) and the alternative predictors. Aside from \( pd_t, bm, ik, \) and dy show significant correlation with \( pr_t \).

### 2.3 Predicting dividend growth

As shown in Equation (8), 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., \( \epsilon_{pd}^t \)) should forecast dividend growth. We measure dividend growth by the ratio of adjacent, non-overlapping cumulative dividends,

\[
\Delta D_{t,t+12} = \frac{\sum_{i=1}^{12} D_{t+i}}{\sum_{i=1}^{12} D_{t-12+i}}. \tag{13}
\]

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.\(^{11}\) Column (1) shows that \( \epsilon_{pd}^t \), 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 \( \epsilon_{pd}^t \) 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 \( pr_t \) is constructed to capture the information about expected return. Column (3) shows that in comparison with \( \epsilon_{pd}^t \), the dividend predictive power of \( pr_t \) is weaker, with an out-of-sample \( R^2 \)s only 15% of the out-of-sample \( R^2 \) of \( \epsilon_{pd}^t \). Thus, the decomposition of \( pd_t \) into \( pr_t \) and \( \epsilon_{pd}^t \) adequately separates the information on expected return and dividend growth.

\(^{10}\)Dividends are calculated from the difference between cum- and ex-dividend S&P index levels.

\(^{11}\)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)).
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 (14). We consider four the right-hand side variables (i.e., predictors), the residuals of \( p_d_t \) after regressing on \( p_r_t \) (\( \epsilon^{pd}_t \)), \( p_d_t \), \( p_r_t \), the equity yield (\( \ln \left( \frac{D_{t+1}}{P_{t+1}} \right) \)), and the results are reported in Column (1) to (4) respectively. The estimated predictive coefficient (\( \beta \)) 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)).

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>( \epsilon^{pd}_t )</th>
<th>( p_d_t )</th>
<th>( p_r_t )</th>
<th>( \ln \left( \frac{D_{t+1}}{P_{t+1}} \right) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newey-West t</td>
<td>(3.204)</td>
<td>(0.247)</td>
<td>(-2.005)</td>
<td>(-3.395)</td>
</tr>
<tr>
<td>Hodrick t</td>
<td>[5.153]</td>
<td>[0.642]</td>
<td>[-3.990]</td>
<td>[-6.767]</td>
</tr>
<tr>
<td>Stambaugh bias adjusted ( \beta )</td>
<td>0.316</td>
<td>0.025</td>
<td>-0.024</td>
<td>-0.118</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.349</td>
<td>0.003</td>
<td>0.057</td>
<td>0.233</td>
</tr>
<tr>
<td>OOS ( R^2 )</td>
<td>0.304</td>
<td>0.045</td>
<td>0.046</td>
<td>0.222</td>
</tr>
<tr>
<td>( p(ENC) )</td>
<td>&lt; 0.01</td>
<td>&gt; 0.10</td>
<td>&lt; 0.10</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>( p(CW) )</td>
<td>0.011</td>
<td>0.418</td>
<td>0.054</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Our analysis of return and dividend predictability echoes the observation of Cochrane (2007) that price-dividend ratio must either predict return or dividend growth. Our results show an even richer story: the predictive information on return and dividend cancels out each other within \( p_d_t \). Once we distill the information on future return, the rest of \( p_d_t \) has a much stronger dividend forecasting power than \( p_d_t \) itself. The intuition behind this result is related to the countercyclicality of expected return. When the economy is booming and dividend grows fast, expected return tends to be low, and when dividend grow slowly, expected return tends to be high. Through the lens of our state space model, the factors that drive expected return and dividend growth have strong correlation with each other, so when combined together in \( p_d_t \), they cancel each other out.

\( \epsilon^{pd}_t \) is closely related to the “equity yield” in Binsbergen, Hueskes, Koijen, and Vrugt (2013), especially its dividend predictive power. Following that paper, we define equity yield
as \( \ln \left( \frac{D_t}{P_t^{\pi}} \right) \), i.e., the log ratio of past dividend to short-term dividend price. The following equation directly decomposes \( pd_t \) into \( pr_t \) and the equity yield:

\[
pd_t = \ln \left( 1 + e^{pr_t} \right) - \ln \left( \frac{D_t}{P_t^{\pi}} \right) \approx \kappa_0 + \kappa_1 pr_t - \kappa_2 \ln \left( \frac{D_t}{P_t^{\pi}} \right),
\]

where the linearization coefficients are \( \kappa_1 = \frac{\exp(pr_t)}{1+\exp(pr_t)} \), \( \kappa_2 = 1 \), and \( \kappa_0 = \ln \left( 1 + \exp (pr_t) \right) - \kappa_2 pr_t \).

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 (14) 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 \( \epsilon_t^{pd} \).

### 2.4 Structural model: return predictability and cash flow process

The economic intuition behind our results on return predictability can be understood by imposing more structure on the state space model. Specifically, we construct a model of 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. As will be shown later, this model is nested by the general state space model that motivates our decomposition of the price-dividend ratio. To streamline the exposition, here we define one unit of time as one year, instead of one month as in the empirical exercises.

**Model structure.** The economy has three independent shocks: a shock to dividend growth, a shock to expected dividend growth, and a preference shock. A \( 3 \times 1 \) vector \( \varepsilon_{t+1} \) record these standard normal shocks. The aggregate dividend is assumed to evolve according to

\[
\Delta d_{t+1} = g + z_t + \sigma_d \varepsilon_{t+1},
\]

\[\text{\textsuperscript{12}}\text{Our sample length differs from Binsbergen, Hueskes, Koijen, and Vrugt (2013) (October 2002 and April 2011), who use dividend derivatives to derive short-term dividend prices, so the estimates of predictive coefficient are different.}
where $z_t$ follows the AR(1) process

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

with $|\phi_z| < 1$. The conditional mean of dividend growth is $g + z_t$. Row vectors $\sigma_d$ and $\sigma_z$ multiply the shocks on dividend growth and $z_{t+1}$. The conditional standard deviation of $\Delta d_{t+1}$ 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 $\Delta d_{t+1}$ 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

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

where $|\phi_x| < 1$ and $\sigma_x$ is a $1 \times 3$ 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 et al. (2004), we assume that 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 equals

$$M_{t+1} = \exp\{-r^f - \frac{1}{2} x_t^2 - x_t \varepsilon_{d,t+1}\}.\quad (18)$$

$\ln E_t [M_{t+1}] = -r^f$, 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.

Solving the price ratio and expected return. We solve $pr_t$, which equals $\ln \left( \frac{S_t}{P_t} - 1 \right)$. 

23
First, we solve $P^T_t$, the price of one-year dividend following Lettau and Wachter (2007):

$$P^T_t = D_t \exp \{ A^T - B^T x_t + C^T z_t \}, \quad (19)$$

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., $\kappa_0 + \kappa_1 p_d t + \Delta d t + 1$, and no-arbitrage condition (details in the Appendix):

$$S_t = D_t \exp \{ A + B x_t + C z_t \}, \quad (20)$$

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

$$B = -\frac{\sigma_d \sigma_x^T}{\sigma_d^2} \kappa_1 C + \|\sigma_d\|, \quad \text{and} \quad C = \frac{1}{1 - \kappa_1 \phi_x}. \quad (13)$$

The return predictor $p_{r_t}$ is a function of $S_t/P^{T-}_t$, which in turn depends on $x_t$ and $z_t$:

$$\frac{S_t}{P^{T-}_t} = \exp \{ A - A^T + (B - B^T) x_t + (C - C^T) z_t \}. \quad (21)$$

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

$$\mathbb{E}_t [r_{t+1}] = A^r + B (\kappa_1 \phi_x - 1) x_t, \quad (22)$$

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

\textsuperscript{13}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.
Equation (15) and (22) 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.

**Source of return predictability.** From Equation (22), we know that to predict market return, all we need is a proxy for \( x_t \). The fact that \( pr_t \) outperforms \( pd_t \) in return prediction suggests that \( pr_t \) is a better signal of \( x_t \). \( pd_t \) and \( pr_t \) are functions of \( x_t \) and \( z_t \) (expected dividend growth). It has been shown that information regarding future dividends compromises the return predictive power of price-dividend ratio (e.g., Menzly, Santos, and Veronesi (2004); Lettau and Ludvigson (2005)). For example, in the calibration of Lettau and Wachter (2007), the shock correlation between \( x_t \) and \( z_t \) is zero, so \( z_t \) adds pure noise. Therefore, \( pr_t \) is likely to be a better return predictor, if it is less responsive to changes in \( z_t \) than \( pd_t \) is. In fact, when the persistence parameter of expected dividend growth, \( \phi_z \), is zero, \( pr_t \) is a perfect signal of \( x_t \), i.e., a function of \( x_t \) only, because \( \phi_z = 0 \) implies \( C = 1 \) so \( C - C T^- \), the coefficient of \( z_t \) in \( S_t / P_t \), is zero.

**Proposition 1** In an economy with stochastic discount factor given by Equation (18) and aggregate dividend growth given by Equation (15), the price ratio, \( pr_t \), is a function of only the price of risk \( x_t \) if and only if the expected dividend growth is not persistent, i.e., \( \phi_z = 0 \).

In Table 3, we show that \( \epsilon_{pd}^t \), the residuals from regressing \( pd_t \) on \( pr_t \), predicts dividend growth with an out-of-sample \( R^2 \) of 30.4%. This suggests once we take the variation of \( pr_t \) out of \( pd_t \), what remains is a strong signal of expected dividend growth. When \( \phi_z = 0 \) and \( pr_t \) only varies with \( x_t \), \( \epsilon_{pd}^t \) precisely captures the variation of \( pd_t \) that is solely driven by \( z_t \), the expected dividend growth in our structural model, and thus, should predict dividends.

Therefore, through the lens of the model, our finding of strong dividend predictive power of \( \epsilon_{pd}^t \) lends further support to the use of \( pr_t \) as a proxy for the price of risk (and expected return), and that the persistence of expected dividend growth is likely to be weak (i.e., \( \phi_z \) close to zero).
Evidence. In Table 15 in the Appendix, we directly fit the state space model to aggregate dividends, and find the null hypothesis of $\rho = 0$ cannot be rejected (using both S&P 500 dividends and total cash payout of CRSP NYSE/AMEX/NASDAQ Cap-Based index).\footnote{We consider different correlation structure of shocks to expected return and expected cash flow in Figure 12a and 12b.} Figure 13 shows that autocorrelations of $\epsilon_{t}^{pd}$, which is our proxy of cash-flow expectation, are not statistically different from zero across countries, which prelude our findings on return predictability outside the United States in the next subsection.

Discussion. Several discussions are in order. $pd_{t}$ does not predict dividends, which is in line with the existing literature (e.g., Cochrane (2011)). Only after the variation of $x_{t}$ is taken out of $pd_{t}$, can its residual predict dividends. Moreover, while $\ln \left( \frac{P_{t}^{T-}}{D_{t}} \right)$ predicts dividends, its power is weaker than $\epsilon_{t}^{pd}$, which is quite intuitive because, in the structural model, $\ln \left( \frac{P_{t}^{T-}}{D_{t}} \right)$ equals $AT^{-} + BT^{-}x_{t} + CT^{-}z_{t}$, so the variation of $z_{t}$ is masked by the variation of $x_{t}$.

We can further appreciate the critical implications of dividend structure on return predictability by fixing $z_{t}$ to zero or any constant in the structural model. In such a case, dividend growth is i.i.d., so both our price ratio, $pr_{t}$, and the traditional price-dividend ratio, $pd_{t}$, vary with $x_{t}$ only, and both should predict return equally well. Many have proposed variants of growth-adjusted valuation ratios as return predictor (Campbell and Thompson (2008); Lacerda and Santa-Clara (2010); Da, Jagannathan, and Shen (2014); Golez (2014)). Binsbergen and Koijen (2010) and Rytchkov (2012) explicitly use state-space models to filter out and separate the information on expected return and dividend growth. We contribute to this line of research by proposing a model-free return predictor that can be obtained directly from market prices. Prices respond to news more promptly than fundamentals, and thus, $pr_{t}$ reveals high-frequency variation in the expected return.

Note that the dividend predictive power of $pr_{t}$ is stronger than $pd_{t}$ (marginally significant coefficient and an out-of-sample $R^{2}$ equal to 4.5\%), albeit that it is one magnitude
below than $\epsilon_t^{pd}$'s. This can be easily understood within our framework. Even if $pr_t$ is a function of $x_t$ only, any shock correlation between $x_t$ and $z_t$ (i.e., $\sigma_x \sigma'_x \neq 0$) implies that unconditionally, $pr_t$ is correlated with expected dividend growth with the signal of correlation depending on $\sigma_x$, $\sigma_z$, $\phi_x$, $\phi_z$, and $\sigma_d$ that affect $pr_t$’s loading on state variables and the unconditional correlation between $x_t$ and $z_t$.

### 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 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 $\epsilon_t^{pr}$ and $\epsilon_t^{pd}$ country-by-country.

**International return predictability.** We supplements 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

\(^{15}\)In the appendix, Figure 11 plots the futures-to-spot ratio for these four countries.
Table 4: International Panel Return Prediction

This table reports the results of return forecasting regression (Equation (23)) 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), $\epsilon_{pr}^t$ (Column 5 and 6), and $\epsilon_{pd}^t$ (Column 7 and 8) in that country. $\epsilon_{pr}^t$ is the residual of $pr_t$ after regressing on $pd_t$, and $\epsilon_{pd}^t$ 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 ($\beta$) 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 $pd_t$.

<table>
<thead>
<tr>
<th>$pr_t$</th>
<th>-0.189</th>
<th>-0.051</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[-4.497]</td>
<td>[-0.740]</td>
</tr>
<tr>
<td>$pd_t$</td>
<td>-0.109</td>
<td>-0.135</td>
</tr>
<tr>
<td></td>
<td>[-3.843]</td>
<td>[-1.339]</td>
</tr>
<tr>
<td>$\epsilon_{pr}^t$</td>
<td>-0.404</td>
<td>-0.113</td>
</tr>
<tr>
<td></td>
<td>[-7.087]</td>
<td>[-1.438]</td>
</tr>
<tr>
<td>$\epsilon_{pd}^t$</td>
<td>-0.063</td>
<td>-0.083</td>
</tr>
<tr>
<td></td>
<td>[-1.961]</td>
<td>[-1.230]</td>
</tr>
</tbody>
</table>

| Country FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Time FE    | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Obs        | 1,469 | 1,469 | 1,553 | 1,553 | 1,469 | 1,469 | 1,469 | 1,469 |

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.

$$12 r_{t,t+1}^n = \alpha + \beta \left( \frac{1}{12} \sum_{i=0}^{11} x_{t-i}^n \right) + \epsilon_{t,t+1}^n, \quad (23)$$

where $n$ represents a country.\textsuperscript{16} The dependent variable is no longer overlapping. It is now the (annualized) next-month market return, and the predictor, for example $pr_t$, is averaged in the most recent twelve months. Hodrick (1992) points out the difficulties in inference when using overlapping observations, especially the poor small-sample properties of GMM-based autocovariances correction (e.g., Newey-West standard error), and suggest the regression in Equation (23) for drawing inference on long-run return prediction.\textsuperscript{17} We also cluster the

\textsuperscript{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.

\textsuperscript{17}Note that the adjusted $R^2$ from the non-overlapping regression of Equation (23) is not comparable to that of the overlapping regression in Table 2, because in Equation (23), we effectively forecast monthly
standard error by time and country. Therefore, the specification of Equation (23) combines the better small-sample properties of Hodrick (1992) standard error and the advantage of clustered standard errors in panel regression that are robust to cross-country and within-country (time-series) correlation in the error term.

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, $\epsilon^{pd}_t$ is the residuals of $pd_t$ after regressing on $pr_t$), albeit that in comparison with U.S. results, $pd_t$ seems to have some distinct information on future returns.

**International comovement in the expected return.** We introduce time fixed effect in Column (2) that absorbs a global factor from realized 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 that is common across countries. This finding suggests that the variations of expected return across countries tend to comove, which is in line with the literature of global equity market integration (Miranda-Agrippino and Rey (2015)). And in Column (4), we see that any return predictive power of $pd_t$ also disappears once the global factor is absorbed by the time fixed effect. Similar result holds in Column (6) for the regression residuals of $pr_t$ with respect to $pd_t$.

Figure 10 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 variation in $pr_t$. The first principal component (48% of variation) exhibits spikes at the onsets of the global return using the one-year average of predictor, even if the inference we draw from such regression is about the return prediction at one-year horizon. Thus, we do not report the $R^2$ of non-overlapping regression.
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.

**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 Asymmetric Predictability

In this section, we study conditional return predictability, and explore the economic mecha-
nism behind predictability. We find that predictability is asymmetric – stronger following a
down market (Table 5). The results are similar outside the United States (Figure 5). This
suggests the instability of structural parameters in the model of Lettau and Wachter (2007)
that we study previously. Therefore, we consider alternative asset pricing models that imply strong asymmetry in return predictability, and evaluate these models by using their proposed state variables as conditioning variables (Table 6).

### 3.1 Asymmetric return predictability: evidence

**Conditional return prediction.** We decompose $p_{rt}$ into two components: (1) $I_{(r_{t-12,t} < 0)} \times p_{rt}$, the interaction between $p_{rt}$ 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) $I_{(r_{t-12,t} \geq 0)} \times p_{rt}$, the interaction between $p_{rt}$ and the up-market indicator. The return forecasting model is

$$r_{t,t+12} = \alpha + \beta_D I_{(r_{t-12,t} < r_{t-12,t}^f)} \times p_{rt} + \beta_U I_{(r_{t-12,t} \geq r_{t-12,t}^f)} \times p_{rt} + \epsilon_{t,t+12}. \quad (24)$$

Thus, the return predictive power of $p_{rt}$ following a down market is reflected in $\beta_D$, and the return predictive power following an up market is reflect in $\beta_U$.

Column (1) of Table 5 reports the regression results. Following a down market, $p_{rt}$ 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, $\beta_D$ is almost twice $\beta_U$ in both magnitude and the t-statistic. The midpoint between $\beta_D$ and $\beta_U$ is very close to the coefficient of $p_{rt}$ as univariate predictor (Table 5). This decomposition by previous market condition reveals a sharp asymmetry in return predictability.

Column (2) and (3) of Table 5 show that the down-market indicator itself does not predict future returns or together with $p_{rt}$. When both the down-market indicator and $p_{rt}$ are used as predictors, the predictive coefficient on $p_{rt}$ is almost identical to the predictive coefficient in the univariate regression, and the t-statistics and $R^2$s are almost identical.
Table 5: 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 (24). On the right-hand side are 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 table). Column (4) reports the results of the specification of Equation (25). 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 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 \( \beta \) 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.

<table>
<thead>
<tr>
<th>( I_{{r_{t-12,t}&lt;r_{t-12,t}^f}} \times pr_t )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Newey-West t )</td>
<td>-0.180</td>
<td>-3.810</td>
<td>-2.977</td>
<td>-2.735</td>
<td>-2.776</td>
<td>-2.739</td>
</tr>
<tr>
<td>( Hodrick t )</td>
<td>-0.108</td>
<td>-2.981</td>
<td>-1.751</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( I_{{r_{t-12,t} \geq r_{t-12,t}^f}} \times pr_t )</td>
<td>0.257</td>
<td>-0.038</td>
<td>-0.031</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( pr_t )</td>
<td>-0.137</td>
<td>-1.358</td>
<td>-5.496</td>
<td>-4.949</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \left(r_{t-12,t} - r_{t-12,t}^f\right) \times pr_t )</td>
<td>0.269</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( r_{t-12,t} - r_{t-12,t}^f )</td>
<td>-1.037</td>
<td>0.083</td>
<td>0.065</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.261</td>
<td>0.012</td>
<td>0.246</td>
<td>0.264</td>
<td>0.008</td>
<td>0.243</td>
</tr>
</tbody>
</table>

Column (4) of Table 5 reports the results of an alternative specification,

\[
\begin{align*}
\tilde{r}_{t,t+12} = \alpha + \beta pr_t + \rho_0 \left(r_{t-12,t} - r_{t-12,t}^f\right) + \rho_1 \left(r_{t-12,t} - r_{t-12,t}^f\right) \times pr_t + \epsilon_{t,t+12}.
\end{align*}
\]

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 or reversal.** The regression of Equation (25) shows that return autocorrelation is conditional on \( pr_t \). This is related to papers 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. But 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 5, 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 return momentum. When \( pr_t \) is one-standard deviation below its mean, the autocorrelation coefficient is \(-0.106\), exhibiting return reversal. Campbell, Grossman, and Wang (1993) find that daily return autocorrelation depends on trading volume. Our results suggest that at one-year horizon, return autocorrelation depends on the relative valuation of long-term vs. short-term dividends.

**Conditional predictability outside the United States.** Figure 5 reports the results of conditional prediction (regression of Equation (24)) for different countries. The candle graph shows the estimates of \( \beta_D \) (red) and \( \beta_U \) (blue) together 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 downturns.\(^{18}\) Henkel, Martin, and Nardari (2011) show that the return predictive power of price-

\(^{18}\)Cujean and Hasler (2017) build an equilibrium model with counter-cyclical investors’ disagreement to
Figure 5: **Conditional predictability across countries.** Figure 5 reports the results of conditional prediction (regression of Equation (24)) for different countries. The candle graph shows the estimates of $\beta_D$ (red) and $\beta_U$ (blue) together with the one Hodrick (1992) standard error band.

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, return is more predictable. Dangl and Halling (2012) propose a dynamic prediction model with time-varying coefficient to account for conditional predictability. We use past return as conditioning variable instead because in our sample of less than thirty years, there are not many business cycles. 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.

### 3.2 Asymmetric return predictability: related theories

Next, we evaluate two theories based on financial intermediation friction (He and Krishnamurthy (2013)) and behavioral bias (Barberis, Huang, and Santos (2001)) that produce asymmetry in return predictability.

**Intermediary asset pricing.** Panel A of Figure 6 is from He and Krishnamurthy (2013) explain why stock return predictability is concentrated in bad times.
Figure 6: 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.

(Their Panel A of Figure 2). It plots the risk premium against the state variable, 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 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 relatively richer, their delegation capacity is sufficient to satisfy the needs of household, and risk premium varies with the aggregate wealth of the economy, showing little variation (“unconstrained region”). When intermediaries are relatively 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. Our down-market indicator, which spans a period of one year, is closely related to 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.
Table 6: 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 $\eta_t$ 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 $\eta_t$ constructed by He, Kelly, and Manela (2017) whose sample ends in 2012.

<table>
<thead>
<tr>
<th></th>
<th>$r_{t-12,t} - r_{t-12,t}^{f}$</th>
<th>$\eta_t - \bar{\eta}_t$</th>
<th>$z - 1$</th>
<th>$r_{t-12,t} - r_{t-12,t}^{f}$</th>
<th>$\eta_t - \bar{\eta}_t$</th>
<th>$z - 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>$-0.180$</td>
<td>$-0.247$</td>
<td>$-0.195$</td>
<td>$-0.108$</td>
<td>$-0.152$</td>
<td>$-0.101$</td>
</tr>
<tr>
<td></td>
<td>Newey-West $t$</td>
<td>(-3.810)</td>
<td>(-2.340)</td>
<td>(-5.221)</td>
<td>(-2.977)</td>
<td>(-2.101)</td>
</tr>
<tr>
<td>Positive</td>
<td>$-0.108$</td>
<td>$-0.152$</td>
<td>$-0.101$</td>
<td>$-0.291$</td>
<td>$-4.549$</td>
<td>(-3.133)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[-1.751]</td>
<td>[-2.111]</td>
<td>[-1.714]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>$0.257$</td>
<td>$0.279$</td>
<td>$0.382$</td>
<td>$-0.038$</td>
<td>$0.023$</td>
<td>$0.020$</td>
</tr>
<tr>
<td></td>
<td>(1.000)</td>
<td>(0.697)</td>
<td>(2.240)</td>
<td>(-0.750)</td>
<td>(0.473)</td>
<td>(0.541)</td>
</tr>
<tr>
<td></td>
<td>[0.987]</td>
<td>[0.549]</td>
<td>[1.296]</td>
<td>[-0.670]</td>
<td>[0.455]</td>
<td>[0.441]</td>
</tr>
<tr>
<td>$R^2$</td>
<td>$0.261$</td>
<td>$0.314$</td>
<td>$0.265$</td>
<td>$0.012$</td>
<td>$0.006$</td>
<td>$0.004$</td>
</tr>
</tbody>
</table>

Prospect theory. Panel B of Figure 6 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 return 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 gains, he becomes more less loss averse: the prior gains will cushion any subsequent loss, making it more bearable. Conversely, after a prior loss, he becomes more loss averse: after being burned by the initial loss, he is more sensitive to additional setbacks.” Panel B of Figure 6 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.

Evaluation of related theories. Table 6 compares our results of conditional return pre-
diction 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., $\bar{\eta}$) 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 5.

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. The predictive coefficient on the interaction between negative indicator and $pr_t$ is twice as large as the coefficient of the interaction between positive indicator and $pr_t$, and all three specifications in Column (1), (2), and (3) have adjusted $R^2$ of similar magnitude. However, both theories predict that the conditioning variable itself (captured by the negative indicator variable) should also predict returns, which is not the case in data as shown by Column (5) and (6).

4 Variation in Expected Return

Our findings so far suggest a simple asset pricing model is unlikely to generate the expected return dynamics we observe. In this section, we use $pr_t$ as forecasting variable to study the variation of expected return, and examine how it responds to monetary policy shocks (Table 7) and comoves with macroeconomic and financial market conditions (Table 8). Next, since shocks to $pr_t$ drives the variation of expected return, and thus, the investment opportunity set, they should be priced in the cross-section of stocks by the logic of ICAPM. As an economic test of $pr_t$ as a valid return predictor, we find a significant and negative price of $pr_t$ risk (Table 9). High $pr$-beta stocks, i.e., those that have high realized returns when $pr_t$ is high (the expected market return low), have low average returns.
4.1 Monetary policy and macroeconomy

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 a 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.

So far, we have shown that $pr_t$ strongly predicts return at one-year horizon. Next, using $pr_t$ as a proxy of the expected return (or “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 existing literature. We also document significant correlations between the expected returns from our predictive regression and various macroeconomic variables. Our findings suggest that both monetary policy and business cycle fluctuations are important drivers of the expected return in the stock market.

**Monetary policy announcements 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 a FOMC meeting and zero otherwise. We also construct a variable, Pre-FOMC day, that equals one if the next day has a FOMC meeting and zero otherwise, and another variable, Post-FOMC day, that equals one if it is the previous day has a 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.\(^\text{19}\)

\(^{19}\)Monetary policy shock is calculated using a 30-minute window from 10 minutes before the FOMC
Table 7: 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, $pr_t$, the residuals from regressing $pr_t$ on $pd_t$ (i.e., $\epsilon_{pr}^d$), and the residuals from regressing $pd_t$ on $pr_t$ (i.e., $\epsilon_{pd}^d$) 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.

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

Table 7 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 7 shows a tightening of monetary policy (i.e., an increase in MP Shock) leads to a decrease in stock market returns, in line with the evidence in Thorbecke (1997) and Bernanke and Kuttner (2005). After controlling for the monetary policy shock itself, the relation between stock return and FOMC day is weakened. It is a long tradition to 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.
understand the contemporaneous response of stock price to monetary policy. Rozeff (1974) finds that a substantial fraction of the variation in stock returns is related to contemporaneous monetary developments.

Our focus is on the regression of $\text{pr}_t$ on the monetary policy variables, because $\text{pr}_t$ serves as a proxy for the expected stock return (i.e., the discount rate in the stock market). Column (3) of Table 7 shows a negative response of $\text{pr}_t$ to monetary tightening, which translates into an increase in the expected return. In other words, unanticipated increase in the Federal Funds rate tends to raise the expected stock return. Therefore, the decline of stock price, i.e., the decrease of contemporaneous stock return, can be largely attributed to the rising discount rate. Column (6) delivers the same message using the residuals of $\text{pr}_t$ from regressing $\text{pr}_t$ on $\text{pd}_t$. 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)) because the response of $\epsilon_t^{pr}$ to monetary policy shock is missed. Thus, our new return predictor $\text{pr}_t$ contains essential information about how the expected return varies with monetary policy. Finally, since the residuals from regressing $\text{pd}_t$ on $\text{pr}_t$ (i.e., $\epsilon_t^{pd}$) strongly forecast dividend growth (Table 3), we regress it on monetary policy variables. We do not find significant relations.

In sum, monetary policy has strong impact on stock prices mainly through the discount rate instead of cash flow growth – 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 in the long run.

**Correlation with other variables.** The expected stock return is highly correlated with variables that reflect the conditions of macroeconomy, financial markets, financial intermediaries, uncertainty, and sentiment (Table 8). Two versions of the expected returns are the

---

20 Our results are related to Patelis (1997) who documents some return predictive power of monetary policy variables.
Table 8: Correlations between Macro Variables and the Expected Return

This table reports the correlation between the in-sample fitted expected returns of unconditional and conditional models and macroeconomic 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).

<table>
<thead>
<tr>
<th>Variable</th>
<th>( \hat{r}_{pr} )</th>
<th>p-value</th>
<th>( \hat{r}_{pr,cond} )</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Macroeconomic:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDP Growth</td>
<td>0.08 (0.38)</td>
<td>0.19</td>
<td>(0.04)</td>
<td></td>
</tr>
<tr>
<td>IP Growth</td>
<td>0.03 (0.48)</td>
<td>0.16</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>CFNAI</td>
<td>0.07 (0.18)</td>
<td>0.17</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>Unemployment</td>
<td>0.38 (0.00)</td>
<td>0.33</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>Cons. Growth</td>
<td>-0.43 (0.00)</td>
<td>-0.25</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>Business Inventories</td>
<td>-0.08 (0.14)</td>
<td>-0.05</td>
<td>(0.35)</td>
<td></td>
</tr>
<tr>
<td>Nonres. Fixed Investment</td>
<td>-0.47 (0.00)</td>
<td>-0.44</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>Res. Fixed Investment</td>
<td>-0.31 (0.01)</td>
<td>-0.24</td>
<td>(0.05)</td>
<td></td>
</tr>
<tr>
<td>GDP Deflator</td>
<td>-0.35 (0.00)</td>
<td>-0.33</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td><strong>Financial:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term Spread</td>
<td>0.27 (0.00)</td>
<td>0.21</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>Baa-Aaa</td>
<td>0.09 (0.07)</td>
<td>0.01</td>
<td>(0.89)</td>
<td></td>
</tr>
<tr>
<td>cay</td>
<td>0.29 (0.00)</td>
<td>0.22</td>
<td>(0.02)</td>
<td></td>
</tr>
<tr>
<td><strong>Intermediary:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B/D Leverage</td>
<td>-0.55 (0.00)</td>
<td>-0.62</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>B/D 1 Year Avg. CDS</td>
<td>0.21 (0.02)</td>
<td>-0.26</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>B/D 5 Year Avg. CDS</td>
<td>0.32 (0.00)</td>
<td>-0.16</td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>ROA Banks</td>
<td>-0.43 (0.00)</td>
<td>-0.36</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td><strong>Uncertainties:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VIX</td>
<td>-0.25 (0.00)</td>
<td>-0.31</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>EPU</td>
<td>0.17 (0.00)</td>
<td>0.09</td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>CP Recession</td>
<td>-0.03 (0.47)</td>
<td>-0.12</td>
<td>(0.01)</td>
<td></td>
</tr>
<tr>
<td>SPF Recession</td>
<td>0.13 (0.17)</td>
<td>-0.01</td>
<td>(0.95)</td>
<td></td>
</tr>
<tr>
<td><strong>Sentiments:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sentiment Index</td>
<td>-0.43 (0.00)</td>
<td>-0.51</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>Sentiment Index (orth.)</td>
<td>-0.44 (0.00)</td>
<td>-0.52</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>IPO #</td>
<td>0.10 (0.05)</td>
<td>0.17</td>
<td>(0.00)</td>
<td></td>
</tr>
<tr>
<td>Close-end Discount</td>
<td>0.32 (0.00)</td>
<td>0.36</td>
<td>(0.00)</td>
<td></td>
</tr>
</tbody>
</table>
fitted values from unconditional (Equation (12)) and conditional (Equation (24)) predictive regressions respectively. Both versions deliver a consistent message.

The expected return is countercyclical. It is positively correlated with unemployment, and negatively correlated with consumption growth, fixed investments, and GDP deflator, suggesting that a major fraction of variation in the expected return is from 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); Tobias Adrian and Shin (2013)). This indicates that when 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 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

\(^{21}\)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.
Our finding is closely related to Moreira and Muir (2017), who show that a trading strategy that scales up when the expected volatility declines tends to generate profits unexplained by common risk factors. 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).

4.2 \( pr \) risk price

**Revisiting ICAPM.** Investment opportunity set varies over time. Particularly, expected stock market return varies with \( pr_t \). Therefore, shocks to \( pr_t \) affects investors’ marginal value of wealth and consumption through their impact on the investment opportunity.

When \( pr_t \) is high and the expected stock return is low, substitution effect suggests that investors allocate more wealth to consumption rather than savings. Thus, positive \( pr \) shocks lead to higher consumption and lower marginal utility of consumption and wealth. Assets with high loadings on \( pr \) shocks (“high \( pr \)-beta”) pay off when wealth is less valued, so they deliver a high average return in equilibrium. In other words, the price of \( pr \) risk is positive.

However, the deterioration of investment opportunity set (high \( pr_t \)) implies that investors become relatively poorer over time, so wealth effect suggests that investors consume less in the current period. Thus, positive \( pr \) shocks lead to lower consumption and higher marginal utility of consumption and wealth. High \( pr \)-beta assets pay off when additional consumption is highly valued, so they deliver a low average return in equilibrium. The price of \( pr \) risk is negative.

Whether substitution or wealth effect dominates, and whether consumption rises or

---

\(^{22}\)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)).
falls, depends on investors’ intertemporal elasticity of substitution. But either way, shocks to \( pr_t \) affect consumption and the marginal value of wealth, so they should be priced in the cross section of assets. Next, we estimate the price of \( pr \) risk in the cross-section of standard sorted portfolios (i.e., size, value, momentum, investment, and profitability).

Any return predictor should go through this cross-section test. If predictive power is not from spurious relations, shocks to the predictor are shocks to investment opportunity set, and should be priced in the cross section. However, few studies on return predictors conduct this economic test. In this paper, we go beyond statistical procedures, such as standard error correction, bias adjustments, and out-of-sample test. By estimating the price of \( pr \) risk, we test \( pr_t \) as a return predictor using the simple logic of ICAPM (Merton (1973)).

**Estimating \( pr \) risk price.** Our cross-section of assets are the twenty-five Fama-French portfolios (sorted by market equity and book-to-market ratio), twenty-five momentum portfolios (sorted by market equity and prior returns), twenty-five investment portfolios (sorted by market equity and change in total assets), and twenty-five profitability portfolios (sorted by market equity and operating profitability). The data of monthly portfolio returns are from Kenneth R. French’s website.\(^{24}\) We consider this set of portfolios as good proxy for the U.S. stock universe.

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 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 correlated with expected asset returns, because it forecasts market return and expected asset returns are functions of expected market return by CAPM. 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 \( pr \) shocks, we use two procedures. The first is the Fama-
Table 9: Risk Prices

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 \( R^2 \).

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

MacBeth method. The second one is GMM, which corrects potential biases in Fama-MacBeth standard errors (Cochrane (2005)). The parameters are over-identified. For the weight matrix, we use the two-stage efficient weight matrix (Hayashi (2000)). In both 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 (Merton (1973)).

Table 9 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 with size. The price of risk is reported
for both $pr$ shock ($\Delta pr_t$) and market excess return ($r_t - r^f_t$), followed by the t-statistic. We also report the mean absolute pricing error (MAPE) and $R^2$.

The price of $pr$ 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 $pr$ 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 or proxy for expected market return.

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

5 Conclusion

We find strong return predictive power of $pr_t$, the ratio of long-term dividend price to short-term dividend price. It outperforms alternative predictors in terms of statistical significance and out-of-sample performances. Its return predictive power holds in the international sample and reflects the global comovement in the expected stock returns.

Our structural model relates the return predictive power of $pr_t$ to the persistence of expected dividend growth, a key parameter in asset pricing models. The model also sheds light on why the traditional price-dividend ratio underperforms $pr_t$ in return prediction, and why after orthogonalized to $pr_t$, $pd_t$ residuals strongly predict dividend growth.

We also find the return predictive power of $pr_t$ is asymmetric – stronger following a down market. This is consistent with asset pricing theories that emphasize financial or
behavioral frictions. The asymmetry of return predictability has implications on the long-standing debate on stock return autocorrelation – whether the market exhibits time series momentum or reversal can depend on $pr_t$.

Using $pr_t$ as forecasting variable, we find the expected return responds strongly to monetary policy shocks. Through the expected return, monetary policy affects stock price. We find that such impact is likely to be transitory (reverting back in one year). The expected return, from both unconditional and conditional predictive regression, exhibits strong correlation with macroeconomic variables, intermediary balance-sheet, uncertainty, and sentiment. Asset pricing theories based on a subset of those elements can capture only part of the variation in expected return.

Finally, we find shocks to $pr_t$ is priced in the cross-section of stocks. The economically and statistically significant price of $pr$ risk is consistent with the implication of ICAPM that shocks to investment opportunity set, and the expected return in particular, should be priced in equilibrium. This economic test lends extra support to the return predictive power of $pr_t$. 

47
Appendix I: Derivation

I.1 Derivation of the state space model

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

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

By law of iterated expectation, we can replace \( \Delta d_{t+j} \) and \( r_{t+j} \) with their time \( t+j-1 \) expectations:

\[ v_t = \frac{\kappa}{1 - \rho} + \sum_{j=1}^{\infty} \rho^{j-1} \mathbb{E}_t [k_{t+j-1} - \mu_{t+j-1}] \]

\[ = \frac{\kappa}{1 - \rho} + \sum_{j=1}^{\infty} \rho^{j-1} \mathbb{E}_t [(\delta_0 + \delta' F_{t+j-1}) - (\gamma_0 + \gamma' F_{t+j-1})] \]

Define \( \phi_0 \) as \( \frac{\kappa + \delta_0 - \gamma_0}{1 - \rho} \), and stack the factor coefficients into \( \psi = (\delta', \gamma') \). Denote the row vector \( (1, -1) \) as \( \iota \). We can rewrite the equation

\[ v_t = \phi_0 + \sum_{j=1}^{\infty} \rho^{j-1} \mathbb{E}_t [\iota \psi' F_{t+j}] \]

\[ = \phi_0 + \sum_{j=1}^{\infty} \rho^{j-1} \mathbb{E}_t [(\iota \psi' A^j) F_t] \]

\[ = \phi_0 + \iota \psi' \left( \sum_{j=1}^{\infty} \rho^{j-1} A^j \right) F_t \]

\[ = \phi_0 + \iota \psi' (1 - \rho A)^{-1} F_t. \]

Define \( \phi' \) as \( \iota \psi' (1 - \rho A)^{-1} \). 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 \( \mu \) is unconditional mean, the predictor \( x_t \) has mean 0 and variance \( \sigma_x^2 \), independent from the error term \( \varepsilon_{t+1} \). 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

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

The variance of market-timing strategy is

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

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

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

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

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 $R^2$,

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

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

$$s_1^2 = A \cdot \frac{\mu^2 + \sigma_x^2}{\sigma_z^2} = A \cdot \frac{\mu^2 + \sigma_x^2}{(\sigma_x^2 + \sigma_z^2)(1 - R^2)} = A \cdot \frac{s_0^2 + R^2}{1 - R^2}.$$

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

$$s_1 = \sqrt{\frac{s_0^2 + R^2}{1 - R^2}}.$$  \hspace{1cm} (26)

### 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

$$pd_t = \ln \left( \frac{S_t}{D_t} \right) = A + Bx_t + Cz_t.$$

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

$$r_{t+1} = \kappa_0 + \kappa_1pd_{t+1} - pd_t + \Delta d_{t+1},$$

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

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

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

where $\Delta d_{t+1} = g + z_t + \sigma_d \varepsilon_{t+1}$ from Equation (15) 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 (17) and (16) respectively. Therefore, we can take all time-$t$ measurable terms outside of the expectation and only leave $t+1$ shocks in it:

$$\mathbb{E}_t \left[ \exp \left\{ -r^f - \frac{1}{2} x_t^2 + \kappa_0 - pd_t + g + z_t + \kappa_1 A + \kappa_1 B (1 - \phi_x) \bar{x} + \kappa_1 B \phi_x x_t + \kappa_1 C \phi_z z_t \right\} \right] = 1.$$

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

$$\mathbb{E}_t \left[ \exp \left\{ -x_t \varepsilon_{d,t+1} + \kappa_1 B \sigma_x \varepsilon_{t+1} + \kappa_1 C \sigma_z \varepsilon_{t+1} + \sigma_d \varepsilon_{t+1} \right\} \right] = 1. \quad (27)$$

Equation (27) holds only if the coefficient on $x_t$ and $z_t$ are zero, so we have the Collect coefficients of $x_t$ equal to

$$-B + \kappa_1 B \phi_x + (\kappa_1 B \sigma_x + \kappa_1 C \sigma_z + \sigma_d) \frac{\sigma_d'}{\| \sigma_d \|} x_t + \frac{1}{2} x_t^2 = 0,$$

and the coefficient on $z_t$ equal to

$$-C + 1 + \kappa_1 \phi_z C = 0.$$
From these two equations, we solve

\[ B = \frac{\sigma_d \sigma'_d \kappa_1 C + \sigma_d \sigma'_d}{1 - \kappa_1 \phi_x - \frac{\sigma_d \sigma'_d}{\|\sigma_d\|} \kappa_1}, \]

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.

\[
\mathbb{E}_t [r_{t+1}] = \kappa_0 + \kappa_1 E_t [pd_{t+1}] - pd_t + E_t [\Delta d_{t+1}]
\]

\[
= \kappa_0 + \kappa_1 A + \kappa_1 B E_t [x_{t+1}] + \kappa_1 C E_t [z_{t+1}] - A - B x_t - C z_t + g + z_t
\]

\[
= \kappa_0 + \kappa_1 A + \kappa_1 B (1 - \phi_x) \bar{x} + \kappa_1 B \phi_x x_t + \kappa_1 C \phi_x z_t - A - B x_t - C z_t + g + z_t
\]

\[
= \left[ \kappa_0 + (\kappa_1 - 1) A + g + \kappa_1 B (1 - \phi_x) \bar{x} \right] + B (\kappa_1 \phi_x - 1) x_t + [C (\kappa_1 \phi_x - 1) + 1] z_t
\]

\[
= \left[ \kappa_0 + (\kappa_1 - 1) A + g + \kappa_1 B (1 - \phi_x) \bar{x} \right] + B (\kappa_1 \phi_x - 1) x_t.
\]

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 7 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
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.
Table 10: One-month Return Prediction

This table reports the results of predictive regression (Equation (12)). 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 $pd_t$ ($ϵ^{pr}_t$), and the residuals of $pd_t$ after regressing on $pr_t$ ($ϵ^{pd}_t$), 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-sample 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)).

<table>
<thead>
<tr>
<th></th>
<th>$pr_t$</th>
<th>$pd_t$</th>
<th>$ϵ^{pr}_t$</th>
<th>$ϵ^{pd}_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$β$</td>
<td>-0.010</td>
<td>-0.015</td>
<td>-0.011</td>
<td>0.004</td>
</tr>
<tr>
<td>Hodrick t</td>
<td>[-2.197]</td>
<td>[-2.006]</td>
<td>[-0.959]</td>
<td>[0.215]</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.017</td>
<td>0.014</td>
<td>0.005</td>
<td>0.000</td>
</tr>
<tr>
<td>OOS $R^2$</td>
<td>0.009</td>
<td>0.007</td>
<td>-0.009</td>
<td>-0.012</td>
</tr>
<tr>
<td>$p(ENC)$</td>
<td>&lt; 0.10</td>
<td>&lt; 0.10</td>
<td>&gt; 0.10</td>
<td>&gt; 0.10</td>
</tr>
<tr>
<td>$p(CW)$</td>
<td>0.078</td>
<td>0.129</td>
<td>0.427</td>
<td>0.170</td>
</tr>
</tbody>
</table>

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 performance 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 8: $p r_t$ from Futures and Option Data. This graph reports $p r_t$ constructed from futures and option data (from Binsbergen, Brandt, and Koijen (2012) from January 1996 and October 2009).
Figure 9: **Predictive Coefficient Stability.** This graph plots the predictive coefficient (and its 95% confidence band in shade) estimated using one-year rolling window of daily observations. The first rolling window ends in December 1988.
Figure 10: **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 11: **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.
Figure 12: \( \phi_z \) estimates from state space model with correlated shocks. This figure reports the expected dividend growth \( \phi_z \) point estimates and t-value in an unrestricted state space model as in Section 2.4 with different correlations of \( \Delta d \) and \( z \) shocks. The correlations of \( \Delta d \) and \( z \) shocks range from -0.9 to 0.9 and the volatility of \( \Delta d \) 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 13: Autocorrelations of cash-flow expectation, $\epsilon_{t}^{pd}$, across countries. This graph plots autocorrelations of $\epsilon_{t}^{pd}$ 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.
## 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 factor 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.

\[
\begin{array}{c|cc}
 & pr & pd \\
\hline
pr & 1.000 & 0.874 \\
pd & 0.874 & 1.000 \\
bm & -0.790 & -0.827 \\
tbl & -0.173 & -0.199 \\
lty & -0.378 & -0.393 \\
ntis & -0.041 & 0.134 \\
infl & -0.108 & -0.073 \\
ltr & 0.005 & -0.042 \\
svar & 0.161 & -0.041 \\
csp & 0.355 & 0.427 \\
cay & -0.384 & -0.381 \\
igi & 0.616 & 0.631 \\
ep & -0.554 & -0.442 \\
de & -0.247 & -0.472 \\
dfy & -0.067 & -0.284 \\
dfr & -0.031 & -0.012 \\
tms & -0.241 & -0.214 \\
dy & -0.883 & -0.989 \\
SII & 0.047 & -0.047 \\
kp & -0.728 & -0.642 \\
SVIX & 0.047 & -0.295 \\
ZCB1Y & -0.205 & -0.215 \\
\end{array}
\]
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 (12)) and conditional (Equation (24)) 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.

<table>
<thead>
<tr>
<th></th>
<th>US</th>
<th>UK</th>
<th>FRA</th>
<th>ESP</th>
<th>JPN</th>
<th>AUS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Unconditional predictions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$pr_t$</td>
<td>-0.138</td>
<td>-0.195</td>
<td>-0.050</td>
<td>-0.093</td>
<td>-0.040</td>
<td>-0.123</td>
</tr>
<tr>
<td>Newey-West t</td>
<td>(-4.719)</td>
<td>(-2.613)</td>
<td>(-0.889)</td>
<td>(-1.837)</td>
<td>(-2.520)</td>
<td>(-9.938)</td>
</tr>
<tr>
<td>Obs</td>
<td>344</td>
<td>280</td>
<td>203</td>
<td>261</td>
<td>272</td>
<td>167</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.238</td>
<td>0.127</td>
<td>0.009</td>
<td>0.054</td>
<td>0.027</td>
<td>0.152</td>
</tr>
<tr>
<td><strong>Panel B: Conditional predictions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_{r_{t-12,t}&lt;r'<em>{f</em>{t-12,t}}}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$pr_t$</td>
<td>-0.012</td>
<td>-0.051</td>
<td>-0.086</td>
<td>-0.102</td>
<td>-0.097</td>
<td>0.020</td>
</tr>
<tr>
<td>(-0.398)</td>
<td>(-1.149)</td>
<td>(-1.156)</td>
<td>(-1.684)</td>
<td>(-1.211)</td>
<td>(0.528)</td>
<td></td>
</tr>
<tr>
<td>[-0.249]</td>
<td>[-0.919]</td>
<td>[-1.026]</td>
<td>[-1.434]</td>
<td>[-1.258]</td>
<td>[0.365]</td>
<td></td>
</tr>
<tr>
<td>Obs</td>
<td>344</td>
<td>280</td>
<td>203</td>
<td>261</td>
<td>272</td>
<td>167</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.255</td>
<td>0.162</td>
<td>0.058</td>
<td>0.102</td>
<td>0.066</td>
<td>0.156</td>
</tr>
<tr>
<td>$I_{r_{t-12,t}&lt;r'<em>{f</em>{t-12,t}}}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$pr_t$</td>
<td>-0.109</td>
<td>-0.078</td>
<td>-0.055</td>
<td>-0.075</td>
<td>-0.035</td>
<td>-0.066</td>
</tr>
<tr>
<td>(-3.828)</td>
<td>(-5.265)</td>
<td>(-0.767)</td>
<td>(-2.588)</td>
<td>(-2.120)</td>
<td>(-9.371)</td>
<td></td>
</tr>
<tr>
<td>Obs</td>
<td>344</td>
<td>280</td>
<td>203</td>
<td>261</td>
<td>272</td>
<td>167</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.255</td>
<td>0.162</td>
<td>0.058</td>
<td>0.102</td>
<td>0.066</td>
<td>0.156</td>
</tr>
<tr>
<td>$I_{r_{t-12,t}&gt;r'<em>{f</em>{t-12,t}}}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$pr_t$</td>
<td>-0.069</td>
<td>-0.044</td>
<td>-0.002</td>
<td>-0.049</td>
<td>-0.056</td>
<td>-0.053</td>
</tr>
<tr>
<td>(-3.145)</td>
<td>(-2.102)</td>
<td>(-0.825)</td>
<td>(-1.424)</td>
<td>(-2.537)</td>
<td>(-1.400)</td>
<td></td>
</tr>
<tr>
<td>Obs</td>
<td>344</td>
<td>280</td>
<td>203</td>
<td>261</td>
<td>272</td>
<td>167</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.255</td>
<td>0.162</td>
<td>0.058</td>
<td>0.102</td>
<td>0.066</td>
<td>0.156</td>
</tr>
</tbody>
</table>
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 shock is measured by AR(1) residual (\( \epsilon_t^{pr} \)) 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 followed 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 \( R^2 \).

<table>
<thead>
<tr>
<th></th>
<th>All (1)</th>
<th>Fama-French 25 (2)</th>
<th>Momentum 25 (3)</th>
<th>Investment 25 (4)</th>
<th>Profitability 25 (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fama-MacBeth</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \epsilon_t^{pr} )</td>
<td>-0.202***</td>
<td>-0.203*</td>
<td>-0.382***</td>
<td>0.099</td>
<td>-0.108</td>
</tr>
<tr>
<td></td>
<td>(-4.067)</td>
<td>(-2.203)</td>
<td>(-3.764)</td>
<td>(1.686)</td>
<td>(-1.751)</td>
</tr>
<tr>
<td>( r_t - r_t^f )</td>
<td>0.009***</td>
<td>0.010***</td>
<td>0.010***</td>
<td>0.010***</td>
<td>0.010***</td>
</tr>
<tr>
<td></td>
<td>(4.130)</td>
<td>(4.147)</td>
<td>(4.000)</td>
<td>(4.207)</td>
<td>(4.103)</td>
</tr>
<tr>
<td>MAPE</td>
<td>0.206%</td>
<td>0.184%</td>
<td>0.238%</td>
<td>0.218%</td>
<td>0.241%</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.771</td>
<td>0.458</td>
<td>0.941</td>
<td>0.973</td>
<td>0.703</td>
</tr>
</tbody>
</table>

| **GMM**          |         |                    |                |                   |                      |
| \( \epsilon_t^{pr} \) | -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: Correlation 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 obtained from Kenneth R. French’s website. We consider two versions of $pr_t$ shocks, the first difference ($\Delta pr_t$) and AR(1) residual ($\epsilon^p_t$) estimated using full sample.

<table>
<thead>
<tr>
<th></th>
<th>Mkt-RF</th>
<th>SMB</th>
<th>HML</th>
<th>RMW</th>
<th>CMA</th>
<th>Momentum</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta pr_t$</td>
<td>0.104</td>
<td>-0.019</td>
<td>-0.052</td>
<td>-0.057</td>
<td>-0.062</td>
<td>-0.127</td>
</tr>
<tr>
<td>$\epsilon^p_t$</td>
<td>0.081</td>
<td>-0.006</td>
<td>-0.031</td>
<td>-0.034</td>
<td>-0.035</td>
<td>-0.113</td>
</tr>
</tbody>
</table>
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.

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


