# Supplemental Appendix for "Superstitious" Investors 

September 5, 2021

## Contents

A Analysis of the source of volatility ..... 3
B Could investors believe dividends were predictable? ..... 4
C Predictability of Treasury bond excess returns and survey data ..... 10
C. 1 Model and simulation ..... 10
C. 2 Survey data ..... 14
C. 3 Results ..... 15
D Additional results for equities ..... 20

## A Analysis of the source of volatility

We address the question of the volatility decomposition in (17). In the main text, we claimed that nearly all the volatility in returns arises from the volatility in expected dividends, as represented by $b_{n-1}^{2} \sigma_{v}^{2}$. Here we explain why this is so. First note that $\sigma_{u}^{2}$ is the volatility of realized dividends. This $0.07^{2}$ per annum in postwar data. On the other hand, the volatility of shocks to $x_{t}, \sigma_{v}$, and the unconditional volatility of $x_{t}, \sigma_{x}$, are unobserved. To understand the magnitude of the remaining terms, we turn to the prices of dividend claims, normalized by current dividends. These are denoted by $\Phi_{n}\left(x_{t}\right)$ and given in (7) and (8).

Recall that the price-dividend ratio on the market is a sum of these component pricedividend ratios. Furthermore, even if the persistence $\phi$ is high, decay is geometric, and so for $n$ sufficiently large, $b_{n} \approx(1-\phi)^{-1}$. If we let $\sigma_{p d}^{2}$ be the variance of the log price-dividend ratio on the market, roughly speaking ${ }^{1}$

$$
\sigma_{p d}^{2} \equiv \lim _{n \rightarrow \infty} \operatorname{Var}\left(\log \Phi_{n}\left(x_{t}\right)\right)=\frac{\sigma_{x}^{2}}{(1-\phi)^{2}}
$$

Then, for long-maturity equity strips (which, due to the properties of geometric decay, best represents the return on the market) the decomposition (17) takes the form

$$
\begin{align*}
\lim _{n \rightarrow \infty} \operatorname{Var}\left(\log \left(1+R_{n t}\right)\right) & =\sigma_{x}^{2}+\frac{\sigma_{v}^{2}}{(1-\phi)^{2}}+\sigma_{u}^{2} \\
& \approx(1-\phi)^{2} \sigma_{p d}^{2}+\left(1-\phi^{2}\right) \sigma_{p d}^{2}+\sigma_{u}^{2} \tag{A1}
\end{align*}
$$

While $\sigma_{u} \approx 0.07, \sigma_{p d} \approx 0.42$. The persistence $\phi$ will equal the persistence of the pricedividend ratio. At $\phi=0.92$, the first term in A1) equals $(0.08 \times 0.42)^{2}$, whereas the second

[^0]Because of geometric decay, $b^{*} \approx(1-\phi)^{-1}$.
term equals $(0.39 \times 0.42)^{2}$. The second term, representing the effect of innovations to $x_{t}$ is thus roughly 25 times larger than the term representing $x_{t}$ itself, and roughly 5 times larger than the term representing dividend volatility ${ }_{2}^{2}$ Finally note that these terms add up to $(0.18)^{2}$, thus (roughly) accounting for the annual volatility in stock returns.

## B Could investors believe dividends were predictable?

A possible objection to our model is that, over time, investors would learn that dividends are in fact unpredictable. If investors did learn the correct distribution, prices would remain volatile, but return predictability would dissipate. In this section, we confront the hypothesized beliefs with data. We consider an investor whose prior beliefs include the possibility of dividend growth predictability. The agent updates these beliefs given the historical time series, seen through the lens of the likelihood implied by equation (1)-(3) in the main paper. Our evidence speaks to the difficulty of learning the true process for dividend growth.

We assume, as in our model, the agent believes that dividend growth contains a predictable component. Should this predictable component exist, it follows from the reasoning in our model that it should be captured by the price-dividend ratio ${ }^{3}$ The agent therefore considers the predictive system:

$$
\begin{align*}
\Delta d_{t+1} & =\beta \hat{x}_{t}+u_{t+1}  \tag{B1}\\
\hat{x}_{t+1} & =\hat{\phi} \hat{x}_{t}+\hat{v}_{t+1} \tag{B2}
\end{align*}
$$

[^1]where $\hat{x}_{t}=p_{t}-d_{t}$, the log price-dividend ratio, and where
\[

\left[$$
\begin{array}{c}
u_{t}  \tag{B3}\\
\hat{v}_{t}
\end{array}
$$\right] \stackrel{i i d}{\sim} N\left(0,\left[$$
\begin{array}{cc}
\sigma_{u}^{2} & 0 \\
0 & \hat{\sigma}_{v}^{2}
\end{array}
$$\right]\right) .
\]

We refer to the predictor variable as $\hat{x}_{t}$ in contrast to $x_{t}$. Up to linearization error, the assumptions in Section 2 imply that $\hat{x}$ and $x$ differ only by a scale factor, approximately equal to $1 /(1-\phi)$. For convenience, we de-mean both variables..$^{4}$

It suffices to consider a prior on the parameters of the dividend process and the marginal likelihood for the dividend process, taking observations on $\hat{x}_{t}$ as given. That is, the timeseries regression (B1) for dividend growth is, in this case, equivalent to standard OLS in which the regressor is strictly exogenous.

We assume a prior inverse-gamma distribution for $\sigma_{u}^{2}$ and, conditional on $\sigma_{u}^{2}$, a normal distribution for the predictive coefficient $\beta$ :

$$
\begin{align*}
\beta \mid \sigma_{u} & \sim N\left(\beta_{0}, g^{-1} \sigma_{u}^{2} \Lambda_{0}^{-1}\right)  \tag{B4}\\
\sigma_{u}^{2} & \sim I G\left(a_{0}, b_{0}\right) \tag{B5}
\end{align*}
$$

We set parameters $a_{0}$ and $b_{0}$ so that the prior on $\sigma_{u}^{2}$ is diffuse ${ }^{5}$ Equation B5 implies a conjugate prior on $\beta$ (Zellner, 1996). As explained below, $\Lambda_{0}$ is a scale factor that will allow us to interpet $g$ as indexing the strength of the prior.

Given the priors (B4) and (B5), and the likelihood defined by (B1 B3), the agent forms a posterior. Let $\hat{\mathbf{x}}_{t}=\left\{\hat{x}_{0}, \ldots, \hat{x}_{t}\right\}$, namely the set of observations on $\hat{x}_{s}$, up to and including time $t$. Let $\mathbf{y}_{t}=\left\{\Delta d_{1}, \ldots \Delta d_{t}\right\}$ be the dividend growth observations up to and including

[^2]time $t$. The agent calculates
\[

$$
\begin{equation*}
p\left(\beta, \sigma_{u} \mid \hat{\mathbf{x}}_{t}, \mathbf{y}_{t}\right) \propto \mathcal{L}\left(\mathbf{y}_{t} \mid \hat{\mathbf{x}}_{t}, \beta, \sigma_{u}\right) p\left(\beta, \sigma_{u}\right) \tag{B6}
\end{equation*}
$$

\]

where $p\left(\beta, \sigma_{u}\right)$ is the prior specified in (B4) and (B5) and $\mathcal{L}\left(\mathbf{y}_{t} \mid \hat{\mathbf{x}}_{t}, \beta, \sigma_{u}\right)$ is the likelihood of observing the dividend growth data given the predictor variable and the parameters.

We fix time $T$ as the last data point observed. We stack the observations on $\hat{x}_{t}$ and $\Delta d_{t}$ into vectors:

$$
Y=\left[\begin{array}{c}
\Delta d_{1} \\
\vdots \\
\Delta d_{T}
\end{array}\right], \quad X=\left[\begin{array}{c}
\hat{x}_{0} \\
\vdots \\
\hat{x}_{T-1}
\end{array}\right]
$$

Note that the OLS estimate of $\beta$ equals

$$
\hat{\beta}=\left(X^{\top} X\right)^{-1} X^{\top} Y
$$

and that (B1) implies

$$
Y=\beta X+U
$$

where $U \sim N\left(0, \sigma_{u}^{2} I\right)$, and $I$ is the $T \times T$ identity matrix. It follows that the posterior (B6) is given by

$$
p\left(\beta, \sigma_{u} \mid, \hat{\mathbf{x}}_{T}, \mathbf{y}_{T}\right) \propto \sigma_{u}^{-n} \exp \left\{-\frac{1}{2 \sigma_{u}}(Y-X \beta)^{\top}(Y-X \beta)\right\} \sigma_{u}^{-1} \exp \left\{-\frac{g \Lambda_{0}\left(\beta-\beta_{0}\right)^{2}}{2 \sigma_{u}^{2}}\right\}
$$

where $\propto$ means up to a proportionality factor that does not depend on $\beta$ and $\sigma_{u}$. Completing the square implies

$$
\begin{equation*}
p\left(\beta, \sigma_{u} \mid, \hat{\mathbf{x}}_{T}, \mathbf{y}_{T}\right) \propto \sigma_{u}^{-1} \exp \left\{-\frac{\left(X^{\top} X+g \Lambda_{0}\right)(\beta-\bar{\beta})^{2}}{2 \sigma_{u}^{2}}\right\} \times p\left(\sigma_{u} \mid \hat{\mathbf{x}}_{T}, \mathbf{y}_{T}\right) \tag{B7}
\end{equation*}
$$

where

$$
\begin{aligned}
\bar{\beta} & =\left(g \Lambda_{0}+X^{\top} X\right)^{-1}\left(g \Lambda_{0} \beta_{0}+X^{\top} Y\right) \\
& =\left(g \Lambda_{0}+X^{\top} X\right)^{-1}\left(g \Lambda_{0} \beta_{0}+\left(X^{\top} X\right) \hat{\beta}\right),
\end{aligned}
$$

and where $p\left(\sigma_{u} \mid \hat{\mathbf{x}}_{T}, \mathbf{y}_{T}\right)$ is a term that does not depend on $\beta$ and is therefore the marginal posterior of $\sigma_{u}$ (see Zellner, 1996, Chapter 8) for more detail). It is clear from (B7) that the posterior of $\beta$ conditional on $\sigma_{u}$ is normal with posterior mean $\bar{\beta}$. Note also that $\bar{\beta}$ is a weighted average between the prior mean $\beta_{0}$ and the sample mean $\hat{\beta}$, with the weights determined by the precisions of the prior and of the sample respectively.

If we, ex post, set $\Lambda_{0}=X^{\top} X$, then $g$ corresponds to the weight on $\beta_{0}$ as a percent of the weight on $\hat{\beta}$, so that $g=0.1$ implies that the prior receives $1 / 10$ of the weight of the sample, and $g=0.01$ means it receives $1 / 100$ of the weight. We set the prior mean of $\beta$ to a value consistent with the agent's beliefs in Section 2. For comparability with Tables 588 , which show regressions on the dividend-price ratio, Figure B1 shows the negative of the posterior mean of $\beta$. We consider an informative prior, with $g=0.10$, and a diffuse prior, with $g=0.01$.

Figure B1 shows that the agent does indeed revise her prior beliefs, at least at first. She revises it to imply more, not less predictability of dividend growth. Indeed, from the 1930s to the 1970s, it appears that dividend growth was more predictable than later in the sample $]^{6}$ Only when nearly the full sample is used, namely around 2000 , does the posterior mean converge to the sample estimate, which happens to be close to, though implying slightly more predictability than, the prior. Note that the convergence implies that the prior does not matter when the full sample is used.

Thus an agent, viewing the evidence on annual dividend growth rates in isolation, would

[^3]Figure B1: Predicting dividend growth using the dividend-price ratio


This figure shows the posterior mean of the predictive coefficient in a regression of one-year ahead dividend growth on the dividend-price ratio. The posterior mean is calculated using Bayesian methods, assuming an informative prior, where $g$ indexes the degree of informativeness. For each year in the sample, the agent uses all available data to form a posterior for the predictive coefficient. Data begin in 1927. A prior parameter of $g=0.1$ implies that the prior mean of the coefficient receives a weight of $10 \%$ relative to the sample estimate, whereas a prior parameter of $g=0.01$ implies that the prior mean receives a weight of $1 \%$. Shaded areas denote plus and minus 2 posterior standard deviations.
be justified in maintaining a belief that dividend growth rates are predictable. This agent, however, is not fully rational. He incorrectly extrapolates the predictability from the oneyear horizon to long horizons. Moreover, he fails to notice that excess returns are also predictable. $7^{7}$

[^4]
# C Predictability of Treasury bond excess returns and survey data 

## C. 1 Model and simulation

Assume that investors believe that the continuously-compounded short-term interest rate $r_{t}$ follows a first-order autoregressive process, so that

$$
\begin{equation*}
\Delta r_{t+1}=(\phi-1)\left(r_{t}-\bar{r}\right)+v_{t+1} \tag{C1}
\end{equation*}
$$

where $\Delta r_{t+1}=r_{t+1}-r_{t},|\phi|<1, \bar{r}$ is the unconditional mean of $r_{t}$, and $v_{t+1} \stackrel{i i d}{\sim} N\left(0, \sigma_{v}^{2}\right)$. Note that $\phi$ is the first-order autocorrelation of $r_{t} .8$

As with dividend growth, investors believe that changes in interest rates are more forecastable than they are in reality. That is, while (C1) represent beliefs, the true process is governed by

$$
\begin{equation*}
\Delta r_{t+1}=(\zeta-1)\left(r_{t}-\bar{r}\right)+v_{t+1} \tag{C2}
\end{equation*}
$$

with

$$
\begin{equation*}
|\zeta-1|<|\phi-1| . \tag{C3}
\end{equation*}
$$

We focus on the case where $\zeta, \phi \in[0,1]$ so that C 3$)$ implies $\zeta>\phi$. In forecasting next

[^5]and
$$
z_{t+1}=\phi z_{t}+v_{t+1}
$$
with $u_{t+1}$ and $v_{t+1}$ distributed as in (3). The interest rate $r_{t}$ then solves
$$
E_{t}\left[\delta e^{-\Delta \pi_{t+1}+r_{t}}\right]=1
$$

Under these assumptions, the analysis proceeds exactly as described.
period's interest rate, (C3) implies that investors put more weight on previous values of the interest rate than they should. Alternatively stated, interest rates are closer to a random walk (they mean revert more slowly) in the data than investors believe ( $\zeta>\phi$ ).

We consider risk-neutral pricing for bonds. The dynamics thus far define a discrete-time Vasicek (1977) model. ${ }^{9}$ Let $B_{n}\left(r_{t}\right)$ denote the price of the $n$-period bond as a function of the riskfree rate between periods $t$ and $t+1$. Then bond prices satisfy the recursion

$$
\begin{equation*}
B_{n}\left(r_{t}\right)=E_{t}^{*}\left[e^{-r_{t}} B_{n-1}\left(r_{t+1}\right)\right] \tag{C4}
\end{equation*}
$$

with $B_{0}\left(r_{t}\right)=1$ and $B_{1}\left(r_{t}\right)=e^{-r_{t}}$. It follows that

$$
\begin{equation*}
\log B_{n}\left(r_{t}\right)=-a_{n}-b_{n} r_{t} \tag{C5}
\end{equation*}
$$

with

$$
\begin{align*}
& a_{n}=a_{n-1}+b_{n-1}(1-\phi) \bar{r}-\frac{1}{2} b_{n-1}^{2} \sigma_{v}^{2}  \tag{C6}\\
& b_{n}=1+b_{n-1} \phi
\end{align*}
$$

and $a_{0}=b_{0}=0$. Note that $a_{1}=0$ and $b_{1}=1$, so that $B_{1}\left(r_{t}\right)=e^{-r_{t}}$. The solution for $b_{n}$ is again

$$
\begin{equation*}
b_{n}=\frac{1-\phi^{n}}{1-\phi} \tag{C7}
\end{equation*}
$$

Defining the continuously compounded yield on the $n$-period bond as

$$
y_{n t}=-\frac{1}{n} \log B_{n}\left(r_{t}\right)
$$

[^6]It follows from (C7) that the yield spread equals

$$
\begin{equation*}
y_{n t}-y_{1 t}=\mathrm{constant}+\left(\frac{1}{n} \frac{1-\phi^{n}}{1-\phi}-1\right) r_{t} \tag{C8}
\end{equation*}
$$

(recall that $y_{1 t}=r_{t}$ ). The (continuously compounded) holding period return on the $n$-period bond is given by

$$
r_{n, t+1}=\log B_{n-1}\left(r_{t+1}\right)-\log B_{n}\left(r_{t}\right)
$$

(note that $r_{1, t+1}=r_{t}$ ). Substituting in for (C5), (C7), and for the physical evolution of $r_{t}$, (C2), we find the following equation for continuously-compounded excess returns:

$$
r x_{n, t+1}=r_{n, t+1}-r_{1, t+1}=\mathrm{constant}+(\zeta-\phi) \frac{b_{n-1}}{1-(1 / n) b_{n}}\left(y_{n t}-y_{1 t}\right)+b_{n-1} v_{t+1} .
$$

When $\zeta=\phi$, we recover the equilibrium with correct beliefs in which excess returns are unpredictable. However, when $\zeta>\phi$, the yield spread will predict excess returns with a positive sign, as in the data.

The economic intuition is similar to that of predictability of stock returns by the pricedividend ratio. Long-term bond yields fluctuate due to changing forecasts of future shortterm interest rates. When long-term yields are high relative to short-term yields, it is because (in this model), investors expect short-term yields to rise. However, short-term yields are not as predictable as investors believe, and thus on average, short-term yields will rise less than anticipated (or even fall). This leads to a positive excess return on the long-term bond.

The ability of the yield spread to forecast excess bond returns was first noted in the data by Campbell and Shiller (1991). According to the expectations hypothesis of interest rates, yields on long-term bonds should reflect forecasts of future short-term interest rates ${ }^{10}$

[^7]Table C1: Moments of Bond Yields

|  | Maturity in Years |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | 1 | 2 | 3 | 4 | 5 |
| Panel A: Data 1952-2019 |  |  |  |  |  |
| $\beta_{n}$ |  | 1.58 | 2.09 | 2.33 | 2.44 |
| $t$-stat |  | [2.90] | [3.46] | [3.73] | [3.52] |
| $\sigma\left(y_{n}\right)$ | 3.08 | 3.04 | 2.97 | 2.93 | 2.87 |
| $A C\left(y_{n}\right)$ | 0.88 | 0.90 | 0.91 | 0.91 | 0.92 |
| $\sigma\left(y_{n}-y_{1}\right)$ |  | 0.33 | 0.53 | 0.69 | 0.81 |
| $A C\left(y_{n}-y_{1}\right)$ |  | 0.41 | 0.46 | 0.52 | 0.55 |
| Panel B: Model |  |  |  |  |  |
| $\beta_{n}$ |  | 1.48 | 1.31 | 1.19 | 1.10 |
| $\sigma\left(y_{n}\right)$ | 2.83 | 2.05 | 1.56 | 1.23 | 1.01 |
| $A C\left(y_{n}\right)$ | 0.86 | 0.86 | 0.86 | 0.86 | 0.86 |
| $\sigma\left(y_{n}-y_{1}\right)$ |  | 0.78 | 1.27 | 1.60 | 1.82 |
| $A C\left(y_{n}-y_{1}\right)$ |  | 0.86 | 0.86 | 0.86 | 0.86 |

Panel A of the table reports the volatility and the first-order autocorrelation of zero-coupon bond yields and yields spread, as well as the regression coefficients $\beta_{n}$ as in $r x_{n, t+1}=\alpha_{n}+$ $\beta_{n}\left(y_{n t}-y_{1 t}\right)+\epsilon_{t+1}$, where $r x_{n, t+1}$ is the return of $n$-year bond in excess of $y_{1}$ over period $t+1$. The $t$-statistics adjust for heteroskedasticity. Panel B report the percentiles of those moments computed over 1000 simulations, each with 66 years of length. Data are from 1952 to 2019 .

Indeed, the recursion ( (C4) implies

$$
y_{n t}=-\frac{1}{n} \log E_{t}^{*}\left[e^{-\sum_{\tau=0}^{n-1} r_{t+\tau}}\right] .
$$

If investors correctly anticipate yields, then bond returns will be unpredictable. However, Campbell and Shiller (1991), Fama and Bliss (1987) and a large subsequent literature show that excess bond returns are strongly forecastable. We replicate this finding in Table C1, which reports coefficients from regressing bond returns on yield spreads using the Fama-Bliss data set for zero-coupon bonds.

As an illustrative calculation, we calibrate $\sigma_{v}$ and $\phi$ to jointly match the volatility and
first-order autocorrelations of yields. This implies $\sigma_{v}=1.5 \%$ per annum and an annual autocorrelation $\zeta$ of (roughly) 0.90. Given these parameters, $\phi=0.45$ gives us roughly the amount of predictability in the data.

Table C1 shows results from historical data and from simulating 1000 samples of length 70 years. We run the regression

$$
r x_{n, t+1}=\alpha_{n}+\beta_{n}\left(y_{n t}-y_{1 t}\right)+\epsilon_{t+1}
$$

for zero-coupon bonds for maturities ranging from 2 to 5 years. Bond excess returns are strongly predictable in both data and model.

In addition to the moments in Table C1, the model makes predictions that can be tested with survey data. In the subsection below we describe the survey data and then perform the tests.

## C. 2 Survey data

Our primary survey data soruce is interest rate forecasts from Blue Chip Financial Forecasts (BCFF). This data source contains survey forecasts for a variety of interest rates in the US, in particular the Treasury rates. Behind each Treasury rate consensus are forecasts provided by tens of major banks and financial institutions, e.g. J.P. Morgan and S\&P Global. This data source goes back to Q4 of 1982.

An alternate data source of interest rate forecasts to Blue Chip Financial Forecasts (BCFF) is Survey of Professional Forecasters (SPF). This is a quarterly survey contains a large number of economic variables, including the 3-month Treasury rates. The contributors to these surveys are economists of a variety of backgrounds. The interest rate forecast data go back to Q3 of 1981, which is similar to BCFF. While SPF is not a specialized interest rate data source and contains only the 3-month Treasury rate forecasts, it is useful as a robustness check on top of the BCFF data. Here, the correlations between interest rate
forecasts and earning growths forecasts computed with BCFF data, as shown in Figure C1 are very similar to those computed with the SPF data in Figure C2,

## C. 3 Results

Equation C1, C2, and C8 predict that the term spread be very positively associated with forecasted change of interest rate and less positively associated with its realization. Column 1-2 and 4-5 of Table C2 show that this does not hold in the survey data. Contrary to the model's prediction, we see that the term spreads are more positively associated with realized interest rate changes in the future. Similar results are seen also in the SPF data, shown in Table C3.

A counterfactual aspect of the model is that the level of the interest rate and the term spread are perfectly negatively correlated. It therefore also predicts that the level of the interest rate should be very negatively associated with forecasted interest rate change and less negatively associated with its realization. Column 3 and 6 of Table C2 and C3 show that this is also not the case either. The model is therefore soundly rejected by the survey data.

Figure C1: Forecasted earnings growth versus forecasted interest rates (BCFF data)


This figure plots log forecasted 1-year earnings growth against forecasted 3-month Treasury bill rate 4 quarters away. Data are quarterly from 1982-2018.

Figure C2: Forecasted earnings growth versus forecasted interest rates (SPF data)


This figure plots log forecasted 1-year earnings growth against forecasted 3-month Treasury bill rate 4 quarters away. Data are quarterly from 1981-2019.

Table C2: Term Spreads, Short Rates, and Forecasted Changes in Short Rates-BCFF

|  |  | $\begin{gathered} (2) \\ { }_{t}\left[r_{3, t+4}\right]- \end{gathered}$ | $\overline{(3)}$ | (4) | $\begin{gathered} \hline(5) \\ r_{3, t+4}-r_{3} \end{gathered}$ | (6) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $y_{60, t}-r_{3, t}$ | $\begin{gathered} 0.315 * * * \\ {[4.36]} \end{gathered}$ |  |  | $\begin{gathered} 0.425^{*} \\ {[1.68]} \end{gathered}$ |  |  |
| $y_{120, t}-r_{3, t}$ |  | $\begin{gathered} 0.269^{* * *} \\ {[5.10]} \end{gathered}$ |  |  | $\begin{gathered} 0.357^{* *} \\ {[2.12]} \end{gathered}$ |  |
| $r_{3, t}$ |  |  | $\begin{gathered} -0.105^{* * *} \\ {[-3.81]} \end{gathered}$ |  |  | $\begin{gathered} -0.139^{* *} \\ {[-2.47]} \end{gathered}$ |
| constant | $\begin{gathered} -0.076 \\ {[-0.60]} \end{gathered}$ | $\begin{gathered} -0.153 \\ {[-1.12]} \end{gathered}$ | $\begin{gathered} 0.726^{* * *} \\ {[5.68]} \end{gathered}$ | $\begin{gathered} -0.743^{*} \\ {[-1.97]} \end{gathered}$ | $\begin{gathered} -0.836^{* *} \\ {[-2.18]} \end{gathered}$ | $\begin{aligned} & 0.327 \\ & {[1.37]} \end{aligned}$ |
| N | 145 | 145 | 145 | 145 | 145 | 145 |

Column 1 of this table reports results of the follow quarterly time-series regression: $\hat{E}_{t}\left[r_{3, t+4}\right]-r_{3, t}=\alpha+\beta\left(y_{60, t}-r_{3, t}\right)+\epsilon_{t}$. Here $\hat{E}_{t}\left[r_{3, t+4}\right]-r_{3, t}$ is the forecasted change of 3 -month Treasury bill rates. $r_{60, t}-r_{3, t}$ is the 5 -year Treasury bond rate subtracting 3month Treasury bill rate in quarter $t$. Column 2 instead uses the 10 -year $/ 3$-month term spread as the independent variable. Column 3 instead uses the 3-month Treasury bill rate as the dependent variable. Column 4-6 are analogous regressions with the dependent variables changed to realized short rate changes. Data are quarterly from 1982Q4-2018Q4. T-stats calculated using Newey-West standard errors with 6 lags are reported in the square brackets.

Table C3: Term Spreads, Short Rates, and Forecasted Changes in Short Rates-SPF

|  | $\overline{(1)}$ | $(2)$ $\left[r_{3, t+4}\right]$ | $\overline{(3)}$ | (4) | (5) $r_{3, t+4}-r_{3}$ | (6) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $y_{60, t}-r_{3, t}$ | $\begin{gathered} 0.329 * * * \\ {[3.94]} \end{gathered}$ |  |  | $\begin{gathered} 0.477^{* *} \\ {[2.38]} \end{gathered}$ |  |  |
| $y_{120, t}-r_{3, t}$ |  | $\begin{gathered} 0.228^{* * *} \\ {[3.17]} \end{gathered}$ |  |  | $\begin{gathered} 0.463^{* *} \\ {[2.53]} \end{gathered}$ |  |
| $r_{3, t}$ |  |  | $\begin{gathered} -0.051 * * \\ {[-2.50]} \end{gathered}$ |  |  | $\begin{gathered} -0.212^{* * *} \\ {[-3.29]} \end{gathered}$ |
| constant | $\begin{aligned} & -0.181 \\ & {[-1.26]} \end{aligned}$ | $\begin{aligned} & -0.161 \\ & {[-1.00]} \end{aligned}$ | $\begin{gathered} 0.482^{* * *} \\ {[5.10]} \end{gathered}$ | $\begin{gathered} -0.958^{* *} \\ {[-2.44]} \end{gathered}$ | $\begin{gathered} -1.184^{* *} \\ {[-2.53]} \end{gathered}$ | $\begin{gathered} 0.552^{*} \\ {[1.91]} \end{gathered}$ |
| N | 150 | 150 | 150 | 150 | 150 | 150 |

Column 1 of this table reports results of the follow quarterly time-series regression: $\hat{E}_{t}\left[r_{3, t+4}\right]-r_{3, t}=\alpha+\beta\left(y_{60, t}-r_{3, t}\right)+\epsilon_{t}$. Here $\hat{E}_{t}\left[r_{3, t+4}\right]-r_{3, t}$ is the forecasted change of 3 -month Treasury bill rates. $y_{60, t}-r_{3, t}$ is the 5 -year Treasury bond rate subtracting 3 month Treasury bill rate in quarter $t$. Column 2 instead uses the 10 -year $/ 3$-month term spread as the independent variable. Column 3 instead uses the 3-month Treasury bill rate as the dependent variable. Column 4-6 are analogous regressions with the dependent variables changed to realized short rate changes. Data are quarterly from 1981Q4-2018Q4. T-stats calculated using Newey-West standard errors with 6 lags are reported in the square brackets.

## D Additional results for equities

In our model of the cross section of equities, the value spread is the absolute value of $z_{t}$. The return to a HML strategy is $-\operatorname{sign}\left(z_{t}\right) v_{z, t+1}$. The aggregate valuation ratio is $x_{t}$. With the addition of stochastic volatility, because the value spread and HML returns are both driven by the same set of shocks, their volatilities should be highly correlated. However, the aggregate valuation ratio is driven by $x_{t}$ and thereby $v_{x, t+1}$. Our model therefore predicts weak correlations between the volatility of aggregate valuation and those of the value spread and HML returns.

Panel A of Figure D1 plots quarterly volatility of aggregate E/P ratio versus that of the value spread. They have a low correlation of 0.17 . Panel B plots quarterly volatility of aggregate E/P ratio versus that of the HML return. They have a low correlation of 0.22 . Panel C plots quarterly volatilites of the HML returns and the value spread we constructed. They have a reasonably high correlation of 0.53 . These results are broadly consistent with the model's predictions.

In our model, the value spread is the absolute value of $z_{t}$, and aggregate valuation ratio is $x_{t}$. Because $x_{t}$ and $z_{t}$ are based on difference iid Gaussian shocks, the model predicts a correlation of 0 between these two measures. Also, return to a HML strategy is $-\operatorname{sign}\left(z_{t}\right) v_{z, t+1}$. That to a market timing strategy trading on the aggregate valuation ratio is $-x_{t} v_{x, t+1}$. Again, because they are driven by two different sets of iid Gaussian shocks, our model predicts that they are uncorrelated.

Figure D1: Relation between market risk and risk to a value strategy

Panel A: Volatility of aggregate $E / P$ versus volatility of value spread


Panel B: Volatility of aggregate $E / P$ versus volatility of HML returns


Panel C: Volatility of HML returns versus volatility of the value spread


Notes: HML is a portfolio that is long the high E/P ratio quintile and short the low E/P ratio. The value spread is defined as the difference of $\mathrm{E} / \mathrm{P}$ ratio of bin 5 and bin 1 scaled by the aggregate E/P ratio. Data are quarterly from 1971-2020.

Figure D2: Relation between the value anomaly and aggregate valuations

Panel A: Value spread versus deviation of aggregate dividend-to-price from its mean


Panel B: Returns to timing the market versus HML returns


Notes: The value spread is defined as the difference between the $\mathrm{D} / \mathrm{P}$ ratio of bin 5 in a value sort and that of bin 1 , scaled by the aggregate $\mathrm{D} / \mathrm{P}$ ratio. Data are monthly from 1926-2020.

## References

Campbell, J. Y. (1986). A defense of traditional hypotheses about the term structure of interest rates. Journal of Finance, 41:183-193.

Campbell, J. Y. and Shiller, R. J. (1991). Yield spreads and interest rate movements: A bird's eye view. Review of Economic Studies, 58:495-514.

Dai, Q. and Singleton, K. (2002). Expectations puzzles, time-varying risk premia, and affine models of the term structure. Journal of Financial Economics, 63:415-442.

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. Journal of Finance, 57:369-443.

Fama, E. F. and Bliss, R. R. (1987). The information in long-maturity forward rates. American Economic Review, 77:680-692.

Jagannathan, R. and Liu, B. (2019). Dividend dynamics, learning, and expected stock index returns. The Journal of Finance, 74(1):401-448.

Piazzesi, M., Salomao, J., and Schneider, M. (2015). Trend and cycle in bond premia. Working Paper, Stanford University and University of Minnesota.

Skinner, B. F. (1948). 'Superstition' in the pigeon. Journal of Experimental Psychology, 38(2):168.

Vasicek, O. (1977). An equilibrium characterization of the term structure. Journal of Financial Economics, 5:177-188.

Zellner, A. (1996). An introduction to Bayesian inference in econometrics. John Wiley and Sons, Inc., New York, NY.


[^0]:    ${ }^{1}$ Note that the $\log$ price-dividend ratio equals

    $$
    p d=\log \sum_{n=1}^{\infty} \Phi_{n}\left(x_{t}\right) \approx \sum_{n=1}^{\infty} a_{n}+b_{n} x_{t}=a^{*}+b^{*} x_{t}
    $$

[^1]:    ${ }^{2}$ This will also be true in a rational model with prices driven by discount rate variation. Most of the variation in realized returns comes from innovations in the discount rate, which are unpredictable. Very little comes from the variation in the discount rate itself.
    ${ }^{3}$ To the extent that the price-dividend ratio fails to capture this component, we are biased against finding dividend growth predictability, and therefore proving the beliefs to be less justifiable than otherwise.

[^2]:    ${ }^{4}$ De-meaning the variables simplifies the analysis, and only affects the conclusions through a degree-offreedom adjustment that becomes negligible as the same size grows.
    ${ }^{5}$ Because our focus will be on the posterior mean of $\beta$, these play no further role in our analysis.

[^3]:    ${ }_{6}^{6}$ Jagannathan and Liu (2019) also show that dividend growth predictability features striking instability over the sample, declining after 1970.

[^4]:    ${ }^{7}$ While we do not model reinforcement learning (which is a feature of Skinner (1948)), these results suggest that the agent would have received positive reinforcement, throughout the sample, in the sense that he or she would have predicted cash flow growth with relative accuracy. While returns would have been different than expect, the low $R^{2}$ in return predictability regressions suggests that reinforcement learning through this channel would not have been significant.

[^5]:    ${ }^{8}$ The analysis in this section takes the short-term interest rate $r_{t}$ as a given. Perhaps the simplest way to micro-found variation in this rate is to consider a risk-neutral investor with discount rate $\delta$ and an exogenous inflation process $\Delta \pi_{t+1}$ such that

    $$
    \Delta \pi_{t+1}=\bar{\pi}+z_{t}+u_{t+1}
    $$

[^6]:    ${ }^{9}$ A substantial literature on latent factor models strongly rejects a single-factor model in favor of multifactor alternatives (Dai and Singleton, 2002, Duffee, 2002). Piazzesi et al. (2015) show how subjective expectations can be incorporated into a model with richer dynamics. For the purpose of illustrating our mechanism, however, this simple model suffices.

[^7]:    ${ }^{10}$ There are slight differences depending on whether this hypothesis is articulated in logs or levels (Campbell, 1986).

