

# Forecasting Dividend Growth to Better Predict Returns

Filipe Lacerda<sup>1</sup>    Pedro Santa-Clara<sup>2</sup>

This version: February 2010<sup>3</sup>

## Abstract

The dividend-price ratio changes over time due to variation in expected returns and in forecasts of dividend growth. We adjust the dividend-price ratio to isolate the fluctuations that are due to variation in expected returns from those that are due to changing forecasts of dividend growth. This adjusted dividend-price ratio is statistically significant in predictive regressions and yields an in-sample  $R^2$  of 16.27% and an out-of-sample  $R^2$  of 12.35%, which compare with 7.88% and -2.94% for the unadjusted multiple. Structural estimation of our model obtains even higher measures of fit. Our results are robust across subsamples.

---

<sup>1</sup>PhD student at the University of Chicago Booth School of Business, e-mail: fl@chicagobooth.edu

<sup>2</sup>Millennium Chair in Finance, Universidade Nova de Lisboa and NBER. Campus de Campolide, 1099-032 Lisboa, Portugal, e-mail: psc@fe.unl.pt

<sup>3</sup>We would like to thank John Cochrane for extensive and very helpful comments. We also want to thank Ralph Koijen, Lubos Pastor, Pietro Veronesi, Mungo Wilson, and Motohiro Yogo for their valuable comments. The latest version of this paper is available at <http://docentes.fe.unl.pt/~psc>

# 1 Introduction

Variation in the dividend-price ratio is the result of two things: fluctuations in expected returns and changes in investors' forecasts of cash-flows. If investors expect to receive higher cash-flows, then stocks will be worth more today. If investors require a higher rate of return, future cash-flows will be more heavily discounted, and stocks will be worth less today.

This relation between the dividend-price ratio and expected returns justifies using the dividend-price ratio to forecast returns. This has been done by Dow (1920), Campbell (1987), Fama and French (1988), Hodrick (1992), and more recently by Campbell and Yogo (2006), Ang and Bekaert (2007), Cochrane (2008), and Binsbergen and Koijen (2009). However, the evidence of return predictability has been questioned by Goyal and Welch (2008), among others, who show that the dividend-price ratio, along with several other variables, has no ability to forecast stock returns out of sample.<sup>4</sup> Whether returns are predictable is still an open debate.

We provide strong evidence that returns are indeed predictable. Our point is that changes in forecasts of dividends need to be taken into account when forecasting returns with the dividend-price ratio. We use a simple present-value model to propose an adjustment to the dividend-price ratio that isolates the component due to expected returns from that caused by changing forecasts of dividend growth. The adjusted and unadjusted versions of the dividend-price ratio are positively correlated but the former is far more volatile than the latter.

We compare the adjusted and unadjusted versions of the dividend-price ratio to forecast returns with predictive regressions and find a significant difference in performance. In sample, the adjusted multiple has an  $R^2$  of 16.27% whereas the unadjusted ratio has an  $R^2$

---

<sup>4</sup>Other references against return predictability are Nelson and Kim (1993), Cavanagh, Elliott, and Stock (1995), Stambaugh (1999), Lewellen (2004), and Torous, Valkanov, and Yan (2004).

of 7.88%. Out of sample, the difference is even more impressive since the adjusted ratio has an  $R^2$  of 12.35% whereas the unadjusted ratio as a negative  $R^2$  of -2.95%. The coefficient of the adjusted ratio in predictive regressions is statistically significant at the 1% level.

We attribute the success of our approach to the fact that we are able to pin down part of the variation in investors' forecasts of future dividend growth in a robust way. We provide evidence of the importance of past lags of dividend growth in capturing variation in future growth rates. By taking a linear combination of past growth rates we are able to identify part of the variation in forecasts of future dividend growth, and therefore better forecast stock returns.

Finally, when we estimate our structural model, we obtain an even more impressive out-of-sample  $R^2$  of 18.62%. The parameter estimates we obtain imply that expected returns are extremely persistent, and can, for practical purposes, be approximated by a random walk.

## 2 A simple model

Our economy has a simple setup. We assume that investors' expectations of future stock market returns follow the simplest persistent time-series process, an AR(1), and that the parameters governing this process are known to them. It appears sensible to assume that agents fully know the dynamics of conditional expected returns since these result from the solution of the investors' own problem of intertemporal utility maximization. For simplicity, similarly to Pastor and Stambaugh (2009) and to Binsbergen and Koijen (2009), instead of specifying a utility function and deriving the dynamics for expected returns, we assume that preferences are such that conditional expected returns follow this auto-regressive process.<sup>5</sup>

However, we assume that agents do not know the true process for the dividend growth rate but have to forecast it from past data. Our assumption is that investors forecast future

---

<sup>5</sup>Contrary to Pastor and Stambaugh (2009), we assume that investors have perfect information about the true process for expected returns.

dividend growth from an average of past dividend growth rates.<sup>6</sup> This assumption again strikes us as sensible. Because dividend growth is the result of the complex interaction of technology, the business cycle, financial leverage, and management decisions, it is reasonable to assume that investors do not know *ex ante* the expected growth rate. In this case, using a simple average of past growth seems a reasonable approach to forecast the future. Finally, we assume that investors price the stock market given their forecasts of dividend growth to deliver the required expected returns.

In the literature, dividend growth has been modelled in different ways. Cochrane (2008) models dividend growth as an *i.i.d.* process, whereas Bansal and Yaron (2004) and Binsbergen and Koijen (2009) model it as a persistent process. Cochrane (2008) argues that if the dividend-price ratio reflected variation in expected dividend growth, this predictability should be captured when running predictive regressions of dividend growth on the lagged dividend-price ratio. Because it is not, Cochrane (2008) models dividend growth as an *i.i.d.* process when studying predictability from the perspective of an econometrician who only observes returns, dividend growth and dividend-price ratios. However, this is a strong assumption since taking expected dividend growth to be known and constant implies that investors had perfect foresight and were fully aware in the 1940's that future growth would be significantly brighter than in previous history.

Our assumption that investors forecast dividend growth from a past average is consistent with the evidence for the existence of a persistent component in expected dividend growth presented in Bansal and Yaron (2004), Lettau and Ludvigson (2005), Menzly, Santos and Veronesi (2006), and Binsbergen and Koijen (2009). The main advantage of our choice is that it provides a simple way to capture persistence in forecasts of growth rates while yielding reasonable results.

---

<sup>6</sup>Whether we can find variables that provide a better proxy for investor growth forecasts, and therefore increase forecasting power, remains an open question. Presumably, that would lead to even better estimates of expected returns.

Our setup is more formally described in the following paragraphs. Let  $\mu_t = E_t[r_{t+1}]$ . We assume that:

$$\mu_{t+1} = a + b\mu_t + \varepsilon_{t+1}^\mu \quad (1)$$

where  $\varepsilon_{t+1}^\mu$  is a zero mean *i.i.d.* shock.

At time  $t$ , agents know the expected return they demand as compensation for bearing risk at any future horizon and price assets accordingly. These expected future returns are implicit in equation (1) as can be seen from iterating it forward and taking expectations conditional on information available at time  $t$ . The following relation gives us agents' expected return from time  $t+k$  to time  $t+k+1$  at time  $t$ :

$$E_t[\mu_{t+k}] = \frac{a}{1-b} + b^k \left( \mu_t - \frac{a}{1-b} \right). \quad (2)$$

At time  $t$ , agents forecast the dividend growth rate from time  $t+k$  to time  $t+k+1$  from the average of past dividend growth rates:

$$E_t[\Delta d_{t+k}] = \bar{g}_t. \quad (3)$$

Our model for investors' forecasts of dividend growth is extremely simple and does not take into account any sort of Bayesian updating of these forecasts.

Finally, we assume that the present-value identity relating the log dividend-price ratio to expected future discount rates and dividend growth derived in Campbell and Shiller (1988) holds. Start with the standard definition of realized returns:

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t}.$$

If we multiply both numerator and denominator of the right-hand side by the price at  $t+1$ ,

take logs on both sides, and then sum and subtract the log of dividend growth from  $t$  to  $t + 1$ , we obtain the following identity, where small letters denote logs of the original variables, e.g.

$$dp_t = \ln\left(\frac{D_t}{P_t}\right):$$

$$r_{t+1} = \ln\left(1 + \frac{D_{t+1}}{P_{t+1}}\right) - dp_{t+1} + dp_t + \Delta d_{t+1}.$$

By taking a first-order Taylor expansion around the estimated mean of the dividend-price ratio we can obtain the Campbell and Shiller (1988) approximation:<sup>7</sup>

$$r_{t+1} \simeq (1 - \bar{\rho}_t) \kappa_t - \bar{\rho}_t dp_{t+1} + dp_t + \Delta d_{t+1} \quad (4)$$

where:

$$\kappa_t = \left[ -\frac{\bar{\rho}_t}{1 - \bar{\rho}_t} \ln(\bar{\rho}_t) - \ln(1 - \bar{\rho}_t) \right], \quad \bar{\rho}_t = \frac{1}{1 + \frac{\bar{D}}{P_t}},$$

and  $\frac{\bar{D}}{P_t}$  is the historical average of the dividend-price ratio up to time  $t$ .

An important note is that Campbell and Shiller take the Taylor expansion around the true unconditional mean of the dividend-price ratio whereas we take it around the sample mean at time  $t$  since that is how we assume investors estimate that ratio.

If we assume that equation (4) holds exactly and rewrite it with  $dp_t$  on the left-hand side, we obtain a recursive equation that we can iterate indefinitely and upon which we can take conditional expectations to obtain:

$$dp_t = \sum_{k=0}^{+\infty} \bar{\rho}_t^k (E_t[r_{t+k+1}] - E_t[\Delta d_{t+k+1}]) - \kappa_t.$$

Using the law of iterated expectations on  $E_t[r_{t+k+1}]$ , we can rewrite it as a function of  $\mu_{t+k}$ . Since we know its process from equation (1), the previous equation is then:

---

<sup>7</sup>In their paper, Campbell and Shiller offer evidence that the log-linear approximation is quite good.

$$dp_t = \sum_{k=0}^{+\infty} \bar{\rho}_t^k (E_t [\mu_{t+k}] - E_t [\Delta d_{t+k+1}]) - \kappa_t. \quad (5)$$

This identity is one of the key arguments for using the log dividend-price ratio to forecast stock returns. Cochrane (2008) argues that with no variation in forecasts of future dividend growth then all the variation in the dividend-price ratio is exclusively due to changes in expectations of future returns. In our framework, however, since the expected dividend growth changes over time, it also plays a role in fluctuations of the valuation multiple.

### 3 Uncovering expected returns

Our first goal is to use this simple model to estimate investors' expected returns. We now take the role of the econometrician, who, unlike the investors, does not know the true parameters of the process for expected returns. Our model has strong implications for the relation between the log dividend-price ratio, conditional expected returns, and forecasts of dividend growth. We use this relation in the best possible way to estimate expected returns. We achieve that by plugging equations (2) and (3) into equation (5). We start with:

$$dp_t = \sum_{k=0}^{\infty} \bar{\rho}_t^k \left( \frac{a}{1-b} + b^k \left( \mu_t - \frac{a}{1-b} \right) - \bar{g}_t \right) - \kappa_t$$

which can be simplified to:

$$dp_t = \frac{a\bar{\rho}_t}{(1-\bar{\rho}_t)(1-b\bar{\rho}_t)} + \frac{1}{1-b\bar{\rho}_t} \mu_t - \frac{\bar{g}_t}{1-\bar{\rho}_t} - \kappa_t. \quad (6)$$

This equation relates the current log dividend-price ratio to current one-period expected returns, the historical averages of the dividend-price ratio, the average of log dividend growth, as well as parameters  $a$  and  $b$ .

If we rearrange equation (6), we find that conditional expected returns are a function of the current log dividend-price ratio, the historical mean of the log dividend growth, and the historical mean of the dividend-price ratio.

$$\mu_t = (1 - b\bar{\rho}_t) \left( dp_t + \frac{\bar{g}_t}{1 - \bar{\rho}_t} \right) + (1 - b\bar{\rho}_t) \kappa_t - a \frac{\bar{\rho}_t}{(1 - \bar{\rho}_t)} \quad (7)$$

The previous equation shows what changes with fluctuations in expected returns. When expected returns increase, prices go down because future cash-flows are discounted at a higher rate, and the dividend-price ratio goes up. On the other hand, expected returns and  $\bar{\rho}_t$  are negatively related. An increase in the dividend-price ratio raises its sample mean and decreases  $\bar{\rho}_t$ .

Notice that the causality in equation (7) is not from right to left. Instead, the right-hand side changes because conditional expected returns vary. Conditional expected returns are determined by investors as they set market prices and therefore reflect investor preferences. However, in our model, changes in the forecast of the dividend growth rate do not affect expected returns. The presence of  $\bar{g}_t$  in equation (7) ensures that fluctuations in the dividend-price ratio that are solely due to changing forecasts of dividend growth are not reflected in expectations of future returns. Intuitively, the decrease in  $dp_t$  that is exclusively due to a higher forecast of the rate of dividend growth is fully offset by the increase in  $\bar{g}_t$  and the decrease in  $\bar{\rho}_t$ .

If investor forecasts of dividend growth were constant (as well as the estimate of the unconditional mean of the dividend price ratio) then equation (7) would only depend on the current level of the dividend-price ratio. In that case, expected returns should be perfectly estimated by a predictive regression of returns on the log dividend-price ratio.

However, because investor forecasts of dividend growth change over time, the dividend-price ratio alone is not a perfect predictor of expected returns. Instead, when running pre-

dictive regressions, we should use an adjusted version where the dividend growth term in equation (5) is added to the log dividend-price ratio. We thus create the following variable:

$$x_t = dp_t + \frac{\bar{g}_t}{1 - \bar{\rho}_t}. \quad (8)$$

The reason why the adjusted version of the dividend-price ratio works better than using the unadjusted ratio is that  $x_t$  varies less than  $dp_t$  due to changes in investors' forecasts of future cash-flow growth. By adding the second term in  $x_t$ , we remove part of the noise that is caused by variation in forecasts of future dividend growth rates. When investor expectations of future cash-flow growth increase (decrease), current prices go up (down), leading the dividend-price ratio downwards (upwards). This happens even when there is no variation in expected returns. In the adjusted version of the dividend-price ratio,  $x_t$ , the change in  $dp_t$  is partially offset by the change in  $\frac{\bar{g}_t}{1 - \bar{\rho}_t}$ , making  $x_t$  less sensitive to fluctuations in investor forecasts of dividend growth. We purge  $dp_t$  from the effects of changes in  $\bar{g}_t$  to isolate the effects of fluctuations in expected returns.

The next sections verify that the adjusted version of the dividend-price ratio that we advocate is indeed better than the unadjusted version and that assuming that agents do not know the true process for dividend growth rates is relevant for return forecasting.

## 4 Predictive regressions

In this section we study our model empirically. We start by comparing the adjusted and unadjusted versions of the dividend-price ratio and analyzing the difference between the two. Finally, we compare the forecasting performances of  $x_t$  and  $dp_t$  in predictive regressions.

All the empirical results in this paper are obtained from the dataset constructed by

Goyal and Welch (2008).<sup>8</sup> The dataset comprises information about returns, dividends, and prices on the S&P 500 index. We use these series to construct each of the variables that are relevant to our model from 1927 until 2007. The average yearly return is 9.43% and the standard deviation of returns is 19.40%. To construct the series for  $\bar{g}_t$ , we start in  $t = 1937$ , which means we lose the first nine data points and compute the 10-year moving average of the log of the dividend growth rate. Implicitly, this assumes that investors predict dividend growth based on the previous 10 years, which roughly corresponds to a full business cycle.<sup>9</sup> We estimate the unconditional mean of the dividend-price ratio from the sample mean.

## 4.1 The adjusted dividend-price ratio

Looking at some plots helps us grasp what's going on. Figure 1 compares the adjusted and unadjusted versions of the dividend-price ratio. Both versions,  $x_t$  and  $dp_t$ , are highly correlated but  $x_t$  is more volatile than  $dp_t$ . The coefficient of correlation between  $x_t$  and  $dp_t$  is 0.52 whereas the standard deviations of  $x_t$  and  $dp_t$  are 0.69 and 0.46, respectively. One explanation for why  $x_t$  is more volatile than  $dp_t$  is that expected returns are positively correlated with forecasts of dividend growth. Additionally, the log of the current dividend-price ratio is positively related to expected returns but negatively related to forecasts of dividend growth. Therefore, shocks to forecasts of dividend growth that accompany shocks to expected returns dampen the change in the dividend-price ratio. By adding  $\frac{\bar{g}_t}{1-\rho_t}$  to  $dp_t$ , we attenuate this dampening effect.<sup>10</sup>

---

<sup>8</sup>The series are drawn from Goyal's website: <http://www.bus.emory.edu/AGoyal>. See Goyal and Welch (2008) for a complete description of the variables and their sources.

<sup>9</sup>We obtain similar results using the average since the sample begins.

<sup>10</sup>To be more precise, we can expand the variance of  $x_t$  and check the conditions under which it is higher than the variance of  $dp_t$ .  $Var(x_t) > Var(dp_t) \Leftrightarrow \frac{1}{2}Var\left(\frac{\bar{g}_t}{1-\rho_t}\right) < Cov\left(\frac{\bar{g}_t}{1-\rho_t}, \frac{\mu_t}{1-b\rho_t}\right) + Cov\left(\frac{\bar{g}_t}{1-\rho_t}, \frac{a\bar{p}_t}{(1-\rho_t)(1-b\bar{p}_t)} - \kappa_t\right)$ .

If we assume that the second part of the right-hand side of the second inequality is negligible, then we can say that if the dividend growth and the expected returns part of the dividend-price ratio are positively related and their covariance is higher than half of the variance of the dividend growth part, then  $x_t$  is more

[INSERT FIGURE 1 HERE]

Another interesting plot to look at is that of the difference between the adjusted and the unadjusted versions of the dividend-price ratio. This series gives us the weighted sum of the stream of forecasts of future dividend growth,  $\frac{\bar{g}_t}{1-\rho_t}$ . As Figure 2 shows, in the first 20 years, comprising the Great Depression and World War II, the difference was very low and even became negative. According to our model, in that that period, investors were extremely pessimistic about growth prospects. However, as those times were gradually forgotten, investor forecasts of future growth became more optimistic and the difference between  $x_t$  and  $dp_t$  became substantially positive.

[INSERT FIGURE 2 HERE]

This can be clearly seen in Figure 3, which plots investor forecasts of future dividend growth from a 10-year moving average. Investors were extremely pessimistic about the path of future cash-flows in the first years of the sample due to extreme negative economic events. Because such extreme events didn't occur again in the sample, investor forecasts became gradually more optimistic. In the last 40 years of the sample,  $\bar{g}_t$  oscillated between 3% and 8% whereas in the first decade it ranged from -5% to 5%.

[INSERT FIGURE 3 HERE]

## 4.2 Predictability

To show that our adjustment to the dividend-price ratio improves forecasting ability, both in- and out-of-sample, we do two things. First, we run predictive regressions with either  

---

volatile than  $dp_t$ .

$dp_t$  or  $x_t$  on the right-hand side. We compare the performance of the predictors in sample by looking at their statistical significance and goodness-of-fit, using t-statistics and  $R^2$ . We run the following regressions:

$$r_{t+1} = \alpha_{dp} + \beta_{dp} dp_t + \varepsilon_{t+1}^{dp} \quad (9)$$

and

$$r_{t+1} = \alpha_x + \beta_x x_t + \varepsilon_{t+1}^x \quad (10)$$

Second, we compare the out-of-sample performance by running regressions (9) and (10) with an expanding sample and examining the squared forecast errors. We run regressions with observations up to time  $t$  and use the estimated coefficients to forecast the return from  $t$  to  $t + 1$ , for  $t = 1958, \dots, 2007$ . We use the out-of-sample  $R^2$  as an evaluation metric. This measure gives us an idea of how well predictor variables perform as compared to using the historical sample mean of returns up to time  $t$  to forecast returns at time  $t + 1$ . The metric, as defined by Goyal and Welch (2008) is:

$$R_{OOS}^2 = 1 - \frac{MSE_A}{MSE_M}.$$

where  $MSE_A$  is the out-of-sample mean squared forecast error from predictive regressions and  $MSE_M$  is the mean squared forecast error from using the historical mean. This measure takes negatives values when the predictor underperforms the current sample mean, the simplest of return predictors.

Table 1 below shows the in- and out-of-sample performance of both variables. The new variable,  $x_t$ , clearly outperforms the log dividend-price ratio.  $x_t$  is able to explain 16.27% of the variation in returns whereas  $dp_t$  only explains 7.88%. In terms of statistical significance,

the t-statistic<sup>11</sup> for the OLS estimate of  $\beta_x$  equals 5.724, compared to a t-statistic of 2.253 for  $\beta_{dp}$ . These results strongly indicate the need to isolate variation in the dividend-price ratio that is due to variation in expected returns from that which is due to changing forecasts of the dividend growth rate.

[INSERT TABLE 1 HERE]

The slope in the regression using the adjusted version of  $dp_t$  ( $\beta_x = 0.084$ ) is very close to that of the unadjusted version ( $\beta_{dp} = 0.095$ ). This fact, together with the result that  $x_t$  is more volatile than  $dp_t$ , implies that estimates of future returns from the adjusted dividend-price ratio vary more than those of the unadjusted ratio. This increase in variability of return forecasts is the reason underlying the superior performance of the adjusted version.

The difference in the out-of-sample performance of the estimators is remarkable. The adjusted version survives the critique of Goyal and Welch (2008) in that it delivers a positive and very large out-of-sample  $R^2$  of 12.35% as opposed to the unadjusted version which has a negative  $R^2_{OOS}$  of  $-2.94\%$ . Figure 4 below plots both out-of-sample forecasts and realized returns. Although the magnitudes by which forecasts of future returns change are not as large as the magnitudes of realized returns, the sign of the change is typically the same.

[INSERT FIGURE 4 HERE]

Our regressions of returns on the the modified dividend-price ratio can be seen as applying restricted least squares to a multiple regression of returns on the simple dividend-price ratio and on the dividend adjustment  $\frac{\bar{g}_t}{1-\rho_t}$ , imposing that both variables have the same

---

<sup>11</sup>These statistics are computed from Newey-West adjusted standard errors.

coefficient. Our simple model provides the economic rationale for imposing such a restriction - it arises from the structure for expected returns and expected dividend growth that we imposed in our economy.

Implicitly, we ran the following regressions imposing that  $\beta_1 = \beta_2$ :

$$r_{t+1} = \beta_0 + \beta_1 dp_t + \beta_2 \frac{\bar{g}_t}{1 - \bar{\rho}_t} + \varepsilon_{t+1}^{Free} \quad (11)$$

Ex ante, if our model provided a poor description of reality, we would expect a considerable gain in the in-sample goodness of fit from estimating regression (11) instead of regression (10). However, unrestricted least squares estimation yields an in-sample  $R^2$  of 16.96%, which is not very different from the  $R^2$  of the restricted regression, 16.27%. Out-of-sample, the unrestricted version works far worse. When we estimate regression (11) we obtain an out-of-sample  $R^2$  of 1.72%, versus an  $R^2$  of 12.35% when we run regression (10). The likely explanation for this is sampling error due to the fact that we are using a relatively small sample. This result stresses another advantage of our model, its out-of-sample robustness.

## 5 Interpreting our results

A close look at equation (5) reveals that we are only able to improve our forecasts of future returns by finding a way to measure variation in future dividend growth rates. We now look at the data and verify that lags of past dividend growth are indeed important to forecast future dividend growth. We report regression  $R^2$  statistics from OLS estimation of a VAR model where the log of the dividend price ratio, log dividend growth and log returns are regressed on combinations of lags of the dividend-price ratio, dividend-growth rates and our measure of future dividend growth,  $\bar{g}_t$ .

The results, displayed in table 2 below, provide some evidence of the importance of using lags of past dividend growth. As we depart from the baseline case where only a lag of the dividend-price ratio is used to forecast dividend growth, the  $R^2$  increases substantially from 3% to around 22%, when one lag of  $\Delta d_t$  is added, and to around 42%, when ten lags of  $\Delta d_t$  are added. This is reflected in the ability to forecast returns, albeit not as strongly, where the  $R^2$  increases from 9.7% to 14.4%. When we use  $\bar{g}$ , implicitly restricting the coefficients on all lags of dividend growth to be the same, the gain in terms of predictability of dividend growth is substantially lower, with an  $R^2$  of 13.3%. However, the improvement in forecasting returns is still very significant when compared to the case of using only the lagged dividend-price ratio as forecasting variable.

[INSERT TABLE 2 HERE]

To be able to say something about the statistical significance of the  $R^2$  statistics we reported, we now test the joint significance of the extra regressors we add when forecasting dividend growth. As can be seen from table 3 below, the increase in goodness of fit from adding lags of past dividend growth to forecast dividend growth is backed by the fact that these regressors are statistically significant. The first row of table 3 tells us that it is better to include a lag of dividend growth – it is significant at the 1% level, and the second row says the same about including 10 lags of dividend growth. The third row tells us that adding ten lags is marginally better than only adding one, since the extra nine lags are still significant at the 10% level. Finally, the last row shows that adding  $\bar{g}_t$  to forecast dividend growth is statistically significant.

[INSERT TABLE 3 HERE]

Ultimately, these results tell us that past dividend growth matters for predicting future dividend growth, and that we need more than just one lag to do it. A 10-year moving average of dividend growth is a simple, estimation-free way to do it. This is especially important in predicting returns out of sample where estimation error can deeply affect the robustness of the forecasts.

## 6 Estimating the full model

So far we have not made full use of all the model's implications. We only used the model to inspire a correction to the dividend-price ratio which better reflects fluctuations in expected returns. We now go deeper and fully estimate the model from the data.

In this section, we derive a relation between the dividend-price ratio, forecasts of dividend-growth, and the sample mean of the dividend-price ratio that allows us to estimate the values of the parameters governing the process for conditional expected returns,  $a$  and  $b$ . With these parameters in hand, we compute estimates of expected returns. We assess the forecasting performance of these estimates of expected returns both in- and out-of-sample. Finally, we test the statistical significance of our estimates of  $a$  and  $b$ .

### 6.1 Estimation by non-linear least squares

When we plug equation (7), identifying expected returns, into equation (1) we obtain the following relation:

$$(1 - b\bar{\rho}_{t+1})(x_{t+1} + \kappa_{t+1}) - \frac{a\bar{\rho}_{t+1}}{(1-\bar{\rho}_{t+1})} = a + b \left[ (1 - b\bar{\rho}_t)(x_t + \kappa_t) - \frac{a\bar{\rho}_t}{(1-\bar{\rho}_t)} \right] + \varepsilon_{t+1}^\mu. \quad (12)$$

Although this equation is non-linear in  $a$  and  $b$ , these parameters can be estimated via non-linear least squares. We require that  $|b| \leq 1$  to rule out explosive expected returns.

Additionally, we require that  $a > 0$  to ensure that the unconditional mean of expected returns is positive.<sup>12</sup> The problem we solve is:

$$\operatorname{argmin}_{a \geq 0, |b| \leq 1} \left\{ \frac{1}{T-1} \sum_{t=1}^{T-1} (\varepsilon_{t+1}^\mu)^2 \right\}$$

where:

$$\varepsilon_{t+1}^\mu = (1 - b\bar{\rho}_{t+1}) [x_{t+1} + \kappa_{t+1}] - b(1 - b\bar{\rho}_t) [x_t + \kappa_t] - a \left( 1 + \frac{\bar{\rho}_{t+1}}{(1 - \bar{\rho}_{t+1})} - b \frac{\bar{\rho}_t}{(1 - \bar{\rho}_t)} \right).$$

We estimate the parameters first by using the full-sample and later by using expanding samples as we did for the previous section. Interestingly, the results for  $a$  and  $b$  are always the same. Our estimates of  $a$  and  $b$  are:<sup>13</sup>

$$\hat{a} = 0 \quad \hat{b} = 1$$

This result would imply that the process for  $\mu_{t+1}$  is given by:

$$\mu_{t+1} = \mu_t + \varepsilon_{t+1}^\mu \tag{13}$$

which is the expression of a random walk without drift. This is consistent with many studies that find that conditional expected stock returns are highly persistent.<sup>14</sup> Predictive regressions for stock returns are usually based on using highly persistent variables to capture variation in investors' expectations regarding future returns. These include variants of the

---

<sup>12</sup> $E[\mu_t] = \frac{a}{1-b}$ .

<sup>13</sup>Even when we use unrestricted NLS, the results do not vary significantly in subsamples. In that case, our results for  $a$  range from  $-0.0128$  to  $-0.0116$  and for  $b$  range from  $1.053$  to  $1.055$ .

<sup>14</sup>The first paper to look more deeply into the relation between the persistence of conditional expected returns and realized returns was Fama and French (1988).

dividend yield, the earnings-price or book to market ratios, yields on T-bills and yield spreads, among others.<sup>15</sup> Many of these have high estimated autocorrelation coefficients.

The idea that expected returns follow a random walk is controversial. On the one hand, it is sensible since it implies that, at time  $t$ , agents expect future 1-year returns to be the same in any given future year. This means investors expect to require the same compensation for risk ten years from now as they do for the current year. For this not to happen, they would need to expect that their preferences towards risk would shift in the future. It seems reasonable to assume that the variation in the degree of risk aversion is unpredictable. The controversy arises when we look at the statistical implications of a random walk in expected returns for stock prices and expected returns volatility. A random walk has unbounded volatility, which implies the possibility of unrealistically high or low values for expected returns. However, we can interpret our estimates of a random walk in expected returns as a computationally simple, and reasonably approximate way of capturing the extremely high persistence of expected returns –  $b$  very close to 1 – and obtain good out-of-sample performance.

## 6.2 Forecasting performance

With the model parameters at hand, we can estimate expected returns and use them to forecast realized returns. Replacing  $\hat{a} = 0$  and  $\hat{b} = 1$  into equation (7) yields the following equation for expected returns:

$$\hat{\mu}_t = \bar{g}_t + (1 - \bar{\rho}_t)(dp_t + \kappa_t). \quad (14)$$

Interestingly, this equation is closely related to the sum-of-the-parts approach in Ferreira and Santa-Clara (2009). In fact, equation (14) is the first-order Taylor expansion of  $\mu_t =$

---

<sup>15</sup>An extensive list of all these variables and a critique of their performance can be found in Goyal and Welch (2007)

$\bar{g}_t + \ln\left(1 + \frac{D_t}{P_t}\right)$  around the sample mean of the dividend-price ratio.

This equation allows us to compute expected returns from the investor forecast of dividend growth, the historical average of the dividend-price ratio, and the current level of the dividend-price ratio. If investor forecasts of dividend growth were constant, expected returns would be a linear function of the current dividend-price ratio. In that case, we would not be able to improve forecastability with equation (14). However, using this equation to forecast returns yields an  $R^2$  of 9.22% in sample and an out-of-sample  $R^2$  of 18.62%.

To further check the robustness of our results, we split the sample into two halves and compute the same statistics. The in-sample  $R^2$ 's are 11.92% for the first subsample and 5.60% for the second subsample. The out-of-sample  $R^2$  remain very positive: 22.73% for the 1938 – 1977 subsample and 12.55% for the 1978 – 2007 sample. Table 4 summarizes the results.

[INSERT TABLE 4 HERE]

The last two columns of Table 4 show that the averages of our estimates of expected returns are inferior to the averages of realized returns in all samples. This means investors were positively surprised by the performance of the stock market, a conclusion supported by the results in Fama and French (2002).

### 6.3 Testing the random walk in expected returns

In this section, we assess the significance of our parameter estimates. We do it in two ways. First, we test the statistical significance of our estimates by using a Lagrange multiplier test. We find that our results are statistically significant. Then, we test the hypothesis that expected returns follow a random walk using the Augmented Dickey-Fuller and Phillips-

Perron tests on the estimated expected returns.<sup>16</sup> We find support for the random walk hypothesis in both tests.

The fact that we reach a corner solution demands that we test the significance of the restrictions that  $a = 0$  and  $b = 1$ . We can test these two restrictions based on the Lagrange multiplier statistic.<sup>17</sup> Under the null hypotheses, the LM statistic has a limiting chi-squared distribution with one degree of freedom (the number of linear restrictions). Table 5 summarizes the results. The 95% critical value of a  $\chi_1^2$  distribution is 3.84, which is clearly larger than the values obtained for the LM statistic both using the full sample and using each of the expanding subsamples.<sup>18</sup>

[INSERT TABLE 5 HERE]

To test the hypothesis that conditional expected returns follow a random walk we look for a unit-root in the process followed by  $\hat{\mu}_t = \bar{g}_t + (1 - \bar{\rho}_t)(dp_t + \kappa_t)$ . We use the Augmented Dickey-Fuller and the Phillips-Perron tests. The first test adds lags of the first difference of expected returns to the auto-regression and the second adjusts the t-statistic to take serial correlation of the differenced series into account. Both of them fail to reject the hypothesis that  $\hat{\mu}_t$  follows a random walk at the 5% level. The results are in Table 6. These results provide evidence for the case of an extremely persistent process for expected returns –  $b$  very close but less than 1 – supporting our simple and intuitive model.

[INSERT TABLE 6 HERE]

---

<sup>16</sup>See Hamilton (1994), chapter 17.

<sup>17</sup>See Greene (2008) for a summary of the usage of LM statistics in non-linear least squares estimation.

<sup>18</sup>The significance of our results is not surprising since the unrestricted estimates are not much different from the restricted ones. The unrestricted estimates using the full sample were  $a = -0.0126$  and  $b = 1.0530$ .

## 7 Conclusion

We propose a simple process for expected returns and an even simpler, yet reasonable, proxy for investor forecasts of dividend growth rates. Our model implies that we can separate the component of the dividend-price ratio that varies with expected returns from that which varies with forecasts of dividend growth. When we remove the dividend growth component from the dividend-price ratio and use the resulting variable to forecast returns, the results are far superior to those obtained when using the unadjusted dividend-price ratio. These results indicate that forecasting dividend growth matters when predicting stock returns and provide strong evidence for return predictability. Our conclusions have obvious implications for portfolio allocation decisions and equity valuation.

## 8 References

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

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

Binsbergen, Jules van, and Ralph Koijen, 2009, Predictive regressions: A present-value approach, *Journal of Finance*, forthcoming.

Campbell, John, 1987, Stock returns and term structure, *Journal of Financial Economics* 18, 373—399.

Campbell, John, and Robert Shiller, 1988, Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661—676.

Campbell, John, and Motohiro Yogo, 2006, Efficient tests of stock return predictability, *Journal of Financial Economics* 81, 27—60.

Cavanagh, Christopher, Graham Elliott, and James Stock, 1995, Inference in models with nearly integrated regressors, *Econometric Theory* 11, 1131—1147.

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

Dow, Charles, 1920, Scientific stock speculation, *The Magazine of Wall Street*.

Fama, Eugene, and Kenneth French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3—25.

Fama, Eugene, and Kenneth French, 2002, The equity premium, *Journal of Finance* 57, 637—659.

Ferreira, Miguel and Pedro Santa-Clara, 2009, Forecasting Stock Market Returns: The Sum of the Parts is More than the Whole, FEUNL working paper.

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

Greene, William, 2008, *Econometric Analysis*, 6th Edition, Prentice Hall.

Hamilton, James, 1994, *Time-Series Analysis*, Princeton University Press.

Hodrick, Robert, 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Review of Financial Studies* 5, 357—386.

Lettau, Martin and Sydney Ludvigson, 2005, Expected Returns and Expected Dividend Growth, *Journal of Financial Economics*, 76, 583-626;

Lewellen, Jonathan, 2004, Predicting returns with financial ratios, *Journal of Financial Economics* 74, 209—235.

Menzly, Lior; Santos, Tano and Pietro Veronesi, 2004, Understanding Predictability, *Journal of Political Economy*, 112, 1-47;

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

Pastor, Lubos and Stambaugh, Robert F., Predictive Systems: Living with Imperfect Predictors, *Journal of Finance* 64, 1583–1628.

Stambaugh, Robert, 1999, Predictive regressions, *Journal of Financial Economics* 54, 375—421.

Torous, Walter, Rossen Valkanov, and Shu Yan, 2004, On predicting returns with nearly integrated explanatory variables, *Journal of Business* 77, 937—966.

**Table 1**  
**Predictive regressions**

Table 1 gives us the slopes, associated t-statistics, in-sample and out-of-sample  $R^2$ 's from running the regressions in equations (9) and (10) with 1-year horizon. The sample starts in 1938 because the first forecasts of future dividend growth and of  $\rho$  are done with the first 10 observations. The t-stats are computed with Newey-West adjusted standard errors. The first estimation for the computation of  $R^2_{OOS}$  is done with 20 observations from 1937 to 1957.

Predictor Variable	Slope	t-statistic	$R^2$	$R^2_{OOS}$
$x_t$	0.0894	5.724	16.27%	12.35%
$dp_t$	0.0949	2.253	7.88%	-2.95%

**Table 2**  
**Regression R<sup>2</sup> from VAR model**

Table 2 presents the regression R<sup>2</sup>s from OLS estimation of a VAR model where the log of the dividend price ratio, log dividend growth and log returns are regressed on combinations of lags of the dividend price ratio, dividend growth, and the 10-year moving average of dividend growth,  $\bar{g}_t$ .

Regressors	Regressands		
	$dp$	$\Delta d$	$r$
1 lag $dp$	87.6%	3.0%	9.7%
1 lag $dp$ and 1 lag $\Delta d$	87.8%	22.0%	10.0%
1 lag $dp$ and 10 lags $\Delta d$	89.1%	41.8%	14.4%
1 lag $dp$ and $\bar{g}$	88.6%	13.3%	12.6%

**Table 3**  
**Model comparisons**

Table 3 presents  $F$ -statistics that allow us to compare different model specifications for forecasting dividend growth. We compare restricted versus unrestricted regressions and test whether the additional regressors are jointly significant. Under the null hypothesis that the restricted model is the true one, the  $F$ -stat follows an  $F(n - k, J)$  distribution where  $k$  is the number of regressors in the unrestricted model,  $n$  is the number of observations, and  $J$  is the number of extra regressors. The p-value columns that are significant at the 1%, 5% and 10% level are marked with one, two, and three (\*), respectively.

Models being compared				
Restricted		Unrestricted	F-stat	p-value
1 lag $dp$	vs.	1 lag $dp$ and 1 lag $\Delta d$	14.364	0.04% <sup>(*)</sup>
1 lag $dp$	vs.	1 lag $dp$ and 10 lags $\Delta d$	3.341	0.22% <sup>(*)</sup>
1 lag $dp$ and 1 lag $\Delta d$	vs.	1 lag $dp$ and 10 lags $\Delta d$	1.898	7.39% <sup>(***)</sup>
1 lag $dp$	vs.	1 lag $dp$ and $\bar{g}$	7.019	1.03% <sup>(**)</sup>

**Table 4****Predictability from the full-blown estimation of the model**

Table 2 gives us some measures of predictive performance of using  $\hat{\mu}_t = \bar{g}_t + (1 - \bar{\rho}_t)(dp_t + \kappa_t)$  to forecast returns. Three samples are considered.  $\bar{\mu}$  and  $\bar{r}$  are the sample averages of expected returns and realized returns, respectively.

Sample	$R_{OOS}^2$	In-sample $R^2$	$\bar{\mu}$	$\bar{r}$
Full-sample	18.62%	9.22%	8.03%	10.96%
1938 – 1977	22.73%	11.92%	8.11%	9.96%
1978 – 2007	12.55%	5.60%	8.19%	11.65%

**Table 5****Expected returns: NLS estimation**

Table 3 gives us the Lagrange multiplier statistics associated with the null hypotheses that  $a = 0$  and  $b = 1$  when estimating equation (12) via non-linear least squares. Under the null, the LM statistics have a limiting chi-squared distribution with one degree of freedom. The 95% critical value of the  $\chi_1^2$  distribution is 3.84.  $\text{Max } \lambda_{LM}$  is the maximum value attained by the LM statistic when the model was estimated with an expanding sample window.

Null hypothesis	Full-sample $\lambda_{LM}$	Max $\lambda_{LM}$
$a = 0$	0.0867	0.1059
$b = 1$	0.0000	0.2246

**Table 6**

**A unit-root in expected returns**

Table 4 gives us the test statistics for the Phillips-Perron and the Augmented Dickey-Fuller tests. Both tests were conducted for the no intercept no trend case. The DF test was undertaken with two lags of the first difference. Critical values are from Mackinnon (1996).

Test	Statistic	10% Critical value
Phillips-Perron	-0.923	-1.618
Augmented Dickey-Fuller	-0.652	-1.618

**Figure 1**

**Adjusted vs. unadjusted dividend-price ratio**

This graph plots end-of-year values for the unadjusted and adjusted versions of the log dividend-price ratio,  $dp_t$  and  $x_t$ .

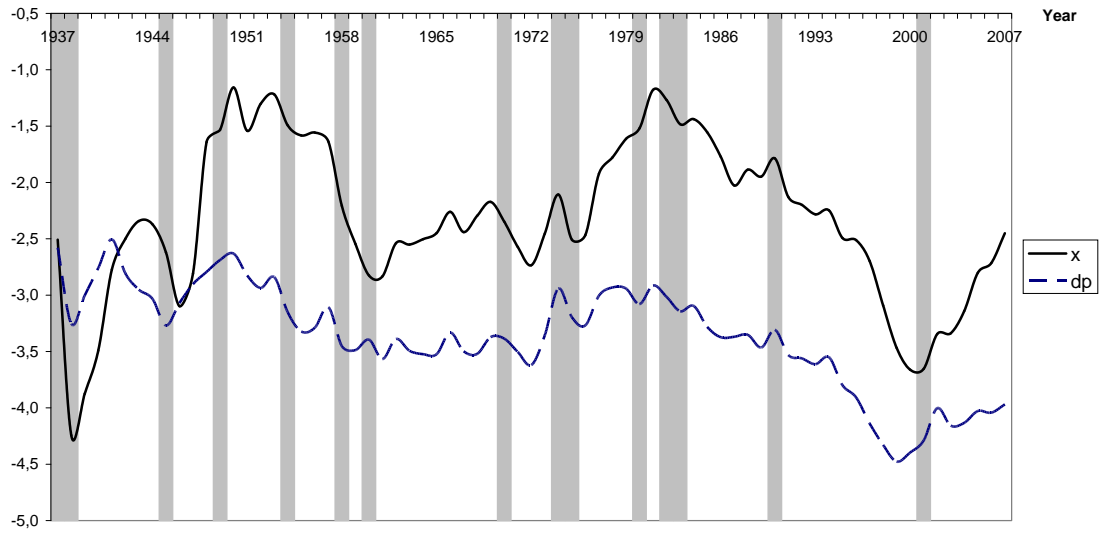
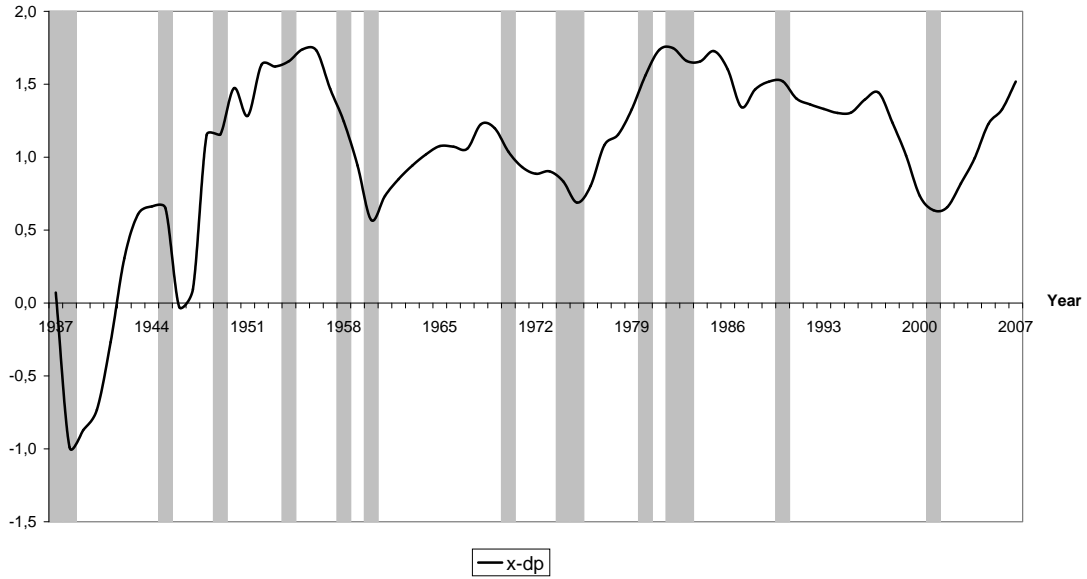


Figure 2

Stream of future dividend growth

This graph plots the difference between  $x_t$  and  $dp_t$ , which is given by the stream of expected future cash-flow growth rates at time  $t$ ,  $\frac{\bar{g}_t}{1-\bar{\rho}_t}$ .



**Figure 3**

**Forecasts of dividend growth**

This graph plots forecasts of dividend growth at time  $t$  from using a 10-year equally-weighted moving average,  $\bar{g}_t$ .

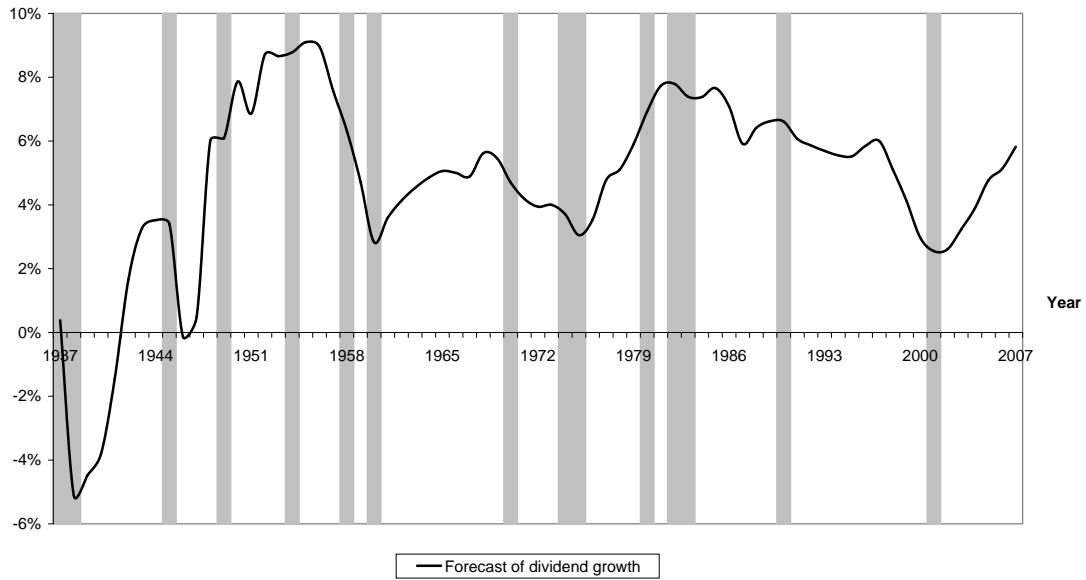


Figure 4

Expected vs. realized returns

This graphs plot realized yearly returns and expected returns from different estimators. The first graph plots expected returns from expanding window predictive regressions using  $x_t = dp_t + \frac{\bar{g}_t}{1-\rho_t}$  as the predictor variable. The second graph plots expected returns from using  $dp_t$  as predictor variables. Finally, the third graph plots estimates of expected returns from the sample mean of realized returns.

