The Predictive Power of the Dividend Risk Premium

Abstract

We show that the dividend growth rate implied by the futures market is informative about (i) the expected dividend growth rate and (ii) the expected dividend risk premium. We model the dividend risk premium and explore its implications for the predictability of dividend growth and aggregate stock returns. We show that accounting for the dividend risk premium strengthens the predictability of dividend growth and aggregate returns both in- and out-of-sample. Economically, we find that a market timing investor who accounts for the time varying dividend risk premium realizes an additional utility gain of 1.43 % per year.

JEL classification: C22, C53, G12, G13, G17

Keywords: Dividend risk premium, dividend strip, predictability, present value model

I Introduction

The dividend growth forecast implied by the futures market is informative about the *risk-adjusted* expectations of future dividend growth. More specifically, the implied dividend growth rate (ig) contains information about (i) the expected dividend growth rate and (ii) the expected dividend risk premium. This insight raises a number of questions. For instance, is ig mainly informative about the expected dividend growth rate or the expected dividend risk premium? What are the theoretical implications of the expected dividend risk premium for the predictability of dividend growth rates and aggregate stock returns?

Addressing these questions is important because a time varying expected dividend risk premium confounds the information content of ig for the expected dividend growth rate. Thus, it might be important to account for these variations when using ig to forecast dividend growth. Furthermore, the logic of present value models suggests that the dividend price (dp) ratio reveals information about the difference between expected stock returns and expected dividend growth rates (Campbell and Shiller, 1988). To the extent that the expected dividend growth rate is time-varying, we need to correct the standard dp ratio for these variations in order to strengthen the predictability of stock returns (Campbell, 2008).

This paper makes three contributions to the literature. First, we formally show that ig contains information about the future (i) dividend growth rate and (ii) dividend risk premium. Using a dataset of intraday futures transaction prices covering the period 1997–2014, we show that 29 % and 71 % of the fluctuations in ig are related to the dividend growth and the dividend risk premium (drp), respectively. This leads us to conclude that the drp does not only move over time but it is also the main driving force of ig.

Second, we propose a model for the dynamics of the drp. In particular, we assume it depends on the lagged ig and the lagged drp. Although admittedly simple, this 2-factor model achieves a satisfactory empirical performance. This is evidenced by an R^2 of around 60 %. We use this parsimonious model to analyze the predictability of dividend growth. We show that the lagged dividend risk premium corrected implied growth rate (ig^{corr}), a linear combination of the lagged ig and the lagged drp, should predict dividend growth with a slope coefficient equal to 1. We find empirical evidence in support of this prediction. A regression of 1-month dividend growth rates on a constant and the lagged ig^{corr} yields a positive and statistically significant slope estimate (0.97). We test the hypothesis that the slope parameter equals 1 and find that we cannot reject this null. Examining the predictive power of ig and ig^{corr}, we find that they yield R^2s of 6.69 % and 11.41 %, respectively. The superior forecasting performance of ig^{corr} is discernible not only in-sample but also out-of-sample. By accounting for the lagged drp, we are able to significantly reduce the mean squared error (MSE) of ig by 7.45 %.

Third, we develop a present value model to study the predictability of aggregate stock returns. Our model predicts that the lagged corrected dividend price (dp^{corr}) ratio, an affine function of (i) the lagged standard dp ratio, (ii) the lagged ig and (iii) the lagged drp, forecasts returns with a positive sign. A regression of 1-month returns on a constant and the lagged dp^{corr} ratio yields a positive and statistically significant slope estimate (0.18). We compare the predictive power of the standard dp ratio (which ignores the lagged values of ig and drp), the dp^{ig} ratio (which ignores the lagged drp) and the dp^{corr} ratio. Our results reveal that the dp^{corr} ratio delivers the highest R^2 (1.91 %) of all three forecasting variables in-sample. Out-of-sample, we find that the dp^{ig} ratio reduces the MSE of the standard dp ratio by 1.17 %. More importantly, the dp^{corr} ratio leads to a reduction in the MSE that is twice larger than afforded by the dp^{ig}. This improvement matters from economic standpoint. Relative to a strategy based on the dp^{ig} ratio, an investor who uses the dp^{corr} ratio as timing signal realizes additional utility gains of 1.43 % per year. Collectively, these results highlight the relevance of the lagged drp.

Our paper is most germane to the innovative work of Golez (2014), who uses ig to correct the standard dp ratio. In a similar vein, Bilson et al. (2015) and Zhong (2016) show that the dividend yield implied by derivatives prices predicts returns. A common feature of these studies is that they assume that dividend risk is not priced.

Our main contribution is to provide a formal treatment of the expected *drp*. We develop a framework that allows us to study its implications for the predictability of dividend growth and aggregate returns.

Our paper also relates to the literature on dividend forecasting. Lintner (1956), Marsh and Merton (1987) and Garrett and Priestley (2000) propose to use accounting data, e.g. earnings, to predict dividend growth rates. We complement this body of works by showing how to obtain dividend growth forecasts from equity futures prices. Because futures prices are (i) forward-looking and (ii) available at high-frequencies (relative to accounting data), our framework could help researchers obtain more timely dividend growth forecasts at fairly high frequencies, e.g. daily. This could prove very useful when performing event studies for example.

Our work contributes to a broader research agenda emphasizing that derivatives prices are informative about *risk-neutral* expectations, whereas for most practical purposes, one is interested in the physical expectations. The risk premium drives a wedge between the two expectations. Borovicka et al. (2015) and Ross (2015), among others, discuss conditions under which it may or may not be possible to "recover" the physical probability distribution from derivatives prices. Several studies rely on historical data to pin down the dynamics of the risk premium. For instance, Piazzesi and Swanson (2008) focus on the Fed fund futures market and propose a parsimonious time-series model for the expected risk premium. They then use their model to correct the forecasts implied by the Fed fund futures market. Chernov (2007) and Prokopczuk and Wese Simen (2014) show how to correct for the variance risk premium when using implied variance to predict realized variance. Our paper is similar in spirit to these works. We posit a time-series model for the expected *drp* and analyze its implications for the predictability of dividend growth and aggregate returns.

The remainder of this paper proceeds as follows. Section II presents our theory and describes the dataset. Sections III and IV discuss our main empirical results. Finally, Section V concludes.

II Methodology and Data

This section begins by presenting our methodology. We formally show that *ig* contains information about (i) the expected dividend growth rate and (ii) the expected *drp*. We then propose a parsimonious model to capture the dynamics of the *drp* and present an empirically testable model of dividend growth rates and returns. Finally, we introduce our dataset.

II.*A*. Methodology

The starting point of our methodology is the cost-of-carry relationship, which posits that the market price of a futures contract can be obtained as follows:

$$F_t = P_t e^{rf_t} - \mathbb{E}_t^Q(D_{t+1}) \quad (1)$$

where F_t is the price at time t of the futures contract that expires at the end of the next period, i.e. $t+1$. P_t is the price of the underlying asset at time t. rf_t denotes the 1-period riskless rate observed at t.[1] $\mathbb{E}_t^Q(D_{t+1})$ is the dividend that a risk-neutral (Q) investor expects to receive from the underlying security at expiration.

In order to clearly show the link between the futures price and the next-period dividend, it is useful to introduce the dividend strip. This financial asset entitles the holder to the dividends paid by the underlying index during the life of the strip (van Binsbergen et al., 2012). We can obtain the market price of dividend strips using two valuation methods: the martingale valuation approach and the standard present value method.

According to the martingale valuation framework of Cox and Ross (1976) and Harrison and Pliska (1981), we can price financial assets as if investors were risk-neutral. A direct implication of this result is that the market price of the dividend strip equals the cashflow that the risk-neutral investor expects to receive discounted

[1] Throughout this paper, we adopt the timing convention that interest rates are given the subscripts for the time when they are observed. As a result, our notation indicates that the interest rate is observed at t, even though it is realized at time $t+1$.

to the present at the riskless rate:

$$STRIP_t = e^{-rf_t}\mathbb{E}_t^Q(D_{t+1}) \quad (2)$$

where $STRIP_t$ is the time t market price of the dividend strip expiring at the end of the next period. All other parameters are as previously defined.

Substituting Equation (1) into the expression above yields:

$$STRIP_t = P_t - e^{-rf_t}F_t \quad (3)$$

The standard present value approach determines the market price of assets by directly discounting the expected cashflows (under the physical probability measure) at the expected rate of return. The following expression formalizes this idea:

$$STRIP_t = e^{-\mathbb{E}_t(drp_{t+1})}\mathbb{E}_t(D_{t+1}) \quad (4)$$

where $\mathbb{E}_t(drp_{t+1})$ denotes the conditional expectation of the future rate of return on the dividend strip.[2,3] $\mathbb{E}(D_{t+1})$ is the dividend the investor expects the underlying security to pay at $t+1$.

Putting together Equations (3) and (4), we derive the following result:

$$\log(\mathbb{E}_t(D_{t+1})) - \mathbb{E}_t(drp_{t+1}) = \log(P_t - e^{-rf_t}F_t) \quad (5)$$

Next, we subtract $\log(D_t)$ from both sides of Equation (5) and ignore the Jensen

[2]Throughout this paper, we refer to the discount rate of the dividend strip as the dividend risk premium (*drp*). Strictly speaking, the discount rate is the sum of the dividend risk premium and the riskless rate. Because interest rates display very little variations in the time-series, we commit this slight abuse of terminology. See Cochrane (2011) for a conceptually similar terminology. Note also that in this paper, we take the *drp* to mean the realized (rather than expected) return of the dividend strip. To denote the expected return of the dividend strip, we use the expression "expected *drp*".

[3]It is worth highlighting that, unlike the risk-free rate, the *drp* is only observed ex-post, i.e. at time $t+1$.

inequality term:[4]

$$\mathbb{E}_t(\Delta d_{t+1}) - \mathbb{E}_t(drp_{t+1}) \approx \underbrace{\log(P_t - e^{-rf_t}F_t) - \log(D_t)}_{\text{Implied Growth}}$$

$$\mathbb{E}_t(\Delta d_{t+1}) - \mathbb{E}_t(drp_{t+1}) \approx ig_t \tag{6}$$

where $\mathbb{E}_t(\Delta d_{t+1})$ denotes the time t expectation of the 1-period dividend growth rate: $\mathbb{E}_t(\Delta d_{t+1}) = \mathbb{E}_t(\log(D_{t+1})) - \log(D_t)$. ig_t denotes the dividend growth rate implied by the futures market: $ig_t = \log(P_t - e^{-rf_t}F_t) - \log(D_t)$. All other variables are as previously defined.

The expression above reveals that ig is the risk-adjusted expectation of future dividend growth. In particular, ig is positively related to the expected dividend growth and negatively related to the expected drp. An implication of this result is that a time varying expected drp could potentially obscure the information content of ig for the expected dividend growth.

Despite its clear insights, the expression above is merely an accounting identity that is of limited practical use. The reason for this is that the terms on the left of the equality sign are conditional expectations, which are not directly observable. In order to obtain an empirically testable economic model, one needs to impose a structure on how the conditional expectation of the drp is generated.[5] We simply assume that the drp depends on a constant, the lagged ig and the lagged drp (which is included in the information set at time t):

$$drp_{t+1} = \phi_0 + \phi_1 ig_t + \phi_2 drp_t + \epsilon^{drp}_{t+1} \tag{7}$$

[4]It is standard in the literature to ignore the Jensen inequality term, e.g. Golez (2014). We conduct a simple simulation exercise which reveals that the approximation error is small. Most important for our purposes, it displays very little variations. A constant approximation error will not materially affect our results since we include an intercept in all regression models.

[5]One may wonder why we do not use the expected dividend growth, derived from time-series models for example, and recover the expected drp by manipulating the identity in Equation (6). We do not pursue this approach because, if one already has an estimate of the expected dividend growth, then there is no need to use ig (and correct for the drp). This is because the dividend growth forecasts could be used directly in the present value model. Furthermore, Golez (2014) shows that, in-sample, ig outperforms the model of Lacerda and Santa-Clara (2010), which relies on historical dividend growth rates. Our aim is to further improve the forecasting ability of ig by explicitly accounting for the expected drp.

where ϕ_0, ϕ_1 and ϕ_2 are constant parameters.

One may take the view that this 2-factor model is too simplistic. We agree. We deliberately keep the model in Equation (7) simple in order to facilitate the exposition of the paper. As we shall see later, this parsimonious model adequately captures the dynamics of the *drp*. It is, however, worth pointing out that the framework can easily accommodate additional forecasting variables. If one has a view on other variables that could predict the *drp*, these forecasting variables could be easily included in our framework. For instance, it might be that the default and the term spreads are related to the expected *drp*. Including these variables could improve the empirical results presented in this paper. In future work, it would be interesting to explore this avenue.

We now motivate the choice of the 2 factors. Our assumption that *ig* predicts the *drp* is directly motivated by Equation (6), which shows that *ig* is negatively related to the expected *drp*. We also note that our assumption that the *drp* depends on its lagged observation is in keeping with previous works. Because the *drp* is essentially the return to a buy-and-hold trading strategy, our modelling approach is consistent with previous studies, which typically assume that returns have an autoregressive component (van Binsbergen and Koijen, 2010; Lacerda and Santa-Clara, 2010; Golez, 2014). Armed with the model above and the identity presented in Equation (6), we are now in a position to discuss our first proposition.

Proposition 1: *The lagged corrected implied growth rate (ig^{corr}), an affine function of (i) the lagged implied dividend growth (ig) and (ii) the lagged dividend risk premium (drp), predicts the next-period dividend growth rate.*

$$\Delta d_{t+1} \quad = \quad \underbrace{\phi_0 + (1+\phi_1) ig_t + \phi_2 drp_t}_{ig_t^{corr}} + \epsilon_{t+1}^{\Delta d} \tag{8}$$

Proof: See Appendix A.1.

This proposition presents our first empirically testable prediction: ig^{corr} predicts dividend growth with a positive sign. Hence, we conduct our statistical inference using a 1-sided alternative hypothesis. Moreover, the theory suggests that the coefficient loading on ig^{corr} is not statistically distinguishable from 1. Furthermore, the proposition makes interesting predictions about the slope coefficients of an unconstrained regression of dividend growth on a constant, the lagged ig and the lagged drp. These slope parameters should not be statistically distinguishable from $1 + \phi_1$ and ϕ_2, respectively. We expect the first slope parameter to be lower than 1 because ϕ_1 should be negative. Indeed, economic theory posits a negative relationship between ig and the expected drp (see Equation (6)). It is also reasonable to expect that the drp, which is the return to a buy-and-hold strategy, is not driven by an explosive or unit root process. This suggests that the magnitude of ϕ_2 should also be lower than 1. Using the estimates of ϕ_1 and ϕ_2 (see Equation (7)), we shall empirically test these two theoretical restrictions.

As pointed out by Campbell (2008), the predictability of dividend growth has important implications for forecasts of stock returns. We formally explore these implications by developing a present value model.

We define the next-period return (r_{t+1}) as follows:

$$\begin{aligned} r_{t+1} &= \log\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right) \\ r_{t+1} &= \log\left(1 + e^{dp_{t+1}}\right) + p_{t+1} - p_t \end{aligned}$$

where P_{t+1} and D_{t+1} denote the stock price and dividend at time $t+1$, respectively. Similarly, P_t represents the stock price at t. The lower case variables indicate a logarithmic transformation: $d_{t+1} = \log(D_{t+1})$, $p_{t+1} = \log(P_{t+1})$ and $p_t = \log(P_t)$. Finally, dp_{t+1} is the dividend price ratio at time $t+1$: $dp_{t+1} = d_{t+1} - p_{t+1}$.

Log-linearizing as in Campbell and Shiller (1988), we obtain the following expression:

$$r_{t+1} \approx k + \Delta d_{t+1} + dp_t - \overline{\rho} dp_{t+1}$$

where k is a constant and $\overline{\rho}$ is the linearization constant computed as follows:

$$\overline{\rho} = \frac{1}{1 + e^{\overline{d-p}}} \tag{9}$$

We exploit the linear recursion above to derive the link between the expected stock return, the expected dividend growth rate and the dividend price ratio:

$$\sum_{j=0}^{+\infty} \overline{\rho}^j \left(\mathbb{E}_t(r_{t+1+j}) - \mathbb{E}_t(\Delta d_{t+1+j}) \right) = \frac{k}{1-\overline{\rho}} + dp_t \tag{10}$$

Equation (10) reveals that, to the extent that the expected dividend growth rate is time varying, the standard dividend price ratio is a noisy proxy for the expected return. Thus, it is important to correct the standard *dp* ratio for fluctuations in expected dividend growth in order to improve the predictability of returns (Campbell, 2008).

We decompose the next-period return (r_{t+1}) into an expected return component (μ_t) and a forecast error (ϵ^r_{t+1}). As is standard in the literature, e.g. Golez (2014), we assume that expected returns and the implied growth rate follow AR(1) processes:

$$r_{t+1} = \mu_t + \epsilon^r_{t+1} \tag{11}$$

$$\mu_{t+1} = \alpha_0 + \alpha_1 \mu_t + \epsilon^{\mu}_{t+1} \tag{12}$$

$$ig_{t+1} = \delta_0 + \delta_1 ig_t + \epsilon^{ig}_{t+1} \tag{13}$$

where all error terms are $i.i.d$ with zero mean. All other variables are as previously defined.

Armed with these additional assumptions, it is straightforward to derive the relationship between the 1-period return on the one hand and the lagged values of the *dp* ratio, *ig* and *drp* on the other. Proposition 2 formalizes this link.

Proposition 2: *The lagged corrected dividend price (dp^{corr}) ratio, which is an affine function of (i) the lagged standard dividend price (dp) ratio, (ii) the lagged implied dividend growth (ig) and (iii) the lagged dividend risk premium (drp)* *forecasts the next-period return.*

$$r_{t+1} = \Psi + (1-\bar{\rho}\alpha_1)\underbrace{\left(\underbrace{dp_t + \frac{(1+\phi_1)ig_t}{1-\bar{\rho}\delta_1}}_{dp^{ig}} + \underbrace{\frac{\bar{\rho}\phi_1\phi_2 ig_t}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)} + \frac{\phi_2 drp_t}{1-\bar{\rho}\phi_2}}_{\textit{risk premium correction}}\right)}_{dp_t^{corr}} + \epsilon^r_{t+1} \quad (14)$$

Proof: See Appendix A.2.

This proposition shows that the standard dividend price ratio alone cannot satisfactorily predict returns. Two adjustments are needed. First, one needs to account for ig to obtain the dp^{ig} ratio.[6] Second, one also needs to exploit the information content of the lagged drp. By making these two adjustments, we obtain the dp^{corr} ratio. If one ignores the lagged drp, i.e. $\phi_2 = 0$, then the dp^{corr} and dp^{ig} ratios are exactly the same. Thus, by comparing the performance of these two forecasting variables, we can shed light on the information content of the lagged drp. If the lagged drp plays an important role, then the dp^{corr} ratio should yield better forecasts of returns than both the dp and dp^{ig} ratios.

A subtle implication of Proposition 2 is that, if the expected return process is not an explosive or unit root process (as we would expect from an economic perspective), i.e. $-1 < \alpha_1 < 1$, the dp^{corr} ratio should predict returns with a positive sign.[7] As a result, we conduct our statistical inference using a 1-sided alternative hypothesis. Another implication of Proposition 2 relates to the slopes of an unconstrained regression of 1-period returns, on a constant, the lagged dp, the lagged ig and the lagged drp. These slope parameters should not be significantly different from $1-\bar{\rho}\alpha_1$, $(1-\bar{\rho}\alpha_1)\left[\frac{1+\phi_1}{1-\bar{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)}\right]$ and $(1-\bar{\rho}\alpha_1)\frac{\phi_2}{1-\bar{\rho}\phi_2}$, respectively.

[6] Our definition of the dp^{ig} nests that of Golez (2014). The author implicitly imposes the restriction that $\phi_1 = 0$. As we shall see later in the paper, this restriction is strongly rejected in the data. This is discernible not only in our study but also in the original work of Golez (2014). Although the author does not present evidence of out-of-sample predictability of dividend growth, our own analysis indicates that imposing this restriction results in poor out-of-sample forecasts of the dividend growth rate. Because the present value model relies on realistic dynamics for dividend growth, we feel compelled to allow ϕ_1 to enter the definition of dp^{ig}.

[7] To see this quickly, notice that the linearization constant ($\bar{\rho}$) is bounded between 0 and 1 (see Equation (9)). Thus, it is straightforward to show that the term $1-\bar{\rho}\alpha_1$ is positive.

II.*B*. Data

We obtain intraday transaction prices (stamped to the minute) on S&P 500 futures contracts and the underlying spot index from Thomson Reuters Tick History (TRTH). Our sample covers the period from May 01, 1997 to December 31, 2014. Although the database contains futures prices from January 1996, it is not until May 1997 that we observe futures contracts of time to maturity greater than 12-month on a monthly basis. Since we are interested in ig of 12-month maturity, we start our sample from May 1997.[8] In doing so, we avoid potential biases in the estimation of the autoregressive parameters induced by missing observations.[9] The S&P 500 futures contracts trade on the Chicago Mercantile Exchange (CME). They expire in March, June, September, December, and the following three Decembers.

We process the dataset as follows. First, we retain only transactions observed between 10:00 and 14:00 local time (van Binsbergen et al., 2012). Notice that both the futures and underlying prices are observed during these trading hours. Thus, our analysis does not suffer from the wildcard feature of US derivatives markets.[10] Second, we match each futures transaction price with the spot index price observed on the same day and at the same time (up to the minute level). By taking this step, we aim to tackle the measurement errors that would arise if the spot and futures prices are observed at asynchronous times.[11]

We proxy the riskless rate with the LIBOR curve, which we also obtain from TRTH. We then merge together the time-series of the riskless rate, the spot and futures prices. For each 3-tuple (futures price, spot price and interest rate of corresponding maturity), we obtain the dividend strip price by plugging the relevant values in Equation (3). Thus, we recover the term structure of dividend strips at

[8] We focus on the 12-month maturity in order to avoid issues related to the seasonality of dividend payments. This is standard in the literature. See Fama and French (1988) or Ang and Bekaert (2007) for example.

[9] Untabulated results reveal, however, that starting the sample in January 1996 leads to similar results.

[10] As discussed in Harvey and Whaley (1992), the S&P 500 spot market closes at 15:00 local time, whereas trading in the derivatives market ends at 15:15, introducing biases in studies that require synchronous observations of spot and derivatives prices.

[11] We refer the interested reader to Boguth et al. (2012) for a study of the impact of asynchronous observations on the dynamics of dividend strips.

the minute level. For each minute, we linearly interpolate the 12-month dividend strip. In order to obtain the monthly dividend strip of annual maturity ($STRIP^A$), we average the prices of the 12-month dividend strips observed on the last five days of each calendar month. By taking the average, we attempt to further mitigate the impact of measurement errors (Golez, 2014).

We obtain the time-series of daily dividends and prices related to the S&P 500 index from Bloomberg.[12] We sum all the intra-month dividends to obtain monthly dividend payments (D^M). The time-series of (annualized) monthly returns is computed as:

$$r_{t+1} = 12 \times \log\left(\frac{P_{t+1} + D^M_{t+1}}{P_t}\right) \tag{15}$$

where r_{t+1} is the 1-period annualized return. For the purpose of our empirical analysis, we take 1-period to mean 1-month. P_{t+1} and D^M_{t+1} denote the stock price and monthly dividend payment related to month $t+1$, respectively. Finally P_t is the stock price observed at the end of month t.

As is standard in the literature, e.g. Ang and Bekaert (2007), we base our analysis on annual dividends (D^A), computed by summing monthly dividends over a trailing window of 12 months. Taking this step ensures we address the issue of seasonality in the dividend series. We then compute the (annualized) 1-month dividend growth rate as follows:

$$\Delta d_{t+1} = 12 \times \log\left(\frac{D^A_{t+1}}{D^A_t}\right) \tag{16}$$

where Δd_{t+1} denotes the monthly growth rate of dividends at $t+1$. D^A_{t+1} and D^A_t are the annual dividends for the periods ending at $t+1$ and t, respectively. Relatedly,

[12] By working with daily dividend payments, we follow existing studies and implicitly assume that the investor "holds" the aggregate stock market index. It is however worth pointing out that, if the investor holds the SPY ETF for example, then dividends are typically paid at a quarterly frequency.

we compute the standard dp ratio as:

$$dp_t = \log\left(\frac{D_t^A}{P_t}\right) \quad (17)$$

We then recover the time-series of ig, by computing the difference between the logarithm of the 12-month dividend strip and that of the annual dividend:

$$ig_t = \log(STRIP_t^A) - \log(D_t^A) \quad (18)$$

Next, we obtain the time-series of the drp:

$$drp_{t+12} \;=\; \log(D_{t+12}^A) - \log(STRIP_t^A) \quad (19)$$

Finally, we use all sample information to estimate the parameters δ_1, ϕ_0, ϕ_1 and ϕ_2.[13] In order to obtain the persistence of ig, i.e. δ_1, we follow Golez (2014) and use successive non-overlapping annual samples. Golez (2014) proposes this approach in order to guard against biases induced by the (i) large overlap between consecutive observations of ig and (ii) potential measurement errors in the implied growth series. To be more specific, we calculate the persistence of ig at the monthly level as follows. We sample all observations of ig observed on Januaries and estimate the model in Equation (13). We repeat these steps for all 12 calendar months and save the corresponding slope estimates. We then average the 12 slope estimates. Since this average corresponds to the AR(12) persistence estimate, we then recover the AR(1) parameter by raising it to the power 1/12. In the data, we find $\delta_1 = 0.92$.[14]

The estimation of ϕ_0, ϕ_1 and ϕ_2 is based on Equation (7). As before, we use non-overlapping annual samples to estimate the relevant parameters. We average the parameter estimates across all 12 possible samples of annual data. Unlike the

[13] When we conduct our analysis out-of-sample, we recursively estimate all parameters. This ensures that we only use the information contained in the training sample. The upshot of this is that our out-of-sample analysis does not suffer from any look ahead bias.

[14] Thus, the monthly persistence (0.72) reported in Table 1, which is simply based on monthly observations of ig and is thus subject to the issues discussed above, is much lower than the 0.92 based on samples of non-overlapping observations. This is consistent with Golez (2014).

estimation of δ_1, we do not convert the annual estimates to the monthly horizon. This is because, each month, we are interested in the *drp* expected at the end of the next 12 months.[15] Thus, ϕ_0, ϕ_1 and ϕ_2 relate to the 12-month rather than the 1-month horizon. We find that $\phi_0 = 0.02$, $\phi_1 = -0.60$ and $\phi_2 = 0.26$. Combining these parameter estimates together with the monthly time-series of ig and realized *drp*, we can recover ig^{corr} (see Equation (8)). Next, we compute the linearization constant $\bar{\rho}$ using the whole sample period (see Equation (9)). We find $\bar{\rho} = 0.98$. Equipped with this information, we then compute the time-series of the dp^{ig} and dp^{corr} ratios (see Equation (14)).

Table 1 summarizes the key statistics of various time-series. For the purpose of predictability, the *drp* matters only if it varies over time. Table 1 shows that the volatility of the *drp* (14 %) is twice larger than the magnitude of its mean. We notice that a buy-and-hold investor who purchases a dividend strip of 12-month maturity realizes a negative return (−6 %). The magnitude and sign of this estimate are broadly comparable to those of the 6-month strip (−4.34 %) reported in Golez (2014).[16]

Although consistent with the results of Golez (2014), the negative dividend risk premium is somewhat surprising. One possible explanation for this result might be that our dividend risk premium is essentially an ex-post quantity constructed over a short sample period and dividend surprises could contaminate the results. It may be that investors overestimated the future dividends to be paid by S&P 500 firms and were disappointed by subsequent dividend payments for most of our sample. This could be due to the fact that, on aggregate, there has been a shift from dividends to share repurchases (Fama and French, 2001; Grullon and Michaely, 2002). Another possible explanation may be that investors who hold a long position in dividend strips are typically net short dividend risk. If this is the case, then the dividend

[15] Remember that ig is informative about the risk-adjusted growth rate expected over the next 12-month period. Therefore, we need the dividend risk premium expected over the following 12-month.

[16] It is worth highlighting that Tables 1 (Panel B) and 3 of Golez (2014) reveal that the 6-month realized and implied growth rates average around 2.97 % and 7.31 %, respectively. Thus, the author's own figures indicate a negative and economically large annualized dividend risk premium of −4.34 % at the 6-month maturity.

strip could be a good hedging instrument. Consequently, these investors may be willing to pay a premium, i.e. accept a loss on the dividend strip, to hedge their dividend risk. Testing this hypothesis requires very detailed data about the dividend risk exposure of key market participants. Alas, such dataset is not yet available.

III Dividend Growth Predictability

The discussion in Section II.*A*. shows that, if we have a good model for the *drp*, we should be able to improve our dividend growth forecasts. Thus, a natural starting point would be to assess the empirical performance of the 2-factor model for the *drp* (see Equation (7)). If the model does a good job, the expected *drp* should be positively and highly correlated with the subsequently realized *drp*.

Figure 1 displays the dynamics of the realized and expected *drp*. The expected *drp* is the forecast generated by the following equation: $\mathbb{E}_t(drp_{t+12}) = 0.02 - 0.60 ig_t + 0.26 drp_t$. We observe that the two series comove strongly. Our untabulated analysis reveals that a regression of the realized *drp* on a constant and the expected *drp* yields a satisfactory R^2 of 59.96 %.

Additionally, we estimate the following forecasting model:

$$drp_{t+12} = \gamma_0 + \gamma_1 X_t + \epsilon^{drp}_{t+12} \tag{20}$$

We consider three distinct cases. First, we assume $X_t = ig_t$. Second, we assume $X_t = drp_t$. Third, we assume that X is a matrix that contains observations of both *ig* and *drp*. Table 2 presents these results. Throughout this paper, we use a significance level of 5 %. The results of the univariate regression models suggest that each of the two factors contains information about the future *drp*. Furthermore, the explanatory power of the multivariate model (60.42 %) confirms that the 2-factor model does a satisfactory job. We next proceed to analyze its implications for the predictability of dividend growth rates.

III.*A*. In-Sample Analysis

We start with the in-sample analysis. This investigation is motivated by Proposition 1, which posits that the lagged ig^{corr}, a linear combination of the lagged ig and the lagged drp, predicts the dividend growth rate with a positive sign. We test this prediction by regressing the time-series of 1-month dividend growth rates on a constant and the lagged predictive variable X_t:

$$\Delta d_{t+1} = \gamma_0 + \gamma_1 X_t + \epsilon_{t+1}^{\Delta d} \tag{21}$$

where γ_0 and γ_1 are the intercept and slope parameters, respectively. X is the forecasting variable. We first consider the scenario where $X = ig$. Then, we analyze the case $X = ig^{corr}$. By comparing the regression results of the two forecasting models, we are able to shed light on the importance of the lagged drp. Figure 2 displays the dynamics of both forecasting variables.

Table 3 summarizes the regression results. The figures in brackets correspond to the Newey–West corrected test statistics.[17] We test H_0: $\gamma_1 = 0$ against the alternative hypothesis H_1: $\gamma_1 > 0$. The 1-sided t-test is interesting for at least two reasons. From a theoretical point of view, our model predicts a positive relationship between the forecasting variable and next-period's dividend growth rate. For instance, Proposition 1 posits a positive relationship between the lagged ig^{corr} and dividend growth. From a statistical standpoint, Inoue and Kilian (2004) show that 1-sided t-tests substantially improve the power of tests of predictability. Examining the t-statistic, we can see that the null hypothesis is always rejected, suggesting that each of the two variables predicts the dividend growth rate.

The regression results reveal that ig predicts the dividend growth rate with a slope of 0.29. This result is consistent with Golez (2014), whose analysis suggests a

[17] We follow earlier studies, e.g. Rangvid (2006) and Ang and Bekaert (2007), and set the lag length equal to $h+1$, where h denotes the forecasting horizon in months. Since we are forecasting monthly returns, we set the lag length equal to 2. Our results are robust to the choice of the lag length.

slope of 0.19.[18] This slope coefficient has an important interpretation. It reveals the share of variations in ig that is attributable to the dividend growth rate. Exploiting Equation (6), we can show that:

$$\begin{aligned} Var(ig_t) &= Cov(ig_t, ig_t) \\ &= Cov(\mathbb{E}_t(\Delta d_{t+1}) - \mathbb{E}_t(drp_{t+1}), ig_t) \\ Var(ig_t) &= Cov(\mathbb{E}_t(\Delta d_{t+1}), ig_t) - Cov(\mathbb{E}_t(drp_{t+1}), ig_t) \end{aligned}$$

Dividing both sides of the Equation above by Var(ig_t), we obtain:

$$1 = \frac{Cov(\mathbb{E}_t(\Delta d_{t+1}), ig_t)}{Var(ig_t)} - \frac{Cov(\mathbb{E}_t(drp_{t+1}), ig_t)}{Var(ig_t)} \quad (22)$$

The expression above shows that we can decompose the variation in ig into two components related to (i) the expected dividend growth and (ii) the expected drp. The first term to the right of the equality sign is essentially the slope coefficient of a regression of the dividend growth rate on a constant and the lagged implied growth rate.[19] If ig is mainly informative about the dividend growth rate, we would expect to see a very large slope estimate. The second term to the right of the equality sign is the slope estimate of a regression of the drp on a constant and the lagged ig (see Table 2).

Table 3 reveals that only 29 % of variations in ig can be linked to news about expected cashflows. As already discussed, this estimate is broadly similar to that of Golez (2014) who studies the 6-month ig and find a figure of around 19 %. These figures indicate that it is the expected drp, rather than the expected dividend growth rate, that is the main driving force of ig. This conclusion holds irrespective of whether one studies the 12-month ig as we do or the 6-month ig as Golez (2014)

[18] In comparing our results to those of Golez (2014), it is worth keeping in mind that the author regresses the monthly dividend growth rate on implied growth, which is an annualized quantity. Thus, the adapted estimate of the 0.0157 loading on ig at the 1-month horizon shown in Table 4 of Golez (2014) corresponds to $0.0157 \times 12 \approx 0.19$ in our set-up.

[19] As Proposition 1 shows, we can express the dividend growth rate as the sum of the expected dividend growth rate and an independent shock. Assuming that the shock is independent of ig, the slope estimate is the same regardless of whether the dependent variable in the regression model is the realized dividend growth or the expected dividend growth.

does.

If Proposition 1 holds, then we would expect to find that ig^{corr} predicts the next-period dividend growth with a slope of 1. Table 3 reports that ig^{corr} enters the regression model with a positive and statistically significant slope of 0.97. Clearly, the estimated slope is very close to the value of 1 predicted by the theory. Using the t-statistic, we can formally test the hypothesis that the slope equals 1. Our untabulated analysis reveals that the slope estimate is not significantly different from 1, thus supporting the model's prediction. Comparing the two forecasting models, we observe that including the lagged drp lifts the R^2 from 6.69 % (ig) to 11.41 % (ig^{corr}). This result further establishes the relevance of the lagged drp.

The preceding analysis directly imposes the restrictions implied by theory. Recognizing that the two forecasting models discussed above are restricted versions of a more general model, one could estimate the unconstrained model first and then test the restrictions imposed by theory. The unconstrained model is as follows:

$$\Delta d_{t+1} = \gamma_0 + \gamma_1 ig_t + \gamma_2 drp_t + \epsilon_{t+1}^{\Delta d} \tag{23}$$

where γ_0, γ_1 and γ_2 are the parameters to estimate.

Proposition 1 makes several empirically testable predictions regarding the slope of the more general model: $\gamma_1 = 1 + \phi_1$ and $\gamma_2 = \phi_2$. As discussed in the previous section, our estimation results suggest that $\phi_1 = -0.60$ and $\phi_2 = 0.26$. Thus, the theory predicts that $\gamma_1 = 0.40$ and $\gamma_2 = 0.26$. Estimating the regression model above, we obtain $\gamma_1 = 0.38$ and $\gamma_2 = 0.28$. Table 4 summarizes these results. We can see that there is very little to distinguish between the estimated and theoretical sets of coefficients. Using the t-statistic, we formally test each of the two theoretical predictions. If the null hypothesis related to the loading on the variable [name in row] cannot be rejected, we report a checkmark (✓) on the last column of Table 4. We do not reject any of the two hypotheses, lending credence to our model's predictions. We also implement an F-test to jointly test both hypotheses. The (untabulated) F-statistic indicates that we fail to reject the null hypothesis. Overall, this set of results reveals that our model adequately describes the dynamics of monthly dividend growth.

The R^2s of the restricted (Table 3) and unrestricted (Table 4) models provide useful information. If our model provides an accurate description of the data, the R^2 of the restricted model should be similar to that of the unrestricted model. We observe that the unconstrained and constrained models yield very similar R^2s of 11.49 % and 11.41 %, respectively. This indicates that the theoretical restrictions—that the lagged ig and the lagged drp predict dividend growth with coefficients $1+\phi_1$ and ϕ_2, respectively—do little damage to the forecast ability.

III.*B*. Out-of-Sample Evidence

We now explore the predictability of dividend growth in an out-of-sample setting. Similar to Campbell and Thompson (2008), we implement a recursive forecasting scheme. We use the first 6 years of data to estimate ϕ_0, ϕ_1 and ϕ_2 (see Equation (7)). Thus, there are no look-ahead biases. We consider two distinct forecasting models. Model 1 is given by $X_t = \phi_0 + (1+\phi_1)ig_t$.[20] Model 2 uses the insights of Proposition 1 to derive the dividend growth forecast: $X_t = \phi_0 + (1+\phi_1)ig_t + \phi_2 drp_t$. A neat feature of this out-of-sample analysis is that it directly imposes the discipline of the theory and avoids the estimation errors typically associated with dividend growth forecasting regressions. We repeat the steps above for each month (except the last month), expanding the training sample by 1 month each time. We then compute the MSE of each forecasting model:

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(\Delta d_{t+1} - X_t)^2 \qquad (24)$$

[20] One may instead want to assume that ig provides an unbiased forecast of dividend growth, i.e. $X_t = ig_t$, as in Golez (2014). We do not focus on this model because the null that ig predicts dividend growth with a slope of 1 is strongly rejected in the data. As Section III.*A*. of this paper shows, we find a slope (0.29) that is significantly different from 1. This is consistent with the in-sample results of Golez (2014). Thus, the assumption of unbiasedness is an important source of misspecification, which results in poor forecasting performance. Our untabulated analysis reveals that the forecasting model $X_t = \phi_0 + (1+\phi_1)ig_t$, which allows for departures from the unbiasedness restriction, reduces the MSE of the model based on $X_t = ig_t$ by a striking 26.89 %. This sheds light on the extent of the misspecification.

where all variables are as previously defined.

Table 5 summarizes the evidence. We observe that Model 2, which is based on ig^{corr}, performs better than its rival as it yields a lower MSE. The magnitude of the improvement is noteworthy. In relative terms, ig^{corr} reduces the MSE of Model 1 by 7.45 %. We attribute this improvement in forecasting performance directly to the lagged drp.

The entries in Table 5 allow us to compute the $MSE-F$ statistic of McCracken (2007):

$$MSE\text{–}F \quad = \quad N \times \frac{MSE_B - MSE_C}{MSE_C} \qquad (25)$$

where N is the total number of 1-step ahead forecasts. MSE_B is the MSE of the benchmark model, i.e. the restricted model. MSE_C is the MSE of the competing model, i.e. the unrestricted model.

This test statistic enables us to formally test the null that the MSE of the restricted model, i.e. Model 1, is smaller than or equal to that of the unrestricted model, i.e. Model 2. The alternative hypothesis is that the MSE associated with the unrestricted model is lower than that of the restricted model. We find that $MSE - F = 10.30$. Clearly, the large and positive magnitude of the statistic indicates that we can reject the null hypothesis. This highlights the relevance of the lagged drp.

Up to this point, our out-of-sample results are based on an initial training sample of 6 years. One may wonder how robust are the results to the length of the initial training sample? In order to shed light on this question, we consider different initial sample split dates ranging from our initial 6-year period to 10 years. Figure 3 shows by how much Model 2 is able to reduce the MSE of Model 1. We observe that, irrespective of the split date, Model 2 yields more accurate forecasts of dividend growth.

IV Stock Return Predictability

Having established the importance of the lagged *drp* for the predictability of dividend growth, we now explore the implications for return predictability. This analysis is guided by Proposition 2, which makes several empirically testable predictions. We start by examining the predictability of returns in-sample and then turn our attention to the out-of-sample evidence.

IV.*A.* In-Sample Evidence

We regress the time-series of monthly returns on a constant and the lagged forecasting variable X_t:

$$r_{t+1} = \gamma_0 + \gamma_1 X_t + \epsilon^r_{t+1} \tag{26}$$

where γ_0 and γ_1 are the intercept and slope coefficients, respectively. X_t is the return forecasting variable. We examine the following variables in turn: dp, dp^{ig} and dp^{corr}. Comparing the results for the first two forecasting variables sheds light on the importance of accounting for ig. Similarly, by contrasting the results for the last two forecasting variables, we can learn about the relevance of the lagged *drp*. Figure 4 shows the dynamics of all 3 variables. We notice that both dp^{ig} and dp^{corr} are more volatile than the standard dp ratio. It is also worth noticing that these two forecasting variables behave in a manner that is reminiscent of ig. This is mainly due to the high magnitude of δ_1 (0.92), which amplifies any shock to ig (see Proposition 2).

Table 6 reports our regression results. We test H_0: $\gamma_1 = 0$ against the 1-sided alternative hypothesis suggested by theory, i.e. $\gamma_1 > 0$. We reject the null hypothesis for both dp^{ig} and dp^{corr}. This indicates that both variables predict returns. Economically, the slope parameter is informative about the persistence of expected returns, i.e. α_1 (see Equation 12). As Proposition 2 shows, the slope $\gamma_1 = 1 - \bar{\rho}\alpha_1$. Since $\bar{\rho} = 0.98$, the loadings on dp^{ig} and dp^{corr} imply that the persistence of expected returns is close to 0.92 and 0.83, respectively.[21] We also observe that dp^{corr} displays the highest explanatory power (1.91 %), suggesting that the lagged drp helps improve the predictability of returns. We note that the improvement for the (in-sample) predictability of returns is not as striking as in the case of dividend growth.

Intuitively, the return forecasting models discussed above may be viewed as special cases of an unrestricted regression of monthly returns on a constant and the lagged values of dp, ig and drp:

$$r_{t+1} = \gamma_0 + \gamma_1 dp_t + \gamma_2 ig_t + \gamma_3 drp_t + \epsilon^r_{t+1} \tag{27}$$

Proposition 2 yields the following testable hypotheses: $\gamma_1 = 1 - \bar{\rho}\alpha_1$ and $\gamma_2 = (1 - \bar{\rho}\alpha_1)\left[\frac{1+\phi_1}{1-\bar{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)}\right]$ and $\gamma_3 = (1 - \bar{\rho}\alpha_1)\frac{\phi_2}{1-\phi_2}$. Note that $1 - \bar{\rho}\alpha_1$ corresponds to the slope parameter (0.18) of the regression of future returns on a constant and the lagged dp^{corr} ratio presented in Table 6. Recall also that $\phi_1 = -0.60$, $\bar{\rho} = 0.98$, $\delta_1 = 0.92$ and $\phi_2 = 0.26$. Thus, the model yields the following predictions for the slope parameters of Equation (27): $\gamma_1 = 0.18$, $\gamma_2 = 0.48$ and $\gamma_3 = 0.06$.

Table 7 reports the estimated slope parameters. Using the t-test, we separately test each of the 3 hypotheses mentioned above. If the null hypothesis related to the loading on the variable [name in row] cannot be rejected, we report a checkmark (✓) in the last column of Table 7. We observe checkmarks for each slope parameter, indicating that we cannot reject any hypothesis. We also perform a joint hypothesis test, simultaneously imposing all 3 restrictions. We obtain an F-statistic of 1.32, indicating that we fail to reject the null. This suggests that imposing the theoretical restrictions does not materially affect the performance of the unrestricted model.

IV.*B.* Out-of-sample Evidence

We now conduct our analysis out-of-sample. We estimate ϕ_1, ϕ_2, δ_1 and $\bar{\rho}$ recursively. We use the first 6 years of data as our initial training sample period. We exploit all

[21] To get α_1, we look at the slope coefficient of the return forecasting regression. Since theory predicts that the slope equals $1 - \bar{\rho}\alpha_1$, we rearrange the expression to recover α_1.

the information in our training sample to estimate the return forecasting regression shown in Equation (26). Equipped with the intercept and slope estimates, we use the last observation of the forecasting variable (in the training sample) to predict the next-period return. We repeat these steps for each month and for each of the two forecasting variables: dp^{ig} and dp^{corr}.

Table 8 reports the MSE (expressed in basis points) of each return forecasting model. We find that augmenting the dp ratio with ig helps reduce the MSE of the standard dp ratio by 1.17 %. Thus, accounting for implied growth improves the forecasting performance of the standard dp ratio. This finding echoes the result of Golez (2014). Analyzing the results for dp^{corr}, we find that it yields the lowest MSE of all three forecasting variables. Indeed, it reduces the MSE of the standard dp ratio by 2.88 %. We test whether the difference between the MSE of the dp^{corr} model and that of the dp^{ig} forecasting model is significant. Our (untabulated) calculation yields an $MSE - F$ statistic equal to 2.21, which is larger than the corresponding critical value. This result is consistent with our model's prediction.

We also consider alternative initial sample split dates. The blue line of Figure 5 tells us by how much, in %, an agent who uses a forecasting model based on dp^{ig} would be able to reduce the MSE of the model based on the standard dp ratio for different sample split dates. There is a clear improvement regardless of the length of the initial sample split date. The red line relates to a similar analysis with the difference that it focuses on dp^{corr} rather than dp^{ig}. We can see that the dp^{corr} ratio delivers the highest improvement in forecast accuracy. This is true for all sample split dates considered. The upshot of this analysis is that our results are not driven by the choice of the initial sample split date.

IV.*C*. The Economic Value of Return Predictability

We explore the implications of the evidence of return predictability for the portfolio choice of an investor willing to use the dp^{corr} ratio as a timing signal for a quantitative strategy. In particular, the market timing strategy allocates a fraction of wealth w_t to the risky stock and the remainder to the riskless asset. The risky asset has expected return μ_t and expected volatility $\hat{\sigma}_t$. The riskless asset yields a return rf_t. We assume that the investor has a quadratic utility function, thus giving rise to the following optimization problem:[22]

$$\max_{w_t} w_t \mu_t + (1 - w_t) rf_t - \frac{\gamma}{2} w_t^2 \hat{\sigma}_t^2 \tag{28}$$

The optimal allocation to the risky asset is given by:

$$w_t = \frac{\mu_t - rf_t}{\gamma \hat{\sigma}_t^2} \tag{29}$$

For each return forecasting model, we compute the expected return on the risky asset and determine the allocation to the risky and riskless assets, respectively. In computing the portfolio weights, we use the 1-month LIBOR rate as our proxy for the riskless rate.[23] Following Campbell and Thompson (2008), we use the previous 5 years of monthly returns data to estimate the variance of the stock returns. We also impose the restriction that the weight has to be positive and not greater than 1.5 (Campbell and Thompson, 2008). Finally, we consider different values for the coefficient of risk aversion, e.g. 4, 6, 8 and 10. Equipped with the portfolio weights and the time series of realized stock returns, we finally compute the time series of realized portfolio returns.

We analyze the certainty equivalent rate (CE) of return, which is the risk-free rate of return that the investor is willing to accept rather than following a risky

[22] The optimization problem of an investor with quadratic utility is equivalent to maximizing a linear combination of mean and variance. This is true irrespective of the distribution of asset returns. We refer the interested reader to Campbell and Viceira (2002) for an excellent treatment of this topic.

[23] As previously discussed, it is standard in the derivatives pricing community to proxy the riskless rate with the LIBOR rate. Consistent with this practice, and thus the earlier part of our study, we use the 1-month LIBOR rate as the risk-free rate proxy. Because the return predictability literature also analyzes the 3-month Treasury bill rate, e.g. Goyal and Welch (2003), one may wonder what impact, if any, does the proxy for the riskless rate have on our portfolio results. To investigate this, we obtain the time series of 3-month T-bill from the website of the Federal Reserve of St. Louis and repeat our analysis. Unreported tabulations show that the riskless rate proxy has very little bearing on our core results.

market timing strategy:

$$CE = \bar{r}_p - \frac{\gamma}{2}\sigma_p^2 \quad (30)$$

where $\bar{r}_p$ is the average of the realized portfolio returns. σ_p is the realized volatility of the portfolio returns. All other variables are as previously defined.

Several results emerge from Table 9. We observe that an investor who uses the dp^{corr} ratio achieves the highest certainty equivalent. How much would an investor pay in order to switch from a quantitative strategy that is based on the standard dp ratio to a timing strategy that relies on the novel dp^{corr} ratio? Our results indicate that an investor with a risk aversion coefficient equal to 4 would pay up to 1.56 % per year. This fee speaks directly to the importance of accounting for (i) the implied growth rate and (ii) the dividend risk premium. In order to understand the contribution of each component to this result, we also examine the timing strategy based on the dp^{ig} ratio. Computing the difference between the certainty equivalent rate of return of the timing strategy based on the dp^{ig} ratio and that of the strategy based on the standard dp ratio, we find that, for that same investor, the dp^{ig} ratio leads to a utility gain of 0.13 % per year. This result reveals that accounting for the drp further elevates the utility gain from 0.13 % to 1.56 %.

In summary, our evidence suggests that an investor who uses the dp^{corr} ratio instead of the standard dp ratio substantially improves the out-of-sample performance of her portfolio. Dissecting the empirical evidence, we find that a sizable proportion of this improvement is related to the correction for the drp.

V Conclusion

We show that the dividend growth rate implied by the futures market contains information about (i) the expected dividend growth rate and (ii) the expected drp. We propose a simple model for the drp and study its implications for the predictability of dividend growth and aggregate returns.

Our empirical analysis establishes that accounting for the expected drp

strengthens the predictability of dividend growth and returns. Our main results hold both in- and out-of-sample. Analyzing the implication of our results for the portfolio choice of an investor, we find that a market timing investor who accounts for the time varying dividend risk premium realizes an additional utility gain of 1.43 % per year. Overall, our study highlights, both theoretically and empirically, the importance of the dividend risk premium for the predictability of dividend growth and returns.

References

Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *Review of Financial Studies*, 20(3):651–707.

Bilson, J., Kang, S. B., and Luo, H. (2015). The term structure of implied dividend yields and expected returns. *Economics Letters*, 128:9–13.

Boguth, O., Carlson, M., Fisher, A. J., and Simutin, M. (2012). Leverage and the limits of arbitrage pricing: Implications for dividend strips and the term structure of equity risk premia. *Working Paper*.

Borovicka, J., Hansen, L. P., and Scheinkman, J. A. (2015). Misspecified recovery. *Forthcoming Journal of Finance*.

Campbell, J. Y. (2008). Viewpoint: Estimating the equity premium. *Canadian Journal of Economics*, 41(1):1–21.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, 21(4):1509–1531.

Campbell, J. Y. and Viceira, L. M. (2002). *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*. Oxford University Press.

Chernov, M. (2007). On the role of risk premia in volatility forecasting. *Journal of Business and Economic Statistics*, 25(4):411–426.

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance*, 66(4):1047–1108.

Cox, J. C. and Ross, S. A. (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3(1):145–166.

Fama, E. F. and French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1):3–25.

Fama, E. F. and French, K. R. (2001). Disappearing dividends: changing firm characteristics or lower propensity to pay? *Journal of Financial Economics*, 60(1):3–43.

Garrett, I. and Priestley, R. (2000). Dividend behaviour and dividend signaling. *Journal of Financial and Quantitative Analysis*, 35(02):173–189.

Golez, B. (2014). Expected returns and dividend growth rates implied by derivative markets. *Review of Financial Studies*, 27(3):790–822.

Goyal, A. and Welch, I. (2003). Predicting the equity premium with dividend ratios. *Management Science*, 49(5):639–654.

Grullon, G. and Michaely, R. (2002). Dividends, share repurchases, and the substitution hypothesis. *Journal of Finance*, 57(4):1649–1684.

Harrison, J. M. and Pliska, S. R. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11(3):215–260.

Harvey, C. R. and Whaley, R. E. (1992). Market volatility prediction and the efficiency of the S&P 100 index option market. *Journal of Financial Economics*, 31(1):43–73.

Inoue, A. and Kilian, L. (2004). In-sample or out-of-sample tests of predictability: Which one should we use? *Econometric Reviews*, 23(4):371–402.

Lacerda, F. and Santa-Clara, P. (2010). Forecasting dividend growth to better predict returns. *Chicago Booth Working Paper*.

Lintner, J. (1956). Distribution of incomes of corporations among dividends, retained earnings, and taxes. *American Economic Review*, 46(2):97–113.

Marsh, T. A. and Merton, R. C. (1987). Dividend behavior for the aggregate stock market. *Journal of Business*, 60(1):1–40.

McCracken, M. W. (2007). Asymptotics for out of sample tests of granger causality. *Journal of Econometrics*, 140(2):719–752.

Newey, W. K. and West, K. D. (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Piazzesi, M. and Swanson, E. T. (2008). Futures prices as risk-adjusted forecasts of monetary policy. *Journal of Monetary Economics*, 55(4):677–691.

Prokopczuk, M. and Wese Simen, C. (2014). The importance of the volatility risk premium for volatility forecasting. *Journal of Banking & Finance*, 40:303–320.

Rangvid, J. (2006). Output and expected returns. *Journal of Financial Economics*, 81(3):595–624.

Ross, S. (2015). The recovery theorem. *Journal of Finance*, 70(2):615–648.

van Binsbergen, J., Brandt, M., and Koijen, R. (2012). On the timing and pricing of dividends. *American Economic Review*, 102(4):1596–1618.

van Binsbergen, J. and Koijen, R. S. (2010). Predictive regressions: A present-value approach. *Journal of Finance*, 65(4):1439–1471.

Zhong, J. (2016). Predictive regressions based on ex ante index futures market information. *Working Paper*.

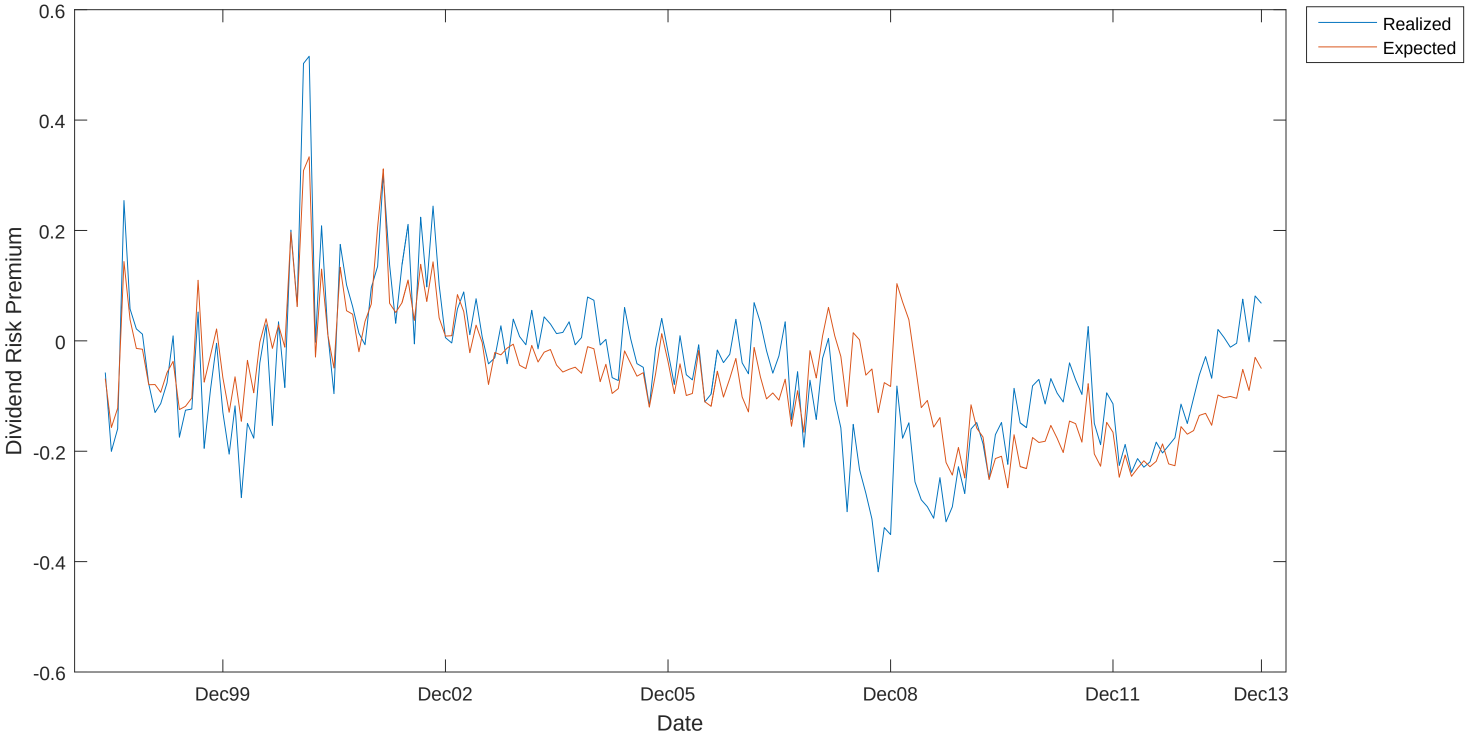

Figure 1: **Realized v.s. Expected** drp

This figure shows the dynamics of the realized and expected drp during our sample period. The expected drp is the forecast generated by the following equation: $\mathbb{E}_t(drp_{t+12}) = 0.02 - 0.60ig_t + 0.26drp_t$. *For ease of exposition, we align the realized and expected drp. The horizontal axis displays the observation date of the realized drp. The vertical axis shows the magnitude of the risk premia. All figures are annualized.*

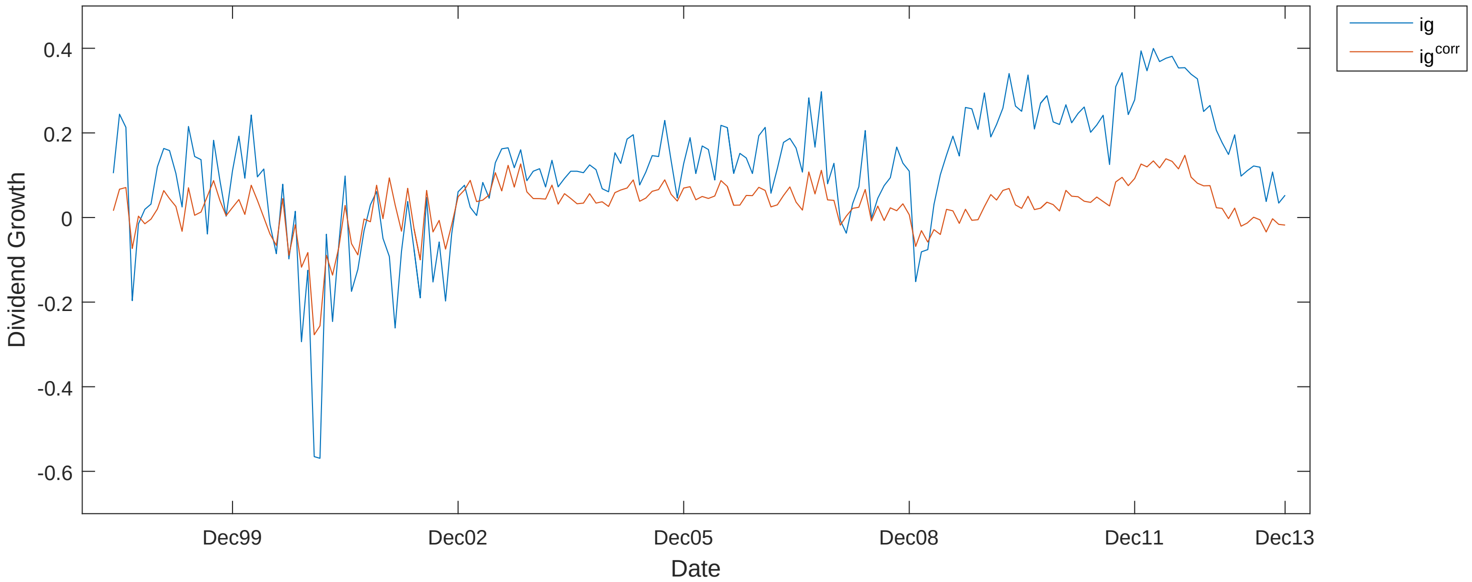

Figure 2: **The Dynamics of ig and ig^{corr}**

This figure plots the time-series dynamics of annualized ig and ig^{corr}, where $ig_t^{corr} = \phi_0 + (1 + \phi_1)ig_t + \phi_2 drp_t$. In the data, we find that $\phi_0 = 0.02$, $\phi_1 = -0.60$ and $\phi_2 = 0.26$. Armed with these parameters, we can construct ig^{corr}. The horizontal axis shows the observation date. The vertical axis indicates the (annualized) implied dividend growth rate.

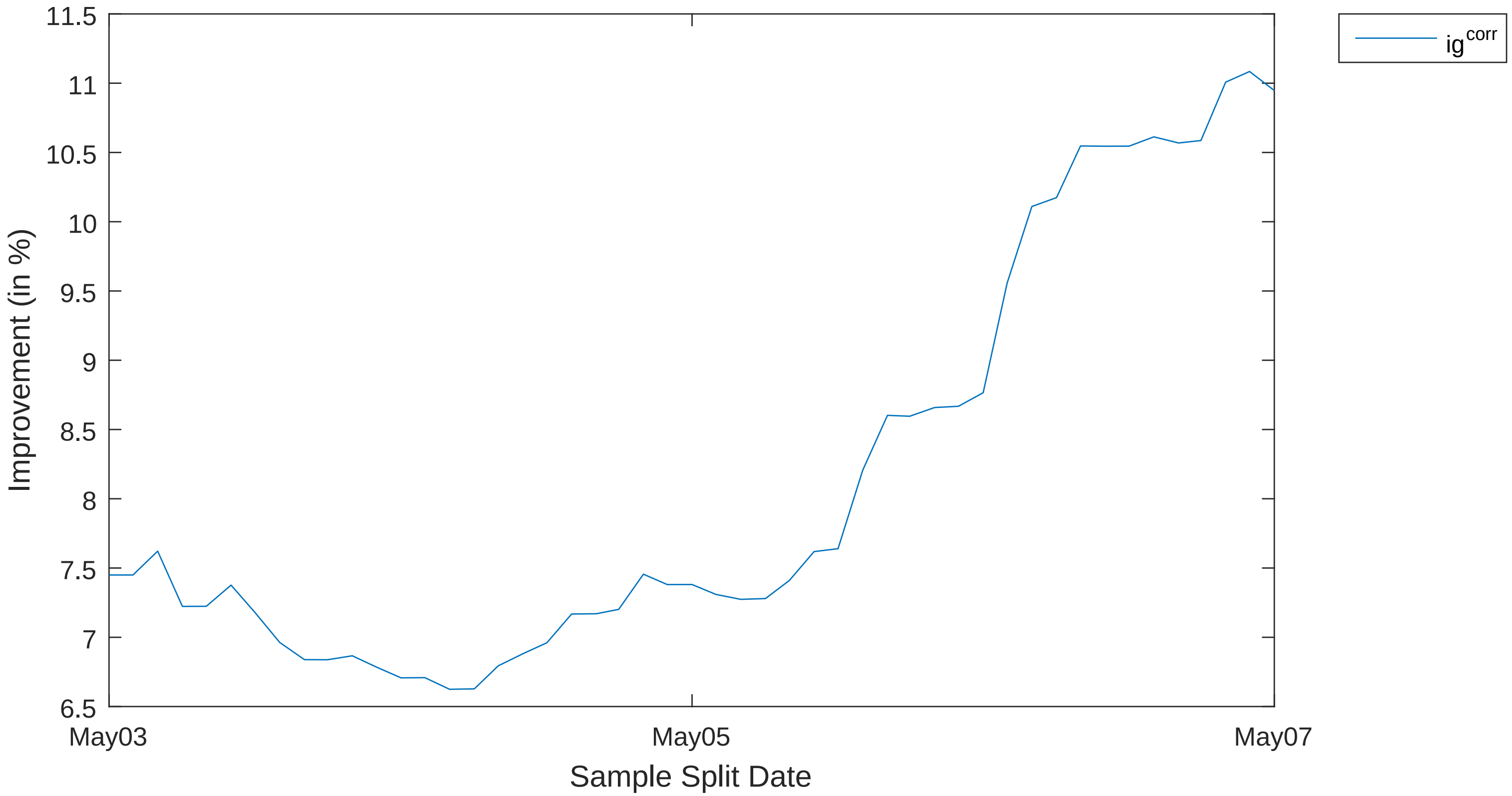

Figure 3: **Improvement in MSE by Sample Split Date**

This figure sheds light on the importance of the lagged drp. The vertical axis tells us by how much, in %, Model 2 reduces the MSE of Model 1 for different initial sample split dates. The forecast generated by Model 1 is given by $X_t = \phi_0 + (1 + \phi_1)ig_t$. The forecast generated by Model 2 takes into account the drp: $X_t = \phi_0 + (1 + \phi_1)ig_t + \phi_2 drp_t$. The parameters ϕ_0, ϕ_1 and ϕ_2 are estimated recursively for each training sample.

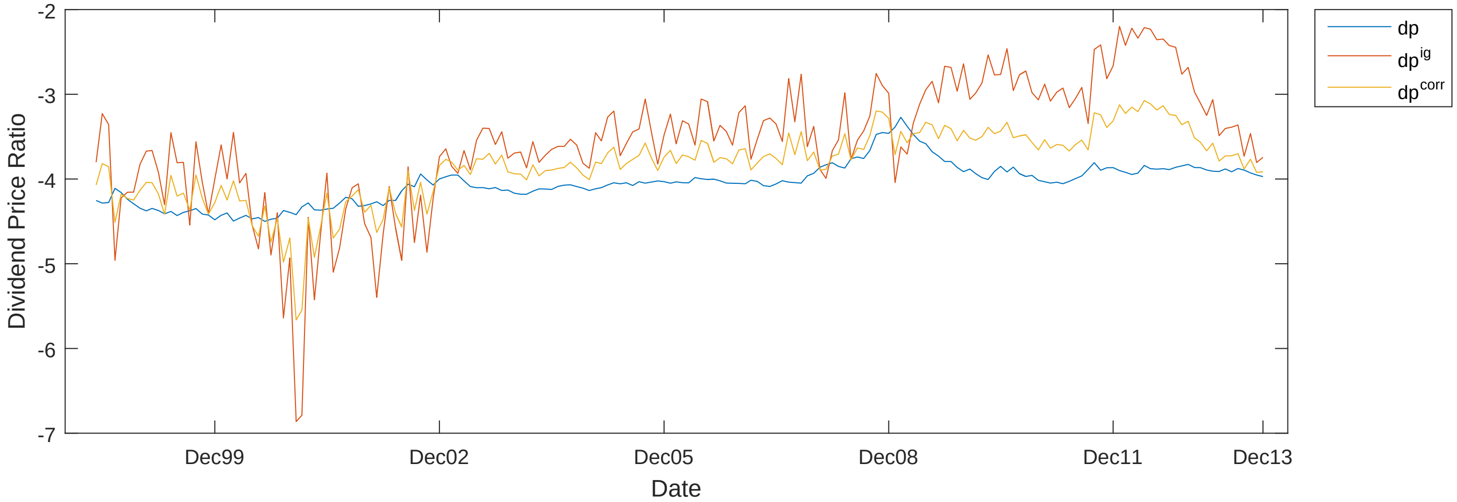

Figure 4: **The Dynamics of** dp**,** dp^{ig} **and** dp^{corr}

This figure plots the time-series dynamics of dp*,* dp^{ig} *and* dp^{corr}*.* dp *is the logarithm of the trailing sum of 12-month dividends over the stock index price.* $dp^{ig} = dp_t + \frac{(1+\phi_1)ig_t}{1-\bar{\rho}\delta_1}$ *and* $dp^{corr} = dp_t + \frac{(1+\phi_1)ig_t}{1-\bar{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2 ig_t}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)} + \frac{\phi_2 drp_t}{1-\bar{\rho}\phi_2}$*. In the data, we find that* $\phi_1 = -0.60$*,* $\bar{\rho} = 0.98$*,* $\delta_1 = 0.92$ *and* $\phi_2 = 0.26$*. With these parameter values, we can construct the relevant time-series. The horizontal axis shows the observation date. The vertical axis shows the magnitude of the ratios.*

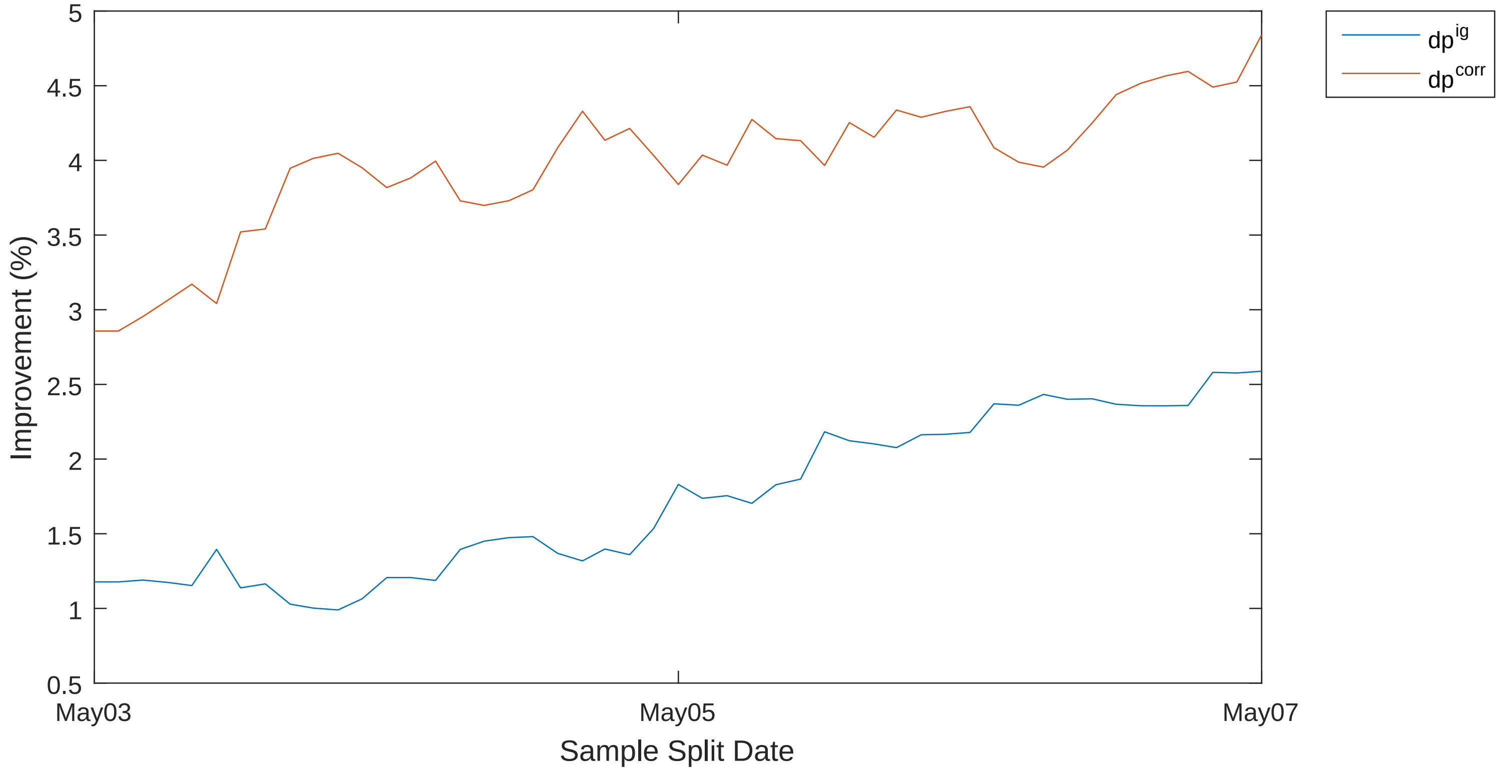

Figure 5: **Improvement in MSE by Sample Split Date**

This figure shows the reduction (in relative terms) of the MSE of the forecasting model based on the dp ratio achieved when the forecaster relies on (i) the dp^{ig} (blue line) and (ii) the dp^{corr} (red line) ratios for different initial sample split dates. dp is the logarithm of the trailing sum of 12-month dividends over the stock index price. $dp^{ig} = dp_t + \frac{(1+\phi_1)ig_t}{1-\bar{\rho}\delta_1}$ *and* $dp^{corr} = dp_t + \frac{(1+\phi_1)ig_t}{1-\bar{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2 ig_t}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)} + \frac{\phi_2 drp_t}{1-\bar{\rho}\phi_2}$. *For each training sample, we recursively estimate the parameters* ϕ_1, $\bar{\rho}$, δ_1 *and* ϕ_2. *We use these parameters to compute the relevant forecasting variables dp,* dp^{ig} *and* dp^{corr}. *We estimate a return forecasting regression using all information from the training sample. We then use the estimated parameters together with the most recent observation of the forecasting variable to generate the forecast for the next-period.*

Table 1: **Summary Statistics**

This table reports the summary statistics of several time-series. Δd denotes the time-series of (annualized) monthly dividend growth. r denotes the time-series of (annualized) monthly S&P 500 returns. ig relates to the implied growth rate. drp refers to the dividend risk premium. dp is the standard dividend price ratio. ig^{corr} is the dividend risk premium corrected implied growth rate. dp^{ig} relates to the growth adjusted dividend price ratio. dp^{corr} denotes the corrected dividend price ratio. The column entitled "Mean" reports the average of the time-series [name in row]. Similarly, "Std", "Skew" and "Kurt" relate to the standard deviation, skewness and kurtosis of the series [name in row]. AR(1) reports the first order autocorrelation. Finally, "Nobs" shows the number of observations.

	Mean	Std	Skew	Kurt	AR(1)	Nobs
Δd	0.06	0.16	-0.56	5.21	0.19	188
r	0.07	0.55	-0.84	4.45	0.09	188
ig	0.11	0.15	-1.22	6.83	0.72	188
drp	-0.06	0.14	0.58	5.24	0.69	188
dp	-4.05	0.23	0.36	3.83	0.98	188
ig^{corr}	0.03	0.06	-1.57	8.60	0.66	188
dp^{ig}	-3.57	0.73	-1.08	6.06	0.80	188
dp^{corr}	-3.84	0.43	-1.04	5.10	0.86	188

Table 2: **The Predictability of the Dividend Risk Premium**

This table summarizes the results of the predictability of the dividend risk premium. We first regress the time-series of the drp on a constant and the 12-period lagged ig. Next, we regress the time-series of the drp on a constant and the 12-period lagged drp. Finally, we regress the time-series of the drp on a constant and the 12-period lagged ig and drp. Although all regressions are estimated with an intercept, we report the slope estimates only. The entries in brackets indicate the Newey–West (1987) adjusted t-statistics computed with 13 lags.

ig	-0.70		-0.65
	(-11.04)		(-7.74)
drp		0.41	0.19
		(3.93)	(1.96)
R^2	57.10%	16.75%	60.42%

Table 3: **The (In-Sample) Predictability of Dividend Growth**

This table summarizes the results of the predictability of monthly dividend growth. We regress the time-series of monthly dividend growth on a constant and the lagged predictive variable. We consider two distinct predictive variables. The first one, ig, is the implied dividend growth rate. The second predictor, ig^{corr}*, is the dividend risk premium corrected implied growth rate:* $ig_t^{corr} = \phi_0 + (1+\phi_1)ig_t + \phi_2 drp_t$*. In the data, we find that* $\phi_0 = 0.02$*,* $\phi_1 = -0.60$ *and* $\phi_2 = 0.26$*. Armed with these parameters, we can construct* ig^{corr}*. Although all regressions are estimated with an intercept, we report the slope estimates only. The entries in brackets indicate the Newey–West (1987) adjusted t-statistics computed with 2 lags.*

ig	0.29 (4.30)	
ig^{corr}		0.97 (5.05)
R^2	6.69%	11.41%

Table 4: **Unconstrained Dividend Growth Forecasting Regression**

This table shows the results of the unconstrained regression of monthly dividend growth on a constant, the lagged ig, and the lagged drp. We report the point estimates of the regression model. We also show in square brackets the values predicted by our theory. The model predicts that $\gamma_1 = 1 + \phi_1$ *and* $\gamma_2 = \phi_2$*. In the data, we find that* $\phi_1 = -0.60$ *and* $\phi_2 = 0.26$*. Thus, the theoretical values of the slope are* $\gamma_1 = 0.40$ *and* $\gamma_2 = 0.26$*. We conduct a t-test to test the null that the estimated parameter [name in row] equals its theoretical value. Entries marked "✓" indicate that we cannot reject the null hypothesis.*

	Parameters	**Hypothesis Test**
ig	0.38 [0.40]	✓
drp	0.28 [0.26]	✓
R^2	11.49%	

Table 5: **The (Out-of-Sample) Predictability of Dividend Growth**

This table presents out-of-sample evidence on the predictability of monthly dividend growth. We consider two forecasting models. Model 1 derives the forecast as follows: $X_t = \phi_0 + (1 + \phi_1)ig_t$. *Model 2 derives the forecast as:* $X_t = \phi_0 + (1 + \phi_1)ig_t + \phi_2 drp_t$. *This forecast corresponds exactly to* ig_t^{corr}. *We use a recursive window to estimate the parameters* ϕ_0, ϕ_1 *and* ϕ_2. *We report the mean squared error* (MSE) *of each model in basis points.*

Model 1	**Model 2**
310.17	287.07

Table 6: **The (In-Sample) Predictability of Returns**

This table summarizes the results of the predictability of monthly returns. We regress the time-series of monthly returns on a constant and the lagged predictive variable. We consider three distinct predictive variables. The first one, dp, is the standard dividend price ratio. The second predictor, dp^{ig}, *is the implied growth augmented dividend price ratio:* $dp^{ig} = dp_t + \frac{(1+\phi_1)ig_t}{1-\bar{\rho}\delta_1}$. *The third predictor,* dp^{corr}, *is the corrected dividend price ratio:* $dp^{corr} = dp_t + \frac{(1+\phi_1)ig_t}{1-\bar{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2 ig_t}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)} + \frac{\phi_2 drp_t}{1-\bar{\rho}\phi_2}$. *Using the following information,* $\phi_1 = -0.60$, $\bar{\rho} = 0.98$, $\delta_1 = 0.92$ *and* $\phi_2 = 0.26$, *we compute the relevant forecasting variables. Although all regressions are estimated with an intercept, we report the slope estimates only. The entries in brackets indicate the Newey–West (1987) adjusted t-statistics computed with 2 lags.*

dp	0.19		
	(0.74)		
dp^{ig}		0.10	
		(1.72)	
dp^{corr}			0.18
			(1.67)
R^2	0.67%	1.86%	1.91%

Table 7: **Unconstrained Return Forecasting Regression**

This table summarizes the results of the predictability of monthly returns. We regress the time-series of monthly returns on a constant and the lagged dp, the lagged ig and the lagged drp.

$$r_{t+1} = \gamma_0 + \gamma_1 dp_t + \gamma_2 ig_t + \gamma_3 drp_t + \epsilon^r_{t+1}$$

The present value model yields the following testable hypotheses: $\gamma_1 = 1 - \bar{\rho}\alpha_1$ *and* $\gamma_2 = (1 - \bar{\rho}\alpha_1)\left[\frac{1+\phi_1}{1-\bar{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)}\right]$ *and* $\gamma_3 = (1-\bar{\rho}\alpha_1)\frac{\phi_2}{1-\phi_2}$*. It is worth noticing that* $1-\bar{\rho}\alpha_1$ *corresponds to the slope parameter of the regression of future returns on a constant and the lagged* dp^{corr} *ratio presented in Table 6. Recall also that* $\phi_1 = -0.60$, $\bar{\rho} = 0.98$, $\delta_1 = 0.92$ *and* $\phi_2 = 0.26$*. Thus, the model yields the following predictions for the slope parameters:* $\gamma_1 = 0.18$, $\gamma_2 = 0.48$ *and* $\gamma_3 = 0.06$*. We show in square brackets the values predicted by the theory. Although all regressions are estimated with an intercept, we report the slope estimates only. Entries marked "✓" indicate that we cannot reject the null hypothesis that the estimated slope parameter associated with the variable [name in row] is equal to its theoretical value presented in square brackets.*

	Parameters	**Hypothesis Test**
dp	0.13 [0.18]	✓
ig	0.48 [0.37]	✓
drp	0.21 [0.06]	✓
R^2	2.10%	

Table 8: **The (Out-of-Sample) Predictability of Returns**

This table summarizes the evidence of the predictability of returns out-of-sample. We consider the dp, the dp^{ig} *and the* dp^{corr} *ratios, in turn. The last two forecasting variables are computed using the following formulas:* $dp^{ig} = dp_t + \frac{(1+\phi_1)ig_t}{1-\bar{\rho}\delta_1}$ *and* $dp^{corr} = dp_t + \frac{(1+\phi_1)ig_t}{1-\bar{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2 ig_t}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)} + \frac{\phi_2 drp_t}{1-\bar{\rho}\phi_2}$*. For each training sample, we recursively estimate the parameters* ϕ_1, $\bar{\rho}$, δ_1 *and* ϕ_2*. We use these parameters to compute the relevant forecasting variables. We estimate a return forecasting regression using all information from the training sample. We then use the estimated parameters together with the most recent observation of the forecasting variable to generate the forecast for the next-period, which we subsequently compare to the realized return. For each of the three models, we compute and report the mean squared error (MSE). These values are expressed in basis points.*

dp	dp^{ig}	dp^{corr}
18.74	18.52	18.20

Table 9: **The Economic Value of Return Predictability**

This table presents results of the out-of-sample portfolio performance of an investor who attempts to exploit the predictability of returns by devising market timing strategies. We assume that the investor has a quadratic utility function with a coefficient of relative risk aversion equal to γ. *The first column shows the different values of* γ, *i.e.* $\gamma = 4, 6, 8$ *or* 10. *At the end of each month, we compute the optimal allocation of the investor to the risky stock and the riskless asset. These weights depend on the forecasting model for expected returns. The investor considers three distinct forecasting variables:* dp, dp^{ig} *and* dp^{corr}. *Given these weights, we compute the realized return on the portfolio. We do this for each calendar month and return forecasting variable. We then compute and report the annualized certainty equivalent (CE) of the strategy based on the predictive variable [name in column].*

γ	dp	dp^{ig}	dp^{corr}
4	0.25%	0.38%	1.81%
6	0.79%	0.87%	1.83%
8	1.06%	1.12%	1.84%
10	1.22%	1.27%	1.85%

Appendix

A Proofs

This appendix presents the detailed proof of the propositions presented in the main text. In order to facilitate the exposition of the derivations, it is useful to re-state our main assumptions:

$$r_{t+1} = \mu_t + \epsilon^r_{t+1} \tag{A.1}$$

$$\mu_{t+1} = \alpha_0 + \alpha_1 \mu_t + \epsilon^\mu_{t+1} \tag{A.2}$$

$$ig_{t+1} = \delta_0 + \delta_1 ig_t + \epsilon^{ig}_{t+1} \tag{A.3}$$

$$drp_{t+1} = \phi_0 + \phi_1 ig_t + \phi_2 drp_t + \epsilon^{drp}_{t+1} \tag{A.4}$$

where all error terms are $i.i.d$ with zero mean.

A.1 Proposition 1

To derive the first proposition of our model, we start from the accounting identity linking together the expected dividend growth rate, the expected drp and the implied growth rate:

$$\mathbb{E}_t(\Delta d_{t+1} - drp_{t+1}) = ig_t$$

This implies that

$$\begin{aligned} \mathbb{E}_t(\Delta d_{t+1}) &= \mathbb{E}_t(drp_{t+1}) + ig_t \\ &= \mathbb{E}_t(\phi_0 + \phi_1 ig_t + \phi_2 drp_t + \epsilon^{drp}_{t+1}) + ig_t \\ \mathbb{E}_t(\Delta d_{t+1}) &= \phi_0 + (1 + \phi_1) ig_t + \phi_2 drp_t \end{aligned} \tag{A.5}$$

Recall that the realized dividend growth can be decomposed into an expected component and a shock:

$$\Delta d_{t+1} = \mathbb{E}_t(\Delta d_{t+1}) + \epsilon_{t+1}^{\Delta d} \tag{A.6}$$

$$\Delta d_{t+1} = \phi_0 + (1+\phi_1) ig_t + \phi_2 drp_t + \epsilon_{t+1}^{\Delta d} \tag{A.7}$$

This completes the proof of Proposition 1. ■

A.2 Proposition 2:

For ease of exposition, let us restate Equation (10):

$$\sum_{j=0}^{+\infty} \overline{\rho}^j \mathbb{E}_t(r_{t+1+j}) - \overline{\rho}^j \mathbb{E}_t(\Delta d_{t+1+j}) = \frac{k}{1-\overline{\rho}} + dp_t \tag{A.8}$$

Using Equations (A.1) and (A.2), we can compute the first summation term on the left-hand side of Equation (A.8):

$$\begin{aligned} \sum_{j=0}^{+\infty} \overline{\rho}^j \mathbb{E}_t(r_{t+1+j}) &\equiv k_r + \sum_{j=0}^{+\infty} \overline{\rho}^j \alpha_1^j \mu_t \\ \sum_{j=0}^{+\infty} \overline{\rho}^j \mathbb{E}_t(r_{t+1+j}) &\equiv k_r + \frac{\mu_t}{1-\overline{\rho}\alpha_1} \end{aligned} \tag{A.9}$$

where k_r is a constant that depends on α_0 and α_1.

Similarly, we combine the result of Proposition 1 together with Equations (A.3) and (A.4) to compute the infinite sum of expected dividend growth rates:

$$\begin{aligned} \sum_{j=0}^{+\infty} \overline{\rho}^j \mathbb{E}_t(\Delta d_{t+1+j}) &= k_{\Delta d} + \sum_{j=0}^{+\infty} \overline{\rho}^j \left[\delta_1^j (1+\phi_1) + \phi_1 \phi_2 \frac{\delta_1^j - \phi_2^j}{\delta_1 - \phi_2} \right] ig_t + \sum_{j=0}^{+\infty} \overline{\rho}^j \phi_2^{j+1} drp_t \\ \sum_{j=0}^{+\infty} \overline{\rho}^j \mathbb{E}_t(\Delta d_{t+1+j}) &\equiv k_{\Delta d} + \left[\frac{1+\phi_1}{1-\overline{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)} \right] ig_t + \frac{\phi_2 drp_t}{1-\overline{\rho}\phi_2} \end{aligned} \tag{A.10}$$

where $k_{\Delta d}$ is a constant that depends on δ_0, δ_1, ϕ_0, ϕ_1 and ϕ_2.

Substituting Equations (A.9) and (A.10) into Equation (A.8) yields:

$$
\begin{aligned}
dp_t &= -\frac{k}{1-\overline{\rho}} + \sum_{j=0}^{+\infty} \overline{\rho}^j \left(\mathbb{E}_t(r_{t+1+j}) - \mathbb{E}_t(\Delta d_{t+1+j}) \right) \\
&= \underbrace{-\frac{k}{1-\overline{\rho}} + k_r - k_{\Delta d}}_{k_1} + \frac{\mu_t}{1-\overline{\rho}\alpha_1} - \left[\frac{1+\phi_1}{1-\overline{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)} \right] ig_t - \frac{\phi_2 drp_t}{1-\overline{\rho}\phi_2} \\
dp_t &\equiv k_1 + \frac{\mu_t}{1-\overline{\rho}\alpha_1} - \left[\frac{1+\phi_1}{1-\overline{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)} \right] ig_t - \frac{\phi_2}{1-\overline{\rho}\phi_2} drp_t \qquad \text{(A.11)}
\end{aligned}
$$

Similarly, we can express the next-period dividend price ratio as:

$$
dp_{t+1} = k_1 + \frac{\mu_{t+1}}{1-\overline{\rho}\alpha_1} - \left[\frac{1+\phi_1}{1-\overline{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)} \right] ig_{t+1} - \frac{\phi_2}{1-\overline{\rho}\phi_2} drp_{t+1}
$$

Using Equations (A.3) and (A.4), we can show that:

$$
\begin{aligned}
dp_{t+1} &= k_1 + \frac{\mu_{t+1}}{1-\overline{\rho}\alpha_1} - \left[\frac{1+\phi_1}{1-\overline{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)} \right] ig_{t+1} - \frac{\phi_2}{1-\overline{\rho}\phi_2} drp_{t+1} \\
&= k_1 + \frac{\alpha_0 + \alpha_1\mu_t + \epsilon^{\mu}_{t+1}}{1-\overline{\rho}\alpha_1} - \left[\frac{1+\phi_1}{1-\overline{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)} \right] (\delta_0 + \delta_1 ig_t + \epsilon^{ig}_{t+1}) \\
&\quad - \frac{\phi_2}{1-\overline{\rho}\phi_2} (\phi_0 + \phi_1 ig_t + \phi_2 drp_t + \epsilon^{drp}_{t+1}) \\
&= k_1 + \underbrace{\frac{\alpha_0}{1-\overline{\rho}\alpha_1} - \left[\frac{1+\phi_1}{1-\overline{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)} \right] \delta_0 - \frac{\phi_0\phi_2}{1-\overline{\rho}\phi_2}}_{k_2} + \frac{\alpha_1\mu_t + \epsilon^{\mu}_{t+1}}{1-\overline{\rho}\alpha_1} \\
&\quad - \left[\frac{1+\phi_1}{1-\overline{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)} \right] (\delta_1 ig_t + \epsilon^{ig}_{t+1}) - \frac{\phi_1\phi_2 ig_t}{1-\overline{\rho}\phi_2} - \frac{\phi_2^2 drp_t + \phi_2\epsilon^{drp}_{t+1}}{1-\overline{\rho}\phi_2} \\
&\equiv k_1 + k_2 + \frac{\alpha_1\mu_t}{1-\overline{\rho}\alpha_1} - \left[\frac{1+\phi_1}{1-\overline{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)} \right] \delta_1 ig_t - \frac{\phi_1\phi_2 ig_t}{1-\overline{\rho}\phi_2} \\
&\quad - \frac{\phi_2^2 drp_t}{1-\overline{\rho}\phi_2} + \underbrace{\frac{\epsilon^{\mu}_{t+1}}{1-\overline{\rho}\alpha_1} - \left[\frac{1+\phi_1}{1-\overline{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)} \right] \epsilon^{ig}_t - \frac{\phi_2\epsilon^{drp}_{t+1}}{1-\overline{\rho}\phi_2}}_{\epsilon^{dp}_{t+1}} \\
&\equiv \alpha_1 k_1 + \frac{\alpha_1\mu_t}{1-\overline{\rho}\alpha_1} - \left[\frac{1+\phi_1}{1-\overline{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)} \right] (\alpha_1 + \delta_1 - \alpha_1) ig_t \\
&\quad - \frac{\phi_1\phi_2 ig_t}{1-\overline{\rho}\phi_2} - \frac{\alpha_1 + \phi_2 - \alpha_1}{1-\overline{\rho}\phi_2} \phi_2 drp_t + k_2 + (1-\alpha_1)k_1 + \epsilon^{dp}_{t+1}
\end{aligned}
$$

$$
\begin{aligned}
dp_{t+1} &= k_2 + (1-\alpha_1)k_1 + \alpha_1 dp_t - \left[\frac{1+\phi_1}{1-\overline{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)}\right](\delta_1-\alpha_1)ig_t \\
&\quad -\frac{\phi_1\phi_2 ig_t}{1-\overline{\rho}\phi_2} - \frac{\phi_2-\alpha_1}{1-\overline{\rho}\phi_2}\phi_2 drp_t + \epsilon^{dp}_{t+1} \qquad \text{(A.12)}
\end{aligned}
$$

Following the steps of Campbell and Shiller (1988), it is straightforward to show that

$$
r_{t+1} \approx k + \Delta d_{t+1} + dp_t - \overline{\rho} dp_{t+1} \qquad \text{(A.13)}
$$

The final step of the proof consists in substituting Equations (A.7), (A.11) and (A.12) in Equation (A.13):

$$
\begin{aligned}
r_{t+1} &= \mathbb{E}_t\left(k + \Delta d_{t+1} + dp_t - \overline{\rho} dp_{t+1}\right) + \epsilon^r_{t+1} \\
&= k + \phi_0 + (1+\phi_1)ig_t + \phi_2 drp_t + dp_t - \overline{\rho}\left(k_2 + (1-\alpha_1)k_1\right) + \epsilon^r_{t+1} \\
&\quad -\overline{\rho}\left(\alpha_1 dp_t - \left[\frac{1+\phi_1}{1-\overline{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)}\right](\delta_1-\alpha_1)ig_t - \frac{\phi_1\phi_2 ig_t}{1-\overline{\rho}\phi_2} - \frac{\phi_2-\alpha_1}{1-\overline{\rho}\phi_2}\phi_2 drp_t\right) \\
&= \underbrace{k + \phi_0 - \overline{\rho}\left(k_2 + (1-\alpha_1)k_1\right)}_{\Psi} + (1-\overline{\rho}\alpha_1)\left(dp_t + \left[\frac{1+\phi_1}{1-\overline{\rho}\delta_1} + \frac{\bar{\rho}\phi_1\phi_2}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)}\right] ig_t\right) \\
&\quad + (1-\overline{\rho}\alpha_1)\left(\frac{\phi_2 drp_t}{1-\overline{\rho}\phi_2}\right) + \epsilon^r_{t+1} \\
r_{t+1} &\equiv \Psi + (1-\overline{\rho}\alpha_1)\underbrace{\left(dp_t + \frac{1+\phi_1}{1-\overline{\rho}\delta_1} ig_t + \frac{\bar{\rho}\phi_1\phi_2 ig_t}{(1-\bar{\rho}\delta_1)(1-\bar{\rho}\phi_2)} + \frac{\phi_2}{1-\overline{\rho}\phi_2} drp_t\right)}_{dp^{corr}} + \epsilon^r_{t+1} \qquad \text{(A.14)}
\end{aligned}
$$

This completes the proof of Proposition 2. ■