CFFI Working Paper No.26-06

A New Approach to Connecting the Dividend-Price Ratio and Stock Returns

Wenting Liao

liaowenting@ruc.edu.cn

School of Finance,

Renmin University of China

Beijing, China

A New Approach to Connecting the Dividend-Price Ratio and Stock Returns

Wenting Liao

Abstract

There is a long-lasting debate on the performance of dividend-price ratio on the stock returns predictability. Most of the literature argues that the predictability decreases after the 1990s. Since Campbell-Shiller decomposition shows that the dividend-price ratio contains the information of both the future returns and future dividend growth, a linear predictive regression of stock returns on the dividend-price ratio may generate biased results due to the measurement error or omitted variables issues. Therefore, this paper proposes a new approach to study the nonlinear Granger causality of dividend-price ratio on stock returns. We conduct an unobserved component model and connect stock returns and the dividend-price ratio through their innovations. We show that our model can be represented by a reduced-form ARMAX process, and it can increase the in-sample predictability for all the sample periods.

1 Introduction

The dividend-price ratio contains the information of future stock returns. Campbell and Shiller (1988) show the relationship between dividend-price ratio, future returns, and future dividend growth through a log-linearization approximation. Since then, a large literature successfully predicts stock returns with the dividend-price ratio. However, the return predictability seems disappeared after the 1990s. Researchers attribute this forecasting ability deterioration to structural breaks, the dividend payout policy changes, or the nonlinearity of the prediction model. In linear predictive regression,

$$r_t = \alpha + \beta dp_{t-1} + \varepsilon_t, \tag{1}$$

where stock return r_t is regressed on lagged dividend-price ratio, dp_{t-1}, the expected return μ_t is a linear function of the predictor variable,

$$\mu_t = E[r_t] = \alpha + \beta dp_{t-1}. \tag{2}$$

Then the correlation between the expected return and the dividend-price ratio, $corr(E[r_t], dp_{t-1})$, equals 1. However, sometimes $corr(E[r_t], dp_{t-1}) < 1$ due to measurement error, structure breaks, omitted variables and some other reasons. The linear predictive model cannot find significant Granger causality and lose its out-of-sample predictability. Thus we propose a novel ARMAX model to capture the nonlinear Granger causality. Besides, the dividend-price ratio is highly persistent, which could make things even worse. Instead of regressing the stock return directly on the dividend-price ratio, our proposed model connects these two variables through their innovations, avoiding the persistence issue. We show how to conduct the model to state-space form and estimate through a two-step estimation. Then we design a simulation exercise to show that the proposed model performs better both in the size of the test and in the power of the test if the true DGP is an unobserved component model. Finally, we report that the proposed model significantly increases in-sample predictability compared with the linear predictive regression. However, the out-of-sample predictability is ambiguous.

The rest of the paper is organized as follows. Section 2 reviews the related literature. Section 3 introduces the proposed model and the estimation method. 4 shows the simulation evidence. 5 reports both the in-sample and out-of-sample analysis. 6 concludes the paper.

2 Literature Review

There is a large body of literature shows that the stock returns can be predicted by dividend-price ratio, see Rozeff (1984), Campbell and Shiller (1988), Bekaert and Hodrick (1992), Cochrane (2011), Cochrane (2008), Campbell and Yogo (2006), Fama and French (1988), and Menzly, Santos and Veronesi (2004). However, some literature argues that the dividend-price ratio predicts stock returns poorly in terms of the out-of-sample performance, especially after 1990s (Welch and Goyal (2008), Goyal and Welch (2003)). There are two main strands of literature studying this phenomenon.

The first strand attributes the disappearing return predictability to the structural breaks in the predictor variables. Rapach and Wohar (2006) find strong evidence of structural breaks in dividend-price ratio. They also find that the predictive power of the dividend-price ratio drops dramatically after 1990Q3. Lettau and Van Nieuwerburgh (2008) show empirical evidence for shifts in the steady-state mean of financial ratios. They further report that the adjusted dividend-price ratio increase, accounting for the shifts, could increase the predictability. In our paper, we also check the change of the predictability by splitting the full sample into different subsamples according to the breakpoints provided in the existing literature.

The other strand of literature argues that, according to Miller and Modigliani (1961), the stock valuation is affected by all the cash flows but not only the dividend. (Robertson and Wright (2006), Boudoukh et al. (2007)). In other words, using the dividend-price ratio as a predictor may have a measurement error issue. Therefore Boudoukh et al. (2007) investigate the predictability of different measures of payout yield and find both the payout yields and net payout yields work well in forecasting the stock returns. Robertson and Wright (2006) construct an adjusted dividend yield which includes all cash flows to shareholders, and find it has significant predictability. Another advantage of these alternative measures is that, unlike the dividend instability caused by changes of policies (Fama and French (2001), Allen and Michaely (2003)), they are stable and do not subject to the structural break issue. Instead of finding alternative cash flows, Kim and Park (2013) directly estimates a time-varying long-run relationship between dividend and stock price, and they find predictive regression with their adjusted series as the predictor can beat the random-walk model both in-sample and out-of-sample. In our model, we assume the dividend-price ratio predicts the stock returns though a latent factor, which filters out the measurement error.

3 Methodology

The stock return is defined as

$$R_{t+1} \equiv (P_{t+1} + D_{t+1})/P_t, \tag{3}$$

where R_{t+1} is the stock return at time period $t+1$, P_{t+1} and D_{t+1} are stock price and dividend paid at time period $t+1$. From this identity, Campbell and Shiller (1988) propose a linear approximation of log divide-price ratio,

$$dp_t \simeq \sum_{j=1}^{k} \rho^{j-1}(r_{t+j} - \Delta d_{t+j}) + \rho^k dp_{t+k}, \tag{4}$$

where the log dividend-price ratio, dp_t, is approximately equal to the sum of the present value of future stock returns, r_{t+j}, and future dividend growth rates, Δd_{t+j}, and a constant, $\rho^k dp_{t+k}$. If we assume future stock returns and dividend growth rates can be predicted by some underlying factors whose laws of motion follow AR(1) processes,

$$r_{t+1} = \mu_t^r + \varepsilon_{t+1}^r,\ \varepsilon_{t+1}^r \sim i.i.dN(0, \sigma_r^2), \tag{5}$$

$$\Delta d_{t+1} = \mu_t^d + \varepsilon_{t+1}^d,\ \varepsilon_{t+1}^d \sim i.i.dN(0, \sigma_d^2), \tag{6}$$

$$(1 - \psi_r L)\mu_t^r = \omega_t^r,\ \omega_t^r \sim i.i.dN(0, \sigma_{\omega r}^2), \tag{7}$$

$$(1 - \psi_d L)\mu_t^d = \omega_t^d,\ \omega_t^d \sim i.i.dN(0, \sigma_{\omega d}^2). \tag{8}$$

Then at time period t, the expected future returns and dividend growth can be expressed as follows,

$$E(r_{t+1+j}) = \psi_r^j \mu_t^r \tag{9}$$

$$E(\Delta d_{t+1+j}) = \psi_d^j \mu_t^d. \tag{10}$$

Substitute equation (9) and (10) into the Campbell-Shiller decomposition (equation (4)),

$$dp_t \simeq \frac{1}{1 - \rho\psi_r^j}\mu_t^r - \frac{1}{1 - \rho\psi_d^j}\mu_t^d, \tag{11}$$

where the dividend-price ratio is a linear combination of two underlying factors. If we consider a extreme case, where future dividend growth is not predictable (Cochrane (2008)), which implies $Var(\omega_t^d) = 0$, and

$$dp_t = \frac{1}{1 - \rho\psi_r^j}\mu_t^r. \tag{12}$$

Then substitute equation (12) into the dynamics of the future stock return (equation (5)),

$$r_{t+1} = (1 - \rho\psi_r^j)dp_t + \varepsilon_{t+1}^r, \tag{13}$$

where the dividend-price ratio is perfectly correlated with μ_t^r, which is implicitly assumed in the linear predictive regression (equation (1)). However, in general, both $Var(\omega_t^d)$ and $Var(\omega_t^r)$ are great than zero. Then substitute equation (11) into the dynamics of the future stock return (equation (5)),

$$r_{t+1} = (1 - \rho\psi_r^j)dp_t + \frac{1 - \rho\psi_r^j}{1 - \rho\psi_d^j}\mu_t^d + \varepsilon_{t+1}^r. \tag{14}$$

This shows that in the predictive regression, the dividend-price ratio is not perfectly correlated with the future returns. Therefore, we have to capture the nonlinear Granger causality.

Even we assume $Var(\omega_t^d) = 0$ is the real situation, we still have the same issue caused by the measurement error. Due to the regulation or payout policies, the observed dividend is an imperfect measure of the underlying theoretical payout and subject to measurement error.

3.1 The Model

First we consider that the stock return r_t is predicted by the unobserved expectation, μ_{t-1}, the unexpected shock, ε_t, and a constant a:

$$r_t = \alpha + \mu_{t-1} + \varepsilon_t, \varepsilon_t \sim i.i.dN(0, \sigma_\varepsilon^2). \tag{15}$$

The expectation μ_t follows a stationary AR(1) process, and ω_t is the shock to the expectation on time t.

$$\mu_t = \psi\mu_{t-1} + \omega_t, \omega_t \sim i.i.dN(0, \sigma_\omega^2). \tag{16}$$

The dynamics of the dividend-price ratio, dp_t , follows a stationary AR(1) process, and v_t is the shock to the predictor variable, b is the unconditional mean.

$$dp_t - b = \phi(dp_{t-1} - b) + v_t, v_t \sim i.i.dN(0, \sigma_v^2). \tag{17}$$

We assume the innovations are correlated with each other only with the same time period:

$$E(\varepsilon_t \omega_t) = \sigma_{\varepsilon,\omega} \tag{18}$$

$$E(\varepsilon_t v_t) = \sigma_{\varepsilon,v} \tag{19}$$

$$E(\omega_t v_t) = \sigma_{\omega,v} \tag{20}$$

$$E(\varepsilon_t \omega_{t+j}) = 0,\ \ for\ j \neq 0 \tag{21}$$

$$E(\varepsilon_t v_{t+j}) = 0,\ \ for\ j \neq 0 \tag{22}$$

$$E(\omega_t v_{t+j}) = 0,\ \ for\ j \neq 0 \tag{23}$$

Then we connect the stock returns to the dividend-price ratio through the correlations between these innovations. To illustrate this connection, we first make an orthogonal projection of ω_t on v_t:

$$\omega_t = \gamma v_t + \omega_t^*. \tag{24}$$

We also make an orthogonal projection of ε_t on v_t:

$$\varepsilon_t = \delta v_t + e_t. \tag{25}$$

Then we show that our model can be represented by a unique reduced-form ARMAX model. Multiply both side of equation 15 by $(1 - \psi L)$ to obtain:

$$(1 - \psi L) r_t = (1 - \psi)\alpha + \omega_{t-1} + (1 - \psi L)\varepsilon_t \tag{26}$$

Substitute equation 24 and 25 into equation 26:

$$(1-\psi L)r_t = (1-\psi)\alpha + \gamma v_{t-1} + \omega^*_{t-1} + (1-\psi L)(\delta v_t + e_t) \tag{27}$$

$$= (1-\psi)\alpha + (\gamma - \psi\delta)v_{t-1} + \delta v_t + (\omega^*_{t-1} + (1-\psi L)e_t) \tag{28}$$

We further use a MA(1) process u_t to replace the last term of the right hand side of the above equation:

$$(1-\psi L)r_t = (1-\psi)\alpha + (\gamma - \psi\delta)v_{t-1} + \delta v_t + (1+\theta L)u_t \tag{29}$$

Therefore, the coefficients before v_{t-1} captures the nonlinear Granger causality between the dividend-price ratio and stock returns.

3.2 Estimation Method

We apply a two-step estimation to estimate the coefficients of the proposed ARMAX model. In the first step, we estimate equation 17 by OLS, and collect the residuals $\hat{v}_t$. In the second step, we replace v_t in equation 29 with $\hat{v}_t$ and conduct it into a state-space model. The measurement equation is

$$\begin{bmatrix} r_t \\ \hat{v}_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_t \\ u_t \\ \hat{v}_t \end{bmatrix} \tag{30}$$

$$(\tilde{r}_t = H\tilde{\xi}_t) \tag{31}$$

The transition equation is

$$\begin{bmatrix} r_t \\ u_t \\ \hat{v}_t \end{bmatrix} = \begin{bmatrix} (1-\psi)\alpha \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \psi & \theta & \Gamma \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} r_{t-1} \\ u_{t-1} \\ \hat{v}_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & \delta \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_t \\ \hat{v}_t \end{bmatrix} \tag{32}$$

$$(\tilde{\xi}_t = A + F\tilde{\xi_{t-1}} + RU_t, U_t \sim i.i.d.N(0, \Omega_U)) \tag{33}$$

where, $\Gamma = \gamma - \psi\delta$.

4 Simulation Evidence

Before we start our empirical study, we check the performance of our proposed model through a simulation exercise. We assume the DGP is as follows,

$$y_t = \mu_{t-1} + \varepsilon_t \tag{34}$$

$$\mu_t = \psi\mu_{t-1} + \omega_t \tag{35}$$

$$x_t = \phi x_{t-1} + v_t \tag{36}$$

$$\begin{bmatrix} \varepsilon_t \\ \omega_t \\ v_t \end{bmatrix} \sim i.i.d.N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\varepsilon^2 & \rho_{\varepsilon\omega}\sigma_\varepsilon\sigma_\omega & \rho_{\varepsilon v}\sigma_\varepsilon\sigma_v \\ \rho_{\varepsilon\omega}\sigma_\varepsilon\sigma_\omega & \sigma_\omega^2 & \rho_{\omega v}\sigma_\omega\sigma_v \\ \rho_{\varepsilon v}\sigma_\varepsilon\sigma_v & \rho_{\omega v}\sigma_\omega\sigma_v & \sigma_v^2 \end{bmatrix} \right) \tag{37}$$

where y_t is the variable we are interested in, which can be predicted by an underlying factor μ_{t-1}. We assume both the underlying factor μ_t and the predictor x_t follow an AR(1) process. The predictor connects with y_t and the underlying factor through their innovations. We explore different sets of parameters. We set $\phi = 0.95$, which captures the high persistence of the predictor. For the AR parameter in the underlying factor law of motion, ψ, we assign it different values to study how the different dynamics affect the estimation. Then we assume the predictor, the underlying factor, and the innovation to y_t have the same variance, $\sigma_x^2 = \sigma_\mu^2 = \sigma_\varepsilon^2 = 1$. Since μ_{t-1} is orthogonal to ε_t by construction, we also assume there is no correlation between ε_t and ω_t, $\rho_{\varepsilon\omega} = 0$. Note that in the DGP, x_t does not predict y_{t+1} directly, but through the relationship between their innovations. $\rho_{\omega v}$ determines the predictability through the latent factor channel, and $\rho_{\varepsilon v}$ determines the predictability through the i.i.d error term. Thus, we choose different values for $\rho_{\varepsilon v}$ and $\rho_{\omega v}$. The sample size $T = 300$, and the number of simulations is 5000. We compare the estimation results of our proposed model with those of linear predictive regressions.

4.1 Case 1: Low Persistence ($\psi = 0.3$)

Firstly, we consider the case where the underlying factor μ_t has low persistence and assume $\psi = 0.3$. To test for the asymmetric effect, we consider both a positive value and a negative value for the correlation $\rho_{\varepsilon v}$. Table 1 reports the simulation result of the benchmark linear predictive regression. Column (1) lists the variables in the simulation. $\hat{\beta}$ is the OLS estimator, $\hat{SE}(\hat{\beta})$ is the corresponding standard error. To deal with the small sample bias issue in the predictive regressions, we follow Amihud and Hurvich (2004). They firstly estimate the autocorrelation coefficient ρ of the predictor x_t,

$$x_t = \phi x_{t-1} + v_t. \tag{38}$$

Then they correct the bias in $\hat{\phi}$, and collect the corrected $\hat{v}_t^c$,

$$\hat{\phi}^c = \hat{\phi} + (1 + 3\hat{\phi})/n + 3(1 + 3\hat{\phi})/n^2, \tag{39}$$

where n is the number of the observations and $\hat{v}_t^c$ is the residual in equation (38) when ϕ is replaced by $\hat{\phi}^c$. Finally, they insert $\hat{v}_t^c$ to the linear predictive regression to get the bias corrected estimator $\hat{\beta}^c$. $S\hat{E}^c(\hat{\beta}^c)$ is the corrected standard error, calculated by the following equation,

$$S\hat{E}^c(\hat{\beta}^c) = \sqrt{[\frac{\hat{\rho}_{\varepsilon v}\hat{\sigma}_\varepsilon}{\hat{\sigma}_v}]^2 V\hat{a}r(\hat{\phi}^c) + [S\hat{E}(\hat{\beta}^c)]^2} \tag{40}$$

The left panel (Column (2) to (4)) reports the case where $\rho_{\varepsilon v}$ is positive, and the right panel (column (5) to (7)) reports the other case where $\rho_{\varepsilon v}$ is negative. Column (2) is the mean of the estimators. Column (3) shows the rejection rates, which are the numbers of cases, where $|t| > 1.96$, divided by the total number of simulations. The null hypothesis is that x_{t-1} does not Granger cause y_t, which implies $\beta = 0$. Column (4) reports the mean squared error. This table shows the rejection rates of the benchmark model are more than 8% for all the cases, which are greater then the theoretical level (5%). The bias correction procedure helps to decrease the R^2 and MSE for both cases. However, its effect on the rejection rate is ambiguous. Since this procedure only shifts the entire estimator distribution toward the same direction, it eliminates biased estimator on the one side, however, at the same time, it will generate new biased estimators on the other side.

[Table 1 about here.]

Table 2 reports the simulation result of the proposed ARMAX model. The parameters with a superscript $'c'$ are the estimators that we use $\hat{v}_t^c$ instead of $\hat{v}_t$ in the second step estimation. The left panel (Column (2) to (5)) reports the case where $\rho_{\varepsilon v}$ is positive, and the right panel (column (6) to (9)) reports the other case where $\rho_{\varepsilon v}$ is negative. The extra columns (column (2) and (6)) in the table report the true value of the estimator, implied by the DGP. The null hypothesis of column (4) and (8) is the estimator equals its true value. If we capture the Granger causality by parameter γ, the rejection rates are slightly lower than those in the benchmark model. However, the rejection rates are still greater than the theoretical level.

Since δ is well estimated, the high rejection rate of $\hat{\Gamma}$ is caused by the poor estimation of ψ when the predictor has a low persistence. If we want to capture the Granger causality by the parameter Γ, we have to do another exercise, where $\rho_{\varepsilon v} = \rho_{\omega v} = 0$.

[Table 2 about here.]

Then we investigate the probability that the t-test can reject the null hypothesis when x_{t-1} Granger causes y_t. Table 3 reports the power of the test of the linear predictive model, and Table 4 reports that of the proposed ARMAX model. The left panel reports the case where $\rho_{\varepsilon v}$ is positive, and the right panel reports the other case where $\rho_{\varepsilon v}$ is negative. The null hypothesis for table 3 is $\beta = 0$, and for table 4 is $\gamma = 0$. The numbers in the bracket are the size-adjusted t-statistics. These two tables show that the proposed ARMAX model increases the rejection rates significantly, and it also increases the R^2. These results do not change due to the sign of $\rho_{\varepsilon v}$.

[Table 3 about here.]

[Table 4 about here.]

4.2 Case 2: High Persistence ($\psi = 0.9$)

When the underlying factor is highly persistent, we conduct a similar simulation exercise as in case 1 but only change ψ in the DGP from 0.3 to 0.9 to produce the high persistence series. Table 5 and table 6 report the size of the test for the benchmark linear predictive regression and the proposed ARMAX model. In this high persistence case, all the parameters in the ARMAX model are well estimated. The rejection rates are all close to the 5% theoretical level, and all the MSEs are very small. The rejection rates in the benchmark model are all above 25%, which implies our proposed model works better in the high persistence situation.

[Table 5 about here.]

[Table 6 about here.]

Table 7 and table 8 report the power of the test. The null hypothesis in table 7 is $\beta = 0$, and in table 8 is $\gamma = 0$. The rejection rates of the proposed ARMAX are greater than those of the benchmark predictive regression.

[Table 7 about here.]

[Table 8 about here.]

5 Empirical Results

5.1 Data

The stock return is constructed from the Center for Research in Security Press (CRSP) value-weighted monthly return. The dividend-price ratio is calculated from the portfolio returns with and without dividends. Define the portfolio returns with dividend at time period $t+1$, R_{t+1}, as

$$R_{t+1} \equiv \frac{P_{t+1} + D_{t+1}}{P_t} - 1, \tag{41}$$

and portfolio returns without dividend at time period $t+1$, R^x_{t+1}, as

$$R^x_{t+1} \equiv \frac{P_{t+1}}{P_t} - 1 \tag{42}$$

Then, the dividend yield is as follows,

$$\frac{D_{t+1}}{P_t} = R_{t+1} - R^x_{t+1}. \tag{43}$$

Finally, the dividend-price ratio at time period t+1 can be calculated as

$$\frac{D_{t+1}}{P_{t+1}} = \frac{R_{t+1} - R^x_{t+1}}{1 + R^x_{t+1}}. \tag{44}$$

The full sample is quarterly data from 1953Q3 to 2020Q4.

5.2 In-sample analysis

We first check the in-sample predictability. The left panel (Column (2) and (3)) in table 9 reports the OLS estimation of the predictive regression, equation (1). The right panel (Column (4) and (5)) reports the results from the proposed ARMAX model, equation (29). Here we use the $\hat{\Gamma}$ to measure the nonlinear Granger causality. In our proposed model, the innovation to lagged dividend-price ratio predicts the stock returns through two channels. Equation (24) describes the underlying factor channel, where γ is the effect of the innovation to the dividend-price ratio on the innovation to the underlying factor. Equation (25) describes the unexpected shock channel, where δ is the effect of the innovation to the dividend-price ratio on the innovation to the unexpected shock. Recall that $\Gamma = \gamma - \psi\delta$, which is the aggregate effect of v_{t-1} on the stock return r_t. The relative importance of these two effects depends on the size of ψ. If ψ increases, the effect δ becomes more important in the aggregate effect. It is also easy to see from equation (16). When ψ increases, the underlying factor will become more persistent, and the variance of ω decreases. Since β in the predictive regression measures the total Granger causal effect, it is reasonable to report $\hat{\Gamma}$ instead of $\hat{\gamma}$.

Table 9 reports the in-sample analysis. The first three rows replicate the results in Rapach and Wohar (2006), where they focus on the after-war data and set a breakpoint at 1990Q4. From 1953Q3 to 2000Q4, the OLS estimator is not significant, while the coefficient in our proposed model is significant. Our R^2 is greater than the OLS method. The second row reports the period before the 1990s. The estimators in both methods are significant, and both R^2 are high. The third row reports the result after the 1990s. The OLS estimator loses its predictability. However, the coefficient in the proposed model is still significant. The R^2 in the proposed model is slightly greater than that in the linear predictive model. Then we extend the sample period to 2007Q2 when is just before the financial crisis. We find that the R^2 increases for both models. Finally, we report the period of the most up-to-date data, which is from 1990Q4 to 2020Q4. The R^2 in the proposed model is greater than that of the OLS estimation. Overall, the right panel shows that $\hat{\Gamma}$ is significant in all periods. Besides, the values of R^2 are greater than that of the OLS method for all sample periods. Figure 1 shows the estimated y_t of the two models against the original data, where the proposed model captures the dynamics of the return better than the OLS etimation.

[Table 9 about here.]

[Figure 1 about here.]

6 Conclusion

This paper proposes a novel approach to capture the nonlinear Granger causality in a predictive system. We show that our model can be represented as an ARMAX model. Then we explain how to estimate the model through a two-step estimation strategy. Through a simulation exercise, we prove that our model works better when the underlying factor is highly persistent. Finally, we show that the proposed model increases the in-sample predictability.

References

Allen, Franklin, and Roni Michaely. 2003. "Payout policy." *Handbook of the Economics of Finance*, 1: 337–429.

Amihud, Yakov, and Clifford M Hurvich. 2004. "Predictive regressions: A reduced-bias estimation method." *Journal of Financial and Quantitative Analysis*, 813–841.

Bekaert, Geert, and Robert J Hodrick. 1992. "Characterizing predictable components in excess returns on equity and foreign exchange markets." *The Journal of Finance*, 47(2): 467–509.

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R Roberts. 2007. "On the importance of measuring payout yield: Implications for empirical asset pricing." *The Journal of Finance*, 62(2): 877–915.

Campbell, John Y, and Motohiro Yogo. 2006. "Efficient tests of stock return predictability." *Journal of financial economics*, 81(1): 27–60.

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

Cochrane, John H. 2008. "The dog that did not bark: A defense of return predictability." *The Review of Financial Studies*, 21(4): 1533–1575.

Cochrane, John H. 2011. "Presidential address: Discount rates." *The Journal of finance*, 66(4): 1047–1108.

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

Fama, Eugene F, and Kenneth R French. 2001. "Disappearing dividends: changing" rm characteristics or lower propensity to pay?" *Journal of Financial Economics*, 60(3): 43.

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

Kim, Chang-jin, and Cheolbeom Park. 2013. "Disappearing dividends: Implications for the dividend–price ratio and return predictability." *Journal of Money, Credit and Banking*, 45(5): 933–952.

Lettau, Martin, and Stijn Van Nieuwerburgh. 2008. "Reconciling the return predictability evidence: The review of financial studies: Reconciling the return predictability evidence." *The Review of Financial Studies*, 21(4): 1607–1652.

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

Miller, Merton H, and Franco Modigliani. 1961. "Dividend policy, growth, and the valuation of shares." *the Journal of Business*, 34(4): 411–433.

Rapach, David E, and Mark E Wohar. 2006. "Structural breaks and predictive regression models of aggregate US stock returns." *Journal of Financial Econometrics*, 4(2): 238–274.

Robertson, Donald, and Stephen Wright. 2006. "Dividends, total cash flow to shareholders, and predictive return regressions." *Review of Economics and Statistics*, 88(1): 91–99.

Rozeff, Michael S. 1984. "Dividend yields are equity risk premiums." *Journal of Portfolio management*, 68–75.

Welch, Ivo, and Amit Goyal. 2008. "A comprehensive look at the empirical performance of equity premium prediction." *The Review of Financial Studies*, 21(4): 1455–1508.

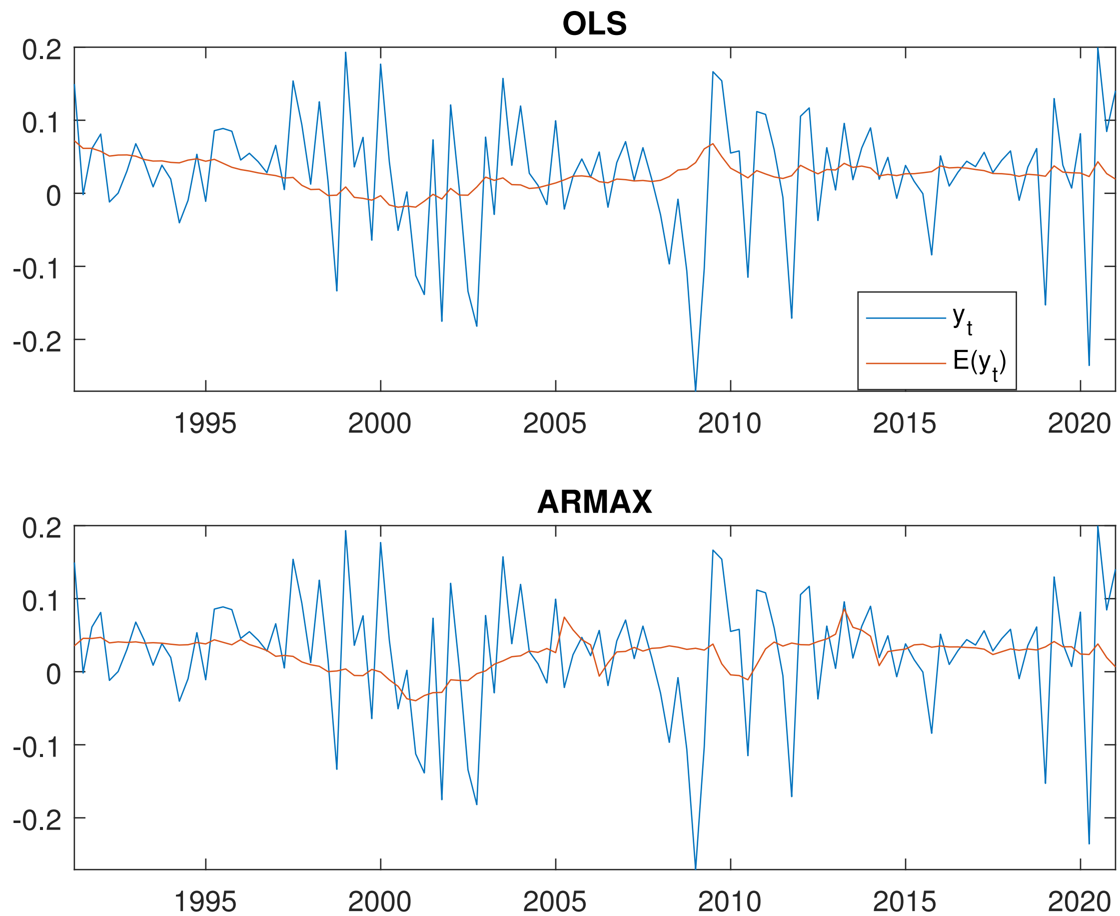

Figure 1: $E(y_t)$ against y_t

Table 1: Low Persistence, Benchmark Model, size of the test($\rho_{\omega v} = 0$)

(1)	(2)	(3)	(4)	(5)	(6)	(7)
		$\rho_{\varepsilon v} = 0.7$			$\rho_{\varepsilon v} = -0.7$	
	Mean	Rej Rate	MSE	Mean	Rej Rate	MSE
$\hat{\beta}$	-0.0311	8.86%	0.0133	0.0308	8.12%	0.0133
	(0.1112)			(0.1109)		
R^2	0.0049			0.0048		
$\hat{SE}(\hat{\beta})$	0.0993			0.0995		
$\hat{\beta}^c$	-0.0023	8.58%	0.0125	0.0020	8.26%	0.0124
	(0.1117)			(0.1114)		
R^2	0.0045			0.0044		
$\hat{SE}^c(\hat{\beta}^c)$	0.0987			0.0989		

note: Rejection rate is the number of case, where $|t| > 1.96$, divided by the total number of simulation. The number in the parentheses is the standard deviation of the simulation distribution. The standard error is the Newey–West standard error.

Table 2: Low Persistence, Proposed ARMAX Model, size of the test ($\rho_{\omega v} = 0$)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$\rho_{\varepsilon v} = 0.7$				$\rho_{\varepsilon v} = -0.7$			
	True	Mean	Rej Rate	MSE	True	Mean	Rej Rate	MSE
$\hat{\gamma}$	0	-0.0349	7.64 %	0.0610	0	0.0367	8.06 %	0.0612
		(0.2445)				(0.2446)		
R^2		0.0060				0.0057		
$\hat{SE}(\hat{\gamma})$		0.2287				0.2287		
$\hat{\gamma}^c$	0	-0.0054	7.44 %	0.0598	0	0.0071	7.70 %	0.0599
		(0.2445)				(0.2447)		
R^2		0.0059				0.0057		
$\hat{\psi}$	0.3000	0.2486	16.68 %	0.1020	0.3000	0.2485	17.76 %	0.1055
$\hat{\psi}^c$	0.3000	0.2489	17.08 %	0.1026	0.3000	0.2478	17.76 %	0.1061
$\hat{\theta}$	-0.1055	-0.0593	17.50 %	0.1070	-0.1055	-0.0600	18.18 %	0.1109
$\hat{\theta}^c$	-0.1055	-0.0595	17.54 %	0.1078	-0.1055	-0.0590	18.24 %	0.1113
$\hat{\delta}$	2.2418	2.2411	5.16 %	0.0505	-2.2418	-2.2459	4.88 %	0.0495
$\hat{\delta}^c$	2.2418	2.2395	5.10 %	0.0505	-2.2418	-2.2443	4.80 %	0.0494
$\hat{\Gamma}$	-0.6725	-0.5931	15.42 %	0.5816	0.6725	0.5956	16.74 %	0.6146
$\hat{\Gamma}^c$	-0.6725	-0.5644	15.22 %	0.5706	0.6725	0.5645	16.18 %	0.6033

note: $\hat{\Gamma}$ is the coefficient before the v_{t-1} in the reduced form ($\Gamma = \gamma - \psi\delta$).
In the MLE, we estimate $\hat{\Gamma}$ directly and back out $\hat{\gamma}$.

Table 3: Low Persistence, Benchmark Model, power of the test($\rho_{\omega v} = 0.5$)

(1)	(2)	(3)	(4)	(5)	(6)	(7)
		$\rho_{\varepsilon v} = 0.7$			$\rho_{\varepsilon v} = -0.7$	
	Mean	Rej Rate	MSE	Mean	Rej Rate	MSE
$\hat{\beta}$	0.2041	66.48%	0.0462	0.2664	68.40%	0.0899
	(0.0673)	[66.24%]		(0.1377)	[52.48%]	
R^2	0.0205			0.0316		
$\hat{SE}(\hat{\beta})$	0.0908			0.1041		
$\hat{\beta}^c$	0.2331	78.26%	0.0588	0.2379	59.50%	0.0757
	(0.0673)	[64.22 %]		(0.1383)	[52.04%]	
R^2	0.0201			0.0313		
$\hat{SE}^c(\hat{\beta}^c)$	0.0908			0.1030		

note: Rejection rate is the number of case, where $|t| > 1.96$, divided by the total number of simulation. The number in the parentheses is the standard deviation of the simulation distribution. The standard error is the Newey–West standard error. The number in the bracket is the size-adjusted t-test.

Table 4: Low Persistence, Proposed ARMAX Model, power of the test ($\rho_{\omega v} = 0.5$)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$\rho_{\varepsilon v} = 0.7$				$\rho_{\varepsilon v} = -0.7$			
	True	Mean	Rej Rate	MSE	True	Mean	Rej Rate	MSE
$\hat{\gamma}$	1.5275	1.4989	100.00 %	2.2919	1.5275	1.5566	100.00 %	2.4684
		(0.2128)	[100.00%]			(0.2129)	[100.00%]	
R^2		0.1156				0.1219		
$\hat{SE}(\hat{\gamma})$		0.2106				0.2094		
$\hat{\gamma}^c$	1.5275	1.5194	100.00 %	2.3534	1.5275	1.5278	100.00 %	2.3800
		(0.2119)	[100.00%]			(0.2140)	[100.00%]	
R^2		0.1146				0.1205		
$\hat{SE}(\hat{\gamma}^c)$		0.2090				0.2099		

note: Rejection rate is the number of case, where $|t| > 1.96$, divided by the total number of simulation. The number in the parentheses is the standard deviation of the simulation distribution. The number in the bracket is the size-adjusted t-test.

Table 5: High Persistence, Benchmark Model, size of the test($\rho_{\omega v} = 0$)

(1)	(2)	(3)	(4)	(5)	(6)	(7)
		$\rho_{\varepsilon v} = 0.7$			$\rho_{\varepsilon v} = -0.7$	
	Mean	Rej Rate	MSE	Mean	Rej Rate	MSE
$\hat{\beta}$	-0.0308	25.60%	0.0515	0.0316	25.38%	0.0498
	(0.2249)			(0.2210)		
R^2	0.0211			0.0199		
$\hat{SE}(\hat{\beta})$	0.1343			0.1337		
$\hat{\beta}^c$	-0.0021	26.14%	0.0512	0.0028	25.00%	0.0495
	(0.2263)			(0.2224)		
R^2	0.0207			0.0195		
$\hat{SE}^c(\hat{\beta}^c)$	0.1339			0.1333		

note: Rejection rate is the number of case, where $|t| > 1.96$, divided by the total number of simulation. The number in the parentheses is the standard deviation of the simulation distribution. The standard error is the Newey–West standard error.

Table 6: High Persistence, Proposed ARMAX Model, size of the test($\rho_{\omega v} = 0$)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$\rho_{\varepsilon v} = 0.7$				$\rho_{\varepsilon v} = -0.7$			
	True	Mean	Rej Rate	MSE	True	Mean	Rej Rate	MSE
$\hat{\gamma}$	0	-0.0297	5.86 %	0.0312	0	0.0322	5.78 %	0.0308
		(0.1741)				(0.1727)		
R^2		0.0051				0.0050		
$\hat{SE}(\hat{\gamma})$		0.1692				0.1689		
$\hat{\gamma}^c$	0	-0.0000	5.60 %	0.0311	0	0.0026	5.02 %	0.0306
		(0.1763)				(0.1751)		
R^2		0.0049				0.0048		
$\hat{SE}(\hat{\gamma}^c)$		0.1714				0.1710		
$\hat{\psi}$	0.9000	0.8797	4.40 %	0.0024	0.9000	0.8800	3.66 %	0.0023
$\hat{\psi}^c$	0.9000	0.8800	4.26 %	0.0023	0.9000	0.8803	3.70 %	0.0023
$\hat{\theta}$	-0.5268	-0.5121	6.06 %	0.0060	-0.5268	-0.5125	5.14 %	0.0057
$\hat{\theta}^c$	-0.5268	-0.5124	5.88 %	0.0060	-0.5268	-0.5128	5.06 %	0.0057
$\hat{\delta}$	2.2418	2.2438	5.10 %	0.0309	-2.2418	-2.2418	5.38 %	0.0306
$\hat{\delta}^c$	2.2418	2.2435	5.02 %	0.0308	-2.2418	-2.2414	5.40 %	0.0305
$\hat{\Gamma}$	-2.0176	-2.0036	5.32 %	0.0434	2.0176	2.0048	5.16 %	0.0425
$\hat{\Gamma}^c$	-2.0176	-1.9743	5.90 %	0.0450	2.0176	1.9756	5.54 %	0.0442

note: $\hat{\Gamma}$ is the coefficient before the v_{t-1} in the reduced form ($\Gamma = \gamma - \psi\delta$).
In the MLE, we estimate $\hat{\Gamma}$ directly and back out $\hat{\gamma}$.

Table 7: High Persistence, Benchmark Model, power of the test($\rho_{\omega v} = 0.5$)

	$\rho_{\varepsilon v} = 0.7$			$\rho_{\varepsilon v} = -0.7$		
	Mean	Rej Rate	MSE	Mean	Rej Rate	MSE
$\hat{\beta}$	0.4646	90.18%	0.2464	0.5209	89.66%	0.3194
	(0.1748)	[68.06%]		(0.2192)	[63.20%]	
R^2	0.1074			0.1293		
$\hat{SE}(\hat{\beta})$	0.1164			0.1262		
$\hat{\beta}^c$	0.4935	92.58%	0.2744	0.4922	87.18%	0.2909
	(0.1758)	[67.72%]		(0.2205)	[63.20%]	
R^2	0.1071			0.1289		
$\hat{SE}^c(\hat{\beta}^c)$	0.1159			0.1257		

note: Rejection rate is the number of case, where $|t| > 1.96$, divided by the total number of simulation. The number in the parentheses is the standard deviation of the simulation distribution. The standard error is the Newey–West standard error. The number in the bracket is the size-adjusted t-test.

Table 8: High Persistence, Proposed ARMAX Model, power of the test($\rho_{\omega v} = 0.5$)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$\rho_{\varepsilon v} = 0.7$				$\rho_{\varepsilon v} = -0.7$			
	True	Mean	Rej Rate	MSE	True	Mean	Rej Rate	MSE
$\hat{\gamma}$	0.6980	0.6727	97.66 %	0.4793	0.6980	0.7317	99.80 %	0.5594
		(0.1636)	[98.36%]			(0.1547)	[99.54%]	
R^2		0.0280				0.0307		
$\hat{SE}(\hat{\gamma})$		0.1599				0.1514		
$\hat{\gamma}^c$	0.6980	0.7039	98.56 %	0.5215	0.6980	0.7039	99.50 %	0.5200
		(0.1614)	[98.54%]			(0.1567)	[99.48%]	
R^2		0.0244				0.0269		
$\hat{SE}(\hat{\gamma}^c)$		0.1573				0.1537		

note: Rejection rate is the number of case, where $|t| > 1.96$, divided by the total number of simulation. The number in the parentheses is the standard deviation of the simulation distribution. The number in the bracket is the size-adjusted t-test.

Table 9: In-Sample diagnostics

(1)	(2)	(3)	(4)	(5)
	OLS		ARMAX	
Sample	$\hat{\beta}$	R^2	$\hat{\Gamma}$	R^2
1953Q3-2000Q4	0.0138	0.0190	0.9482	0.0266
	(0.0190)		(0.0360)	
1953Q3-1990Q3	0.1331	0.0938	0.9952	0.1021
	(0.0300)		(0.0207)	
1990Q4-2000Q4	0.0395	0.0393	1.0100	0.0396
	(0.0280)		(0.0551)	
1990Q4-2007Q2	0.0737	0.0791	0.9909	0.0914
	(0.0337)		(0.0612)	
1990Q4-2020Q4	0.076	0.0473	0.8627	0.0667
	(0.0333)		(0.0585)	

The number in the parentheses is the standard error. For the OLS estimation, we report Newey–West. The standard errors of the proposed ARMAX model are from the inverse Hessian matrix.