That is not my dog: Why doesn't the log dividend-price ratio seem to predict future log returns or log dividend growth?*

Philip H. Dybvig

Washington University in Saint Louis

Huacheng Zhang

Southwestern University of Finance and Economics†

First draft: August, 2016

Current version: December 18, 2018

*We are grateful for helpful comments from Michael Brennan, Pingyang Gao, Jun Liu, Zhongzhi Song, Rossen Valkanov, Russ Wermers, and especially Jason (Huafeng) Chen. We are also grateful for helpful comments from seminar participants at the SFS Cavalcade Asia-Pacific 2017 (Beijing), the GCFC 2018 Conference (Xiamen), the CICF 2018 Conference (Tianjin), the Southwestern University of Finance and Economics, the University of Alabama, and Washington University at Saint Louis.

†corresponding author: zhanghuacheng@gmail.com

That is not my dog: Why doesn't the log dividend-price ratio seem to predict future log returns or log dividend growth?

By Philip H. Dybvig and Huacheng Zhang

Abstract

Campbell and Shiller's "accounting identity" implies that changes in the log dividend-price ratio must predict either future returns or future log dividend growth. However, neither quantity seems to be predictable — a well-known puzzle in the literature. We examine this puzzle step-by-step from theoretical derivation through empirical testing. Stationarity of the log dividend-price ratio is an important assumption behind the accounting identity, but Campbell and Shiller's test justifying this assumption does not make sense, and a corrected test does not reject non-stationarity. Nonetheless, a truncated accounting identity works reasonably well in the existing sample, and we find that the log dividend-price ratio predicts log dividend growth, not returns. Traditional tests using one or a few lags have trouble detecting predictability of log dividend growth because predictability is spread over many periods. Unfortunately, predictability of log dividend growth is not robust to subsamples, and it seems unwise to rely too much on the estimates given that the entire sample includes only five non-overlapping observations.

Key words: return predictability, dividend-price ratio, stationarity test. [JEL G12 G17]

Clouseau: Does your dog bite?

Innkeeper: No.

Clouseau: Nice doggy.

(Clouseau tries to pet the dog on the floor and is bitten)

Clouseau (angry): I thought you said your dog did not bite.

Innkeeper: That is not my dog. [1]

Often, like Clouseau, we get into trouble because we ask the wrong question. This paper examines the failure of the log dividend-price ratio (hereafter LDPR) to predict either future log returns or future log dividend growth. Campbell and Shiller's (1988) "accounting identity" asserts that the current LDPR is approximately equal to a constant plus the sum of "present values" of future log returns minus the sum of "present values" of future log dividend growths.[2] This implies that current LDPR should be able to predict future log returns, or log dividend growth rates, or both. Empirically, however, the LDPR seems to predict neither. This puzzle was examined by Cochrane (2008), who side-stepped the puzzle by assuming a just-identified model, using the analogy of the "dog that didn't bark" from Sherlock Holmes, but that is not our dog. Since, like Clouseau, we do not know which question to ask, we go step-by-step through the entire argument to uncover where the problem is. We find that the analysis fails in two places. First, Campbell and Shiller assume that the long-term mean LDPR exists and justify this assumption by an empirical test that does not actually test this at all. Our correctly specified test fails to reject the null that the long-term mean LDPR does not exist. However, in spite of this, the finite-horizon version of their accounting identity is a very good approximation within our sample. Second, performing a regression having the form of the finite-horizon version within our sample (instead of the usual VAR with one or a few lags) has significant predictability of future log dividend growth but not future log returns. However, we do not want to read too much into this result because the sample has only five nonoverlapping data points, the results change on subsamples, and the estimators are inconsistent if the LDPR is indeed nonstationary.

The Campbell-Shiller "accounting identity" can be derived by starting with the single period definition of returns as the sum of dividends and capital gains. Using algebra, taking logs, and

[1] Edwards (1976).

[2] The "present values" in the accounting identity are computed using an artificial interest rate computed using the long-term mean LDPR, not the market interest rate.

doing a Taylor series expansion around some typical value for the LDPR, we obtain a one-period approximation formula linking log returns and log dividend growth with beginning- and end-of-period LDPR. By telescoping this approximation over many periods, current LDPR can be linearly approximated in a finite horizon by the sum of weighted log returns and the sum of weighted log dividend growth rates over future periods plus weighted LDPR in the final period. The accounting identity is derived by assuming that the long-term mean LDPR exists and expanding around the long-term mean, which implies that the final term disappears when we take the infinite limit of the finite horizon versions.

Campbell and Shiller (1988) purport to show the existence of the long-term mean LDPR, but they use a flawed test of stationarity. They use an augmented Dickey-Fuller test of Phillips and Perron (1988) in which the null is nonstationarity plus a trend, and the alternative is stationarity plus a trend. Unfortunately, the trend implies that the process is stationary and has a long-term mean under neither the null hypothesis nor the alternative hypothesis. Obviously, we cannot test for stationarity if both the null and alternative hypotheses imply nonstationarity. Instead, we conduct an original Dickey-Fuller stationarity test in which the null hypothesis is nonstationarity without a trend, which implies the long-term mean does not exist, and the alternative hypothesis is stationary without a trend, which implies the long-term mean does exist. Using annual data for the S&P 500 index (with Shiller's backfill[3] using data from Cowles (1939)) from 1871 to 2015, we cannot reject the null of nonstationarity, which suggests the long-term mean does not exist. In principle, this is a big problem for the Campbell-Shiller approximation, which is based on an expansion around the long-term mean, but the truncated version of the approximation still works well in our sample looking 30 years out. It is also a big problem for any asymptotic interpretations of the statistical tests, since nonstationarity of LDPR would mean that the approximation error in the derivation will become unbounded over time and we also lose the asymptotic justification of estimates and standard errors.

Although we are skeptical about the existence of the long-term mean LDPR, perhaps the approximation is still useful if we expand around some reasonable value, for example, the sample mean in our current sample. We perform a regression of the LDPR on the sums over future 30 years and the final log dividend-price ratio. All coefficients on independent variables are close to the theoretical value (one or minus one) and the R^2 is close to 100% (98.91%). The approximation is

[3]The backfill is described in Shiller (1981) and Campbell and Shiller (1988).

worse but still acceptable if we drop the final LDPR in year 30 (the R^2 drops to 82.25%) . Therefore, the source of the puzzle is the lack of power in previous tests rather than any intrinsic problem with the approximation, at least for the truncated identity in our current sample.

We test the predictability of stock returns and dividend growth rates using the finite approximation itself with many terms rather than using one or a few lags as is common in the literature. Our estimation uses statistical corrections for the correlation in error terms and for spurious regression bias. The Campbell-Shiller approximation implies that the current LDPR is able to predict either future log returns or log dividend growth or both; we find that log dividend growth is significantly predictable, but log returns are not. The results are robust to expanding the log dividend-price ratio around alternative points rather than the sample mean.

As noted by Cochrane (2008), dividends are smooth. He concludes that log dividend growth is not predictable implying under the model restriction that returns are predictable. However, it seems more accurate to assert that the predictability of log dividend growth is spread over many maturities and that nearby dividends are not very predictable because dividends are smooth. What is happening is that there is small predictability of dividend growth spread over many periods, which is buried by noise in conventional simple regression or vector-autoregressive (VAR) estimation.

The limitation inherent in using a small number of lags to search for predictability of dividend growth seems to be a solution of the puzzle of why the theory (based on an accounting identity and an approximation that is not so bad in the current sample) is hard to verify. In general, the predictability of log dividend growth is difficult to find because of the large prediction error introduced by the unpredictable part of future log returns, which is a common factor with future log dividend growth that cancels in the accounting identity.

Although the best evidence (based on our whole sample) suggests that the LDPR predicts log dividend growth but not log returns, this result seems fragile. For one thing, this result is reversed on the second half-period, consistent with Chen's (2009) "tale of two periods" and explaining an apparent inconsistentcy with a similar regression of Cochrane (2008, Section 7.2). We also worry about the statistical properties of the estimators, both because the whole sample has only about five non-overlapping observations (and subsamples even fewer) and because of the apparent instability over time. If the LDPR is indeed nonstationary, the estimation will not improve as the sample size gets larger, because the Taylor series expansion will become much less accurate as the range of values increases over time. One interesting aspect of the accounting identity is that it is not an economic model since its derivation uses only manipulation of identities and approximations. If we think about the economics, Modigliani-Miller suggests that to first order dividends are irrelevant, which is consistent with instability of these relationships over time. Therefore, both statistical and economic arguments call into question the asymptotic stability of the estimation.

The rest of this paper is organized as follows. We review the approximation leading to the accounting identity in Section 1 and test the quality of this approximation in Section 2. We use a model-implied approach to test the predictability of returns and dividend growth in Section 3. In Section 4, we further analyze the approximation error and how that would change if the LDPR moved away from the current range. Section 5 concludes the paper.

1 Dividend-Price Decomposition

1.1 Theory

We begin by specifying the standard definition relating return, current and future prices, and dividend payment. Define gross investment return over one period as:

$$
\begin{aligned}
(1) \qquad 1+R_{t+1} &= \frac{P_{t+1}+D_{t+1}}{P_t} \\
&= \frac{P_{t+1}}{P_t}\left(1+\frac{D_{t+1}}{P_{t+1}}\right),
\end{aligned}
$$

where P_t and P_{t+1} denote start-of-period and end-of-period stock prices, R_{t+1} denotes the net return over the period, and D_{t+1} denotes the end-of-period dividend.[4] This may seem like a strange way to write the return, since we would normally look at gross capital gains P_{t+1}/P_t and dividend yield D_t/P_t, with information known at the beginning of the period in the denominator. For our purpose,

[4] For empirical tests, we might treat all dividend payments during the period as coming at the end of the period. Alternatively, we could try to construct a more accurate return calculation that takes into account the timing of the dividends and the returns within each period. In practice, these approaches are likely to yield very similar results. Empirical work must also account for splits and distributions other than dividends, although the data we use has already made these adjustments and including them in our notation here would not change the substance of the analysis.

we simply manipulate accounting identities to express returns in this unconventional manner because placing P_t in both denominators would give us a telescoping series in which the final term cannot vanish. Taking logs on both sides, (1) becomes:

$$(2) \quad \log(1+R_{t+1}) = \log\left(\frac{P_{t+1}}{P_t}\right) + \log(1+\exp(\delta_{t+1})),$$

where $\delta_{t+1} \equiv \log(D_{t+1}/P_{t+1})$. We approximate (2) by a first-order Taylor series expansion around $\delta_{t+1} = \delta$. Traditionally, the constant δ is taken to be the long-term mean of the log dividend-price ratio, $\log(D_t/P_t)$, but we will take a broader perspective and view δ as taking on some reasonable value. This distinction could be important, given the evidence later in this section that the long-term mean may not exist. Letting $\rho \equiv 1/(1+\exp(\delta))$, then

$$\begin{aligned} \left.\frac{d\log(1+\exp(\delta_{t+1}))}{d\delta_{t+1}}\right|_{\delta_{t+1}=\delta} &= \left.\frac{\exp(\delta_{t+1})}{1+\exp(\delta_{t+1})}\right|_{\delta_{t+1}=\delta} \\ &= 1-\rho. \end{aligned}$$

Therefore, letting $\kappa \equiv \log(1+\exp(\delta)) - (1-\rho)\delta$, the Taylor approximation is:

$$\begin{aligned} (3) \quad \log(1+R_{t+1}) &= \log\left(\frac{P_{t+1}}{P_t}\right) + \log(1+\exp(\delta_{t+1})) \\ &\approx \log(\frac{P_{t+1}}{P_t}) + \log(1+\exp(\delta)) + (1-\rho)(\delta_{t+1}-\delta) \\ &= \log(P_{t+1}) - \log(P_t) + \log(1-\exp(\delta)) \\ &\quad +(1-\rho)(\log(D_{t+1}) - \log(P_{t+1})) - (1-\rho)\delta \\ &= \kappa + \rho\log(P_{t+1}) + (1-\rho)\log(D_{t+1}) - \log(P_t) \\ &= \kappa - \rho\log(\frac{D_{t+1}}{P_{t+1}}) + (\log(D_{t+1}) - \log(D_t)) + (\log(D_t) - \log(P_t)). \end{aligned}$$

Rewrite (3) as

$$(4) \quad \log\left(\frac{D_t}{P_t}\right) \approx -\kappa + \log(1+R_{t+1}) + \rho\log\left(\frac{D_{t+1}}{P_{t+1}}\right) - \Delta\log(D_{t+1}).$$

We can use (4) itself, changing t to $t+1$, to substitute in for the term $\rho\log(D_{t+1}/P_{t+1})$ on the right-hand side. Doing this repeatedly (changing t to $t+1, t+2, ..., T-1$ in turn), the expression telescopes to become:

$$(5) \quad \log\left(\frac{D_t}{P_t}\right) \approx -\frac{\kappa}{1-\rho}(1-\rho^{T-t}) + \sum_{s=t+1}^{T} \rho^{s-t-1}(\log(1+R_s) - \Delta\log(D_s)) + \rho^{T-t}\log\left(\frac{D_T}{P_T}\right).$$

This is the essential formula we work with. Campbell and Shiller assume the LDPR is stationary and set δ equal to its mean. Since $0 < \rho < 1$, if the LDPR is stationary, the final term vanishes (converges to 0 in probability) as T increases, and we have the asymptotic expression:

$$(6) \quad \log\left(\frac{D_t}{P_t}\right) \approx -\frac{\kappa}{1-\rho} + \sum_{s=t+1}^{\infty} \rho^{s-t-1}(\log(1+R_s) - \Delta\log(D_s)),$$

often referred to in the literature as the *accounting identity*. This identity states that, subject to the quality of the approximation, today's log dividend-price ratio $\log(D_t/P_t)$ is *identically equal to* a constant plus a linear combination of future log returns $\log(1+R_s)$ and future changes in log dividend $\Delta\log(D_s)$. This implies that the log dividend-price ratio must predict one or both of them. It may seem strange that this is an identity (subject to the approximation) and there is no expectation in (6), since it would be naive to think that weighted future log returns are perfectly predictable. The point is that the part of future returns that is not predictable cancels the part of log dividend growth that is not predictable, as will be explored further in Section 3. What we can say is that if the LDPR is not constant, then the LDPR must predict either future log returns or future log dividend growth. This follows from taking the expectation of both sides of (6) conditional on information at time t, since both D_t and P_t are known at time t.

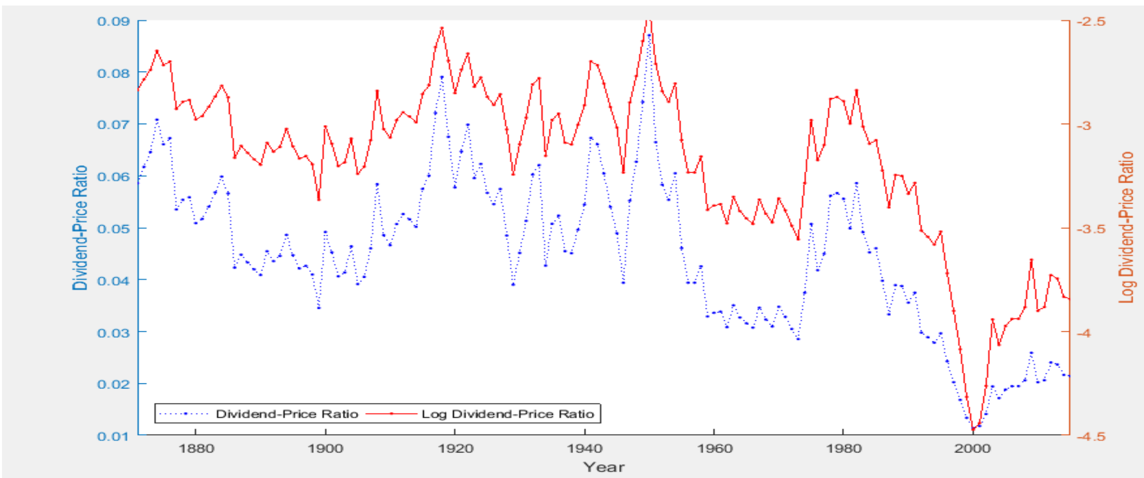
1.2 Stationarity of the LDPR

We now test whether the LDPR is stationary and the long-term mean exists, and the short answer is no, the LDPR is not stationary and its long-term mean does not seem to exist. Figure 1 shows the time series of both the annual dividend-price ratio for the backfilled S&P 500 index (dashed line) and the corresponding LDPR (solid line) between 1871 and 2015. Both are much smaller towards the end than in the first half of the sample.[5] Over the sample period, the average dividend-price ratio is 4.47%

[5] The data were obtained from Robert Shiller's website at http://www.econ.yale.edu/~shiller/data.htm. The early observations are backfilled using data collected by Cowles (1939); see Campbell and Shiller (1988) for details. We

(corresponding to a ρ of 0.95) with a standard deviation of 1.52% while the average log dividend-price ratio is -3.18 with a standard deviation of 0.40. The dividend-price ratio varies around 5% during the Campbell-Shiller period (1871–1986), but declines to around 2% during the post-Campbell-Shiller period (1987–2015). In short, Figure 1 suggests that the long-term means of the dividend-price and log dividend-price ratios may not exist, consistent with the results of our "appropriate" stationarity test.

Figure 1: Dividend-Price Ratio and Log Dividend-Price Ratio (LDPR)



We call our test "appropriate" because unlike the augmented Dickey-Fuller test used by Campbell and Shiller, we use an original Dickey-Fuller test that actually tests for stationarity. For our test, the LDPR is nonstationary under the null and stationary under the alternative, unlike the Campbell-Shiller test under which the LDPR is nonstationary under both null and alternative. The main difference between the two tests is the inclusion of a trend in the augmented Dickey-Fuller test used by Campbell and Shiller. Stationarity around a trend is still nonstationary (except in the degenerate case that the trend is flat, and is not imposed under the null of the augmented Dickey-Fuller test), and implies that the long-term mean does not exist. The stationarity test is specified as $\log(D_t/P_t) = \alpha + \beta \log(D_{t-1}/P_{t-1}) + \varepsilon_t$. The results are reported in Table 1. The Dicky-Fuller statistic is -15.8 over the whole sample period with a 5% critical value of -16.3, so we fail to reject the hypothesis that the LDPR is a non-stationary series, and nonstationarity would imply the long-

focus on annual data because monthly dividend payments are linearly interpolated from annual and quarterly dividend payments, and we do not want to deal with the approximation error this might entail.

term mean LDPR does not exist. It will certainly be a problem over time if the long-term mean does not exist and the LDPR gets more and more dispersion that will make the Taylor approximation disintegrate. This also calls into question the asymptotic justifications of estimators and standard errors. Interestingly, the correct test rejects non-stationarity over the Campbell-Shiller period (statistic -33.0 with 5% critical value -16.3) but not over the later period (statistic -4.8 with 5% critical value -14.6), consistent with Chen's (2009) "tale of two periods." Possible nonstationarity of the LDPR weakens the interpretation of the sample mean as an approximation to the long-term mean, but it doesn't necessarily invalidate the analysis using the current data. In the following analyses, we investigate what we can learn about the approximation and predictability with the possibly nonstationary LPDR data series.

Table 1: LDPR Stationarity Tests

This table reports the results of Dickey-Fuller tests of stationarity of the annual series of the log dividend-price ratio for the back-filled S&P 500 index. The stationarity test is specified as $\log(D_t/P_t) = \alpha + \beta \log(D_{t-1}/P_{t-1}) + \varepsilon_t$. We fail to reject nonstationarity for the entire sample and post-Campbell-Shiller, but but we reject nonstationarity for the Campbell-Shiller subperiod.

dependent variable: $log(D_t/P_t)$	$1871--2015$ entire sample	$1871--1986$ Campbell-Shiller	$1987--2015$ post-Campbell-Shiller
constant	-0.16	-0.39	-0.34
	(0.06)	(0.09)	(0.15)
$log(D_{t-1}/P_{t-1})$	0.89	0.71	0.82
	(0.04)	(0.07)	(0.09)
Dicky-Fuller stat	-15.78	-32.95	-4.76
Dicky-Fuller 5% critical	-16.30	-16.30	-14.60
N	139	111	26
Adj$-R^2$	77.14	49.55	72.37
Reject unit root	No	Yes	No

1.3 Stationarity and the Accounting Identity: Theory

We have seen that the accounting identity (6) of Campbell and Shiller depends only on definitions, algebra, and two approximations: deleting the final term of (5) in the limit and the Taylor series expansion (3). In this subsection, we consider what connection (if any) the two approximations have with stationarity tests. Interestingly, the stationarity tests are not well-connected to dropping the final term, but our test without a trend is closely connected to the Taylor series expansion.

To analyze these assumptions under the lens of the stationarity tests, it is useful to write down the following statistical model of the LDPR:

$$\delta_t = z_t + mt \tag{7}$$

where

$$dz_t = -kz_t + sdZ_t, \tag{8}$$

for some constants m, $k \geq 0$, and $s > 0$. Under this structure, the LDPR δ_t can be stationary ($m = 0$ and $k > 0$), nonstationary without a trend ($m = 0$ and $k = 0$), stationary around a trend ($m \neq 0$ and $k > 0$), or nonstationary around a trend ($m \neq 0$ and $k = 0$). Perhaps this is not the most interesting economic model of the LDPR, but it corresponds well with the stationarity tests, and as we vary the parameter values, we obtain processes in both the null and the alternative of both the original Dickey-Fuller test and the augmented Dickey-Fuller test.

The final term can be eliminated if $\mathrm{plim}_{T\uparrow\infty}\rho^{T-t}\delta_T = 0$. For the Taylor approximation, consider the error in (4), which equals

$$\xi_t = \log(1+\exp(\delta_{t+1})) - \log(1+\exp(\delta)) - (1-\rho)(\delta_{t+1}-\delta), \tag{9}$$

as can be verified by subtracting ξ_t from the r.h.s. of (4) and simplying using the definition of κ. This error ξ_t as a function of δ_{t+1} is strictly convex, nonnegative, and has slope and value equal to zero at $\delta_{t+1} = \delta$.

The following theorem says that the final term vanishes for all of the processes we are currently considering. However, whether the Taylor approximation blows up over time depends on the stationarity of the LDPR.

Theorem 1 *Suppose the LDPR δ_t follows the dynamics (7) and (8), for some constants m, $k \geq 0$, and $s > 0$. Further suppose that $\rho \in (0,1)$ (as follows from its definition $\rho \equiv 1/(1+\exp(\delta))$). Then the* *final term in (5) vanishes in probability as $T \uparrow \infty$, i.e.* $\text{plim}_{T\uparrow\infty} \rho^{T-t}\delta_T = 0$. *Furthermore, the Taylor approximation error (9) blows up in the sense that* $\text{plim}_{T\uparrow\infty} \xi_t = \infty$ *unless δ_t is stationary ($m = 0$ and $k > 0$), in which case it does not blow up.*

Proof: From (7) and (8), we have that

$$\delta_T = \delta_0 \exp(-kT) + mT + s\int_{t=0}^{T} e^{-k(T-t)} dZ_t,$$

and since m, $k \geq 0$, and $s > 0$ are all constant, δ_T is normally distributed with mean $\delta_0 \exp(-kT) + mT$ and variance

$$\begin{aligned} V_T &\equiv \int_{t=0}^{T} s^2 exp(-2k(T-t))dt \\ &= \begin{cases} \frac{s^2}{2k}(1-\exp(-2kT)) & \text{for } k < 0 \\ s^2 T & \text{for } k = 0 \end{cases} \end{aligned}$$

which is positive for $T > 0$ and increasing in T. When $k > 0$, $V_\infty \equiv \lim_{T\uparrow\infty} V_T = 1/(2k) < \infty$, while if $k = 0$ then $\lim_{T\uparrow\infty} V_T = \infty$.

When does the final term vanishing as $T \uparrow \infty$? Since δ_T is distributed $N(\delta_0 \exp(-kT) + mT, V_T)$, the final term $\rho^{T-t}\delta_T$ is $N(\rho^{T-t}(\delta_0 \exp(-kT) + mT), \rho^{2(T-t)}V_T)$. Whatever the parameters m, $k > 0$, and $s > 0$, both the mean and the variance of the final term converges to 0 as $T \uparrow \infty$, which implies that the final term converges in L^2 and therefore in probability.

Now consider whether the Taylor approximation blows up. The Taylor approximation error converges in probability to infinity if, for all $K > 0$, $\lim_{T\uparrow\infty} \text{prob}(\xi_T < K) = 0$. Now ξ_t is for all t the same strictly convex differentiable function of δ_{t+1}, where ξ_t achieves a minimum with zero value and zero slope at $\delta_{t+1} = \delta$. Therefore, there exists a finite nondegenerate interval $[\delta^D, \delta^U]$ (where $\delta_{t+1} = \delta^D$ and $\delta_{t+1} = \delta^U$ are the two solutions to $\xi_t = K$) such that $\xi_t \leq K$ if and only if $\delta_{t+1} \in [\delta^D, \delta^U]$. Therefore, $\text{plim}_{T\uparrow\infty} \xi_t = \infty$ if and only if $\lim_{T\uparrow\infty} \text{prob}(\delta_T \in [\delta^D, \delta^U]) = 0$. Now

$$\text{prob}(\delta_T \in [\delta^D, \delta^U]) = N\left(\frac{\delta^U - \delta_0 e^{-kT} - mT}{\sqrt{V_T}}\right) - N\left(\frac{\delta^D - \delta_0 e^{-kT} - mT}{\sqrt{V_T}}\right).$$

For $m = 0$ and $k > 0$, the right-hand side converges to $N(\delta^U/\sqrt{V_\infty}) - N(\delta^U/\sqrt{V_\infty}) > 0$, and the approximation error does not blow up. For $k = 0$, the right-hand side has the same limit as $N(-\sqrt{T}m/s) - N(-\sqrt{T}m/s)$, which is $0 - 0 = 0$ if $m > 0$, $1/2 - 1/2 = 0$ if $m = 0$, and $1 - 1 = 0$ if $m < 0$. Finally, if $k < 0$ and $m \neq 0$, the right-hand side has the same limit as $N((\delta^U - mT)/\sqrt{V_\infty}) - N((\delta^D - mT)/\sqrt{V_\infty})$, which is $0 - 0 = 0$ if $m > 0$ and is $1 - 1 = 0$ if $m < 0$. Therefore, the approximation blows up in all the cases except when $m = 0$ and $k > 0$, in which case it does not blow up. ■

The theorem says that failure to reject nonstationarity in either the augmented Dickey-Fuller test used by Campbell and Shiller or the original Dickey-Fuller test we use is not compelling evidence about the impact of discarding the final term in the limit, even leaving issues of power aside. We do not want to interpret this result as saying that we think the term would disappear for any reasonable assumptions; recall that (7) and (8) are chosen to be in the spirit of the Dickey-Fuller tests, not on economic grounds.

The theorem also says that for this set of processes, our original Dickey-Fuller test should identify correctly when the approximation error blows up over time, since a process with a trend will look like a nonstationary process to the test. Specifically, failure to reject nonstationarity is evidence that the approximation will blow up over time, although there is an issue of power (for example, if mean reversion is weak and is significant only over a much larger time scale than our sample).

Although the results of this subsection assume a particular class of processes defined by (7) and (8), the results are easy to generalize using an ergodic theorem. For example, for the stationary case (without trend) $m = 0$ and $k > 0$, we could assume instead that the LDPR is a stationary ergodic process and the ergodic theorem would give us the necessary asymptotics. For the nonstationary cases (with or without a trend), we could assume instead stationary ergodic changes.

2 Approximation Test

In this section, we test the quality of the LDPR approximation in (5) using the current data. Campbell and Shiller (1988) suggest a vector autoregression (VAR) approach to test (6). They find that the LDPR series is persistent and able to predict both future stock returns and future dividend growth, but the associated R^2s in their tests are small. Unfortunately, the VAR approach suffers several shortcomings. A VAR procedure with a limited number of lags imposes a restriction that does not sufficiently capture the long-term relationship among current dividend-price ratio, future returns and future dividend growth rates. Cochrane (2011) shows that VAR estimates can be biased and significantly different from those in the true linear regressions. In general, the analysis in Campbell and Shiller (1988) does not tell us whether (5) holds empirically.

An improved approach that avoids such shortcomings is to conduct a test of the approximations (5) and (6) with and without the final term. We could perform this test taking all the right-hand side coefficients to be free parameters, but instead, we perform a parsimonious regression of the current LDPR on the sum of weighted future log returns, the sum of weighted log dividend growth rates and the weighted log dividend-price ratio in the last period:

$$
\begin{aligned}
(10)\quad \log\left(\frac{D_t}{P_t}\right) &= \alpha+\beta_1\left(\sum_{s=t+1}^{T}\rho^{s-t-1}(\log(1+R_s))\right)+\beta_2\left(\sum_{s=t+1}^{T}\rho^{s-t-1}\Delta\log(D_s)\right)\\
&\quad +\beta_3\left(\rho^{T-t}\log(\frac{D_T}{P_T})\right)+\varepsilon_t.
\end{aligned}
$$

We choose ρ to be the value implied by setting δ equal to the sample mean LDPR; we show later (in Table 4) that the results are insensitive to this choice. By construction, this regression overcomes the shortcomings of both conventional linear and VAR estimations and is more parsimonious. If the approximation of (5) is accurate, we should expect that the estimated β_1 and β_3 in (10) to have values close to one and β_2 close to minus one. The corresponding R^2 should be close to 100%.

We take $T-t$ to be 30 years, which is reasonably long and gives us 115 overlapping observations (years) for analysis. We use (10) to test (5), and the results are reported in the first regression in Table 2 and suggest that the LDPR approximation is accurate, with an R^2 of 99%. The sum of discounted log returns and the final term both have coefficients close to one, and the discounted log dividend growth rates has a coefficient close to minus one, all as predicted by the theory. All coefficients are statistically significant at the 1% level. This regression uses Newey and West (1987) (hereafter Newey-West) standard errors to adjust for serial correlation (including that due to overlapping observations) and heteroscedasticity.[6] The high R^2 suggests that current LDPR predicts at least one regressor but does

[6]We report the Newey-West standard errors with 4 lags. In an untabulated analysis, we find that the Newey-West standard errors based on 10, 20 or 30 lags are similar.

Table 2: Test of the LDPR Approximation

This table reports the results of a regression testing the accuracy of the approximations of the log dividend-price ratio in (5) and (6) using (10) with and without the final term. The results are based on the annual prices and dividend payments of the S&P 500 index from 1871 to 2015, with $T-t$ set to be 30 years. The associated Newey-West standard errors, computed using four lags, are in parentheses. *** denotes statistical significance at the 1% level.

dependent variable: $\log(D_t/P_t)$	with the final term	without the final term
constant	-2.96^{***}	-3.64^{***}
	(0.05)	(0.08)
$\sum_{s=t+1}^{T}\rho^{s-t-1}(\log(1+R_s))$	0.96^{***}	0.88^{***}
	(0.02)	(0.08)
$\sum_{s=t+1}^{T}\rho^{s-t-1}\Delta\log(D_s)$	-0.99^{***}	-1.08^{***}
	(0.02)	(0.08)
$\rho^{T-t}\log(\frac{D_T}{P_T})$	1.19^{***}	
	(0.05)	
N	115	115
Adj-R^2 (%)	98.91	82.25

not indicate which one(s). Note that we have not adjusted for spurious regression bias (which could be caused by low frequency series on both sides). For now, it suffices to note that the fit is very good. We will correct for spurious regression bias when we conduct predictive regressions.

We use (10) to test the approximation without the final term, which can be viewed as a test of (6) using a finite approximation. The results are reported in the second regression in Table 2 and suggest that leaving out the LDPR 30 years from now has an impact that is neither trivial nor particularly large. The coefficients on the sum of discounted log returns and log dividend growth are still very significant and relatively close to the theoretical values (1 and -1). The R^2 drops significantly by 17%, from 99% to 82%, but is still large enough to suggest that moves in today's LDPR should predict future log returns, log dividend growth, or both. This seems to address the concern Kleidon (1986), Marsh and Merton (1986), and Merton (1987) expressed about dropping the final term, at least in the finite approximation in the current sample. Our failure to find evidence of stationarity of the LDPR suggests that their concern about dropping the final term in the asymptotic result (6) is likely to become more important over time.

3 Predictability

3.1 Predictability Test

The high R^2 (82.25%) in the test of the approximation without the final term (the second regression in Table 2) suggests that the variation of current LDPR is not primarily from the error term ε_t in (10), but rather from either the variation of cumulative future log returns or the variation of future cumulative log dividend growth rates, or both, implying that either future log returns or future dividend growth or both are predictable by current LDPR. This argument motivates the following log return and log dividend predictive analyses:

$$(11) \quad \sum_{s=t+1}^{T} \rho^{s-t-1}(\log(1+R_s)) = c_1 + \lambda_1 \log\left(\frac{D_t}{P_t}\right) + \eta_{T,1}.$$

$$(12)\quad \sum_{s=t+1}^{T} \rho^{s-t-1} \Delta \log(D_s) = c_2 + \lambda_2 \log\left(\frac{D_t}{P_t}\right) + \eta_{T,2}.$$

For completeness, we also test the predictability of LDPR 30 years out:

$$(13)\quad \rho^{T-t} \log\left(\frac{D_T}{P_T}\right) = c_3 + \lambda_3 \log\left(\frac{D_t}{P_t}\right) + \eta_{T,3}.$$

The log return predictive regressions in (11) and (12) are presented in Cochrane (2008), although his regressions are based on a subset of our sample period (1926–2004 rather than 1871–2015). Cochrane's sample period is similar to our second half-sample analysis. A significant coefficient on $\log(D_t/P_t)$ in (11) (resp. in (12) or (13)) suggests that the LDPR predicts future log returns (resp. future dividend growth or final term).

The use of cumulative present values of the predicted variable in future periods has advantages over a conventional predictive specification, in which one-period leading predicted variable is mostly used, in that it can capture the total predictability across horizons. In other words, (11) and (12) capture both short-run and long-run predictabilities (if any).

The finite horizon version of accounting identity in (5) implies that the sum of the absolute coefficients of the LDPR across all three predictive tests (i. e. $\lambda_1 - \lambda_2 + \lambda_3$) should be one if our predictive specification is exact. Over the sample period from 1871 to 2015, Panel A in Table 3 shows that the sum is indeed close to one, $(0.19-(-0.60)+0.17=0.96)$. Future log dividend growth rates are significantly predictable but future log returns are not. The lack of predictability of stock returns supports the argument by Lanne (2002), Valkanov (2003), and Boudoukh, Richardson and Whitelaw (2008) that conventional analysis of long-term predictability of stock returns is spurious.

It may be surprising that predictability of the LDPR 30 years out (Panel A of Table 3) is both economically and statistically significant, leading us to ask what we know now about what will happen 30 years in the future. This view might be compelling if we took the dividend process as exogenous, but as Modigliani and Miller (1958) point out, dividends are somewhat arbitrary (and in their model almost completely arbitrary). Although new information today may be primarily about cash flows in the coming ten years, this cash may go into repurchasing shares rather than paying dividends, meaning that the future increase of dividends may be spread over decades. The predictability of the LDPR out 30 years only depends on (1) the predictability of cash flows over a short horizon, and (2) firm policies implying that it takes a very long time for these increased cash flows to appear in dividends. Consistent with the arguments of Chen, Da, and Priestley (2012), all of these observations are consistent with the smoothness of dividends as noted by Lintner (1956) and others.

Panels B and C in Table 3 present the predictive analysis results for the two equal-long subsample periods. In the first subsample, covering the period from 1871 to 1928 (Panel B), we see that cumulative discounted dividend growth rates are significantly predictable by the current LDPR but cumulative discounted returns are not, consistent with the results over the whole sample period. However, the results for the second subsample from 1929 to 1985 (Panel C) reverse the results for the whole sample. In this subsample, the cumulative discounted dividend growths are not predictable by the LDPR while the cumulative discounted log returns are significantly predictable. The results in the second subperiod are consistent with a similar test in Cochrane (2008, Section 7.2) on a similar sample period.

Given the inconsistent findings across subperiods, we do not want to push any of the results. We have only five non-overlapping observations over the whole sample period and even fewer nonoverlapping observations (2 1/2 instead of 5) over each subsample period, casting doubt on on the the asymptotic properties of the statistical estimates. In particular, the significance of coefficients in Table 3 may tell us more about the size of the test than about the predictability we are trying to test. In Section 1, we could not reject that the LDPR is a non-stationary process. However, both the original approximation and asymptotic properties of the statistical estimates in Table 3 depend on the stationarity of the LDPR. If pressed to take a stand, we are more comfortable with predictability tests over the whole sample period, but it is not very reassuring to say we have five nonoverlapping observations instead of 2 1/2.

We use two adjustments to the inference in the predictability tests: a Newey-West adjustment of the standard errors for heteroskasticity and serial correlation of the errors, and a Stambaugh (1999) adjustment of the coefficients for spurious regression bias (SRB). Serial correlation of the errors is likely to be present given the moving averages used in the estimation, and might be present even without the moving averages. We report the Newey-West standard errors with 4 lags in Table 3. The

Table 3: Predictability Tests

This table reports the results of three regressions testing whether the current log-dividend-price ratio is able to predict the sum of discounted future log returns, the sum of discounted future log dividend growths, or the discounted terminal log dividend-price ratio, all with $T-t$ set to 30 years. Panel A contains the results based on the annual S&P 500 index data from 1871 to 2015. Panels B and C report the results over equal non-overlapping subsamples. The spurious regression bias-adjusted coefficient on $\log(D_t/P_t)$ — labeled SRB in the panels — is estimated following Stambaugh (1999). The associated Newey-West standard errors with four lags are in parentheses. ***, ** and * denote statistical significance at the 1% , 5% and 10% levels, respectively.

dependent variable	constant	$\log(D_t/P_t)$	SRB	Adj-R^2 (%)
Panel A: Whole Sample Period				
$\sum_{s=t+1}^{T}\rho^{s-t-1}(\log(1+R_s))$	1.98^{**}	0.19	0.17	0.43
	(0.76)	(0.24)	(0.24)	
$\sum_{s=t+1}^{T}\rho^{s-t-1}\Delta\log(D_s)$	-1.26^{*}	-0.60^{***}	-0.56^{***}	14.79
	(0.66)	(0.20)	(0.20)	
$\rho^{T-t}\log(\frac{D_T}{P_T})$	-0.19	0.17^{**}	0.17^{**}	17.31
	(0.19)	(0.07)	(0.07)	
Panel B: First Subsample Period (1871–1928)				
$\sum_{s=t+1}^{T}\rho^{s-t-1}(\log(1+R_s))$	1.64^{*}	0.18	0.17	0.66
	(0.86)	(0.29)	(0.29)	
$\sum_{s=t+1}^{T}\rho^{s-t-1}\Delta\log(D_s)$	-2.21^{**}	-0.85^{***}	-0.81^{***}	33.01
	(0.91)	(0.31)	(0.31)	
$\rho^{T-t}\log\left(\frac{D_T}{P_T}\right)$	-0.74^{***}	-0.03	-0.03	0.33
	(0.16)	(0.05)	(0.05)	
Panel C: Second Subsample Period (1929–1985)				
$\sum_{s=t+1}^{T}\rho^{s-t-1}(\log(1+R_s))$	4.00^{***}	0.75^{***}	0.69^{***}	34.15
	(0.39)	(0.13)	(0.13)	
$\sum_{s=t+1}^{T}\rho^{s-t-1}\Delta\log(D_s)$	0.59^{**}	-0.08	-0.07	-1.01
	(0.23)	(0.07)	(0.07)	
$\rho^{T-t}\log(\frac{D_T}{P_T})$	-0.25	0.17^{**}	0.16^{**}	25.09
	(0.20)	(0.07)	(0.07)	

Newey-West standard errors based on 10, 20 or 30 lags are similar. SRB, studied by Granger and Newbold (1974), Stambaugh (1999), and Ferson, Sarkissian, and Simin (2003), is a small sample bias for linear regressions with lagged stochastic regressors. Stambaugh shows that this bias is pronounced in the predictive coefficient but not in the standard error of the predictive coefficient or the R^2. By assuming $\log\left(\frac{D_t}{P_t}\right)$ to be a first-order autoregressive process as $\log\left(\frac{D_t}{P_t}\right) = c + \tau \log\left(\frac{D_{t-1}}{P_{t-1}}\right) + \nu_t$, Stambaugh shows that the magnitude of SRB in the predictive coefficient in (11) equals $-\frac{\sigma_{\eta\nu}}{\sigma_\nu^2}(\frac{1+3\tau}{N})$, where $\sigma_{\eta\nu}$ is the covariance of η_t and ν_t, σ_ν^2 the variance of ν_t, and N the number of observations of the sample. Interestingly, the Stambaugh adjustment does not seem very important for any of these tests.

3.2 Why is Predictability Much Weaker than the Approximation?

It is worth exploring why the coefficient on stock returns in Table 2 is consistently close to one and statistically significant, while the predictor coefficient (of the LDPR) in the stock return predictability test in Table 3 is small and insignificant. After all, in a univariate regression, the standard algebra implies the R^2 is unchanged if you interchange dependent and independent variables. Furthermore, since the univariate and multivariate regression coefficients are the same when the independent variables are uncorrelated, we know the explanation must come from nonzero correlation between the independent variables. The explanation lies in the high correlation between the weighted sum of future stock returns and the weighted sum of future dividend growth rates rather than any information in the LDPR about future stock returns. We can think of the significant coefficient on future log returns in Table 2 as a correction to the future log dividends (by removing common noise) rather than any correlation between today's LDPR and future log returns. To illustrate this argument, let us start with the assumption that the weighted sum of future log returns is just equal to some noise Z that is uncorrelated with the LDPR:

$$(14) \quad \sum_{s=t+1}^{T} \rho^{s-t-1}(\log(1+R_s)) \approx Z_t.$$

Now, use this expression and (5) without the final term and ignoring the constant (which does not affect variance and covariance) to approximate the weighted sum of future dividend growth rates:

$$(15)\quad \sum_{s=t+1}^{T} \rho^{s-t-1} \Delta \log(D_s) \approx -\log(\frac{D_t}{P_t}) + Z_t.$$

Then the covariance matrix between the LDPR and sum of discounted future dividend growth rates is:

$$(16)\quad \operatorname{var}\left(\log\left(\frac{D_t}{P_t}\right), \sum_{s=t+1}^{T} \rho^{s-t-1} \Delta \log(D_s)\right) = \begin{pmatrix} \sigma_\delta^2 & -\sigma_\delta^2 \\ -\sigma_\delta^2 & \sigma_\delta^2 + \sigma_Z^2 \end{pmatrix}.$$

When we regress current LDPR on the sum of discounted future dividend growth rates alone, the regression coefficient (ignoring estimation error) is $\beta = -\sigma_\delta^2/(\sigma_\delta^2 + \sigma_Z^2)$. The coefficient is biased towards zero compared to what it would be without the noise Z_t (the standard errors-in-variables result), and when the noise σ_Z^2 in stock returns is large compared to σ_δ^2 (which is consistent with the data), then the bias is large. However, if we run the LDPR on both the weighted sum of future log returns and the weighted sum of future log dividend growth rates, we obtain a coefficient of -1 on the dividend sum and a weight of 1 on the return sum (a perfect fit given our approximations (14) and (15)). Including returns allows the regression to cancel the noise in the dividend sum.

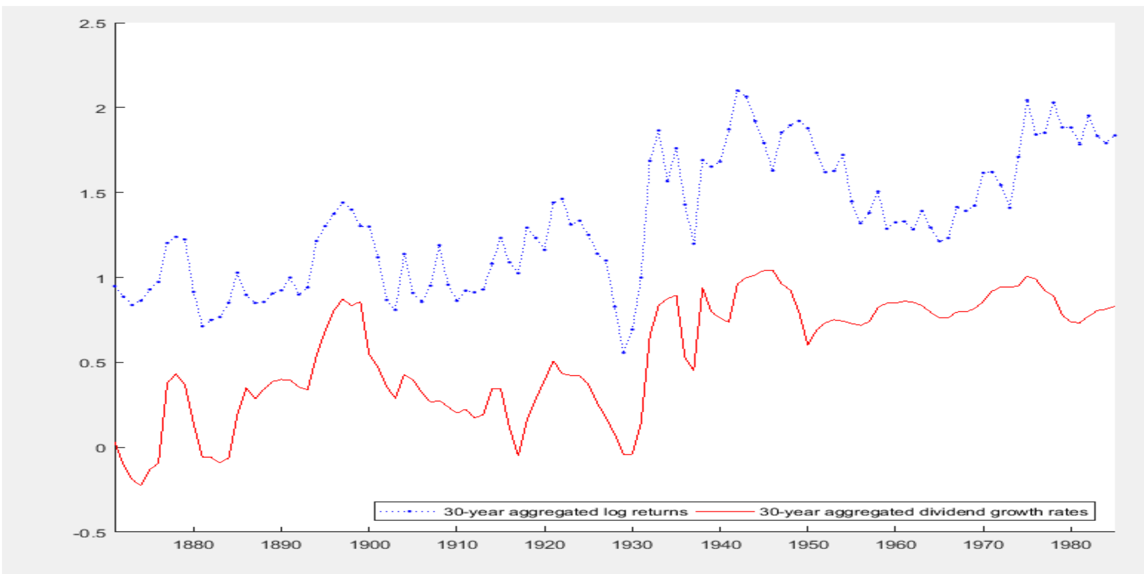
To confirm the common factor in returns and dividends, consider this both theoretically and empirically. Given (14) and (15), we have that:

$$(17)\quad \operatorname{var}\left(\sum_{s=t+1}^{T} \rho^{s-t-1}(\log(1+R_s)), \sum_{s=t+1}^{T} \rho^{s-t-1} \Delta \log(D_s)\right) = \begin{pmatrix} \sigma_\delta^2 & \sigma_Z^2 \\ \sigma_Z^2 & \sigma_\delta^2 + \sigma_Z^2 \end{pmatrix}.$$

$$(18)\quad \operatorname{corr}\left(\sum_{s=t+1}^{T} \rho^{s-t-1}(\log(1+R_s)), \sum_{s=t+1}^{T} \rho^{s-t-1} \Delta \log(D_s)\right) = \frac{\sigma_Z^2}{\sigma_Z^2 + \sigma_\delta^2}.$$

When σ_Z^2 is large compared to σ_δ^2, the correlation between the return sum and the dividend sum should be large. Figure 2 shows that the evolutions of cumulative log returns and log dividend growth rates are closely correlated. In fact, the correlation between the log return and the log dividend growth is 0.63, and the correlation between $(\sum_{s=t+1}^{T}\rho^{s-t-1}(\log(1+R_s)))$ and $(\sum_{s=t+1}^{T}\rho^{s-t-1}\Delta\log(D_s))$ is 0.84. Moreover, the standard deviations of the cumulative returns and log dividend growth rates are respectively 38.1% and 34.3% which implies that the sum of the two terms' variances is as high as 37.2%, or 61.0% in terms of standard deviation, while the standard deviation of the log dividend-price ratio over the same period is 23.6%. This relationship is also observed by Ferson, Sarkissian and Simin (2003) and Valkanov (2003) with simulated data.

Figure 2: Cumulative Discounted Log Returns and Log Dividend Growth Rates



One concern about the results in this section is that our sample includes only about five non-overlapping observations of the weighted sums. Although it is impressive that the estimates (with Newey-West and Stambaugh corrections) are significant in spite of this, this puts a lot of demand on the Newey-West adjustment and we are far from its asymptotic justification.

4 Alternative Expansion Points

In the log linear approximation in (3), we approximate $\log(1+\exp(\delta_{t+1}))$ around some value δ using a first-order Taylor expansion. After telescoping this expression and dropping the final term, we arrive at (6) or its finite horizon version (5). In this section, we examine the impact of the approximation error and dropping the final term on the approximation and predictability tests, with a special focus on how the error depends on the expansion point δ. This is especially important because the LDPR may not be stationary.

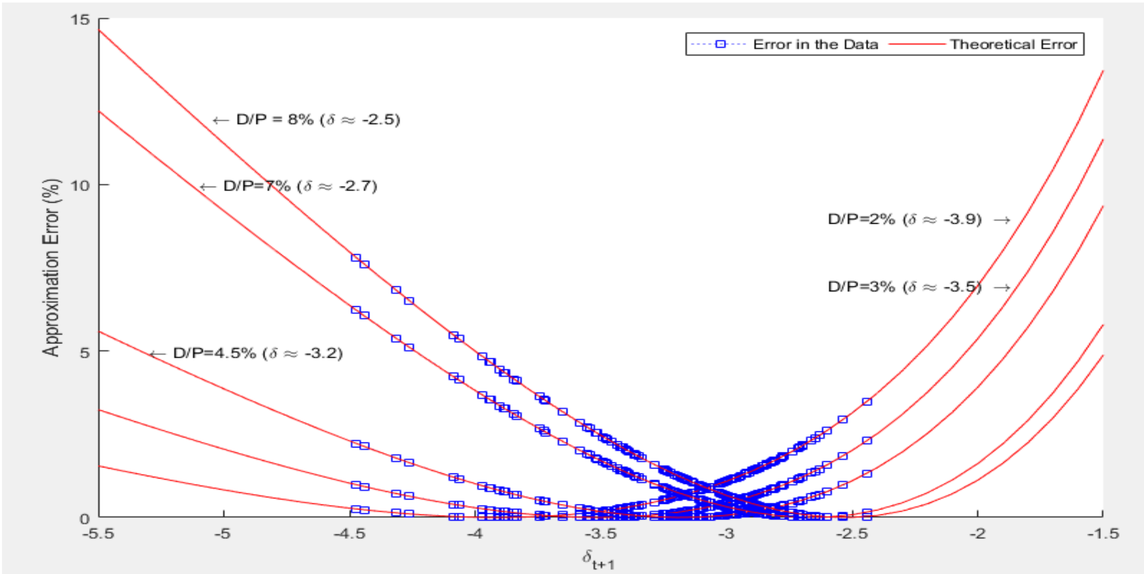
4.1 Single Period Taylor Approximation Error

We first look at the possible magnitude of the approximation error in a single period. Recall that the accounting identity (6) was derived by taking a limit of the discrete version (5) and discarding the final term. In turn, (5) was derived from telescoping the single-period approximate LDPR expression (4). The error in (4) comes from the Taylor expansion and we wrote it down explicitly in (9).

The approximation error ξ_t is zero when $\delta_{t+1} = \delta$ and positive everywhere else. Given δ, ξ_t is a convex function of δ_{t+1} that gets more positive as δ_{t+1} moves away from δ.[7] To see how ξ_t is influenced by the selection of expanding point δ as well as the LDPR (δ_{t+1}), we plot the relationship of ξ_t as a function of δ_{t+1} for different δ, theoretically as well as using the data. The data land exactly on the theoretical curves because we have an exactly expression for the error. Specifically, we consider expanding δ_{t+1} around its sample mean, which is around -3.18 (corresponding to a dividend-price ratio of 4.17%), as well as four alternative expanding points to take into account the declining trend in the dividend-price ratio: 2%, 3%, 7% and 8%, which correspond to LDPRs of -3.91, -3.51, -2.66 and -2.53. The results are illustrated in Figure 3. Regardless of expanding points, Figure 3 shows that the approximation error ξ_t is close to zero when the LDPR is close to the expanding point. However, Figure 3 also shows that ξ_t is far from zero if the LDPR deviates too much from the expanding point. Consider -3.88 (or a corresponding dividend-price of 2.06%), the LDPR in 2008, as an example. The approximation error ξ_t in this year is almost zero when the LDPR is expanded around -3.9, and becomes 0.8% for expansion around its sample mean (-3.18), and 4.45% for expansion around -2.53 (corresponding to a dividend-price ratio of 8%).

[7] When δ_{t+1} is far from δ, ξ_t is roughly affine, with slope $\rho - 1$ if $\delta_{t+1} \ll \delta$ and slope ρ when $\delta_{t+1} \gg \delta$.

Figure 3: Return Approximation Error



4.2 Multiple-Period Approximation Error and Dropping the Final Term

In the case of approximation over multiple periods as shown in LDPR equation (5) or (6), we define the approximation error as the following:

$$(19) \quad \zeta_t = -\frac{\kappa}{1-\rho} + \sum_{s=t+1}^{T} \rho^{s-t-1}(\log(1+R_s) - \Delta\log(D_s)) - \log(\frac{D_t}{P_t}),$$

where $\rho \equiv 1/(1+\exp(\delta))$ and $\kappa \equiv \log(1+\exp(\delta)) - (1-\rho)\delta$.

(19) and (6) suggest that ζ_t contains two sources of errors: the accumulation of the approximation errors (9) in (4), and the omission of the final term $\rho^{T-t}\log(\frac{D_T}{P_T})$ in (5). When δ approaches negative infinity and ρ approaches one, the weight on the final term, ρ^{T-t}, tends to be one, which may lead to a large ζ_t. Table 4 gives us some idea how much deterioration we can expect in the approximation if the LDPR continues to wander away from the past values in the next one or two hundred years. Overall, the approximations still seem useful. The most striking problem is that as δ falls, omitting the final term has more and more impact. In Table 5, we see that changing the expansion point δ has little impact on the predictability regressions in our current sample.

Table 4: LDPR Approximation Test: Alternative Expansion Points

This table reports the sensitivity to the expansion point δ of the tests of the approximation tests in Table 2. The regression is specified as: $\log(D_t/P_t) = \alpha + \beta_1(\sum_{s=t+1}^{T}\rho^{s-t-1}(\log(1+R_s))) + \beta_2(\sum_{s=t+1}^{T}\rho^{s-t-1}\Delta\log(D_s)) + \beta_3(\rho^{T-t}\log(D_T/P_T)) + \varepsilon_t$, with and without the final term. The results are based on annual data from the S&P 500 index from 1871 to 2015, with $T-t$ set to be 30 years. The associated Newey-West standard errors with four lags are in parentheses. *** denotes statistical significance at the 1% level.

Model	constant	$\sum_{s=t+1}^{T}\rho^{s-t-1}(\log(1+R_s))$	$\sum_{s=t+1}^{T}\rho^{s-t-1}\Delta\log(D_s)$	$\rho^{T-t}\log(\frac{D_T}{P_T})$	Adj-R^2 (%)
Panel A: Expanding point: $\delta = -3.91$ ($\rho \approx 0.98$, $D/P = 2\%$)					
w/final	-2.71^{***}	0.92^{***}	-0.87^{***}	0.82^{***}	93.79
term	(0.08)	(0.04)	(0.04)	(0.05)	
w/o final	-3.60^{***}	0.62^{***}	-0.80^{***}		55.62
term	(0.07)	(0.08)	(0.08)		
Panel B: Expanding point: $\delta = -3.51$ ($\rho \approx 0.97$, $D/P = 3\%$)					
w/final	-2.82^{***}	0.95^{***}	-0.93^{***}	0.92^{***}	98.12
term	(0.04)	(0.02)	(0.02)	(0.05)	
w/o final	-3.66^{***}	0.74^{***}	-0.93^{***}		67.37
term	(0.09)	(0.07)	(0.08)		
Panel C: Expanding point: $\delta = -2.66$ ($\rho \approx 0.94$, $D/P = 7\%$)					
w/final	-3.00^{***}	0.95^{***}	-1.02^{***}	1.53^{***}	96.17
term	(0.08)	(0.04)	(0.04)	(0.14)	
w/o final	-3.58^{***}	0.93^{***}	-1.13^{***}		85.67
term	(0.07)	(0.09)	(0.08)		
Panel D: Expanding point: $\delta = -2.53$ ($\rho \approx 0.93$, $D/P = 8\%$)					
w/final	-3.02^{***}	0.95^{***}	-1.03^{***}	1.80^{***}	94.36
term	(0.09)	(0.05)	(0.05)	(0.22)	
w/o final	-3.54^{***}	0.94^{***}	-1.14^{***}		85.83
term	(0.06)	(0.09)	(0.08)		

Table 5: Predictability Test: Alternative Expansion Points

This table explores the robustness of the results of three regressions in Table 3 to the choice of expansion point for the Taylor approximation. The results are based on annual data from the S&P 500 index from 1871 to 2015. The associated Newey-West standard errors with four lags are in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

dependent variable	constant	$\log(D_t/P_t)$	SRB	Adj-R^2 (%)
Panel A: Expanding point: $\delta = -3.91$ ($\rho \approx 0.98$, D/P =2%)				
$\sum_{s=t+1}^{T} \rho^{s-t-1}(\log(1+R_s))$	1.78*	−0.08	−0.07	−0.75
	(0.98)	(0.32)	(0.32)	
$\sum_{s=t+1}^{T} \rho^{s-t-1}\Delta\log(D_s)$	−1.48*	−0.76***	−0.70***	14.83
	(0.86)	(0.27)	(0.27)	
$\rho^{T-t}\log\left(\frac{D_T}{P_T}\right)$	−0.48	0.43**	0.41**	17.31
	(0.47)	(0.17)	(0.17)	
Panel B: Expanding point: $\delta = -3.51$ ($\rho \approx 0.97$, D/P=3%)				
$\sum_{s=t+1}^{T} \rho^{s-t-1}(\log(1+R_s))$	1.87**	−0.08	−0.08	−0.87
	(0.90)	(0.29)	(0.29)	
$\sum_{s=t+1}^{T} \rho^{s-t-1}\Delta\log(D_s)$	−1.42*	−0.71***	−0.65***	15.14
	(0.79)	(0.25)	(0.25)	
$\rho^{T-t}\log(\frac{D_T}{P_T})$	−0.36	0.32***	0.30***	17.31
	(0.35)	(0.12)	(0.12)	
Panel C: Expanding point: $\delta = -2.66$ ($\rho \approx 0.94$, D/P=7%)				
$\sum_{s=t+1}^{T} \rho^{s-t-1}(\log(1+R_s))$	2.01*	0.28	0.26	2.36
	(0.67)	(0.21)	(0.21)	
$\sum_{s=t+1}^{T} \rho^{s-t-1}\Delta\log(D_s)$	−1.14*	−0.53***	−0.51***	13.80
	(0.58)	(0.18)	(0.18)	
$\rho^{T-t}\log\left(\frac{D_T}{P_T}\right)$	−0.14***	0.10***	0.10***	17.31
	(0.18)	(0.04)	(0.04)	
Panel D: Expanding point: $\delta = -2.53$ ($\rho \approx 0.93$, D/P=8%)				
$\sum_{s=t+1}^{T} \rho^{s-t-1}(\log(1+R_s))$	2.01***	0.31	0.28	3.48
	(0.64)	(0.20)	(0.20)	
$\sum_{s=t+1}^{T} \rho^{s-t-1}\Delta\log(D_s)$	−1.08**	−0.49***	−0.46***	13.14
	(0.54)	(0.17)	(0.17)	
$\rho^{T-t}\log\left(\frac{D_T}{P_T}\right)$	−0.09	0.08***	0.08***	17.31
	(0.08)	(0.03)	(0.03)	

5 Conclusion

Whether stock returns are predictable is an important and challenging question for both academia and industry. Campbell and Shiller (1988) argue, based on accounting definitions and some approximations, that the log dividend-price ratio must predict future log returns, future log dividend growth, or both. However, in past literature neither prediction has been found to be economically or statistically significant, creating a well-known puzzle. We check each step of Campbell and Shiller's argument, from the accounting definition through the approximation to the statistical tests. We find that Campbell and Shiller's test of existence of the long-term average log dividend-price ratio is fatally flawed. This is important because existence of the long-term mean is a critical assumption in deriving their main result. A correct test reverses the Campbell-Shiller result and cannot reject the null hypothesis that the long-term mean does not exist.

If the long-term mean log dividend-price ratio does not exist, Campbell and Shiller's "accounting identity" will not hold, and the approximation in Campbell and Shiller's theory can be expected to get worse and worse over time. However, a truncated version of the LDPR approximation may still be useful in our existing sample if we replace the long-term mean in the theory by a reasonable value, such as the mean in the sample we do have. We find that the approximation does work reasonably well in our sample and is not too sensitive to the choice of log dividend-price ratio to expand around. Our findings show that the source of the failure to find a significant relationship arises from a mismatch between the small lags in the traditional tests and the many terms in the theoretical expression. When we conduct a test closer to the theoretical expression, with appropriate correction for serial correlation due to overlapping data, possible heteroscedasticity, and spurious regression bias, we find that future log dividend growth is significantly predictable but future returns are not, which seems to resolve the puzzle.

While this is the best conclusion given the data currently available, this result does not seem to be robust for several reasons. For one, there are only a few (about five) non-overlapping observations of the truncated identity for the whole period, so we are asking a lot of the Newey-West adjustment. Also, the results are different over two subperiods, which calls into question our assumption of a stable relationship over time and reliance on asymptotic properties of the statistical estimates. Perhaps we should not expect stability of the dividend process over time, since, according to Modigliani and Miller, dividends are irrelevant. Even if Modigliani-Miller's arguments should not be taken too literally, they do mean that seemingly small changes in taxes or transaction costs can have a big impact on dividend policy and affect the time series properties of log returns, log dividend growth, and log dividend-price ratios. Possible nonstationarity of the LDPR, which we cannot reject for the whole sample or for the second half of the sample, is a serious problem for the theory because the Taylor series approximation worsens as the range of the log dividend-price ratio increases. For these reasons, it seems that the limitations of this approach may be intrinsic, and the accounting identity may never tell us much about return predictability, even as we collect more and more data.

References

Boudoukh, Jacob, Matthew Richardson, and Robert F. Whitelaw. 2008. The Myth of Long-Horizon Predictability. The Review of Financial Studies 21 (4): 1577-1605.

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.

Chen, Long. 2009. On the Reversal of Return and Dividend Growth Predictability: A Tale of Two Periods. Journal of Financial Economics 92: 128-151.

Chen, Long, Zhi Da, and Richard Priestley. 2012. Dividend Smoothing and Predictability. Management Science 58 (10): 1834-1853.

Cochrane, John H. 2011. Presidential Address: Discount Rates. The Journal of Finance 66 (4): 1047-1108.

Cochrane, John H. 2008. The Dog That Did Not Bark: A Defense of Return Predictability. The Review of Financial Studies 21 (4):1533-1575.

Cowles, Alfred. 1939. *Common Stock Indexes* (2nd edition), Principal Press, Bloomington, Indiana.

Dickey, David A., and Wayne A., Fuller. 1981. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. Econometrica 49 (4): 1057-1072.

Edwards, Blake (Producer & Director). 1976. *The Pink Panther Strikes Again*. United Kingdom: Amjo Productions Limited.

Ferson, Wayne E., Sergei Sarkissian, and Timothy T. Simin. 2003. Spurious Regressions in Financial Economics? The Journal of Finance 58 (4): 1393-1413.

Granger, C.W.J., and P. Newbold. 1974. Spurious Regressions in Econometrics. Journal of Econometrics 2 :111-120.

Kleidon, Allan W. 1986. Variance Bounds Tests and Stock Price Valuation Models. The Journal of Political Economy 94 (5), 953-1001.

Lanne, Markku. 2002. Testing the Predictability of Stock Returns. The Review of Economics and Statistics 84 (3): 407-415.

Lintner, John. 1956. Distribution of Incomes of Corporations Among Dividends, Retained Earnings, and Taxes. The American Economic Review 46 (2): 97-113.

Marsh, Terry A., and Robert C. Merton. 1986. Dividend Variability and Variance Bounds Tests for the Rationality of Stock Market Prices. The American Economic Review 76 (3): 483-498.

Merton, Robert C. 1987. On the Current State of the Stock Market Rationaity Hypothesis, in Dornbusch, Rudiger, Stanley Fischer and John Bossons, Macroeconomics and Finance: Essays on Honor of Franco Modigliani, the MIT Press, 1987.

Modigliani, Franco, and Merton H. Miller. 1958. The Cost of Capital, Corporation Finance and the Theory of Investment. The American Economic Review 48(3): 261-297.

Newey, Whitney K., and Kenneth D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica, 55 (3): 703-708.

Phillips, Peter, C.B., and Pierre Perron. 1988. Testing for a unit root in time series regression. Biometrika 75 (2): 335-346.

Shiller, Robert J. 1981. Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? The American Economic Review 71 (3):421-436.

Stambaugh, Robert F. 1999. Predictive Regressions. Journal of Financial Economics 52: 375-421.

Valkanov, Rossen. 2003. Long-Horizon Regressions: Theoretical Results and Applications. Journal of Financial Economics 68: 201-232.