The Modified Dividend-Price Ratio

Ioannis M. Neokosmidis* and Vassilis Polimenis†‡

Abstract

After failing to reject the null of a unit root in the classical dividend-price ratio (dp), we show that a cointegrating relationship between dividends and prices exists. We then define the modified dividend-price ratio (mdp), as the long run trend deviation between log dividends and prices. Using S&P 500 data for the period 1926 to 2009, we show that mdp provides substantially improved forecasting results over the classical dp ratio. Actually, when both ratios are present in a multivariate setting, the classical dividend price ratio fails to bring any extra information about future returns. Thus, in the presence of the modified ratio the only thing left for the classical dividend-price ratio to predict is its forward looking value. Out of sample, while the classical dp ratio cannot outperform the “mean” benchmark for any useful horizon, an investor who employs the modified dp ratio will do better in forecasting 5- and 7-year returns with an $R^2_{OS}$ between 52-54%. Further, with 1-year and 3-year $R^2_{OS}$ of 12% and 39% respectively, mdp addresses a major weakness in dp, namely its presumed inability in revealing business cycle variation in expected returns and equity risk-premia. We show that the gain of our modified ratio in forecasting returns is mainly due to its enhanced ability to forecast their risk free component, and argue that it can be considered as a de-noising of the classical ratio.

* Department of Economics, School of Law and Economics, Aristotle University of Thessaloniki. inekosm@econ.auth.gr
† Division of Business Administration, School of Law and Economics, Aristotle University of Thessaloniki. polimenis@econ.auth.gr
‡ Corresponding author.
The ability to forecast returns can easily be regarded as the most significant question for asset allocation, and one of the most important issues in the entire financial economics. After an early period where return predictability was approached with simplistic or brute-force methods, during the late 80s and early 90s, the literature proposed more sophisticated and smart ways to measure the ability of valuation ratios and other statistics in predicting aggregate stock returns. Motivated mainly by practitioner views, starting with the classic Graham and Dodd (1934), that high valuation ratios should carry positive information about future returns, Fama and French (1988) find that economically substantial return predictability at a long horizon exists. Long-horizon forecasts are the mechanical result of short horizon same-direction forecastability combined with a highly persistent forecasting variable. The persistence of a predictor variable leads to increased slope coefficients for longer horizons.

This analysis is lately aided by a clever tool, the Campbell-Shiller (CS) relationship, which ties the current dividend-price ratio to future returns and growth, and analytically describes and quantifies what was known long ago: if high stock prices in an efficient market do not predict large increases in dividends (i.e. a strong growth) then they either forecast poor future returns, or are part of a stock market bubble.

The dividend price ratio has special importance as a forecasting variable due to the straightforward participation of the dividend yield in return formation, its place in the CS identity and its highly persistent dynamics which provides predictability in long forecasting horizons. At a practitioner level, it is not an over-statement to say that empirical analysis of
the behaviour of the dividend yield has transformed the way in which analysts perceive the
evolution of financial markets.

provides a new way to understand the intricate relationship between return forecastability and
the variability of financial ratios such as the dividend-price ($dp_t$) ratio. More specifically, for
long horizons, long-run return and/or dividend growth predictability have to coincide with the
variability of financial ratios. Thus, the old belief in the stability of expected stock returns has
now been replaced with a new philosophy, summarized in the following three stylized facts:

- The dividend yield is a highly persistent variable that has the ability to predict
  future market returns when they are measured in sufficiently large time horizons

- It is not possible to predict the changes in dividends.

- The largest part of the variation in dividend yields is due to their ability to predict
  future market returns.

Cochrane (2008) defends return predictability by arguing that when testing a null hypothesis
where returns are not forecastable, we are also testing a hypothesis that dividend growth is
forecastable, and thus lack of growth forecastability must imply that the dividend price ratio
is actually forecasting long-run returns.

While some of these facts are hard findings embedded in the data, some are softer economic
interpretations based on two main assumptions: the assumed ability to recursively extend the
Campbell-Shiller relation to infinity, and the stationarity of dividend yields.
But there are some major problems with the predicting performance of the dividend yield. Firstly, its weak performance in predicting returns and risk premia outside the sample used to determine the slope coefficient. Secondly, an inability in revealing high to medium frequency variation (i.e. business cycles) in expected returns and equity risk-premia. Over shorter than 7-10 year horizons, dividend-price ratios mainly predict themselves (Goyal and Welch, 2003). Cochrane (2008) reports that even when artificially setting a null in which all dividend ratio variability is driven by returns’ forecasts, given the small short horizon predictability no serious gains can be achieved for real-time market timing.

The poor Out-of-Sample (OS) performance of dividend-price ratio is exhibited in Goyal and Welch (2003), Welch and Goyal (2008) and Campbell and Tompson (2008) who summarize the power in forecasting monthly and annual returns and equity premia for a pool of common financial and accounting variables. Campbell and Tompson introduce a coefficient of determination $R_{OS}^2$, which measures the effect of OS performance of a predictor variable against the mean benchmark. Firstly, they report that the earnings yield and smoothed earnings yield are the only financial ratios which have a good IS and OS performance. By imposing restrictions on slope coefficients, they further enhance OS performance for other variables that would not outperform the mean benchmark otherwise. Rapach et.al (2010) argue that the univariate performance of common predictors, can be enhanced if they are combined in a multivariate setting.

The unquestionable economic requirement that prices cannot be far from fundamentals for too long has been interpreted in a strict sense that the dividend-price ratio is stationary either in the full sample or at least in specific subsamples. Thus, most contemporary literature de
facto assumes that the classical dividend-price ratio \((dp_t)\) is a stationary process. Yet, the majority of econometric studies on return predictability, cannot reject statistically (if not economically) the hypothesis of the presence of a unit root in dividend-price ratio (Goyal and Welch, 2003; Lettau and Van Nieuwerburgh, 2008; Lettau and Ludvigson, 2005 among others). Lettau and Van Nieuwerburgh (2008) allow the presence of nonstationary components in their state space representation. Cochrane (2008) discusses the case of a unit root in the dividend-price ratio without rejecting it statistically, and briefly mentions that the nonstationarity of dividend yield does not violate the CS identity as such, but would cause trouble since that approximation is only accurate near the expansion point.

This paper modifies the dividend-price ratio by considering the true cointegration vector between dividends and prices. More specifically, we move away from the economic requirement that classical dividend-price ratios \((dp_t)\) should be stationary, and we view the classical dividend-price ratio only as a possible cointegration relationship \((d_t - p_t)\) with a cointegration vector of \([1,-1]\) that is then tested in the data. Having failed to detect a purely stationary behaviour in \(dp\), and also explicitly rejecting the \([1,-1]\) as a cointegrating vector, we show that econometrically there is no evidence to reject the null of a unit root in \(dp\), i.e. that \(dp\) is of the form \(dp = I(0) + I(1)\).

Since we relax the stationarity assumption for the classical \(dp_t\), the next logical step, which retains a fundamentals’ based asset pricing philosophy, is to assume a deterministic long run relation between dividends and prices; i.e. assume a cointegration vector of the form \(d_t = \alpha + \beta p_t\), and then allow the data to reveal the “true” cointegration vector \([1,-\beta]\). The stationary trend deviation among dividends and prices, \(d_t - \beta p_t\), is defined as the modified
dividend-price ratio ($mdp_t$), and is shown here to econometrically provide an impressively better performance in terms of forecasting future returns both in-sample and out of sample.

Our paper is related to the literature that “corrects” the dividend-price ratio (Boudoukh et al, 2007 and Boudoukh et al, 2008; Lettau and Van Nieuwerburgh, 2008; among others) in order to achieve a better $R^2$ performance. But in those cases, higher in-sample $R^2$ does not produce a higher out-of-sample performance, and, much like the classical dp ratio, those new measures cannot outperform the “mean” benchmark for any useful horizon. On the contrary, an investor who employs the $mdp$ ratio will improve Out-of-Sample forecasting of 5- and 7-year returns with an $R^2_{5S}$ between 52-54%. Further, with 1-year and 3-year $R^2_{5S}$ of 12% and 39% respectively, $mdp$ addresses a major weakness in dp, namely its presumed inability in revealing business cycle variation in expected returns and equity risk-premia.

Based on the infinite version of the Campbell-Shiller identity, $dp_t \approx \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} - \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}$, it has become clear that a variable that improves our predictions about future returns, without helping us predict long-run dividend growth, should signal a change in the term structure of expected returns. But, in a changing dividend-price ratio world, the CS identity is only a good approximation, and taking limits to infinity of an approximation may be mathematically tricky. If we consider that $dp_t$ is changing, or rather persistent, enough, so that we are not allowed to take limits to infinity, the story behind the enhanced performance of $mdp_t$ is that it filters out the non-informational unit root component in $dp_t$. True information does not come from the $l(1)$ “level”-type variables, but from the $l(0)$ changes to integrated levels. Thus, in a variable like the classical $dp_t$ that includes both $l(0)$ and a small
An $I(1)$ embedded component, $dp_t = mdp_t + (\beta - 1)p_t$, true information is the by-construction stationary $mdp_t$ component\(^1\).

One significant new insight of this paper is that part of the high persistence\(^2\) $\phi=0.93$ of a non-stationary $dp_t$ is due to the small embedded unit root, but unlike the $\phi=0.70$ “useful” persistence in $mdp_t$, this extra persistence carries no real predicting power. Thus, the true forecasting horizon is determined by the lower $mdp_t$ persistence. The artificially longer horizon for $dp$, that one gets by mechanically extending short period $dp$ predictability into the distant future, is a statistical artifact of the non-stationary noise embedded in $dp$ and of no real forecasting value.

The surprisingly strong performance gain of the modified $dp$ ratio, versus the classical ratio, in predicting future returns is considerably toned down when using $mdp$ to predict equity premia. This implies that the enhanced performance of $mdp$ comes mainly from predicting the risk free component of returns. Such a forecasting behaviour of the modified ratio could be consistent with a clientele explanation of dividend policy; if companies that consistently pay dividends attract a certain type of yield seeking investor, they get away with low dividend yields when such low payouts coincide with low current and future risk free yields. In such low-yield states of the economy, income seeking investors will not allocate their portfolios out of low dividend yield stocks because they have nowhere to go. One way to understand, why the classical ratios don’t share such foresting ability with their modified counterparts, is if we view the modified ratio as a de-noised $dp$. Even though this yield

---

\(^1\) This is similar to the fact that not much information is provided when we are told that a stock price yesterday closed at $103$. It is the fact that it went up by 3% (i.e. opened at $100$) that provides information.

\(^2\) $\phi=.93$ with dividend reinvestment and .86 with simple dividend summation.
information is embedded in the classical ratio as well (as its stationary part), the $I(1)$ “noise” component needs to be removed first. A strong positive correlation of the modified ratio with the risk free returns equal to 0.42 completely changes the picture from the misleading almost orthogonality of the classical $dp$ with yields. Our $dp$ modification then can be considered economically as a de-noising of the classical dividend-price ratio.

The added benefit of $mdp$ is that the fairly high, but substantially lower than one, persistence of $mdp$ also avoids known inference problems of the very persistent dividend yield\(^3\). This effect, of understating a p-value, gets stronger when the regressor is a “near non-stationary” variable and when the innovations among the returns and predictor variable are correlated.

The rest of the paper is organized as follows. In the next section we discuss the data and present the two methodologies and the results of estimating the trend deviation of a long-run relationship between dividends and prices, and form the modified dividend price ratios ($mdp$ and $mdp'$). In Section II, we further elaborate on the economics of the non-stationarity of the classical dividend price ratio. We then move on in Section III to present in-sample predictability for the modified ratios and compare the results with the classical ratios. In section IV we use the Campbell-Shiller identity to perform multivariate analysis where both $dp$ and $mdp$ are present on the right hand side. Finally, in Section V the robust out-of-sample performance gains of the modified ratios are documented. Section VI concludes.

---

\(^3\) Stambaugh (1999) argues that the assumed return predictability may not be as obvious as it seems, since there is a lot of finite sample bias even for the one period return forecasting coefficient when a highly persistent forecasting regressor such as $dp$ is used. Under the zero return forecastability hypothesis, he reports that the probability to see something as high as in his tables by chance, considering two subsamples of 70 years and 20 years respectively, is about 17% and 15% rather than 6% and 2% respectively.
I. Data Analysis and Construction of the Modified Ratio

In this section we describe the two available econometric techniques (DLS and ADL) in forming the modified dividend price ratios \((mdp_t\) and \(mdp'_t\) respectively). High quality return data for the S&P 500 index, with and without dividends, are available from CRSP since 1926. In the appendix, we show how we use (total and ex-dividend) monthly returns’ data for the S&P 500 in order to formulate annual dividend and price level series, and the classical dividend-price ratios. Our sample\(^4\) spans the most recent 84 year period that ranges from January of 1926 to December of 2009.

When constructing the \(dp\) ratio, we need to employ an annual horizon in order to cancel the strong dividend seasonality. Depending on how one forms annual dividends at the end of month \(t\), from the 12 preceding monthly dividends, the classical dividend price ratio may be calculated with two different accounting methodologies: The straightforward dividend-price ratio with annual dividends formed as sum of twelve monthly dividends \((DP_t)\), and the one where annual dividends are formed by immediately reinvesting monthly dividends as they arrive at the end-of-month prices \((YP_t)\). The annual sum construction for the \(dp\) is a common composition employed by the literature, and is less volatile, representing more purely dividend policy decisions of firms. The latter technique may be more appropriate from a conceptual point of view, but transfers to dividends some of the market volatility for the year, and may thus be of less value for practical purposes as it distorts true cash made available to shareholders during the period. The Appendix has an analytical description of data

\(^4\) The data are from the Goyal and Welch database, available at http://www.hec.unil.ch/agoyal
construction. The reinvestment construction echoes Cochrane (2011) in an effort to retain the Campbell-Shiller return identity for an annual horizon.

Having determined the unit root nature for all \((d, y, p)\) series, we may proceed to test for the existence of deterministic cointegration between both types of dividends \((d \text{ or } y)\) and prices \((p)\). We firstly apply the Phillips-Ouliaris-Hansen procedure, and additionally, in order to better represent the d.g.p. among dividends and prices, and correct for possible autocorrelation in the errors, we test for cointegration via the ADL model methodology. Only if such a cointegration relationship is identified, a regression analysis between dividend and price levels imparts meaningful (and not spurious) information about their long-run equilibrium.

While both testing methodologies find strong cointegration evidence (at the 1% and 5% level, respectively) between summed dividend series \((d)\) and prices \((p)\), both find a weaker deterministic long run relation between (the not commonly used) reinvested dividends \((y)\) and prices. For this reason, construction of the modified dividend-price ratio centers on the relation between standard summed dividends \(d\), and \(p\). Having identified a strong cointegration relationship, we may proceed to regression analysis of the form \(d_t = \alpha + \beta p_t + z_t\), in order to construct the modified dividend-price ratio. Then the modified dividend-price ratio \((mdp_t)\) will be defined as the trend deviation from the established long-run equilibrium between dividends and prices

\[
mdp_t = d_t - \beta p_t
\]
The first construction of the modified $dp$ ratios is based on the *Dynamic Least Squares* methodology (DLS) by Stock and Watson (1993), and is similar to the formulation of the cay variable in Lettau and Ludvigston (2001). A second method to estimate the modified ratio, the *Autoregressive Distributed Lag* (ADL) method, allows for a better representation of the data generating process (d.g.p.), since it corrects for autocorrelation in the series’ shocks. In this case, the ratio is constructed based on the long run solution which is implied by the ADL model. Tables (AII) and (AIII) in the Appendix present the statistical results from cointegration analysis.

Both techniques also enable us to test (and reject) the hypothesis that there is a long run equilibrium under the [1,-1] specification of trend deviation. This test is a more powerful way to establish the non-stationarity of the classical $dp$, and thus deal with the low power of unit root tests against highly persistent alternatives. While we may proceed directly in testing with the ADL estimated parameters, with DLS the hypothesis can be tested by modifying the standard t-statistic. The Appendix B provides the necessary details on cointegration testing procedures and results based on these two methodologies (DLS, ADL). Here we provide the basic philosophy of the methods, and results about the estimation of the modified ratios.

The difficulty with directly estimating a cointegrating relation by $d_t = \alpha + \beta p_t + z_t$ is that we get non-standard distributions for hypothesis testing due to possibly non-zero correlations between $z_t$ and returns $\Delta p_t$. The DLS methodology is a univariate estimation procedure that corrects for possible endogeneity of the regressors, by adding leads and lags of the first difference of prices in the static regression (for more details see Hamilton, 1994: p.608 and

---

5 It is also well known that existing breaks will lower the power of unit root tests (Perron (1989)), thus making stationary processes with breaks difficult to distinguish from those including a unit root.
Campbell and Perron, 1991: p.189). More specifically, DLS employs the following model in order to estimate the trend deviation among log dividends and log prices

\[ d_t = \alpha + \beta p_t + \sum_{i=-k}^{k} b_i \Delta p_{t-i} + u_t \]  

(2)

where \( u_t \) by construction is uncorrelated with \( \Delta p_{t-i} \) for \( i = -k, -k + 1, \ldots k \). Estimation of (2) by OLS gives superconsistent estimators of the true cointegrating parameters (see rule 21 of Campbell and Perron, 1991). We then calculate the modified dividend-price ratio as,

\[ mdp_t = d_t - \hat{\beta} p_t \]  

(3)

In order to estimate (2), we need to specify the optimal lead and lag depth \( k \). In order to neutralize all possible seasonality in the data, we consider an initial depth of eleven leads and lags \( (k = 11) \). Then we condition down to a more parsimonious representation. This is done by removing pairs with an insignificant lead and/or lag, starting from the upper bound, until we get the first jointly significant lead-lag pair. Based on AIC criterion optimal \( k \) is 10.

Using the Philips-Ouliaris-Hansen procedure, we find strong cointegration evidence (details on Appendix) among annual sum dividends \( (d) \) and prices \( (p) \). Reinvested dividends \( (y) \) are too volatile, probably carrying an exogenous time growth that does not let a strong cointegrating relationship with the prices to emerge under standard assumptions for the dp ratio (such as the log dividend-price ratio does not include trend, Lettau and Ludvingston, 2005).

In order to perform hypotheses tests on the long run estimators derived by (2), we must correct for nuisance parameters such as the autocorrelation of residuals. Generally, the static
regression as estimated by (2) corrects for regressor endogeneity, but does not correct the residuals for possible autocorrelation. Thus, with DLS, the hypothesis that there is a long run equilibrium under the [1,-1] specification of trend deviation can be tested by modifying the standard t-statistic based on a procedure described in Hamilton (1994, p.610). We reject the hypothesis that there is long run equilibrium among annual sum of dividends and prices under the [1,−1] cointegration vector (see Appendix, Table AIII).

A different procedure in order to estimate the trend deviation among dividends and prices is the Autoregressive Distributed Lag (ADL) method. These models are an additional tool to univariate cointegration analysis, allowing for a richer d.g.p. that incorporates autocorrelation in the residuals. It has the advantage of an optimal specified model which we can carry standard hypothesis test on long run coefficients (the long run parameters are asymptotically normal, something not true for the statically estimated DLS cointegrating parameters, see rule 22 of Cambell and Perron, 1991: p.186).

We employ the following ADL model in order to estimate the long run solution between annual sum of dividends and prices,

\[ d_t = b_0 + \sum_{i=1}^{l} b_i d_{t-i} + \sum_{j=0}^{l} c_j p_{t-j} + \epsilon_t \]  

(4)

The true lag length is not known in (4). If we consider a more complicated structure, with more dynamic components, then it is possible to deal with the problem of multicollinearity. If we consider a more comprehensive representation (with fewer dynamic terms) it is likely that some residual autocorrelation noise still remains. Due to superconsistency, this finite sample bias does not pose a parameter estimation problem, but may be a concern for hypothesis
testing (see Inder, 1993). In order to estimate (4) for the dividend and price series, we consider three lags for both dividends and prices.

The optimal ADL model in (4), with \( l = 3 \), is then transformed to a long run estimated solution of the type

\[
d_t = \hat{\alpha} + \hat{\beta} p_t
\]

with \( \hat{\alpha} = \hat{b}_0 / (1 - \sum_{i=1}^{3} \hat{b}_i) \) and \( \hat{\beta} = \sum_{j=0}^{3} \hat{c}_j / (1 - \sum_{i=1}^{3} \hat{b}_i) \).

This transformation of an ADL model to its corresponding long run solution can only be accepted under the cointegration assumption between the series. The latter depends on a unit root test of the form \( H_0: (1 - \sum_{i=1}^{3} \hat{b}_i) = 0 \). If we accept the null, then clearly the model does not represent (and thus cannot be transformed to) a long run solution. Under the alternative hypothesis, using the \( \beta \) estimated by the ADL methodology (5) in (3) allows us to estimate a second modified ratio \( mdp' \). Any hypothesis imposed in the ADL-estimated long run parameters, such as the slope coefficient \( (\beta) \), can asymptotically be tested under standard \( t \)-statistic distribution. Similarly to DLS, we reject the null of \([1, -1]\) cointegration vector for (5). (see Appendix Table AII for cointegration results based on ADL)

The estimated modified ratios based on the above procedures are given by,

\[
DLS: mdp_t = d_t - 0.7811 p_t
\]

\[
ADL: mdp'_t = d_t - 0.8402 p_t
\]
II. The Nonstationarity of Dividend-Price Ratio

We can see from summary statistics presented in Table (1), that the reinvested dividend-price ratio has an autocorrelation of 0.94 (with $\phi=0.86$ for dp). Clearly, this is a local alternative that unit root tests have no power against. Furthermore, it is known, as early as Kendall (1954), that typical estimation methods will tend to highly underestimate true persistence in finite samples.\(^6\) In the last section, we presented robust econometric evidence against the stationarity of the classical dp. Not only is stationarity rejected via a straightforward ADF testing for the two dividend-price ratios ($dp$ and $yp$), but using the more powerful test of a restriction on the cointegration vector for d and p we reject the hypothesis that dividends and prices are linked with a long run relationship of the form $(d-p)$.\(^7\) Econometrically, dividend-price ratios are at best near non stationary processes.

By not de facto assuming an unreliable rejection of the non-stationarity null for dp, the modified ratios present a more reliable alternative, which allows for a richer representation of the d.g.p. Also, at $\phi=0.70$, mdp still has enough persistence in order to provide forecastability in long horizons. Before diving into a set of econometric tests that will undoubtedly establish the superiority of using our trend-corrected modified dividend-price ratio, in forecasting long-run returns, it is worth to first approach the economic ramifications of a non-stationary $dp$ from a qualitative point of view.

Miller and Modigliani (1961) offered the path breaking argument that dividend policy is irrelevant, and that stock prices should be driven by the “real” variable which is the earnings

---

\(^6\) Actually, even the Kendall bias correction for autocorrelation $- (1 + 3\phi)/T$ is low.

\(^7\) That the $[1,-1]$ vector spans the cointegration space.
power of corporate assets. Yet, dividends are a critical component of the total return an investor enjoys from her stock holdings, and the extant literature has empirically concluded that the dividend-price ratio can actually predict aggregate returns, but surprisingly has no information about future dividend growth. Thus, dividend yields have now become one of the most significant forecasting variables, and a widely accepted view is that almost all variation in dividend yields is driven by variation in discount rates.

The theoretical consideration of stationary (or at least bounded) \( dp \) is driven by a general modelling philosophy or fundamental way of thinking about the behaviour of valuation ratios and financial markets. According to this school, we should develop models with prices that may not move arbitrarily away from fundamentals. Thus, the financial ratios such as \( dp \) should not have to include any kinds of trends (either deterministic or stochastic). Generally speaking though, this is an economic requirement, which depends on a particular sample, rather than a hard fact.

Another asymptotic point of view, as Cochrane (1991, 2005) points out, is that the \( dp \) ratio is truly nonstationary only if there is a bubble. The possibility of a nonstationary \( dp \) is then technically ruled out by the assumption that in infinity, only the stationary component of the Campbell-Shiller identity (i.e. the \( \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} - \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} \) part) survives the process. Again, this hypothesis is econometrically un-testable for any finite horizon, and in turn this effect is actually a way to interpret the data, based on a finite sample, and extrapolate econometric findings to infinity. Actually, in a changing dividend-price ratio world, the CS identity is only an approximation, although a good one, and mathematically we are not allowed to take limits to infinity of an approximation. Actually, the error that a unit
root in dp might inject into the identity, is a strong argument in only using medium (5- or 7-year) horizons when deploying the CS-identity.

Econometrically, most researchers argue that the \( dp \) is a stationary process based on infinite sample or asymptotic arguments, and take \( dp \) stationarity as a given assumption. But neither the data sets that we actually use, nor the time horizons that we use to evaluate our models’ performance are infinite. More importantly, corporate managers have large discretion over payout options, and such discretion might impart unexpected structure into the dynamics of the dividend yield. The fact that, over any finite period of time, dividends (and dividend growth) can be arbitrary, and delinked from asset prices, means we should neither be dogmatic about the time series properties of the dividend yield nor about its inability to predict dividend growth. Yet, generally speaking, both academia and practice have avoided tackling head-on the possibility of nonstationary dynamics in valuation ratios such as the dividend-price ratio, despite the fact that the hypothesis of a unit root in long horizon samples cannot be statistically rejected. The economical source of such nonstationarity in dividend yields is not easily understood. It could be the result of changes in dividend policy such as dividend smoothing, use of share repurchases in lieu of cash payments, or it could be induced by other changes of investors’ attitudes toward dividends and taxes.

In any case, such changes in dividend policy will emerge in the data as a slope differential between dividends and prices. When we move away from dividend yield stationarity, assuming a deterministic long run equilibrium relation between dividends and prices of the form \( d = \alpha + \beta p \) is the next logical step still satisfying the fundamental asset pricing philosophy.
In this setup, $\beta$ gives a consistent estimate of the drift ratio between $d$ and $p$. Roughly speaking, a $\beta<1$ implies that dividends have been growing more slowly than prices. Having motivated the possibility for such a slope differential, and thus a non-stationary $dp$, the important question with respect to understanding the true dynamics of $dp$ is whether such nonstationarity is only due to a deterministic time trend or it includes a unit root. The problem is that, as is now well understood, this question is inherently unanswerable for any finite sample (see Blough 1992) since for any unit root process, and sample size $T$, there exists a stationary process that is indistinguishable. Another way to understand this issue is that the question of the inclusion of a unit root in the process is equivalent to finding whether the population spectrum at zero is zero or attains any positive value. This is clearly unanswerable, since in any sample there is no information about cycles of a period larger than the sample size. A realistic target for the financial economist should rather be to describe the data in a parsimonious way with low order autoregressions, since they are easier to estimate than high order moving average processes.

In the above long-run relation, we define the modified dividend-price ratio as the stationary cointegration error of this long-run equilibrium, $mdp = d - \beta p$. We may then think of $\beta$ as the unique population parameter that “fine tunes” $dp$ by revealing a (possibly) small but stationary trend deviation between dividends and prices. This modified ratio ($mdp$) is more informative than its non-stationary counterpart, the classical $dp$ ratio. Effectively, in our analysis, the classical $dp$ can be thought as the modified ratio, $mdp$, plus a noise term.

$$dp_t = mdp_t + (\beta - 1)p_t$$ (7)
Another method that may sometimes produce a good in sample fit to the data is to allow for occasional breaks to the levels or the slope of an otherwise stationary process (e.g. Fama and French (2002) consider a mean reverting dp within different regimes). Furthermore, it is well known that existing breaks will lower the power of unit root tests (Perron (1989)), thus making stationary processes with breaks difficult to distinguish from those including a unit root. But allowing for breaks that are impossible to predict ex ante has little value for return predictability and forecasting, and will produce a meagre out of sample $R^2$ (see Lettau and Van Nieuwerburgh (2008)). This is clearly not the case with our parsimonious approach that, as we show here, produces significant out-of-sample forecasting gains.

III. The Modified Ratio Dominates the Classical dp

In this section, we present the basic univariate forecasting regressions based on both the classical dividend-price ratios ($dp$ and $yp$), and the modified ratio ($mdp$) respectively.\textsuperscript{8} We formulate continuously compounded returns, equity premia, and dividend growth for 1, 3, 5 and 7-year horizons using monthly S&P 500 data.

In Table (2) we present the slope coefficients, t-statistics under the null of unpredictable series, and the coefficient of determination. Long horizon series is formulated by the use of a rolling window of overlapping monthly observations, and standard errors are GMM corrected based on the Hansen-Hodrick formula.

***insert Table 2 around here***

\textsuperscript{8} For completeness, we also present predictability results for the second modified ratio ($mdp'$) computed using the ADL methodology.
As we can see, for all ratios (classical and modified) as we increase the horizon moving from 1 to 7 years out, the slope coefficient and the coefficient of determination are increasing for returns. Long-horizon forecasts are the mechanical result of short horizon same-direction forecastability combined with a highly persistent forecasting variable. This is a well understood effect in the literature starting as early as Fama and French (1988), and explains how the persistence of a predictor variable leads to increased slope coefficient for longer horizons.

The new insight of this paper is that part of the high persistence of a non-stationary \( dp \) is due to the small embedded unit root in \( dp = mdp + (\beta - 1)p \). This extra persistence though, unlike the “useful” persistence in \( mdp \), carries no real predicting power. Thus, the true forecasting horizon is determined by the lower \( mdp \) persistence. The artificially longer horizon of \( dp \), that one gets by mechanically extending short period \( dp \) predictability into the distant future, is an artifact of the non-stationary noise embedded in \( dp \) and of no real forecasting value.

For all return horizons, modified ratios achieve impressive improvements in all three dimensions: slope size (b), significance (t-stats), and long-return explanatory power (R\(^2\)). Modified ratio performance strictly dominates classical ratios in all horizons, and furthermore, this modified ratio dominance gets more pronounced with an increasing horizon. For example, in forecasting returns five years out, and while classical slopes are about 0.40, the modified slopes have already attained their Cochrane\(^9\) “theoretical limit” of 1. When extending the forecast horizon to seven years, classical slopes have gone from 0.40 to 0.50,

---

\(^9\) Cochrane (2011) forcibly argues that all \( dp \) variability comes from expected return volatility, and none from dividend growth; “…What we expected to be zero is one; what we expected to be one is zero.”
still only half the size of their Cochrane limit. Furthermore, with a t-statistic as large as three
times the classical t-statistic, modified ratios explain an impressive 40% of the five year
future return, and an even more impressive 50% of the 7-year future return.

The strong performance of mdp in predicting future returns is considerably toned down when
using mdp in explaining equity premia. Actually, the performance of mdp in forecasting
equity premia is comparable to the performance of the classical ratios. Since total equity
return is composed of the risk free return plus the equity premium, we can intuitively deduce
that the performance of mdp in predicting future returns comes from its robust capacity in
predicting returns from money invested in risk free securities. Indeed as shown in Table 2b,
in all tested horizons, 1-,3-,5-, and 7-year risk free returns are forecasted by mdp but not dp.

***insert Table 2b around here***

It is important to economically explain its surprising ability to forecast future risk free returns
in Table 2b. We know that, given the high persistency of short term yields, T-bill returns are
highly forecastable. If interest rates (and hence one-year risk free returns) are currently low,
they are likely to remain low for the next years as well. The forecasting behaviour of
modified dp is driven by its strong positive correlation (0.42) with risk free returns, as
opposed to the near orthogonality of the classical dp (-0.10). Economically this could be
consistent with a *cliente*le explanation of dividend policy. If companies that consistently pay
dividends attract a certain type of yield seeking investor, then such companies can get away
with low dividend yields when such low payouts coincide with low current and future (due to
their high persistency) risk free yields. In such low-yield states of the economy, income
-seeking investors will not allocate their portfolios out of low dividend yield stocks because
they have nowhere to go. One way to understand why the classical ratios don’t share such foresting ability with their modified counterparts is if we view the modified dp as a *de-noised* dp. Even though this yield information is embedded in the classical ratio as well (as its stationary part), the I(1) “noise” component needs to be removed before such information can be harnessed for forecasting risk free returns.

Figure (1) plots the 5-year future returns against the *dp* and *mdp* ratios. As we can see, the *mdp* ratio explains long run returns better by removing both stochastic and deterministic trends from the classical *dp*. Additionally, the *mdp* captures better the business cycle variation in expected returns. In particular we note the surprising ability of mdp to avoid the excessively low dp print in the early 2000s. This happens because, in a world where some dividend policy trend (e.g. an increasing use of share repurchases) has induced non-stationarity in dp yields, mdp captures the true deviation from long run equilibrium between prices and fundamentals, by properly factoring out the non-stationarity inducing dynamic.

Neither the modified (*mdp*) nor the classical (*dp*) ratios can forecast dividend growth for any horizon. This is a first sign, similar to Cochrane’s results for the classical dp series, that the variability of the modified dp as well comes from its ability to forecast long run returns and not the dividend growth. To properly ground this discussion, we need to use the CS-relationship as follows next.
IV. The Informational Content of the Classical dp Ratio

In this section we use a multivariate analysis in order to investigate the sources of finite horizon dp variability in the presence of mdp. This analysis can be delivered by the use of the well known Campbell-Shiller (CS) relationship\(^\text{10}\),

\[
dp_t \approx \sum_{j=1}^{h} \rho_{j-1} r_{t+j} - \sum_{j=1}^{h} \rho_{j-1} \Delta d_{t+j} + \rho^h dp_{t+h}
\]  

(8)

The powerful feature of the CS relation is that it holds not only ex ante, as an expectation about the future, but is also valid ex-post on a path-by-path basis.

The dividend-price ratio completely determines this linear combination between long run returns, dividend growth and future dividend yields; i.e. by attending dp today we know completely the investors' expectations for this sum. As the relation holds ex post, no new information can disturb this sum, and inclusion of any extra information \(I_t\) about long run terms can only re-arrange these terms in a way that respects the CS-sum as it is calculated by the current dp,

\[
pd_t \approx E\left[\sum_{j=1}^{h} \rho_{j-1} \Delta d_{t+j} - \sum_{j=1}^{h} \rho_{j-1} r_{t+j} + \rho^h pd_{t+h} \mid I_t\right]
\]

(9)

For example, if new information in another variable forecasts an increased long run return it will either have to also forecast an increased long run dividend growth and/or higher future prices. Cochrane (2011) suggests taking regressions of long run terms on the dividend-price ratio in order to reveal the source of dp variation. As the horizon increases, the source of dp

\(^{10}\) In this section, to enhance readability, we denote the dividend-price ratio as \((dp_t)\), while we always run regressions based on \((yp_t)\) constructed for an annual horizon. This is because, in order to retain the CS identity in annual horizon, it’s necessary to use the reinvested dividends \((y)\) for dividend-price ratio and dividend growth construction (see Cochrane, 2011).
variation needs to correspond to fundamentals (i.e. returns and/or dividend growth). Cochrane argues that when looking as far as 15 years ahead, the variation of returns completely accounts for dp variation.

Unlike the bubble term $\rho^h dP_{t+h}$ that vanishes\(^{11}\) by being multiplied by $\rho^h$, when expanding the CS relation in a long horizon, the approximation error does not vanish, but rather successive errors get cumulated\(^{12}\) at decreasing weights. Thus, a highly persistent dp that deviates far from the expansion point in a particular sample may produce significant cumulative approximation residuals in the long horizon application of the formula. The theoretical properties of the error induced by the long horizon application of the identity remain largely unknown. Still, in economical terms consideration of the CS-relation is important and justified as it represents a robust way to quantify the basic forward looking pricing principle: high prices today that are not followed by robust growth will either lead to low future returns or be part of a rational bubble. Yet, the limited power in detecting near non-stationarity in dp, and the uncontrollable error that a unit root in dp might inject into the identity, is a strong argument that favors the use of limited (5- or 7-year) instead of infinite horizons when analyzing the performance of the CS-identity.

***insert Table 3 around here***

Table (3) shows the CS analysis of the total dp variation for 5 and 7 years ahead based on weighted forecasting regressions of future returns, dividend growth and the bubble

\(^{11}\) As long as dp doesn’t grow faster than 4% per annum.

\(^{12}\) This is already discussed in the original Campbell and Shiller (1989) paper. In order to evaluate the magnitude of the approximation error, Campbell and Shiller compare actual returns to the ones predicted by their approximation and find it small and almost constant. Yet, their empirical analysis is using data up to 1986, and most of the interesting behaviour of the dp ratio occurs around 2000.
component on dp. We use annual data that are constructed from the original monthly observations. One may in principle use the CS identity in monthly horizons, but would then have to deal with the problem of strong seasonality in dividend payments. As we can see, the variation of returns can account significantly for 35% and 46% of the total variability for 5 and 7 years ahead respectively. At these horizons the “bubble” component still has a strong effect on the dp roughly explaining another 60% and 46% of its variability respectively.

An important observation is here in order: If the CS identity was holding exactly, the three slope coefficients should sum up to one. The departure from the theoretical 100% limit is due to the approximation error in the CS sum, and the size of this deviation is a way to quantify the approximation error in the particular horizon for that path. For example, at a 7-year horizon, the slope sum is below 96%, implying a significant approximation error. Actually, the fact that the slope coefficients do not sum up to one is even more significant, because it is evidence that the CS error is not only sizeable but also exhibits significant variation and correlation with dp itself. Measuring the slope sum deviation from 1 is actually the proper way to measure the economic significance of the CS error. On the contrary, when measuring the modified slopes, in Table (4), the respective sum gets very close to zero and always insignificant. This is evidence that mdp, unlike dp, is actually un-correlated to the CS error.

***insert Table 4 around here***

In Table (4) we present the findings of a multivariate analysis with the modified ratio mdp placed along the classical dividend price ratio (yp). We see that for both horizons the modified dividend-price ratio completely drives out the dividend-price ratio. While in all

13 35 and 83 basis points for 5- and 7-year horizon respectively.
cases the modified ratio attains a return coefficient \((c_r)\) close to one and very significant, the dividend-price ratio slope \((b_r)\) is always insignificant and close to zero. This is a significant conceptual departure from the prevalent point of view that all dp variability explains variation in future long run returns, and needs to be carefully analyzed. More specifically, the contemporary doctrine is that when forming (sufficiently) long-horizon weighted predictive regressions all dp variability is explained by changes in long run returns; that is since \((\rho \phi)^h \rightarrow 0, b_r \rightarrow 1, b_d \rightarrow 0\). In a stationary dividend-price ratio world, this analysis then taken to a limit of a very long horizon implies that since\(^{14}\) \((c_r + c_d + c_e = 0)\), all an extra forecasting variable can do, when failing to predict dividend growth \((c_d = 0)\), is explain the so called term structure of long run returns.

Table (4) cannot be explained by the above analysis that has to be further refined as follows. The picture in Table (3) is that for medium term horizons, some of the dividend-price variation is driven by changes in long run returns and the rest is driven by the autocorrelation with the bubble term. When the modified ratio is added in the picture, in a multivariate setting, the superiority of its filtered information is so powerful that the dividend price ratio completely fails to bring any extra information about future returns to the table \((b_r = 0)\).

Thus, in the presence of the modified ratio the only thing left for the classical dividend-price ratio to predict is its forward looking value through its autocorrelation; i.e \(b_r = b_d = 0, b_e = 1\).

\(^{14}\)With \(c_r, c_d, c_e\), the mdp slope coefficients for returns, dividend growth and bubble factor \((\rho^h dp_{t+h})\) respectively.

25
The problem with the third slope coefficient \((b_e)\) is that, in a classical environment (i.e., stationary dividend-price ratio), this term will have to go to zero for long horizons (since single-period autocorrelation gets multiplied by \(\rho^h\)). But in a nonstationary dividend-price world as the one that we assume (and we have econometrically failed to reject), the picture is more complicated since the error in the CS identity is not bounded anymore. Since \(mdp\) is stationary by construction, if \((b_e)\) is to survive for long horizons it can only do so through the nonstationary “noise” in \(dp\) (i.e. \((\beta - 1)p_e\).) But if \(b_e = 1\) for long horizons, and since the only thing that survives in large horizons is the nonstationary I(1) component in \(dp\), we can conclude that the right economic picture for the long horizon weighted forecasting slope coefficients is that \((b_r = b_d = 0, b_e = 1)\) and \((c_r = 1, c_d = 0, c_e = -1)\). Then, the modified ratio captures economically all the fundamental variation of \(dp\) and leaves the “noise” part to explain the variation of \(dp\) which comes from the “bubble” component.

A critical issue is to understand where the enhanced ability of the modified ratios to predict long run returns comes from. One method to shed some light is to break down future returns on the risk free returns plus the equity premia \((r_t = re_t + rf_t)\). Long run returns from investing in T-bills should be much easier to predict than future equity premia, since it is well known that T-bill returns are predictable due to the slowly moving nature of yields. We run multivariate long run forecasting regressions based on the expanded CS identity,

\[
dp_t \approx E \left[ \sum_{j=1}^{k} \rho^{j-1}(re_{t+j} + rf_{t+j}) - \sum_{j=1}^{k} \rho^{j-1}\Delta d_{t+j} + \rho^k(d_{t+k} - p_{t+k}) \bigg| l_t \right] \quad (10)
\]

and present the findings in Table (5).
For 5-year horizon, we see that the ability of modified ratio $mdp_t$ ($mdp'_t$) to predict 86% (103%) of future returns is due to a 48% (60%) in capturing future equity premia and 38% (43%) in explaining T-bill returns. Not surprisingly, as the horizon gets longer to 7 years ahead, an even larger fraction equal to 50% (57%) of the impressive power of $mdp_t$ ($mdp'_t$) to still explain 91% (105%) of future returns is due to its power to forecast future T-bill returns. Even at 7-years ahead, the modified ratios can still explain 41% (48%) of total future equity premium variation.

***insert Table 4b around here***

V. Out-Of-Sample Evaluation

In this section, we evaluate the ability of modified dividend-price ratios to forecast Out-of-Sample (OS) returns and equity premia. The evaluation is done by comparing it against the forecasting ability of the “mean” benchmark for a real time investor. Campbell and Tompson (2008), who summarize the forecasting power for a pool of common financial and accounting variables, introduce the Out-of-Sample coefficient of determination via their $R^2_{\text{os}}$ statistic,

$$R^2_{\text{os}} = 1 - \left[ \frac{\sum_{k=1}^{T}(r_{t+k} - \bar{r}_{t+k})^2}{\sum_{k=1}^{T}(r_{t+k} - \bar{r}_{t+k})^2} \right]$$

(11)

This measures the effect of OS performance of a predictor variable against the mean benchmark. The OS coefficient of determination $R^2_{\text{os}}$ effectively asks if we could do a better forecasting job than someone who just expects that “…returns will always be the same”.

When compared with the squared Sharpe ratio, a positive $R^2_{\text{os}}$ directly measures the welfare benefits (for a mean-variance investor with a given risk aversion coefficient) of the increased portfolio returns by using one of the predictor variables.
We compare the results with the classical dp ratios in four forecasting horizons of one year, three years, five years and seven years ahead. We choose the estimation and forecasting periods based on the Welch and Goyal (2008) criteria, where it is necessary to have enough initial data in order to provide reliable OLS estimators and at the same time a large evaluation period for reliable OS appraisal. We divide the initial data sample in two periods, the “estimation period” (1926:1 to 1945:12) and the “evaluation period” (1946:1 to 2009:12).

To illustrate, Table (5) provides the results. As we can see, the classical dp ratios (both the standard annual sum dividends, and the ones with reinvesting) cannot provide positive $R_{OS}^2$ values, meaning that they fail to outperform the mean benchmark in all short to medium term horizons. To get the dp ratios to (marginally) outperform the “mean”, one needs to utilize a seven year long horizon. On the other hand, the modified ratios beat the “average” as fast as the one year ahead return offering a surprising 12% Out-of-Sample $R^2$ statistic. Thus, use of the mdp addresses a major weakness in dp, namely its presumed inability in revealing high to medium frequency (i.e. business cycle) variation in expected returns and equity risk-premia. As the investing horizon gets longer, the modified ratio $R_{OS}^2$ is increasing and reaches an astonishing 54% gain for seven years ahead. The statistical significance of $R_{OS}^2$ does not matter here because we believe that the sign is clear. Even a small $R_{OS}^2$ can provide great investment benefits for investors who otherwise thought that “…returns will be as they always have…” (see Campbell and Thompson 2008; Rapach et.al, 2010). Levels of OS
performance comparable to our modified ratios (i.e. with over 50% for a 7-year horizon) have not been achieved by any other forecasting indicator that we know of.

The cointegration coefficients for both modified ratios mdp and mdp', are estimated using the full sample. Given that the modified ratios have significant forecasting power, a concern is whether a practitioner operating in the early part of our sample and without access to the whole sample to estimate population cointegration coefficients, could have exploited such forecasting power to his advantage. This “look ahead” concern, when we try to examine the out-of-sample power of modified ratios, is well documented by Lettau and Ledvigson (2001) in the similar case of evaluating the performance of their cay variable. There is an inherent difficulty in addressing this issue, since subsample analysis (such as out-of-sample forecasting tests) entails a loss of information, and may fail to reveal the full forecasting ability measured with in-sample tests. For reasons explained in Lettau and Ludvingston (2001), the appropriate estimation strategy for measuring the full forecasting power of the modified ratios, is to use the full sample, because sufficiently large samples of data are necessary to recover the true cointegration coefficients. Since cointegration coefficients are super-consistent, converging to their true values at a rate proportional to the sample size T, they may be treated as population values during the second-stage forecasting regressions. On the other hand re-estimating the cointegration coefficient, each time using only data up to a certain point t, carries greater sampling errors, in contrast with the population coefficients based on the full sample, and puts the theory at a great disadvantage.

A robustness check proposed in Lettau and Ludvingston (2005) for the case of cay is that, instead of estimating the cointegration relation among d and p in a first-stage regression and
then using the trend deviation mdp as the single right-hand side variable, to consider single-equation, multivariate regressions of the form,

$$r_t(h) = a + b_1 d_t + b_2 p_t + u_t(h)$$

(12)

Under the null hypothesis that the left-hand-side variable is stationary, while the right-hand side variables are I(1) with a single cointegration relation, the limiting distributions for $b_1$ and $b_2$ will be standard, implying that the above forecasting regression will produce valid $R^2$ and t-statistics. Since this procedure does not require any first-stage estimation of cointegration parameters, it is clear that the forecasting $R^2$ statistics, are true indications of forecasting power.

The $R^2$ of the multivariate regressions on long run returns for one, three, five and seven years ahead are 6%, 23%, 41%, and 49% respectively, showing that modified ratios have true forecasting power and do not carry “forward looking” information by being estimated in a first stage, using data from the whole sample period.
VI. Conclusion

While dividends are a critical component of the total return an investor enjoys from her stock holdings, and dividend-price ratio can predict returns, extant literature has largely avoided tackling head-on the possibility of a nonstationary dividend-price ratio. After failing to reject the null of a unit root in the classical dividend-price ratio (dp), we assume away dividend yield stationarity, and show that a cointegrating relationship, not spanned by [1,-1], between dividends and prices exists. We estimate a relation of the type \( d = \alpha + \beta p \), and define the modified dp ratio as the stationary cointegration error of this long-run equilibrium. We think of \( \beta \) as the unique parameter that “fine tunes” dp, and reveals the stationary trend deviation between \( d \) and \( p \), by removing a possibly small I(1) “noise” component. We argue that, the error that a unit root in dp might inject into the CS-identity is a strong argument in using medium term (5- or 7-year) horizons in the analysis. Indeed, using S&P 500 data for the period 1926 to 2009, we show that mdp is more informative than its non-stationary counterpart, and provides substantially improved forecasting results over the classical dp ratio for medium horizons. We show that the gain of the modified ratio in forecasting returns is mainly due to its enhanced ability to forecast their risk free component. Out of sample, an investor who employs the modified dp ratio will enjoy a dramatic 52-54% improvement in forecasting 5- and 7-year returns. Further, with 1-year and 3-year \( R^2_{DS} \) of 12% and 39% respectively, our mdp addresses a major weakness in dp, namely its presumed inability in revealing business cycle variation in expected returns and equity risk-premia.
Appendix

A. Data Description

All series start in January of 1926 and extend to December of 2009. Capital letters (D and P) denote the level of the series, and lowercase letters (d and p respectively) denote the log of the series. When constructing the DP ratio, we need to employ an annual horizon in order to cancel the dividend seasonality. Depending on how one forms annual dividends at the end of month t, from the 12 preceding monthly dividends, the dividend price ratio may be computed with two different methodologies.

Throughout the paper, a subscripted time index \( t \) is used to express an annual horizon ending at the end of the \( t^{th} \) month. In the appendix, in order to avoid confusing monthly with annualized quantities ending at the same month, monthly quantities for the \( t^{th} \) month are always indexed with time in parenthesis. For example, using this notation, the ending at time \( t \) annual dividend \( D_t \) as a simple sum of the 12 preceding monthly dividends is written as

\[
D_t = \sum_{i=0}^{11} D(t - i)
\]

The most common annual dividend-price ratio is based on the following computation

\[
DP_t = \frac{D_t}{P_t} = \frac{\sum_{i=0}^{11} D(t - i)}{P_t}
\]

When we are endowed with monthly gross returns with and without dividends, \( R(t) = \frac{p_t + D(t)}{p_{t-1}} \), and \( X(t) = P_t / P_{t-1} \) respectively, the monthly dividend for month \( t \) is also given by
Another method that takes into consideration investing opportunities for interim dividends, is then to form a dividend-price ratio \((YP_t = Y_t/P_t)\), where the annual dividend \(Y_t\) of the 12-month period ending at time \(t\) is formed from the immediate reinvestment of the interim preceding monthly dividends \(D(t-1), D(t-2), \ldots, D(t-11)\) at the prevailing end-of-month market prices.

\[
Y_t = \left( \prod_{i=0}^{11} \frac{R(t-i)}{X(t-i)} - 1 \right) P_t
\]

Again, if we are careful in differentiating between annual gross returns with and without dividends \(R_t = \prod_{i=0}^{11} R(t-i)\) and \(X_t = \prod_{i=0}^{11} X(t-i)\) from their monthly counterparts \(R(t)\), and \(X(t)\) respectively, we may directly calculate this new dividend-price ratio with reinvested dividends as

\[
YP_t = \frac{R_t}{X_t} - 1
\]

Additionally, we formulate the true (i.e. de-seasonalized) dividend growth series by using the YP series through

\[
\frac{Y_{t+1}}{Y_t} = X(t+1) \frac{YP_{t+1}}{YP_t}
\]
B. Econometric Analysis

In this section, we provide the basic econometric analysis of the series containing the unit root and cointegration test results. Table AI provides the results for the unit root analysis for S&P 500 dividends and prices based on the Ng-Perron test. When compared with other unit-root testing methodologies, this test is characterized by good power and size properties. Most well known unit root tests (like ADF), have many serious power and size distortions for long samples with many possible structural breaks in the series. Our data consist of the log price levels \((p_t)\), the log reinvested dividends \((y_t)\) and the log 12 month moving sum dividends \((d_t)\). For more details on the test statistics of the Ng-Perron procedure see (Ng and Perron, 2001). The optimum lag selection for the testing regressions is based on the modified AIC criterion with 11 maximum initial lags, while we consider both trend and constant as deterministic components.

***insert Table AI around here***

Having determined the unit root nature for all \((d, y, p)\) series in Table AI, we then proceed to test for the existence of deterministic cointegration between both types of dividends \((d\ or\ y)\) and prices \((p)\). We firstly apply the Phillips-Ouliaris-Hansen procedure and the results are shown in part (a) of Table AII. Additionally, in order to correct for possible autocorrelation in the errors, we test for cointegration via the ADL model methodology as has already been discussed in the text. ADL results are presented in part (b) of Table AII. While both testing methodologies find cointegration between summed dividend series \((d)\) and prices \((p)\), both fail to find a deterministic long run relation between reinvested dividends \((y)\) and prices.
In the final Table AIII, we provide results of testing $dp$ stationarity via two methods: the first based on the straightforward ADF testing for the two dividend-price ratios ($dp$ and $yp$), and the second as a more powerful test of the $[1, -1]$ restriction on the cointegration vector for d and p. We can accept the ADF null that the classical ratios are nonstationary, and reject the restriction that the long run relationship is of the form $[1, -1]$.

The first methodology is the direct unit root testing procedure by ADF in the dividend-price ratios considering only a constant in the testing regression\textsuperscript{15}. \textit{A priori}, one could proceed in ADF testing for $dp$ (and $yp$) that includes a trend. \textit{A posteriori}, having seen, in Table AII, that $d$ and $p$ are deterministically cointegrated, accepting that $dp$ includes a trend is a less informative hypothesis and should thus be avoided per the Pantula principle. The first panel in Table AIII shows the results from the ADF test on the classical $dp$ and $yp$ ratios, considering a constant and no trend on the testing regression. The lag depth is the optimal for autocorellation based on the AIC criterion. As shown in AIII, the non-stationarity null is not rejected for both classical ratios.

Having accepted the null of cointegration between summed dividends (d) and prices (p), we can provide a more powerful testing for $dp$ stationarity: a testing sequence of the null hypothesis that the estimated cointegration vector is of form $[1, -1]$. The second panel of AIII shows the results under the restriction that the $[1, -1]$ vector spans the cointegration

\textsuperscript{15} The Pantula principle postulates the acceptance of the most restrictive null. Directly introducing a trend in dp implies \textit{stochastic} cointegration, and is less preferable than testing for deterministic cointegration between d and p.
space of d and p.\textsuperscript{16} If \([1,-1]\) is a cointegration vector between dividends and prices, there would be no reason to estimate a general cointegration relationship of the form \([1,-\beta]\). Thus, in Table AIII, we provide the evidence that the vector \([1,-1]\) does not span the cointegration space that characterizes the long run relationship between the series.

When using DLS to estimate the long-run \(\beta\) we need to parameterize the usual t-statistic in order to take into account possible serial correlation in the residuals of the DLS regression between dividends and prices, and then find that the restriction \([1,-1]\) is rejected in the regression coefficients (for further discussion see Hamilton, 1994 p.610). Specifically, we estimate the DLS regression and fit an AR(q) model to the residuals, where \((q)\) is the optimum lag length based on Campbell and Perron rule. Then, the modified t-statistic is given by,

\[
mod\_t\_stat = t\_stat \times \left(\frac{s}{\lambda}\right)
\]

Where \((s)\) is the standard error of the DLS regression and \((\lambda)\) is a modification factor which is a function of the autoregressive coefficients, and the standard error from the AR fitted regression to the DLS residuals.

Finally, we provide test results of the \([1,-1]\) restriction for the long run solution of the ADL estimator, for which t-statistics follow the standard normal distribution and do not need to be corrected. The results are given in Table AIII, and as we can see the \([1, -1]\) null is rejected for both estimation methodologies.

\textsuperscript{16} We don’t do the same for yp, since no cointegration is found between y and p as is shown in Table AII.
References


Table 1: Summary Statistics

We present the summary statistics for annual returns, equity premia, risk free rates, dividend-price ratios and modified dividend-price ratios. The table shows the correlation matrix among the series as well as the mean, standard deviation and the autocorrelation coefficient based on AR(1) fitted model. Data are annual from 1926 to 2009.

<table>
<thead>
<tr>
<th></th>
<th>$r_t$</th>
<th>$re_t$</th>
<th>$rf_t$</th>
<th>$yp_t$</th>
<th>$dp_t$</th>
<th>$mdp_t$</th>
<th>$mdp_t'$</th>
<th>Mean</th>
<th>Std</th>
<th>AR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_t$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$re_t$</td>
<td>0.99</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$rf_t$</td>
<td>0.04</td>
<td>-0.11</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$yp_t$</td>
<td>-0.03</td>
<td>-0.01</td>
<td>-0.11</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$dp_t$</td>
<td>-0.25</td>
<td>-0.24</td>
<td>-0.10</td>
<td>0.97</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$mdp_t$</td>
<td>-0.34</td>
<td>-0.40</td>
<td>0.42</td>
<td>0.54</td>
<td>0.61</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$mdp_t'$</td>
<td>-0.34</td>
<td>-0.37</td>
<td>0.24</td>
<td>0.78</td>
<td>0.84</td>
<td>0.94</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Predictability of returns, equity premia and dividend growth

This table presents the results of return, equity premium and dividend growth predictability for S&P 500 based on the following forecasting regression,

\[ g_t(h) = \sum_{j=1}^{h} g_{t+12j} = a + bx_t + u_t(h) \]

The left hand variable is the time-\( t \) future log return (\( r \)), equity premium (\( re \)) or dividend growth (\( \Delta d \)) for one, three, five and seven years ahead (\( h = 1,3,5,7 \)). The predictor variable represents either one of the classical dp ratios (\( dp \) and \( yp \)), or one of the modified ratios (\( mdp \) and \( mdp' \)). We use overlapping monthly data in order to formulate the corresponding series for horizons greater than one month, and thus standard errors are GMM corrected based on the Hansen-Hodrick formula. Data are annualized constructed from monthly observations with an overlapping rolling window from 1926 to 2009.
<table>
<thead>
<tr>
<th>$r_i(1)$</th>
<th>b</th>
<th>t(b)</th>
<th>$R^2$</th>
<th>$r_i(3)$</th>
<th>b</th>
<th>t(b)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>yp$_i$</td>
<td>0.11</td>
<td>2.17</td>
<td>0.05</td>
<td>0.26</td>
<td>3.18</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>dp$_i$</td>
<td>0.10</td>
<td>1.63</td>
<td>0.04</td>
<td>0.25</td>
<td>2.94</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>mdp$_i$</td>
<td>0.21</td>
<td>2.71</td>
<td>0.06</td>
<td>0.66</td>
<td>4.38</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>mdp'$_i$</td>
<td>0.19</td>
<td>2.56</td>
<td>0.06</td>
<td>0.58</td>
<td>5.75</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>$r_i(5)$</td>
<td></td>
<td></td>
<td></td>
<td>$r_i(7)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yp$_i$</td>
<td>0.39</td>
<td>3.58</td>
<td>0.14</td>
<td>0.52</td>
<td>3.44</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>dp$_i$</td>
<td>0.41</td>
<td>3.81</td>
<td>0.17</td>
<td>0.50</td>
<td>3.25</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>mdp$_i$</td>
<td>1.05</td>
<td>7.92</td>
<td>0.41</td>
<td>1.17</td>
<td>10.47</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>mdp'$_i$</td>
<td>0.92</td>
<td>8.54</td>
<td>0.37</td>
<td>1.06</td>
<td>10.42</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>re$_c(1)$</td>
<td></td>
<td></td>
<td></td>
<td>re$_c(3)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yp$_i$</td>
<td>0.12</td>
<td>2.30</td>
<td>0.05</td>
<td>0.28</td>
<td>2.97</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>dp$_i$</td>
<td>0.10</td>
<td>1.79</td>
<td>0.04</td>
<td>0.28</td>
<td>3.01</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>mdp$_i$</td>
<td>0.16</td>
<td>1.96</td>
<td>0.04</td>
<td>0.51</td>
<td>2.71</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>mdp'$_i$</td>
<td>0.17</td>
<td>2.19</td>
<td>0.05</td>
<td>0.50</td>
<td>3.72</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>re$_c(5)$</td>
<td></td>
<td></td>
<td></td>
<td>re$_c(7)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yp$_i$</td>
<td>0.42</td>
<td>3.03</td>
<td>0.17</td>
<td>0.58</td>
<td>2.97</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>dp$_i$</td>
<td>0.45</td>
<td>3.34</td>
<td>0.20</td>
<td>0.57</td>
<td>3.06</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>mdp$_i$</td>
<td>0.80</td>
<td>3.71</td>
<td>0.25</td>
<td>0.85</td>
<td>3.98</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>mdp'$_i$</td>
<td>0.79</td>
<td>4.58</td>
<td>0.27</td>
<td>0.90</td>
<td>5.00</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>$\Delta$d$_{(1)}$</td>
<td></td>
<td></td>
<td></td>
<td>$\Delta$d$_{(3)}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yp$_i$</td>
<td>0.00</td>
<td>0.03</td>
<td>0.00</td>
<td>-0.02</td>
<td>-0.23</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>dp$_i$</td>
<td>0.03</td>
<td>0.68</td>
<td>0.00</td>
<td>-0.00</td>
<td>-0.01</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>mdp$_i$</td>
<td>0.09</td>
<td>1.29</td>
<td>0.02</td>
<td>0.08</td>
<td>0.71</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>mdp'$_i$</td>
<td>0.07</td>
<td>1.20</td>
<td>0.01</td>
<td>0.05</td>
<td>0.45</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>$\Delta$d$_{(5)}$</td>
<td></td>
<td></td>
<td></td>
<td>$\Delta$d$_{(7)}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yp$_i$</td>
<td>-0.03</td>
<td>-0.38</td>
<td>0.00</td>
<td>-0.04</td>
<td>-0.36</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>dp$_i$</td>
<td>0.02</td>
<td>0.24</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>mdp$_i$</td>
<td>0.20</td>
<td>1.20</td>
<td>0.03</td>
<td>0.14</td>
<td>0.99</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>mdp'$_i$</td>
<td>0.14</td>
<td>1.06</td>
<td>0.01</td>
<td>0.09</td>
<td>0.78</td>
<td>0.01</td>
<td></td>
</tr>
</tbody>
</table>
Table 2b: Univariate forecasting of long run risk free rates

We run univariate regressions among long run risk free rates, $rf_t(h) = \sum_{j=1}^{h} rf_{t+12j}$, with the competing dividend-price ratios ($yp_t$, $mdp_t$) as regressors. Data are annualized constructed from monthly observations with an overlapping rolling window from 1926 to 2009. Standard errors are GMM corrected based on the Hansen-Hodrick formula.

<table>
<thead>
<tr>
<th>$rf_t$</th>
<th>$yp_t$</th>
<th>$mdp_t$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(1)$</td>
<td>-0.01</td>
<td>-0.69</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>2.62</td>
<td>0.17</td>
</tr>
<tr>
<td>$(3)$</td>
<td>-0.02</td>
<td>-0.46</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>2.11</td>
<td>0.20</td>
</tr>
<tr>
<td>$(5)$</td>
<td>-0.04</td>
<td>-0.46</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>0.24</td>
<td>2.14</td>
<td>0.20</td>
</tr>
<tr>
<td>$(7)$</td>
<td>-0.07</td>
<td>-0.57</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>0.32</td>
<td>2.18</td>
<td>0.18</td>
</tr>
</tbody>
</table>
Table 3: Univariate analysis based on CS identity: Variability of dp

We run weighted forecasting regressions based on the following univariate model,

\[ wg_t(h) = \sum_{j=1}^{h} \rho^{j-1} g_{t+12j} = a + by_{t} + u_t(h) \]

The left hand variable represents the weighted long-run returns or dividend growth, or the future dividend-price ratio for five and seven years ahead. Data are annualized constructed from monthly observations with an overlapping rolling window from 1926 to 2009. Standard errors are GMM corrected based on the Hansen-Hodrick formula.

<table>
<thead>
<tr>
<th>wr(5) b</th>
<th>t(b)</th>
<th>R^2</th>
<th>wr(7) b</th>
<th>t(b)</th>
<th>R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.35</td>
<td>3.43</td>
<td>0.14</td>
<td>0.46</td>
<td>3.34</td>
<td>0.22</td>
</tr>
<tr>
<td>wΔy(5)</td>
<td></td>
<td></td>
<td>wΔy(7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.02</td>
<td>-0.39</td>
<td>0.00</td>
<td>-0.04</td>
<td>-0.43</td>
<td>0.00</td>
</tr>
<tr>
<td>(0.965) * y_{t+5}</td>
<td></td>
<td></td>
<td>(0.967) * y_{t+7}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.60</td>
<td>10.11</td>
<td>0.48</td>
<td>0.46</td>
<td>8.47</td>
<td>0.33</td>
</tr>
</tbody>
</table>
Table 4: Multivariate analysis based on CS identity

We run weighted forecasting regressions based on the following multivariate model,

\[
wg_t(h) = \sum_{j=1}^{h} \rho^{j-1} g_{t+12j} = a + b y_{p_t} + c z_{t} + u_t(h)
\]

Where \((z_t)\) represents the modified ratio \(mdp_t\) (or \(mdp_t')\) estimated through the DLS (or ADL respectively) methodology. The left hand variable represents either the weighted returns, or dividend growth, or the long run dividend-price ratio for a horizon \(h\) of five or seven years ahead. Data are annualized constructed from monthly observations with an overlapping rolling window from 1926 to 2009. Standard errors are GMM corrected based on the Hansen-Hodrick formula.

<table>
<thead>
<tr>
<th>(wr(5))</th>
<th>slope</th>
<th>t-stat</th>
<th>(R^2)</th>
<th>(w\Delta y(5))</th>
<th>slope</th>
<th>t-stat</th>
<th>(R^2)</th>
<th>((0.96^5)*y_{p+5})</th>
<th>slope</th>
<th>t-stat</th>
<th>(R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y_{p_t})</td>
<td>0.02</td>
<td>0.15</td>
<td>0.37</td>
<td>-0.11</td>
<td>-0.85</td>
<td>0.03</td>
<td>0.83</td>
<td>11.24</td>
<td>0.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(mdp_t)</td>
<td>0.86</td>
<td>4.01</td>
<td>0.24</td>
<td>1.07</td>
<td>0.04</td>
<td>0.99</td>
<td>9.64</td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(y_{p_t})</td>
<td>-0.22</td>
<td>-1.01</td>
<td>0.35</td>
<td>-0.20</td>
<td>-1.03</td>
<td>0.04</td>
<td>-0.71</td>
<td>-5.14</td>
<td>0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(mdp_t')</td>
<td>1.03</td>
<td>4.03</td>
<td>0.32</td>
<td>1.18</td>
<td>0.02</td>
<td>0.74</td>
<td>9.82</td>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(wr(7))</th>
<th>(w\Delta y(7))</th>
<th>((0.96^7)*y_{p+7})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y_{p_t})</td>
<td>0.10</td>
<td>0.58</td>
</tr>
<tr>
<td>(mdp_t)</td>
<td>0.91</td>
<td>4.47</td>
</tr>
<tr>
<td>(y_{p_t})</td>
<td>-0.13</td>
<td>-0.55</td>
</tr>
<tr>
<td>(mdp_t')</td>
<td>1.05</td>
<td>4.34</td>
</tr>
</tbody>
</table>
Table 4b: Multivariate slope breakdown analysis based on CS identity

We run weighted forecasting regressions based on the following multivariate model,

$$ wr_t(h) = \sum_{j=1}^{h} \rho^{j-1} g_{t+j} = a + b y_p_t + c z_t + u_t(h) $$

where $z_t$ represents one of the modified ratios ($mdp_t$ or $mdp'_t$) respectively. The left hand variable represents either the weighted equity premia or risk free returns for a horizon $h$ of five or seven years ahead. Standard errors are GMM corrected based on the Hansen-Hodrick formula. Data are annual from 1926 to 2009.

<table>
<thead>
<tr>
<th>$wre_t(5)$</th>
<th>slope</th>
<th>$wre_t(7)$</th>
<th>slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_p_t$</td>
<td>0.20</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>$mdp_t$</td>
<td>0.48</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>$y_p_t$</td>
<td>0.05</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>$mdp'_t$</td>
<td>0.60</td>
<td>0.48</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$wrf_t(5)$</th>
<th>$wrf_t(7)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_p_t$</td>
<td>-0.18</td>
</tr>
<tr>
<td>$mdp_t$</td>
<td>0.38</td>
</tr>
<tr>
<td>$y_p_t$</td>
<td>-0.27</td>
</tr>
<tr>
<td>$mdp'_t$</td>
<td>0.43</td>
</tr>
</tbody>
</table>
We examine the OS forecasting ability of classical and modified dp ratios. We present the Campbell-Thomson OS coefficient of determination, for predicting returns and equity premia for annual, 3-, 5- and 7-year horizons. Initially, we utilize a 20-year minimal estimation period (1926-1946). The remaining sample, which extends beyond the estimation period, constitutes the evaluation period (initially 1946-2009) respectively. We denote with double prime ("") predictors that have been found In-Sample insignificant for the given horizon. Data are annualized constructed from monthly observations with an overlapping rolling window from 1926 to 2009.

<table>
<thead>
<tr>
<th></th>
<th>( r_t(1) )</th>
<th>( R^2_{os} )</th>
<th>( r_t(3) )</th>
<th>( R^2_{os} )</th>
<th>( r_t(5) )</th>
<th>( R^2_{os} )</th>
<th>( r_t(7) )</th>
<th>( R^2_{os} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( yp_t )</td>
<td>-0.056</td>
<td>-0.438</td>
<td>-0.294</td>
<td>0.014</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( dp_t )</td>
<td>0.009(&quot;&quot;)</td>
<td>-0.190</td>
<td>-0.214</td>
<td>0.063</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( mdp_t )</td>
<td>0.119</td>
<td>0.388</td>
<td>0.520</td>
<td>0.538</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( mdp'_t )</td>
<td>0.112</td>
<td>0.288</td>
<td>0.385</td>
<td>0.477</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>( r_e(1) )</th>
<th>( r_e(3) )</th>
<th>( r_e(5) )</th>
<th>( r_e(7) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( yp_t )</td>
<td>-0.047</td>
<td>-0.421</td>
<td>-0.259</td>
<td>0.074</td>
</tr>
<tr>
<td>( dp_t )</td>
<td>0.021(&quot;&quot;)</td>
<td>-0.161</td>
<td>-0.180</td>
<td>0.117</td>
</tr>
<tr>
<td>( mdp_t )</td>
<td>0.058</td>
<td>0.163</td>
<td>0.205</td>
<td>0.212</td>
</tr>
<tr>
<td>( mdp'_t )</td>
<td>0.076</td>
<td>0.171</td>
<td>0.236</td>
<td>0.311</td>
</tr>
</tbody>
</table>
Table AI: Univariate Ng & Perron unit root test for S&P 500 dividend and price series

This table shows the results from the Ng and Perron (2001) unit root testing of the price and dividend series. We also test for unit roots in the first difference series in order to verify the I(1) assumption for the data. All series are I(1). A (**) denotes rejection of non-stationarity at the 1% rejection level. Data are overlapping annual spanning the period, 1926:1-2009:12.

<table>
<thead>
<tr>
<th>Ng&amp;Perron_stats</th>
<th>p_t</th>
<th>d_t</th>
<th>y_t</th>
</tr>
</thead>
<tbody>
<tr>
<td>MZA</td>
<td>-6.37</td>
<td>-10.13</td>
<td>-4.77</td>
</tr>
<tr>
<td>MZt</td>
<td>-1.77</td>
<td>-2.24</td>
<td>-1.50</td>
</tr>
<tr>
<td>MSB</td>
<td>0.28</td>
<td>0.22</td>
<td>0.31</td>
</tr>
<tr>
<td>MPT</td>
<td>14.31</td>
<td>9.05</td>
<td>18.83</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Δp_t</th>
<th>Δd_t</th>
<th>Δy_t</th>
</tr>
</thead>
<tbody>
<tr>
<td>MZA</td>
<td>-19.20**</td>
<td>-25.50**</td>
<td>-630.93**</td>
</tr>
<tr>
<td>MZt</td>
<td>-3.09**</td>
<td>-3.48**</td>
<td>-17.76**</td>
</tr>
<tr>
<td>MSB</td>
<td>0.16**</td>
<td>0.14**</td>
<td>0.03**</td>
</tr>
<tr>
<td>MPT</td>
<td>1.30**</td>
<td>1.27**</td>
<td>0.04**</td>
</tr>
</tbody>
</table>
We apply the P-O-H residual testing procedure assuming trending series and no trend on cointegration relationship. The pair \((d_t, p_t)\) tests for a cointegration relationship between the 12 month moving sum dividends \((d)\) and prices, while \((y_t, p_t)\) tests the cointegration relationship between the reinvested dividend series \((y)\) and prices. All data are in logs. The second panel gives the cointegration results based on a dynamic procedure (ADL) as described in the text. A \(*, **\) denotes the 5% and 1% rejection level respectively. The critical values for the ADL test are based on (Ericsson and MacKinnon, 2002). Data are overlapping annual spanning the period, 1926-2009.

<table>
<thead>
<tr>
<th>coint.pairs</th>
<th>POH_test</th>
<th>Lags</th>
<th>ADL_test</th>
<th>Lags</th>
</tr>
</thead>
<tbody>
<tr>
<td>((d_t, p_t))</td>
<td>-4.30**</td>
<td>9</td>
<td>-3.34*</td>
<td>3</td>
</tr>
<tr>
<td>((y_t, p_t))</td>
<td>-2.93</td>
<td>10</td>
<td>-1.84</td>
<td>3</td>
</tr>
</tbody>
</table>
Table AIII: Testing the stationarity of the classical dp

We test dp stationarity via two methods: the first based on the straightforward ADF testing for the two dividend-price ratios (dp and yp), and the second as a more powerful test of the [1,-1] restriction on the cointegration vector for d and p. The first panel shows the results from the ADF test on the classical dp and yp ratios, considering a constant and no trend on the testing regression. The lag depth is the optimal for autocorrelation based on the AIC criterion. The second panel shows the results under the restriction that the [1,-1] vector spans the cointegration space of d and p, since no cointegration is found between y and p as is shown in Table AII. We firstly show results for the modified t-test as described in Hamilton (1994, p.610). Additionally, we provide the results of the usual t-test for the [1,-1] assumption in the long run solution derived by the ADL model. (** ) denotes the 1% rejection level. Data are overlapping annual spanning the period, 1926-2009.

<table>
<thead>
<tr>
<th>coint.pairs</th>
<th>ADF_test</th>
<th>Lags</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d_t, p_t)</td>
<td>-2.45</td>
<td>5</td>
</tr>
<tr>
<td>(y_t, p_t)</td>
<td>-2.19</td>
<td>8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>coint.pairs</th>
<th>mod_t_stat</th>
<th>ADL (d_t, p_t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d_t, p_t)</td>
<td>-3.97**</td>
<td>-3.11**</td>
</tr>
</tbody>
</table>
Figure 1: Future realized long run returns plotted along dp and mdp ratios. $r_t(5)$ represents the 5-year realized return, and $dp_t, mdp_t$ the classical and modified ratios respectively. Both series are plotted on the same time horizon. Data are annual spanning the period 1926-2009.